Title: The Classification of South American Languages

Short title: Classification of S. American Languages

Author Name: Lev Michael

Author affiliation: University of California, Berkeley

ORCID: 0000-0002-9612-5996

email: levmichael@berkeley.edu

Corresponding author contact information:

1212 Dwinelle Hall #2350
University of California, Berkeley
Berkeley, CA 94720
levmichael@berkeley.edu

Abstract

With some 108 independent genealogical units, South America is the linguistically most diverse region of our planet, and presents a particular challenge to linguists seeking to understand the genealogical relationships among human languages. Recent years have seen a resurgence of interest in the internal classification of South American language families, and this article provides a critical overview of research in this very active area, focusing on the seven largest language families of the continent: Arawakan, Cariban, Jê, Panoan, Quechuan, Tukanoan, and Tupian. The strengths and weaknesses of major classification proposals are examined, and directions for future research discussed. A number of long-distance relationship proposals that South Americanists are actively debating, including Carib-Tupian, Pano-Tacanan, Quechumaran, TuKaJê, and Macro-Jê, are also examined.

Keywords: Language Classification, South American Languages, Historical Linguistics, Comparative Linguistics

1. Introduction

1.1 Overview

South America is the most linguistically diverse region of the world, exhibiting some 108[1] independent genealogical units, or one quarter of the world's linguistic diversity, which

[1] This number of course varies somewhat depending on the judgments of particular scholars, and advances in the historical linguistics of the continent.

encompasses some 420 living languages (Campbell 2012), and corresponds to some 574 languages that existed when Europeans arrived in the Americas (Seifart and Hammarstrom 2017). This diversity both inspires and challenges efforts to understand the genealogical relationships among these languages, both in terms of developing internal classifications of well-established genealogical units, and developing proposals for 'long-distance' relationships among languages whose genealogical relatedness is less certain.[2] The accurate classification of South American languages is of paramount importance not only for linguistics, where a better understanding of genealogical relationships among these languages underpins research into novel processes of language change and contact (Epps 2009), but also for fields such as anthropology, archaeology, and ethnohistory, that draw on genealogical classifications and the products of historical linguistic research for insights into the deep social and cultural history of the continent.

This article provides an overview of the state-of-the-art in the classification of South American languages, with the goal of equipping readers to engage critically with this literature. For reasons of space I focus on the seven largest language families,[3] and do not provide exhaustive classifications of any these families, referring readers instead to recent works that do so (e.g. Campbell 2012; for specific families, see Sections 3.1-3.7). I similarly do not address the history

[2] Use of the unmodified terms 'relationship' and 'related' is shorthand for 'genealogical relationship' and 'genealogical relatedness', respectively.

[3] Chibchan, a family distributed across Central America and northwestern South America, is not included, as the proposed Proto-Chibchan homeland and the majority of its members are found in Central America (Constenla-Umaña 2012)

of classification of the families I examine, which in some cases runs to several centuries in length, but instead focus on major modern works.[4]

I begin in Section 2 with a brief overview of the types of classifications found in the literature, the methodologies they employ, and an evaluation of their relative strengths and weaknesses, and then turn to a critical overview of extant classifications in Section 3. Of the 108 units identified by Campbell (2012), 53 are isolates, and 48 are families with 6 or fewer members, leaving seven larger families: Quechuan, Arawakan, Tupian, Cariban, Macro-Jê, Panoan, and Tukanoan. Sections 3.1-3.7 are devoted to these languages, in this order.

It is worth noting that some of the most interesting recent changes to our understanding of the classification of South American language families have involved careful application of the comparative method to determine that certain long-recognized language families actually need to be split into distinct language families, as the earlier erroneous classification having resulted from misleading similarities arising from language contact. This includes 'Makúan' (cited, e.g., in Campbell (2012)), which Epps and Bolaños (2017) argue needs to be split into at least distinct two, if not three, language families, and Witotoan, which Seifart and Echeverri (2015) argue needs to be split into a 'narrow' Witotoan family and Boran.

[4] Campbell (2012, pp. 63-67) presents a historical overview of large-scale classifications of South American languages from the earliest Jesuit classifications through the present, while Adelaar (2012a) presents a similarly broadly-gauged overview of research on the languages of the continent.

Works such as this cast an interesting light on recurrent efforts to attempt to reduce the apparent genealogical diversity of South America – perhaps inspired by expectations regarding diversity influenced by experience with more genealogically homogeneous regions of the world – through long-distance classification proposals of varying degrees of daring. The fact that increasingly careful historical work on South American languages has in some cases resulted in *increased*, rather than *decreased* diversity, should encourage caution and careful evaluation of intriguing long-distance proposals such as Adelaar's (2000) Katukinan-Harakmbut proposal, and Pache's (2016) Pumé-Chocoan proposal.

1.2 A note on language and language family names

Many South American languages and language families appear in the scholarly literature with more than one name. One source of variation is linguists taking greater care in recent decades to use language names preferred by the speakers of those languages, and to avoid pejorative names for languages and language families, as e.g. in the case of the *Naduhup* family (formerly *Makú*, from the Arawakan term meaning 'without language (lit. tongue)' (Epps and Bolaños 2017)), and the *Chicham* family (formerly *Jivaroan*, from Spanish *jíbaro*, a term with pejorative connotations (Deshoullière and Utitiaj Paati 2019)). Differences also arise due to different family naming conventions among Anglophone and Iberophone linguists, with the former group often using family names bearing the adjectival suffix *-an*, and the latter not (e.g. Panoan vs. Pano). This article follows the Anglophone practice, except in cases where it is typically

eschewed even by English-speaking linguists, e.g. '(Macro-)Jê', which is much more common than '(Macro-)Jêan'.

2 Types of classifications

There are four major types of linguistic classification, distinguished by the methods they employ, given here in order of their typical reliability: 1) classification based on the comparative method; 2) on computational phylogenetic methods; 3) expert classifications; and 4) those based on mass comparison.

The **comparative method** (CM) employs long-established methods to reconstruct linguistic features of proto-languages (Rankin 2003, a.o.). Such reconstructions allow linguists to identify shared innovations among daughter languages, which serve as the only reliable basis for identifying subgroups. Critical to CM is its reliance on: 1) systematic sound correspondences as evidence for cognacy, which distinguish cognates from borrowings or chance resemblances; and 2) the directionality of sound change to reconstruct proto-sounds, and thereby proto-forms. These features make CM the only reliable means for evaluating hypotheses of genealogical relatedness among languages (Campbell and Poser 2008, a.o.).

Computational phylogenetic methods (CPM) encompass a range of quantitative methods for classifying languages, and fall into two major classes: character-based methods and distance-based ones (see Dunn 2015; Michael and Chousou-Polydouri 2019; and Nichols and Warnow

2008, for overviews). Character-based methods produce classifications by evaluating how well different trees account for the distribution of a set homologous linguistic features (i.e., features shared by virtue of descent from a common ancestor; e.g., exhibiting members in a given cognate set) among a group of genealogically related languages, in light of some optimization criterion (e.g. minimizing the number of independent innovations). Rooting phylogenetic trees (i.e., identifying the ancestral node of the entire tree) results in classifications based on shared innovations, making character-based CPMs compatible with the basic insight of CM. Distance-based methods classify languages based on overall similarity metrics, conflating shared innovations, shared retentions, and parallel innovations, meaning that distance-based classifications cannot be interpreted as genealogical classifications. Most applications of CPMs in linguistics have been based on lexical data, but these methods can be applied to any set of homologous features. Note that CPMs can yield plausible classifications in cases where the CM cannot yet do so, due to the fact that the innovations identified by particular applications of CM are too common be reliably treated as shared innovations.

Expert classifications are based on a scholar's deep knowledge of a family, but are distinguished from the preceding methods by the lack of an explicit methodology and, typically, the lack of presentation of explicit evidence or argumentation (see examples in sections below). Expert classifications are difficult to evaluate when they conflict with each other, or when they conflict with classifications based on explicit methodologies.

Mass comparison is intended as a method for identifying distant genealogical relationships among languages (Greenberg 1987). The method consists of considering morphemes in different languages with similar forms and (loosely) similar meanings to be cognate, which serves as evidence for relatedness among the languages in question. Mass comparison has been extensively criticized for abandoning the comparative method's reliance on sound correspondences as evidence for cognacy, and Greenberg's proposed long-distance relationships among the languages of the Americas have won little support among Americanists (see Campbell and Poser 2008, pp. 266-279; Campbell 1997, pp. 210-213, and works cited therein), and will not be discussed further here.

3 Classification of major families

3.1 Quechuan and Quechumaran

3.1.1 Quechuan

At their greatest extent, Quechuan languages were spoken through much of the habitable Andes, and extended into both the Pacific coastal and Amazonian lowlands (Adelaar 2004, pp. 168-170). With 8.5-10 million speakers (ibid., p. 168), the languages of this family are among the largest indigenous languages of South America.

The most influential classifications of the Quechuan varieties stem from the work of Torero (1964, 1970, 1983) and Parker (1963, 1969a-d, 1971), who reconstructed Proto-Quechuan (PQ) phonology and many aspects of PQ morphology, allowing them to propose internal

classifications based on a combination of phonological and morphological innovations (Cerron-Palomino 1987). The classification of Quechuan languages has been complicated by contact among different Quechuan varieties, and with Aymaran languages.

