Fernando O. de Carvalho
Federal University of Amapá (UNIFAP), Brazil; fernaoorphao@gmail.com

Evaluation of cognation judgments undermines computational phylogeny of the Arawakan language family[1]

The goal of this paper is to critically examine the linguistic analyses underlying Walker & Ribeiro (2011), a widely cited computational phylogenetic study of the Arawakan language family. To the extent that their claims concerning the internal classification of this vast group of languages hinges on their cognation judgments, and that their more ambitious claims concerning prehistoric migration routes of Arawakan-speaking peoples depend, in turn, on this proposed internal classification, I show that outright rejection of their results is highly commendable. Errors include both false negatives, where cognation relations between comparanda were missed, as well as false positives, where non-cognate material in different languages were judged to be reflexes of single Proto-Arawakan etyma. No clear pattern seems to emerge from their cognation decisions, and the resulting judgments seem to be, in many cases, so strongly detached from even impressionistic assessments of similarity that the resulting distribution of cognation scores could have been produced independently of the data. The paper ends with a plea for greater sobriety in the historical linguistics of native South America, which should focus on clear and well-supported applications of the comparative method before embarking on endeavors that depend on this traditional work for their success. More importantly, though, South Americanists should avoid groundless statements on the supposed uselessness, or exhaustion, of the comparative method as a tool for uncovering the linguistic history of the continent.

Keywords: Arawakan languages; cognation; etymology; computational phylogeny.

1. Introduction

This paper seeks to present the results of the task of revising a sample of the cognate judgments used as input data for a computational phylogenetic study of the Arawakan language family of South America. Walker & Ribeiro (2011), henceforth, WR, attempt to advance claims on the migrations responsible for the attested geographic spread of Arawakan-speaking peoples throughout South America, basing themselves on a phylogeny of the Arawakan languages derived from comparative vocabulary data or, more precisely, from patterns of distribution of cognate elements across 60 compared languages. Given their ambitious goals and the popularity of their study, I must at the outset announce that the results of this review are critically damaging and entirely negative to the goals set by WR.

WR has been cited in the molecular anthropology literature (see e.g. Barbieri *et al*. 2014; Scliar *et al*. 2014; Nogueira-Santamaría *et al*. 2015) and has been brought up in more general discussions on the application of computational phylogenetic methods in historical linguistics

---

(Pereltsvaig & Lewis 2015: 56). It was described as one of many 'successful' applications of quantitative methods for phylogeny inference in language (Honkola *et al.* 2013: 1245), or of Bayesian character-based methods for phylogenetic inference more specifically (Galucio *et al.* 2015). It has been cited by studies applying novel phylogeny-inference tools to languages outside of South America (Fris Boegh, K. *et al.* 2016) and its results have informed more ambitious approaches to the history of the Arawakan languages and peoples (see Carling *et al.* 2012; Ericksen & Danielsen 2014). Many of these synthetic classifications of the Arawakan language family, and attendant inferences of the pre-historic events such as rates of diversification or diffusion, have been built around a comparison of the WR classification with other, independent classifications based on grammatical and morphological features (Ericksen & Danielsen 2014; Epps 2015). Given this accolade, it is perhaps remarkable that, in the nine years since its publication, no independent scrutiny or checking of the analyses underlying WR has been produced. One can only guess at the reasons for this neglect, but it certainly stems from a perceived underestimation of the effects errors in data analysis may have for lexicostatistical and other quantitative approaches to language classification, as noted by Geisler & List (2010: 2–3).

It goes without saying that WR's cognation judgments do not exhaust those aspects of their work that could be examined or criticized. These are, however, the usual focus of attention and immediate concern of linguists (see e.g. Nichols & Warnow 2008: 769 for this general observation, and Eska & Ringe 2004 for a particularly telling example). The critical importance of having sound cognation judgments is not due, however, to accidents of expertise. Computational phylogenetic analyses "are only ever as good as the data and analysis that underlie them" (McMahon & McMahon 2005: 48). As nearly everyone in the field agrees, reliable phylogenetic inference of language histories builds on solid linguistic work supporting cognate identification and the derivation of 'phyletic characters' (Ringe, Warnow & Taylor 2002; Eska & Ringe 2004: 570). This dependence, and its implications for the evaluation of work in computational phylogenetics of language, is aptly expressed by McMahon & McMahon (2005: 97):

> "It is important, especially for non-linguists, to appreciate this fact, since in many disciplines the data in comparisons are considerably more direct, with numbers derived directly from measured similarities between the items. In historical linguistics (…) the data are actually judgments made on the basis of those similarities in the light of historical knowledge and investigation: but they are crucially judgments, rather than algorithmically derived measures of similarity. Quantification is therefore strictly two steps away from the original linguistic material, being a numerical transformation of linguistic cognacy scores."

For this reason, I decided to subject the cognation judgments in WR to careful scrutiny. This investigation started from an analysis of the WR coded matrix where each cognate decision is entered as either 1 or 0. The character matrix was published as a supplementary material (Supplementary Material Nr. 3) to their paper. I checked the specific vocabulary items WR used in their judgments, by looking at their wordlists, kindly provided to me by Robert S. Walker (via e-mail, in August 28, 2017). In sampling from their matrix, I gave particular attention to the languages I happen to know best — Terena, the Mojeño dialects (Ignaciano and Trinitario), and the languages of the Purus branch (Yine, Apurinã and Iñapari). There is no reason, however, to think that the shortcomings and limitations attested here are limited to this specific sample of languages in the exact semantic glosses I have chosen for scrutiny. At certain points of the discussion, observations are also made on other languages, such those of the Caribbean branch (Lokono, Island Carib/Garifuna and Wayuunaiki) or the languages of the Kampan branch[2].

---

[2] A reviewer brings up the important point that the label Kampan (Kampa, Campa) is considered pejorative by some speakers of some of the relevant languages. Michael (2020: 97, fn. 2) proposes *Nihagantsi*, the Proto-

This paper is organized as follows: Section **2** briefly discusses the nature of WR's study, in its most relevant aspects. Section **3** is concerned with the evaluation of WR's claims and, in particular, with a critical evaluation of their cognation judgments. This section is divided in headings, each of which stands for an individual semantic gloss or meaning that constitutes a sub-section in itself, where specific cognation judgments made by WR will be evaluated. In section **4** I offer an overall evaluation of the cognation judgments by WR, stressing the visible lack of any clear standard of formal comparability that could have guided their decisions. Finally, section **5**, by way of conclusion, relates the nature of WR's study and its limitations to some perceived features of the current context of the historical and comparative linguistics of indigenous South America, ending with a plea for a research effort focused on the application of the traditional comparative method instead of the production of ambitious yet groundless claims that depend, for their success, on the conclusion of basic work still awaiting to be done.

2. WR on the Arawakan languages and peoples

WR relied on vocabulary data to build an internal classification of 60 Arawakan languages/dialects. The vocabulary data consisted of cognation judgments over a 100-item Swadesh meaning list. WR's method is thus essentially *lexicostatistic*, as it depends on the assumed existence of a *basic vocabulary* and on the premise that shared retentions in the vocabularies of different languages reflect their degrees of genetic relationship. The specific aspect or step in the WR methodology that must be clarified in the present study concerns their *cognation decisions*, subsequently coded in the data matrix published as supplementary material to their paper. Using Geisler & List's (2010: 2) presentation of the working procedure for lexico-statistical approaches for inferring language phylogenies, presented in (1) below, the present discussion will deal mainly with Step 3 of the procedure:

(1) *Procedures for Lexicostatistical approaches to language phylogeny*
   Step 1: Compilation
   Step 2: Translation
   Step 3: Cognate Judgments
   Step 4: Coding
   Step 5: Computation

Steps 1 and 2 relate to a number of important issues in lexicostatistical work concerning both the composition of meaning lists and the selection of comparanda based on their meanings. For instance, different researchers diverge on the importance they place on meaning identity: while some researchers would allow solely for 'lexicostatistical cognates' (with required meaning identity among compared items), others would not oppose the use of 'etymological cognates' (not requiring meaning identity) as well. In fact, members of the latter group will often insist that the exclusive reliance of classical lexicostatistical work on the former constitutes a shortcoming that can and should be avoided (see e.g. Heggarty 2010: 315–316 for discussion and some references). I have chosen not to address these matters in the evaluation of WR's work for the simple reason that no consistent practice seems to have been followed by WR in their sampling of comparanda; the authors advance cognation claims both for lexicostatistical and for etymological cognates. An example of the latter is WR's inclusion of Terena *sîni*

Kampan form for 'language', as a new label for this group of languages. I have not considered it reasonable to introduce this innovation in the present paper, but the issue is nevertheless worthy of mention.

'jaguar' in their entry for 'dog', instead of the semantically-proper match *tamúku*, as *sîni* 'jaguar' is certainly a much better candidate for being a cognate of forms such as Baniwa *ʧíinu* 'dog'. If I bring this issue up at any moment in the remainder of this paper, it will certainly have a secondary role only. On step 3 itself, the demonstration of cognation, a commentary is needed on the approach to be followed here. Many of the incorrect cognation judgments by WR could be identified as such based on simple inspection. Thus, for 'tree', WR judge Terena *tikóti* and Apurinã *ããmɨna* as cognates, which does not seem like a compelling proposal even as a 'first-pass etymology', *sensu* Watkins (1990). Nevertheless, I have taken great pains to elaborate on the formal factors (e.g. regular sound correspondences and morphological analysis of the comparanda) that ground either the acceptance or the rejection of the claims being evaluated. Besides the rather obvious fact that rigor is always a desirable feature of argumentation for or against etymological proposals, it must be kept in mind that Arawakan is an under-researched language family from the historical point of view, and this strategy has the benefit of contributing additional insights in an attempt to help filling relevant gaps in our knowledge. This issue will be touched on again in sections **4** and **5**, where the incipient status of Arawakan historical linguistics and the role of the comparative method in advancing rigorous etymologies will be discussed within a broader context.

