Andrey Nikulin
Núcleo Takinahakỹ de Formação Superior Indígena — Federal University of Goiás; nikulin@ufg.br

Lexical evidence for the Macro-Jê–Tupian hypothesis

The Macro-Jê and Tupian language families of Eastern South America have long been thought to be distantly related, mainly based on morphological evidence. This article assembles lexical evidence for the Macro-Jê–Tupian hypothesis. Reconstructed Proto-Macro-Jê and Proto-Tupian forms are compared, with special attention to the distribution of the etyma in each family, morphosyntactic behavior of the comparanda, and semantic and phonological plausibility of the proposed etymologies. Although the total number of possible cognates is very limited, the fact that they show recurrent sound correspondences renders the Macro-Jê–Tupian hypothesis promising and worthy of further research.

Keywords: Macro-Jê; Tupian; comparative method; South American indigenous languages.

The goal of this contribution is to present lexical evidence for the hypothesis whereby the Macro-Jê and Tupian languages are considered to be distantly related. Macro-Jê and Tupian are two major language families of Eastern South America, whose geographic spread coincides to a great extent. Both are present south of the Amazon River in what is now Brazil and Eastern Bolivia. Northern Argentina, Paraguay, and (formerly) Uruguay are home to a few peoples that speak Tupian languages of the Guaranian branch, though in the past two Macro-Jê languages—Ingain and Kaingang—were spoken there, too. In addition, due to post-Columbian migrations a few Tupian languages—Wajãpi, Teko, Zo'e, and Nheengatu—, are now spoken north of the Amazon River in French Guiana, Brazil, and Venezuela.

Sections **1** and **2** present the Macro-Jê and Tupian families, respectively, with an emphasis on the state-of-the-art reconstructions of the respective protolanguages. Section **3** surveys the extant scholarship on the Macro-Jê–Tupian hypothesis. The potential cognates are discussed in section **4**, and the respective sound correspondences are dealt with in section **5**. Section **6** concludes the paper. The provenance of linguistic data is indicated at the end of the paper, before the list of abbreviations and the acknowledgments.

Throughout this article, I employ the Macro-Jê Alphabet (Nikulin 2020: 50–53) for reconstructed forms of Proto-Macro-Jê and other Macro-Jê (proto)languages that lack an established practical orthography. For Proto-Tupian and Proto-Cariban, the International Phonetic Alphabet is used, except that */ɾ/, */β̞/, */ɛ/ are written as **r*, **β*, **e*. Practical orthography is used for contemporary or historically attested languages, when possible. In Mundurukú, Yudja, and Mondé forms, tone is indicated despite being unmarked in the respective practical orthographies (ˊ for high tone, ˋ for creaky voice; cf. Pinheiro et al. 2020); in Tuparí, the stress position is likewise indicated by means of an acute accent.

1. Macro-Jê

Approximately 35 languages are classified as Macro-Jê, though only 12 of these (Karajá, Maxakalí, Laklãnõ, Kaingang, Akwẽ-Xerente, Xavante, Panará, Pykobjê–Krĩkatí, Canela–Krahô, Apinajé, Mẽbêngôkre, and Khĩsêtjê) currently serve as the main means of communication in the respective communities. Their classification is shown in Figure 1. Note that **Southern Ka-**

makã is a cluster composed of three dialects or closely related languages (Menien, Kotoxó = Mongoyó, and Kamakã proper); **Core Maxakalian** includes at least six varieties (Maxakalí = Tikmũ'ũn, Ritual Maxakalí, Makoní, Pataxó, Pataxó-Hãhãhãe, and Koropó); **Southern Jê** includes two languages, Kaingang (with its at least five dialects) and Laklãnõ = Xokleng; **Akuwẽ** includes four languages (Xavante, Akwẽ-Xerente, Xakriabá = Krẽká, and Akroá); **Northern Jê** includes ca. 7 languages (Kajkwakhrattxi, Khĩsêtjê, Mẽbêngôkre with its two extant dialects, Apinajé, Parkatêjê, Pykobjê–Krĩkatí with its two dialects, and Canela–Krahô with its no less than three dialects); **Karajá** has four dialects (Southern Iny, Northern Iny, Javaé, and Xambioá = Ixỹbiòwa); **Chiquitano** is composed of three dialects, or maybe three closely related languages (Bésɨro, Migueleño, and Eastern).

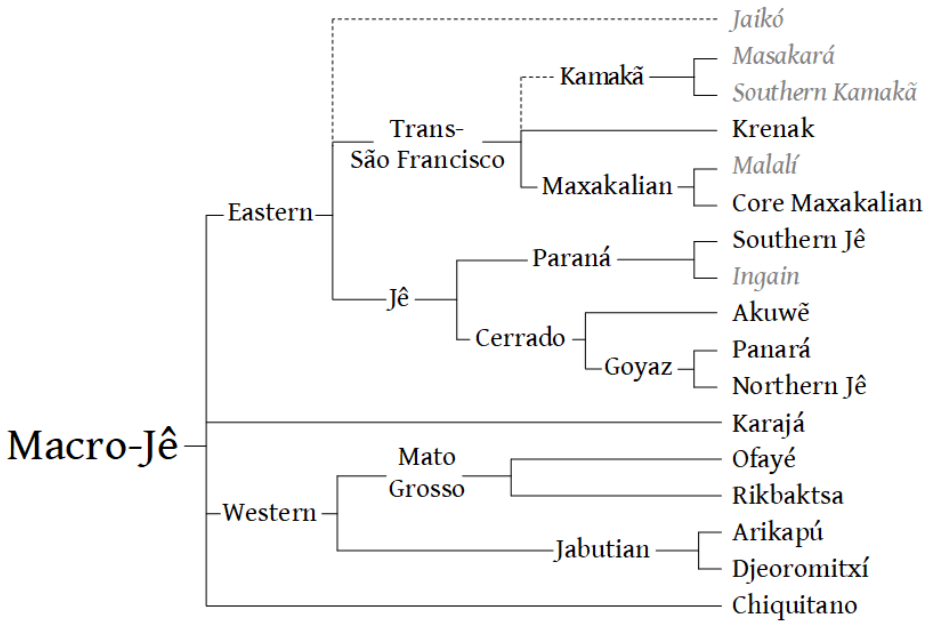

Figure 1. Macro-Jê *Stammbaum* (adapted from Nikulin 2020: 178)[1]

The only extant study that deals with the reconstruction of Proto-Macro-Jê phonology, lexicon, and morphology is Nikulin 2020. In that proposal, 11 consonants (*/p m w t n r c ñ j k ŋ/) and at least 16 vowels (*/a â ə ə̃ ə̂ y ỹ o ô u ũ e ẽ ê i ĩ/)—and possibly even more, as indicated by subscript digits—are reconstructed for Proto-Macro-Jê. The maximal syllable was */CrVC°/, where /°/ stands for the so-called echo vowel[2]. Complex onsets were composed of a peripheral (labial or velar) non-continuant and a rhotic: */pr mr kr ŋr/. Underlying nasal onsets surfaced as postoralized preceding an oral nucleus: */m mr n ñ ŋ ŋr/ were thus pronounced as *[mb mbɾ nd ɲɟ ŋg ŋgɾ] before oral vowels. For example, PMJ */mi$_1$n°/ 'water' was likely pronounced as *['mbini]. In Nikulin's (2020) PMJ reconstructions, these allophonic realizations are represented by means of the combinations **mb*, **mbr*, **nd*, **nĵ*, **ŋg*, **ŋgr*, as in **mbi$_1$n°*. Likewise,

[1] The classification presented here differs from Nikulin 2020 in that Chiquitano is considered here a branch of Macro-Jê rather than an outgroup. This change is motivated by the absence of clear innovations that would define non-Chiquitano Macro-Jê languages as a clade. The labels in gray italics refer to scantly attested languages.

[2] A reviewer has inquired whether postnuclear consonants followed by an echo vowel are syllabified as codas or onsets. The answer depends on the level of analysis. On the surface, the echo vowel is indeed realized as a regular segment, with the preceding consonant syllabified as its onset (at least in some daughter languages). However, the underlying status of the echo vowels is less clear. It may be argued that their occurrence is best represented by a timing-related feature, whereby the release of the nucleus gesture is delayed until the release of the coda gesture.

underlying */j/ surfaced as *[ɲ] preceding a nasal nucleus, as in the genitive adposition PMJ */-jũk/ *[-ˈɲũk]. This allophone is represented as **ñ* in Nikulin's (2020) PMJ reconstructions, as in **-ñũk*.

Proto-Macro-Jê was a head-final language. An important fact about its morphosyntax is that PMJ stems were subdivided into two classes, known as relational and absolute stems. Relational stems required their internal argument to be expressed immediately to the left of the stem, either as a noun phrase or as a person index of the so-called internal series (one of **a-* 2, **i-* 3*NCRF*, **ta-* 3*CRF*). Note that the internal series lacked dedicated first-person indices, and pronouns were employed instead for expressing first-person internal arguments. Conversely, absolute stems lacked an internal argument and thus were not capable of taking person indices. Nouns, verbs, and adverbs/adpositions were lexically specified either as absolute or as relational; relational stems are indicated by means of a hyphen before the stem.

Another important division, which cross-cuts all relational stems, is whether their initial segment was the thematic consonant */j/ followed by a vowel (class II stems) or not (class I stems). The thematic consonant */j/ in class II stems was deleted upon the accretion of a person index; the person indices, in turn, had special allomorphs in class II stems: **∅-* 2, **c-* 3NCRF, **t-* 3CRF). Class I stems started with consonants other than */j/. It is tempting to analyze class II stems as underlyingly vowel-initial (cf. Rodrigues 2012), but Salanova (2011) shows that the thematic consonant */j/ is best understood as a part of the stem in at least some Macro-Jê languages.

Proto-Macro-Jê roots are commonly monosyllabic, though some disyllabic roots can be reconstructed as well. A frequent evolution pathway, especially common in Jê and Chiquitano, is the fossilization of prefixes or incorporated roots, whose semantics cannot be identified with precision at all times, at the left margin of stems, especially verbal ones. These fossilized elements have been variously labeled as formatives (Oliveira 2005: 82) or transitivity prefixes (Nikulin & Salanova 2019: 539–540) in Jê studies, and as classifiers in Chiquitano studies (Ciucci 2020).

One outstanding aspect of Proto-Macro-Jê phonology is the frequent occurrence of stem-final consonants, which may be followed or not by an echo vowel. These consonants were often lost in many contemporary languages. Nikulin & Silva (2020) establish that three branches of Macro-Jê are useful for reconstructing PMJ codas. **Maxakalí** (alongside other Maxakalian languages) is particularly conservative regarding the place of articulation of Proto-Macro-Jê codas, but not their manner of articulation or the echo vowels; synchronically, the language distinguishes between four codas, symbolized as /*P T C K*/, which are underspecified for features other than place of articulation (Silva 2015, 2020). **Krenak** is conservative in that it preserves stops as stops and nasals as nasals in the coda position, but erstwhile alveolar codas merge with velar ones (**-t*, **-n* > *-k*, *-ŋ*), and erstwhile palatal codas become alveolar (**-c*, **-ñ* > *-t*, *-n*); echo vowels are not preserved. **Proto-Jê** preserves most PMJ codas intact, but many of them are lost in individual Jê branches, sometimes leaving traces such as vowel lengthening[3] or morphophonological alternations. For the development of codas in other branches of Macro-Jê, see Nikulin (2020: 158sqq.). Taking PMJ codas into account is crucial for any attempts at exploring the external connections of Macro-Jê, especially given the fact that most

[3] This is a novel finding, not described in Nikulin & Silva (2020) or Nikulin (2020). More specifically, the nasal codas **-n* and **-ñ* followed by echo vowels are deleted in the Goyaz branch with compensatory lengthening, as in PCerr **pryn* 'road', **-mbyn* 'tail', **-mbên* 'liquid', **-jarên* 'root', **-ŋgôn* 'louse', **mãñ* 'greater ema', **-jwañ* 'tooth', **-kwañ* 'hole', **-ŋgoñ* 'wet', **-ŋgrôñ* 'embers' > Proto-Goyaz **pry:*, **-mby:* 'penis' (cf. **-jamby:* 'tail'), **-mbê:*, **-jarê:*, **-ŋgô:*, **-mã:*, **-ĵwa:*, **-kwa:*, **-ŋgo:*, **-ŋgrô:*. Note that in the Macro-Jê Alphabet echo vowels are unmarked in PCerr reconstructions (by contrast, their absence is marked by means of an apostrophe). The vowel length is most consistently reflected in Pykobjê–Krĩkatí as documented by Pries (2008).

Proto-Macro-Jê roots are monosyllabic. Evidently matches involving **C(r)VC* structures are more reliable for demonstrating common origin of languages than those involving **C(r)V* structures.

As shown in Figure 1 above, the Macro-Jê family is currently thought to include four first-level branches. The **Eastern** branch includes Jê, Maxakalian, Krenak, and possibly the poorly known Jaikó and Kamakã languages. These languages, except for a few Jê languages, are spoken east of the Araguaia River. The **Western** branch includes several languages spoken west of the Araguaia River: Rikbaktsa, Ofayé, and the Jabutian languages Djeoromitxí and Arikapú. **Karajá**, spoken along the Araguaia River, and **Chiquitano**, spoken mostly in the Chiquitano Dry Forest region in Bolivia and adjacent areas of Brazil, do not appear to form a clade with any other Macro-Jê group. Therefore, I consider that a given form can be reconstructed for Proto-Macro-Jê if its reflexes are present in at least two major subdivisions of Macro-Jê (Eastern, Western, Karajá, or Chiquitano).

2. Proto-Tupian

The Tupian language family includes approximately 70 languages, of which ca. 45 serve as primary means of communication in the respective communities. The subgrouping of Tupian is shown in Figure 2.

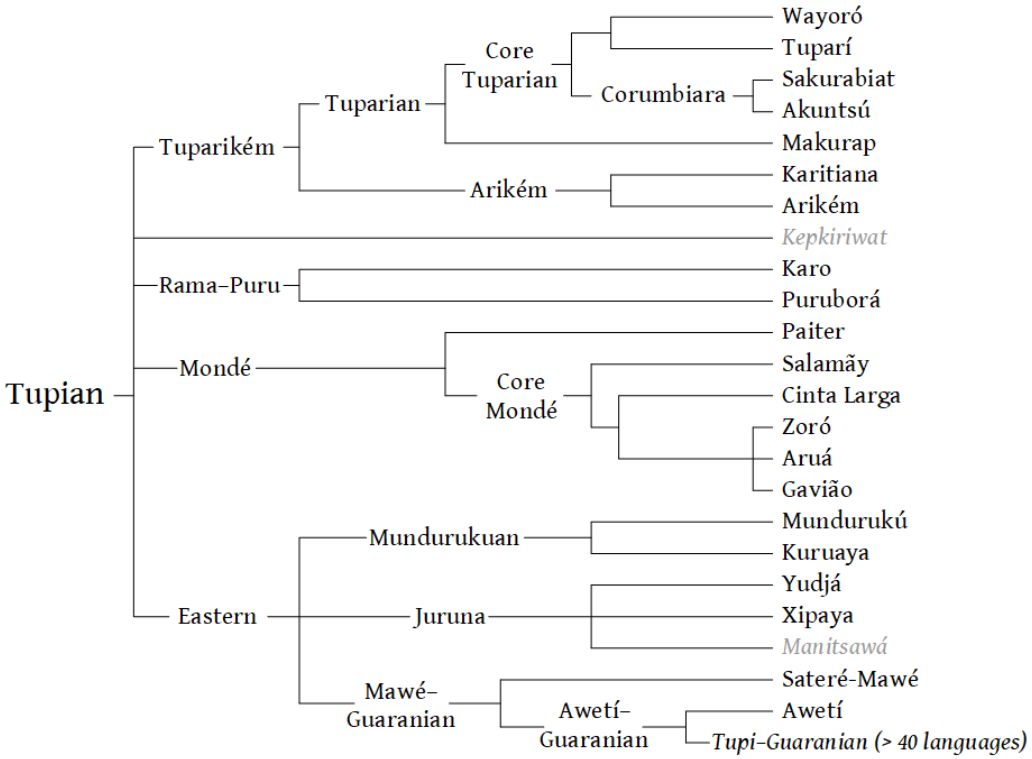

Figure 2. Tupian *Stammbaum* (based on Nikulin & Carvalho 2022: 20–21)[4]

[4] Nikulin and Carvalho's (2022) proposal differs from a more conservative proposal by Galucio et al. (2015) in that it posits a clade consisting of Tuparian and Arikém (based on three shared innovations involving Proto-Tupian **ɨ* and **ə*), dubbed "Tuparikém", and reinstates the so-called Eastern clade, originally proposed by Rodrigues (2005).

For a significant period of time, the only attempt at a phonological and lexical reconstruction of Proto-Tupian had remained that of Aryon Dall'Igna Rodrigues, with an early version thereof found already in Hanke et al. 1958. Its elements are presented in a significant number of publications by Rodrigues and his students, with Rodrigues 2005, 2007 and Corrêa-da-Silva 2010 being the most complete sources. Rodrigues' proposal has been criticized for failing to follow the principles of bottom-up reconstruction; for his overreliance on data of one single branch, Tupi–Guaranian, and especially the Old Tupí language; for misrepresentation of phonological facts of specific languages; and for positing typologically implausible developments (Meira & Drude 2015: 290–291; Singerman 2018: 390–392; Nikulin & Carvalho 2019: 276–278, 2022, among others). Moreover, recent years have seen considerable progress in phonological and lexical reconstruction of the protolanguages of individual Tupian branches, such as Proto-Tupi–Guaranian (Carvalho 2022, 2023, forthc.; Carvalho & Birchall 2022); Proto-Mawé–Guaranian (Meira & Drude 2015), Proto-Mundurukuan (Picanço 2019), Proto-Juruna (Fargetti & Rodrigues 2008, 2021; Carvalho 2019), Proto-Tuparian (Galucio & Nogueira 2012; Nogueira et al. 2019; Nikulin & Andrade 2020), Proto-Tuparikém (Nikulin forthc.). Taking into account recent progress in comparative studies of Tupian, Nikulin and Carvalho (2022) proposed an updated reconstruction of the sound system of Proto-Tupian, with an emphasis on the vowel system, though the reconstruction of Proto-Tupian consonants was also updated with respect to Rodrigues' (2007) proposal.

The inventory of Proto-Tupian onsets posited by Nikulin and Carvalho (2022) includes */p m β w t n ð r tʲ c j k kʲ ɻ ŋ ʔ/. Of these, the consonant */tʲ/ is rare but well-supported, while the reconstruction of */β/ and */ð/ is more dubious. The phonological and phonetic properties of */k kʲ ɻ/ are a matter of speculation. PT */k/ yields velar reflexes in all branches; */kʲ/ yields velar reflexes in all branches except Tuparian and Kepkiriwat, which reflects it as */ʔ/ or zero; */ɻ/ yields velar reflexes in Tuparian and Kepkiriwat, but */ʔ/ or zero in other branches. Since it is unclear whether */kʲ/ was actually articulated as a palatalized velar stop, I will henceforth employ the *ad hoc* character */ǩ/ for the character in question; I also replace */ɻ/ with the *ad hoc* character */ƙ/ so as to avoid unwanted associations with ejective or uvular stops. Similarly to Proto-Macro-Jê, the underlying nasals */m n ŋ/ were articulated as postoralized *[mb nd ŋg] before oral vowels; for example, */məC/ 'snake' was likely articulated as *[mbəc̚]. This is represented in Nikulin & Carvalho's (2022) Proto-Tupian reconstructions (as in **mbəC*), following Moore and Galucio's (1994: 124) representation conventions for Tuparian. No complex onsets can be reconstructed for Proto-Tupian.

The inventory of Proto-Tupian codas includes only four possibilities: */P T C K/. The use of small caps signals that these codas were underspecified for features other than place of articulation, just like in Maxakalí (Silva 2015, 2020). This is still the case in many daughter languages, such as Gavião (Moore 1984: 230), Proto-Tuparian (Moore & Galúcio 1994: 123), Sakurabiat (Galucio 1994: 998–992), Puruborá (Galucio 2005: 170–171), Awetí (Drude 2009), Tuparí (Singerman 2016), and many other languages for which such an analysis has never been proposed but is certainly possible. Major deviations from this pattern are found in Juruna, where erstwhile codas are now syllabified as onsets of the following syllables, and in Karo and Mundurukuan, where codas now contrast for nasality. In both Karo and Mundurukuan, codas are usually oral after oral vowels, and nasal after nasal vowels. However, nasal codas also occur after oral vowels, mostly at morpheme boundaries (as is the case with two homonymous suffixes in Mundurukú, *-m* 'instrumental' and *-m* 'imperfective'; Picanço 2005: 158–163), as a result of morphophonological processes (such as /-t-t/ → /-n/ in Karo; Gabas Jr 1999: 58–59), or due to late vowel denasalization, as in Proto-Tupian **-jĩ:K* > Proto-Mundurukuan **-ðiŋ* 'smoke' (Picanço 2019: 139). Karo is unique among Tupian languages in allowing oral codas after nasal

vowels, as in *-jakõp* ‘warm’ or *-pãt* ‘beautiful’ (Gabas Jr 1999: 49), a fact unaccounted for by Nikulin and Carvalho (2022).

Seven vowel qualities are reconstructed for Proto-Tupian: */a ə ɨ e i o ɯ/. Each of them had a nasal counterpart. This proposal differs from the traditional reconstruction in Rodrigues 2005, which posited only six vowel qualities (*/a ɨ e i o u/), in having */ə/ instead of his */o/ (and sometimes */e/), */ɨ ɯ/ instead of his */ɨ/, and */o/ instead of his */u/. There is evidence that vowel length may have been contrastive; it is best preserved in the Tuparikém branch and possibly in Sateré-Mawé and Mondé.

Mundurukuan, Juruna, Mondé, and Karo languages are tonal, and there is evidence that pitch accent may be contrastive in Makurap; in addition, lexically specified stress has been described for Tuparí and Akuntsú (see Nikulin & Andrade 2020: 286). This suggests that Proto-Tupian may have also been a tonal language, but no attempts have been made at reconstructing its prosody.

A typical Proto-Tupian morpheme had one or two syllables, and morpheme-internal codas appear to have been rare (though existent, as in **jaCjo* ‘armadillo’). Just like in Proto-Macro-Jê, stems were subdivided into relational and absolute, with relational stems obligatorily taking a complement immediately to its left (signaled by means of a hyphen before the stem), and absolute stems disallowing them[5]. The class of relational stems was further subdivided into two classes. Class I stems started with consonants, took the allomorph **i-* of the third-person index, and did not require any thematic element when their internal argument was expressed by means of a noun phrase. Class II stems, conversely, took the allomorph **c-* of the third-person index, and occurred with the thematic consonant **j-* when their internal argument was expressed by means of a noun phrase (or a person index other than the third-person one). The original configuration is most faithfully preserved in Makurap, Mundurukuan, and Sateré-Mawé.

Although the differences between Rodrigues’ (2005, 2007) and Nikulin and Carvalho’s (2022) proposals are significant, the consequences of preferring one proposal over the other are minimal for the purposes of establishing long-range connections with other families. This is so because most differences are related to the manner of articulation of the consonants and to specific vowel features, but the reconstructed forms are still quite similar across proposals, as shown in Table 1.

	‘to grind’ = ‘larva’	‘arrow’	‘leaf’	‘to seize’	‘door’	‘armadillo’
Rodrigues	**ček^{w}*	**ek^{w}ʔɨp*	**ep^{w}*	**pɨčɨk*	**ek^{w}en*	**tajtu*
Nikulin & Carvalho	**-təK*	**əʞɯP / *-jəʞɯP*	**əP / *-jəP*	**-pɨtɨK*	**ək-ẽT / *-jək-ẽT*	**jaCjo*

Table 1. Rodrigues’ (2005, 2007) and Nikulin & Carvalho’s (2022) Proto-Tupian reconstructions

As shown in Figure 2 above, the Tupian family is currently maintained to include no less than five first-level branches. The **Eastern** branch includes the Juruna, Mundurukuan, and Mawé–Guaranian groups (the latter is further subdivided into Sateré-Mawé and Awetí–Guaranian, and Awetí–Guaranian is in turn subdivided into Awetí and Tupi–Guaranian). This branch reaches its highest diversity between the lower Madeira and the lower Iriri Rivers. The **Tuparikém** branch includes the Tuparian and Arikém groups, which includes seven lan-

[5] There were also pairs of relational and absolute stems, which some authors have analyzed as constituting an inflectional paradigm. Examples include **-pɨ* ‘foot (rel.)’ / **mbɨ* ‘foot’ (abs.), **-ja:pe* ‘path (rel.)’ / **pe* ‘path’ (abs.), and **-jəK* ‘house (rel.)’ / **əK* ‘house’ (abs.). I prefer envisaging such pairs as derivational.

guages spoken in what is now the Brazilian state of Rondônia. The **Mondé** branch includes a handful of languages spoken in Rondônia and in adjacent areas of the Mato Grosso state. The **Rama-Puru** branch includes two languages, Karo and Puruborá, both spoken in Rondônia. The extinct **Kepkiriwat** language was also spoken in Rondônia. The languages of the latter four branches are therefore spoken in the same area, which facilitates lexical diffusion. I consider that a given form can be securely reconstructed for Proto-Tupian if its reflexes are present in the Eastern branch and at least one of the Rondonian branches (Tuparikém, Mondé, Rama-Puru, and Kepkiriwat). If the Eastern branch lacks a cognate, the requirement is that reflexes be present in at least three branches. Cognate sets whose distribution is restricted to two Rondonian branches (say, Tuparikém and Mondé) are likely to involve horizontal transmission. For example, one could technically reconstruct PT *-*aka:T* or *-*aǩa:T* 'to bite' based on Karitiana *-okoot*, Paiter *-ákar*, and Salamãy *-áka:l*, but given the absence of cognates outside the Tuparikém and Mondé branches, this verb is unlikely to have existed in Proto-Tupian.

3. Macro-Jê–Tupian hypothesis

Possible external relations of Macro-Jê and Tupian are still debated. Both families have figured in a number of partially overlapping long-range proposals, and even the limits of the Macro-Jê family are not universally agreed upon. Macro-Jê languages have been linked to, or claimed to include as a constituent branch, language groups such as Bororoan, Yaathê, Karirian, Purian, Guató, and Otí (Guérios 1939; Davis 1968; Greenberg 1987; Rodrigues 1993, 1999; Ribeiro 2002, 2011; Ribeiro & Voort 2010; Martins 2009, 2011; Nikulin & Carvalho 2018; Silva forthc.). Other long-range proposals have connected Macro-Jê to language families such as Cariban (Rodrigues 2000, 2009; Meira et al. 2010: 512–515; Nikulin & Carvalho 2018); Chibchan (Pache 2023); Mapudungun and Katukina–Harakmbut (Adelaar 2008: 11); Mataguayan and Guaicuruan (Viegas Barros 2005; Nikulin & Carvalho 2018), Payaguá and Guachí (Viegas Barros 2005), Zamucoan (Nikulin & Carvalho 2018), and even the putative Nostratic macrofamily (Aikhenvald-Angenot & Angenot 1992). Tupian has been most notably compared with Cariban (Rodrigues 1985, 2000, 2009; Meira et al. 2010: 512–515; Nikulin & Carvalho 2018), but also Bororoan (Nikulin & Carvalho 2018), Yaathê (Silva forthc.), Karirian (Ribeiro 2002; Nikulin & Carvalho 2018), Mataguayan, Guaicuruan, and Zamucoan (Nikulin & Carvalho 2018). In addition, the aforementioned families were thought by Greenberg (1987) to be part of a much larger Amerind macrofamily, with Macro-Jê classified as a member of the so-called Ge–Pano–Carib branch, and Tupian as a member of the so-called Equatorial subgroup of the Andean–Equatorial branch.

This study, however, focuses on one specific proposal, whereby Macro-Jê and Tupian are considered to be related to each other (though possibly also to other language families). Although some lexical lookalikes have been identified already by Davis (1968: 47), the most widely known claim on the possible relation of these two families is found in Rodrigues (2000, 2009), who proposes that Macro-Jê, Tupian, and Cariban are all ultimately related (note that in Rodrigues' definition the Macro-Jê family encompasses language groups such as Bororoan, Purian, Karirian, Yaathê, and Guató, whose inclusion is not supported by Nikulin's 2020 study). The proposal has had a moderately positive reception in the scholarly community (cf. Meira et al. 2010: 512–515; Ribeiro 2002: 41–42, 2011: 107–109; Nikulin & Carvalho 2018) and sometimes goes by the label "TuKaJê".

The evidence that substantiates the TuKaJê hypothesis is largely morphological and morphophonological in nature. Most notably, Macro-Jê, Tupian, and Cariban share a pattern

whereby stems capable of taking an internal argument — directly possessable nouns, postpositions, and at least some classes of verbs in at least some constructions — are subdivided into two large classes, commonly referred to as "class I" (which typically includes consonant-initial stems) and "class II" (vowel-initial stems). Class I stems do not undergo any alternations in their paradigm, and they combine with the allomorph **i-* of the third-person index in Proto-Macro-Jê, Proto-Tupian, and Proto-Cariban. By contrast, class II stems are preceded by the element **j-* when they take an internal argument expressed by a noun phrase in its canonical position (i.e., immediately to the left from the head), again in Proto-Macro-Jê, Proto-Tupian, and Proto-Cariban. This element has been variously analyzed as a so-called "contiguity relational prefix" (in works by Rodrigues and his students), as a "thematic consonant" (Nikulin 2020), or as the initial segment of the stem (Salanova 2011; Meira & Drude 2013, 2015). When the internal argument is expressed by a third-person index, the latter takes the allomorph **c-* in Proto-Macro-Jê and Proto-Tupian (**∅-* in Proto-Cariban), and the element **j-* is not present. With other person indices, **j-* may be present or absent depending on the language family and the person. This is illustrated below in example (1) (Proto-Macro-Jê and Proto-Tupian reconstructions are mine; the Proto-Cariban paradigm is from Meira et al. 2010). Note the outstanding similarities in the person indices themselves, which are particularly strong between Tupian and Cariban.