Torero and Parker distinguished two major branches: Quechua I, which occupies the Central Peruvian Andes, and Quechua II, which brackets Quechua I, extending north to Ecuador and Colombia, and South to Bolivia, Chile, and Argentina.[5] Even at this early stage of investigation challenges arose for a purely arboreal classification (Cerrón Palomino 1987, pp. 223-224; Parker 1969; 1971, pp. 45-47, Torero 1968, p. 291), since many innovations useful for defining subgroups form complexly partially-overlapping isoglosses across distinct varieties and dialect continua within each branch. The consensus that has emerged is that of the three putative subgroups of QII (i.e. QIIA, B, and C), QIIA is not defined by shared innovations that group it decisively with either QII or QI (Adelaar 2004, p. 186). Similar observations are made regarding Pacaraos Quechua, leading to the suggestion that these varieties might form branches coordinate with QI and QII (ibid.). Some scholars have even argued that the distinction between QI and QII is untenable (e.g. Heggarty 2005; Landerman 1991; Mannheim 1991; Pearce and Heggarty 2011), although Quechuanists have generally found the argumentation of these critiques unconvincing (Adelaar 2013, pp. 58-62).

Despite the increasing availability of descriptive material on Quechuan languages, there has been relatively little CM-based work since the pioneering efforts of Torero and Parker.

[5] Parker (1963 a.o.) referred to these two branches as Quechua A and B, respectively.

However, Floyd's (to appear) recent clarification that Ecuadorean Quechuan varieties are descended from Cuzco Quechua (or very similar varieties), but heavily influenced by contact with Barbacoan languages, indicates the promise of such work. Clarifying the status of the varieties that are not clearly members of QI or QII are an obvious priority, and since the innovations used to argue for some major subgroups do not form entirely consistent isoglosses, character-based phylogenetic analyses may prove useful, since they permit objective assessments of competing subgrouping evidence.

3.1.2 Quechumaran

Quechuan and Aymaran languages have enjoyed a long history of mutual influence, leading to significant convergences in their phonology, morphosyntax, and lexicon (Adelaar 2012b). Similarities among these languages has inspired the hypothesis that Quechuan and Aymaran families are related as subgroups of a larger Quechumaran family, while the effects of contact have complicated its evaluation. The Aymaran family consists of Aymara, with over 1.5 million speakers in Bolivia, Chile and Peru, and two small languages found in Southern Peru, Jaqaru and Kawki.

A former consensus that Quechuan and Aymaran are related was significantly undermined by Torero's and Parker's reconstruction of Proto-Quechuan as lacking the three-way contrast between ejective, aspirated, and plain stops found in Aymaran languages and the neighboring major Quechuan languages, especially Bolivian and Cuzco Quechua (Cerrón-Palomino 1987, pp.

351–374). In defense of the Quechumaran hypothesis, Orr and Longacre (1968) reconstructed precisely this three-way contrast for Proto-Quechumaran, but this conclusion met with significant skepticism (Cerrón-Palomino 1987, Campbell 1995). Critics observed that the Quechuan data was mostly drawn from QII varieties adjacent to Aymaran languages, and thus biased towards the small number of Quechuan languages exhibiting the three-way contrast in question, and those most likely to have been influenced by contact with Aymaran. Furthermore, no effort was made to exclude the vast number of loanwords between Aymaran and Quechuan languages.

As Campbell (1995; 1997, pp. 273-283) observes, however, the justified criticisms of Orr and Longacre's reconstruction, and the evidence of abundant grammatical and lexical borrowing between Quechuan and Aymaran languages, do not rule out a genealogical relationship between the two families: rather, they indicate the importance of carefully selecting the languages to compare, and weeding out borrowing in the comparative dataset (Adelaar 1986). Emlen and Adelaar (2017) makes significant progress on the latter front, additionally providing reconstructions of forms that can be safely traced by to Proto-Quechuan or Proto-Aymaran, while Emlen and Dellert (to appear) clarifies how Proto-Quechuan roots were modeled on Aymaran forms. Given the importance of the Quechumaran hypothesis for understanding the history of the Andean region, we can hope that renewed interest in Aymaran-Quechuan language contact (e.g., Emlen 2017) will serve as a rigorous basis for its evaluation (Cerrón-Palomino 1987, pp. 374-375).

3.2 Arawakan, Macro-Arawakan, and Arawakan Linguistic Matrix Hypothesis

3.2.1 Arawakan

Arawakan is the largest language family of the Americas, with its ~80 historically-attested members being dispersed across a vast region extending from the Caribbean to the Argentinean pampas, and from the Atlantic coast of Brazil to the eastern Andes (Aikhenvald 1999a; Ramirez 2001). Taino, the Arawakan language of the Greater Antilles and Bahamas, was the first indigenous language of the Americas encountered by Europeans, and is the source for many Spanish flora, fauna, and cultural terms specific to the Americas (e.g., *maíz* `maize, corn', *iguana*, and *cacique* 'indigenous chief').

Consensus on the delimitation of the Arawakan family and its low-level subgroups was reached by the 1990s, as summarized in Payne's (1991) comparative study, which reconstructs the segmental inventory of Proto-Arawakan (PA) and 203 PA etyma. This delimitation included the recognition that two small language families, Harakmbut and Arawan, are not Arawakan, although they exhibit lexical borrowings from their Arawakan neighbors (for Arawakan-Arawan borrowings, see Facundes and Brandão (2011)), and the recognition that two languages, Yanesha' and Resígaro, are Arawakan, despite their affiliation being obscured by contact with their Andean and Boran/Witotoan neighbors, respectively (for Yanesha', see Adelaar 2007; Wise 1976; for Resígaro, see Payne 1985).[6]

[6] Payne (2005) also makes a compelling case the now-extinct Apolista is Arawakan.

The most influential classifications of Arawakan have been expert ones (Kaufman 1994, Aikhenvald 1999a), which have overall exhibited much similarity. In general, the lower-level subgroups of the family are obvious, but higher level relationships among them are unclear, which is reflected in expert classifications positing relatively flat classifications. One noteworthy exception to this general caution is their positing a top-level split between Northern and Southern branches (with some variation in the structure and names of the branches).

A number of classifications using CPMs have been developed for the family, with Payne's (1981) and Ramirez' (2001) being the empirically best-grounded distance-based ones, and Walker and Ribeiro (2012) being the sole character-based one. As already mentioned, distance-based CPMs exhibit inherent weaknesses, but Payne and Ramirez furthermore do not employ principled clustering methods, leading to some arbitrariness in subgrouping decisions. Walker and Ribeiro's (2012) Bayesian phylogenetic analysis of root-meaning sets is methodologically sounder, but its reliance on a short concept list (100-word Swadesh list for 60 Arawakan languages) means that most higher-level nodes exhibit low posterior probabilities, such that their classification are effectively rake-like near the root. These works generally recover the low-level subgroups posited in expert classifications, but none find evidence for a top-level Northern-Southern split, although Walker and Ribeiro (2012) intriguingly find the Palikur-Marawan subgroup emerging as sister to the rest of the family. Their analysis also yields some intriguing mid-level subgroups, such as the Waraiku-Marawa subgroup forming a subgroup with the well-established Caribbean subgroup, and finding reasonably strong support for a large subgroup consisting of most northeastern Arawakan languages.

An obvious priority is improving lexical phylogenetic analyses of the family by increasing the size of concept lists and rooting techniques employed. Stark (2018) showed that increasing the size of the comparative concept list from 100, used by Walker and Ribeiro (2012), to 726 dramatically increased the posterior probabilities in the analysis of the Caribbean subgroup, and similar improvements can be expected elsewhere. Improvements can also be made in rooting the tree. Since no out-group exists for Arawakan, Walker and Ribeiro (2012) rooted their tree with a clock, but both their choice of evolutionary model (a Yule prior) and the lack of either internal calibration points or tip date points raise questions about the accuracy of the rooting (see Michael and Chousou-Polydouri (2019) for discussion). Addressing these issues would contribute to a more trustworthy tree root and classification.

3.2.2 The Arawakan Linguistic Matrix Hypothesis

It has been observed that many peoples speaking Arawakan languages share certain cultural practices – an 'Arawakan matrix' (Santos-Granero 2002) – including an avoidance of endo-warfare, and a tendency to form regional trade and political networks. The 'Arawakan linguistic matrix hypothesis' (ALMH)[7] is a linguistic hypothesis inspired by these observations, and posits that Proto-Arawakan (PA) was a trade language or lingua franca spoken in much of lowland South America, whose diversification was inhibited over this large area until relatively recently

[7] I adopt this term to distinguish the theory relating specifically to PA and the diversification of its daughter languages from more general observations about a shared 'Arawakan ethos'.

by a continent-spanning Arawakan-dominated trade network (Hornborg 2005; Eriksen 2011). The ALMH is articulated most clearly by Eriksen and Danielsen (2014) (see also Danielsen, Dunn, and Muysken (2011)), who propose that Proto-Arawakan only began to diversify only after the break-up of the trade network, at some point after 600AD.[8]

The ALMH thus posits that the time depth of Arawakan is similar to that of Romance, which is highly implausible, given that the internal lexical and grammatical diversity of the family is comparable to language families like Tupian, which is commonly assumed to have a time depth of 5000 years (Rodrigues and Cabral 2012). In this light, it is worth noting the linguistic evidence presented to support the ALMH is quite weak, namely, that the NeighborNet network for a set of grammatical typological features is star-like in shape. Beyond the general weakness of distance based methods for developing internal classifications, there is little reason to suppose that the best explanation for the lack of clear higher-level subgrouping structure in the NeighborNet is the language contact scenario posited by the ALMH, rather than, say, that the chosen features are simply insufficiently informative for higher-level classification purposes.