Each etymology or set of cognate forms in WR's dataset defines a character and each language will, for each such character, score either a 1, in which case the language contributes a witness to this etymology, or 0, when the language has no cognate of this form. WR's cognate decisions are in turn aligned in relation to entries in meaning list, so that each meaning entry or gloss has at least one and potentially many more characters (i.e. cognate sets) within it. WR's cognation judgments produce a data matrix which is, in turn, the input data WR use to infer a phylogeny (an evolutionary history of diversification), using 'a recent Bayesian technique' for this purpose (Walker & Ribeiro 2011: 2). One of the novel and exciting aspects of WR's study is the promise of unravelling not only the temporal pattern of diversification and split of the Arawakan languages but also their *spatial* pattern of evolution, ultimately yielding the most likely point of origin (or homeland) of the Arawakan peoples. Certainly, all the steps necessary in producing this final inference — a hypothesis on the dispersions of the Arawakan-speaking peoples derived from a linguistic phylogeny and from geographic data — could be subject to scrutiny. However, the whole edifice built by WR depends, as noted in the preceding section, on the cognation judgments represented by their data matrix and it is these judgments that constitute the focus of this paper.

3. Problems with the Walker & Ribeiro (2011) study

There is hardly any point in subjecting WR's data matrix to the time-consuming task of examining *all* the cognation judgments underlying it. Section **4**, which discusses some of the broader generalizations on WR's cognation decisions, is based on the reasonable assumption that the issues identified in the sample of decisions examined here are representative of the whole set of decisions on which WR's study is ultimately based.

3.1. On singleton cognate sets

One of the most striking claims made by WR is the statement that "the words 'I', 'you', 'we', 'know' and 'sun' were coded as having only a single cognate across all Arawak languages" (Walker & Ribeiro 2011: 2). As shown below, this claim is either crucially imprecise

(as in some cases the relevant relation may be, at best, one of *partial* cognation) or obviously incorrect, as no grounds can be offered for treating the comparanda as cognates.

Arawakan languages typically have three sets of syntactically and morphologically distinct (and specialized) forms grammatically indexing, or expressing, the category of person (and, in some cases, number and gender as well): Cross-referencing prefixes, which code features of a possessor in nouns, of the A[3] argument in transitive verbs and of the S argument of a subset of intransitive verbs; cross-referencing suffixes coding a O argument in transitive verbs and the S of some intransitive verbs; and independent personal pronouns (see Payne 1987: 61–65; 1991: 375–389; Aikhenvald 1999)[4]. The latter are usually derived by the attachment of the cross-referencing prefixes to a 'pronominal base'. Examination of the data used by WR reveals that it is exactly the latter, the independent personal pronouns, that were used in their comparison.

For the discussion of the singleton sets, that is, those semantic glosses for which WR claim a single cognate set exists for the entire Arawakan family, one may initially focus on data such as that in Table 1, presenting relevant forms from a representative subset of Arawakan languages, all included in the WR sample. Data in Table 1 come from Nies (1986: 551, 559, 573, 581) for Yine, Klumpp (1995: 140, 150, 156, 158) for Piapoco, Pet (2011: 181, 183, 196, 201) for Lokono, Launey (2003: 65) for Piapoco and Shaver (1996: 288, 322, 327) for Nomatsigenga. Terena data come from my own fieldwork materials.

		Piapoco	Terena	Palikur	Lokono	Nomatsigenga	Yine
I		*nú-a*	*û^{n}di*	*nah*	*de*	*na-ro*	*hita*
You		*pí-a*	*îti*	*pis*	*bi*	*obi-ro*	*pica*
We		*wí-a*	*ûti*	*wis*	*we*	*eí-ró*	*wica*
Sun		*èeri*	*káʃe*	*kamuw*	*(h)adali*	*paba*	*tkatʃi*
Know		*nú-alía-ka-wa*	*é-ʃ-o*	*hijak*	*eitʰi-n*	*i-gó-k-e-ri*	*rumat-lɨ*

Table 1

Examination of the comparanda within each set is enough to suggest that only with the recognition of regular sound correspondences, and not simply on the basis of impressionistic evaluations of rough phonetic similarity, would it be plausible to postulate, for instance, that Terena *káʃe*, Palikur *kamuw*, Lokono *(h)adali* and Nomatsigenga *paba*, all meaning 'sun', are in fact cognate[5]. This proposed cognate set will be evaluated in section **3.2** along with other lexical comparisons; suffice it to say, for now, that WR's cognation decisions reveal a total lack of adherence to the basic cannons of the comparative method and even defy understanding or justification on the basis of gross assessments of phonetic similarity.

---

[3] A denotes the most active/controlling argument of prototypically transitive verbs, O denotes the least active/ affected argument of prototypically transitive verbs, while S refers to the single argument of intransitive verbs.

[4] Cross-referencing prefixes realize the person-number features of possessors in nouns and cross-reference an A argument, that is, the 'Subject' of transitive verbs. Cross-referencing suffixes can apparently be reconstructed for Proto-Arawakan (see Payne 1991) though many of the daughter languages have impoverished systems implying total loss or great simplification of this set of markers. Such suffixes usually code an O argument, that is, the 'Object' of transitive verbs, or So, the single argument of a stative intransitive, in those languages with split intransitive systems. Note also that many primary descriptive sources on Arawakan languages describe these prefixes and suffixes as 'clitics' instead. As nothing in the present discussion depends on this particular issue, I will apply the labels *prefix* and *suffix* throughout.

[5] Lokono *hadali* and *adali*, for 'sun', are dialectal variants (see Patte 2011: 40), with *hadali* being considered the most common variant (see Patte 2011: 13).

On the pronominal forms, note, first, that, as indicated by morphological segmentations in Table 1 above, Arawakan independent pronouns are usually derived from two elements: The first, leftmost, is formally and functionally identical to the person cross-referencing prefixes that occur in nouns (coding the possessor in possessive constructions) and in verbs (coding the A argument of transitive verbs and the S argument of a subset of intransitive verbs)[6]. The second element, however, likely originates in certain deictic elements to which these cross-referencing prefixes were attached and, as discussed in Carvalho (2016a), there is evidence that these deictic pronominal bases (or, better yet, their selection as bases for the independent pronouns) are branch- or language-specific innovations, rather than retentions from Proto-Arawakan. This would make Arawakan independent pronouns, at best, partial cognates, and losing sight of their origin in these deictic bases amount to missing information that may turn out to be vital for internal classification within this family. A second point, now related to the prefixes that appear in these independent pronouns, concerns the well-known distinction between *ta*-Arawakan and *nu*-Arawakan languages, based on the form of the first person singular prefix attested in a subset of the Caribbean Arawakan languages (Lokono and Wayuunaiki/Añun) and that attested in the rest of the family, respectively. So far, few researchers have entertained the idea that the *nu-* and *ta-* prefixes could be cognate (see Payne 1985, 1987; Oliver 1989: 152). Others simply 'list' both forms as Proto-Arawakan first person singular prefixes (see Aikhenvald 2009: 61). Finally, the Yine first person singular pronoun *hita* is certainly an oddball in the comparison, lacking even internal partial cognates: while the first person singular prefix in the language is *no-*, a clear cognate of the marker found in the other '*nu*-Arawakan' languages, the independent first person singular pronoun has the form *hita*, presumably *hi-ta*, which is seemingly unrelated to the *no-* prefix (see Hanson 2010: 45–49). In conclusion, then, there are clear grounds for a strong skepticism regarding the claim that single cognate sets can be established throughout the 60 Arawakan languages compared by WR for the meanings 'I', 'You' and 'We'.

3.2. Evaluating cognation judgments

The discussion below will pinpoint many cognate decisions underlying WR coding scheme that simply defy any principled formulation, being at times over-relaxed — that is, coding as cognates forms that are not — and being at times overstrict — that is, ignoring demonstrably cognate material. All in all, the cognation decisions seem to have been carried in the absence of any historical comparative knowledge of the Arawakan languages, and without applying any objective method to guide and motivate such decisions.

Sun

A single cognate set for this meaning slot was found for the 60 Arawakan languages compared by WR. What is striking about this is not only the total absence of regular correspondences that could support the recognition of formally very distinct comparanda as cognates (as noted above in section **3.1**), but, and this is more telling, the fact that plenty of evidence exists on the secondary character of many terms for 'sun' in Arawakan languages.

---

[6] The Nomatsiguenga form for the 1PL pronoun in the WR data available to me is *kiro*. This must be a typo, as no entry with this form exists in the Shaver (1996) dictionary, used by WR as their data source on the language. Only *eíró* is attested for the meaning 'we'. For another Kampan language, Michael (2008: 373) mentions the formation of the Nanti personal pronouns by the prefixation of person-number prefixes to a base *-ro*, which he calls a suffix.

For the languages of the Purus branch of the Arawakan family we have: Yine *tkaʧi* 'sun' (Nies 1986: 559), Apurinã *atokaʧi* (Facundes 2000: 655) and Iñapari *tuʔatí* (Parker 1995: 66). These are, in turn, reflexes of the Proto-Purus etymon **atukatsi* 'sun', which lacks semantically comparable cognates elsewhere in the family, and likely constitutes a lexical isogloss of this particular branch (see Carvalho, forthcoming)[7].

For the languages of the Kampan branch, WR include comparanda such as Nomatsigenga *paba* 'sun' (Shaver 1996: 314) in their family-wide cognate set. Not only is the form widely diverging in form from its translational equivalents in other Arawakan languages, but clear evidence exists on its secondary status. Baer (1994: 100), in a discussion of the cosmology and religion of the Matsigenka (another Kampan group), notes that some of the *Tasorintsi*, the Kampan name for (originally) anthropomorphous supernatural agents or entities, are addressed with the use of the term for 'father', which in Matsigenka he transcribes as *Pa'vaa*, a clear cognate of Nomatsigenga *paba*, which is also attested with the meaning 'father' in the latter (Shaver 1996: 291). Among these *Tasorintsi* so named, one finds not only *Ka'shiri*, the Moon, but also *Porea'tsiri*, the Sun. It is thus likely that Nomatsigenga *paba* 'sun' continues an earlier Kampan term for 'father', used in reference to the sun in these specific contexts.

Lokono *hadali ~ adali* 'sun' (Patte 2011: 311) is certainly isolated, lacking semantically matched cognates in other languages of the family (see Taylor 1958: 154 on this). However, Lokono *(h)adali* is plausibly analyzable as including the masculine suffix *-li*, leaving a root *ada-* which has cognates elsewhere, notably in Island Carib *h-ára* 'hot, warm' (see, again, Taylor 1959: 134 on this point).