(1)	Proto-Macro-Jê	Proto-Tupian		Proto-Cariban
class I	*NP *ŋgyn°*	*NP *ŋgɯP*	'NP's louse'	*NP *C…*
	**i-ŋgyn°*	**i-ŋgɯP*	'her/his louse'	**i-C…*
	**(∅-)ŋgyn°*	**o-ŋgɯP*	'my louse'	**u-C…*
	**a-ŋgyn°*	**e-ŋgɯP*	'your louse'	**ə-C…*
	**ta-ŋgyn°*	**tə-ŋgɯP*	'her/his own louse'	**tɨ-C…*
	**u-ŋgyn°*	—	'our (INCL) lice'	**kɨ-C…*
class II	*NP *j-uñ°*	*NP *j-ãC*	'NP's tooth'	*NP *j-V…*
	**c-uñ°*	**c-ãC*	'her/his tooth'	**∅-V…*
	**(∅-)j-uñ°*	**o-j-ãC*	'my tooth'	**u-j-V…*
	**∅-uñ°*	**e-j-ãC*	'your tooth'	**ə-j-V…*
	**t-uñ°*	**tə-j-ãC*	'her/his own tooth'	**t-V…*
	**u-j-uñ°*	—	'our (INCL) teeth'	**k-V…*

Another morphological similarity, identified by Ribeiro (2002: 41–42), involves the morphology employed for converting absolute (unpossessable) nouns to relational (possessable) ones in a subset of Macro-Jê and Tupian languages. In a few languages belonging to the Cerrado branch of the Jê group — Xavante, Akwẽ-Xerente, and possibly Panará — this is attained by means of a prefix or an adposition whose Proto-Cerrado form may be reconstructed as **-ñĩm-* (> Xavante *-nhim-/-nhib-/-nhi-*, Akwẽ-Xerente *-nĩm*, Panará *-jĩ-*), as shown in (2).

(2) a. Xavante < Akuwẽ < Cerrado < Jê < Macro-Jê (Estevam 2011: 163)

dzeru	→	*wa-nhib-dzeru-wawẽ*
money		1SG-PSSD-money-AUG
'money'		'our plentiful money'

b. Akwẽ-Xerente < Akuwẽ < Cerrado < Jê < Macro-Jê (Xerente 2019: 77)

tka	→	*ĩ-nĩm=*	*tka*
land		1SG-PSSD=	land
'land'		'my land'	

c. Panará < Cerrado < Jê < Macro-Jê (Dourado 2001: 72)[6]

inkwa	→	*kjẽ-mẽra*	*jĩ*	*kwa*
house		I-PL	PSSD	house
'house'		'our house'		

As for the Tupian family, a likely cognate prefix, dubbed 'indirect possession mediator' in Rodrigues et al. 2006: 23, is found in three major branches: Tuparian (Makurap *-xep-* 'alienable possession marker'), Mundurukuan (Mundurukú *-e-*, bearing high tone after noun phrases and low tone after person indices), and Mawé–Guaranian (Sateré-Mawé *-e-*, or *-he-* after some person indices; Awetí *te-* / *-e-*; PTG *te(p)-* / *-re(p)-*). I follow Rodrigues et al. 2006: 23 in reconstructing its Proto-Tupian form as **-eP-*. Its final **P* is preserved in Makurap as well as in the TG relational stem for 'container' (3i). It is deleted before consonant-initial roots in TG, and before all roots in Mundurukú, Sateré-Mawé, and Awetí. The Makurap, Awetí, and Tupi-Guaranian reflexes suggest the reconstruction **jep-* instead of **ep-*. One may surmise that reflexes of **j-* in the latter set of languages were inserted due to the fact that vowel-initial possessable (relational) stems are otherwise uncommon in Tupian. Some examples follow in 3.

(3) Makurap < Tuparian < Tupian (Braga 2005: 42–43)

a.

-pia-t	→	*o=xe-pia-t*
-liver-PSSD		1SG=ALZ-liver-PSSD
'liver'		'my liver (an animal's liver belonging to me)'

b.

xau	→	*o-xep-xau-t*
flour-PSSD		1SG=ALZ-flour-PSSD
'flour'		'my flour'

Mundurukú < Mundurukuan < Tupian (Picanço 2005: 259)

c.

kòbé	→	*ayácát*	*é-kòbé*
canoe		woman	PSSD-canoe
'canoe'		'woman's canoe'	

d.

nobánő́	→	*wuy-e-nobánő́*
rifle		1+2-PSSD-rifle
'rifle'		'our (*INCL*) rifle'

Sateré-Mawé < Mawé–Guaranian < Tupian (Ribeiro 2010: 67, 85, 90, 91)

e.

kui'a	→	*uru-e-kui'a*
bowl		1+2-PSSD-bowl
'calabash bowl'		'our (INCL) calabash bowl'

f.

sokpe	→	*u-he-sokpe*
clothes		1-PSSD-clothes
'clothes'		'my clothes'

[6] Dourado (2001: 71–72) claims that *-jĩ* is only found in elders' speech, and that the more common genitive postposition (or rather a genitive noun in her analysis) is *-jõ*, with cognates all across Macro-Jê (Ribeiro 2002, 2009) that reflect Proto-Macro-Jê **-ñũk* (Nikulin 2020: 404). However, the very existence of Panará *-jĩ* is doubtful: all instances of this form in the cited word are accompanied with the noun transcribed as *koa* by Dourado (2001: 71–72, 77), whose form is attested as *inkwa* /ŋwa/ [ĩŋˈkwa] in more recent works (Bardagil-Mas 2018: 51). It is thus possible that the combination *-jĩ kwa* in Dourado (2001) is simply a phonetic variant, or even a mistranscription, of *-jõ inkwa*. More recent sources on Panará do not report the existence of *-jĩ* either.

Awetí < Awetí–Guaranian < Mawé–Guaranian < Tupian (Drude 2011: 178)

g. *ky* → *Mopot* *e-ky*
ax Mopot PSSD-ax
'ax' 'Mopot's ax'

Old Tupí < TG < Awetí–Guaranian < Mawé–Guaranian < Tupian (Barbosa 1956: 110–111)

h. *mbetar-a* → *te-mbetar-a*
tembetá-REF PSSD-tembetá-REF
'tembetá' 'one's tembetá'

i. *uru-∅* → *abá* *rep-uru-∅*
container-REF person PSSD-container-REF
'container' 'indigenous person's container'

Despite the morphosyntactic, semantic, and phonological similarity between the aforementioned morphemes, Ribeiro's (2002: 41–42) hypothesis is rendered less plausible than it could have been by the very limited distribution of **-ñĩm-* on the Macro-Jê side of the comparison: its reflexes are only found in the Cerrado branch of the Jê group (or, if the alleged Panará reflex is shown to be a ghost morpheme—as suggested in footnote 6—, in its Akuwẽ subbranch), and an entirely different postposition **-ñũk* is reconstructed in the same meaning for Proto-Macro-Jê.

Finally, Rodrigues (2000: 101), Ribeiro (2002: 42), and Rodrigues et al. (2006: 34–35) point out the similarity between morphemes of similar shape in some Macro-Jê languages (Xavante *-nhimi-*, Akwẽ-Xerente *-nmĩ-* < Proto-Akuwẽ **-ñĩmĩ-*) and in some Tupian languages (PTG unpossessable **mbi-*, possessable **te-mbi-* / **-re-mbi-*, with cognates in Aweti, Sateré-Mawé, and possibly other branches, such as Tuparian), whose function has been variously described as a patient nominalizer or an antipassive nominalizer. In both language families, it attaches to transitive verbs (more specifically, to their nonfinite forms in the case of the Macro-Jê languages), and takes a possessor encoding the notional agent of the verb. This is unusual, since in both language families it is typically the absolutive participant — and not the ergative one — that shares the coding strategy with possessors of nouns. Cf. the illustrations in (4):

(4) a. Xavante < Akuwẽ < Cerrado < Jê < Macro-Jê (Estevam 2011: 330)
romhu-ri → *wa-nhimi-romhu-ri*
work-NF 1PL-NMLZ.ANTP-work-NF
'work.NF' 'our work'

b. Sateré-Mawé < Mawé–Guaranian < Tupian (Ribeiro 2010: 67, 71)
-koi → *mi-koi*
plant_V NMLZ_P-plant_V
'to plant' 'plant (noun)'

c. Apyãwa < TG < Awetí–Guaranian < Mawé–Guaranian < Tupian (Almeida et al. 1983: 35)
-'o → *xe-re-mi-'o-∅*
eat 1SG-PSSD-NMLZ_P-eat-REF
'to eat' 'my food'

Once again, the Macro-Jê–Tupian comparison is undermined by the distribution of the alleged cognates on the Macro-Jê side, with reflexes being restricted to the Akuwẽ subbranch of

the Cerrado branch of the Jê group. Prefixes with similar properties in other Macro-Jê languages, such as the Proto-Chiquitano inverse voice marker *-*ij*-, bear no formal similarity to Proto-Akuwẽ *-*ñĩmĩ*- or to the Tupian forms.[7]

Regarding the lexical evidence, as mentioned above, Davis (1968: 47) identified ten similarities between his own Proto-Jê reconstructed forms and Proto-Tupian reconstructions extracted from Hanke et al. 1958 as well as forms representing Guajajara, a Tupi–Guaranian language of the Tupian family. Five of his cognate pairs — 'liver', 'husband', 'foot', 'to eat', 'hand' / 'arm' — are accepted as valid in this study, and are discussed in **4.1**. The remaining five comparisons are rejected because of a mismatch in stem-final consonants, not always reconstructed by Davis (1968) for Proto-Jê but identified in later comparative work ('water', 'louse', 'head', 'path', 'one'). This is summarized in Table 2. The updated reconstructions are from Nikulin 2020 for the Macro-Jê languages, whereas for the Tupian languages they are based on Nikulin & Carvalho 2022. Guajajara forms have been checked against Harrison & Harrison's (2013) dictionary.

Rodrigues & Cabral (2010) make another attempt at identifying lookalikes involving Macro-Jê and Tupian languages. These authors take it for granted that languages such as Bororoan, Karirian, Purian, Yaathê, and Guató are part of the Macro-Jê family (cf. Rodrigues 1999), a position not confirmed by more recent studies; as a consequence, multiple proposed cognate sets do not include data of languages classified as Macro-Jê beyond reasonable doubt. A serious shortcoming of Rodrigues & Cabral's (2010) study is that they consider data of contemporary Macro-Jê languages rather than reconstructed forms. Once the phonological history of individual languages is taken into account, some problems in Rodrigues & Cabral's (2010) proposal become apparent. A case in point is their comparison of several Kaingang forms containing *f* /ɸ/ with Old Tupí forms containing **p*. Kaingang *-fa* 'leg', *-fór* 'full', *-for* 'thrown away', *-fo* 'pus' (whence *-fó-m* 'to suppurate'), *-fyr* 'extremity', *-fár* 'skin, bark', *-fi* 'to give, to lay' are thus compared to Old Tupí *-py* 'foot', *-por-* 'full', *-por-* 'to jump', *-peu* 'pus', *-apyr-* 'tip', *-pir-* 'skin', *t-epy* 'payment, price'. However, as observed by Ribeiro (2004a: 94, fn. 3), Kaingang *f* /ɸ/ is known to go back to a coronal consonant, reconstructed in Nikulin 2020 as PSJ **θ* < PJ **c* < PMJ **c*, which entails that the Kaingang–Tupian lookalikes are fortuitous.

The absence of a phonological reconstruction of Proto-Macro-Jê has for long remained a major obstacle in further entertaining the Macro-Jê–Tupian hypothesis. This gap has now been filled (Nikulin 2020), as discussed in section **1**. Furthermore, Nikulin and Carvalho (2022) proposed a revision of the reconstruction of Proto-Tupian, as stated in section **2**. We are therefore now in position to compare reconstructed Proto-Macro-Jê and Proto-Tupian forms.

[7] The Proto-Chiquitano inverse voice marker *-*ij*- almost certainly goes back to a patient nominalizer, since the Chiquitano verbs in the inverse voice encode their notional patient by means of suffixal person indices, reminiscent of those used for nominal predication (and not found elsewhere in the verbal paradigm), whereas their notional agent is encoded by means of prefixes of the absolutive/genitive series. That way, I surmise that the Chiquitano inverse construction (as in **a-ij-arapá-ta* 2SG-INV-pour-F.3SG$_P$ 'you pour it') originated as a nominal predication ('it is your poured thing'). This is formally and functionally similar to the evolution of patient nominalizations in the Tuparian languages, which currently employ erstwhile patient nominalizations in the object focus construction (Galucio & Nogueira 2018). I do not discard the possibility that Proto-Chiquitano *-*ij*- is cognate with the Tuparian (erstwhile) patient nominalizers: Wayoro, Akuntsu, Sakurabiat *-i-*, Tuparí *-y(')-*, Makurap *-yĩ-* (the correspondences between these forms are not entirely regular, making it difficult to unambiguously reconstruct the Proto-Tuparian form).

gloss	Proto-Jê (Davis 1968)	updated reconstruction	Proto-Tupian or Guajajara (as in Davis 1968)	updated form or reconstruction	comparison status
'liver'	**ma*	PJ **-mba* < PMJ **-mbâ*	PT **pɨa*	PT **-pɨʔa*, abs. **mbɨʔa*	accepted
'husband'	**mzɛn*	PJ **-mbê$_2$n'* < PMJ **-mbi$_2$n*	PT **men*	PT **-mẽT*	accepted
'water'	**ŋo*	PCerr **ŋgôj'*, PSJ **ŋgôj* (irregular correspondence)	PT **ɨgɨ*	PT **ʔɯ* 'water'; **-ŋgɯ* 'liquid' (Tuparikém branch only)	rejected (coda mismatch + poor distribution)
'louse'	**ŋo*	PJ **-ŋgə̂$_1$n* < PMJ **-ŋgy$_1$n°*	PT **ŋkɨv*	PT **-ŋgɯP*	rejected (coda mismatch)
'arm' / 'hand'	**pa* 'arm'	PJ **-pa* 'arm, branch' < PMJ **-pa*[8]	PT **po* 'hand'	PT **-pə*, abs. **mbə* 'hand'	accepted
'foot'	**par*	PJ **-par* < PMJ **-pâr°*	PT **pɨ*	PT **-pɨ*, abs. **mbɨ*	accepted
'head'	**krã*, **krãñ*	PJ **-krỹj'* < PMJ **-krỹñ°*	Guaj. *kaŋ-*	Guaj. *-àkàg* 'head' < PTG **-ʔa-kãK* 'head' < PT **-ʔa* 'head' + PTG **-kãK* 'bone'	rejected (coda mismatch + wrong morphological segmentation)
'to eat'	**ku*, **kur*	PJ **-ku$_2$* < PMJ **-ko$_2$*	Guaj. *-ʔu*	Guaj. *-'u* < PTG **-ʔu* < PT **-ko*	accepted
'path'	**prɨ*	PJ **pryn* < PMJ **prən°*	Guaj. *pɛ*	Guaj. *pe* / *-rape* < PTG **pe* / **-rape* < PT **pe* / **-ja:pe*	rejected (coda mismatch)
'one'	**pɨci*, **pɨcit*	PCerr **-p^{ɔ}ji* < PMJ **-p(V)jet*	Guaj. *pitci*	Guaj. *pitài ~ pitei ~ mitài ~ pitàz ~ petei* < **pe-tẽ-C*[9]	rejected (multiple issues)

Table 2. Davis' (1968) Jê–Tupian etymologies and their current status

4. Possible cognates

This section presents the lexical evidence for the Macro-Jê–Tupian hypothesis. Seeking to reduce the number of false positives, I adopt a stringent approach to cognate identification. In order to qualify as a likely cognate set, the reconstructed Proto-Macro-Jê and Proto-Tupian morphemes must show a full match between all consonants with respect to the place of articulation (i.e., Proto-Macro-Jê labial consonants can only correspond to Proto-Tupian labial consonants, and so on), whereas back vowels in one protolanguage are not allowed to correspond

[8] Nikulin (2020: 369) reconstructs a palatal coda in this word (PMJ **-paj ~ *-paj°*), based on the Akuwẽ reflexes: Xavante *-pai-hi* 'arm', *-pa-nõ* [-paˈnːõ] /-paj-dõ/ 'arm', Akwẽ-Xerente *-pai-nõ* 'arm'. Note, however, that the palatal coda does not show up in Akwẽ-Xerente *-pa-krta* // *-pa-krda* 'arm', nor is it visible in Xavante *pa* 'creek', *-pa* or *wede-pa* 'branch, root'; Akwẽ-Xerente *-pa* or *wdê-pa* 'root'. Therefore, the grounds for reconstructing a palatal coda in PMJ are rather weak.

[9] The reconstruction **petẽC* is shallower than Proto-Tupi–Guaranian: the reflexes of this form are only found in a few Tupi–Guaranian languages, such as Tapiete *pente*, Mbyá *peteĩ ~ teĩ*, Ka'apor *peteĩ*. This is likely a fossilized derivative from the PTG root **pe* (followed by **-(e)te* 'true' and by a diminutive suffix). Most Tupi–Guaranian languages reflect different derivatives of **pe*, such as **o-je-pe* (with a 3rd person active prefix **o-* and the reflexive prefix **-je-*); see Schleicher 1998: 12–13.

to front vowels in another protolanguage. Some slight deviations from this principle are duly justified. At this stage, semantically shifted cognates are not considered. In what follows, I list the Macro-Jê–Tupian lookalikes that satisfy the aforementioned criteria separated into four groups: etyma that are clearly reconstructible both to Proto-Macro-Jê and Proto-Tupian (**4.1**), etyma that are clearly reconstructible to Proto-Macro-Jê but have a deficient distribution in Tupian (**4.2**), etyma that are clearly reconstructible to Proto-Tupian but have a deficient distribution in Macro-Jê (**4.3**), and etyma that have a limited distribution both in Macro-Jê and Tupian (**4.4**). **4.5** lists some lookalikes that are best interpreted as loans or accidental resemblances.

In what follows, rather than citing reflexes in all daughter languages to support the reconstructed forms, I provide data from representative languages of each branch: typically Bésɨro for Chiquitano, Djeoromitxí for Jabutian, Maxakalí for Maxakalian, Xavante for Akuwẽ, Khĩsêtjê for Goyaz, Kaingang for Southern Jê, Makurap or Wayoró for Tuparian, Karitiana for Arikém, Paiter for Mondé, Yudja for Juruna, Mundurukú for Mundurukuan, Sateré-Mawé and Apyãwa for Mawé–Guaranian.

4.1. Good distribution in both families

3NCRF prefix: PMJ ****i-* / **c-*** : PT ****i-* / **c-***

The Proto-Macro-Jê reconstructions are from Nikulin (2020: 393, 423), who claims that **i-* was used with class I stems, and **c-* with class II stems. The reflex of **i-* are found in all Macro-Jê branches, including Chiquitano (Bésɨro *i-*), Western (Rikbaktsa *i-*, Arikapú *i-*, possibly also Ofayé *ə̃-*), Karajá (*i-*), and Eastern (Xavante *ĩ-*, Maxakalí *ũ-*). The reflexes of **c-* are found in at least one Western language (Ofayé *h-*), in Karajá (*t-*/*tx-*), and in several Eastern languages (Khĩsêtjê *s-*, Xavante *ts-*), but possibly also in Chiquitano (Bésɨro *∅-*). The original distribution is still clearly preserved in Karajá and the Akuwẽ languages,[10] possibly also in Ofayé and Chiquitano.

The Proto-Tupian prefixes **i-* and **c-* are likewise used with class I and class II stems, respectively. The original distribution is most clearly seen in the Mundurukuan and Mawe–Guaranian languages of the Eastern branch (Mundurukú *i-*/*y-* and *t-*, Sateré-Mawé *i-* and *h-*, Apyãwa *i-* and *h-*/*∅-*) and in one Tuparikém language (Makurap *∅-*/*y-* and *t-*). In the Tuparikém branch, the prefix **i-* is mostly preserved in all languages, with special reflexes before vowel-initial roots in Tuparian (Makurap and Wayoró *y-*, Tuparí *s-*/*y-*, Akuntsú *t-*/*ɲ-*, Sakurabiat *s-*); in Makurap, it was irregularly lost before consonants, thus yielding forms such as *∅-tur-et* ‘her/his spade’ or *∅-kar-et* ‘her/his body’ (Braga 2005: 51) instead of the expected **i-tur-et*, **i-kar-et*. However, in all Tuparikém languages except Makurap the prefix **i-* was also extended to erstwhile class II stems, replacing **c-* altogether. It is possible that the prefixes **i-* and **c-* are also reflected in the Mondé languages, but I am unaware of a coherent account of their evolution in that particular branch of Tupian.

The reflexes of this person index are opposed to those of PMJ **ta-*, PT **tə-* in some Macro-Jê branches (Karajá, Western) and in some Tupian languages (Tuparikém branch, Sateré-Mawé) in that the indexed argument has a disjoint reference with some other participant (typically the subject).

[10] The Akuwẽ languages have innovated by extending the prefix *ĩ-* (originally used with class I stems) to most class II stems, resulting in the allomorphs Xavante *ĩts-*, Akwẽ-Xerente *ĩs-* (instead of *ts-*/*s-*). The conservative allomorphs *ts-*/*s-* are found, for example, in the perlative postposition (Xavante *-dzô*, Akwẽ-Xerente *-zô*; third-person form *ts-ô*/*s-ô*).

‘meat, flesh’: PMJ ***ĩt* / *-ñĩt*** : PT ***ẽT* / *-jẽT***

The Proto-Macro-Jê reconstruction *-*ñĩt* is from Nikulin 2020: 407. The root is preserved in all first-level branches of Macro-Jê, including Chiquitano (Bésiro *n-{a}ñé-se*), Western (Djeoromitxí *-nĩ*, Rikbaktsa *-ni*), Karajá (*dèè*), and Eastern (Maxakalí *-yĩn*, Khĩsêtjê *-nhi*, Xavante *-nhi*, Kaingang *-nĩ*). The reconstruction of the coda *-*t* is based on the evidence from the Trans-São Francisco languages, where Krenak *-ñik* preserves its manner of articulation (with the regular change from an alveolar to a velar), and Maxakalí *-yĩn* preserves its place of articulation. The correspondences are regular. As for the absolute form **ĩt*, it is preserved only in the Maxakalí compound *ĩn-mõ-xa* ‘the Ĩnmõxa monster’, analyzed in Silva 2020: 184 as ‘the flesh going out’; it must be an archaism, since *-*ñĩt* ‘meat’ has extra morphology — the ancient relationalizing prefix */-j-/ — compared to it.

The Proto-Tupian reconstruction is based on Proto-Tuparian *-*ɲẽTʔã* (Nikulin & Andrade 2020: 296) and Proto-Mundurukuan *-*ẽn* (Picanço 2019: 137), with reflexes present in all languages of the respective branches (Wayoró *-yẽra*, Munduruku *-ẽn*, etc.); see Galucio et al. 2015: 253 for a selection. Proto-Tuparian shows a fossilized formative *-*ʔa* (originally a classifier for spherical objects, but found in other Proto-Tuparian terms as well) and the relationalizing prefix */-j-/, which surfaces as *-*ɲ-* before a nasal vowel. The correspondences are otherwise regular.

In Macro-Jê, the reflexes of *-*ñĩt* belong to class II in the languages of the Goyaz branch of the Jê group, but to class I in the languages of the Akuwẽ branch of Jê (see Estevam 2011: 138 for Xavante) and in Karajá (see Ribeiro 2012a: 216 for an example). It must have originally belonged to the less productive class II. In Tupian, *-*jẽT* must have originally belonged to class II, as attested for Makurap by Braga (2005: 208; note that she uses the label “class I” for my class II); other Tuparian languages have lost the distinction. Mundurukuan has apparently reanalyzed the erstwhile absolute stem *ẽT ‘meat/flesh (unpossessed)’ as relational.

‘to stand’: PMJ ***ja*** (nonfinite ***-ja-m***) : PT ***-ja*** or ***-ʔãP***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 398. The etymon is preserved in the Western (Ofayé *-xeːʔ ~ -heːʔ*, possibly Rikbaktsa *-sa* ‘to start walking /a developmental milestone/’), Karajá (nonfinite *-lma*), and Eastern (singular only: Maxakalí *-xip*, irrealis *xihip*; Khĩsêtjê *ta*, nonfinite *-tãm*; Xavante *dza*, nonfinite *-dzam*; Kaingang *jẽ*, nonfinite *jẽg*) branches. The correspondences are regular. The original finite form was lost in Maxakalí, whose irrealis form has been remodeled after the realis form (*-xip* < *-ja-m*; the expected irrealis form would be **xihi* < **ja*), and in Karajá, which now shows a suppletive finite form *-ỹi* of unknown origin.

The Proto-Tupian reconstructions correspond to two distinct etyma, which could be compared to PMJ finite and nonfinite forms, respectively.

Proto-Tupian *-*ja* is preserved in two Rondonian branches, Tuparikém (as an auxiliary only: Karitiana *ty-so* ‘IMPF:standing’, Sakurabiat *ta-t* ‘PRG.PRS:standing’, *ta-a* ‘PRG.PST:standing’) and Rama-Puru (Karo *-ja* ‘to stand’, with a possible cognate in Puruborá; Galucio et al. 2015: 258). The correspondences are regular. Note that Proto-Tuparikém can be reconstructed as having a series of no less than three auxiliares contrasting for position only, as shown in Table 3. These correspond to lexical verbs for ‘to sit’ and ‘to stand’ in Rama-Puru or other Tupian languages; the term for ‘to lie’ is noncognate in Rama-Puru (*-*mboP* > Karo *-mbop*, Puruborá *-bɔp-a*), but clear cognates are found elsewhere in Tupian, as in Old Tupí *tub-/-rub-* ‘to lie.NF’ (Barbosa 1956: 305).

Proto-Tupian *-*ʔãP* is preserved in at least one Tuparikém language (Akuntsu *-ãP*) and in most Mawé–Guaranian languages (Eastern branch), such as Sateré-Mawé *-’am* ‘to go up’, Old Tupí *-am* ‘to stand’, Kamayurá *-’am* ‘to stand’, among many other cognates. The verb for ‘to stand up’ in Mawé–Guaranian languages is evidently derived from this root: Sateré-Mawé *-poi’am* ‘to stand up’, Old Tupí *-puam* ‘to stand’, Kamayurá *-uhwam* ‘to stand’, among others. The correspondences are regular.

	Proto-Tuparikém	Sakurabiat	Karitiana	Proto-Rama-Puru	Karo
	AUX	PRG (PRS / PST)	IMPF	lexical verb	lexical verb
		Galucio 2001: 58	Rocha 2022: 239		Gabas Jr 1989: 16, Galucio et al. 2015: 257–258
lying	**joP*	*to(o)p-∅ / to-a*	*ty-syp*	*(*-mboP)*	*(-mbop)*
sitting	**jẽ*	*yẽ-t / y-ã*	*ty-j̃a*	**-jə̃*	*-yã*
standing	**ja*	*ta-t / ta-a*	*ty-so*	**-ja*	*-ya*

Table 3. Tuparikém auxiliaries and Rama-Puru lexical verbs

In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class II relational stem. This is clearly seen in the Khĩsêtjê reflex: the finite stem *ta* is absolute, and the nonfinite stem *-tãm* takes the full set of the person prefixes (1 *i-tãm*, 2 *a-tãm*, 3 *s-ãm*), where *-t-* is a thematic consonant. In Tupian, the morphosyntactic behavior of **-ja* and **-ʔãP* cannot be reconstructed with certainty. The former is reflected as an auxiliary in the Tuparikém languages, where it combines with other morphemes (such as *-t* 'present' and *-a* 'past' in Sakurabiat; *ty-* 'imperfective' in Karitiana), whereas the Karo and Puruborá reflexes are only marginally attested in the available data. The latter is mostly known from Mawé–Guaranian languages, where the reflexes are active class I intransitive verbs. Therefore, there is a class membership mismatch between the PMJ class II stem **-ja-m* and the PT class I stem **-ʔãP*.

'name': PMJ ***-jet** : PT ***-jeT**

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 405. The root is preserved in Karajá (*nii*) and in many languages of the Eastern branch (Maxakalí *-xu-xet-'ax*/*-ã-xet-'ax*, Khisêtjê *-nhinti*, Xavante *-nhitsi*, Kaingang *-jiji*). In all these languages, the root is preceded by a prefix whose PMJ shape is difficult to reconstruct: Karajá and the Cerrado languages point to PMJ **-ñĩ-jet*, the Southern Jê languages to **-ji-jet* or maybe **-jy-jet*, and Maxakalí shows an alternating pair of prefixes, whose choice depends on the syntactic context. The coda **-t* is reconstructed based on the correspondence between Maxakalí /-t/ and Jê zero. The semantic equivalents in Chiquitano (**-tsɨri* / **-ɨri*), Ofayé (*-xirêʔ*), and Krenak (*-unĵak*) show some superficial resemblance to PMJ **-jet*, but are unlikely to be cognate due to lack of regular sound correspondences.

Proto-Tupian **-jeT* is preserved in most Tupian languages, including the Tuparikém (Makurap *-xet*, Karitiana *-sat*), Mondé (Paiter *-léd*), and Eastern (Sateré-Mawé *-set*, Apyãwa *ter-a* / *-rer-a*) branches; see Galucio et al. 2015: 261 for a selection of reflexes. The PT reconstruction is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian **-tʲet*, or—in my notation—**-ceT* (Meira & Drude 2015: 294) and Proto-Tuparian **-jeT* (Nikulin & Andrade 2020: 295). The correspondences are regular, with the possible exception of probable exceptions in the Juruna languages, such as Yudja *-zá* (the regular reflex of PT **-T* is Yudja *l* /ɮ/, not *z*).

In Tupian, the root is reconstructed as a class II relational stem. In Macro-Jê, it is always accompanied with derivational prefixes, and the inflectional properties of the bare root are thus not recoverable.

'father': PMJ ***-jo₂m** : PT ***-joP**

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 401. The root is preserved in most branches, including Western (Djeoromitxí *ho{txi}* 'father', *-ro* 'father; male', Rikbaktsa *-zo*, Ofayé *-xəw ~ -xôw ~ -xew*), Karajá (3rd person *t-by*), and Eastern (Canela–Krahô *-xũm* 'male',

3rd person *h-ũm* ‘father’, Kaingang *-jóg*). The Correspondences are regular, except that the Ofayé reflex shows irregular vowels in Eduardo Ribeiro’s (*ə*) and Sarah G. Gudschinsky’s (*e*) attestations. In addition, no traces of the coda **-m* are seen in the alleged cognates in the Northern Jê languages of the Trans-Tocantins subgroup: Apinajé *-xũr*, Mẽbêngôkre *djũnũ* or *djũn-wa* ‘father (voc.)’, Kajkwakhrattxi and Khĩsêtjê *turê* ‘father (voc.)’, though the Mẽbêngôkre term for male — 3rd person *∅-ũm-ti-re* — does show the expected *-m*. In Karajá and in the Northern Jê languages of the Timbira branch, only the third-person form (PMJ **c-o₂m*) is used in the meaning ‘father’, though the Timbira languages preserve the uninflected form **-jo₂m* in the meaning ‘male’.