[8] Proponents of the ALMH do provide a precise date for the break-up of PA, but Eriksen (2011:272) suggests that the 'Arawak regional exchange system' was at its height between 200-600 CE, while Eriksen and Danielsen (2014:170) claim that the 'Arawakan matrix' reached its maximum extent by 800 CE, and that the 'Arawak cultural matrix' was dominant in South America until 1000 CE (ibid.:172).

Note that although evidence adduced for the ALMH may not be compelling, this does not entail that theories regarding the Arawakan ethos are flawed, but simply that the Arawakan ethos was not underpinned by linguistic uniformity.

3.2.3 Macro-Arawakan

During the 20th century a number of individual languages and small language families were considered to be related to what we now consider the Arawakan family. These include Arawan, Guahiban, Timotean, Tiniguan, and Harakmbut languages, as well as the Andean language Puquina, and the Amazonian isolate Yurakaré (see, e.g., Payne 1991, p. 365; Kaufman 1990, p. 58). Following Kaufman (1990), I refer to this variable penumbra of languages, together with Arawakan, as 'Macro-Arawakan'.

None of the above proposals have significant support among Arawakanists now, and they appear to have arisen from suggestive but misleading morphological similarities, or by lexical borrowings that were not identified as such. De la Grasserie (1894, p. 10), for example, declared Puquina to be "incontestably connected to the great Maypure [i.e., Arawakan] family" principally on the basis of the similarity of its 1st and 2nd person nominal and verbal person markers: *no-* and *po-*, respectively, to *n-* and *p*-initial 1st and 2nd person markers in most Arawakan languages. We now know that the *n-/p-* 1st/2nd person pattern has a circum-Pacific distribution (Nichols and Peterson 1996; Zamponi 2017), and cannot be considered compelling evidence for an Arawakan affiliation.

The delimitation of Macro-Arawakan has also generated some confusion and controversy in the literature regarding the terms ‘Arawak’ and ‘Arawakan’. The family I here call ‘Arawakan’ was first known as ‘Maipure’ or ‘Maipur(e)an’, with the names ‘Arawak’ and ‘Arawakan’ additionally coming into use as synonyms toward the beginning of the 20th century (Brinton 1891; Chamberlain 1907). However, when the membership of the family was clarified in the 1980s, Payne (1991, p. 363) suggested that set of indisputably related languages be called ‘Arawak’ (here: ‘Arawakan’), and the larger set of more dubiously related languages be called ‘Arawakan’ (here: ‘Macro-Arawakan). Payne’s proposal has been honored more in the breach, however, with the undisputedly related set of languages being variably called ‘Arawak’, ‘Arawakan’, and ‘Maipur(e)an’ in the literature. Although there has been some vigorous defense of Payne's original terminological proposal (e.g., Aikhenvald 2012, p. 23), the actual onomastic practice of linguists is sufficiently inconsistent that it seems clearest to use the term ‘Macro-Arawak(an)’ when referring to the larger, more speculative grouping, and countenance the above-mentioned terms as synonyms when referring to the smaller set of indisputably related languages.

3.3. Tupian, Tupi-Cariban, and TuKaJê

3.3.1 Tupian

The Tupian family is second in size only to Arawakan in South America, with its ~70 member languages located in Brazil, and in adjacent areas of Argentina, Paraguay, Bolivia, Peru, and

French Guyana (see Cabral and Rodrigues 2012 for an overview). Significantly, ~45 of these languages belong to a single subgroup, Tupi-Guarani (TG), which includes Paraguayan Guarani, one of the national languages of Paraguay, with almost five million speakers, and the now-extinct Tupinambá, the language spoken along much of the Brazilian Atlantic coast when Europeans arrived, and an important source for words denoting neo-tropical flora, fauna, and foodstuffs in European languages.

Although in recent years phylogenetic methods have been applied to both Tupian and Tupi-Guarani, expert classifications of both groups have been very influential. And while the phonological inventories and parts of the lexicon of both groups have been reconstructed (Tupian: Cabral and Rodrigues 2012; Rodrigues 2005, pp., 35-46; Rodrigues 2007, p. 11; TG: not presented explicitly, but implicit in Rodrigues (1984/5) and Cabral and Rodrigues (2002)), they have not played a significant role in developing classifications, since the sound changes involved are generally too common to serve as a basis for subgrouping. The Tuparí subgroup is among the best-studied non-TG Tupian subgroups, with Moore and Galucio (1994) reconstructing Proto-Tuparí phonology and lexical items, and Galucio and Nogueira (2011) extending this reconstruction, and arguing on the basis of morphological innovations that Makurap was the first language in the subgroup to branch off.

The most conservative expert classification of the Tupian family is due to the Tupi Comparative Project, based at the Museu Paraense Emilio Goeldi, which posits a rake-like organization of seven subgroups at the highest level of the family, including the Mawetí-Tupí-Guaraní

subgroup, with TG successively embedded under Awetí, and then Mawé (Galucio et al. 2015). Cabral and Rodrigues (2012), in contrast, propose that Tupian is comprised of two symmetric branches, an Eastern one and a Western one, each consisting of five successively nested subgroups. The Western one contains TG as its most deeply nested subgroup, followed by successively larger subgroups including additional subgroups in this order: Awetí, Mawé, Mundurukuic, and Jurunan. Significantly, Galucio et al.'s (2015) distance-based phylogenetic analysis of Tupian lends some support for Cabral and Rodrigues' proposed structure for the Western clade, although the support for the Jurunan subgroup forming a clade with the remaining languages is somewhat weak. However, Galucio et al. (2015) find no evidence for a distinct Eastern subgroup, and support for nodes above that corresponding to Rodrigues and Cabral's (2012) Western branch is in general rather weak, meaning that much uncertainty remains about the higher-level organization of the family. Given this uncertainty, an obvious medium-term priority for South American historical linguistics is to develop a robust character-based phylogenetic classification of the family.

The languages of the TG subgroup have been an important focus of historical and descriptive work on South American languages, with Rodrigues' (1984/5) rake-like expert classification of the TG family into eight *subconjuntos* ('subsets', rather than *subgrupos* 'subgroups') having been particularly influential. Rodrigues' deliberately cautious choice of *subconjunto* over *subgrupo* reflects the often underappreciated fact that while Rodrigues identifies certain sound changes associated with each subset, he is explicit that this classification is not a genealogical one (ibid., p. 48), but one based on perceived similarity (i.e., it is an expert classification).

Significantly, the sound changes in question are in general neither probative for subgrouping (since many are extremely common ones), exclusive to particular subsets, nor shared by all members of a subset.

Slight modifications of the eight *subconjuntos* have been proposed by Mello (2000; 2002) and Rodrigues and Cabral (2002), where the latter classification also proposes additional higher-level structure, grouping the eight *subconjuntos* into three major branches. Some arguments are provided for this higher-level structure on the basis of sound change and morphology, but in general the evidence is weak, consisting either of shared retentions, or of extremely common sound changes (O'Hagan, Chousou-Polydouri, and Michael 2019). Lemle (1971) represents an earlier attempt to develop a TG internal classification using reconstructions based on data from 10 members of the family, but is likewise hampered by the scarcity of informative sound changes. Similarly, neither Schleicher's (1998) TG phonological reconstruction, nor Jensen's (1998) reconstruction of TG morphosyntax clarify the internal structure of the subgroup.

The most detailed classification of TG is Michael et al.'s (2015) lexical phylogenetic classification of the family, which yields considerable well-supported high-level structure for the family. This classification recovers five of Rodrigues' (1984/5) eight *subconjuntos* as proper subgroups, and one as a paraphyletic group,[9] with only two of the *subconjuntos* not being confirmed in some way. They identify Kamayurá as sister to the rest of the family, which they dub 'Nuclear TG',

[9] Paraphyletic groups are those who consist of only a subset of the daughters of an ancestral language.

which itself splits into three major branches, two smaller northern Amazonian ones, and a third that embeds all the southern TG languages as a monophyletic subgroup.

3.3.2 TuKaJê and Tupi-Cariban

Rodrigues has advanced two long-distance genealogical relationship hypotheses involving the Tupian family: one positing that the Tupian, Cariban, and Jê languages are related (Rodrigues 2009), and the second that only the Tupian and Cariban languages are related (Rodrigues 1985); the TuKaJê and Tupí-Cariban hypotheses, respectively.

The TuKaJê hypothesis rests principally on the observation that languages of all three families exhibit 'relation prefixes', a set of putative prefixes analyzed as indicating whether or not a head is preceded by its associated complement, e.g. a possessum by its possessor in a possessive NP, or a transitive verb by its object. Recent work, however calls into doubt the validity of the relational prefix analysis in TG languages (Meira and Drude 2013), where they have been best studied, and these arguments are equally troubling for the relational prefix analysis in other putative TuKaJê languages. In brief, Meira and Drude (ibid.) observe that so-called relational prefixes are an artifact of treating morphophonological root-initial segment alternations as if they were morphological alternations, which involves segmenting the alternating segment off from the rest of the root, and attributing meanings of 'contiguity' (when preceded by a complement) and 'non-contiguity' (when not) to the two alternants. Crucially, they show that comparative data from Mawé and Awetí, which together form the

Mawetí-TG (MATG) clade, reveals that the supposed 'relational prefixes' are simply the initial segments of Proto-MATG roots, where the segmental alternations are conditioned by presence or absence of preceding heads. As Meira and Drude observe (ibid., p. 10), this behavior resembles patterns of initial consonant mutation in Celtic languages, *raddoppiamiento sintatico* in certain Italian dialects, and *sandhi* in Sanskrit. From this vantage point, it appears that what Tupian, Cariban, and Jê languages share is not a set of 'relational prefixes', but rather a morphophonological process, namely root-initial segmental alternations conditioned by the morphosyntactic environment. While suggestive, this is relatively weak evidence for a genealogical relationship among the three families.