In Payne's (1991) comparative work on the Arawakan language family, two etyma for the meaning 'sun' are reconstructed: **kamui* 'Sun (summer)' and **keʧɨ* 'Sun2 (day)' (Payne 1991: 420). Reflexes of these separate etyma include Palikur *kamuw* 'sun' (Launey 2003: 237) and Mojeño Ignaciano *sáʧe* 'sun' (Ott & Ott 1983: 625), respectively. Some languages, in fact, show reflexes of both etyma, as in Piapoco *èeri* 'sun' and *kamuí* 'summer' (Klumpp 1995: 153, 157). The correct decision, then, is to recognize for the meaning 'sun' at least two family-wide sets, in turn reconstructable to two semantically close yet independent etyma. Coupled with the known innovation of branch/language-specific forms for 'sun' in different Arawakan languages, WR's decision to have a single cognate set for this meaning is entirely unprincipled and lacking in empirical support.

Last but not least, note that WR claim to have 'relied heavily' on Payne's (1991) comparative study of Arawakan, is an attempt to ground their cognation decisions in what stands, to this day, the sole study devoted to an application of the comparative method to the Arawakan family at large. It is surprising, therefore, to learn that they have chosen (incorrectly) *to drastically change* Payne's (1991) work by merging his two sets, **kamui* 'Sun (summer)' and **keʧɨ* 'Sun2 (day)', into a single etymology. Further examples of similar missteps on the part of WR will be examined below.

Man

In WR's matrix, the Apurinã form for 'man', *kɨkɨ* (Facundes 2000: 651), is judged *not cognate* with the forms found in its two closest relatives within the Purus branch (Payne 1991) of the Arawakan family: Iñapari and Yine. Discussion of the two binary comparisons (Apurinã-

---

[7] See that Proto-Purus **atukatsi* 'sun' is likely a partial cognate of non-Purus forms such as Terena *káʃe* 'sun, day'. What is claimed here, tentatively, is that the formation **atu-katsi*, where **atu-* is a morph of unclear meaning, is an exclusive innovation of the Purus branch.

Iñapari and Apurinã-Yine) will be presented in separate below, as they instantiate distinct problems with the WR approach to the evaluation of similarities between comparanda. However, before discussing the reasons for treating Apurinã *kɨkɨ* 'man' and its translational equivalents in the other Purus languages as cognate elements, see that this cognation was recognized in Facundes (2000:662), which also happens to be WR main source on Apurinã. It is legitimate to infer then, that in this case as well, WR have chosen to either ignore or directly contradict a statement found in the available comparative Arawakan literature.

First, despite their formal differences, cognation between Apurinã *kɨkɨ* 'man' and Iñapari *ehɨ* 'man' (see Parker 1995: 46) can be established based on an understanding of the formal relations between cognates in these two languages and, in particular, of regular segmental correspondences. Synchronically, Iñapari is unusual, both areally and genetically, in lacking the velar stop *k* (Parker 1999: 2,7). This gap is explained historically by the recognition of specific developments of the velar stop **k*, reconstructed for Proto-Purus Arawakan, the shared ancestor of Apurinã, Yine and Iñapari: In the latter language it has the reflexes *h* preceding *ɨ*, *i* and *e* in any position, but ∅ preceding other vowels initially and *ʔ* preceding other vowels in medial position. Examples of these are given below:

	Man	Nose	To fear
Proto-Purus	**-kekɨ*	**-kiri*	**-pika*
Yine	*çeçi*	*hi-çri*	*-pika*
Apurinã	*kɨkɨ*	*-kiri*	*-pĩka*
Iñapari	*ehɨ*	*hirí-ti*	*i-piʔá-ma*

Table 2. Reflexes of Proto-Purus **k*

In view of the stated correspondences, Iñapari *ehɨ* 'man' may seem exceptional, as an initial *h* would be expected as a reflex of **k* preceding *e*. Here, familiarity with the data available on the language furnishes the required explanation: Loss of word-initial *h* in *ehɨ* 'man' seems to be a recent and isolated development, as earlier documents on the language, discussed in Valenzuela (1991: 214), record <*hehɨ*> 'hombre', with the expected initial *h*.

As for the diverging vocalism in the comparison of Apurinã *kɨkɨ* 'man' and Iñapari *ehɨ* 'man' (Apurinã *e* : Iñapari *ɨ* in the first syllable), recognition of the cognation is again backed up by parallels. In Apurinã, the regressive assimilation **e > ɨ /_Cɨ*, is attested not only in the form for 'man', but also in the forms for 'snake' and 'tongue', as shown below, with the etymological *e* retained in Iñapari and, for 'man', in Yine as well:

	TONGUE	MAN	SNAKE
Proto-Purus	**-nenɨ*	**kekɨ*	**-himenɨ*
Yine	*hi-nni*	*çeçi*	*himni*
Apurinã	*-nɨnɨ*	*kɨkɨ*	*imɨnɨ*
Iñapari	*-není-pa-ti*	*ehɨ*	*himení*

Table 3. Evidence for **e > ɨ /_Cɨ* in Apurinã

Note that the reflexes of **-nenɨ* 'tongue' show syncope in Yine and regular **ɨ > i* in Iñapari. This further supports the hypothesis that the vocalism of the proto-form is **e-ɨ*, and not **ɨ-ɨ* as attested for Apurinã, since, if the Apurinã vocalism was inherited, Iñapari **ɨ > i* would not

have applied[8]. Iñapari *-pa* is probably an independent morpheme (cf. *-není-ʧa-ti* 'palate'; Parker 1995: 56).

Now for the Apurinã-Yine equation. Apurinã *kɨkɨ* 'man' was compared by WR to Yine *jinerɨ*, both with the same meaning 'man' (see Nies 1986: 321). It is reassuring, one might say, to know that WR reject a cognation hypothesis in this case. Yine *jine-rɨ* contains a base for 'person' (actually, the same as in the name 'Yine') and a Masculine suffix *-rɨ* (see e.g. Hanson 2010: 107–114). Yine does have, however, a cognate of Apurinã *kɨkɨ* 'man', one that was already introduced in Table 2 above: *çeçi* 'man' (Nies 1986: 470). For Yine, Proto-Purus **k > ç* before **ɨ*, **i* and **e*. Though in this particular case the cognation judgment in WR was not incorrect, Yine and Apurinã were incorrectly coded as lacking cognates for the meaning 'man'. A final observation on this set underscores the generalized lack of pattern in WR's approach to sources and to how these have been harvested for comparative data. A direct look at their comparative database reveals, as noted above, that instead of using *çeçi* for Yine 'man', they have opted for using *jine-rɨ* instead. In the case of 'woman', however, WR selected the simple root *sɨco* 'woman' for comparison, instead of the feminine derivative of *jine-*, *jine-ro* (see Nies 1986: 506). The motivations behind this distinct treatment are not stated either in WR's paper or in their supplementary materials.

Heart

Apurinã and Iñapari are judged as having cognates for this meaning, but Yine is given as showing no cognate with these forms. This set is interesting because the Apurinã and Iñapari comparanda are arguably *less similar* formally than was the case in the set for 'man' examined above; and yet, cognation is recognized. Also relevant is the fact that some of the same regular correspondences discussed above are relevant for recognizing cognation in this case as well. The Yine form is also cognate, *contra* WR, as argued below.

Apurinã *hãkɨpa* 'heart' (Facundes 2000: 649) and Iñapari *ahɨpati* 'heart' (Parker 1995: 37) are correctly identified as cognate by WR. See that in this case the final syllable *-ti* of the Iñapari comparandum, in fact the absolute suffix *-ti*, though lacking a match in the Apurinã form, was not detrimental to their evaluation of the forms as cognate. Elsewhere, however, WR seems to have failed to factor out the presence of this marker, including it in comparisons and apparently grounding incorrect cognation judgments on its presence (see 'head' below). This starkly unsystematic character of WR's cognate decisions will become more obvious as the discussion of other sets is presented, and will be discussed further in section 4.

Another recurring problem of the WR treatment of the comparative data, a cavalier approach to sources, is found in the Yine comparandum for this set. In the WR dataset, the Yine (Piro) form is given as *rahi* 'heart'. Note, however, that there are two errors here, one relating to the way Yine orthography has been re-transcribed or adapted, and the other involving morphological under-analysis of the compared form. First, although Nies (1986: 403) gives <raji> for 'heart', Yine orthographic <j> does not stand for the glottal fricative *h*, as in most Spanish-based orthographies, but to a palatal fricative *ç* instead[9]. Second, *raçi* (<rahi>) is actu-

---

[8] Iñapari shows a context-specific development **ɨ > i* after the coronal consonants **r*, **n* and **t*, unless the vowel of the preceding syllable was also **ɨ* (cf. **hunɨ > huní* 'water', **kajatɨ > ajáʧi* 'paca' (rodent sp.), but **kasɨrɨ > aɨrɨ* 'moon', **-tɨrɨ > -tɨrɨ* 'son'). See Carvalho (Forthcoming) for more details on Proto-Purus Arawakan.

[9] Since Yine *ç* and *h* participate in distinct correspondence sets with the other Purus Arawakan languages, the apparently minor confusion involving these consonants could have more significant implications for cognation decisions, in particular if WR had concerned themselves with the identification of sound correspondences (or if

ally a third person singular possessive form, with the prefix *r-* (cf. the absolute/non-possessed form *haçᵽi* 'heart'; Nies (1986:403)).

If we recognize in Apurinã *hãkɨpa* 'heart' and Iñapari *ahɨpa-* the same morpheme *-pa* noted in the preceding section when discussing the Iñapari form for 'tongue' (cf. Iñapari *není-pa-ti* 'tongue', but *není-tʃa-ti* 'palate'), we end up with a compelling etymology matching Yine *-haçi*, Apurinã *hãkɨ-* and Iñapari *ahɨ-*, all with the meaning 'heart'. Correction of the orthographic confusion noted above in the WR Yine form turns out to be vital here for a proper account of the relations, since the Yine palatal fricative *ç* as discussed above in relation to 'man', is a reflex of **k* in the context of a following *ɨ*. Given that this **k* reconstructable to Proto-Purus yields Apurinã *k* in all contexts and Iñapari *h* medially, there is little if any impediment to the conclusion that the Purus Arawakan forms for 'heart' are, *contra* WR, in fact part of the same cognate set[10].