Proto-Tupian **-joP* ‘father’ is preserved in most Tupian languages, including Kepkiriwat (‹xuá›), Tuparikém (Wayoró *-ndop*, Karitiana *-syp* ‘father of a woman’), Mondé (Paiter *-lob*), and Eastern (Yudja *-pá*, Kuruaya *-lop*, Awetí *tup*/*-up*, Apyãwa *tow-a*/*-row-a*). The reconstructed form is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian **-tʲup*, or — in my notation — **-cuP* (Meira & Drude 2015: 293) and Proto-Tuparian **-joP* (Nikulin & Andrade 2020: 295). The correspondences are regular; Alves (2004: 180) documents Tuparí *-hòp*, with an unexpected long vowel (symbolized by means of a grave accent in the practical orthography), but the expected form with a short vowel is attested in Singerman 2018: 50. There is also a homonymous stem PT **-joP* ‘fish roe, pus’, whose reflexes have at times been claimed to belong to the same etymology as **-joP* ‘father’ (cf. Meira & Drude 2015: 293); its reflexes are found in the Rama-Puru (Karo *-xop* ‘dirt’, Puruborá *-tɔP* ‘fish roe’), Mondé (Paiter *-lób* ‘pus’), and Eastern (Sateré-Mawé *-sup* ‘sperm’, *win sup* ‘fly maggots’, Apyãwa *ipira-ow-a* ‘fish roe’) branches.

In both language families, the root is reconstructed as a class II stem, with the following provisos. In Macro-Jê, it appears to have shifted to class I in Ofayé (3rd person *ə̃-xəw ~ ə̃-xôw* instead of the expected **h-əw ~ *h-ôw*; Oliveira 2006: 97; Ribeiro n/d). In Tupi–Guaranian, **tuP*/**-ruP* belongs to the so-called subclass IIb, which includes a handful of kinship terms; it differs from other class II subtypes in having a third-person form identical to the absolute one (**tuP* ‘her/his father’). It thus contrasts with nouns such as **-ruP* ‘fish roe’, whose third-person form is reconstructed in my proposal as **θ-uP* (> Apyãwa *h-ow-a*).

‘pus’: PMJ ****-jo₂w°*** : PT ****-joP*** ‘fish roe, pus’

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 401. The root is preserved in the Western (Djeoromitxí *-ro ~ -ro{o}* ‘sap, pus, mucus’) and Eastern (Maxakalí *-xapa*, Khĩsêtjê *-tu{ru}*, Xavante *-dzub{rui}* // *-dzub{ru}*, Kaingang *{f}o* ‘pus’, *{f}ó-m* ‘to suppurate’) branches. The correspondences are mostly regular, except that the Akuwẽ reflexes show an unexpected palatal coda in the utterance-medial allomorph **-ĵubruj*. In addition, the PMJ coda **-w°* is reconstructed exclusively in order to account for the correspondence PJ **-P* ~ Maxakalí *-pV*. If the Maxakalí datum turns out to be noncognate, the PMJ reconstruction can be updated to **-jo₂p*. The Cerrado languages reflect a derived form, **-jup-r*, which can be interpreted as an erstwhile nonfinite form of the verb ‘to suppurate’. The Southern Jê languages reanalyzed the third-person index **c- > *θ-* as a part of the stem (Ribeiro 2004a: 95).

Proto-Tupian **-joP* ‘fish roe, pus’ is reflected in the Rama-Puru (Karo *-xop* ‘dirt’, Puruborá *-tɔP* ‘fish roe’), Mondé (Paiter *-lób* ‘pus’), and Eastern (Sateré-Mawé *-sup* ‘sperm’, *win sup* ‘fly maggots’, Apyãwa *ipira-ow-a* ‘fish roe’) branches. The correspondences are regular. At least in Paiter, it contrasts with the nearly homonymous term for ‘father’ in having high tone (Bontkes 1978: 5), suggesting that the tonal contrast was also present in PT. In this study, I do not make an attempt at reconstructing PT tone.

In both language families, the root is reconstructed as a class II stem, as evidenced by the third-person forms such as Khĩsêtjê *s-u{ru}* and Apyãwa *h-ow-a*.

‘tooth’: PMJ ***-juñ°*** : PT ***-jãC***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 403. The etymon is preserved in all Macro-Jê branches, including Chiquitano (e.g. Bésɨro *-só’o*), Western (Djeoromitxí *hü*, Ofayé *-xe:ʔ*), Karajá (*juu*), and Eastern (Khĩsêtjê *-twa*, Xavante *-’wa*, Kaingang *-jã*, Maxakalí *-xox*). The reflexes in all daughter languages are regular, except that Xavante *-’wa* and Akwẽ-Xerente *-kwa* unexpectedly lack utterance-medial allomorphs with a palatal coda (Xavante *-’wai*, Akwẽ-Xerente *-kwai*), or at least such allomorphs have not been attested in the literature. Possibly the utterance-final allomorph, which regularly loses the underlying palatal coda, has been generalized in the history of the Akuwẽ languages (see Nikulin 2017: 155–158 on utterance-medial and utterance-final allomorphs in Akuwẽ). The palatal nasal coda followed by an echo vowel is reconstructed based on the reflexes in the Maxakalian languages (Maxakalí *-xox* and Pataxó-Hãhãhãe ‹-tei›, ‹-tóy›, ‹-ˈtʰʊi›, ‹-txũi› point to a palatal coda), as well as in Krenak (*-jun*, with *n* clearly going back to PMJ **ñ* or **ñ°*), Pykobjê–Krĩkatí (*-xwaa*, with the long vowel suggesting an erstwhile **-n°* or **-ñ°*), and Ofayé (*-xe:ʔ*, with the plural and diminutive forms attested in Oliveira 2006: 79 strongly suggesting the presence of an underlying nasal coda).

Proto-Tupian **-jãC* is preserved in all branches of Tupian, including Kepkiriwat (‹nhain›, ‹-nhai-›), Tuparikém (Makurap *-yãy*, Karitiana *-j̃oj̃*), Rama-Puru (Karo *-yãy*), Mondé (Zoró *-jẽẽj*), and Eastern (Munduruku *-nũy*, Apyãwa *tỹj-a*/*-rỹj-a*); see Galucio et al. 2015: 254 for a selection of reflexes. The correspondences are completely regular, except that those Mondé languages that preserve this etymon — Aruá, Gavião, and Zoró — unexpectedly show a long front vowel /ẽ:/ as the reflex of PT **ã*.

In both families the stem is reconstructed as relational, class II. This is clearly seen in the third-person (singular) forms, with no thematic consonant: Bésɨro *∅-o’ó-xɨ*, Karajá *tx-uu*, Khĩsêtjê *s-wa* < PMJ **c-uñ°*; Makurap *t-ãy*, Munduruku *t-ũy*, Apyãwa *h-ỹj-a* < PT **c-ãC* (Ribeiro 2012a: 119; Santos 1997: 39; Braga 2005: 50; Picanço 2005: 262; Almeida et al. 1983: 26–27).

‘to ingest’ = ‘to eat/drink’: PMJ ***-ko₂*** : PT ***-ɨko***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 410. Reflexes are found in most Macro-Jê branches: Western (Djeoromitxí *-ko* ‘to eat’, Ofayé *-hô* ‘to eat something solid’), Karajá (*-ky* ‘to eat grains’), and Eastern (Khĩsêtjê *-khu* ‘to eat.PL’, Xavante *-hu* ‘to ingest.PL’, Kaingang *-ko* ‘to eat, to use’). The correspondences are regular. Rikbaktsa *-ku* ‘to drink’ is viewed as a reflex of PMJ **-ko₂* in Nikulin 2020, but it could be alternatively considered cognate with Proto-Goyaz **ij-kõ* (nonfinite **-kõ-m*) ‘to drink’ (> Khĩsêtjê *i-khõ*, *-khõm*).

The Proto-Tupian reconstruction is mentioned *in passim* in Nikulin & Carvalho 2022: 16; see Galucio et al. 2015: 256 for a selection of reflexes. The root is preserved in most branches of Tupian, including Kepkiriwat (‹-qu-›), Tuparikém (Tuparí *-ko*, Karitiana *-’y*), Rama-Puru (Karo *-’o*, Puruborá *-ʔɔ*), and Eastern (Munduruku *-’o*, Apyãwa *-’o*). The correspondences are regular.

In both language families, the root is a class I stem. In Old Tupí and possibly some other TG languages, this verb is unusual in that it does not take the third-person accusative prefix *îo-* when finite (Barbosa 1956: 305). In the languages of the Cerrado branch of the Jê group, the verb **-ku* takes indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is **-ku-r’*. Note that in almost all Tupian languages the reflexes cover the entire semantic domain of eating and drinking; in Macro-Jê, this is synchronically the case in the Akuwẽ languages (compare Xavante *-hu* ‘to eat.PL’ and *ö-hu* ‘to drink.PL’).

‘tree, tree-like object (leg, horn, bone)’: PMJ ****(-)kɣ₁m°*** : PT ****(-)kɯP***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 409. The root is preserved in all first-level branches of Macro-Jê, including Chiquitano (Bésɨro *-{tápa}kɨ* ‘horn’), Western (Djeoro-

mitxí *ku* ‘tree’, *{me}ku* ‘horn’, Ofayé *həuɥ* ‘tree’, *-həuɥ* ‘horn’), Karajá (female speech *kòò*, male speech *òò* ‘wood, horn’), and Eastern (Maxakalí *-kup* ‘stick, bone, leg’, *-ptox-kup* ‘horn’, Khĩsêtjê *khô* ‘club’, *-khô* ‘grove’, Xavante *-õmo* // *-u* ‘horn’, Kaingang *ka* ‘tree’, *-{nĩ}ka* ‘horn’). The reconstruction of a labial nasal coda followed by an echo vowel is based on the evidence from the Akuwẽ languages (Proto-Akuwẽ *-*kõmõ* // *-*ku* ‘horn’) and corroborated by Maxakalian, which preserves its place of articulation. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31; see Galucio et al. 2015: 252 for a selection of reflexes. The root is preserved in all first-level branches of Tupian, including Kepkiriwat (‹quêp›/‹queb-› ‘tree, wood’, ‹-nécubá› ‘elbow’, ‹cü-ümarã› ‘leg garter’, ‹ócüpe› ‘stud’, ‹-mbátoquêp› ‘index finger’, ‹jaácupe› ‘maize cob’), Tuparikém (Makurap *kup* ‘tree’, *-api-kup* ‘horn’, Karitiana *’ep* ‘tree’, *-’ep* ‘bone’), Rama-Puru (Karo *ma-’ûp* ‘tree’, Puruborá *ʔɨᴘ* ‘tree’), Mondé (Paiter *ihb* ‘tree’), and Eastern (Yudja *epá* ‘stick’, Mundurukú *íp* ‘tree’, *-’ip* ‘tree/wood (classifier)’, Sateré-Mawé *aria-’yp* ‘tree’, *-’yp* ‘tree (of a concrete species), handle’, Apyãwa *-’yw-a* ‘leg, handle, tree (of a concrete species)’); see Galucio et al. 2015: 252 for a selection of reflexes. The correspondences are regular.

In both language families, the root occurs both as a class I relational stem and as an absolute stem. It is thus reconstructed as relationally labile (i.e., the possessor is optional). Note the closely matching semantics of the reflexes in Macro-Jê and Tupian: ‘tree’ is the most recurring meaning, but ‘leg’, ‘horn’, and ‘bone’ are also attested across both families.

‘liver’: PMJ ***-mbâ*** : PT ***-pɨ(-)ʔa / *mbɨ(-)ʔa***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 375. The root is preserved at least in the Western (Djeoromitxí *-bä*, Rikbaktsa *-py*, Ofayé *-ɸa(h)*), Karajá (*maa*), and Eastern (Khĩsêtjê *-mba*, Xavante *-pa*, Kaingang *-tỹ-mẽ*) branches. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32, where the element *-*ʔa* is given as a part of the root. It is semantically and morphologically plausible to analyze *-*ʔa* as a formative for spherical objects. The root is preserved in most first-level branches, including Kepkiriwat (‹-piá›), Tuparikém (Makurap *-pia*), Rama-Puru (Karo *-pía*, Puruborá *-bia*), and Eastern (Yudja *-bï’á*, Mundurukú *-psà*, Sateré-Mawé *-py’a*/*my’a*, Apyãwa *-py’ã-∅*/*my’ã-∅*); see Galucio et al. 2015: 256 for a selection of reflexes. The correspondences are regular.

In Macro-Jê, the reflexes of *-*mbâ* belong to class I. In Tupian, *-*pɨʔa* is reconstructed as a relational class I stem, and *mbɨʔa* as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. The erstwhile absolute stem *mbɨʔa* is preserved in the Mawé–Guaranian languages but was apparently lost in all other branches.

‘smoke’: PMJ ***-ñĩjə̂k*** : PT ***-jĩːᴋ***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. The etymon is preserved in the Paraná and Akuwẽ branches of the Jê groups (Xavante *-nhidzé*, Akwẽ-Xerente *-nĩze*, Laklãnõ *nẽjó*, Kaingang *nĩja*) as well as in Karajá *he-dà* (from *hèè* ‘firewood’). The reconstruction of a stem-final velar stop is supported by the Kaingang derivative *nĩjãg* ‘to produce smoke’. Karajá *hedà* (Palha 1942: 25; Ribeiro 2012a: 105) is not the main term for ‘smoke’ in the modern language, where *wàdàsi* ‘smoke’ is found instead.

Proto-Tupian *-*jĩːᴋ* is preserved in most Tupian languages, including Kepkiriwat (‹iá-in›), Tuparikém (Wayoró *-yĩing*, Karitiana *-ĩjing*), Mondé (Paiter *mokây-ñiĩ̃*), and Eastern (Mundurukú *-diĩ̃*, Sateré-Mawé *y-hiĩ̃*, Awetí *taza-ting*, Kawaiwete *tata-sing*); see Galucio et al. 2015: 259 for a selection of reflexes. The reconstructed form is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian *-*tʲiŋ*, or—in my notation—*-*ćĩᴋ* (Meira & Drude 2015: 294), Proto-Mundurukuan *-*ðiŋ* (Picanço 2019: 139), Proto-Tuparian *-*ɲĩːᴋ* (Nikulin & Andrade 2020:

296), with the reconstruction of a long vowel based on evidence from Tuparian languages, such as Wayoró and Sakurabiat. The correspondences between these forms are mostly regular. The denasalization of **ĩ* in Mundurukuan could be regular, as the sequence ***ðĩ* was banned in Proto-Mundurukuan (Picanço 2005: 173). So could be the second stage of the purported development **jĩ* > **cĩ* > **ćĩ* in Proto-Mawé–Guaranian.[11] Somewhat problematic are the alleged reflexes in the Juruna languages (such as Yudja *-xi'ã* < Proto-Juruna **-ʃi-ʔã*). Nikulin and Andrade (2020: 296, fn. 30) discuss several difficulties with the reflexes in individual Tuparian languages. Finally, an irregular reflex of PTG **-K* is seen at least in the Apyãwa form *tata-xin-a* /tãtã-tɕĩT-a/.

Both in Macro-Jê and Tupian, the stem is reconstructed as relational, class II. This is clearly seen in the third-person forms, with no thematic consonant: Xavante *{ĩ}ts-idzé* < PJ **c-ĩjə̂*$_2^K$ (Lachnitt 1987: 79); Mundurukú *t-iɡ̃* < PT **c-ĩ:K* (Picanço 2005: 320).

This comparison deviates from my stringent criteria in that the PMJ sequence **jə̂* is not matched to any PT segment. However, the fact that the PT form is reconstructed with a long vowel makes the comparison somewhat more plausible: it is easy to imagine a contraction of an **ĩjV* sequence into **ĩ:*.

<u>'feces'</u>: PMJ ***-ñṼt°** : PT ***-jṼT**

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 407. The etymon is preserved in Chiquitano (Bésɨro *-a'a*) and in the Eastern branch (Maxakalí *-yõn* 'feces, to defecate', Khĩsêtjê *-nhin* // *-nhini*, Xavante *-nhana*). Note that in multiple Macro-Jê languages the reflexes of **-ñṼt°* are polysemous and can refer not only to feces, but also to bowels (Mẽbêngôkre, Parkatêjê, Pykobjê–Krĩkatí, Canela–Krahô, Xavante); in other languages, terms for 'bowel' or 'small bowel' are derived from the respective root (Bésɨro *-an-terere*, Mẽbêngôkre *-nhĩn kra*, Parkatêjê *-jĩn-kra*, Pykobjê–Krĩkatí *-jẽhn cra*, Canela–Krahô *-jĩn kra*, Akwẽ-Xerente *-nnã hi-rê*). The correspondences involving the consonants are regular, except that the Xavante and Akwẽ-Xerente reflexes unexpectedly lack utterance-medial allomorphs with a voiceless stop coda (Xavante **-nhatã*, Akwẽ-Xerente **-ntã*), possibly due to analogy with the regular utterance-final allomorphs *-nhana* [-ˈɲãːnã] / *-nnã* [-nʱˈnã]. By contrast, the vowels across Macro-Jê show no regular correspondence whatsoever. Maxakalí *õ* /ũ/ points to PMJ **ũ*; Khĩsêtjê *i* /ĩ/ < Proto-Goyaz **ĩ* suggests PMJ **ĩ*; Xavante *a* /ə̃/ < Proto-Akuwẽ **ə̃* can go back to PMJ **ə̃* or **ỹ*. The Chiquitano reflex shows a nasal vowel /ã/ in the Migueleño and Eastern varieties; the Macro-Jê origins of Proto-Chiquitano **ã* have not been established yet, but it could technically be the regular reflex of PMJ **ə̃* or **ỹ*.

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém and Eastern branches, including Wayoró *-yẽn* (< Proto-Tuparian **-ɲẽT*; Nikulin & Andrade 2020: 296),[12] Karitiana *-ĩjin* (< Proto-Arikém **-ɲĩT*), Mundurukú *-nũn* (< Proto-Mundurukuan **-ðãn*; Picanço 2019: 139), Xipaya *-súna*, Yudja *unã*, and Sateré-Mawé *-ĩjun*. Just like in Macro-Jê, many of its reflexes either colexify 'feces' with 'bowel' (e.g. Wayoró *-yẽn*) or use derivatives of the root in question in the meaning 'bowel' (Karitiana *-ĩjin-py*, Mundurukú *-nũn-pú* < PT **-jṼT-pə*). In addi-

[11] PT **j* normally yields Proto-Mawé–Guaranian **c* (> Sateré-Mawé *s*, Awetí and PTG zero word-medially); the reflex **ć* (> Sateré-Mawé *h*, Awetí *t*, PTG **t* word-medially) is otherwise known to occur following an **i* or a *C by progressive palatalization. But the sequence **ci* is not reconstructed for any Proto-Mawé–Guaranian morpheme (at least in Meira & Drude 2015), and may have been subject to regressive palatalization in pre-Proto-Mawé–Guaranian.

[12] Tuparian has a similar root **-ɲẽ:T* (also **kɨ-ɲẽ:T*) 'ashes', which, however, must be unrelated to **-ɲẽT* 'feces' (*pace* Nikulin & Andrade 2020: 296), since its reflexes are documented with a long vowel in most daughter languages (Galucio et al. 2015: 259).

tion, PTG **tũT/*-rũT* ‘black’ (**-ũT* in compounds, as in **-piθ-ũT* ‘black skin’) regularly corresponds to Sateré-Mawé *-j̃un* ‘feces’. It is reflected, for example, as Apyãwa *-ron*, *-pi-on*; Siriono *-rö* ‘muddy’, *-ɨ-sö* ‘dark’, etc. Despite the semantic discrepancy, the evolution ‘feces’ > ‘dirty’ > ‘black’ seems feasible. The correspondences involving the consonants are regular. However, the vowels correspond in a unique way in this cognate set: Juruna and Mawé–Guaranian point to Proto-Tupian **-jõT*, Mundurukuan to **-jãT*, Tuparian to **-jẽT*, and Arikém to **-jĩT*.

In both families the stem is reconstructed as relational, class II. This is clearly seen in the third-person forms, with no thematic consonant: Khĩsêtjê *s-ĩn* // *s-ĩni*, Xavante *ts-ãna* < PMJ **c-ṼT*; Mundurukú *t-ũn* < PT **c-ṼT* (Nonato et al. 2012: 7; Lachnitt 1987: 74; Picanço 2005: 151). The fact that the vowel correspondences are highly irregular both in Macro-Jê and Tupian can be possibly accounted for by reconstructing a low-frequency nasal vowel for both protolanguages. The colexification of the meanings ‘feces’ and ‘bowel’, found in both language families, renders the cognation hypothesis particularly plausible.

‘earth’: PMJ ****ŋgyN°*** : PT ****ƙuc***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 418. The etymon is preserved in the Chiquitano (Bésɨro *kɨ́-xɨ*), Western (Ofayé *həuɲẽʔ*) and Eastern branches (Kaingang *ga*). The Ofayé reflex, not listed in Nikulin 2020, points to a nasal coda, as suggested by the plural form *hɔ:-ñe* and the allomorph *hət-*, found in compounds (Oliveira 2006: 79). PNJ **ŋgyᵊ* ‘clay, mud’ could be related, but the origin of the diphthong **yᵊ* is unclear; the regular reflex of PMJ **yN* would be PNJ **ô:* or **ô:*. Karajá *sùù* (underlying /θʊ/) does not appear to be cognate with the aforementioned forms, since PMJ **y* is normally reflected as *ò* /ɔ/ in Karajá; the reflex of PMJ **ŋg* in Karajá is presently unknown (but **ŋgr* is indeed regularly reflected as *s* /θ/ in Karajá).

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31. Its reflexes are found in the Kepkiriwat (‹cuitá-á›, ‹queitaá›), Tuparikém (Makurap *kux*, Wayoró *kuy*, Karitiana *’ej*), Rama-Puru (Puruborá *ʔɨc*), and Eastern branches (Yudja *etá* ‘sand, beach’, Sateré-Mawé *’yi*, Apyãwa *yj-a*).

In both Macro-Jê and Tupian, the reflexes of PMJ **ŋgyN°* and PT **ƙuc* are typically absolute (unpossessable) nouns, though in some languages they are optionally possessed and behave as class I relational stems, as in Bésɨro *n-í-kɨ ma-monkó-ka* ‘Chiquitanía’ (literally ‘the land of the Monkóxɨ nation’).

‘arm’: PMJ ****-pa*** ‘arm, branch’ :
PT ****-pə / *mbə*** ‘hand, vine-like’, ****-pə-ʔa / *mbə-ʔa*** ‘arm’

The Proto-Macro-Jê reconstruction adopted here differs slightly from the one in Nikulin 2020: 369, where it is reconstructed as **-paj ~ *-paj°*. Reflexes are found in most branches, including Chiquitano (Bésɨro *-pa* ‘arm, wing’), Western (Djeoromitxí *{ha}pa* ‘arm’, *ku-{ra}pa* ‘branch’, Rikbaktsa *-pa-* ‘arm (in compounds)’, *-{tsi}pa* ‘arm’, *-sara-pa* ‘branch’, Ofayé *-ɸe* ‘arm, wing’), and Eastern (Khĩsêtjê *-hwa* ‘arm, branch’, Xavante *-pai-hi* ‘arm’, *-pa-nõ* [-paˈnːõ] /-paj-dõ/ ‘arm’, Kaingang *-pẽ* ‘arm’, *ka pẽ* ‘branch’). The only reason for reconstructing a palatal coda are the reflexes in the Akuwẽ languages: Xavante *-pai-hi* ‘arm’, *-pa-nõ* [-paˈnːõ] /-paj-dõ/ ‘arm’, Akwẽ-Xerente *-pai-nõ* ‘arm’. However, no palatal coda is found in Xavante *pa* ‘creek’, *-pa* or *wede-pa* ‘branch, root’; Akwẽ-Xerente *-pa-krta* // *-pa-krda* ‘arm’, *-pa* or *wdê-pa* ‘root’. Therefore, the grounds for reconstructing a palatal coda in PMJ are rather weak. The meanings ‘arm’ and ‘branch’ were probably colexified in PMJ **-pa*, as shown by evidence from Jabutian, Rikbaktsa, and Jê. The meaning ‘wing’, seen in Chiquitano and Ofayé, is probably innovative, since a distinct root for ‘wing, armpit’ is otherwise reconstructed (PMJ **-jar°*; Nikulin 2020: 399).

The Proto-Tupian reconstruction *-*pə* / **mbə* is from Nikulin & Carvalho 2022: 31. The correspondences are regular. Reflexes are found in all branches of Tupian, including Kepkiriwat (*-mbo* 'CL:long': ‹umbó› 'my guts', ‹uhembó› 'my neck', ‹boi uarumbó› 'anaconda'), Tuparikém (Wayoró *mbo* / *-wo* 'hand', Karitiana *-py* 'hand'), Rama-Puru (Karo *=pĩ'* 'CL:cylindrical+small', Puruborá *-bə* 'CL:vine-like'), Mondé (Paiter *-pá-be*), and Eastern (Yudja *-wá* 'hand' < Proto-Juruna **-bu-á*, Mundurukú *-pu* 'hand, finger; *CL*:vine-like', Sateré-Mawé *-po*/*mo* 'hand', *-po-'yp* / *mo-'yp* 'arm', Apyãwa *-pa-∅* / *ma-∅* 'hand'). The reflexes in languages such as Kepkiriwat, Karo, Puruborá, and Mundurukú clearly show that PT *-*pə* occurred not only as a body part term, but also as a second element in compounds designating long, vine-like objects, such as vines (PT **ɯtɯ-pə*), roots (PT **-ja-pə*, Eastern branch only), and possibly threads, snakes, cords, fingers, etc. The term for 'arm' is reconstructed as *-*pə-ʔa* / **mbə-ʔa*, whose second element appears to be *-*ʔa* 'head, *CL*:spherical'; it has known reflexes in Rama-Puru (Karo *-pá-be* 'hand', Puruborá *-ba* 'arm') and Eastern branches (Mundurukú *-pà* 'arm; *CL*:cylindrical+thick' < Proto-Mundurukuan *-*pa̱*; Picanço 2019: 136). The reflexes in Mundurukuan clearly point to PT **V*ʔ*a*, and the quality of the vowel that precedes the glottal stop is inferred based on the possible morphological relation to *-*pə*/**mbə*. Unlike in Macro-Jê, Tupian shows a distinct root for 'branch', PT **-jãŋã* (Wayoró *kuw-angã* 'branch', *mbo-angã* 'wrist'; Karitiana *-j̃ongõ ~ -j̃ongo* 'arm, branch'; Mundurukú *-dákʉ́ ~ -nákʉ́* 'branch'; Apyãwa *-rakỹ-∅* 'branch'). There is also an alternate candidate for the main term for 'arm', PT *-*nẽ*, with reflexes in Kepkiriwat, Tuparian, Arikém, Mondé, and Mundurukuan (compare also PT *-*nẽ-pɨ* 'armpit').

In Macro-Jê, the reflexes of *-*pa* belong to class I. In Tupian, *-*pə* is reconstructed as a relational class I stem, and **mbə* as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. In Kepkiriwat and Tuparian, the erstwhile absolute stem **mbə* > **mbo* 'hand (unpossessed)' was apparently reanalyzed as relational, whereas Karo and possibly some other languages have lost the allomorph **mbə* entirely.

'foot': PMJ ***-pâr°*** : PT ***-pɨ* / *mbɨ***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 370. Reflexes are found in the Western (Rikbaktsa *-pyry*, Ofayé *-ɸar*) and Eastern (Maxakalí *-pata*, Khĩsêtjê *-hwaj* // *-hwaji*, Xavante *-para*, Kaingang *-põn*) branches. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32. The root is preserved in all first-level branches, including Kepkiriwat (‹-mbi› 'leg', ‹-mbitecaiã› 'heel'), Tuparikém (Makurap *-mi*, Karitiana *-pi*), Rama-Puru (Karo *-pi-be*', Puruborá *-ʃi-bɛ*), Mondé (Paiter *-pí-pe*), and Eastern (Mundurukú *-i*, Sateré-Mawé *-py*/*my*, Apyãwa *-py-∅*/*my-∅*); see Galucio et al. 2015: 255 for a selection of reflexes. The correspondences are regular.

In Macro-Jê, the reflexes of *-*pâr*° belong to class I. In Tupian, *-*pɨ* is reconstructed as a relational class I stem, and **mbɨ* as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. In Kepkiriwat and Tuparian, the erstwhile absolute stem **mbɨ* > **mbi* 'foot (unpossessed)' was apparently reanalyzed as relational, whereas Arikém, Mundurukuan, and some other languages have lost the form **mbɨ* entirely.

This comparison deviates from my stringent criteria in that a PMJ coda is not matched to any PT segment. However, the correspondences are otherwise recurrent, and the semantic match is perfect; the rhotic codas in PMJ are in any case infrequent.

'to burn, to set on fire': PMJ ***(-)py₁k° ~ *(-)py₁ŋ°*** : PT ***-pɯK***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 371. The etymon is preserved in two first-level branches of Macro-Jê, Western (Rikbaktsa *-pok* 'to set on fire') and Eastern (Maxakalí *-puk* 'to burn (intr.)', Canela–Krahô *pôr*, nonfinite *-hpôc* 'to burn (intr.)'). The correspondences

are regular, including the sound change *-*k* > *-*r* in finite forms of intransitive verbs, typical of the Cerrado languages (cf. Nikulin & Salanova 2019: 544). A difference in valency between the Rikbaktsa verb and its Eastern Macro-Jê cognates is a problem for the comparison, though hardly an insurmountable one. The uncertainty between the reconstruction of *-*k*° or *-*ŋ*° is due to the absence of a cognate in Krenak, the only Macro-Jê language that is known to preserve the distinction (cf. Nikulin 2020: 159).

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 34. The etymon is preserved in three branches of Tupian: Tuparikém (Wayoró *-pug{a}* 'to cook'), Rama-Puru (Karo *-pâk* 'to burn'), and Eastern (Mundurukú *-pik* 'to burn').

In all said languages, the verb is a relational class I stem, except for the finite form in the Cerrado languages (finite intransitive verbs are absolute).