More recently, Nikulin (2015) has argued that lexical evidence supports the TuKaJê hypothesis, comparing reconstructed forms for Proto-Tupian (PT), Proto-Cariban (PC), and Proto-Macro-Jê (PMJ). The validity of the reconstructed forms is somewhat unclear, however, as the phonological and lexical reconstructions are presented somewhat telegraphically, and many of the putative TuKaJê cognate sets are not entirely convincing, since they do not yield systematic sound correspondences between PT, PC, and PMJ. Despite these difficulties, the basic strategy Nikulin adopts is precisely the one to pursue in decisively evaluating the TuKaJê hypothesis in the future: reliable reconstructions of PT, PC, and PMJ forms, and reconstruction of TuKaJê proto-forms on the basis of regular sound correspondences among cognate sets built from these reconstructed PT, PC, and PMJ forms.

The evidence for Tupí-Cariban is somewhat more promising. Rodrigues (1985) presents 121 possible cognate sets and extracts sounds correspondences from them, with Rodrigues making an effort to identify loans between TG languages and Nheengatú on the one hand, and northern Amazonian Cariban languages on the other hand, as well as loans from Wayana (Cariban) into the neighboring Wayampi (TG). Unfortunately, Rodrigues does not compare reconstructed PT forms with reconstructed PC ones, but instead compares forms from three Tupian languages (Mundurukú, Tuparí, and Tupinambá), with those from eight Cariban languages (plus Island Carib, an Arawakan language with a large number of Cariban loanwords). This procedure leaves the door wide open for undetected loanwords and especially *Wanderwörter* (Epps 2014), a doubt that is magnified by Rodrigues' suggestion of similarities with forms in certain Jê languages. In addition, Rodrigues does not use his correspondence sets to reconstruct Proto-Tupi-Cariban phonology and etyma, leaving doubt as to whether or not the correspondence sets yield coherent reconstructions.

Given the importance that establishing a genealogical link between Tupian and Cariban would have for our understanding of lowland South American linguistic and social history, an evaluation of the Tupi-Cariban hypothesis that compares reconstructed PT and PC is an obvious priority

3.4 Cariban

Approximately 50 Cariban languages were spoken when Europeans first arrived in the Americas, with 25 languages spoken by 60,000-100,000 individuals today (Gildea 2012, p. 441). These mainly lie in a broad band centered on the border region between Brazil and the Guyanas and Venezuela, and extending to the Caribbean littoral, with the exception a string of languages distributed southwards towards central Brazil, mainly in or near the Xingu River basin, and Carijona, which is spoken in central Colombia.

Cariban classification has been marked by sharply differing proposals even at the level of relatively low-level subgroups (Gildea 2012), with historical phonology having played an important role in overturning previous expert classifications. The Proto-Cariban (PC) phonological inventory was first reconstructed by Girard (1971), who proposed 14 low-level subgroups on the basis of shared phonological innovations, most of which have proved robust. This reconstruction was revised by Meira and Franchetto (2005), who, observing that much of Girard's (1971) data was problematic, developed comparative lists with modern data for eight languages. More recent advances include the reconstruction of PC /ô/ (a central or back unrounded mid vowel) and its subsequent evolution (Gildea, Hoff, and Meira 2010), and the reconstruction of a stem-initial ablaut process conditioned by a third person marker and a putative 'relational prefix' (Meira, Gildea, and Hoff 2010; see Section 3.3.2).

The two most recent classifications of Cariban are Gildea's (2012) state-of-the-art classification and Meira, Birchall, and Chousou-Polydouri's (2015) phylogenetic classification, which are useful to juxtapose. The former is a "somewhat speculative" expert classification that is

significantly informed by results stemming from the comparative method, when available, and reflects the current consensus among Caribanists. It is rake-like near the root, positing five major branches (in decreasing order of confidence): Parukatoan, Pekodian, Venezuelan, Nahukwa, and Guianan, plus a 'Residue' group. The phylogenetic classification presents a Bayesian analysis of lexical data from 34 Cariban varieties, based on 100 word Swadesh lists, and is likewise effectively rake-like near the root, exhibiting relatively low posterior probabilities for most higher level nodes, quite possibly due to the shortness of the comparative concept lists used (Michael and Chousou-Polydouri 2019). Only Parukotoan and Pekodian, the two highest-confidence groups in the expert classification, are recovered by the phylogenetic analysis, with Venezuelan, Nahukwa, and Guianan failing to emerge as monophyletic. These phylogenetic results correlate with the support that each subgroup in the state-of-the-art classification receives from CM results.

Parukatoan, for example, is defined by a clear sound change: PC **jô > Proto-Parukatoan *jo (Meira, Hoff, and Gildea 2010). Similarly, Pekodian was first proposed by Meira and Franchetto's (2005), who evaluated the validity of a Southern branch posited in expert classifications of the family (Derbyshire 1999, Durbin 1977, and Kaufman 1994). They found multiple shared phonological innovations that split the putative Southern languages in two groups, Pekodian (Arara, Ikpeng, and Bakairi) and Nahukwa (Kuikuro and several co-dialects), but no innovations that define Southern as a whole.

Venezuelan, as proposed by Gildea (2003), is defined by the unconditioned split of PC **o > *o,

ə, and one morphological innovation. Unconditioned splits, however, are generally viewed skeptically, except where they are seen as the outcome of incomplete lexical diffusion (see, e.g., Guy 2008), casting doubt on this proposed shared phonological innovation.[10] The proposed morphological innovation consists of the replacement of the PC third person possessive marker ***y-* by **it-* for vowel-initial nouns in Proto-Venezuelan, but as Gildea observes, the reconstructed PC form itself is not uncontroversial, raising the question of which form is in fact innovative. Given the preceding observations, and that Venezuelan does not emerge as a subgroup in Meira, Birchall, and Chousou-Polydouri's (2015) analysis, Gildea's (2012) characterization of Venezuelan as an "untested hypothesis" is suitably cautious.

Guianan, consisting of Taranoan (reconstructed by Meira 2000), Wayana, and Kari'nja, the most widely spoken Cariban language, is identified on the basis of a number of shared features (Meira 2005). The fact that the Meira, Birchall, and Chousou-Polydouri's analysis does not return this group, however, raises the question of how many of these are in fact shared innovations.

Caribanists have made significant strides in phonological and morphological reconstruction in comparison to colleagues working with many other South American families (see Gildea 2012 for overview), providing a valuable foundation for future work. Since it appears that sound changes may not be capable of providing a fine-grained classification of the family,

[10] Doubts about lexical diffusion as a significant pathway for sound change have mounted in recent years, however (see, e.g., Labov (2020)).

morphological reconstruction and further phylogenetic work with larger comparative lists are promising avenues for improving the classification of the family.

3.5 Jê and Macro-Jê

3.5.1 Jê

Jê is a family of ~17 historically-attested languages spoken near the eastern edge of the Amazon basin and adjacent areas outside the Amazon basin proper. Many of the languages spoken closer to the Atlantic coast became extinct during the colonial period and are only fragmentarily attested.

Empirically-grounded classifications of Jê languages date to Davis' (1966) reconstruction of Proto-Jê, which although based on a modest quantity of lexical data from only five languages, supported the previously uncertain inclusion of Kaingang in the family. The latter language is one of two living members of the Southern branch in Rodrigues' (1986, 1999) expert classification of the family into Northern, Central, and Southern branches. Davis did not seek to classify the Jê languages, and both his cognacy judgments and analytical decisions were subsequently criticized (Nikulin 2016; Ribeiro and van der Voort 2010), but it has only been relatively recently that scholars have resumed investigations into historical Jê phonology. Jolkesky (2010) provides a phonological and lexical reconstruction of Proto-Southern Jê. Nikulin (2016) provides a reconstruction of Proto-Northern-Jê phonology and Nikulin (2017) provides a

reconstruction of Proto-Cerrado (a proposed subgroup consisting of Northern and Central Jê) phonology. Given the importance that PJ reconstructions have for evaluating the membership of Macro-Jê (see next section), these recent advances are particularly important. It will be critical for transparency and verifiability of these reconstructions for future works to explicitly and systematically present cognate and correspondence sets, and provide explicit argumentation for segmental and lexical reconstructions.

Jolkesky (2010) and Nikulin (2017) provide distance-based classifications of Southern Jê and Northern Jê, respectively. Re-analysis of the extant cognate sets used in these analyses using character-based methods is obvious low-hanging fruit that could yield important insights into the internal classification of Jê.

3.5.2 Macro-Jê

Macro-Jê (MJ) is a linguistic family of ~30 historically-attested languages, which has the Jê family as its largest subgroup (Ribeiro 2006; Rodrigues 1999). The precise membership of MJ is the active focus of current research, with certain subgroups having been decisively shown to be members of the family only relatively recently. Note that there is some ambiguity in the use of the term 'Macro-Jê', as it can be employed in different contexts for the smaller core set of languages that a given author posits to be related, as well as include the larger penumbra of possibly related languages; it is typically clear from context which sense is intended.

Evidence that ranges between compelling and suggestive has been presented for six first-order subgroups in Macro-Jê, in addition to Jê proper: Aimorean, Maxakalian, Jabutian, Karajá, Ofayé, and Rikbatsa. Davis (1968) provides comparative evidence for the inclusion of Maxakalí and Karajá by presenting forms from these languages that are cognate with PJ forms, and identifying regular sound correspondences between Maxakalí, Karajá, and PJ. Ribeiro (2012a) strengthens the case for the inclusion of Karajá by providing additional phonological and morphological correspondences with Jê languages. Maxakalí is the sole surviving member of a larger group of modestly historically-attested related languages, as discussed by Ramirez, Vegini, and de França (2015), who present additional lexical evidence for its MJ affiliation (see also Nikulin and da Silva (2020) and Ribeiro (2012b)).