Woman

For this gloss, the main problems identified in the cognation judgments would have been avoided by a more careful, judicious treatment of sources, and, again, by consideration of regular patterns of correspondence, coupled with the recognition of the derived (non-etymological) character of certain formal mismatches.

The cognation between the Baure noun *eton* 'woman' (see e.g. Danielsen 2007: 28) and its translational equivalents in Ignaciano, *esena* (Ott & Ott 1983: 192), and Trinitario *ʔseno* (Gill 1970: 52) both with the meaning 'woman', is not recognized by WR. Note that cognation between the Baure and Mojeño forms for 'woman' becomes more obvious by the use of the earliest attestations of this noun, and by considering dialectal information. As noted by Danielsen (2007: 52), a source of Baure data used by WR, a more conservative form *eteno* 'woman' is attested and/or inferable based on the testimony of the more conservative dialects of the language and from earlier documentation.

The formal mismatches between the recovered 'Pre-Baure' form *eteno* 'woman' and the comparanda in the Mojeño dialects, Trinitario *ʔseno* 'woman' and Ignaciano *esena* 'woman', could be a complicating factor for recognizing cognation, in particular if a purely impressionistic assessment of similarities is employed. Nevertheless, both the anlaut correspondence between an initial cluster in Trinitario (more properly, a zero) and a vowel in Baure and in Ignaciano, and the vowel correspondence Ignaciano *a* : Baure/Trinitario *o*, can be related to the operation of regular sound change. In the comparison between the two Mojeño dialects, the word-initial consonant cluster of the form *ʔseno* 'woman' of the Trinitario variety is derived. A simple inspection of Ignaciano and Trinitario cognates where a trisyllabic form appears in Ignaciano show the regular character of the loss of the initial vowel in the Trinitario cognate (see see table 4; Carvalho & Rose 2018 for details).

Given the non-predictable nature of the Ignaciano vowel that lacks a correspondent in the Trinitario cognates, these are clearly inherited and, hence, Trinitario *ʔseno* 'woman' actually derives from **eseno* 'woman' at the Proto-Mojeño level. Finally, concerning the final *a* in Ignaciano, this is also secondary. Ignaciano shows *a* corresponding to both *o* and *a* in Trinitario and, moreover, Trinitario features a contrast between two back rounded vowels *u* and *o* that is simply lacking in Ignaciano (see table 5).

---

other, interested researchers, were to do so on the basis of the comparative data amassed by WR. In fact, Carling *et al*. (2012: 38) used WR's data set in their own study of Arawakan phylogenetic differentiation.

[10] I am grateful to an anonymous reviewer for noting that this cognation relation was previously noticed by Facundes (2000:661), thus showing that here as well WR could have relied on past published work for their cognation judgments.

Gloss	Trinitario	Ignaciano
Person	*ʔtʃane*	*aˈtʃane*
Woman	*ʔseno*	*eˈsena*
Ant	*ˈktʃiru*	*kaˈtʃiru*
Jaguar	*ʔtʃini*	*iˈtʃini*
Peccary	*ˈsmoru*	*siˈmaru*
Toucan	*ˈhnore*	*haˈnare*
Smoke	*ˈçhore*	*kiˈhare*
North	*ˈkhoʔo*	*kaˈhaʔa*

Table 4

	Ignaciano	Trinitario
Fog	*ijaru*	*ʔjoru*
Cloud	*uka*	*uko*
Earth, mud	*mate*	*mote*
Grandmother	*-atse*	*-otse*
Wife	*-jena*	*-jeno*
Shoulder, arm	*-pawa*	*-powo*
Tooth	*-aʔe*	*-oʔe*
Sky	*anu-ma*	*anu-mo*
Stone, stony floor	*mari*	*mari*
Sun	*satʃe*	*satʃe*
Person	*atʃane*	*ʔtʃane*
Son-in-law	*tʃina*	*tʃina*
Name	*-iha*	*-iha*
Hear	*-sama*	*-samo*

Table 5

The best explanation is that Ignaciano underwent a merger of **o* and **a* as *a*. This shows that Ignaciano *esena* 'woman' likewise reflects **eseno* 'woman'. The conclusion is that the proper comparison should be one between Baure *eteno* 'woman', available in the very same source consulted by WR, and a Common or Proto-Mojeño form **eseno* 'woman', a comparison more conducive to the recognition of the cognation relation missed by WR.

Bird

This is one of the most striking cases of false positives in cognate identification in WR's data, since even in the hypothetical absence of any knowledge of the languages compared, a simple *reading* of existing literature on the Arawakan language family would have prevented this mis-taken cognation judgment.

The forms for 'Bird' in Garífuna, *dunuru* (see Sabio & Ordóñez 2006: 75), and in Island Carib *túnulu* (Taylor 1956: 401) were judged to have cognates in other Arawakan languages, for instance, in Achagua *míʃidu* 'bird' (see e.g. Meléndez Lozano 2011: 40). In this case, the similar word-final syllables were apparent enough for these comparanda to be recognized as

cognates, while the obviously different four initial segments were deemed inconsequential. It is well-known, however, that these Garífuna and Island Carib forms are **loans from Cariban languages** (see Taylor 1956: 401), where, for instance, Kari'ña/Carib has *tonoro* 'bird' (Courtz 2008: 389) and Proto-Taranoan has **tonoro/*torono* (Meira 2000: 140). In fact, the misidentification of *dunuru/tunulu* 'bird' as an inherited Arawakan element was a feature of some of the earliest attempts at a comparative understanding of the Arawakan languages, and was noticed by Douglas Taylor (1966: 304, fn. 3) in his insightful review of Noble's *Proto-Arawakan*.

Tree (and Fire)

For the gloss 'Tree', Terena and Apurinã forms appear in the same cognate set, while Ignaciano and Trinitario appear in a distinct one. As shown below, this is misleading both as a statement of the relations between the forms compared and as evidence on the relations between the languages under comparison. A particularly odd feature of WR's cognation judgments appears in a rather transparent manner in these particular cases: their surprising decision to judge semantically and formally **identical** forms as non-cognate, while judging formally very divergent forms as cognate.

First, it is far from clear how Terena *tikóti* 'tree' (not *tikôti*, as in WR's data) could be judged cognate with Apurinã *ããmɨna* 'tree' (see Facundes 2000: 656). Although a detailed account of the relevant diachronic developments is currently unavailable, the Apurinã form fits quite clearly in the etymology proposed by Payne (1991: 423) as reflexes of his PA **anda[mɨ][na]* 'tree'. Not even Payne (1991), however, despite his often over-relaxed standards for regular correspondences, could fit Terena *tikóti* in this same set. The Terena noun *tikóti* 'tree' is of unclear etymology, lacking obvious parallels in any other Arawakan language I am aware of[11].

WR correctly identify Ignaciano *jukuki* 'tree' (see Ott & Ott 1983: 483, 599) and Trinitario *jkugi* (see Gill 1970: 5) as cognates. Again correctly, the semantically equivalent form *tikóti* 'tree' of Terena is judged not cognate. There are, however, two observations on the formal and the semantic aspects of the Mojeño (Ignaciano and Trinitario) forms. Formally, these reflect a Proto-Mojeño etymon **jukuki* 'tree' (see Table 6) which is, in turn, morphologically analyzable as **juku-ki*, **-ki* being a form-based classifier for nouns with cylindrical and rigid forms such as branches, arms and trees (see Olza Zubiri *et al.* 2002: 288–290). This morphological structure is still transparent both in Ignaciano and Trinitario. Semantically, Proto-Mojeño **juku-ki* means not only 'tree', but also 'wood' and, crucially, 'firewood' as well (see Ott & Ott 1983: 460 for Ignaciano; Marbán 1701: 265 for Old Mojeño). The meaning 'firewood' associated with **juku-ki* bears an obvious relation to the meaning of the base of this derivative form, as witnessed by the base meaning of Proto-Mojeño **juku* 'fire', with formally and semantically identical reflexes in all Mojeño dialects (see Ott & Ott 1983: 459 for Ignaciano; Gill 1970: 19 for Trinitario, and Marbán 1701: 239 for Old Mojeño).

Given the existence, in Terena, of a form *júku* meaning both 'fire' and 'firewood', it cannot be denied that an etymological relation exists between Terena *júku* and the Mojeño form for 'tree, wood, firewood'. This fact should be somehow expressed in the WR cognation judgments for the gloss 'tree', in particular given the fact that in preparing their comparative sets WR were not limited to lexicostatistical comparison with exact semantic identity (as noted before in section **2**). The data discussed so far are summarized in Table 6 below:

---

[11] A non-obvious parallel that may hold the key for the etymology of *tikóti* 'tree' is the Old Mojeño form <ticooiray> 'fruit-bearing tree' ('árbol fructífero'; Marbán 1701: 143). Identification of the Nominalizing suffix *-ray* (see Olza Zubiri *et al.* 2002: 674) and of the Mojeño third person prefix *ti-* leaves a stem *-cooi-*, likely derived from the Attributive prefix reconstructed as **ko-* for Proto-Mojeño (Carvalho & Rose 2018) and the noun **-oʔi* 'fruit'.

	Forms			Source
Terena	*tikóti* 'tree'	*júku* 'firewood'	*júku* 'fire'	Author field data
PM	**juku-ki*	**juku-ki*	**juku*	Carvalho & Rose (2018)
Ignaciano	*juku-ki*	*juku-ki*	*juku*	Ott & Ott (1983: 161)
Trinitario	*jku-çi*	*jku-çi*	*juku*	Gill (1970: 5)
Old Mojeño	<yucuqui>	<yucuqui>	<yucu>	Marbán (1701: 143, 239)

Table 6. Forms for 'tree' and 'firewood' in Terena and Mojeño

Since we have now reached the meaning 'fire', associated, either via morphological derivation or as a matter of lexical polysemy with other notions such as 'firewood' and 'wood' in Terena and in Mojeño, it would be interesting to assess WR's cognation judgments for 'fire' as well, an independent datapoint in the comparative database. Here, WR cognate decisions also look far from reasonable: despite having semantically and formally identical forms for the meaning 'fire' (Table 6), Terena and the Mojeño dialects are *coded as having no cognate element for this meaning*. This perplexing result is, nevertheless, paralleled by other cases where WR have coded semantically and formally identical forms in different languages as not cognate, and add to the impression that no systematic principle or rule has been consistently followed in the creation of the matrix of cognation judgments used as an input to their phylogenetic inference tool.