3CRF prefix: PMJ ****ta-*** : PT ****tə-***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 253–260, 383. The etymon is preserved in Karajá (*ta-* with class I stems, *t-* with class II stems) and in two languages of the Western branch (Rikbaktsa *ta-*, Arikapú *ta-*). The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 383. PT **tə-* is preserved in the Tuparikém branch (Wayoró *te-*, Karitiana *ta-*) and in at least one Eastern language (Sateré-Mawé *to-*); a possible reflex with an unexpected vowel is also seen in the Rama-Puru branch (Karo *to-*). The correspondences are otherwise regular. In addition, Mondé and Awetí–Guaranian languages have 3CRF indices that point to PT **ə-* rather than **tə-* (Awetí *o-*/*w-*, Gavião *a-*; Sabino 2016: 71–72, 146; Moore 1984: 30), a fact I am presently unable to account for.

In both language families, the morpheme in question is a third-person index which signals coreferentiality with another participant (typically the subject). In all languages where it occurs, it can encode the possessor of a relational noun, but in some languages it can also encode the patient of a transitive verb or a complement of an adposition (as in Rikbaktsa), or else the subject of a intransitive verb (as in Wayoró), or of a subclass of intransitive verbs (as in Arikapú). In the latter use, the person index is taken to be coreferential with the noun phrase expressing the subject.

'to give': PMJ ****-ũp*** : PT ****-õP***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 422. Reflexes are found in most Macro-Jê branches: Western (Djeoromitxí *-õ*), Karajá (*-õ*), and Eastern (Khĩsêtjê *-ngõ*, Maxakalí *-hõm*). The labial stop in the coda position is reconstructed based on evidence from Maxakalí (*-hõm* /-hũP/), Krenak (*-um̥*), and the nonfinite form found in the Akuwẽ languages (Xavante *-nh-om-ri*). The correspondences are mostly regular, including the consonantal epenthesis in onsetless stressed syllables in Maxakalian and Jê; however, the origins of the voiceless nasal *m̥* in Krenak are unclear.

The Proto-Tupian reconstruction is mentioned *in passim* in Nikulin & Carvalho 2022: 16; see Galucio et al. 2015: 258 for a selection of reflexes. The root is preserved in the Tuparikém (Tuparí *-om*) and Eastern (Sateré-Mawé *-um*) branches. The Mundurukú reflex *-ũm*/*-g̃-ũm* unexpectedly shows an unrounded vowel /ɨ̃/ (represented as *ũ* orthographically), but the rounded reflex is found in the closely related Kuruaya (*-õm* / *-n-õm*; Galucio et al. 2015: 258). The correspondences are otherwise regular.

In both language families, the root is vowel-initial, with no thematic consonant, and is thus classifiable as class I. In the languages of the Cerrado branch of the Jê group, the verb *-*gõ* takes indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is a class II stem *-*ñ-õp-r*', with a thematic consonant and a

suffix of nonfiniteness. Since the verb typically takes an inanimate theme, it frequently occurs with a third-person index (PT **i-*) in Tupian languages, which typically takes a consonantal allomorph before a vowel-initial root.

‘to go up, to rise’: PMJ ***-we(C)*** : PT ***-we(ˀ)P***

The Proto-Macro-Jê reconstruction is given as **-wi(C)* in Nikulin 2020: 382, which is an unfortunate typo (cf. Nikulin 2020: 148): PNJ **i* can only go back to PMJ **e*. Reflexes are found in the Western (Ofayé *-wi*, possibly Djeoromitxí *{hu}wi*) and Eastern (Khĩsêtjê *a-pi*, nonfinite *-tá-pi-ri*). The correspondences are regular. The Northern Jê reflexes continue PNJ **aː-pi*, nonfinite **-jəː-pi-r*; the alternating prefixes **aː-* (finite) and **jəː-* (nonfinite), found in a number of intransitive verbs, are of unclear origin, but they are clearly distinct from the antipassive prefixes **a-*/**ap-* (finite) and **jə-*/**ju-* (nonfinite), which has a short vowel. The absence of clear cognates in diagnostic languages, such as Maxakali, Krenak, Xavante, or Akwẽ-Xerente, makes it impossible to determine whether the Proto-Macro-Jê verb had a final consonant. If Maxakalí *-ã-pep*/*-xu-pep* ‘to leave/arrive.SG’ is cognate, the PMJ reconstruction can be amended to **-wep ~ *-wem°*, but the semantic discrepancy renders the comparison uncertain.

The Proto-Tupian reconstruction is based on reflexes such as Wayoró *-ngwep* (< Proto-Tuparian **-weP*; Nikulin & Andrade 2020: 299), Karitiana *-haap* ‘to rise (of the sun)’ (< Proto-Arikém **-häːP*), Paiter *-web-á* ‘to swell’, and Awetí *-tep* (attested in Reiter 2011: 205). The correspondences are regular except for the mismatch between the short vowel in Tuparian and the long vowel in Arikém.

The class membership of Proto-Macro-Jê **-we(C)* is difficult to determine based on direct evidence: the Ofayé reflex is only marginally attested, whereas in other languages only a prefixed derivative was preserved. In Tupian, the verb is a class I stem.

4.2. Good distribution in Macro-Jê only

‘hole’: PMJ ***-kuñ°*** : Proto-Mundurukuan ***-kɰ̃j***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 411. Reflexes of the bare root are found in the Western (Djeoromitxí *-kü*) and Eastern (Maxakalí *-kox*, Canela–Krahô *kwa* ‘well, spring’) branches. In derivatives, such as the terms for ‘sky’ and ‘mouth’, it is preserved in even more languages, as in Karajá (female speech *biku*, male speech *biu* ‘rain, sky’), Khĩsêtjê (*-jajkhwa* ‘mouth’), or Kaingang (*jãnkã* ‘door’).

The Proto-Mundurukuan form, reflected in Mundurukú as *-kũ̀y* /-kɰ̃j/ and in Kuruaya as *-kɰ̃j*, is from Picanço 2019: 136. It lacks known cognates in other Tupian languages. If it is shown to be of Proto-Tupian original, its PT etymon must have been **-kãʔãC*, **-kãʔəC*, **-ǩãʔãC*, **-ǩəʔəC*, **-ŋãʔãC*, or **-ŋəʔəC*. Two alternative candidates for the Proto-Tupian term for ‘hole’ are **-ǩɨ̃PʔəC* and **-ƙaP*, but both have their distribution limited to two Rondonian branches only. The former has reflexes in Tuparikém (Tupari *-apsi-kúm’e* ‘inner ear’, *-kúm’e* ‘vagina’, Akuntsú *-api-tep-kɨ́mã* ‘inner ear’ with an irregular final vowel, Karitiana *emã ~ emmã* ‘pit’) and Mondé (Paiter *-îwã* ‘hole, buttock’). The latter has reflexes in Tuparikém (Karitiana *-’op* ‘hole, channel’ and Kepkiriwat (‹uapicáp› /u-api-kaP/ ‘inner ear’).[13]

In both Macro-Jê and Mundurukuan, the roots in question are relational class I stems.

[13] Otherwise, each Tupian branch employs its own root(s) for the meaning ‘hole’: Makurap *pun*; Tuparí *-áu’am*; Karitiana *-’op*; Karo *-xâk* ~ Puruborá *ʃɛK*; Aruá ‹ñiñap›; Proto-Juruna **-ku(-)á* and **karapú*; Sateré-Mawé *-kaʔa*; Proto-Awetí–Guaranian **-kʷaT*.

It must be noted that PMJ *-*kuñ*° shows similarity with yet another root, found in Mondé only: Paiter *-koy* in *gôy-koy* 'pit' (from *gõhy* 'earth'). This root cannot be cognate with Proto-Mundurukuan *-*kə̃j*, and could in principle be equated, at least etymologically, with the directional suffix *-koy* 'towards'.

'ripe': PMJ ***-ndêp°*** : Tuparí ***-tep***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 390. The etymon is best known for its reflexes in the Eastern branch (Maxakalí *tep-ta* 'banana', Khĩsêtjê *-ndep-txi* 'red', Apinajé *-nep* 'ripe', Canela–Krahô *-ntep-ti* 'ripe, red'), but a likely reflex is also found in the Western branch (Rikbaktsa *-{ne}ne* 'ripe').

Tuparí *-tep* is documented, for example, in Alves 2004: 257, 258. No cognates in other Tupian languages are known. It can technically go back to a variety of forms, such as PT *-*teP*, *-*təP*, *-*ndeP*, *-*ndəP*, *-*ðeP*, or *-*ðəP*. In many other Tupian languages, the concept 'ripe' is expressed by a reflex of *-*woP* 'red, ripe' instead: compare Wayoró *-ngop* 'red, ripe' (Nogueira et al. 2021: 103), Akuntsú *-kop* 'red, ripe' (Aragon 2014: 104, 131), Paiter *-ób* 'red, ripe' (Bontkes 1978: 14), Yudja *-upa* 'ripe' (Fargetti 2001: 281–283), Mundurukú *-op* 'ripe' (Crofts 1985: 99), etc. It is possible that *-*woP* was primarily used a color term, whereas the highly hypothetical form PT *-*teP*, *-*təP*, *-*ndeP*, *-*ndəP*, *-*ðeP*, or *-*ðəP* could have been a dedicated term for 'ripe', ousted in most daughter languages by reflexes of *-*woP*.

Both in Jê and Tuparí the stem in question is a relational class I stem. Maxakalí *tep-ta* is an absolute stem, and Rikbaktsa *-{ne}ne* is an intransitive verb (the language no longer has a class I / class II distinction).

'to kill': PMJ ***-wĩ*** : Karo ***-wĩ***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 383. Reflexes are found in the Western (Ofayé *-wĩ* 'to shoot'), Karajá (*-wè-* 'to sting, to penetrate, to stab'), and Eastern (Ritual Maxakalí *-mĩ-y*, irrealis *-mĩ* 'to kill', Maxakalí *-mĩ-y*, irrealis *-mĩ* 'to make', Khĩsêtjê *-pĩ*, nonfinite *-pĩ-rĩ* 'to kill.SG', Xavante *-wĩ*, nonfinite *-wĩ-rĩ* 'to kill.SG'). The correspondences are regular. The meaning 'to kill' is attested only in Ritual Maxakalí and in most languages of the Cerrado branch (singular only), whereas Ofayé, Karajá, and spoken Maxakalí all show deviant meanings. Even the Cerrado languages Canela–Krahô and Pykobjê–Krĩkatí do not use the reflexes of PMJ *-*wĩ* as the basic verb for 'to kill'; instead, they are used figuratively, e.g. as 'to extinguish a fire', 'to kill by drowning (of water)', 'to suffocate'.

The Karo verb *-wĩ* 'to kill' (Gabas Jr 1999: 48, 57) can technically go back to PT *-*wĩ*, *-*wĩC*, *-*wɨ̃*, or *-*wɨ̃C* (note that PT **ĩ* and **ɨ̃* merge in all Tupian languages except Juruna and Mawé–Guaranian, whereas **C* is deleted after a front high vowel in these languages; see the cognate set for 'heavy' in **4.3**). It is likely related to Puruborá *-wi* 'to kill' (Galucio et al. 2015: 257), but the absence of vowel nasality in the putative Puruborá cognate is unaccounted for. Karitiana *-mĩ* 'to beat' is technically comparable with Karo *-wĩ* 'to kill', given that **m* and **w* merge as *m* before nasal vowels in Arikém, but it could likewise be cognate with Proto-Tuparian *-*mĩ* (> Tupari *-mĩ* 'to stab, to sting', Sakurabiat and Akuntsú *-mĩ* 'to kill'); in the latter case, the Proto-Tuparikém form must be reconstructed as *-*mĩ*, thus showing no regular correspondence with Karo *-wĩ*. Alternatively, one could reconstruct Proto-Tupian *-*wĩ*, *-*wĩC*, *-*wɨ̃*, or *-*wɨ̃C* based on reflexes in the Rama-Puru and Tuparikém branches and posit an irregular sound change **w* > **m* in Proto-Tuparikém, Proto-Tuparian, or Proto-Core Tuparian. If such a verb existed in Proto-Tupian, it was likely distinct from PT *-*ʔaoka* ~ *-*ʔaoka̋* 'to kill, to beat', with reflexes in Kepkiriwat, Mondé, Mundurukuan, and Mawé–Guaranian, in that the latter

prototypically referred to beating to death, whereas the former probably referred to killing by stabbing or piercing (e.g. with an arrow), as suggested by the Tuparí reflex.

In Macro-Jê, the root is a class I stem. In the languages of the Cerrado branch of the Jê group, its reflexes take indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is PCerr *-*wĩ-r*'. Karo does not have a class I/class II distinction.

4.3. Good distribution in Tupian only

'bitter': PT ***-ðəP*** : PCerr ***-ndap*** 'sour, bitter'

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 27, with reflexes in the Tuparikém (Tuparí *-tép-'a* 'bitter', *-tép-'ut* 'sour', Karitiana *-taap*) and Eastern (Mundurukú *-cúp*, Sateré-Mawé *-nop*, Awetí *-lop*, Apyãwa *-rap*) branches, as well as possibly in Mondé (Paiter *-{pe}txáb*). The correspondences involving the nucleus and the coda are regular. However, the correspondence between Proto-Tuparian/Proto-Arikém **t* and Sateré-Mawé *n*, Awetí *l*, PTG **r* is unprecedented: Proto-Tuparian/Proto-Arikém **t* points to PT **t*, Sateré-Mawé *n* suggests PT **nd* (allophone of */n/), whereas Awetí *l* : PTG **r* is not known to go back to any specific PT consonant[14]. The Mundurukuan reflexes are uninformative, since **t* and **nd* are not otherwise distinguished in Mundurukuan. I reconstruct **ð* for the correspondence in question and assume that it was a low-frequency phoneme in Proto-Tupian, just like its apparent reflex /l/ in Awetí. However, other solutions are also imaginable, such as the existence of a hypothetical alternation between the allomorphs **-təP* (relational) and **ndəP* (absolute), with the subsequent generalization of the former in Tuparikém and of the latter in Mawé–Guaranian.

The Proto-Cerrado reconstruction is from Nikulin 2020: 456. Reflexes include Akwẽ-Xerente *-{wam}t(a)pa/-{wam}tap* 'bitter, sour', Mẽbêngôkre *-nap* 'sour', Pykobjê–Krĩkatí *-ntap* 'sour, ripe', and possibly Khĩsêtjê *-ndap* // *-ndawy* 'smooth'. The correspondences are regular. The semantic reconstruction is not straightforward: there are other terms for 'sour' and 'bitter' in the Goyaz languages, which are also quite old: Proto-Goyaz **-ĵwa* 'sour' and **-ĵə̂* 'bitter' (< PJ **-jə̂K* 'sour, salty', **-jə̂$_2$K* 'bitter'). If PCerr **-ndap* is shown to be of Macro-Jê origin, its erstwhile form should be reconstructed as PMJ **-ndap°* or **-ndâp°*[15].

Both PT **-ðəP* and Proto-Cerrado **-ndap* are reconstructed as relational class I stems.

'to do, to say, to be like this': PT ***-ɬe*** : PSJ ****kê // *ke***

The Tupian reconstruction is based on reflexes in the Tuparikém (Makurap *-ke*, Karitiana *-'a*), Rama-Puru (Karo *-'e*), and Eastern (Sateré-Mawé *-'e*, Apyãwa *-'ẽ/-e*) branches. The correspondences are regular, except that the Awetí–Guaranian branch has innovated some irregular inflected forms: the third-person form is reconstructed as **eʔi* (rather than the expected ***o-ʔe*), whereas the second-person singular form is attested as *e'i* (rather than **e-'e*) in Awetí and reconstructed as **ere* (rather than ***ere-ʔe*) in PTG. Although other forms are regular (PTG 1 **a-ʔe*, 1+2 **ja-ʔe*, 1+3 **oro-ʔe*, 2+3 **peʔj-e*), some daughter languages show extra irregularities, such as

[14] It is conceivable that Awetí *l* and PTG **r* are the regular reflexes of PT **nd*. Unfortunately, PT **ndo(:)* 'hill' and **-ndoK* 'to eat.INTR' lack known reflexes in these languages, making it difficult to determine the evolution pathways of PT **nd* in the Awetí–Guaranian branch.

[15] Other Macro-Jê languages show noncognate forms for 'sour' and 'bitter': Krenak *-rə* 'sour', *-ñãŋgrok* 'bitter'; Maxakalí *-xupyãg* 'sour', *-xũĩy* 'pain, sour, bitter, spicy'; Karajá 3 *tx-ubrèrè* 'sour'; Ofayé 3 *h-əʃə̂* 'sour', 3 *ə̃-xahtə* 'bitter'; Rikbaktsa *-bui* 'sour', *-sikpia ~ -spia* 'bitter'; Proto-Jabutian **-ǰombi* 'pain; sour', **-wəwə* 'bile' (whence Djeoromitxí *-wäwä-rü* 'bitter') or Arikapú *-oay ~ -way* 'bitter'; Proto-Chiquitano **ókor-* 'to be sour', **pičar-* 'to be bitter'.

the nasalization **e* > *ẽ* in the Apyãwa forms *ã-'ẽ*, *xa-'ẽ*, *ara-'ẽ* or the analogical vowel raising in Guarasugwe (*éɾɨ*, *ɨ-ʔɨ* instead of **ére*, **é-ʔɨ*), Kawaiwete (*a-'i* instead of **a-'e*), or Kamayurá (*i-'i* instead of **e-'i*).

On the Macro-Jê side of the comparison, one finds only PSJ **kê* // **ke* (the latter allomorph appears utterance-finally due to a general process of vowel lowering), reflected, for example, in Kaingang *ke* // *ké* 'to do, to say'. This verb lacks known cognates elsewhere in Macro-Jê. It is semantically close to its Tupian counterparts in that it is used both for actions and speech acts. However, it cannot be a Tupian loan, since the only Tupian languages that have a velar relfex of PT **ɨ* — Tuparian and Kepkiriwat — are spoken 1,500 km northwest from the Southern Jê-speaking zone. If it goes back to Proto-Macro-Jê, the protoform must have been **-ki(C)*.

'white': PT ***-ǩɨT*** : PCerr ***-ka:***

The Proto-Tupian reconstruction is based on reflexes in Rama-Puru (Karo *-kût*), Mondé (Paiter *-kír*), Eastern (Sateré-Mawé *-kyt{'i}*, *-kyt{siɡ̃}*), and apparently Tuparikém (Wayoró *-{y}ir{a}* 'white', though the main root for 'white' in Tuparikém is **-pa(:)K*); see Galucio et al. 2015: 260 for more reflexes.

The Proto-Cerrado form for 'white' is reconstructed as **-ka* by Nikulin (2020: 467), who does not recognize the existence of contrastive vowel length in that protolanguage. However, it is now clear that long vowels in Pykobjê–Krĩkatí (and Canela–Krahô, whose long vowels are however not so thoroughly documented) correspond to long vowels in Xavante, where they are preserved utterance-medially only, as documented by McLeod & Mitchell (1977). Therefore, long vowels must have existed in Proto-Cerrado. Xavante *-'a* /-ʔa:/ 'white' has an underlying long vowel, as seen in the example *tsi'a hã pi'õ* [si:ʔa: hã piʔõ] /ci:-ʔa: hã piʔõj/ 'the chicken (lit. white bird) is female' (McLeod & Mitchell 1977: 107), and so does Pykobjê–Krĩkatí *-jacaa* /-jakʰa:/ 'white'. Other reflexes include Khĩsêtjê *-jakha*, Mẽbêngôkre *-jaka*, and Akwẽ-Xerente *-ka*. The Northern Jê reflexes contain the element **-ja-*, which could have historically been a plural prefix. The updated Proto-Cerrado reconstruction is, therefore, **-ka:*. No cognates elsewhere in Macro-Jê are known, but no stronger candidates for the Proto-Macro-Jê root for 'white' are known either[16]. If **-ka:* is shown to be of Macro-Jê origin, its protoform can be reconstructed as **-kaC°* or as **-kâC°*, with an unidentified coda.

In both language families, the term in question is a class I relational stem.

'husband': PT ***-mẽT*** : PMJ ***-mbi₂n*** (Eastern)

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32. Reflexes are found in the Tuparikém (Wayoró *-mẽn*, Karitiana *-man*), Rama-Puru (Puruborá *-mẽT*), Mondé (Gavião *-met*), and Eastern (Yudja *-menά*, Sateré-Mawé *-men*, Apyãwa *-men-a*) branches. The correspondences are regular.

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 380. Reflexes are found in the Eastern branch only and include Maxakalí *-pit* 'male', Khĩsêtjê *-mdjên* // *-mdjêni* 'husband', Panará *inpin-pjâ* 'husband', and Kaingang *-mén* 'husband'. The correspondences are regular.

In both language families, the noun in question is a class I relational noun.

'I': PT ***o-*** : PCerr ***wa***

In Proto-Tupian, **o-* is reconstructed as an absolutive/genitive first-person prefix, from which the pronoun **õT* is derived, just like the pronoun **ẽT* is derived from the second-person prefix **e-*.

[16] Each Macro-Jê branch employs its own root(s) for this meaning: Krenak *-jirum*; PSJ **kupri*; Proto-Karajá **-kûrã*; Ofayé *-kãĕʔ* and *-gə̆teʔ*; Rikbaktsa *-baraza*; Arikapú *-mäw* ~ *-mäo*; Djeoromitxí *-känõrü*; Proto-Chiquitano **purusuβíi*.

It has reflexes in almost all Tupian languages. Before consonant-initial stems, it is reflected as Kepkiriwat ‹u-›, Makurap *o-*, Wayoró *o-*, Karitiana *y-*, Karo *o-*, Puruborá *ɔ-*, Paiter *o-*, Yudja *u-*, Mundurukú *o-*, Sateré-Mawé *u{i}-*, Apyãwa *w{e}-* ‘1*CRF*’, among many other reflexes. Before vowel-initial stems, it shows asyllabic allomorphs in some languages, such as Wayoró *m(b)-/∅-* (before rounded vowels) or Yudja *w-/∅-*. In Mawé–Guaranian, it is unexpectedly reflected as **uC-* rather than **u-*. The TG reflex is only used anaphorically, particularly when a first-person possessor on a noun or a first-person argument of a gerund of an intransitive verb is coreferential with some other participant. As for noncoreferential uses, it has been ousted by the clitic **ice=* in the TG languages.

The Proto-Cerrado pronoun **wa* ‘I’ is reflected as Xavante *wa hã*, Akwẽ-Xerente *wa (hã)*, Khĩsêtjê *pa* (topical) / *wa* (nominative), Mẽbêngôkre *ba*, Canela–Krahô *pa* (topical) / *wa* (nominative), among other reflexes. It is erroneously reconstructed as **waj*’ in Nikulin 2020: 451, where the palatal coda is claimed to have been present in the reconstructed form based on the Akwẽ-Xerente reflexes *waĩmẽ* ‘with me’, *waĩtê* ‘mine’, mistakenly segmented as *waĩ-mẽ*, *waĩ-tê*. Instead, the correct segmentation must be *wa=ĩ-mẽ*, *wa=ĩ-tê*, where *ĩ-* is a first-person prefix preceded by the cliticized pronoun *wa*. Compare also the second-person forms *kaimẽ* ‘with you’ and *kaitê* ‘yours’, analyzable as *ka=ai-mẽ*, *ka=ai-tê*, where *ka* is a pronoun and *ai-* is a second-person prefix. Some Northern Jê languages show distinct reflexes of **wa* when stressed (topical) and unstressed (nominative); at least in Khĩsêtjê this is the expected consequence of the conditioned split that affected PNJ **b*.

Proto-Cerrado **wa* has no known cognates in other Macro-Jê languages. If it is shown to go back to Macro-Jê, its original form can be stipulated to have been **u(C)*. Nikulin (2020: 187–193) reconstructs a case paradigm consisting of PMJ **iñ* (first-person internal case pronoun) and **a* (first-person agentive case pronoun), but does not discard the possibility that the pronominal case paradigm included even more cases. It is possible that Proto-Cerrado **wa* reflects a PMJ first-person pronoun inflected for some other case, whose original function is yet to be identified.

‘to wake up’: PT ***-paK*** : Proto-Jabutian ***-pa***

Proto-Tupian **-paK* is reconstructed based on its reflexes in the Tuparikém (Wayoró *-{e}pak*), Rama-Puru (Karo *-{pe}pak*), Mondé (Paiter *-páká-tẽ* ‘to wake smb. up’, *-pák{o}* ‘to be awake’), and Eastern (Yudja *-pak-*, Apyãwa *-pãk*) branches. The correspondences are regular. The elements *e-* in Tuparian and *pe-* in Karo are, at least etymologically, intransitivizing and impersonal passive markers, respectively.

The Proto-Jabutian reconstruction is from Nikulin 2020: 542. The root is preserved both in Arikapú and Djeoromitxí as *-pa*. It is hardly borrowed from the neighboring Tuparian languages, since all Tuparian languages show the element *e-* found in Wayoró. If the Jabutian root is shown to be of Macro-Jê origin, the protoform can be reconstructed as **-pa(C)*, **-pə(C)*, or possibly **-pỹ(C)*. No stronger candidates for the Proto-Macro-Jê root for ‘to wake up’ are known[17].

Both in Tupian and Jabutian the verb is a relational class I stem.

‘heavy’: PT ***-pətɨC*** : Maxakalí ***-ptux***

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 34. Reflexes are found in most Tupian languages, including Tuparikém (Makurap *-poti*, Karitiana *-pyti*), Rama-Puru

[17] Each Macro-Jê branch employs its own root for this meaning: Krenak *-mrət ~ -mrək* (intr.), *-mrãŋ* (tr.); Maxakalí *-koxa-k*, irrealis *-koxa* (tr.); PJ **-rĩt°* ‘to wake up, to look’ (intr.), cf. also PNJ **-mbra:* (nonfinite **-mbra:-r*) ‘to wake up’ (tr.); Proto-Karajá **-eθî-θã* (intr., with the reflexive prefix **eθî-*); Ofayé *-xêrê-ge* / *-xêhõê ~ -xêhôj*; Rikbaktsa *-popo* (tr.); Proto-Chiquitano **syto-pyr-* (intr., from **-sýto ~ *-súto* ‘eye’).

(Karo *-pi'ti*), Mondé (Paiter *pati-ga*), and Eastern branches (Yudja *-padít-*, Sateré-Mawé *-potyi*, Apyãwa *-pooj*). The correspondences are mostly regular, except that Wayoró *-pooti ~ -poti* (Nogueira et al. 2021: 105) has an unexpected variant with a long vowel; Mundurukú *-poxí* (Picanço 2005: 264) has an irregular rounded vowel (unlike in the closely related Kuruaya); and Karo *-pi'ti* (Gabas Jr 1999: 15) has an unexpected vowel in the initial syllable followed by a glottal stop. In addition, the vowel of the final syllable has been attested as *e* in the Juruna branch (Yudja *-padét-*, Xipaya *-padet-*; Fargetti & Rodrigues 2008: 562), but at least the Yudja form is mistranscribed in that source. The actual Yudja form is *-padít-*, as attested elsewhere (Lima 2014: 28) and confirmed by native speakers.

On the Macro-Jê side, one finds Maxakalí *-ptux* /-ptɨC/ 'heavy' (Silva 2020: 96), with the allomorph *-putux* ocurring after consonants. It lacks known cognates elsewhere in Macro-Jê, though technically it can be quite old, given that no other Proto-Macro-Jê term for 'heavy' can be reconstructed[18]. The hypothetical PMJ form could then start with **pr*, **m(b)r*, **pVt*, **pVn(d)*, **pVr*, **mbVt*, **mbVn(d)*, **mbVr*, **wVt*, **wVn(d)*, or **wVr*; the nucleus could be either **ɔ̂*, **y*, or **ỹ*; the coda could be any palatal coda, with or without an echo vowel. It is unlikely that the Maxakalí form was borrowed from Tupian. Although Maxakalí has a handful of well-known loanwords from a Tupian language, these come from Old Tupí, or from a closely related variety (Ribeiro 2012b: 91). However, Old Tupí, just like all TG languages, does not preserve Proto-Tupian **t* as a stop, and has the form *-posyî* /-pɔsɨC/ 'heavy' as the reflex of PT **-pətɨC.* Such a form would have been borrowed into Maxakalí as **-poxux* */-pucɨC/, or perhaps as **-pxux* */-pcɨC/ (assuming a diachronic loss of unstressed /u/, as in *-pxet* 'one' and *-ptox* 'head'; see Silva & Nikulin 2021: 36). From a phonological point of view, non-TG Tupian languages would be more suitable candidates, but all of these languages are spoken thousands of kilometers west of the current Maxakalí area.

'to go', 'to come': PT ***-tẽP*** 'to exit', ***-ʔatẽP*** 'to arrive'
: PMJ ***tẽ*** (nonfinite ***-tẽ-m*** or ***-tẽ-n***) 'to go, to come' (Eastern)

The Proto-Tupian reconstruction **-tẽP* 'to exit' is based on reflexes in the Tuparikém (Karitiana *-tam* 'to fly') and Eastern (Mundurukú *-cẽm*, Sateré-Mawé *-tem*, Awetí *-tem*, Kawaiwete *-em*) branches. The Proto-Tupian reconstruction **-ʔatẽP* 'to arrive' is based on reflexes in the Tuparikém (Karitiana *-otam*), Rama-Puru (Puruborá *-anẽm-ã*), and Eastern (Mundurukú *-àjẽm*, Awetí *-{to}tem*, Apyãwa *-{w}aem*) branches. The erstwhile presence of PT **ʔ* is recoverable based on the creaky voice in Mundurukú. All TG languages show a fossilized element **w-*, which is likely to have originated in an active third-person prefix. The correspondences are regular for both verbs. The former appears to have split into two different verbs in the Guaranian branch of TG: **-θẽP* (> Tapiete, Mbyá *-ẽ*) and **-cẽP* (> Paraguayan Guaraní *-sẽ*, Mbyá *-xẽ* 'to leave definitely'), with Mbyá showing reflexes of both with different meanings. This is likely a result of horizontal transmission between (pre-Proto-)Guaranian varieties. PT **t* is otherwise known to have two reflexes in Proto-Guaranian in the default position, which I reconstruct as **θ* and **c*, but the conditioning environments for this purported split have not been established so far[19].