Evidence for the inclusion of Aimorean was initially due to a brief work by Seki (2002) that compares forms in Krenak, the best-attested of the Aimorean languages, with Davis' (1966) PJ forms, identifying a number of generally modestly attested correspondences, but providing neither phonological or lexical reconstructions. Recent work by Nikulin and da Silva (2020) significantly strengthens the case for the MJ affiliation of Aimorean and furthermore provides evidence that Maxakalian and Aimorean languages form a subgroup within MJ by reconstructing proto-Maxakalí-Krenak phonology and lexical items and proposing a number of shared phonological, lexical, and grammatical innovations for the subgroup.

The inclusion of Rikbatsá in MJ is based on comparative work by Boswood (1973), who identifies regular sound correspondences with Davis' reconstruction of PJ, while the inclusion of

Ofayé is based on a brief work reconstructing ‘Ofaie-Jê’ on the basis of sound correspondences between Ofayé, and again Davis' reconstruction. Ribeiro and van der Voort (2008, p. 29-30) express methodological concerns about both works, suggesting that it would be timely to revisit this early comparative work.

Finally Ribeiro and van der Voort (2010) employ CM to decisively demonstrate that the small Jabutian language family, consisting of Arikapú and Djeorimitxi, both spoken near the Brazilian-Bolivian border, is Macro-Jê, despite it not figuring in a number of expert classifications as MJ, including Rodrigues (1999).

Beyond these subgroups, for which some compelling comparative evidence has been presented, Rodrigues' (1999) expert classifications of MJ additionally lists Bororoan, Guató, Kamakãn, Karirian, Purian, and Yatê, as possible MJ subgroups under a broader conception of MJ that he characterizes as a "working hypothesis". In general, as the relationship of these linguistic groupings to MJ have come under greater scrutiny, the less credible an MJ affiliation has come to seem.

Ribeiro (2012a, p. 263) is pessimistic about the MJ status of Guató, observing that "the evidence presented by Greenberg (1987) and Rodrigues (1986; 1999) was rather superficial and inconclusive, a recent study ... (Martins 2011), purporting to present additional evidence based on a perusal of the entire corpus available of Guató, fails to provide any convincing new arguments."

Ramirez, Vegini, and de França (2015) examine several poorly-attested languages of eastern Brazil, including the Kamakã, Puri-Coroado, and Koropó, and seek to clarify their relationships with each other and languages of the Maxakalí and Aimorean subgroups of MJ. Turning first to the Purian family, they argue that there is little evidence for distinguishing Purí and Coroado as distinct languages, and then turning to a comparative list of 90 Puri-Coroadan items, find an insufficient number of possible cognates with either Krenak or Maxakali to permit them to identify regular sound correspondences between Puri-Coroado and a more reliably MJ language. The third language treated as Purian by Rodrigues (1999), Koropó, they argue on the basis of significant lexical data, is not related to Puri-Coroado, but is rather a Maxakalian language with a large number of Puri-Coroado loans. Ramirez, Vegini, and de França (2015) thus cast significant doubt on Purian as an MJ language. Interestingly, they find evidence for a close relationship between Kamakãn varieties and Maxakalí, leading them to propose that they form a subgroup of Maxalian, and are thus members of MJ, but not as an independent subgroup.

One language that is not included in Rodrigues' (1999) expert classification, but which a number of MJ specialists have come in more recent years to consider a plausible candidate for inclusion in the MJ family is Besɨro (a.k.a. Chiquitano), due to suggestive morphological and lexical similarities to Jê languages (Adelaar 2008; Santana 2006). Ribeiro (2011) also argues that relational prefixes in Chiquitano and several reliably MJ languages appear to have a common source, and that they appear on roots that are "very probably cognate" (p.117). Regardless of the ultimate validity of relational prefix analyses more generally (see Section 3.3.2), Ribeiro's

optimism regarding cognacy is a call to action to evaluate this hypothesis rigorously employing the comparative method.

With the exception of Nikulin's and da Silva's (2020) proposal that Maxalian and Aimorean form a subgroup, and Ramirez, Vegini, and de França's (2015) suggestion that the Kamakã varieties belong to the Maxalian subgroup, MJ is generally treated as having a rake-like structure below the root.[11] As is clear from this overview, the solid delimitation of MJ awaits the careful and systematic application of the comparative methods within and among its major candidate subgroups.[12] This project may yield an internal classification, in addition to successfully delimiting the family, but in the likely eventuality that sound changes will not fully resolve the MJ tree, the application of phylogenetic methods will be essential.

3.6 Panoan and Pano-Takanan

3.6.1 Panoan

Panoan is a family of some 33 languages spoken in eastern Peru, north-eastern Bolivia, and adjacent areas of western Brazil, of which some 25 are still spoken (Valenzuela and Zariquiey to appear; Valenzuela and Guillaume 2017; Fleck 2013). Panoan languages exhibit some marked

[11] Nikulin (2015:46) presents a relatively articulated MJ tree based on lexicostatistical methods, without an explicit clustering method.
[12] Nikulin (2015) provides a preliminary reconstruction of PMJ phonology and lexicon, although the brevity of the presentation makes the reconstructions difficult to evaluate.

similarities in grammar and lexicon with the Takanan languages of Bolivia, leading to the hypothesis of a Pano-Takanan family.

As Oliveira's (2014) cogent critical overview makes clear, the literature on Panoan internal classification has been dominated by expert classifications (e.g., Fleck 2013; Loos 1999) and problematic quantitative studies (e.g. d'Ans 1973; Ribeiro 2006). An important exception is Shell's (1965) phonological and lexical reconstruction of 'Reconstructed Pano',[13] based on seven languages. On the basis of shared phonological innovations, Shell proposes a classification consisting of three first-order groups: Headwaters (Amawaka, Cashinawa, Marinawa), Mainline (Chákobo, Shipibo-Konibo, Kapanawa), and Western (Kashibo). Oliveira (2014) and Valenzuela and Guillaume (2017) summarize work reconstructing of aspects of Proto-Panoan, including Oliveira's (2014) extension of Shell's (1965) reconstruction by including data from 19 languages, including northern Panoan languages. These reconstructions do not in general contribute to an improved understanding of the internal classification of the family, however.

The most recent work on Panoan internal classification is Valenzuela and Zariquiey's (to appear) distance-based phylogenetic study of 200 item Swadesh lists for 20 Panoan languages (plus five Takanan languages that serve as outgroup languages to root the NeighborNet network). The authors tentatively propose three major branches: 1) Northern, consisting of Matses/Mayoruna, Kulina, Korubo, and Matis; 2) Southeastern, consisting solely of Kasharari;

[13] Since Shell was unable include data from northern Panoan languages, she cautiously declined to characterized her results as a reconstruction of Proto-Panoan *per se*.

and 3) Central-Southern, which contains the remaining of the family. The authors do not employ an explicit clustering technique to identify subgroups, however, and appear to have been guided by the rough consensus in previous influential expert classifications in inferring subgroups from the NeighborNet visualization. As a result, the relationship between structure present in the NeighborNet and the subgroups given in the classification is somewhat unclear. An obvious next step for advancing the classification of Panoan languages is to employ character-based CPMs to analyze the valuable cognate set information employed in Valenzuela and Zariquiey's analysis.

3.6.2 Pano-Takanan

Takanan is a family of five languages spoken in Bolivian Amazonia (Guillaume to appear) which have long been speculated to be genealogically related to the languages of the Panoan family (Valenzuela and Guillaume 2017). Key (1968) presented potential Panoan and Takanan cognates and sound correspondences, but Girard (1971) remains the sole application of the comparative method to evaluating the Pano-Takanan hypothesis. Girard identified a significant number of regular sound correspondences from cognates sets of reconstructed Proto-Panoan and Proto-Takanan forms, and reconstructs 166 Proto-Pano-Takanan etyma. Most of the correspondences are only supported by one or two cognate sets, however, and Girard himself remarks on the modest number of cognates that support the correspondences, as well as identifying some inconsistencies in the sound correspondences. This result is clearly promising, but specialists have split, with some regarding the Pano-Takanan relationship very likely or

certain (e.g., Kaufman 1997; Campbell 1997, 2012), and others expressing doubt (e.g., Loos 1999; Fleck 2013).

Supporters of the Pano-Takanan hypothesis have in recent years been encouraged by some striking grammatical similarities that have been identified between languages of the two families, including in their associated motion (Guillaume 2017), transitivity harmony (Valenzuela 2017), and ditransitive alignment systems (Zariquiey 2017). Whether these similarities reflect a genealogical relationship or contact at an early stage in the diversification of the two families remains an open question, whose resolution will ultimately depend on further systematic application of CM to the large quantity of high quality lexical and grammatical data that has become available for both Panoan and Takanan languages since the appearance of Girard (1971).

3.7 Tukanoan

Tukanoan is a family of some 29 languages (Chacon 2014) that are spoken in two areas: one centered on the Vaupés River border region of eastern Colombia and northwestern Brazil, and another centered on middle Putumayo River, spanning parts of the border areas of Ecuador, southern Colombia, and Peru. Most of the languages of the former area participate in a system of obligatory linguistic exogamy (Jackson 1982; Grimes 1985; Sorensen 1967; Stenzel 2005), which has driven linguistic convergence among these languages in ways that are pervasive (Aikhenvald 1999b; 2002), but still not well understood.