Water

The problem with this etymology lies seemingly in an overstrict criterion for cognate identification. The Terena form for 'water', *úne*, is coded as having a cognate only in Kinikinau, a closely related co-dialect (see Carvalho 2016b).

As it happens, obvious cognates of this form are found almost everywhere in the Arawakan language family. Ironically, this is a much more representative, pan-Arawakan cognate set than is, say, 'sun', judged wrongly by WR to have a single cognate throughout the family. Exemplar forms are given below for a representative set of languages:

Language	Form	Source
Palikur	*un*	Launey (2003: 237)
Lokono	*oni*	Patte (2011: 266)[12]
Wapixana	*wini*	WLP (2000: 153)
Piapoco	*uni*	Klumpp (1995: 108)
Terena	*úne*	Author field data
Resígaro	*hooní*	Allin (1979: 460)
Paresi	*one*	Rowan (2008: 64)
Mojeño Ignaciano	*une*	Ott & Ott (1983: 429)
Waujá	*uni*	Postigo (2014: 237)

Table 7. Cognates for 'water' in a set of Arawakan languages

---

[12] Lokono *oni* actually means 'eau de pluie, pluie', while *oniabo*, with the possessed stem *-oni-a*, means 'eau, eau potable, eau de pluie, bain, océan' (Patte 2011: 176–177).

The forms above, all semantically equivalent and formally nearly identical, would be judged cognate even on an impressionistic basis. It is far from clear on which principles can the cognation between Terena *úne* 'water' and the other forms in Table 7 be denied, and this raises the question of how this fact was missed in WR cognation judgments. Not surprisingly, all of these forms were judged cognate by Payne (1991: 425), where a single etymology for this meaning is presented and Proto-Arawakan **uni* 'water' is reconstructed.

Eat

The shortcomings in WR's cognation judgments for this set include two cases of false negatives, that is, statements that cognate forms cannot be found in the relevant language, while, in fact, they can. These incorrect judgments seem to stem from a superficial knowledge of the languages' morphology, and from a cavalier treatment of primary sources.

A form like **-niko* (or **-ni-ko*) is the most general Arawakan verb stem for the meaning 'to eat', and Payne (1991: 402) reconstructs **-nika* 'eat' for Proto-Arawakan. Apurinã *-ɲika* 'to eat' (Facundes 2000: 648) and Iñapari *iniʔama* 'he is eating' (Parker 1995: 36) are part of this set; and yet, WR judge these forms not cognate. Note that the Iñapari form is not entirely comparable, as it includes the third person, Masculine 'Subject' prefix *i-*, as well as a verbal suffix *-ma*. With the root *-niʔa-* 'to eat' excised, its cognation to Apurinã *-nika* is obvious. Moreover, as noted above, **k* was systematically eliminated in Iñapari, having a reflex *ʔ* in intervocalic contexts. Since the WR data for Iñapari simply give *iniʔama* for the meaning 'eat' (as it appears in Parker 1995: 36), without indication of morpheme boundaries, I presume that the partial cognation relation was missed, and that the form as a whole was, on the basis of an impressionistic inspection, judged not cognate with Apurinã *-ɲika* 'to eat'.

In relation to the other member of the Purus branch of Arawakan, WR code Yine as lacking a cognate of either the Apurinã or the Iñapari form. Inspection of the Yine data used by WR shows that the chosen comparandum for this meaning in Yine was *natʃnewlɨ*, which is in fact attested in Nies (1986: 398) for the meaning 'comer' ('to eat'). However, under the same entry in the same dictionary one finds *niklɨ* (given as an 'infinitive') as well as *nika* 'he/she eats' ('come', in Nies 1986: 398). The fact that Yine *-nika* is a cognate of Apurinã *-ɲika* and of Iñapari *-niʔa* is hardly worthy of discussion, and it is surprising that this Yine cognate form was missed by WR despite being present in the same dictionary entry from which the non-cognate comparandum was chosen.

Ear

Terena *kêɲo* (not *kênjo*, as in the authors' data) is judged a cognate of Trinitario *tʃoka* 'ear' (Gill 1970: 29) and Ignaciano *tʃaka* 'ear' (Ott & Ott 1983: 597). Despite the semantic identity, there is no way to justify this assertion on formal grounds. Nevertheless, Terena *-kêɲo* 'ear' *does have* cognates in the Mojeño dialects, namely: Ignaciano *-kiɲa* 'ear hole' (Ott & Ott 1983: 596) and Trinitario: *-çiɲo* 'ear hole' (Gill 1970: 28).

What is remarkable about this set is that although the relations between these languages are rightly identified — Terena and the Mojeño dialects do share cognate elements for the required meaning — **this is not borne out by the forms compared by WR.** Here, as in many other analyzed cases, one gets the impression that the comparison of the selected forms and the production of the classification were dissociated from each other.

Head

The nouns for 'head' in the Purus languages (Apurinã, Yine and Iñapari) are correctly identified by WR as cognates. However, their translational equivalents in Terena and in the Mojeño dialects (Ignaciano and Trinitario) are also marked as cognate, which is incorrect. As shown below, the grounds for this misidentification seem to lie in lack of adequate morphological analysis.

The forms appearing in the authors' data are Apurinã *kɨwɨ* 'head' (see e.g. Facundes 2000: 649), Iñapari *hɨwɨti* 'head' (Parker 1995: 32), Yine *çiwʧi* (Nies 1986: 383), Terena *tʰuti* 'head' and, for both Ignaciano and Trinitario, *ʧuti* 'head' [13] (see Ott & Ott 1983: 497; Gill 1970: 8). Despite the semantic identity, the formal grounds for recognizing cognation between the first three forms and the last two seems to rest entirely on the occurrence of a medial rounded vowel *u* (in Terena and the Mojeño dialects) perhaps matching *w* in the three Purus languages, and on the final syllable with a coronal consonant followed by *i* in all languages except Apurinã. Word-final *-ti* in Iñapari and *-ʧi* in Yine are, in fact, nominal suffixes marking these (inalienable) nouns as non-possessed or absolutes (see Payne 1991), the remaining roots, Iñapari *-hɨwɨ* and Yine *-çiw*, being cognates of Apurinã *-kɨwɨ*, all derivable, in turn, from Proto-Purus **-kɨwɨ* (which, in turn, presumably reflect Proto-Arawakan **kɨwɨ* 'head'; Payne 1991: 407). In the Mojeño and Terena forms, word-final *-ti* cannot, on the basis of any independent evidence internal to these languages, be recognized as a discernible morpheme, and, therefore, both Terena *-tûti* 'head' and the Mojeño form *-ʧuti* 'head' are single unanalyzable roots. WR correctly identify the latter two as cognate. Even granting that the relation between internal and external etymologization is a complex one — and that, therefore, the final syllable *-ti* of the Mojeño and Terena forms might be shown to derive in fact from an absorbed absolute suffix — this would leave us with a partial cognation at best, since it would be difficult to relate the root-initial syllable *tu-/ʧu-* to Proto-Purus **-kɨwɨ* 'head'.

Fish

The existence of two widespread cognate sets for 'fish' in the Arawakan family has been known for some time (Oliver 1989: 152–153; Payne 1991: 404) and these seem to correspond to the first two lines in the WR matrix for this meaning slot. Payne (1991: 404) assigns one of these to a Proto-Arawakan etymon **kopaki* and the other to **hima*. Investigation of the etymologies advanced by WR for this meaning slot is informative and relevant for two reasons. First, here as well one observes the curious pattern whereby the relations between languages are correctly recognized, despite there being no support for this on an examination of the actual comparanda found in WR's dataset. Second, as in the case of 'sun', it shows how WR have produced novel cognation judgments by making modifications in the judgments present in Payne (1991), a fundamental source on comparative Arawakan and one which WR claim to have relied upon extensively for their cognation claims.

WR code Terena *hôe* 'fish' as a cognate of Ignaciano *hima* (Ott & Ott 1983: 605) and Trinitario *himo* (Gill 1970: 31) both meaning 'fish' as well. In this WR diverge from Payne (1991: 404), for whom Terena *hôe* is instead a reflex of PA **kopaki*. Sharing only their meanings and an initial consonant, WR's proposal is a very weak etymology. In fact, closer attention to these languages show the equation *hôe : hima, himo* to be spurious. A bilabial nasal stop *m* in both Ignaciano and Trinitario corresponds regularly to *m* in Terena (Ignaciano data from Ott & Ott 1983; Trinitario data from Gill 1970):

---

[13] I will gloss over the fact that Terena *tʰuti* is an incorrect rendition of Terena *-tûti* 'head'. The source of the form with aspiration in the root initial consonant is unknown to me.

	Terena	Ignaciano	Trinitario
Husband	*-îma*	*-ima*	*-ima*
Tapir	*kámo*	*sama*	*samo*
To hear	*-kâmo*	*-sama*	*-samo*
To steal	*-oméʃo*	*-ámetʃa*	*-ómetʃo*

Table 8. Evidence for Terena *m* : Ignaciano *m* : Trinitario *m*

As far as I can tell, the Terena noun for 'fish' lacks cognates elsewhere in the family. That is, the evidence for grouping it under **kopaki*, as Payne (1991) did, also looks thin. The modern Terena form *hôe* 'fish' results from the debuccalization *ʃ > h* to which the fricatives of Early Terena (= Guaná) were regularly subject (see Carvalho 2017). Terena *hôe* (attested as *ʃoe* in late 19th century documents) 'fish' lacks an internal etymology and is, for this reason, suspect of being a loanword; Payne's (1991: 404) attempt at relating Terena *hôe* 'fish' to forms such as Paresi *kohátse* 'fish' and Achagua *kubái* 'fish' can be deemed unsuccessful. In particular, it postulates a diachronic loss of a labial stop, **p > ∅*, that lacks parallels elsewhere in the language, being therefore an *ad hoc* and unsupported proposal (on the isolated status of this correspondence, see Payne 1991:434).