[18] Each Macro-Jê branch employs its own root for this meaning: Krenak *mukran ~ mukraŋ*; Proto-Goyaz **-pytĩ:*; Proto-Akuwẽ **-pirê: // *-pirê*; PSJ **kuθy*; Proto-Karajá **-kutîe*; Ofayé *-wencɔ̃wɔ̃ ~ -encɔ̃wɔ̃*; Rikbaktsa *-tsakyrik*; Proto-Jabutian **-kômỹ ~ *-kumỹ*; Proto-Chiquitano **-ãũmĩ/*-ũũmĩ*. Despite the superficial similarity between the Maxakalí, Proto-Goyaz, and Proto-Akuwẽ forms, as well as between the PSJ and the Proto-Karajá one, none of them are conceivably cognate with each other because of lack of regular correspondences involving vowels.

[19] Some authors have proposed that the distinction between these two consonants is quite old, and project it to the Proto-Tupi–Guaranian (Carvalho 2022) or even Proto-Tupian (Rodrigues 2007) stage. Others assume

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 387. Reflexes are found in the Eastern branch only and include Maxakalí *-nũn* (irrealis *nũ*) 'to come', Krenak *-nĩŋ* (imperative *nĩ*), Khĩsêtjê *thẽ* (nonfinite *-thẽm*) 'to go/come.SG', Xavante *-nem* 'to go/come.DU' (nonfinite only), and Kaingang *tĩ* (nonfinite *-tĩg*) 'to go/come.DU'. The correspondences are mostly regular, except that the nonfinite form in Jê points to PMJ **-tẽ-m*, whereas the realis/indicative forms found in the Trans-São Francisco languages Maxakalí and Krenak point to PMJ **-tẽ-n*. In the Jê languages, the verb in question is used as the generic movement verb (the concepts 'to go' and 'to come' are distinguished by means of centrifugal and centripetal particles), restricted to singular subjects in the Goyaz and Southern branches and to dual subjects in the Akuwẽ branch. For plural subjects, the verb **mũ$_1$* 'to go/come.PL' is used. In the Trans-São Francisco branch, the opposition between the cognates of **tẽ* and **mũ$_1$* is not that of number, but rather of direction: PJ **tẽ* 'to go/come.*SG*' corresponds to **nẽ-n* (irrealis **nẽ*) 'to come', whereas **mũ$_1$* 'to go/come.*PL*' corresponds to **mũ-ŋ* (irrealis **mũ*) 'to go'.

The Tupian verbs are class I verbs. In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class I relational stem. The proposed match is between the nonfinite stem in Macro-Jê and the invariable stem in Tupian; note that Tupian does not have a systematic finiteness distinction in verbal stems except for a handful of verbs in TG, which are usually referred to as irregular verbs (cf. Barbosa 1956: 305–309).

'to arrive': PT ***-wuc*** 'to go out, to arrive' : PCerr ****wôc***, nonfinite ****-wôc***

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém (Karitiana *-hej* 'to go away, to abandon'), Rama-Puru (Karo *-wûy* 'to go out', Puruborá *-wɨ* 'to go out'), and Eastern (Yudja *-wï* 'to arrive') branches. The correspondence between the onset consonants and the vowels is regular. The fact that Karitiana and Karo show a palatal coda, absent in Puruborá and the Juruna languages, remains unexplained. An identical correspondence is observed in the Rama-Puru cognate set for 'to wait' (Karo *-pûy*, Puruborá *-bɨ*), suggesting that at least Puruborá may have regularly lost the palatal coda after an *ɨ*. The polysemy 'to go out' / 'to arrive' is common in the region, and is attested in languages such as Canela–Krahô (*-cato*) or Maxakalí (*-xu-pep* / *-ã-pep*, singular only).

The Proto-Cerrado form is reconstructed as **wôj* (finite), **-wôc* (nonfinite) in Nikulin 2020: 451 based on reflexes such as Khĩsêtjê *pâji* (finite), *-pôt* (nonfinite) and Xavante *wi* (finite, singular only), *-witsi* (nonfinite, singular only). However, the reconstruction can be amended to **wôc* (finite), **-wôc* (nonfinite). The Northern Jê languages show a regular lenition of the stem-final stop in the finite form, yielding the reflex **bôj* as opposed to the nonfinite form **-bôc* (Nikulin & Salanova 2019: 544). In the Akuwẽ languages, the finite form is reconstructed as **wi* 'to arrive.*SG*', but the loss of **-c* is expected in the finite form, since finite forms only occur clause-finally in Akuwẽ, and the utterance-final allophone of Proto-Akuwẽ */c/ is zero.[20] No cognates in other Macro-Jê languages are known, but if **(-)wôc* is shown to be of Macro-Jê origin, the respective PMJ protoform must have been **(-)wy$_1$c°*.

The Tupian verbs are class I verbs. In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class I relational stem.

---

that the distinction is a relatively recent innovation restricted to the Guaranian branch (cf. Schleicher 1998, Meira & Drude 2015, Nikulin & Carvalho 2022). The existence of doublets such as **-θẽP*/**-cẽP* 'to leave' suggests that the distinction between **θ* and **c* cannot continue an ancient Proto-Tupi–Guaranian or Proto-Tupian opposition.

[20] In the utterance-medial position, Proto-Akuwẽ */c/ surfaces as **cV*, as in **-(ⁿ)pêcê* 'good', **-puci* 'to leave.DU.NF', **-ĵaci* 'to enter.DU.NF'. These stems surface as **-(ⁿ)pê*, **-pu*, **-ĵa* in the clause-final position.

4.4. Limited distribution in both families

‘bat’: Proto-Goyaz ****nĵêp*** : PT ****juP*** (Kepkiriwat and Mondé)

Proto-Goyaz **nĵêp* ‘bat’ is reconstructed based on reflexes such as Khĩsêtjê *ntêp-txi* and Panará *{na}nsêpi* (attested as ‹incêp› in the early 20th century). The root lacks known cognates in other Macro-Jê languages. Nikulin’s (2020: 463) comparison of Proto-Goyaz **nĵêp* ‘bat’ with Proto-Akuwẽ **cibi* // **ci:bi* ‘tarantula’ must be rejected not only for semantic, but also phonological reasons: the expected cognate of Proto-Goyaz **nĵêp* in Proto-Akuwẽ should have the form ***cipi* // ***ci:bi* (underlying **/cip/). Since there are no stronger candidates for the Proto-Macro-Jê term for ‘bat’,[21] it is possible that Proto-Goyaz **nĵêp* is a retention from the hypothetical Proto-Macro-Jê form **nĵip°*.

In Tupian, similar terms for ‘bat’ are found in at least two Rondonian branches, Kepkiriwat (‹jêp›) and Mondé (Paiter *líhb*, Zoró *djîp*, among other reflexes; Proto-Mondê **nĵí:P*). In João Barbosa de Faria’s notes, ‹e› or ‹ê› may stand for Kepkiriwat /ɨ/, a sound transcribed by Cândido M. S. Rondon as ‹u› (compare Barbosa de Faria’s ‹queitaá› and Rondon’s ‹cuitá-á› ‘earth’). Therefore, the Kepkiriwat term for ‘bat’ can be restituted as /jɨP/ (/jeP/ is another possibility, but this does not correspond to Proto-Mondé **nĵí:P*). A possible cognate in the Rama-Puru branch is Puruborá *ʃip{ɛ̃}* (Monserrat 2005: 16), though the morphological segmentation is unclear. A much weaker candidate for the Proto-Tupian term for ‘bat’ is seen in the Tuparikém (Makurap *waxariax*, Wayoró *ngwaria*, Tup *wári’a*, Sakurabiat *kwarisa*, Karitiana *asori*, Arikém /ɒjɒɾi/) branch and in Awetí (*tati’a*). However, the correspondences are entirely irregular: Core Tuparian languages point to PT **wariʔa ~ *warɨʔa*, Makurap to **wajariʔaC ~ *wajarɨʔaC*, Karitiana to **wejari ~ *wejarɨ*, Arikém to **aiari ~ *aiarɨ* or the like, and Awetí to **wakeʔa*. This etymology plausibly involves extensive horizontal transmission rather than cognation.

It is unlikely that the similarity between Proto-Goyaz, Kepkiriwat, and Mondé forms is due to contact. Note that the Goyaz languages are mostly spoken more than 1,000 km east from the Kepkiriwat- and Mondé-speaking area. An exception is constituted by the westernmost Goyaz languages, Kajkwakhrattxi (until the 20th century) and Khĩsêtjê (until the 19th century), which used to be spoken in the Tapajós River basin, some 300 km east from the easternmost Mondé territory. However, Kajkwakhrattxi and Khĩsêtjê are known to be newcomers in that region; moreover, these languages characteristically reflect Proto-Goyaz **nĵ* as *nt* /ⁿt̪/.

Both in Goyaz and Tupian, the term for ‘bat’ is an absolute stem.

‘to dig’: PMJ ****-kut*** (Eastern only) : Proto-Mundurukuan ****-ɟe-kot***

The Proto-Macro-Jê reconstruction, taken from Nikulin 2020: 411, is based on reflexes restricted to the Eastern branch, such as Maxakalí *-kot*, Khĩsêtjê *-khwâ* (nonfinite *-khwân*), Laklãnõ *ka* ‘to dig’. The correspondences are regular, except that in the Northern Jê languages the finite form (PNJ **-kwâ*) was analogically remodeled based on the regular nonfinite form **-kwâ-ñ*; the expected reflex of the finite form would have been ***-kwa* (the sound change **wa* > **wâ* normally takes place only in closed syllables).

The Proto-Mundurukuan reconstruction, taken from Picanço (2019: 137), is based on Mundurukú *-je-kot* and Kuruaya *-de-kot* ‘to dig’. This verb includes a middle voice prefix, Mundurukú *je-* / Kuruaya *de-* (Gomes 2007). The root lacks known cognates in other Tupian

[21] Each Macro-Jê branch employs its own root for this meaning: Krenak *kiiñjət ~ hiĩñjət ~ ñ̥əñjət*; Maxakalí *xũnĩm*; PSJ **k(r)yŋθej*; Proto-Akuwẽ **arobo*; Karajá *tyrèhè*; Ofayé *ɸoktaeʔ ~ ɸektajʔ* (underlying /ɸəŋtan°/ or the like); Rikbaktsa *byrizuk*; Arikapú *arokäi*; Djeromitxí *beretxe*; Proto-Chiquitano **šyβijucy- (~ *ši- ~ *šu-)*.

languages, however, there are no stronger candidates for the Proto-Tupian verb for 'to dig'[22]. A semantically close verb *-*kəC* probably rather meant 'to plant', as evidenced by its reflexes in Mondé, Juruna (also 'to bury'), or Sateré-Mawé (Nikulin & Carvalho 2022: 30); only the Awetí–Guaranian languages show the meaning 'to dig', and even then usually in compounds, such as Apyãwa *-'ywy-kaj* (with a historically incorporated root *ywy* 'earth'). Therefore, it is quite possible that the Proto-Mundurukuan root *-*kot* is an archaism. The respective Proto-Tupian form could have been *-*ko(:)T*, *-*ǩo(:)T*, or *-*ŋgo(:)T*. A possible semantically shifted cognate, *kohr{a}* or *kor-kor* 'to paddle', is seen in Paiter. If the Proto-Tupian reconstruction is shown to be *-*ǩoT*, Wayoró *-pi-ot* could be claimed to be a partial cognate (but see fn. 22).

'to enter': PJ ***ŋgê₂*** (plural only) : PT ***-ke ~ *-ǩe*** (Eastern)

The Proto-Jê reconstruction is from Nikulin 2020: 443. It is reflected as Khĩsêtjê *angrê* (nonfinite *-ngrêt*), Xavante *ãdza* (nonfinite *-dzatsi*, dual only), Kaingang *ge*, among other reflexes. In fact, the Cerrado languages show reflexes of three morphologically related verbs: **a:ŋgja* (nonfinite *-*ŋgjac*) 'to enter.PL', *-*ŋgja* (nonfinite *-*ŋgjañ*') 'to insert.PL', and *-*jaŋgja* (nonfinite *-*jaŋgjañ*') 'to wear.PL'. The correspondences are regular, except that the finite forms in Parkatêjê (*akjêj*) and Canela–Krahô (*acjêj*) have been remodeled based on the nonfinite form; the expected finite forms in these languages would be **akjê*/**acjê* or **akia*/**aquij*. Khĩsêtjê *ngr* /ŋɽ/ [ŋgɹ] is possibly the regular reflex of PNJ */ŋj/, though no supporting examples are known. If PJ *$ŋgê_2$ is of Macro-Jê origin, the respective protoform can be reconstructed as *$ŋgi_2$(C)*. No alternative candidate for 'to enter.PL' can be reconstructed. Its singular counterpart is reconstructed as Proto-Macro-Jê **jôp* (Nikulin 2020: 400), based on reflexes in the Cerrado languages (**a:jə*, nonfinite *-*jəpr*) and Ofayé *-xəh*. However, it is equally possible that the Ofayé verb is cognate with Karajá *-lò* 'to enter', Djeoromitxí *hu*/*-ru* 'to enter', and possibly Arikapú *-txu{rü}* 'to enter' (with the unexplained element *-rü*). In this case, one should reconstruct PMJ **jy(C)* 'to enter.SG'. Rikbaktsa *-tsuk* 'to enter.SG' shows no regular correspondence to the aforementioned forms, despite being superficially similar.

In Tupian, the reflexes of *-*ke* ~ *-*ǩe* 'to enter' are only seen in the Eastern branch and include Mundurukú *-je-xé* 'to come home' (with a middle voice prefix), Sateré-Mawé *-(w)e-ke* 'to enter' (with a reflexive prefix), Apyãwa *-ke* 'to enter', among other reflexes. In some TG languages, the verb shows a prefixal alternation between the finite and nonfinite stems. For example, Old Tupí has the finite stem *-ike*, whereas in the nonfinite paradigm the class II stem *teîke* (*-reîke*, 3 *s-eîke*) is found (Barbosa 1956: 307). Similar alternations affect several other *i-/*e-initial verbs; I assume that the alternation in question originated as an absolute/relational alternation, also found in pairs such as PT *ĩrĩ 'hammock (absolute)' and *-j-ẽrĩ 'hammock (relational)'. If this turns out to be an archaism, one can reconstruct PT *-ike ~ *-iǩe (finite, absolute) and *-j-eke ~ *-j-eǩe (non-finite, relational) 'to enter, to come home', with the loss of the initial vowel in languages such as Apyãwa. This verb was in any case distinct from PT *-*wuãP* 'to enter', with reflexes in the Tuparikém (Makurap *-mum*/*-mu-ã*, Wayoró *-ngũ-ã*, Karitiana *-mem*) and Eastern (Mundurukú *-ə̃m* 'to enter') branches[23].

[22] Each Tupian branch employs its own root for this meaning: Tuparí *-ay*, Wayoró *-pi-ot*, Akuntsu *-poro-ka*, Makurap *-kix*, Karitiana *-yt*, Sateré-Mawé *-pan*, Awetí *-koy* (from Proto-Tupian *-*kəC* 'to plant'), PTG *-*ʔɨβɨ-koC* (*ɨβɨ 'earth' is historically an incorporated object, and *-*koC* goes back to *-*kəC* 'to plant'). The element *-ot* in Wayoró *-pi-ot* could be cognate with Karitiana *-yt*, pointing to Proto-Tuparikém *-*oT*, but Nogueira (2019: 175) analyzes the Wayoró verb as 'to go inside', where *-ot* stands for 'to go'.

[23] The Mundurukú reflex *õ* of PT **wuã* is not known to be regular, but a similar sound correspondence is seen in Mundurukú *ó-'a* (< PT **wɨ* 'ax').

In both language families, the finite verb appears to have been originally absolute (and fossilized voice prefixes are seen in the Cerrado languages and in Tupian), and its nonfinite counterpart is reconstructed as a relational stem (class I in Jê, class II in Tupian).

‘to pierce’: PCerr ***-pôk*** (SG), ***-japôk*** (PL) : PTG ***-puK***

The Proto-Cerrado reconstruction is from Nikulin (2020: 446), who also considers the possibility that the singular stem had a distinct finite form *-*pôr* (preserved in Akuwẽ only), but note that the alternation *-*r* (finite) / *-*k* (nonfinite) is otherwise normally found in intransitive verbs only. Reflexes include Canela–Krahô *-{jõ}pôc* ‘to gut’, *-japôc* ‘to pierce.PL’ and Xavante *puru* (finite), *-pu’u* // *-pu* (nonfinite) ‘to pierce.SG, to spill’, *-dzapu’u* // *-dzapu* ‘to pierce.PL’. No cognates in other Macro-Jê languages are known, but no stronger candidates for the Proto-Macro-Jê term for ‘to pierce’ are known either. If this root does go back to Proto-Macro-Jê, its original PMJ form can be reconstructed as *-$py_1k°$ ~ *(-)$py_1ŋ°$.

On the Tupian side of the comparison, one finds Kawaiwete *-fuk* ‘to be pierced’, Apyãwa *-pok* ‘to bleed’, Old Tupí *-puk* ‘to have a hole, to break (intr.)’, pointing to PTG *-puK ‘to be pierced’. If this root does go back to Proto-Tupian, its original PMJ form can be reconstructed as *-*po(:)K* or *-*mbo(:)K*. Phonetically similar verbs in other Tupian languages, such as Sateré-Mawé *-puk* ‘to swell’ or Makurap *-pok* ‘to beat, to kill’, are too semantically distant from the TG verb, and are not considered to be cognate.

The Proto-Cerrado verb is reconstructed as transitive (class I), and the Proto-Tupian as intransitive (class I).

‘son’: Proto-Chiquitano ***´-tsay***
: Proto-Tuparian ***-jaʔɨP*** or Proto-Mawé–Guaranian ***-caʔɨT***

On the Macro-Jê side, one finds Proto-Chiquitano *´-*tsay* ‘son’, where *t*s* is a thematic consonant: compare 1SG *í-*tsay*, 1+2 *ú-*tsay*, with the thematic consonant, and 2SG *∅-áy, 3SG *aý-šy* without it. Reflexes are seen in all Chiquitano varieties, such as Bésiro ´-*saɨ*. No cognates in other Macro-Jê languages have been found, but it could in principle go back to PMJ *-*jay(C)* or a similar protoform.

There are two similar forms in the Tupian languages. Proto-Tuparian *-*jaʔɨP* ‘son, fraternal nephew (male ego)’ has reflexes in all Tuparian languages: Makurap *-xaup* (also ‘sperm’), Wayoró *-ndaup*, Tuparí *-ha’úp*, Sakurabiat *-taup*, Akuntsú *-taɨP* (Nogueira et al. 2019: 43; Nikulin & Andrade 2020: 295). It lacks known cognates in other Tupian languages, but could in principle go back to PT *-*jaʔɯP* or *-*jaǩɯP*; the former could be related to Proto-Chiquitano *´-*tsay*. In the Mawé–Guaranian languages, one finds reflexes of Proto-Mawé–Guaranian *-*caʔɨT* ‘son / fraternal nephew (male ego)’, reflected as Sateré-Mawé *-sa’yr{u}*, Awetí *ta’yt*/*-a’yt*, PTG *ta*ʔɨT* / *-ra*ʔɨT* (Carvalho & Birchall 2022: 27). Unless it is related to Proto-Tupian *-*ƙɯT* ‘child’, with reflexes in Tuparikém and Mundurukuran, it has no known cognates elsewhere in Tupian. Its possible original PT form could be *-*jaʔɨT*, *-*jaʔɯT*, *-*jaƙɨT*, or *-*jaƙɯT*; the former two could be technically related to Proto-Chiquitano *´-*tsay*.

All the aforementioned forms are class II relational stems. In Tupi–Guaranian, *ta*ʔɨP*/*-ra*ʔɨP* belongs to the so-called subclass IIb, which includes a handful of kinship terms; it differs from other class II subtypes in having a third-person form identical to the absolute one (**ta*ʔɨP* ‘his son/fraternal nephew’).

‘sour’: PJ ***-jôK*** ‘sour, salty’ : Karitiana ***-syk***

The Proto-Jê reconstruction is from Nikulin 2020: 437. Reflexes include Khĩsêtjê *-*tɯa* ‘sour’, Canela–Krahô *-xwa* ‘sour, salty’, Pykobjê–Krĩkatí *-xwa* ‘salty’, Kaingang *-{ka}jã* ‘salty, sour’. PNJ **ka:ĵwa* ‘salt’ is likely related. No cognates in other Macro-Jê languages are known, but no

stronger candidates for the Proto-Macro-Jê term for 'sour' are known either (see the discussion under 'bitter' in **4.3**). If this root does go back to Proto-Macro-Jê, its original PMJ form can be reconstructed as *-*juk* (the stem-final velar stop can be recovered based on the Kaingang verbal derivative *-kajã-g* 'to become sour').

On the Tupian side of the comparison, one finds Karitiana *-syk* 'sour, to become sour', attested in Rocha 2011: 218. A possible cognate is Karo *-{xa'}yõk* 'sour; to be drunk', prompting the reconstruction PT *-*joK*. The element *xa'-* in *-xa'yõk* is plausibly a fossilized prefix with an unclear meaning, also found in *xa'kĩn* 'monkey (sp.)', *xa'wût* 'thorn', *xa'wap* 'sun', *a'-xa'pe* 'bark' (compare PT *-*pe* 'bark, skin'); the nasal vowel *õ* is unexpected, but parallels do exist (Karo *-yakõp* 'hot' < PT *-*jakoP*). Alternatively, the Karo form can also be compared to the final syllable of Sateré-Mawé *-j̃ej̃ug̃* 'sour' (only the third-person form *h-ej̃ug̃* is in fact attested in Ribeiro 2010: 58); this would account for the nasal vowel in Karo, but not for the stem-final *k* (Karo **…yõg* would be expected). Karitiana *-syk* is also similar to Proto-Mundurukuan *-*sak* 'to be sour' (Picanço 2019: 138), but there are no regular correspondences between these forms, and the similarity must be accidental. The hypothetical Mundurukuan cognate of Karitiana *-syk* would be **-*ðək*; note that Proto-Mundurukuan **s* normally results from contraction, as in **másɨk* < PT **mãnĩ-ʔəK* (possibly through the stages **mãndjɨk* < **mãndiʔɨk*).

Be it as it may, Karitiana *-syk* and possibly Karo *-xa'yõk* can technically go back to PT *-*joK*. The meaning 'sour' could probably be alternatively expressed by PT *-*atɨ* or *-*jatɨ* 'pain, to hurt' (Nikulin & Carvalho 2022: 29), as is still the case in Tuparian (Makurap *-xati*, Wayoró *-ati*, Tuparí *-así*); compare Yudja *-xadɨ́* and Xipaya *-xadɨ* 'to become sour', and possibly Aruá ‹tatíit›. If the root *-*joK* coexisted with it, its semantics must have been more restricted.

'sweet': PMJ ***-jə̃ñ*** (Eastern) : Tuparí ***-hoy***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. Reflexes are found in the Eastern branch only and include Maxakalí *-xux-pex*, Khĩsêtjê *-tán* // *-táni*, Xavante *-dzei* // *-dze*. The correspondences are regular, except that the Canela–Krahô and Pykobjê–Krĩkatí reflex *-xen* shows an irregular fronted reflex of the vowel **ə* (the third-person forms Canela–Krahô *h-àn* and Pykobjê–Krĩkatí *h-ỳn* are, however, regular).

Tuparí *-hoy* 'sweet' is attested in Alves 2004: 179. No cognates are known elsewhere in Tupian, but if this form is shown to go back all the way to Proto-Tupian, the respective protoform should be reconstructed as *-*joC*. Each Tuparikém language shows a different root for 'sweet' (Wayoró *-tui*, Akuntsú *-kon*, Makurap ‹čan›, Karitiana *-kowot* /-kowt/), and other branches of Tupian also show noncognate forms: Karo *-pewit*, Aruá ‹čiim›, Proto-Juruna *-*etʃãkũ* (> Yudja *-etxãkũ*, Xipaya *-etákũ*), Mundurukú *-kurúkurú*, Proto-Mawé–Guaranian *-*cẽʔẽ* (> Sateré-Mawé *-j̃e'ẽ*, Awetí *-e'ẽ*, Apyãwa *'e'ẽ*, Old Tupí 3 *s-e'ẽ* 'sweet, salty'). Therefore, even though the chances of Tuparí *-hoy* to be an archaism are rather slim, there are no stronger candidates for the Proto-Tupian term for 'sweet' anyway.

The Proto-Macro-Jê stem is a class II relational stem, as seen in Khĩsêtjê 3 *s-án* // *s-áni*, Canela–Krahô *h-àn*, Pykobjê–Krĩkatí *h-ỳn*. Tuparí has lost the class I/class II distinction, but *h*-initial relational stems in that language typically go back to Proto-Tuparian (and Proto-Tupian) class II stems.

4.5. Noncognate lookalikes or loans

'flat': Proto-Mawé–Guaranian ***-pe:P*** and Ofayé ***-ɸiʔ***

The Proto-Mawé–Guaranian reconstruction is from Meira & Drude 2015: 293. Reflexes include Sateré-Mawé *-pẽp*, Awetí *-pep* (in *mõj-pep* 'flat snake', *tatu-pep* 'armadillo', *ywy-pep*

‘ground’), Apyãwa *-pew-a*, all meaning ‘flat’. No cognates elsewhere in Tupian are known.[24]

The Ofayé term for ‘flat’ is very scarcely attested. Gudschinsky documents it only in compounds (*krej-ɸiʔ* ‘blind = eye-flat’; *kə̆nõ:r-ɸiʔ* ‘cockroach = ?-flat’) and states that the bare root did not occur in her corpus, though her consultant did recognize the element *-ɸiʔ* as a term for ‘flat’. Although Ofayé *-ɸiʔ* is technically comparable to Proto-Mawé–Guaranian *-*pe:P*, I propose that it is more likely to be related to Maxakalí *-pex* /-pɛk/ ‘flat’, attested in the compound *kot-pex* ‘beiju = manioc-flat’ (Silva 2020: 260). The Proto-Macro-Jê etymon can then be reconstructed as *-*pek(°)*, *-*pêk(°)*, *-*mbek(°)*, or *-*mbêk(°)*, with a velar coda that is not compatible with the labial coda seen in Tupian.

‘to kill’: Ofayé ***-kãjʔ***, Proto-Chiquitano ****kõ ̃ɔ̃j-*** ‘to kill, to die’, and Awetí ***-kỹj***

If the Ofayé and Chiquitano forms are indeed cognate, the Proto-Macro-Jê form can be tentatively reconstructed as *-*ŋãJ*. Since PMJ **k* yields Ofayé *h* (Nikulin 2020: 108), one is forced to reconstruct the initial consonant as **ŋ*, whose development in Ofayé has hitherto remained unknown. However, this proposal is not compatible with the idea that Ofayé *həuɻĕʔ* ‘earth’ is cognate with Kaingang *ga* ‘earth’, as hypothesized in section **4.1**. The vowel can be reconstructed as **ã* (the only PMJ nasal vowel whose Ofayé reflex is unknown), and the coda can be reconstructed as palatal based on Ofayé *j* and the Chiquitano third-person finite forms such as Bésɨro *kóiñ-o* or Migueleño *kóoñ-o*. Note that Chiquitano shows pervasive patientive lability, and the verb is used both for describing spontaneous death (with an absolutive subject) and unnatural death (with an absolutive patient and ergative agent). A possible cognate in the Eastern branch Maxakalí is *-kux* ‘to finish’ (Silva 2020: 275).

Awetí *-kỹj* ‘to kill’ (Sabino 2016: 56) is superficially similar to the aforementioned data. This root is isolated within Tupian: the meaning ‘to kill’ is rather expressed by reflexes of PT *-*ʔaoka ~ *-ʔaoka̋* ‘to kill, to beat’ in other languages of the Eastern branch. It cannot be a loan from Ofayé or Chiquitano, because the Awetí live far away from the Ofayé (1,000 km to the north) and the Chiquitano (800 km to the northeast), and there is no reason to suspect these peoples have ever been in contact.

‘liquid’: PT ****ʔu* / *-j-u*** and Proto-Jabutian ****-y***

The Proto-Tupian reconstructions are from Nikulin & Carvalho (2022: 30, 37). PT **ʔu* was an absolute noun (the basic term for ‘water’), whereas *-*j-u* was a relational class II noun, used in compounds that denoted liquids. Reflexes are found in all branches, including Kepkiriwat

[24] Gerardi et al. (2022, concept FLAT) propose a number of competing cognate sets involving terms for ‘flat’, but none withstands scrutiny. Their cognate set 6281 includes Mundurukú *-sẽm* ‘smooth’ (Crofts 1973) and reflexes of an unrelated Proto-Mawé–Guaranian etymon *-*tɨ̃P* ‘smooth’ (whence Sateré-Mawé *-tym* ‘smooth’; Ribeiro 2010: 63). Gerardi et al.’s (2022) cognate set 6282 includes Karo *xẽrat* ‘smooth’ (Gabas Jr 1999: 22) and Kuruaya *-korop*, a term obviously noncognate with Karo *xẽrat* and cognate with Mundurukú *-kóróp* ‘smooth’ (Crofts 1973). Their cognate set 6283 lists three cognate terms restricted to the Mondé branch (Gavião *ĩníníĩp*, Mondé *sinĩp*, Suruí-Paiter *firip*) as well as an obviously noncognate Karitiana term *-kỹkyn* ‘smooth’ (Landin 2005: 15). Finally, Gerardi et al.’s (2022) cognate set 6284 lists a form cited as Kamayurá *ojim* — which is obviously a mistranscription of a root whose third-person form is attested in Seki (2000: 413) as *i-jym* ‘it is smooth’ (ultimately a reflex of Proto-Mawé–Guaranian *-*tɨ̃P* ‘smooth’) — and Paraguayan Guaraní *-joja*, which is ultimately derived from the verb *-ja* ‘to stick’ (< Proto-Tupí–Guaranian *-*jaT*) by means of a reciprocal prefix. Needless to say, the Kamayurá and Paraguayan Guaraní forms given in Gerardi et al. (2022) cannot be cognate. This example is illustrative of the general careless approach to etymology, morphological segmentation, and semantics in Gerardi et al.’s (2022) database, which unfortunately cannot be used as a reference source for comparative Tupian studies.

(‹i-ü› ‘water’), Tuparikém (Makurap *u* ‘water’, Wayoró *u-gu* ‘water’, Karitiana *e* ‘rain’, *e-se* ‘water’, *-se* ‘liquid’), Rama-Puru (Karo *i-xû* ‘water’, Puruborá *ĩ* ‘liquid, chicha’), Mondé (Paiter *ih* ‘water’, Aruá ‹endjatji› ‘tear’, ‹namdji› ‘milk’), and Eastern (Mundurukú *i-dì-bí* ‘water’, *-di* ‘liquid’, Apyãwa *’y-∅* ‘water’); see Galucio et al. 2015: 258 for a selection of reflexes.