Distance-based phylogenetic (Waltz and Wheeler 1972) and expert (Barnes 1999, a.o.) classifications of the Tukanoan family were the most influential until Chacon's (2014) reconstruction of the Proto-Tukanoan (PTk) consonant inventory, and internal classification of the family based on shared phonological innovations. Most significantly, Chacon's (2014) classification identifies a two-way top-level split between Eastern (ET) and Western (WT) branches, in contrast with many earlier classifications, which additionally posit a Central (CT) branch that includes Kubeo (e.g., Waltz and Wheeler 1972), and sometimes other languages as well (e.g., Letuama/Returã and Tanimuka, according to Barnes 1999). Chacon's (2014) classification also differs in important ways from earlier classifications in the organization of lower-level subgroups.

The complex distribution of sound correspondences across the family, reflecting the fact that Tukanoan languages have experienced both sound changes as shared innovations and wave-like diffusions of sound change, complicates the use of sound changes for purposes of internal classification (Chacon 2014; Chacon and List 2015; Malone 1986). First, this complexity has led to quite divergent reconstructions of PTk phonology, generating different distributions of sound changes, and thus, classifications. For example, where Chacon (2014) reconstructs a series of voiceless laryngealized stops, others such as Waltz and Wheeler (1992) and Malone (1986) reconstruct voiced stops; correspondingly Malone's (1986) classification is markedly different from Chacon's (2014). In this light, a clear priority is to extend the empirically well-grounded reconstruction efforts pioneered by Chacon (2014) by reconstructing PTk vowels and

supersegmental features such as nasality and tone. Second, the complex set of sound changes resulting from any given reconstruction allows for a variety of different subgrouping proposals. In response to this challenge, Chacon and List (2015) applied parsimony-based phylogenetic methods, with transition networks modeling sound change tendencies, to sound change characters derived from Chacon (2014). The most successful model yielded a classification that preserved the basic ET/WT split, but differed in a number of ways from Chacon (2014), perhaps most strikingly in placing Kubeo and Tanimuka in a subgroup that was sister to the remainder of ET. Chacon and List (2015) conclude by presenting a conservative tree where on the basis of knowledge of language contact among Tukanoan languages, they manually undid some of the subgroups yielded by the phylogenetic methods employed. This last point highlights the importance both of continuing to study Tukanoan language contact and of turning to sources of evidence which may prove reliable as evidence for subgrouping in the context of intense language contact, such as morphological innovations. In this light, it is significant that the position of Kubeo as branching at the ET root, identified by Chacon and List (2015), is supported by Chacon and Michael's (2018) account of the development of ET past/perfective verbal subject agreement, which shows that ET to the exclusion of Kubeo experienced certain nasal harmony processes in the development of its past/perfective suffixes.

4. Conclusion

Recent decades have seen a flowering of documentation and description of South American languages, which puts South Americanist historical linguists in the best position yet to clarify

the complex genealogical relationships among the languages of the continent. It is now feasible to move beyond expert classifications, which as helpful as they were in the past, leave much to be desired from empirical and analytical standpoints. Methodologically, future progress will depend on judicious use of the classical comparative method and computational phylogenetic methods, where the latter are especially valuable for groups of languages where sound changes are incapable of yielding internal classifications. As Michael and Chousou-Polydouri (2019) argue, however, it is specifically character-based phylogenetic methods that should be employed, since distance-based methods conflate shared retentions and shared innovations in producing classifications. Morphological reconstruction will also likely prove critical for identifying shared innovations capable of informing classifications where sounds changes are not informative.

Also critical for progress will be sustaining and strengthening the collaborations between scholars and indigenous communities in documenting and describing South American languages. For purposes of language classification, support for lexical documentation, which is often not valued in academic circles, but is often valued by indigenous communities, is critical. Work on the classification of some language families is indeed significantly hampered by the lack of adequate lexical documentation. Similarly, support for language description, and not simply the development of annotated corpora, as valuable as they are, will be essential in the longer term for the refinement of internal classifications and evaluation of long-distance proposals. Finally, the investigation of genealogical relationships in South America must proceed in tandem with the study of language contact and areality, which loom large as factors

complicating our understanding of language relationships in South America (see, e.g., Epps and Michael 2017).

5 Acknowledgements

I am grateful for the generous feedback of many South Americanist colleagues, including Joshua Birchall, Lyle Campbell, Thiago Chacon, Natalia Chousou-Polydouri, Nicholas Emlen, Simeon Floyd, Spike Gildea, Sérgio Meira, Denny Moore, André Nikulin, Zachary O'Hagan, Andrés Salanova, Pilar Valenzuela, and Roberto Zariquiey. None is responsible for any misuse I may have made of their valuable insights and knowledge.

6 References

Adelaar W. 1986. La relación quechua-aru: Perspectivas para la separación del léxico. *Revista Andina* 4:379–426.

Adelaar W. 2000. Propuesta de un nuevo vínculo genético entre dos grupos lingüísticos indígenas de la Amazonía occidental: Harakmbut y Katukina. In *Actas del I Congreso de Lenguas Indígenas de Sudamérica*, pp. 219–236. Lima, Peru: Universidad Ricardo Palmas.

Adelaar W. 2004. *The languages of the Andes*. New York, NY: Cambridge University Press.

Adelaar W. 2007. The Quechua impact in Amuesha, an Arawak language of the Peruvian Amazon. In *Grammars in contact: A cross-linguistic typology*, eds. A. Aikhenvald, R.M.W. Dixon, pp.290–312. Oxford: Oxford University Press.

Adelaar W. 2008. Relações externas do Macro-Jê: O caso do Chiquitano. *Topicalizando Macro-Jê*, eds. S. Telles. A. de Paula, pp. 9–28. Recife: NECTAR.

Adelaar W. 2012a. Historical overview: Descriptive and comparative researchon South American Indian languages. In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell and V. Grondona, pp. 1-58. Boston: Mouton de Gruyter.

Adelaar W. 2012b. Languages of the Middle Andes in areal-typological perspective: Emphasis on Quechuan and Aymaran. In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell and V. Grondona, pp. 575-624. Boston: Mouton de Gruyter.

Adelaar W. 2013. Quechua I y Quechua II: En defensa de una distinción establecida. *Revista Brasileira de Linguística Antropológica* 5: 45–65.

Aikhenvald A. 1999a. The Arawak language family. In *The Amazonian languages*, eds. R.M.W. Dixon, A. Aikhenvald, pp. 65-106. New York, NY: CUP

Aikhenvald A. 1999b. Areal diffusion and language contact in the Içana-Vaupés basin, North West Amazonia. In *The Amazonian Languages,* eds. R.M.W. Dixon, A. Aikhenvald, pp. 385–415. New York, NY: CUP.

Aikhenvald A. 2002. Language contact in Amazonia. New Your, NY: OUP.

Aikhenvald A. 2012. The languages of the Amazon. New York, NY: OUP.

Barnes J. 1999. Tucano. In *The Amazonian languages*, eds. R.M.W. Dixon, A. Aikhenvald, pp. pp. 207–226. New York, NY: CUP.

Boswood J. 1973. Evidências para a inclusão do Aripaktsá no filo Macro-Jê. *Série Lingüística* 1:67–78.

Brinton D. 1891. The American race: a linguistic classification and ethnographic description of the native tribes of North and South America. New York, NY: N.D.C. Hodges.

Campbell L. 1995. The Quechumaran hypothesis and lessons for distant genetic comparison. *Diachronica* 12:157–200.

Campbell L. 1997. *American Indian languages: the historical linguistics of Native America*. New York, NY: Oxford University Press.

Campbell L. 2012. Classification of the indigenous languages of South America. In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell and V. Grondona, pp.59–166. Boston: Mouton de Gruyter.

Campbell L, Poser W. 2008 *Language classification: history and method*. Cambridge: Cambridge University Press.

Chacon T. 2014. A revised proposal of Proto-Tukanoan consonants and Tukanoan family classification. *IJAL* 80:275–322.

Chacon T, List J-M. 2015. Improved computational models of sound change shed light on the history of the Tukanoan languages. *Journal of Language Relationship* 13:177–203.

Chacon T, Michael L. 2018. The evolution of subject-verb agreement in Eastern Tukanoan. Journal of Historical Linguistics 8:59–94.

Chamberlain A. 1907. South American linguistic stocks. In Congrès international des américanistes, XVe session, pp. 187-204. Québec: Dussault & Prould.

Cerron-Palomino R. 1987. *Lingüística quechua*. Cuzco, Peru: Centro de estudios rurales andinos "Bartolome de Las Casas".

Constenla-Umaña A. 2012. Chibchan Languages. In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell and V. Grondona, pp.391–439. Boston: Mouton de Gruyter.

Danielsen S, Dunn M, Muysken P. 2011. The role of contact in the spreading of Arawak languages. In *Ethnicity in ancient Amazonia: Reconstructing past identities from archaeology, linguistics, and ethnohistory*, eds. A. Hornborg and J. Hill, pp.173–196. Boulder, CO: University Press of Colorado.

D'Ans A-M. 1973. *Estudios glotocronológicos sobre nueve hablas pano.* Documento de trabajo 17, Centro de Investigación de Lingüística Aplicada. Lima, Peru: Universidad Nacional Mayor de San Marcos.

Davis I. 1966. Comparative Jê phonology. *Estudos Lingüísticos: Revista Brasileira de Lingüística Teórica e Aplicada* 1:10–24.

Davis I. 1968. Some Macro-Jê relationships. *IJAL* 34:42–47.

De la Grasserie R. 1894. *Langue Puquina*. Leipzig: K.F. Koehler.