4. Overall evaluation

Table 9 below presents, in synthetic format, the following information organized in rows: a relevant sample of the comparanda considered for each meaning gloss in the preceding sections, WR's decision as to whether the items in question are cognate or not, and a brief comment on the problems with these judgments.

Many of the critical comments on the etymologies proposed in the WR study are the result of recent or ongoing historical linguistic investigations. This fact could be seen as implying that the criticism offered here misses the point by being anachronistic. In the end, all that can be claimed is that WR's cognate judgments can be subjected to improvement and revision, like any etymological hypothesis. This conclusion, if correct, would trivialize the discussion done here in the preceding sections, as all WR could have done was to work with the best historical linguistic work available then.

These attenuating observations are, however, without force. First of all, a series of incorrect cognation judgments on the part of WR could be avoided by simple consideration of the literature existing at the time, that is, comparative studies on Arawakan languages that, in many cases, predate 2011 by a large margin. Thus, the misidentification of Garífuna *dunuru* 'bird', which is a loanword from a Cariban language, as an inherited Arawakan form with cognates in other languages could have been avoided by consultation of the vast literature on the history of Island Carib, such as Taylor (1956), where this exact same form is listed among the stratum of Cariban loanwords in the language. Second, in many cases, despite the authors' claimed reliance on early comparative work on Arawakan (notably Payne 1991, on which WR claim to have 'relied heavily'), the cognate decisions that were fed to WR's particular application of their phylogenetic inference tool stand, in many cases, in *direct contradiction* to established cognate judgments in this same specialized source. In the gloss for 'fish', for instance, while Payne (1991: 404) considered Terena *hôe* 'fish' to be a reflex of his Proto-Arawakan etymon **kopaki* 'fish', WR switched the Terena witness to the other etymology proposed by Payne (1991), advancing, therefore, the claim that *hôe* 'fish' has, among its cognates, forms such as

Ignaciano *hima* 'fish' and Yine *fima* 'fish'. This reveals that WR *did in fact* produce original cognation judgments and were even in a position of confidence to disagree with the existing historical linguistic literature on these matters. The same basic observation can be extended, as seen, for other sets, such as 'sun' and 'water', and in not a single case can WR's decision be supported by the available evidence.

Meaning	Comparanda	Decision in WR	Comment
'Sun'	Nomatsigenga *paba* Yine *tkatfi* Palikur *kamuɯ* Lokono *(h)adali*	Cognate	Incorrect. Individual comparanda can be shown to be either language/branch-specific innovations or to reflect separate Proto-Arawakan etyma.
'Man'	Apurinã: *kɨkɨ* Iñapari: *ehɨ*	Not cognate	Incorrect. Cognation supported by regular segmental correspondences.
'Man'	Apurinã: *kɨkɨ* Yine: *jinerɨ*	Not cognate	Decision is correct, but Yine has a semantically matched cognate of the Apurinã form: *çeçi* 'man'.
'Woman'	Baure *eton* Mojeño *ʔseno*	Not cognate	Incorrect. Earliest attested forms, such as Old Baure *eteno*, and reconstructed Proto-Mojeño **eseno*, show that the forms are in fact cognate.
'Head'	Iñapari *hiwíti* Yine *çiwtfi* Terena *tûti* Mojeño *ʧuti*	Cognate	While the Iñapari and Yine forms are in fact cognate, Iñapari *-ti* and Yine *-tfi* are not comparable to the final syllables of the Terena and Mojeño forms, which belong into a separate set
'Ear'	Terena *kêɲo* Mojeño *ʧoka*	Cognate	The forms are not cognate. However, a cognate of Terena *-kêɲo* does exist in Proto-Mojeño *-kiɲo* 'ear hole'
'Tree'	Terena *tikóti* Apurinã *ããmɨna*	Cognate	Incorrect, the forms cannot be shown to be cognate.
'Fire'	Terena *júku* Mojeño *juku*	Not cognate	Clearly incorrect. Even preliminary lexicostatistics or 'multilateral comparison' would identify such identical items as cognate.
'Water'	Terena *úne* Paresi *one*	Not cognate	Even 'preliminary lexicostatistics' would identify these items, semantically identical and formally nearly so, as cognate.
'Fish'	Mojeño *himo* Terena *hôe*	Cognate	No support from regular correspondences for this decision

Table 9

The third and, perhaps, most fundamental conclusion of this review is that the sample of WR's cognation judgments selected here for criticism reveals a total lack of adherence to the minimal standards of historical linguistics and, in particular, the comparative method; what is more, these judgments seem to have been produced in the absence of the most cursory knowledge of the structure of the languages compared. This finding is critically damaging to their enterprise since, as stressed by Eska & Ringe (2004: 574): "it is possible to prove that forms are cognates by showing that they exhibit multiple regular sound correspondences that recur significantly often throughout the comparative word list. The provability of cognation is one of the cornerstones of scientific historical linguistics" [14].

---

[14] I should note that the technical notion of *proof*, that is, a demonstrative, deductive inference mode of justifying conclusions, does not offer an adequate 'theory' of the reasonings employed in historical linguistics. As cor-

The formal aspects of the comparisons, usually the safest and strongest grounds for either establishing or rejecting etymological proposals, are treated with such a contempt in WR's cognation judgments that one gets the impression that no parameters, even if subjective and implicit, were being followed in the rating of similarities between forms. Their contrasting decisions when faced with an almost total lack of formal similarity in the comparanda and when faced with total identity in form are particularly telling: Thus, in view of WR's **asserted cognation** for comparanda such as Terena *tikóti* 'tree' and Apurinã *ããmɨna* 'tree', and the parallel **denied cognation** for comparanda such as Terena *júku* 'fire' and Ignaciano/Trinitario *júku* 'fire', one gets the impression that WR's matrix of cognation judgments was produced independently of the comparative database. This is all the more telling since in many of the examined cases the relations between the languages was captured correctly — that is, they do, in fact, share cognates for the relevant meaning — but this is not borne out by the particular forms selected by WR for comparison. This was seen in the case of the Yine form for 'eat', and also for the Terena and Mojeño comparison in the meaning 'ear', where the form chosen for Mojeño (reflexes of Proto-Mojeño *-ʧoka* 'ear' in the two Mojeño dialects), are *not* cognate with the Terena comparandum, despite the former having a perfect match for Terena *-kêɲo* 'ear', in the reflexes of Proto-Mojeño *-kiɲo* 'ear hole'.

Last but not least, it should be noted, for once, that it is of course illegitimate to conclude from the present study that WR's conclusion about the eastern Peruvian origin of the 'Arawakan diaspora' is incorrect; all that can be concluded is that WR have failed, contrary to their claims, to present adequate linguistic evidence supporting it. Other authors had early expressed some sympathy towards the hypothesis of a large Arawakan migration originating in southwestern Amazon near the eastern Andean foothills, though again with flawed reasoning. This is, in particular, the case with Urban's (1992: 95–97) conclusion that the original *Urheimat* of the Arawakan peoples should be sought in relatively high altitudes near headwaters, not along the major river courses of the Amazon (as suggested by Lathrap 1970). The main problem with his reasoning is its foundation on the distribution of the sample of languages that happen to have made their way into Payne's (1991) comparative Arawakan study. We know of the existence of a number of Arawakan-speaking groups that lived along major rivers, such as the Manao, the Wainumá, the Mariaté, the Marawá, the Bahuana, the Cariaí and many others along the courses of the Amazon, the Solimões and Negro rivers. That hardly any data survives on the languages of these groups (this being the reason why they have not been included in Payne`s 1991 study) is a consequence of their early extinction/assimilation and should not be seen as implying a lack of Arawakan presence in these regions — and yet, any attempt at identifying geographical centers of diversity that misses this crucial fact is doomed to be unsuccessful or, worse, highly misleading. One can conclude, therefore, that a proper assessment of the historical linguistic evidence on the history of migrations behind the great 'Arawakan diaspora' remains a task for the future.

5. Implications for South American historical linguistics — a plea for greater sobriety

 The historical and comparative linguistics of South American indigenous languages seems to be currently at an impasse. On the one hand, researchers on the languages of the continent seem eager to take part in the nascent 'computational phylogenetic revolution' in historical

---

rectly stressed by Starostin (2014: 3), probabilistic modes of inference offer a much more credible description of the relevant methodological processes.

linguistics, or, more generally, to apply any new methodology that promises to free the investigator from putative 'shackles' attributed to the traditional comparative method of historical linguistics. This novel agenda, along with its imperative, is described in the following terms by Carling *et al.* (2012: 29):

> "(…) linguistic research, aided by new developments in software technology and computer capacity, has moved toward large-scale comparison projects where new ways of measuring linguistic distances are supplying new evidence of genetic relationships as well as contact scenarios."

The same authors, in fact, quote WR's study as one of these "new developments". On the other hand, it is well-known that, as applied to the better understood and long-researched language groups of Eurasia, any innovative technique that has been proposed has been advanced as a *complement or add-on to the traditional methods of historical linguistics*. The problem for South America lies in the fact that, for the vast majority of its indigenous language families, there is little for these non-orthodox methods to complement, as the coverage of existing historical comparative work is uneven and, where investigations have been carried, results are quite limited[15]. As I hope to have shown in the preceding sections, this may force authors of studies that take part in these innovative developments to build veritable houses of cards, perhaps more aptly described as rootless or groundless (phylogenetic) trees.

This situation, in turn, seems to create a kind of dissonance, and a particularly harmful one, once researchers start to derive methodological recommendations or epistemological judgments from it. This dissonance can be exemplified with two citations from a single paper that introduces a reference volume on the languages of South America:

> "With respect to *language classification*, using the comparative method there is a current consensus for some 108 separate language families on the continent, half of which are isolates" (Muysken & O'Connor (2014: 1)).

> "Comparative-historical linguistics in the South American context faces a number of problems. There are few reconstructed proto-languages for comparison at the level of families, and the coverage and quality of documentation is uneven for many languages" (Muysken & O'Connor 2014: 4).