Proto-Jabutian **-y* ‘liquid’ (Voort 2007: 159) is reflected as Arikapú *-ü*, Djeoromitxí *-i*. Chiquitano, Rikbaktsa, and Maxakalí have remotely similar forms: Bésiro *n-∅-iyɨ-xɨ* ‘its juice, broth’, Rikbaktsa *tsik* ‘chicha’, *tsik-*/*-tsik* ‘liquid’ (cf. also *-hik* in *pi-hik* ‘water’), Maxakalí *-hep* ‘liquid, blood’, but these are hardly cognate with Proto-Jabutian **-y*. The Proto-Jabutian vowel **y* has no known Macro-Jê origin, suggesting that **-y* is a likely loan from Tupian. It is however unclear why and how the absolute stem **ʔu* (or its reflex in some specific branch of Tupian) could have been borrowed as a relational stem in Jabutian.

‘louse’: PMJ ****-ŋgy₁n°*** (Eastern only) and Proto-Core Mondé ****giT***

The Proto-Macro-Jê reconstruction is based on reflexes found in the Eastern branch only: Maxakalí *-kut*, Khĩsêtjê *-ngô*, Xavante *-’u*, Kaingang *-ga* ‘louse, maggot’. It is erroneously given as **ŋgy₁t* in Nikulin 2020: 419, but the Pykobjê–Krĩkatí cognate *-ncuu* with its long vowel shows that the protoform must be reconstructed with a nasal coda followed by an echo vowel. The expected reflex of **-ŋgyn°* in Xavante would be **-’õno* in the utterance-medial position and *-’u* in the utterance-final position, but the former is not attested in my sources on Xavante; I assume it was ousted by the utterance-final allomorph.

On the Tupian side of the comparison, one finds Gavião, Aruá, and Zoró *git* (the Cinta Larga term for ‘louse’ is not attested in the sources I am aware of). The stem-final consonant matches the Macro-Jê forms, in stark contrast with all other Mondé and, more broadly, Tupian languages, which uniformly show reflexes of Proto-Tupian **(-)ŋgɯP* (Nikulin & Carvalho 2022: 33): compare Paiter *g̃ib*, Salamãy *gip*, Makurap *gup*, Wayoró *-a-ngup*, Karitiana *ngep*, Puruborá *tɨP*, Yudja *kïpá*, Mundurukú *kíp*, Sateré-Mawé *g̃yp*, Awetí *-’a-kyp*, Apyãwa *-kyw-a*; see Galucio et al. 2015: 252 for a selection of reflexes. I have no explanation regarding the outstanding similarity between Proto-Core Mondé **giT* and the Macro-Jê forms. It is of course possible, but also undemonstrable, that an extinct branch of Macro-Jê that preserved the place of articulation of the PMJ codas was present in the Mondé-speaking area during the period of the independent evolution of the Core Mondé languages (i.e., after the split-off of Salamãy, but before the differentiation of Proto-Core Mondé into dialects), and Proto-Core Mondé could have borrowed the noun **giT* from the speculative Macro-Jê language. In any case, it cannot be cognate with PMJ **-ŋgy₁n°*, since the basic term for ‘louse’ both in Proto-Tupian and Proto-Mondé clearly had a labial coda.

The Macro-Jê noun is a relational class I stem. In Tupian, both relational (Wayoró, Awetí, Apyãwa) and absolute (Makurap, Karitiana, Yudja, Mundurukú, Sateré-Mawé) reflexes are attested, suggesting that the Proto-Tupian root was relationally labile.

‘neck’: PT ****-woT*** and Proto-Cerrado ****-mbut***

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém (Makurap *-wot-kup*, Wayoró *-ngot-kup*, Karitiana *-hyt*), Rama-Puru (Karo *-ot ká’*), and Eastern (Sateré-Mawé *-hut-’yp*, Awetí *-tur-’yp*, Apyãwa *-xor-a*) branches; see Galucio et al. 2015: 255 for a selection of reflexes. The correspondences are regular, except for the reflexes in the Siokweriat dialect of Sakurabiat (*-kut-kup* instead of the expected **-kot-kup*) and Akuntsu (*-pɨT-kɨP* instead of the expected **-koT-kɨP*). The reflexes in the Tuparian languages, in Arikém (but not in Karitiana), Sateré-Mawé, and Awetí point to the compound **-woT-ƙuP*.

The Proto-Cerrado reconstruction is from Nikulin 2020: 449. Reflexes include Khĩsêtjê *-mbut* // *-mburu*, Panará *imputi* ‘nape’, and Xavante *-butu* // *-budu*, and the correspondences are

fully regular. It could technically go back to PJ *-*mbut*° < PMJ *-*mbot*°, but it is unlikely that the root in question is old, given that there are two stronger candidates for the PMJ terms for 'neck'. PMJ *-*ndo*$_1$*ñ* 'neck' (Nikulin 2020: 388) is preserved in Chiquitano (Bésiro -*tɨ*, Migueleño -*tɨɨ*), Ofayé (-*tôə̃ʔ*, underlying /-tôn°/ 'nape'), and Eastern (Kaingang -*nunh*). PMJ *-*jô(C)-cet* ~ *-*jô(C)-cek* ~ *-*jy(C)-cet* ~ *-*jy(C)-cet* (Nikulin 2020: 401) is reflected as Karajá -*lòti* and Rikbaktsa -*soik*. Therefore, it is quite improbable that *-*mbot*° was the basic term for 'neck' in Proto-Macro-Jê.

All the aforementioned forms (except Karajá -*lòti*) are class II relational stems.

'powder, paste': PT ***-jõʔõ**P* and Proto-Jabutian ***-nũ***

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém and Eastern branches, including Wayoró -*yõom* 'powder' < Proto-Tuparian *-*ɲõʔõp* (Nikulin & Andrade 2020: 296), Yudja -*umá* < Proto-Juruna *-*um-á*, Mundurukú -*nõm* < Proto-Mundurukuan *-*ðn̪m* (Picanço 2019: 140), Awetí *ywy-lu'um* 'dirt', -*enta-lu'um* 'rheum', and Apyãwa *to'om-a*/-*ro'om-a* 'paste'. The correspondences are regular, except that Awetí *l* is not a regular reflex of PT *j* > Proto-Mawé–Guaranian *c*.

Proto-Jabutian *-*nũ* 'pamonha, porridge, food' is reconstructed in Voort 2007: 156, who notes the similarity of this term with classifiers for 'pamonha, flour' in different unrelated languages of the Guaporé area. However, this noun is a reflex of PMJ *-*ñũ*$_2$*(C)* 'food' (Nikulin 2020: 403), whence Eastern Chiquitano -*õ'õ*, Karajá *dòò* 'solid food, such as fish, turtle or meat', Khĩsêtjê -*nho* 'food', Xavante -*nho* 'food'.

Despite the similarity in form and the fact that both Tupian and Macro-Jê comparanda are relational class II stems, the semantic difference between the Proto-Tupian and Proto-Macro-Jê forms renders the comparison unattractive.

'thorn': PMJ ***-ñĩn°** ~ ***-ñĩñ°*** and Tuparí ***-ĩ***

The Proto-Macro-Jê reconstruction is given as *-*ñĩ(C)* in Nikulin 2020: 406. Reflexes are found in the Western (Djeoromitxí -*nĩ* 'leaf', Rikbaktsa -*ni*), Karajá (*dè~dè*), and Eastern (Khĩsêtjê -*khrã-nhi*, Mẽbêngôkre *mrỳ-nhĩ*, Apinajé -*nhĩ*, Pykobjê–Krĩkatí *hum-jẽeh*, Akwẽ-Xerente -*krã-nĩ*) branches. The Pykobjê–Krĩkatí reflex with a long vowel suggests that the PMJ form ended in a nasal consonant followed by an echo vowel.

Tuparí -*ĩ* 'thorn, grain' (Alves 2004: 185) lacks known cognates in other Tupian languages. It could technically go back to Proto-Tuparian *-*ɲĩ* and PT *-*jĩ(C)* ~ *-*jɨ̃(C)* (the loss of *ɲ before *ĩ* is regular in that language; see Nikulin & Andrade 2020: 296), but it is unlikely that the root in question is old, given that a different root **wo:* 'thorn' (whence Wayoró *ngoo*, Karitiana *hy*, Sateré-Mawé *hu*, Apyãwa *xo-*∅, etc.) can be reconstructed. Instead, Tuparí -*ĩ* could be an Arikapú borrowing. Mundurukú -*ĩ* 'CL:nuts' (Crofts 1985: 313) is probably unrelated.

5. Regular sound correspondences

Now that 38 Macro-Jê–Tupian possible cognate sets have been identified (**4.1–4.4**; the data from **4.5** are discarded), I proceed to examine the sound correspondences that recur in my comparative corpus (**5.1**). Non-recurrent correspondences may signal that a given comparison is spurious, and should be discarded over the next iteration. In section **5.2**, I address the possibility of identifying additional sound correspondences, which violate the constraints set out in the preamble of **4** — notably the full match between the places of articulation of the onsets and codas — but could nevertheless be regular.

5.1. Main sound correspondences

In this section, I make an attempt at determining the sound correspondences between PMJ and PT. In reproducing the data from the preceding section, I adhere to the following principles.

Whenever the data allow for multiple diachronic interpretations, I choose the option that best matches the correspondence sets whose existence is independently established. For example, the data of the Macro-Jê languages are insufficient to determine whether PMJ **-we(C)* 'to go up' contained a coda or not. In this section, this form is rewritten as **-wep ~ *-wem°*, since these are the only possibilities that can match PT **-we(:)P*.

For PMJ and PT hypothetical reconstructions based on evidence from a single branch (**4.2–4.4**), I employ the symbol # instead of the asterisk. For example, the hypothetical Proto-Tupian ancestor of Proto-Mundurukuan **-kə̃j* 'hole' can, in theory, be reconstructed as **-kãʔãC*, **-kə̃ʔə̃C*, **-ǩãʔãC*, **-ǩə̃ʔə̃C*, **-ŋãʔãC*, or **-ŋə̃ʔə̃C*. Of these, **-ǩãʔãC* is the option that best matches PMJ **-kuñ°*, and it is reproduced in this section as PT *#-ǩãʔãC*.

Table 4 shows the sound correspondences between PMJ and PT onsets. PMJ */c/ and */ñ/ each occur only once in the corpus, hence it is unsurprising that the respective correspondences are not recurrent. In the cognate sets for 'arm', 'foot', and 'liver', Tupian shows an alternation between */p/ in relational stems and */m/ in absolute ones. Macro-Jê would appear to have generalized the relational stems for 'arm' and 'foot', and the absolute one for 'liver'. As for the cognate set PMJ **-ja-m* 'to stand (nonfinite)': PT **-ʔãP* 'to stand', it may be significant that PMJ lacks relational vowel-initial stems, and makes use of the relationalizing prefix **j-* when a vowel-initial root enters a relational stem (see section **3**). See **4.3** for a discussion on the root-medial correspondence in the cognate set for 'smoke'.

Two non-recurrent correspondences are PMJ */ŋ/ : PT */ƙ/ ('earth') and PMJ */ŋ/ : PT */k/ or */ǩ/ ('to enter'). Of these, the former could be due to an erroneous inclusion of PSJ **ŋgə* 'earth' into the comparison; if it turns out to be noncognate, the PMJ term for 'earth' can be reconstructed as **kyñ°* instead (with reflexes in Chiquitano and Ofayé), thus instantiating the recurrent correspondence PMJ **k* : PT **ƙ*. Alternatively, one could surmise that historically PMJ had an alternation between relational */k/-initial stems and absolute */ŋ/-initial stems (a similar alternation is reconstructible for Proto-Tupian based on evidence from Sateré-Mawé and Mondé). Note that PMJ **ŋgyñ°* 'earth' and *#ŋgi*$_2$ 'to enter.PL' (finite) are absolute and do not take prefixes, whereas PMJ **-ko*$_2$ 'to ingest', **(-)ky*$_1$*m°* 'tree(-like)', *#-ki* 'to do, to say', **-kut* 'to dig', **-kuñ°* 'hole', *#-kân°* 'white' all take absolutive or accusative indices (**(-)ky*$_1$*m°* is relationally labile).

Table 5 shows the sound correspondences between PMJ and PT vowels (vowel nasality is ignored at this stage). PMJ */ə̂/ and */ê/ occurred each only once in the corpus, hence it is unsurprising that the respective correspondences are not recurrent. The cognate set for 'feces' presents insurmountable difficulties regarding the reconstruction of its nasal vowel in both protolanguages: the correspondences are unique in both Macro-Jê and Tupian. In the cognate set for 'smoke', Tupian could have contracted a disyllabic sequence into a long vowel, as suggested in **4.3**. The vowel correspondence in the term for 'bat' appears to be truly irregular; recall, however, that the reconstruction of PT **ɯ* in **jɯP* hinges on one's interpretation of Barbosa de Faria's attestation of ‹ê› in Kepkiriwat as an instance of /ɨ/. If ‹jêp› is a representation of /jiP/ rather than /jɨP/, the cognate set for 'bat' instantiates the recurrent correspondence PMJ **i* : PT **i*. The vowel correspondence in the term for 'to pierce' is unique; combined with the discrepancy in the transitivity of the PMJ and PT verbs (transitive and intransitive, respectively), this is a sufficient reason to discard the etymology.

Although oral and nasal vowels are not distinguished in the correspondences in Table 5, there is a systematic tendency for PMJ oral vowels to correspond to PT oral vowels (28 examples), whereas PMJ nasal vowels correspond to PT nasal vowels ('to go/come', 'to kill', 'smoke',

PMJ	PT	examples
*/p/	*/p/	**‘arm’**: PMJ ***-pa*** ‘arm’ : PT ***-pə / *mbə*** ‘hand’ **‘to burn’**: PMJ ***(-)py₁k°*** ~ ***(-)py₁ŋ°*** : PT ***-puK*** **‘foot’**: PMJ ***-pâr°*** : PT ***-pi / *mbi*** **‘heavy’**: PMJ ***#-pVtVJ(°)*** : PT ***-pətiC*** **‘to pierce’**: PMJ ***#-py₁k°*** ~ ***#-py₁ŋ°*** : PT ***#-poK*** **‘to wake up’**: PMJ ***#-paK(°)*** : PT ***-paK***
/m/ ([mb], *[m])	*/m/	**‘husband’**: PMJ ***-mbi₂n*** : PT ***-mẽT*** **‘liver’**: PMJ ***-mbâ*** : PT ***-pi(-)ʔa / *mbi(-)ʔa***
*/w/	*/w/	**‘to arrive’**: PMJ ***#(-)wy₁c°*** : PT ***-wuC*** **‘to go up’**: PMJ ***-wep*** ~ ***-wem°*** : PT ***-we(:)P*** **‘to kill’**: PMJ ***-wĩ*** : PT ***#-wĩ***
*/t/	*/t/	**‘heavy’**: PMJ ***#-pVtVJ(°)*** : PT ***-pətiC*** **3CRF prefix**: PMJ ***-ta-*** : PT ***-tə-*** **‘to go/come’**: PMJ ***-tẽ / *-tẽ-m*** : PT ***-tẽP***
/n/ ([nd], no examples for *[n])	*/ð/	**‘bitter’**: PMJ ***#-ndap°*** : PT ***-ðəP*** **‘ripe’**: PMJ ***-ndêp°*** : PT ***#-ðeP***
*/c/	*/c/	**3NCRF**: PMJ ***-c-*** : PT ***-c-***
*/ñ/	*/j/	**‘bat’**: PMJ ***#-nĵip°*** : PT ***-juP***
/j/ ([j], *[ɲ])		**‘father’**: PMJ ***-jo₂m*** : PT ***-joP*** **‘feces’**: PMJ ***-ñVt°*** : PT ***-jVT*** **‘meat (rel.)’**: PMJ ***-ñĩt*** : PT ***-jẽT*** **‘name’**: PMJ ***-jet*** : PT ***-jeT*** **‘pus’**: PMJ ***-jo₂w°*** : PT ***-joP*** **‘sweet’**: PMJ ***-jə̃ñ*** : PT ***#-joC*** **‘smoke’**: PMJ ***-ñĩjəƙ*** : PT ***-jĩːK*** **‘son’**: PMJ ***#-jayC*** : PT ***#-jaʔuP*** or ***#-jaʔuT*** **‘sour’**: PMJ ***#-juk*** : PT ***#-joK*** **‘to stand’**: PMJ ***-ja*** : PT ***-ja*** **‘tooth’**: PMJ ***-juñ°*** : PT ***-jãC***
*/k/	*/ƙ/	**‘to ingest’**: PMJ ***-ko₂*** : PT ***-ƙo*** **‘tree(-like)’**: PMJ ***(-)ky₁m°*** : PT ***(-)kuP*** **‘to do, to say’**: PMJ ***#-ki*** : PT ***-ke***
	*/ǩ/	**‘to dig’**: PMJ ***-kut*** : PT ***#-ǩoT*** **‘hole’**: PMJ ***-kuñ°*** : PT ***#-ǩãʔãC*** **‘white’**: PMJ ***#-kân°*** : PT ***-ǩiT***
/ŋ/ ([ŋg], no examples for *[ŋ])	*/ƙ/	**‘earth’**: PMJ ***-ŋgyñ°*** : PT ***-kuC***
	*/k/ or */ǩ/	**‘to enter’**: PMJ ***#ŋgi₂*** : PT ***-ke ~ *-ǩe***
*∅	*∅	**‘to give’**: PMJ ***-ũp*** : PT ***-õP*** **‘I’**: PMJ ***#-u*** : PT ***-o-*** **‘meat (abs.)’**: PMJ ***-ĩt*** : PT ***-ẽT***
	*/ʔ/	**‘hole’**: PMJ ***-kuñ°*** : PT ***#-ǩãʔãC*** **‘son’**: PMJ ***#-jayC*** : PT ***#-jaʔuP*** or ***#-jaʔuT***
*/j/		**‘to stand’**: PMJ ***-ja-m*** : PT ***-ʔãP***
	*∅	**‘smoke’**: PMJ ***-ñĩjəƙ*** : PT ***-jĩːK***

Table 4. Sound correspondences between Macro-Jê and Tupian onsets

PMJ	PT	examples
*a	*a	**'to stand'**: PMJ ****-ja / *-ja-m*** : PT ****-ja / *-ʔãP***
		'son': PMJ ***#-jayC*** : PT ***#-jaʔuP*** or ***#-jaʔuT***
		'to wake up': PMJ ***#-paK(°)*** : PT ****-paK***
	*ə	**3CRF prefix**: PMJ ****ta-*** : PT ****tə-***
		'arm': PMJ ****-pa*** 'arm' : PT ****-pə / *mbə*** 'hand'
		'bitter': PMJ ***#-ndap°*** : PT ****-ðəP***
*â	*ɨ	**'foot'**: PMJ ****-pâr°*** : PT ****-pɨ / *mbɨ***
		'liver': PMJ ****-mbâ*** : PT ****-pɨ(-)ʔa / *mbɨ(-)ʔa***
		'white': PMJ ***#-kân°*** : PT ****-ǩɨT***
*ə̂	*o	**'sweet'**: PMJ ****-jə̃ñ*** : PT ***#-joC***
*y	*ɯ	**'to arrive'**: PMJ ***#(-)ɯy₁c°*** : PT ****-ɯɯC***
		'to burn': PMJ ****(-)py₁k° ~ *(-)py₁ŋ°*** : PT ****-pɯK***
		'tree(-like)': PMJ ****(-)ky₁m°*** : PT ****(-)kɯP***
		'earth': PMJ ****ŋgyñ°*** : PT ****kɯC***
		'son': PMJ ***#-jayC*** : PT ***#-jaʔuP*** or ***#-jaʔuT***
*e	*e	**'to go up'**: PMJ ****-wep ~ *-wem°*** : PT ****-we(:)P***
		'name': PMJ ****-jet*** : PT ****-jeT***
		'to go/come': PMJ ****tẽ / *-tẽ-m*** : PT ****-tẽP***
*ê		**'ripe'**: PMJ ****-ndêp°*** : PT ***#-ðeP***
*i	*i	**3NCRF prefix**: PMJ ****i-*** : PT ****i-***
		'to kill': PMJ ****-wĩ*** : PT ***#-wĩ***
		'smoke': PMJ ****-ñĩjə̃k*** : PT ****-jĩːK***
	*e	**'to do, to say'**: PMJ ***#-ki*** : PT ****-ǩe***
		'to enter': PMJ ***#ŋgi₂*** : PT ****-ǩe***
		'husband': PMJ ****-mbi₂n*** : PT ****-mẽT***
		'meat': PMJ ****ĩt / *-ñĩt*** : PT ****ẽT / *-jẽT***
*o	*o	**'father'**: PMJ ****-jo₂m*** : PT ****-joP***
		'to ingest': PMJ ****-ko₂*** : PT ****-ǩo***
		'pus': PMJ ****-jo₂w°*** : PT ****-joP***
*u	*a	**'tooth'**: PMJ ****-juñ°*** : PT ****-jãC***
		'hole': PMJ ****-kuñ°*** : PT ***#-ǩãʔãC***
	*o	**'to dig'**: PMJ ****-kut*** : PT ***#-ǩoT***
		'to give': PMJ ****-ũp*** : PT ****-õP***
		'I': PMJ ***#u*** : PT ****o-***
		'sour': PMJ ***#-juk*** : PT ***#-joK***
non-recurrent		**'bat'**: PMJ ***#nĵip°*** : PT ****jɯP***
		'feces': PMJ ****-ñVt°*** : PT ****-jVT***
		'to pierce': PMJ ***#-py₁k° ~ #-py₁ŋ°*** : **PT** ***#-poK***
		'smoke': PMJ ****-ñĩjə̃k*** : PT ****-jĩːK***

Table 5. Sound correspondences between Macro-Jê and Tupian vowels

'meat', 'to give', 'feces'). An exception is constituted by four cognate sets which show a PMJ oral vowel corresponding to a PT nasal vowel (5).

(5) PMJ PT
  a. 'to stand' *-*ja-m* *-*ʔãP*
  b. 'husband' *-*mbi$_2$n* *-*mẽT*
  c. 'tooth' *-*juñ°* *-*jãC*
  d. 'hole' *-*kuñ°* #-*k̆ãʔãC*

In all these cognate sets, the oral vowel in PMJ is followed by a nasal coda. It is tempting to assume that the respective Proto-Macro-Jê–Tupian etyma likewise contained an oral vowel followed by a nasal coda, and that the vowel became nasal in Tupian by assimilating the nasality of the erstwhile coda. However, there are also several cognate sets which feature an oral vowel followed by a nasal coda in PMJ, yet the PT cognate has an oral vowel (6).

(6) PMJ PT
  a. 'white' #-*kân°* *-*k̆ɨT*
  b. 'sweet' *-*jə̂ñ* #-*joC*
  c. 'tree(-like) *(-)*kɣ$_1$m°* *(-)*ɬɯP*
  d. 'earth' *ŋgɣñ°* *ɬɯC*
  e. 'father' *-*jo$_2$m* *-*joP*

Even though the data are too scarce to warrant a firm conclusion, it is noteworthy that the examples in 5 and 6 involve different vowel qualities: pre-PT **a* and **e* did undergo nasalization to PT **ã* and **ẽ* before an erstwhile nasal coda, whereas PT **ɨ*, **ɯ*, and **o* show no signs of such a process.

In sum, 37 out of 38 candidates for cognate sets (with the exception of 'to pierce') show recurrent sound correspondences, or a reasonable explanation is available as for why the sound correspondences are not demonstrably recurrent.

5.2. Additional sound correspondences

In the preamble of section **4**, I defined the criteria for the cognate search as follows: (i) all PMJ and PT consonants are required to fully match in their place of articulation, (ii) correspondences involving a back vowel in one protolanguage and a front vowel in another are disallowed. Of course, it is perfectly possible that at least some sound correspondences between PMJ and PT violate these constraints: cross-linguistically, it is very common for consonants to diachronically change their place of articulation (or to be lost altogether), and for vowels to diachronically change their backness value. Therefore, any cognates displaying such sound correspondences remained undetected in my initial cognate search. Moreover, my criteria rendered it impossible to detect any Tupian cognates for PMJ stems with complex onsets (*/pr/, **/kr/, **/mr/, **/ŋr/), because they could not be matched to anything in PT, which lacks complex onsets. This section explores the possibility of identifying cognates and sound correspondences that were overlooked in **4** and **5.1** due to the stringency of my initial criteria.

5.2.1. PMJ complex onsets corresponding to PT simple onsets

I start by discussing a group of possible cognates that involve a complex onset in PMJ. At least PMJ */mr/ corresponds to a simple onset */m/ without a rhotic in PT, as shown by the following two examples.

‘ashes’: PMJ ***(-)mbrôŋ*** : PT ***-mboᴋ***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 381. Reflexes are found in the Western (Arikapú *pikä-brä*, Rikbaktsa *poro* ‘bamboo, salt’, Ofayé *kŏtah* /ktah/) and Eastern (Maxakalí *putohok*, Krenak *proŋ* ‘charcoal’, Khĩsêtjê *-mbro*, Xavante *-pro* ‘foam’ in *ö-dzai-pro* ‘foam’, *wede-pro* ‘coffee’, *-dzadai-pro* ‘saliva’, Laklãnõ *mlã*) branches. The correspondences are regular, except that Kaingang *mrẽj* shows an unexpected final consonant, and Canela (and possibly other Timbira varieties) has the relational allomorph *-hpro* instead of the expected **-mpro*, which must be a back-formation from the regular absolute allomorph *pro*.

The Proto-Tupian term for ‘ashes’ is not readily reconstructible (Nikulin 2020:60, fn. 47). One promising candidate is PT **-mboᴋ*. It has semantically shifted reflexes in Tuparikém (Tuparí *-{a}pok* ‘foam’, *-épa-pok* ‘rheum’) and a variety of morphologically complex reflexes in the Eastern branch: Xipaya *-{pu}búk-a* ‘ashes’ (‹-pụβúka› in Nimuendajú 2013: 205), Kawaiwete *-{’}muk* ‘powder’, Parintintin *yvy-mu~{’}mbug* ‘powder-like dust’. The preglottalization in Kawaiwete and Parintintin points to PTG **-ʔmbuᴋ*, a form that probably results from vowel syncope and goes back to earlier **-ʔĩmbuᴋ*. The latter form is most clearly seen in the Proto-Awetí–Guaranian compound **tatʲa-ʔĩpuᴋ* ‘ashes’ (literally ‘fire-powder’), as reflected in Awetí *taza-’ipuk* ‘ashes’ and Ka’apor *tat-imbuk*. Quite surprisingly, TG languages other than Ka’apor do not reflect PTG **tãt-ĩmbuᴋ*, but rather **tãnĩmbuᴋ*: Kawaiwete *tanimuk*, Parintintin *tanimbug* ‘ember’, Apyãwa *tanimok-a*, Old Tupí *tanimbuk-a*, Guarasugwe *tanímɨ* (with an irregular final vowel), etc. Be it as it may, Proto-Awetí–Guaranian **-ʔĩpuᴋ* appears to go back to a morphologically complex form, where the element **-puᴋ* goes back to PT **-mboᴋ*, and the origin of the element **ʔĩ-* is unclear. PT **mb* (underlying */m/) is reconstructed based on the Wayoró reflex *-{a}mbo* ‘foam’, though the velar coda is unexpectedly lacking in that form.

‘snake’: PT ****mbəc*** and Proto-Jabutian ****mrãj***

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31. Reflexes are found in all branches except Tuparikém, including Kepkiriwat (‹bôi›, ‹boi-›), Rama-Puru (Karo *mãy{gãra}*, Puruborá *məŋ{ũ}ᴘ*), Mondé (Zoró *baj*), and Eastern (Yudja *hutá*, Mundurukú *pùy-bu*, Sateré-Mawé *moi*, Apyãwa *maj-a*). The correspondences are regular, except that the Awetí reflex *mõi* has an unexpected nasal vowel.

The Proto-Jabutian term for ‘snake’ is reconstructed as **mrãj* in Voort 2007: 161 based on Arikapú *mrãy* and Djeoromitxí *mẽ*. It lacks known cognates in other Macro-Jê languages; Rikbaktsa *pyryhyk* displays some superficial similarity and could be partially cognate, though details of this etymology have not been worked out. However, no stronger candidates for the Proto-Macro-Jê root for ‘snake’ are known either. It has been proposed that Proto-Core Maxakalian **kãŋã*, Proto-Goyaz **kaŋə̃*, and Ofayé *koni* are cognate (Gudschinsky 1971: 12; Nikulin 2015: 287, 297), but these forms show no regular sound correspondences whatsoever.[25] If the Jabutian root is an archaism, the PMJ form can be hypothesized to have been similar to **mrə̃J*.

I have not succeeded at identifying other plausible cognate sets involving PMJ onsets of the type **Cr*. There is some similarity between PMJ **ŋrṼn°* ‘toucan’ (Nikulin 2020: 420;[26] Eastern only) and the second syllable of PT **jõkãᴛ* ‘toucan’ (the reflexes in the Arikém languages

[25] Other Macro-Jê branches employ clearly noncognate root(s) for this meaning: Krenak *ŋgraŋ*; Malalí ‹checheem›, ‹háhim›; Proto-Akuwẽ **wa:hi* (venomous), **amke* (non-venomous); PSJ **pə̃n*; Proto-Karajá **hemə̃lãlã*; Rikbaktsa *pyryhyk*; Proto-Chiquitano **išoβo- ~ *išoβu-*.

[26] Nikulin (2020: 420) actually reconstructs **ŋrṼt ~ *ŋrṼn°*, but the former variant can be excluded based on the Timbira reflex, which has a long vowel.

point to **juĩʔãkãT*), but this comparison involves too many irregularities to be accepted. Another match that should be viewed as spurious is the comparison between Proto-Cerrado **prə̃m*' 'blackfly' (whence Canela–Krahô *prãm-re*, *prãm-ti*, Akwẽ-Xerente *mrãm-rê*; Nikulin 2020:447) and Proto-Tuparikém **mẽrẽP* 'fly' (whence Makurap *mẽrõ{ã}*, Wayoró *mĩrĩm{a}*, Tuparí *mẽrém{'a}*, Karitiana *mãrãm*, etc.), which shows poor distribution in both language families and non-recurrent sound correspondences.

5.2.2. PMJ palatal coda corresponding to PT zero

There are four pairs of Macro-Jê and Tupian cognate sets with identical or similar meanings where the PMJ (PJ, PCerr) form has a palatal coda, and the Tupian comparanda lack a coda altogether. It is thus possible that some kinds of palatal codas in the hypothetical Proto-Macro-Jê–Tupian language were deleted in the phonological history of Proto-Tupian.