Deshoullière G, Utitiaj Paati S. 2019. Acerca de la declaración sobre el cambio de nombre del conjunto Jívaro. *Journal de la société des américanistes* 105:167–179.

Derbyshire D. 1999. Carib. In *The Amazonian Languages*, eds. R.M.W. Dixon and A. Aikhenvald, pp.22–64. New York, NY: CUP.

Durbin M. 1977. A survey of the Carib language family. In *Carib Speaking Indians, Culture, and Society*, ed. E. Basso, pp.23–38. Tucson, AZ: University of Arizona Press.

Dunn M. 2015. Language phylogenies. In *The Routledge handbook of historical linguistics*, eds. C. Bowern, B. Evans, pp.208–229. London: Routledge.

Emlen N. 2017. Perspectives on The Quechua–Aymara Contact Relationship And The Lexicon And Phonology Of Pre-Proto-Aymara. *IJAL 83*:307–340.

Emlen N, Adelaar W. 2017. Proto-Quechua and Proto-Aymara agropastoral terms. In *Language Dispersal Beyond Farming*, eds. M. Robbeets, A. Savelyev, pp.25-46. Philadelphia: John Benjamins.

Epps P. 2009. Language classification, language contact, and Amazonian prehistory. *Language and Linguistics Compass 3*:581–606.

Epps P, Bolaños K. 2017. Reconsidering the "Makú" language family of Northwest Amazonia. *IJAL* 83: 467–507.

Epps P, Michael L. 2017. The areal linguistics of Amazonia. In The Cambridge Handbook of Areal Linguistics, ed. R. Hickey, pp.934–963. New York, NY: CUP

Eriksen L. 2011. *Nature and culture in prehistoric Amazonia*. PhD dissertation, Lund University, Lund, Sweden. 364pp.

Eriksen L, Danielsen S. 2014. The Arawakan matrix. In *The native language of South America: Origins, Development, Typology*, eds. L. O'Connor, P. Muysken, pp.152–176. New York, NY: CUP.

Facundes S, Brandão AP. 2011. Comparative Arawak linguistics: notes on reconstruction, diffusion and Amazonian prehistory. In *Ethnicity in ancient Amazonia: Reconstructing past identities from archaeology,linguistics, and ethnohistory*, eds. A. Hornborg, J.. Hill, pp.197–210. Boulder, CO: University Press of Colorado.

Fleck D. 2013. Panoan languages and linguistics. *American Museum of Natural History Anthropological Papers 99*. New York, NY: American Museum of Natural History.

Floyd S. to appear. The origins of Ecuadorian Highland Quichua. *IJAL*.

Galucio A, Meira S, Birchall J, Moore D, Gabas Júnior N, Drude S, Storto G, Picanço G, Rodrigues CR. 2015. Genealogical relations and lexical distances within the Tupian linguistic family. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10:229–274.

Galucio A, Nogueira A. 2011. 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*, pp.1-28. Austin, TX: University of Texas at Austin.

Gildea S. 2012. Linguistic studies in the Cariban family. In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell and V. Grondona, pp. 441-494. Berlin: Mouton de Gruyter.

Gildea S. 2003. The Venezuelan branch of the Cariban language family. *Amérindia* 28: 7–32.

Gildea S, Hoff B, Meira S. 2010. The story of *ô in the Cariban family. In *Fieldwork and linguistic analysis in Indigenous languages of the Americas*, eds. A. Berez, D. Rosenblum, J. Mulder, pp.91–123. Honolulu: Language Documentation and Conservation Special Publication 2.

Girard V. 1971. *Proto-Takanan phonology*. University of California Publications in Linguistics 70. Berkeley/Los Angeles: University of California Press.

Girard, Victor.1971 *Proto-Carib phonology*. PhD dissertation. University of California, Berkeley, Berkley, CA.

Greenberg J. 1987. *Language in the Americas.* Stanford, CA: Stanford University Press.

Grimes B. Language attitudes: Identity, distinctiveness, survival in the Vaupes. *Journal of Multilingual & Multicultural Development* 6: 389-401.

Guillaume A. To appear. Tacanan. In Amazonian languages: An international handbook, eds. P. Epps and L. Michael. Berlin: de Gruyter Mouton.

Guillaume A. 2017. Sistemas complejos de movimiento asociado en las lenguas Takana y Pano: perspectivas descriptiva, tipológica e histórico-comparativa. *Amerindia* 39:211–261.

Guy, G. 2008. Variationist approaches to phonological change. In The handbook of historical linguistics, eds. B. Joseph, R. Janda, pp.369–400. Malden, MA: Blackwell.

Heggarty P. 2005. Enigmas en el orígen de las lenguas andinas: aplicando nuevas técnicas a las incógnitas por resolver. *Revista Andina* 40:9–80.

Hornborg A. 2005. Ethnogenesis, Regional Interaction, and Ecology in Prehistoric Amazonia. *Current Anthropology* 46: 589–620.

Jackson J. 1983. *The Fish People. Linguistic Exogamy and Tukanoan Identity in Northwest Amazonia*. Cambridge: Cambridge University Press.

Jensen C. 1998. Comparative tupí-guaraní morphosyntax. In *Handbook of Amazonian languages,* vol. 4, eds. D. Derbyshire, G. Pullum, pp. 489-618. New York: Mouton de Gruyter.

Jolkesky, M. 2010. *Reconstrução fonológica e lexical do Proto-Jê Meridional*. MA thesis, Universidade Estadual de Campinas, Campinas, Brazil.

Kaufman T. 1990. Language history in South America: What we know and how to know more. In *Amazonian Linguistics: Studies in Lowland South American Languages*, ed. D. Payne, pp.13–67. Austin, TX: University of Texas Press.

Kaufman T. 1994. The native languages of South America. In *Atlas of the World's Languages*, eds. C. Moseley, R. E. Asher (eds.), pp.59–93. London: Routledge.

Key, M. R. 1968. *Comparative Tacanan phonology: with Cavineña phonology and notes on Pano-Tacanan relationship*. The Hague: Mouton.

Labov, W. The regularity of regular sound change. *Language* 96: 42-59.

Landerman P. 1991. Quechua dialects and their classification. PhD dissertation, UCLA, Los Angeles, CA.

Lemle, Miriam. 1971. Internal classification of the tupí-guaraní linguistic family. In *Tupí studies I*, ed. D. Bendor-Samuel, pp.107–129. Norman, OK: Summer Institute of Linguistics.

Loos E. 1999. Pano. In *The Amazonian languages*, eds. R.M.W. Dixon, A. Aikhenvald, pp.227-249. New York, NY: CUP.

Malone T. 1986. *Proto-Tucanoan and Tucanoan genetic relationship*. Ms., Lomalinda, Colombia: Instituto Linguístico de Verano.

Mannheim, Bruce. 1991. *The language of the Inca since the European invasion*. Austin, TX: University of Texas Press.

Martins A. 2011. *Uma avaliação da hipótese de relações genéticas entre o Guató e o tronco Macro-Jê*. PhD dissertation, Unversidade de Brasília, Brasília, Brazil. 417pp.

Meira S. 2000. *A reconstruction of Proto-Taranoan: Phonology and morphology*. Munich: LINCOM Europa.

Meira S. 2005. Cariban languages. In *Encyclopedia of Languages and Linguistic Theory*, ed. K. Brown, pp.199–204. Oxford: Elsevier.

Meira S, Birchall J, Chousou-Polydouri S. 2015. *A character-based internal classification of the Cariban family*. Talk presented at the 48th annual meeting of the Societas Linguisticae Europaea, Leiden, September 4, 2015.

Meira S, Drude S. 2013. Sobre a origem histórica dos "prefixos relacionais" das línguas Tupí-Guaraní. C*adernos de Etnolingüística* 5:1-31.

Meira S, Franchetto B. 2005. The Southern Cariban Languages and the Cariban Family. *IJAL* 71:127–192.

Meira S, Gildea S, Hoff B. 2010. On the origin of ablaut in the Cariban family. *IJAL* 76(4):477–515.

Mello A. 2000. *Estudo histórico da família linguística tupí-guaraní: Aspectos fonológicos e lexicais.* PhD dissertation, Universidade Federal de Santa Catarina, Florianópolis, Brazil.

Mello A. 2002. Evidencias fonológicas e lexicais para o sub-agrupamento interno tupí- guaraní. In *Línguas indígenas brasileiras: Fonologia, gramática e história,* vol. 1, eds. A.S. Cabral, A. Rodrigues, pp.338–342. Belém: Editora Universitária, Universidade Federal do Pará.

Michael L, Chousou-Polydouri N. 2019. Computational phylogenetics and the classification of South American languages. *Language and Linguistics Compass 13*: e12358.

Michael L, Chousou-Polydouri N, Bartholomei K, Donnelly E, Wauters V, Meira S, O'Hagan Z. 2015. A Bayesian Phylogenetic Classification of Tupí-Guaraní. *LIAMES* 15:193–221.

Moore D, Galucio A. 1994. Reconstruction of Proto-Tupari consonants and vowels. Survey Reports, Survey of California and Other Indian Languages. Berkeley, CA: University of California, Berkeley .

Nichols J, Peterson D. 1996 The Amerind personal pronouns. *Language* 72:336–371.

Nichols J, Warnow T. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2:760–820.

Nikulin A. 2015. On the genetic unity of Jê-Tupí-Karib. MA thesis, Lomonsonov Moscow State University, Moscow, Russia. 165pp.

Nikulin A. 2016. A phonological reconstruction of Proto-Cerrado (Jê family). *Journal of Language Language Relationship* 14:165-185.