Setting aside the recurrent mistake of seeing the comparative method (= reconstruction of proto-languages) as a tool for establishing genetic kinship among languages (see e.g. Nichols 1996)[16], it is clear that an underlying tension, if not an outright contradiction, jumps out from a comparison of these quotes, three pages apart from one another: it is said that an understanding of language diversity in the continent has been arrived at by 'using the comparative method', but, at the same time, it is asserted that there are very few actual applications of the method ('reconstructed proto-languages') and that paucity of documentation hinders further progress. In trying to understand how this obviously awkward contradiction could have gone unnoticed, it is perhaps relevant to note that these claims appear as nothing but a sort of preparation for the denouncement of 'limitations in the traditional methods'. After an astonishingly brief and selective review of the history behind the recognition of the major genetic unities of South America, the authors go on and say that:

> "However, the overall picture is not one of unification, and this raises issues about the universal applicability of the comparative method" (Muysken & O'Connor 2014: 3).

---

[15] This is not the place to review in detail the state of the art in the historical linguistics of indigenous South America. On the reasons for the limited development of the field, both the limited descriptive coverage and the limited manpower (that is, few people doing historical linguistics) could be presented as reasonable causes.

[16] Elsewhere the authors also refer to yet another incorrect idea on the comparative method, viz.: that it is somehow associated with an intrinsic time depth limit; Muysken & O'Connor (2014: 2).

This is followed by an Edward Sapir quote, mentioning "networks" and "language contact" as possible hurdles for the comparative method. Of course, one is left wondering how "issues about the universal applicability of the comparative method" can be raised by research on the historical linguistics of indigenous South America, where "there are few reconstructed proto-languages", a point on which everybody seems to agree.

The practical implications of this situation are tangible, in fact, in the very study evaluated here. Thus, WR point out that the phylogenetic tree produced by their analysis, which includes posterior probabilities indicating the degree of support for the formation of each clade, is 'broadly consistent with expert classifications by linguists, at least near the tips' (pg. 3). These 'expert classifications', one can imagine, are certainly included in the body of studies that promoted the 'overall picture' of the historical linguistics of indigenous South America mentioned by the Muysken & O'Connor (2014) quote above. Careful consideration of the works WR have in mind — Mason (1950), Noble (1965), Loukotka (1968), Payne (1991), Aikhenvald (1999) and Ramirez (2001) — reveals that they constitute a very weak base for validation of classifications produced by non-traditional methods (and, arguably, for classifications of any kind). On John Alden Mason's classification of the South American native languages, Campbell (2012: 66) notes that 'Mason was in the tradition which sought to reduce the vast diversity among American Indian languages by proposing preliminary but undemonstrated hypotheses of more far-reaching families to be tested by later research'. That is, he provided more of a tentative programmatic proposal in need of testing than a definitive statement. Noble's (1965) work has serious limitations due to the paucity and low-quality of the data he worked on (Taylor & Hoff 1966: 303–304), and for the fact that he ignored a number of important studies, especially the series of papers authored by Taylor, mostly on the Caribbean Arawakan languages (see Payne 1991: 368). To make things worse, Noble (1965) assumed the Arawakan affiliation of many languages now known to have no relation to this family, and this assumption had consequences to some of his cognation judgments and his reconstructed forms (see Taylor & Hoff 1966: 306). As for Loukotka (1968), his work was hampered not only by the lack of reliable data, but his methods have always been the focus of strong criticism, as it consisted basically in the impressionistic visual scanning of an arbitrary list of forty-five 'typical words' (see e.g. Rowe 1954; Campbell 2012: 66).

Payne's (1991) study, whose goal consists in presenting a 'credible internal classification' for the Arawakan language family is certainly the best, family-wide comparative study so far — one on which, as we have seen, WR claim to have relied extensively. Payne's (1991) classification is, however, based on shared proportions of *retentions* from his set of 203 Proto-Arawakan etyma, and thus faces the same methodological problems of any classification not based on putative shared innovations. Aikhenvald's (1999: 75) internal classification is, on the author's own admission, based on an 'areal-geographic principle'. Finally, while Ramirez (2001) employs the comparative method, it does so, in any detail, only for his 'Japurá-Colombia' branch, which includes a number of northwestern Amazonian Arawakan languages. Moreover, Ramirez (2001) is a study that has to be approached with care as to its conclusions, since that are sometimes important problems with the data he presents, as well as with his analyses (see Michael 2009 and Carvalho 2018 for Resígaro in particular).

The perceptive reader, if aware of the overall outlines and recent developments of historical comparative linguistics in a world scale, will see, in the South American situation, the broad outlines of a scenario similar to that where the historical linguistics of Australian languages found itself until some time ago. Though for different reasons — in the case of Australia, an apparent 'diffusionist bias' exposed by certain influential researchers — similar announcements were made about the 'finding' that something out of the line with traditional

methods was at place, in tandem with claims that "innovative methods" are necessary. Needless to say, as soon as people started to get their feet wet and their hands dirty, Australian languages have shown themselves to be amenable to the comparative method (see Evans 2005 and Sutton & Koch 2008 for interesting reviews).

In view of these late pronouncements, it is perhaps wiser to fall back on some notable pre-21th century directions for research on the languages of the continent, such as what Kaufman (1990) had to say in his proposal about 'How to know more on language history in South America':

> 'The 'art' of diachronic linguistics is the method, in any case: the comparative method. This method is fairly well developed at this point, and what is really needed is to apply it rather than to figure out how it could be better, or worse yet, try to short-cut it' (Kaufman 1990: 14).

What was true in 1990 seems to remain true in 2020. It is an urgent task for researchers on the historical linguistics of South America to resort to the full (and largely untapped) potentialities of traditional methods, and to recognize the limitations of grandiose statements about the history of the continent that are advanced in the absence of anything like solid historical linguistic work applying the comparative method, as if conjured out of thin air. Quantitative, computational tools and their use in historical linguistics hold a great promise, and for good reason (see e.g. Verkerk 2017) but their full potential will not be realized in the absence of rigorous and adequately informed investigations on the historical linguistics of language groups.

References

Aikhenvald, Alexandra. 1999. Arawak languages. In: R.M.W. Dixon, Alexandra Aikhenvald (eds.) *Amazonian Languages*: 487–517. Cambridge University Press.

Aikhenvald, Alexandra. 2009. Arawak languages. In: Keith Brown, Sarah Ogilvie (eds.). *Concise Encyclopedia of Languages of the World*: 59–61. Oxford: Elsevier.

Allin, Trevor. 1979. *Vocabulario Resígaro*. Peru: Instituto Lingüístico de Verano (ILV).

Baer, Gerhard. 1994. *Cosmologia y Shamanismo de los Matsiguenga (Perú Oriental). Colección Biblioteca Abya-Yala 15*. Quito: Ediciones Abya-Yala.

Barbieri, Chiara et al. 2014. Between Andes and Amazon: The Genetic Profile of the Arawak-speaking Yanesha. *American Journal of Physical Anthropology* 155 (4): 600–609.

Campbell, Lyle. 2012. Classification of the indigenous languages of South America. In: Lyle Campbell, Verónica Grondona (eds.). *The Indigenous Languages of South America*: 59–166. Berlin: De Gruyter Mouton.

Carling, Gerd, Love Eriksen, Arthur Holmer, Joost van de Weijer. 2012. Contrasting linguistics and archaeology in the matrix model: GIS and cluster analysis of the Arawakan languages. In: Lars Borin, Anju Saxena (eds.). *Approaches to Measuring Linguistic Differences*: 29–56. Berlin/Boston: Mouton de Gruyter.

Carvalho, Fernando O. de. 2016a. The diachrony of person-number marking in the Lokono-Wayuunaiki subgroup of the Arawakan language family. *Language Sciences* 55: 1–15.

Carvalho, Fernando O. de. 2016b. Terena, Chané, Guaná and Kinikinau are One and the Same Language: Setting the Record Straight on Southern Arawak Linguistic Diversity. *LIAMES* 16 (1): 39–57.

Carvalho, Fernando O. de. 2017. Fricative debuccalization and primary split in Terena (Arawak) historical phonology. *International Journal of American Linguistics* 83 (3): 509–537.

Carvalho, Fernando O. de. 2018. Diachronic split and phoneme borrowing in Resígaro (Arawakan). *Canadian Journal of Linguistics* 63 (3): 339–358.

Carvalho, Fernando O. de. Forthc. A comparative reconstruction of Proto-Purus (Arawakan) phonology. *International Journal of American Linguistics*.

Carvalho, Fernando O. de, Françoise Rose. 2018. Comparative reconstruction of Proto-Mojeño and the diversification of Mojeño dialects. *LIAMES* 18 (1): 3–44.

Courtz, Henk. 2008. *A Carib Grammar and Dictionary*. Toronto: Magoria Books.

Danielsen, Swintha. 2007. *Baure: An Arawak Language of Bolivia. Indigenous Languages of Latin America (ILLA)*. Leiden: Research School CNWS.

Epps, Patience. 2015. The Dynamics of Linguistic Diversity: Language Contact and Language Maintenance in Amazonia. *Presentation at the Diversity Linguistics Conference, Max Planck Institute for Evolutionary Anthropology, May 1–3,   2015.*

Ericksen, Love, Swintha Danielsen. 2014. The Arawakan Matrix. In: L. O'Connor, Pieter Muysken (eds.). *The Native Languages of South America*: 152–176. Cambridge University Press.

Evans, Nicholas. 2005. Australian languages reconsidered: A review of Dixon (2002). *Oceanic Linguistics* 44 (1): 216–260.

Eska, Joseph, Donald Ringe. 2004. Recent Work in Computational Linguistic Phylogeny. *Language* 80 (3): 569–582.

Facundes, Sidney. 2000. *The language of the Apurinã people of Brazil*. Doctoral Dissertation: SUNY at Buffalo.

Friis Boegh, Kristoffer, Aymeric Daval-Markussen, Peter Bakker. 2016. A Phylogenetic analysis of stable structural features in West African languages. *Studies in African Linguistics* 45 (1/2): 61–94.

Galucio, Ana V., Sérgio Meira, Joshua Birchall, Denny Moore, Nilson Gabas Júnior, Sebastian Drude, Luciana Storto, Gessiane Picanço, Carmen R. Rodrigues. 2015. Genealogical Relations and Lexical Distances within the Tupian Linguistic Family. *Bol. Mus. Para. Emilio Goeldi. Ciênc. hum.* 10 (2): 229–274.