'urine': PMJ ***-jə̂c*** : PT ***-jɨ(:)***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. Reflexes are found in Chiquitano (Bésɨro *-ɨ'ɨ*, not listed in Nikulin 2020) and the Eastern branch (Maxakalí *-xux* 'to urinate', Khĩsêtjê *-tá*, Xavante *-dzé*, Kaingang *-já-nh* 'to urinate'). Nikulin (2020: 400) also lists a putative Karajá reflex, given as **-lə̂*, but the correct Proto-Karajá reconstruction must be **-ly* 'to urinate', whose reflex is attested e.g. in the form *a-r-i-ly=kre* 'I will urinate (male speech)' (Karajá et al. 2013: 5, for some reason with a transitive prefix *i-*). Proto-Karajá **y* does not correspond to Maxakalí *u* /ɨ/ or Proto-Cerrado **ə*, and **-ly* is thus noncognate with the remaining forms. The Chiquitano reflex shows complexities as well. In addition to the well-attested relational stem **-ɨ̃ʔɨ*, there is also a similar absolute stem **jɨʔɨ̃-ʂɨ* 'urine', whose reflex is attested as ‹yiïs› in the 18th century and as *iü-rch* in the Brazilian variety of Eastern Chiquitano (Santana 2012: 258). In my field data, *∅-yɨ'ɨ-j* (*y-ɨ'ɨ-j*?) is documented as an irregular third-person singular form of *-ɨ'ɨ* (thus 'his/her/its urine'), but I concede that this may be a misanalysis on my part, and it is possible that *yɨ'ɨ-j* could in fact be an absolute (unpossessed) term for 'urine' in Migueleño as well. Moreover, the 18th-century materials suggest that the relational stem for 'urine' takes a thematic consonant and thus has the shape **-cɨ́ʔɨ* (whence ‹ziï› *∅-zɨ'ɨ* 'my urine', ‹oziï› *o-zɨ'ɨ* 'our (*INCL*) urine'), which matches the data of other Macro-Jê languages but not of the contemporary Chiquitano varieties (Migueleño *ixh-ɨ'ɨ* / *iy-ɨ'ɨ* 'my urine (female / male speech)', Bésɨro *n-ixh-ɨ'ɨ* 'my urine').[27]

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 29. Reflexes are found at least in the Tuparikém (Wayoró *-ndi-gu*, Sakurabiat ‹-tĩ›, Karitiana *-si*, Arikém ‹-sĩ›) and Eastern (Sateré-Mawé *-sy*, Apyãwa *ty-∅*) branches, with possible cognates in Rama-Puru and Mondé. The sound correspondences are regular, except that the Sakurabiat and Arikém forms (both attested in pre-modern sources only) show a long vowel, which does not match the evidence from Wayoró, Karitiana, and Sateré-Mawé.

In both language families, the root in question is reconstructed as a class II relational stem, except for the aforementioned complexities in Chiquitano. The palatal coda, reconstructed for PMJ based on the Maxakalí reflex *-xux* and the Southern Jê verbalized reflex (Kaingang and Laklãnõ *já-nh* 'to urinate'), has prevented this comparison from being cited in **4.2**. The correspondence between PMJ **ə̂* and PT **ɨ* is, however, a non-recurrent one (though the only other comparison that instantiates PMJ **ə̂* is quite weak, being represented by just one language on the Tupian side). If the Proto-Macro-Jê etymon of Maxakalí *-ptux* 'heavy' — a possible cognate

[27] I thank Luca Ciucci for bringing my attention to the 18th-century forms.

of PT *-*pətɨc* — is to be reconstructed as *#-pVtâJ(°)* (as opposed to *#-pVtyJ(°)* or *#-pVtỹJ(°)*), one could argue that the correspondence between PMJ **â* and PT **ɨ* is in fact the regular one, whereas the similarity between PMJ **-jân̂* and Tuparí *-hoy* could be spurious.

‘another, a’: PMJ ***-n̥ũc ~ *-n̥ũ$_1$j ~ *-n̥ũ$_1$j°*** : PT ***-nõ***

The Proto-Macro-Jê form is reconstructed as **-nũJ* in Nikulin 2020: 390 based on reflexes in Karajá (*-nõ*), Maxakalí (*-nõy*), and Timbira (Canela–Krahô *-hnõ*, Pykobjê–Krĩkatí *-’no*, Parkatêjê *-nõ*). Ramirez et al. (2015: 256) identify additional cognates in other Jê languages (Apinajé *-hõ*, Mẽbêngôkre *-’õ*, and Kaingang/Laklãnõ *ũ*), a suggestion rejected in Nikulin (2020) due to apparent phonological irregularities. However, it has since been established (Nikulin & Salanova 2022: 138) that the correspondence between Proto-Timbira */ˀn/, Apinajé /ʔ/, Mẽbêngôkre /ʔ/, and Khĩsêtjê and Kajkwakhrattxi /t̪^{h}/ is a regular one, and that it goes back to a distinct segment of the protolanguage, despite being exceedingly rare in the lexicon. Therefore, Nikulin’s (2020) criticism of Ramirez et al.’s (2015) proposal is invalid. Additional cognates are Khĩsêtjê and Kajkwakhrattxi *-thõ*. I amend the PMJ reconstruction to **-n̥ũ$_1$c ~ *-n̥ũ$_1$j ~ *-n̥ũ$_1$j°*. The voiceless nasal */n̥/ has not been posited in earlier works on PMJ phonology. I contend that positing PMJ **n̥* as an independent phoneme helps accounting for the otherwise inexplicable reflexes in Southern Jê (**ũ*, with no onset), in the Karajá male genderlect, and in the Javaé dialect of Karajá (*-õ*; Ribeiro 2012a: 139–141). Therefore, PMJ **n̥* must have been preserved in PJ (as well as in PCerr and Proto-Goyaz) and Proto-Karajá. In PNJ, it evolved into **ˀn̥* > Khĩsêtjê and Kajkwakhrattxi /t^{h}/, Apinajé and Mẽbêngôkre /ʔ/, Parkatêjê /n/, and Canela–Krahô and Pykobjê–Krĩkatí /ˀn/. It was independently lost in PSJ and in the male genderlect of Karajá (and in the Javaé dialect). In Maxakalí and in the female genderlect of Karajá, it yielded *n* (underlying /d/).

The Proto-Tupian form has a limited distribution: it is preserved in two Rondonian branches only, Tuparikém (Wayoró *-nõ* ‘another’ < Proto-Tuparian **-nõ*; Nikulin & Andrade 2020: 306) and Rama-Puru (Karo *-nõ* ‘one of’; Gabas Jr 1999: 30, 2013).

PT **-nõ* is functionally identical to PMJ **-n̥ũc ~ *-n̥ũ$_1$j ~ *-n̥ũ$_1$j°*, and the phonological similarity is striking. Only the palatal coda, reconstructed for PMJ based on the Maxakalí reflex *-nõy*, has prevented this comparison from being cited in **4.2**.

‘wet’: Proto-Cerrado ***-ŋgoñ*** : Akuntsú ***-ko***

The Proto-Cerrado form is given as **-ŋgoj’* in Nikulin 2020: 473, reflected as Khĩsêtjê *-ngo* and Akwẽ-Xerente *-koi* // *-ko*, among others. This reconstruction must be updated to **-ŋgoñ* based on the long vowel in the Pykobjê–Krĩkatí reflex *-ncoo* (see fn. 2). No cognates in other Macro-Jê languages are known, but no stronger candidates for the Proto-Macro-Jê root for ‘wet’ are known either.[28] If Proto-Cerrado **-ŋgoñ* is a retention from Proto-Tupian, the original form must have been **-ŋgôñ°*.

On the Tupian side of the comparison, one finds Akuntsú *-ko* ‘wet’ (Aragon 2014: 138), with no known cognates elsewhere in Tupian; even the closely-related Tuparikém languages show noncognate terms for ‘wet’ (Wayoró *-txuup*, Tuparí *-súm-’e*, Makurap *-wuyo*, Karitiana *-sebok*). No stronger candidate for the Proto-Tupian term for ‘wet’ is known.[29] If Akuntsú *-ko* is an archaism, the original form could have been **-ko*, **-k̆o*, or **-ŋgo*.

[28] Each Macro-Jê branch employs its own root(s) for this meaning: Krenak *hĩñot*; Maxakalí *-pato*; PSJ **-paŋpe* (the root is likely just **-pe*; compare also Kaingang *mrér* ‘wet’); Proto-Karajá **-tuku*; Ofayé ‹penó›; Rikbaktsa *-bibi* and *-hõrõ ~ -hõ*; Arikapú *-ü*; Djeoromitxí *-boi* and *-bu*; Proto-Chiquitano **pãʔã-*.

[29] Each Tupian branch employs its own root(s) for this meaning: Puruborá *i(-)pəc* (Monserrat 2005: 19), Paiter *siab*, Mundurukú *-dírem*, Yudjá *-’úrú* and Xipaya *-súru*, Sateré-Mawé *-’apuk*, PTG **-ãkɨ̃p*.

If the forms PMJ *-ŋgôñ° ‘wet’ and PT *-ko ~ *-ɬko ~ *-ŋgo ‘wet’ actually existed, they may have been cognate. No parallels are known that would reveal the PT correspondence for PMJ *ô, the initial consonants correspond well, and the coda shows a mismatch that is precisely the object of discussion in this subsection.

‘water’ / ‘liquid’: Proto-Cerrado ***ŋgôj’** and and Proto-Tuparikém ***-ŋgɨ̈** ‘liquid’

The Proto-Cerrado term for ‘water’ is from Nikulin 2020: 473, where the apostrophe stands for the absence of an echo vowel. It is reflected, for example, as Khĩsêtjê *ngô*, Panará *inkô*, and Xavante *ui // u* ‘still water’. Despite the striking similarity, it does not regularly correspond to PSJ *ŋgôj ‘water’ (> Kaingang *goj* and Laklãnõ *goj*): the former form points to PJ *ŋgə̂$_1$j’, and the latter to *ŋgu$_1$j, with a different nucleus and a different coda. Similar, but unrelated, are the Proto-Akuwẽ noun *kəj // *kə ‘flowing water’ (< PCerr *wyj’ or *wyñ ‘river, whence Mẽbêngôkre *by-ti-re* ‘Xingu River’), the PCerr form *-ŋgoñ ‘wet’ (see above), and the Maxakalí noun *kõnãg-kox* ‘river’ (possibly from *kõnããg* ‘water’ and *-kox* ‘hole’). Maxakalí *-kux* ‘riverbank’ is phonologically comparable with PCerr *ŋgôj’ ‘water’, but a semantically closer cognate is available, PNJ *{ca}kəc ‘riverbank’ (> Khĩsêtjê *sakhát // sakhárá*). I hesitate at deciding whether PCerr *ŋgôj’ ‘water’ and PSJ *ŋgôj ‘water’ should be considered cognate; in any case, these roots are an innovation, since the PMJ term for ‘water’ is clearly reconstructible as *mbi$_1$n°. If PCerr *ŋgôj’ is a semantically shifted reflex of a PMJ noun, its original form should be reconstructed as *ŋgy$_1$j or *ŋgy$_1$j°.

On the Tupian side of the comparison, one finds Proto-Tuparikém *-ŋgɨ̈ ‘liquid’, reflected, among others, as Wayoró *-ngu* ‘liquid’ and Karitiana *-nge* ‘blood’ (Nikulin & Carvalho 2022: 33). The Karitiana meaning is clearly innovative, since the closely related Arikém retains a pan-Tupian root for ‘blood’ (‹nhaé›, ‹nyaë› < PT *-jəɯ; Nikulin & Carvalho 2022: 29). No cognates in other Tupian languages are known, and the Proto-Tupian term for ‘liquid’ is reconstructed as *-jɯ (Nikulin & Carvalho 2022: 30). If Proto-Tuparikém *-ŋgɨ̈ is nevertheless a semantically shifted reflex of a PT noun, its original form should be reconstructed as *-ŋgɯ.

Proto-Cerrado *ŋgôj’ is reconstructed as an absolute stem, whose relationalized equivalent is *-ga-ŋgôj’ (> Khĩsêtjê *-kangô*, Panará *nankô* / *-rankô*, Xavante *-wa’u*). Proto-Tuparikém *-ŋgɨ̈ is a relational class I stem. The palatal coda in Proto-Cerrado is reconstructed based on the Xavante and Akwẽ-Xerente utterance-medial allomorphs, and it lacks a correspondence in Tuparikém. The correspondence PMJ *ŋg */ŋ/ : PT *ŋg */ŋ/ is unparalleled, but still imaginable, and PMJ *y$_1$ does correspond to PT *ɯ.

5.2.3. PMJ back vowel corresponding to PT *e

Nikulin (2020: 188–189) reconstructs the PMJ second-person pronoun (internal case) as *a, with reflexes such as Ofayé *e:*, *e-*, Krenak *ho-ti*, Panará *ka*, and Kaingang *ã*. In addition, there is a similar second-person index, reconstructed as having two allomorphs, *a- with class I stems and *∅- with class II stems (Nikulin 2020: 208–219). Its proposed reflexes are found in all major branches, including Chiquitano (Bésɨro *a-*/*∅-*), Western (Djeoromitxí *a-*, Rikbaktsa *a-*, Ofayé *ə-*/*∅-*), Karajá (*a-*/*∅-*), and Eastern (Maxakalí *ã-*/*∅-*, Krenak *a-*/*h-*, Khĩsêtjê *a-*/*ng-*/*k-*,[30] Panará *a-*/*k-*,

[30] Khĩsêtjê *k-* (the expected reflex of PMJ *∅- before oral vowels, via *∅ > *ɦ > *ɯ̨ > *g > *k) no longer functions as an inflectional marker: it was ousted by the allomorph *ng-* (originally found before nasal vowels only, as in *ng-ĩnti* ‘your name’ or *ng-ũmndât // ng-ũmndârâ* ‘your wrist’), and forms such as *ng-ahrâ* ‘to play with you’ and *ng-ajkhêrê* ‘you yawn’—instead of the expected *k-ahrâ, *k-ajkhêrê—are found in Khĩsêtjê. The allomorph *k-* is preserved in the triadic kinship terms *k-áthẽng // k-áthẽngẽ* ‘your son, who happens to be my relative’, *k-átôt(-jê) // k-átôrô* ‘your mother, who happens to be my in-law (avoidance woman)’.

Xavante *a-* /a:-/ ‘second-person honorific’). Both cognate sets show somewhat irregular reflexes in some languages. The reconstruction **a* ‘you’ does not account for the Southern Jê reflexes (Kaingang *ã*, Laklãnõ *a* ‘you’), which rather point to PMJ **u*; normally PMJ **a* yields PJ **a* > PSJ **ã* > Kaingang *ẽ*, Laklãnõ *ã*. The reconstruction **a-* ‘second-person index’ fails to account for the reflexes in Ofayé (*ə-* instead of the expected **e-*; points to PMJ **ə̂-* or **y-*) and Krenak (*a-* instead of the expected **o-*; points to PMJ **ə-*). It is important to note that person indices are unstressed in almost all Macro-Jê languages (Chiquitano is an exception), and the development of unstressed PMJ vowels is understudied. In any case, all reflexes of the pronoun and person index point to a PMJ back vowel (be it **a*, **u*, **ə*, **ə̂*, or **y*).

In Proto-Tupian, **e-* is reconstructed as an absolutive/genitive second-person prefix, from which the pronoun **ẽT* is derived, just like the pronoun **õT* is derived from the first-person prefix **o-*. It has reflexes in almost all Tupian languages. Before consonant-initial stems, it is reflected as Makurap *e-*, Wayoró *e-*, Karitiana *a-*, Karo *e-*, Puruborá *ɛ-*, Paiter *e-*, Yudja *e-*, Mundurukú *e-*, Sateré-Mawé *e-*, Apyãwa *e-* ‘2*CRF*’, among many other reflexes. Before vowel-initial stems, it shows asyllabic allomorphs in some languages, such as Yudja *l-*. The Tupi–Guaranian reflex is only used anaphorically, particularly when a second-person possessor on a noun or a second-person argument of a gerund of an intransitive verb is coreferential with some other participant. As for noncoreferential uses, it has been ousted by the clitic **(e)nde=* in the Tupi–Guaranian languages.

The possibility of linking the Proto-Macro-Jê second-person markers and the Proto-Tupian index **e-* was not considered in **4** due to the mismatch in vowel backness. Although the sound correspondence is not recurrent, the cognation hypothesis is still plausible, since irregular vowel changes are otherwise known to be common in grammatical morphemes (as seen in the Macro-Jê cognate set discussed in this subsection).

6. Conclusion

In this article, I have assembled the lexical evidence supporting the proposed common origin of the Macro-Jê and Tupian families. Despite their limited number, the matches analyzed in **4**–**5** show recurrent sound correspondences that are hardly attributable to chance or to language contact: they mostly involve basic vocabulary (including 19 items on the 110-item Swadesh list: ‘meat’, ‘name’, ‘smoke’, ‘tooth’, ‘ingest’ = ‘to eat/drink’, ‘tree’, ‘liver’, ‘foot’, ‘burn’, ‘to give’, ‘to stand’, ‘earth’, ‘to kill’, ‘white’, ‘I’, ‘heavy’, ‘to go/come’, ‘ashes’, ‘snake’) and grammatical morphemes, and multiple proposed cognate sets involve data from Macro-Jê and Tupian branches spoken very far from each other (e.g. Jê and Karitiana). Moreover, the Macro-Jê–Tupian comparanda often involve matching codas, but the Eastern branch of Macro-Jê—the one that most faithfully preserves PMJ codas—is geographically removed from Rondônia, where most (non-Tupi–Guaranian) Tupian languages are spoken. Therefore, the similarities noted above can hardly result from language contact, and common genetic origin is the best explanation available.

Further research will need to concentrate on the lexical reconstruction of Proto-Macro-Jê and Proto-Tupian. In this article, I have proposed multiple hypothetical PMJ and PT forms based on reflexes in only one branch or language; I predict that some of these etymologies can be further strengthened by identifying previously unnoticed cognates in the attested languages. As of now, relatively few Macro-Jê and Tupian etymologies are currently known. Nikulin’s (2020) dissertation lists 188 PMJ reconstructions, of which some are quite dubious, and others involve reflexes in one first-level branch only. Although no comprehensive source on

Tupian etymology exists so far, I am currently working on a Tupian etymological dictionary, and my draft has 255 entries that involve reflexes in more than one first-level branch (including compounds). It is certainly possible to reconstruct many more PMJ and PT forms.

Another direction for further research is to include other language families into consideration. In my opinion, language families and isolates such as Cariban, Bororoan, Karirian, and Yaathê are very likely related to Macro-Jê and Tupian, but a search for possible cognates in these languages is complicated by the fact that Proto-Cariban, Proto-Bororoan, Proto-Karirian, and pre-Yaathê do not have codas (except for the marginal coda *-*j* in pre-Yaathê; Silva forthc.). If some or all of these languages are ultimately related to Macro-Jê and Tupian, it is unclear whether Macro-Jê and Tupian codas should be matched to zero (under the assumption that **C(r)VC*-structures yielded **C(r)V* in these languages), or whether Macro-Jê **C(r)VC*-structures and Tupian **CVC*-structures should be matched to polysyllabic roots in other languages (under the assumption that **C(r)VCV*-structures yielded **C(r)VC* in Macro-Jê and Tupian). This uncertainty leaves too much room for false positives at this stage of investigation. Some other language families — notably Katukina–Harakmbut, Mataguayan, and Guaicuruan — have comparable syllable structures with robust codas. Indeed, there are several promising lookalikes with matching onsets and codas involving these families, as in PT **-ɨkaT* : Proto-Mataguayan **-kɑ̊ˀt-* : Harakmbut *-kot* ‘to fall’; PMJ **(-)mbrôŋ* : PT **-mboK* ‘ashes’ : Proto-Mataguayan **-mɑ̊ˀk* ‘powder’; PT **atʲa* ‘fire’ : Katukina *ita*, Harakmbut *’uta’* ‘firewood’ (but Proto-Mataguayan **ʔítɑ̊χ* ‘fire’, with a uvular coda); Proto-Tupian **ɔK* : Katukina *hak*, Harakmbut *jak* /hak/ ‘house’; PMJ **-mbâ* : Proto-Tupian **-pɨ(-)ʔa*/**mbɨ-ʔa* : Katukina *ma*, Harakmbut *-me’* ‘liver’. If more of such matches are found and if regular sound correspondences are identified linking the aforementioned languages, the Macro-Jê–Tupian hypothesis may turn out to be the tip of an iceberg — quite possibly, the largest macrofamily in the Americas.

Acknowledgments

I acknowledge the generosity of all speakers and researchers of Macro-Jê and Tupian languages who have dedicated their time to answering my questions about the meaning, form, spelling, and inflection of so many words. In particular, I am grateful to Andrés Pablo Salanova, Antônia Fernanda Souza Nogueira, Antonia Socoré Masaí, Chadawa Juruna, Ignacia Tomichá Yopié, Jamthô Suyá, Khawiri Suyá, Luca Ciucci, Mário André Coelho da Silva, Mawaré Juruna, Orengô Tapayuna, Ricardo Capêrkô Canela, Talles Huryty Tapirapé, Tawayaku Kayabi, Tenywaawi Tapirapé, Victoriano Julián Laverán Ramos, and Yrywaxã Tapirapé. I thank an anonymous JoLR reviewer for their comments on the manuscript. An early version of this paper was presented at II Seminário do GELCIA / II SIPLI-Norte (Federal University of Pará, Belém) on November 18, 2019; I thank the audience for their comments.

Provenance of linguistic data

Unreferenced examples come from the following sources.

Akwẽ-Xerente	Krieger & Krieger 1994
Akuntsú	Aragon 2008, 2014
Apinajé	Albuquerque 2012
Apyãwa	Almeida et al. 1983; Tenywaawi Tapirapé, p. c.; Yrywaxã Tapirapé, p. c.
Arikapú	R. Ribeiro 2008; Arikapú et al. 2010
Arikém	Nimuendajú 1932; Rondon & Faria 1948

Aruá	Sekelj 1948
Awetí	Sabino 2016
Bésiro	Parapaino Castro 2008
Canela–Krahô	Grupp 2015
Djeoromitxí	M. Ribeiro 2008
Eastern Chiquitano	Fuss & Riester 1986
Gavião	Moore 1984; Felzke & Moore 2019
Guarasugwe	Ramirez et al. 2017
Harakmbut	Tripp 1995
Ka'apor	Kakumasu & Kakumasu 2007
Kaingang	Wiesemann 2011
Kajkwakhrattxi	Camargo 2010
Kamayurá	Seki 2000
Karajá	Ribeiro 2012a; Karajá et al. 2013
Karitiana	Landin 2005, Rocha 2011; Storto 2019
Karo	Gabas Jr 1999
Katukina	dos Anjos 2011
Kawaiwete	Weiss 2005
Kepkiriwat	Rondon & Faria 1948
Khĩsêtjê	Nonato et al. 2012; Jamthô Suyá, p. c., Khawiri Suyá, p. c.
Krenak	Seki n/d
Kuruaya	Costa 2002; Picanço 2005, 2019
Laklãnõ	Alves Jr 2014
Makurap	Sekelj 1948; Braga 2005
Malalí	Silva & Nikulin 2021
Maxakalí	Silva 2020; Silva, p. c.
Mbyá	Dooley 2006
Mẽbêngôkre	Salanova, p. c.
Migueleño Chiquitano	own field data
Mundurukú	Crofts 1985; Picanço 2005
Ofayé	Hanke 1964; Gudschinsky 1974; Ribeiro 2004b; Oliveira 2006
Old Tupí	Barbosa 1956
Paiter	Bontkes 1978
Panará	Bardagil-Mas 2018
Paraguayan Guaraní	Centurión Servin & Davalos Arce 2009
Parintintin	Betts 1981
Parkatêjê	Araújo 2016
Pataxó-Hãhãhãe	Silva & Nikulin 2021
Proto-Mataguayan	Nikulin & Carol forthc.
Puruborá	Galucio 2005
Pykobjê–Krĩkatí	Pries 2008
Rikbaktsa	Tremaine 2007
Sakurabiat	Galucio 2001; Snethlage 2015
Salamãy	Galucio et al. 2015
Sateré-Mawé	Ribeiro 2010; Silva 2010
Siriono	Gasparini & Dicarere Méndez 2015
Tapiete	Gonzalez 2005
Tuparí	Alves 2004; Singerman 2018
Wayoró	Nogueira 2011, 2019; Nogueira et al. 2021
Yudja	Fargetti 2001; Chadawa Juruna, p. c.
Xavante	McLeod & Mitchell 1977; Lachnitt 1987
Xipaya	Fargetti & Rodrigues 2008
Zoró	Galucio et al. 2015

Abbreviations

X // Y X is the utterance-medial allomorph, Y is the utterance-final allomorph
{X} fossilized material (noncognate part for which cognation is not asserted)
‹X› material cited *verbatim* after premodern sources
X : Y X corresponds to Y

Grammatical abbreviations: 1/2/3 = first/second/third person; ALZ = alienizer; ANTP = antipassive; AUG = augmentative; AUX = auxiliary; CRF = coreferential; DU = dual; F = finite; IMPF = imperfective; INCL = inclusive; INV = inverse; NCRF = noncoreferential; NF = nonfinite; NMLZ = nominalizer; PL = plural; PSSD = possessed; PRG = progressive; PRS = present; PST = past; REF = referentializer; SG = singular; A = agent; P = patient; V = verb; NP = noun phrase.

Phonological abbreviations: C = consonant; J = palatal consonant; K = velar consonant; N = nasal consonant; V = vowel.

Language names: Guaj. = Guajajara (Tenetehara); PCerr = Proto-Cerrado; PJ = Proto-Jê; PMJ = Proto-Macro-Jê; PNJ = Proto-Northern Jê; PSJ = Proto-Southern Jê; PT = Proto-Tupian; PTG = Proto-Tupi–Guaranian; TG = Tupi–Guaranian.

References

Adelaar, Willem F. H. 2008. Relações externas do Macro-Jê. O caso do Chiquitano. In: Stella Telles, Aldir Santos de Paula (orgs.). *Topicalizando Macro-Jê*: 9–28. Recife: Nectar.

Aikhenvald-Angenot, Alexandra Y., Jean-Pierre Angenot. 1992. The South-American Proto-Ge and the Old World. In: Shevoroshkin, Vitaly V. (ed.). *Nostratic, Dene-Caucasian, Austric and Amerind. Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8–12 November, 1988*: 403–18. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.

Albuquerque, Francisco Edviges (org.). 2012. *Dicionário Escolar Apinayé. Panhĩ kapẽr–kupẽ kapẽr. Apinayé–português.* Belo Horizonte: Editora da Faculdade de Letras–UFMG.

Almeida, Antônio, Irmãzinhas de Jesus, Luiz G. de Paula. 1983. *A língua tapirapé.* Rio de Janeiro: Biblioteca Reprográfica Xerox.

Alves, Poliana Maria. 2004. *O léxico do Tuparí: proposta de um dicionário bilíngüe.* Ph.D. dissertation, Universidade Estadual Paulista "Júlio de Mesquita".

Alves Jr, Ozias. 2014. *Parlons xokleng / laklãnõ. Langue indigène du sud du Brésil.* Paris: L'Harmattan.

Aragon, Carolina Coelho. 2008. *Fonologia e aspectos morfológicos e sintáticos da língua Akuntsú.* MA thesis, Universidade de Brasília.

Aragon, Carolina Coelho. 2014. *A grammar of Akuntsú, a Tupían language.* Ph.D. dissertation, University of Hawai'i at Mānoa.

Araújo, Leopoldina. 2016. *Dicionário Parkatêjê–português.* Belém: s/ed.

Arikapú, Mamoa, Nazaré Arikapú, Hein van der Voort. 2010. *Vocabulário Arikapú–português.* Baltimore: Cadernos de Etnolingüística. (Monografias, 1.)

Barbosa, Pe A. Lemos. 1956. *Curso de Tupi Antigo: gramática, exercícios, textos.* Rio de Janeiro: Livraria São José.

Bardagil-Mas, Bernat. 2018. *Case and agreement in Panará (Naamval en congruentie in het Panará).* Ph.D. dissertation, Rijksuniversiteit Groningen.

Betts, LaVera. 1981. *Dicionário Parintintin–português, português–parintintin.* Cuiabá: Sociedade Internacional de Lingüística.

Bontkes, Willem. 1978. *Dicionário preliminar Suruí–português, português–Suruí.* Porto Velho: Summer Institute of Linguistics.

Braga, Alzerinda. 2005. *Aspects morphosyntaxiques de la langue Makurap/Tupi.* Ph.D. dissertation, Université de Toulouse–Le Mirail.

Camargo, Nayara da Silva. 2010. *Língua Tapayúna: aspectos sociolingüísticos e uma análise fonológica preliminar.* MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2010.770701

Carvalho, Fernando O. de. 2019. Revisitando o Proto-Jurúna: a reconstrução das oclusivas orais. In: Edna dos Santos Oliveira, Eduardo Alves Vasconcelos, Romário Duarte Sanches (eds.). *Estudos linguísticos na Amazônia*: 215–236. Campinas: Pontes Editores.

Carvalho, Fernando O. de. 2022. On the Guaranian evidence for two Proto-Tupi–Guarani affricates. *Journal of Language Relationship* 20(1): 81–112.

Carvalho, Fernando O. de. 2023. Proto-Tupi–Guarani did not have a palatalized velar stop. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 18(1): e20220013. doi:10.1590/2178-2547-BGOELDI-2022-0013

Carvalho, Fernando O. de. Forthcoming. *A comparative reconstruction of Proto-Tupi-Guarani phonology* [provisional title]. Berlin: Language Science Press. (Topics in Phonological Diversity.)

Carvalho, Fernando O. de., Joshua Birchall. 2022. A comparative reconstruction of Proto-Tupi–Guarani kinship terminology. *LIAMES* 22: e022001. doi:10.20396/liames.v22i00.8666489

Centurión Servin, Celsa, Juana Davalos Arce. 2009. *Milibri pocket. Castellano–guaraní, guaraní–castellano: grafía tradicional y científica.* Asunción: Milibri Ediciones.

Ciucci, Luca. 2020. *Los clasificadores del chiquitano (bésɨro).* Talk presented at Kusɨribo Oberabakáx ñomé Mumanityakatoe aukí na Nikɨ Mamonkóka (November 13, 2020). doi:10.13140/RG.2.2.30849.68963

Corrêa-da-Silva, Beatriz Carretta. 2010. *Mawé/Awetí/Tupí-Guaraní: relações lingüísticas e implicações históricas.* Ph.D. dissertation, Universidade de Brasília.