Nikulin A. 2017. A phonological reconstruction of Proto-Cerrado (Jê family). *Journal of Language Relationship 15*:147–180.

Nikulin A, da Silva M. 2020. As línguas Maxakalí e Krenák dentro do tronco Macro-Jê. *Cadernos de Etnolingüística* 8:1–64.

O'Hagan Z, Chousou-Polydouri N, Michael L. 2019. Phylogenetic classification supports a Northeastern Amazonian Proto-Tupí-Guaraní homeland. *LIAMES 19*:1-29.

Oliveira S. 2014. *Contribuições para a reconstrução do Proto-Páno*. PhD dissertation, Universidade de Brasília, Brasília, Brazil.

Orr C, Longacre R. 1968. Proto-Quechumaran. *Language* 44:528–555.

Pache M. 2016. Pumé (Yaruro) and Chocoan: Evidence for a new genealogical link in northern South America. *Language Dynamics and Change 6*: 99–155.

Parker G. 1971. Comparative Quechua phonology and grammar V: The evolution of Quechua B. *University of Hawaii Working Papers in Linguistics* 3:45–109.

Parker G. 1969a. Comparative Quechua phonology and grammar I: Classification. *University of Hawaii Working Papers in Linguistics* 1:65–87.

Parker G.1969b. Comparative Quechua phonology and grammar II: Proto-Quechua phonology and morphology. *University of Hawaii Working Papers in Linguistics* 1: 123-147.

Parker G. 1969c. Comparative Quechua phonology and grammar III: Proto-Quechua lexicon. *University of Hawaii Working Papers in Linguistics* 1: 1-61.

Parker G. 1969d. Comparative Quechua phonology and grammar IV: The evolution of Quechua A. *University of Hawaii Working Papers in Linguistics* 1(2): 149-204.

Parker G. 1963. La clasificación genética de los dialectos quechuas*. Revista del Museo Nacional* XXXII: 241-252.

Payne D. 1985. The Genetic Classification of Resigaro. *IJAL* 51:222–231.

Payne D. 1991. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In Handbook of Amazonian Languages, vol. 3, eds. D. Derbyshire, G. Pullum, pp.355–499. Berlin: Mouton de Gruyter.

Payne D. 2005. Apolista (Lapachu) as a Maipuran Arawakan language. *Revista Latinoamericana de Estudios Etnolingüísticos 10*:239-250.

Pearce A, Heggarty P. 2011. "Mining the Data" on the Huancayo-Huancavelica Quechua Frontier. In *History and Language in the Andes*, eds*.* P. Heggarty and A. Pearce*,* pp.87–109. New York, NY: Palgrave Macmillan.

Ramirez H. 2001. *Lińguas arawak da Amazônia setentrional: comparação e descrição.* Manaus, Brazil: Editora da Universidade do Amazonas.

Ramirez H, Vegini V, de França M. 2015. Koropó, puri, kamakã e outras línguas do Leste Brasileiro. *LIAMES 15*: 223–277.

Rankin R. 2003. The comparative method. In *The Handbook of Historical Linguistics*, eds. B. Joseph, R. Janda, pp.183–212. Malden, MA: Blackwell.

Ribeiro E. 2006. Macro-Jê. In *Encyclopedia of Language & Linguistics* (2nd ed.), Vol. 7, ed. K. Brown, pp.422–426. Oxford: Elsevier.

Ribeiro E. 2011. Prefixos relacionais como evidência histórico-comparativa: os casos Chiquitano e Jabutí. *Línguas e Culturas Macro-Jê* 2:105–120.

Ribeiro E. 2012a. *A grammar of Karajá*. PhD dissertation, University of Chicago, Chicago, IL.

Ribeiro E. 2012b. Final consonants in Maxakalí and their comparative status. *LIAMES* 12:189-193.

Ribeiro E, van der Voort H. 2010. Nimuendajú was right: The inclusion of the Jabutí language family in the Macro-Jê stock. *IJAL 76*:517–570.

Ribeiro L. 2006. Uma proposta de classificação interna das línguas da família Pano. *Revista Investigações 19*:157–188.

Rodrigues A. 1984/1985. Relações internas na família lingüística tupí-guaraní. *Revista de Antropologia* 27/28:33–53.

Rodrigues A. 1985. Evidence for Tupi–Carib relationships. In *South American Indian Languages: Retrospect and Prospect*, ed. Harriet E. Manelis Klein and Louisa R. Stark, pp.371–404. Austin: University of Texas Press.

Rodrigues A. 1986. *Línguas brasileiras: para o conhecimento das línguas indígenas*. São Paulo, Brazil: Loyola.

Rodrigues A. 1999. Macro-Jê. *The Amazonian Languages*, ed. R.M.W. Dixon and A. Aikhenvald, pp.165–206. New York, NY: CUP.

Rodrigues A. 2005 As vogais orais do Proto-Tupí. In *Novos estudos sobre línguas indígenas*, eds. A. Rodrigues, A.S. Cabral, pp.35–46. Brasília: Editora Universidade de Brasilia.

Rodrigues A. 2007. As consoantes do Proto-Tupí. In *Línguas e culturas Tupi*, eds. A.S. Cabral. A. Rodrigues, pp.167–203. Campinas: Curt Nimuendaju; Brasília: LALI.

Rodrigues A. 2009. A Case of Affinity among Tupí, Karíb, and Macro-Jê. *Revista Brasileira de Linguística Antropológica* 1:137-162.

Rodrigues A, Cabral AS. 2002. Revendo a classificação interna da família tupí-guaraní. In *Línguas indígenas brasileiras: fonologia, gramática e história*, eds. A.S. Cabral, A. Rodrigues, pp.327–337. Belém: Editora Universitária, Universidade Federal do Pará.

Rodrigues A, Cabral AS. 2012. Tupían. In . In *The indigenous languages of South America: A comprehensive guide*. Vol. 2, eds. L. Campbell, V. Grondona, pp.495-574. Boston: Mouton de Gruyter.

Santana, Á. 2006. *Comparações preliminares entre a língua Chiquitano (Brasil/Bolívia) e o Proto-Jê*. Paper presented at the Symposium "Advances in Native South American Historical Linguistics" at the Fifty-second International Congress of Americanists, Seville.

Santos-Granero F. 2002. The Arawakan matrix: ethos, language, and history in native South America. In *Comparative Arawakan histories: Rethinking language family and culture area in Amazonia*, eds. J. Hill, F. Santos-Granero, pp.25–50. Urbana, IL: University of Illinois Press.

Schleicher C. 1998. *Comparative and Internal Reconstruction of the Tupi-Guarani Language Family.* PhD dissertation, University of Wisconsin – Madison, Madison, Wisconsin. 372 pp.

Seifart F, Echeverri JA. 2015. Proto Bora-Muinane. *LIAMES* 15:279–311.

Seifart F, Hammarström H. 2017. Language isolates in South America. In *Language isolates*, ed. L. Campbell, pp. 260-286. New York, NY: Routledge.

Seki L. 2002. O Krenak (Botocudo/Borum) e as línguas Jê. In *Línguas Je*: *estudos vários*, eds. L. dos Santos, I. Pontes, pp. 15–40. Londrina, Brazil: Editora UEL.

Shell O. 1965. *Pano reconstruction*. PhD dissertation, University of Pennsylvania, Philadelphia, PA. 534pp.

Sorensen, A. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69:670–684.

Stark T. 2018. Caribbean Northern Arawak person marking and alignment: A comparative and diachronic analysis. PhD dissertation, UC Berkeley, Berkeley, CA. 122 pp.

Stenzel K. 2005. Multilingualism in the Northwest Amazon, revisited. *Memorias del Congreso de Idiomas Indígenas de Latinoamérica-II*, pp.1-28. Austin, TX: University of Texas at Austin.

Torero A. 1964. Los dialectos quechuas. *Anales Científicos de la Universidad Agraria* 2:446–478.

Torero A. 1968. Procedencia geográfica de los dialectos quechuas de Ferreñafe y Cajamarca. *Anales Científicos de la Universidad Agraria* 3-4:291–316.

Torero A. 1970. Lingüística e historia de la sociedad andina. *Anales Científicos de la Universidad Agraria* 8:231–264.

Torero A. 1983. La familia lingüística quechua. In *América Latina en sus lenguas indígenas*, ed. B. Pottier, pp.61–92. Caracas: Monte Avila Editores.

Valenzuela P. 2017. Armonía transitiva en las lenguas Pano y Takana. Amerindia 39:407–451.

Valenzuela P, Guillaume A, 2017. Estudios sincrónicos y diacrónicos sobre lenguas Pano y Takana: una introducción. *Ameríndia* 39:1–49.

Valenzuela P, Zariquiey R. to appear. Pano. In *Amazonian languages: An international handbook*, eds. P. Epps, L. Michael. Berlin: De Gruyter Mouton.

Walker R, Ribeiro L. 2011. Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*: 2562–2567.

Waltz N, Wheeler A. 1972. Proto Tucanoan. In *Comparative studies in Amerindian language*, eds. E. Matteson, A. Wheeler, F. Jackson, N. Waltz, D. Christian, pp.119–149. The Hague: Mouton.

Wise MR. 1976. Apuntes sobre la influencia Inca entre los amuesha: factor que oscurece la clasificación del idioma. *Revista del Museo Nacional* XLII:355-366.

Zamponi R. 2017. First-person n and second-person m in Native America: a fresh look. *Italian Journal of Linguistics* 29:189–230.

Zariquiey R. 2017 Construcciones ditransitivas en Pano (y Takana): una aproximación basada en propiedades. *Amerindia* 29:295–322