Geisler, Hans, Johann-Mattis List. 2010. Beautiful Trees on Unstable Ground: Notes on the data problem in lexicostatistics. In: Hettrich, H. (ed.). *Die Ausbereitung des Indogermanischen: Thesen aus Sprachwissenschaft, Archäologie und Genetik [The Diffusion of Indogermanic: Theses from Linguistics, Archaeology, and Genetics]*: 1–10. Wiesbaden: Reichert.

Gil, Wayne. 1970. *Diccionario Trinitario-Castellano, Castellano-Trinitario*. Ms. Misión Nuevas Tribus.

Hanson, Rebecca. 2010. *A Grammar of Yine (Piro)*. Doctoral Dissertation: LaTrobe University.

Heggarty, Paul. 2010. Beyond lexicostatistics: How to get more out of 'word list' comparisons. *Diachronica* 27 (2): 301–324.

Honkola, Terhi, O. Vesakoski, K. Korhonen, J. Lehtinen, K. Syrjänen, N. Wahlberg. 2013. Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* 26: 1244–1253.

Hornborg, Alf, Jonathan Hill. 2011. *Ethnicity in Ancient Amazonia: Reconstructing Past Identities from Archaeology, Linguistics and Ethnohistory*. University Press of Colorado.

Kaufman, Terrence. 1990. Language history in South America: What we know and how to know more. In: Doris L. Payne (ed.). *Amazonian Linguistics: Studies in Lowland South American Languages*: 13–74. Austin: University of Texas Press.

Klumpp, Deloris. 1995. *Vocabulario Piapoco-Español*. Bogotá: Instituto Lingüístico de Verano (ILV).

Lathrap, Donald. 1970. *The Upper Amazon*. Thames & Hudson.

Launey, Michel. 2003. *Awna Parikwaki: Introduction à la langue Palikur de Guyane et de l'Amapá*. Paris: IRD Éditions.

Loukotka, Čestmir. 1968. *Classification of South American Indian Languages*. Los Angeles: University of California.

Marbán, Pedro. 1701. *Arte de la Lengua Moxa, com su vocabulário, y cathecismo*. Lima: Imprenta Real de Joseph de Contreras.

Mason, John A. 1950. The language of South American Indians. *Handbook of South American Indians* 143: 157–317.

McMahon, April, Robert McMahon. 2005. *Language Classification by Numbers*. Oxford University Press.

Meira, Sérgio. 2000. *A reconstruction of Proto-Taranoan phonology and morphology*. München: LINCOM Europa.

Meléndez Lozano, Miguel Ángel. 2011. *Diccionario Achagua-Español, Español-Achagua*. Bogotá: Universidad de los Andes.

Michael, Lev. 2008. *Nanti evidential practice: Language, knowledge and social action in an Amazonian society*. Doctoral dissertation: University of Texas at Austin.

Michael, Lev. 2009. Review of Ramirez 2001. *International Journal of American Linguistics* 75 (3): 447–450.

Michael, Lev. 2020. Rethinking the communicative functions of evidentiality: Event responsibility in Nanti (Arawakan) evidential practice. *Cadernos de Etnolinguistica* 8 (1): 95–123.

Muysken, Pieter, Loretta O. Connor. 2014. Introduction: South American indigenous languages; genealogy, typology, contacts. In: Loretta O'Connor, Pieter Muysken (eds.). *The Native Languages of South America: Origins, Development, Typology*: 1–26. Cambridge University Press.

Nichols, Johanna, Tandy Warnow. 2008. Tutorial on Computational Linguistic Phylogeny. *Language and Linguistics Compass* 2 (5): 760–820.

Nies, Joyce. 1986. *Diccionario Piro*. Instituto Linguístico de Verano (ILV).

Noble, G. Kingsley. 1965. *Proto-Arawakan and its Descendants*. Bloomington: Indiana University Press.

Nogueira-Santamaría, Maria C., Carl Edlund Anderson, Daniel Uricoechea, Clemencia Durán, Ignacio Briceño-Balcázar, Jaime Bernal Vilegas. 2015. Mitochondrial DNA Analysis Suggests a Chibchan Migration into Colombia. *Universitas Scientiarum* 20 (2): 261–278.

Oliver, José. 1989. *The archaeological, linguistic and ethnohistorical evidence for the expansion of Arawakan into northwestern Venezuela and northeastern Colombia*. Doctoral Dissertation: University of Illinois at Urbana-Champaign.

Olza Zubiri, Jesus, Conchita Nuni de Chapi, Juan Tube. 2002. *Gramática Moja Ignaciana (Morfosintaxis)*. Caracas: Universidad Catolica Andrés Bello.

Ott, Willis, Rebecca de Ott. 1983. *Diccionario Ignaciano y Castellano con apuntes gramaticales*. Cochabamba: Instituto Lingüístico de Verano (ILV).

Parker, Steve. 1995. *Datos de la Lengua Iñapari*. Pucallpa: Instituto Lingüístico de Verano (ILV).

Parker, Steve. 1999. A sketch of Iñapari phonology. *International Journal of American Linguistics* 65 (1): 1–39.

Patte, Marie-France. 2011. *La Langue Arawak de Guyane*. Marseille: IRD Éditions.

Payne, David. 1985. The genetic classification of Resígaro. *International Journal of American Linguistics* 51: 222–231.

Payne, David. 1987. Some morphological elements of Maipuran Arawakan: Agreement affixes and the genitive construction. *Language Sciences* 9 (1): 57–75.

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

Pereltsvaig, Asya, Martin Lewis. 2015. *The Indo-European Controversy: Facts and Fallacies*. Cambridge University Press.

Pet, Willem. 2011. *A Grammar Sketch and Lexicon of Arawak (Lokono Dian). SIL E-books 30*. Summer Institute of Linguistics.

Postigo, Adriana. 2014. *Língua Wauja (Arawak)*. Doctoral Dissertation: UNESP.

Ramirez, Henri. 2001. *Línguas Arawak da Amazônia Setentrional*. Manaus: EDUA.

Ringe, Don, Tandy Warnow, Ann Taylor. 2002. Indo-European and Computational Cladistics. *Transactions of the Philological Society* 100 (1): 59–129.

Rowan, Orlando. 2008. *Dicionário Paresi-Português*. Cuiabá: Sociedade Internacional de Linguística.

Rowe, John Howland. 1954. Linguistic Classification Problems in South America. In: C. Chretien, M. Beeler, M. Emeneau, M. Haas (eds.) *Papers from the Symposium on American Indian Linguistics*: 13–19. University of California Press.

Sabio, Fernando, Celia Karina Ordóñez. 2006. *Hereru Wagüchagu*. La Ceiba: Honduras.

Scliar, Maria O. et al. 2014. Bayesian Inferences Suggest that Amazon Yunga Natives diverged from the Andeans less than 5000 yBP: Implications for South American prehistory. *BMC Evolutionary Biology* 14: 174.

Shaver, Harold. 1996. *Diccionario Nomatsiguenga-Castellano, Castellano-Nomatsiguenga*. Lima: Instituto Lingüístico de Verano (ILV).

Starostin, George. 2014. Macro-comparative linguistics in the 21st century: State of the art and perspectives. *Journal of Language Relationship* 11: 1–12.

Sutton, Peter, Harold Koch. 2008. Australian languages: A singular vision. *Journal of Linguistics* 44: 471–504.

Taylor, Douglas. 1956. Language contacts in the West Indies. *Word* 12 (3): 399–414.

Taylor, Douglas. 1958. The place of Island Carib within the Arawakan family. *International Journal of American Linguistics* 24 (2): 153–156.

Taylor, Douglas. 1959. Homophony or Polysemy? *International Journal of American Linguistics* 25 (2): 134–135.

Taylor, Douglas, Berend Hoff. 1966. Review of Proto-Arawakan and its Descendants, by G. Kingsley Noble. *International Journal of American Linguistics* 32 (2): 302–308.

Urban, Greg. 1992. A história da cultura brasileira segundo as línguas nativas. In: Manuela Carneiro da Cunha (ed.). *História dos Índios no Brasil*. São Paulo: Companhia das Letras.

Valenzuela, Pilar. 1991. Comprobación del lugar de la lengua Iñapari dentro de la rama preandina de la familia Arawak. *Revista Latinoamericana de Estudios Etnolingüísticos* VI: 209–239.

Verkerk, Annemarie. 2017. Phylogenies: Future, not fallacy. *Language Dynamics and Change* 7: 127–140.

Walker, Robert, Lincoln A. Ribeiro. 2011. Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings of the Royal Society B* 278: 1–6.

Watkins, Calvert. 1990. Etymologies, equations and comparanda. In: Philip Baldi (ed.). *Patterns of Change, Change of Patterns: Linguistic Change and Reconstruction Methodology*: 289–304. Berlin / New York: Mouton de Gruyter.

Фернанду де Карвальью. Критическая оценка вычислительной филогенетики аравакской языковой семьи на основании анализа предполагаемых когнаций

Целью настоящей статьи является критический анализ лингвистических данных в широко цитируемой работе Walker & Ribeiro (2011), содержащей автоматизированную филогенетическую классификацию языков, относящихся к аравакской семье. Учитывая, что представленные в ней выводы о внутренней структуре этой большой семьи напрямую зависят от принятых в работе этимологических решений, и что достаточно амбициозные выводы авторов относительно доисторических путей миграции аравакоязычных народов целиком основаны на полученной классификации, мы приходим к выводу, что полученные авторами результаты должны быть отвергнуты. Аналитические ошибки авторов включают как ложные негативы (незамеченные родственные связи между формами), так и ложные позитивы (неродственные формы, засчитанные как отражения общей прааравакской праформы). Общая методология расстановки когнационных индексов в целом остается неясной, а конкретные итоги настолько разительно отличаются даже от поверхностных представлений о фонетическом сходстве, что итоговая дистрибуция когнатов выглядит в целом независимой от реальных данных. Настоящая статья призывает к более трезвому анализу исторических данных по языкам коренного населения Южной Америки, основанному на прозрачном и подкрепленном качественными данными применении сравнительно-исторического метода, равно как и к отказу от беспочвенных заявлений относительно бесполезности применения этого метода к анализу языковой истории этого континента.

Ключевые слова: аравакские языки; когнация; этимология; вычислительная филогенетика.