Costa, Raimundo Nonato Vieira. 2002. Fonologia segmental da língua Kuruaya. *Moara* 17: 85–101. doi:10.18542/moara.v1i17.3151

Crofts, Marjorie. 1973. *Gramática Munduruku.* Brasília: Summer Institute of Linguistics.

Crofts, Marjorie. 1985. *Aspectos da língua Mundurukú.* Brasília: Summer Institute of Linguistics.

Davis, Irvine. 1968. Some Macro-Jê relationships. *International Journal of American Linguistics* 34(1): 42–47. doi:10.1086/464994

Dooley, Robert A. 2006. *Léxico Guaraní, dialeto Mbyá com informações úteis para o ensino médio, a aprendizagem e a pesquisa lingüística.* Cuiabá: Sociedade Internacional de Linguística.

dos Anjos, Zoraide. 2011. *Fonologia e gramática Katukina-Kanamari.* Ph.D. dissertation, Vrije Universiteit Amsterdam.

Dourado, Luciana Gonçalves. 2001. *Aspectos morfossintáticos da língua Panará (Jê).* Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2001.206356

Drude, Sebastian. 2009. Nasal harmony in Awetí: a declarative account. *ReVEL — Revista Virtual de Estudos da Linguagem* 7(3).

Drude, Sebastian. 2011. Aweti in relation with Kamayurá: the two Tupian languages of the Upper Xingu. In: Bruna Franchetto (org). *Alto Xingu. Uma sociedade multilíngüe*: 155–192. Rio de Janeiro: Museu do Índio/FUNAI.

Estevam, Adriana Machado. 2011. *Morphosyntaxe du xavante. Langue jê du Mato Grosso (Brésil).* Ph.D. dissertation, Université Paris-Diderot (Paris-VII).

Fargetti, Cristina Martins. 2001. *Estudo fonológico e morfossintático da língua Juruna.* Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2001.222157

Fargetti, Cristina Martins, Carmen L. Reis Rodrigues. 2008. Consoantes do Xipaya e do Juruna — uma comparação em busca do proto-sistema. *Alfa* 52(2): 535–563.

Fargetti, Cristina Martins, Carmen Lúcia Reis Rodrigues. 2021. Vogais do Juruna e do Xipaya — uma comparação. *Moara* 58: 186–204. 10.18542/moara.v0i58.10906

Felzke, Lediane Fani, Denny Moore. 2019. Terminologias de parentesco dos grupos da família linguística Mondé. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 15–32. doi:10.1590/1981.81222019000100003

Fuss, Max, Jürgen Riester. 1986. *Zúbaka. La Chiquitanía: visión antropológica de una región en desarrollo. Tomo I: Vocabulario español–chiquito y chiquito–español.* La Paz: Los Amigos del Libro.

Gabas Jr, Nilson. 1989. *Estudo fonológico da língua Karo (Arara de Rondônia).* MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.1989.38012

Gabas Jr, Nilson. 1999. *A grammar of Karo, Tupi (Brazil).* Ph.D. dissertation, University of California, Santa Barbara.

Galucio, Ana Vilacy. 1994. Fonologia segmental da língua Mekens. In: *Anais do IX Encontro Nacional da ANPOLL (Caxambu, junho de 1994). Área de Lingüística, v. 2*: 988–997. João Pessoa: Associação Nacional de Pós-Graduação e Pesquisa em Letras e Lingüística.

Galucio, Ana Vilacy. 2001. *The morphosyntax of Mekens (Tupi).* Ph.D. dissertation, University of Chicago.

Galucio, Ana Vilacy. 2005. Puruborá: notas etnográficas e lingüísticas recentes. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 1(2): 159–192.

Galucio, Ana Vilacy, Antônia Fernanda de Souza Nogueira. 2012. Comparative study of the Tupari branch of the Tupi family: contributions to understanding its historical development and internal classification. In: *Memorias del V Congreso de Idiomas Indígenas de Latinoamérica, 6–8 de octubre de 2011, Universidad de Texas en*

Austin. Austin: Center for Indigenous Languages of Latin America (CILLA). Available online at: ailla.utexas.org/sites/default/files/documents/Galucio_Nogueira_CILLA_V.pdf, accessed on 24/02/2023.

Galucio, Ana Vilacy, Antônia Fernanda de Souza Nogueira. 2018. From object nominalization to object focus: the innovative A-alignment in the Tuparian languages (Tupian family). *Journal of Historical Linguistics* 8(1): 95–127. doi:10.1075/jhl.16025.gal

Galucio, Ana Vilacy, Sérgio Meira, Joshua Birchall, Denny Moore, Nilson Gabas Júnior, Sebastian Drude, Luciana Storto, Gessiane Picanço, Carmen Reis Rodrigues. 2015. Genealogical relations and lexical distances within the Tupian linguistic family. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 229–274. doi:10.1590/1981-81222015000200004

Gasparini, Noé, Victor Hugo Dicarere Méndez. 2015. *Diccionario Siriono. Siriono–castellano–inglés y castellano–siriono*. Trinidad: Editorial Tiempos del Beni.

Gerardi, Fabrício Ferraz, Stanislav Reichert, Carolina Aragon, Tim Wientzek, Johann-Mattis List, Robert Forkel. 2022. *TuLeD. Tupían Lexical Database* (v0.12). doi:10.5281/zenodo.6572576

Gomes, Dioney Moreira. 2007. Voz média em Mundurukú: uma análise do morfema *je-* (*ʤe-*). In: Ana Suelly Arruda Câmara Cabral, Aryon Dall'Igna Rodrigues (orgs.). *Línguas e culturas Tupí, vol. I*: 317–323. Campinas: Curt Nimuendajú; Brasília: LALLI/UnB.

Gonzalez, Hebe Alicia. 2005. *A grammar of Tapiete (Tupi–Guarani).* Ph.D. dissertation, University of Pittsburgh.

Gudschinsky, Sarah C. 1971. Ofaié-Xavante, a Jê language. In: Sarah C. Gudschinsky (ed.). *Estudos sôbre línguas e culturas indígenas: Trabalhos lingüísticos realizados no Brasil*: 1–16. Brasília: SIL–Instituto Lingüístico de Verão.

Gudschinsky, Sarah C. 1974. Fragmentos de Ofaié: a descrição de uma língua extinta. In: Loraine Irene Bridgeman (ed.). *Série Lingüística, no. 3*: 177–249. Brasília: Summer Institute of Linguistics.

Guérios, Rosário F. Mansur. 1939. O nexo lingüístico Bororo/Merrime-Caiapó (contribuição para a unidade genética das línguas americanas). *Revista do Círculo de Estudos "Bandeirantes"* 2: 61–74.

Greenberg, Joseph H. 1987. *Language in the Americas.* Stanford University Press.

Grupp, Bernhard. 2015. *Dicionário Canela: Canela–português–inglês, português–Canela, inglês–Canela.* 2nd ed. Barra do Corda: Missão Cristã Evangélica do Brasil.

Hanke, Wanda. 1964. *Völkerkundliche Forschungen in Südamerika: Verlöschende Urzeit im Innern Brasiliens (= Kulturgeschichtliche Forschungen).* Braunschweig: Limbach.

Hanke, Wanda, Morris Swadesh, Arion Dall'Igna Rodrigues. 1958. Notas de fonologia Mekens. In: Juan Comas (ed.). *Miscellanea Paul Rivet octogenario dicata, vol. II*: 187–217. Mexico City: Universidad Nacional Autónoma de México.

Harrison, Carl, Carole Harrison. 2013. *Dicionário Guajajara–português.* Anápolis: Associação Internacional de Lingüística–SIL Brasil.

Kakumasu, James Y., Kiyoko Kakumasu. 2007. *Dicionário por tópicos Kaapor–português.* Cuiabá: Associação Internacional de Lingüística–SIL Brasil.

Karajá, Hatawaki, Juanahu Karajá, Leandro Karajá, Teribre Karajá, Wadoi Karajá, Woubedu Karajá. 2013. *Dicionário Enciclopédico Inyrybè/Karajá–português brasileiro.* Rio de Janeiro: Museu do Índio/Prodoclin. Available online at: http://prodoclin.museudoindio.gov.br/images/conteudo/Karaja/produtos_pesquisadores/Karaj%C3%A1_Iny_Lexicon.pdf, accessed on 24/02/2023.

Krieger, Wanda Braidotti, Guenther Carlos Krieger (orgs.). 1994. *Dicionário escolar: Xerente–português; português–Xerente.* Rio de Janeiro: Junta das Missões Nacionais da Convenção Batista Brasileira.

Lachnitt, Georg. 1987. *Romnhitsi'ubumro. A'uwẽ mreme–waradzu mreme. Dicionário xavante–português.* Campo Grande: Missão Salesiana de Mato Grosso.

Landin, David. 2005. *Dicionário e léxico Karitiana / português.* 2nd ed. Cuiabá: Sociedade Internacional de Lingüística.

Lima, Suzi. 2014. *The grammar of individuation and counting.* Ph.D. dissertation, University of Massachusetts Amherst. doi:10.7275/msad-1e04

Martins, Andérbio Márcio Silva. 2009. Um estudo comparativo-lexical das famílias Kamakã e Purí. In: Silvia Lúcia Bigonjal Braggio, Sinval Martins de Sousa Filho (orgs.). *Línguas e culturas Macro-Jê*: 231–238. Goiânia: Vieira.

Martins, Andérbio Márcio Silva. 2011. *Uma avaliação da hipótese de relações genéticas entre o Guató e o tronco Macro-Jê.* Ph.D. dissertation, Universidade de Brasília.

McLeod, Ruth, Valerie Mitchell. 1977. *Aspectos da língua Xavánte.* Brasília: Summer Institute of Linguistics.

Meira, Sérgio, Spike Gildea, B. J. Hoff. 2010. On the origin of ablaut in the Cariban family. *International Journal of American Linguistics* 76(4): 477–515. doi:10.1086/658055

Meira, Sérgio, Sebastian Drude. 2013. Sobre a origem histórica dos "prefixos relacionais" das línguas Tupí–Guaraní. *Cadernos de Etnolingüística* 5(1): 1–30.

Meira, Sérgio, Sebastian Drude. 2015. A summary reconstruction of Proto-Maweti–Guarani segmental phonology. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 275–296. doi:10.1590/1981-81222015000200005

Monserrat, Ruth Maria Fonini. 2005. Notícia sobre a língua Puruborá. In: Aryon Dall'Igna Rodrigues, Ana Suelly Arruda Câmara Cabral (eds.). *Novos estudos sobre línguas indígenas*: 9–22. Brasília: Editora UnB.

Moore, Denny. 1984. *Syntax of the language of the Gavião Indians of Rondônia, Brazil.* Ph.D. dissertation, City University of New York.

Moore, Denny, Ana Vilacy Galucio. 1994. Reconstruction of Proto-Tupari consonants and vowels. In: Margaret Langdon, Leanne Hinton (eds.). *Proceedings of the Meeting of the Society for the Study of the Indigenous Languages of the Americas, July 2–4, 1993, and the Hokan-Penutian Workshop, July 3, 1993, both held at the 1993 Linguistic Institute at Ohio State University in Columbus, Ohio*: 119–137. Berkeley: Survey of California and other Indian Languages.

Nikulin, Andrey. 2015. Apofonia e sistema vocálico do Proto-Jê Meridional: contribuição para estudos comparativos das línguas Jê. *Revista Brasileira de Linguística Antropológica* 7(2): 275–306. doi:10.26512/rbla.v7i2.20601

Nikulin, Andrey. 2017. A phonological reconstruction of Proto-Cerrado (Jê family). *Journal of Language Relationship* 15(3): 147–180. doi:10.31826/jlr-2018-153-404

Nikulin, Andrey. 2020. *Proto-Macro-Jê: um estudo reconstrutivo.* Ph.D. dissertation, Universidade de Brasília.

Nikulin, Andrey. Forthc. Reconstrução fonológica, ou como reconstruir o Proto-Tuparikém. Submitted to: Stella Telles, Cláudio Couto, Emerson Carvalho, and W. Leo Wetzels (eds.). *Tópicos na estrutura sonora das línguas indígenas sul-americanas: uma introdução à análise fonológica.* Available online at: www.academia.edu/85174094/Reconstru%C3%A7%C3%A3o_fonol%C3%B3gica_ou_como_reconstruir_o_Proto_Tuparik%C3%A9m_submetido_, accessed on 24/02/2023.

Nikulin, Andrey, Rafael Andrade. 2020. The rise and fall of approximants in the Tuparian languages. *Journal of Language Relationship* 18(4): 284–319. doi:10.31826/jlr-2020-183-412

Nikulin, Andrey, Javier Carol. Forthc. *Historical phonology of Mataguayan*. Berlin: Language Science Press. (Topics in Phonological Diversity.)

Nikulin, Andrey, Fernando O. de Carvalho. 2018. Prehistoria de las lenguas y familias lingüísticas del Gran Chaco, de la meseta brasileña y cercanías: propuesta de base de datos léxicos y resultados preliminares. In: María Alejandra Regúnaga, Silvia Andrea Spinelli, María Emilia Orden (eds.). *IV Encuentro de Lenguas Indígenas Americanas-ELIA: libro de Actas*: 545–560. Santa Rosa: Universidad Nacional de La Pampa.

Nikulin, Andrey, Fernando O. de Carvalho. 2019. Estudos diacrônicos de línguas indígenas brasileiras: um panorama. *Macabéa — Revista Eletrônica do Netlli* 8(2): 255–305. doi:10.47295/mren.v8i2.1910

Nikulin, Andrey, Fernando O. de Carvalho. 2022. A revised reconstruction of the Proto-Tupian vowel system. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 17(2): e20210035. doi:10.1590/2178-2547-BGOELDI-2021-0035

Nikulin, Andrey, Andrés Pablo Salanova. 2019. Northern Jê verb morphology and the reconstruction of finiteness alternations. *International Journal of American Linguistics* 85(4): 533–567. doi:10.1086/704565

Nikulin, Andrey, Andrés Pablo Salanova. 2022. O enfraquecimento diacrônico de consoantes em Mẽbêngôkre. In: Edna dos Santos Oliveira, Eduardo Alves Vasconcelos, Romário Duarte Sanches (eds.). *Estudos Linguísticos na Amazônia, vol. 2*: 121–143. Campinas: Pontes.

Nikulin, Andrey, Mário André Coelho da Silva. 2020. As línguas Maxakalí e Krenák dentro do tronco Macro-Jê. *Cadernos de Etnolingüística* 8(1): 1–64.

Nimuendajú, Curt. 1932. Wortlisten aus Amazonien. *Journal de la Société des américanistes* 24(1): 93–119. doi:10.3406/jsa.1932.1846

Nogueira, Antônia Fernanda de Souza. 2011. *Wayoro ẽmẽto: fonologia segmental e morfossintaxe verbal.* MA thesis, Universidade de São Paulo. doi:10.11606/D.8.2011.tde-22102012-093617

Nogueira, Antônia Fernanda de Souza. 2019. *Predicação na língua Wayoro (Tupi): propriedades de finitude.* Ph.D. dissertation, Universidade de São Paulo. doi:10.11606/T.8.2019.tde-14082019-101006

Nogueira, Antônia Fernanda Souza, Ana Vilacy Galucio, Nicole Soares-Pinto, Adam Roth Singerman. 2019. Termos de parentesco nas línguas Tuparí (família Tupí). *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 33–64. doi:10.1590/1981.81222019000100004

Nogueira, Antônia Fernanda de Souza, comunidade Wajuru, Clenilson Miranda de Sousa. 2021. Apresentação da primeira versão do vocabulário Wayoro–português. In: Patrícia Goulart Tondineli (org.). *(Re)vitalizar línguas minorizadas e/ou ameaçadas: teorias, metodologias, pesquisas e experiências*: 83–108. Porto Velho: Coleção Pós-Graduação da UNIR–EDUFRO.

Nonato, Rafael, Jamtô Suyá, Kawiri Suyá. 2012. *Dicionário Kĩsêdjê–português.* Rio de Janeiro: Museu do Índio/Prodoclin. Available online at: http://prodoclin.museudoindio.gov.br/images/conteudo/kisedje/dicionario_kisedje1.pdf, accessed on 24/02/2023.

Oliveira, Christiane Cunha de. 2005. *The language of the Apinajé people of Central Brazil.* Ph.D. dissertation, University of Oregon.

Oliveira, Maria das Dores de. 2006. *Ofayé, a língua do povo do mel: fonologia e gramática.* Ph.D. dissertation, Universidade Federal de Alagoas.

Pache, Matthias. 2023. Evidence for a Chibcha–Jê connection. *International Journal of American Linguistics* 89(2): 219–253. doi:10.1086/723641

Palha, Luiz. 1942. *Ensaio de gramática e vocabulário da língua Karajá falada pelos índios remeiros do rio 'Araguaia'*. Rio de Janeiro: Olímpica.

Parapaino Castro, Pablino. 2008. *Isiukiché nikorokó bésɨro. Guía de escritura del idioma bésɨro.* Santa Cruz de la Sierra: Unión de Artesanos de la Tierra — UNIARTE.

Picanço, Gessiane Lobato. 2005. *Mundurukú: phonetics, phonology, synchrony, diachrony.* Ph.D. dissertation, University of British Columbia. doi:10.14288/1.0092991

Picanço, Gessiane Lobato. 2019. *A fonologia diacrônica do Proto-Mundurukú (Tupí).* Curitiba: Appris.

Pinheiro, Edilson, Maria de Nazaré Duarte, Gessiane Picanço. 2020. A ortografia Mundurukú em questão: uma proposta de adaptação com foco em não falantes. *Revista Brasileira de Línguas Indígenas* 3: 28–39. doi:10.18468/rbli.2020v3n2.p28-39

Pries, Stanley T. 2008. *Dicionário Gavião-Krikati.* Mimeo.

Ramirez, Henri, Valdir Vegini, Maria Cristina Victorino de França. 2015. Koropó, puri, kamakã e outras línguas do Leste brasileiro: revisão e proposta de nova classificação. *LIAMES* 15(2): 223–277. doi:10.20396/liames.v15i2.8642302

Ramirez, Henri, Valdir Vegini, Maria Cristina Victorino de França. 2017. O warázu do Guaporé (tupi–guarani): primeira descrição linguística. *LIAMES* 17(2): 411–506. doi:10.20396/liames.v17i0.8647468

Reiter, Sabine. 2011. *Ideophones in Awetí.* Ph.D. dissertation, Christian-Albrechts-Universität zu Kiel.

Ribeiro, Eduardo Rivail. 2002. O marcador de posse alienável em Karirí: um morfema Macro-Jê revisitado. *LIAMES* 2: 29–46. doi:10.20396/liames.v2i1.1403

Ribeiro, Eduardo Rivail. 2004a. Prefixos relacionais em Jê e Karajá: um estudo histórico-comparativo. *LIAMES* 4: 91–101. doi:10.20396/liames.v4i1.1427

Ribeiro, Eduardo Rivail. 2004b. *A língua Ofayé: documentando uma língua ameaçada.* Talk presented at IEL/Universidade Estadual de Campinas on 2/8/2004.

Ribeiro, Eduardo Rivail. 2009. Old data, new cognates: the case of the "marker of alienable possession" in the Kamakã, Purí, and Krenák families. *Cadernos de Etnolingüística. Série Notas* 2: 1–5.

Ribeiro, Eduardo Rivail. 2011. Prefixos relacionais como evidência histórico-comparativa: os casos Chiquitano e Jabutí. In: Ana Suelly Arruda Câmara Cabral, Aryon Dall'Igna Rodrigues, Jorge Domingues Lopes, Maria Risolêta Silva Julião (orgs.). *Línguas e Culturas Macro-Jê, vol. 2*: 105–120. Campinas: Curt Nimuendajú; Brasília: Laboratório de Línguas Indígenas.

Ribeiro, Eduardo Rivail. 2012a. *A grammar of Karajá.* Ph.D. dissertation, University of Chicago.

Ribeiro, Eduardo Rivail. 2012b. Final consonants in Maxakalí and their comparative status. *LIAMES* 12: 189–93. doi:10.20396/liames.v0i12.1489

Ribeiro, Eduardo Rivail, Hein van der Voort. 2010. Nimuendajú was right: The inclusion of the Jabutí language family in the Macro-Jê stock. *International Journal of American Linguistics* 76(4): 517–570. doi:10.1086/658056

Ribeiro, Maria de Jesus Pacheco. 2010. *Dicionário Sateré-Mawé/português.* MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).

Ribeiro, Michela Araújo. 2008. *Dicionário Djeoromitxi–português: registro da diversidade lingüística do povo Jabuti.* MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).

Ribeiro, Rosa Maria de Lima. 2008. *Dicionário Arikapu–português: registro de uma língua indígena amazônica.* MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).

Rocha, Ivan. 2011. *A estrutura argumental da língua Karitiana: desafios descritivos e teóricos.* MA thesis, Universidade de São Paulo. doi:10.11606/D.8.2011.tde-12092012-120027

Rocha, Ivan. 2022. Interpretação temporal em orações não finitas em Karitiana: a contribuição do aspecto. In: Dioney Moreira Gomes, María Alejandra Regúnaga, Arthur Britta Scandelari (orgs.). *Diversidade linguística na América: línguas ameríndias (v. 1)*: 229–263. Brasília: Editora UnB.

Rodrigues, Aryon Dall'Igna. 1985. Evidence for Tupi–Carib relationships. In: Harriet E. Manelis Klein, Louisa R. Stark (eds.). *South American Indian languages: retrospect and prospect*: 371–404. Austin: University of Texas Press.

Rodrigues, Aryon Dall'Igna. 1993. Uma hipótese sobre flexão de pessoa em Bororo. In: *Anais da 45ª Reunião Anual da Sociedade Brasileira para o Progresso da Ciência*: 505. Recife: Sociedade Brasileira para o Progresso da Ciências.

Rodrigues, Aryon Dall'Igna. 1999. Macro-Jê. In: R. M. W. Dixon, Alexandra Y. Aikhenvald (eds.). *The Amazonian Languages*: 165–206. Cambridge University Press.

Rodrigues, Aryon Dall'Igna. 2000. Ge–Pano–Carib x Jê–Tupí–Karíb: sobre relaciones prehistóricas en Sudamérica. In: Luis Miranda (org.). *Actas del I Congreso de Lenguas Indígenas de Sudamérica, tomo I*: 95–105. Lima: Universidad Ricardo Palma.

Rodrigues, Aryon Dall'Igna. 2005. As vogais orais do Proto-Tupí. In: Aryon Dall'Igna Rodrigues, Ana Suelly Arruda Câmara Cabral (orgs.). *Novos estudos sobre línguas indígenas*: 35–46. Brasília: Editora UnB.

Rodrigues, Aryon Dall'Igna. 2007. As consoantes do Proto-Tupí. In: Ana Suelly Arruda Câmara Cabral, Aryon Dall'Igna Rodrigues (orgs.). *Línguas e culturas Tupí, vol. I*: 167–203. Campinas: Curt Nimuendajú; Brasília: LALLI/UnB.

Rodrigues, Aryon Dall'Igna. 2009. A case of affinity among Tupí, Karíb and Macro-Jê. *Revista Brasileira de Linguística Antropológica* 1(1): 139–167. doi:10.26512/rbla.v1i1.12289

Rodrigues, Aryon Dall'Igna. 2012. Flexão relacional no tronco linguístico Macro-Jê. *Revista Brasileira de Linguística Antropológica* 4(2): 267–277. doi:10.26512/rbla.v4i2.20691

Rodrigues, Aryon Dall'Igna, Ana Suelly Arruda Câmara Cabral. 2010. Explorando a hipótese de Rodrigues sobre possíveis conexões genéticas Tupí e Macro-Jê. In: Rosane de Sá Amado (org.). *Estudos em línguas e culturas Macro-Jê*: 143–153. São Paulo: Paulistana.

Rodrigues, Aryon Dall'Igna, Ana Suelly Arruda Câmara Cabral, Beatriz Carretta Corrêa da Silva. 2006. Evidências lingüísticas para a reconstrução de um nominalizador de objeto **-*mi*- em Proto-Tupí. *Estudos da Língua(gem)* 4(2): 21–39. doi:10.22481/el.v4i1.1021

Rondon, Cândido M. S., João Barbosa de Faria. 1948. *Glossário Geral das tribos silvícolas de Mato-Grosso e outras da Amazônia e do Norte do Brasil, tomo I.* Rio de Janeiro: Imprensa Nacional. (Comissão Rondon, 76.)

Sabino, Wary Kamaiurá. 2016. *Awetýza tiʔíngatú: construindo uma gramática da língua Awetý, com contribuições para o conhecimento do seu desenvolvimento histórico.* Ph.D. dissertation, Universidade de Brasília.

Salanova, Andrés Pablo. 2011. A flexão de terceira pessoa nas línguas Jê. *LIAMES* 11: 75–114. doi:10.20396/liames.v0i11.1497

Santana, Áurea Cavalcante. 2012. *Línguas cruzadas, histórias que se mesclam: ações de documentação, valorização e fortalecimento da língua Chiquitano no Brasil.* Ph.D. dissertation, Universidade Federal de Goiás.

Santos, Ludoviko Carnasciali dos. 1997. *Descrição de aspectos morfossintáticos da língua Suyá (Kĩsêdjê), família Jê.* Ph.D. dissertation, Universidade Federal de Santa Catarina.

Schleicher, Charles O. 1998. *Comparative and internal reconstruction of the Tupi–Guarani language family.* Ph.D. dissertation, University of Wisconsin.

Sekelj, Tibor. 1948. *Wordlist Aruá-Makurap-Zaboti-Arikapú-Tupari*. Ms. Available online at: etnolinguistica.wdfiles.com/local--files/sekelj%3A1/Sekelj_Wordlist_Aru%C3%A1_Makurap_Zabot%C3%AD_Arikap%C3%B3_Tupar%C3%AD.pdf, accessed on 24/02/2023.

Seki, Lucy. 2000. *Gramática do Kamaiurá, línguas Tupi–Guarani do Alto Xingu.* Campinas: Editora da Unicamp; São Paulo: Imprensa Oficial.

Seki, Lucy. No date. *Vocabulário português–Botocudo.* Ms.

Silva, Mário André Coelho. 2015. *A coda consonantal em Maxakalí.* MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2015.945929

Silva, Mário André Coelho. 2020. *Tikmũũg yĩy ax tinã xohi xi xahĩnãg. Sons e pedaços da língua Maxakalí: descrição da fonologia e morfologia de uma língua Macro-Jê.* Ph.D. dissertation, Universidade Federal de Minas Gerais.

Silva, Mário André Coelho. Forthc. Propostas para a fonologia do Yaathê e do Pré-Yaathê. *Amerindia*.

Silva, Mário André Coelho, Andrey Nikulin. 2021. Uma aproximação à fonologia e morfologia do Malalí. *Cadernos de Etnolingüística* 9(1): e090107.

Silva, Raynice Geraldine Pereira da. 2010. *Estudo morfossintático da língua Sateré-Mawé.* Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2010.769939

Singerman, Adam Roth. 2016. Nasal harmony and phonotactic well-formedness in Tupari. *International Journal of American Linguistics* 82(4): 453–485. doi:10.1086/688603

Singerman, Adam Roth. 2018. *The morphosyntax of Tuparí, a Tupían language of the Brazilian Amazon.* Ph.D. dissertation, University of Chicago. doi:10.6082/ns2a-cv36

Snethlage, Emil Heinrich. 2015. *Die Guaporé-Expedition (1933–1935): in Forschungstagebuch. Aus dem Nachlass herausgegeben von Rotger M. Snethlage, Alhard-Mauritz Snethlage und Gleice Mere*. Wien /Köln / Weimar: Böhlau Verlag. doi:10.7788/9783412218591

Storto, Luciana. 2019. Kin terms in Karitiana and how they may contribute to the reconstruction of Proto-Tupian kin terminology. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 121–130. doi:10.1590/1981.81222019000100008

Tremaine, Sheila. 2007. *Dicionário Rikbaktsa–português, português–Rikbaktsa.* Cuiabá: Associação Internacional de Lingüística — SIL Brasil.

Tripp, Robert. 1995. *Diccionario amarakaeri–castellano.* Lima: Ministerio de Educación/Instituto Lingüístico de Verano. (Serie Lingüística Peruana, 34.)

Viegas Barros, J. Pedro. 2005. *Algunas semejanzas gramaticales Macro-Guaicurú/Macro-Jê.* Talk presented at IV Encontro de Pesquisadores de Línguas e Culturas Macro-Jê (November 3–5, 2005). Available online at www.etnolinguistica.org/artigo:viegas-barros-2005, accessed on 24/02/2023.

Voort, Hein van der. 2007. Proto-Jabutí: um primeiro passo na reconstrução da língua ancestral dos Arikapú e Djeoromitxí. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 2(2): 133–168. doi:10.1590/S1981-81222007000200007

Weiss, Helga E. 2005. *Dicionário Kayabi–português com um glossário português–Kayabi.* Brasília: Sociedade Internacional de Lingüística.

Wiesemann, Ursula Gojtéj. 2011. *Dicionário Kaingang–português, português–Kaingang.* Curitiba: Esperança.

Xerente, Armando Sõpré. 2019. *Particularidades dos sons, nomes, verbos, advérbios e posposições em Akwẽ (Xerente), família Jê Central, tronco Macro-Jê.* MA thesis, Universidade de Brasília.

*А. В. Никулин. Л*ексические свидетельства в пользу макро-же–тупийской гипотезы

Гипотеза о возможном дальнем родстве двух крупнейших языковых семей восточной Южной Америки — макро-же и тупийской — высказывается уже давно, в основном с опорой на морфологические схождения. В этой статье приводятся лексические свидетельства в пользу макро-же–тупийской гипотезы. При этом сравниваются именно праформы, восстановленные для пра-макро-же и пратупийского языков. Особое внимание уделено дистрибуции рефлексов рассматриваемых этимонов внутри каждой семьи, морфосинтаксическим свойствам сравниваемых форм, а также семантическому и фонологическому правдоподобию предлагаемых этимологий. Хотя количество возможных схождений не очень велико, между ними устанавливаются регулярные звуковые соответствия, что делает макро-же–тупийскую гипотезу привлекательной и достойной дальнейшего рассмотрения.

Ключевые слова: макро-же языки; тупийские языки; сравнительно-исторический метод; коренные языки Южной Америки.