Copyright © 2013 The Author. All Rights Reserved.

A new proposal of Western Tukanoan consonants and internal classification *

Amalia Skilton
Advisor: Claire Bowern

Contents

* An essay submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts in Linguistics, Yale University, Spring 2013

Abstract

The Tukanoan language family, one of the six major families of the Amazon Basin, has traditionally been classified as comprised of two branches, Eastern and Western. Although the Eastern Tukanoan languages and peoples have been subject to much linguistic and anthropological study, little research has taken place on Western Tukanoan (WT). Comparative work on WT is especially scarce, with only three papers (Mason 1950, Waltz and Wheeler 1972, and Chacón forthcoming) available on the history and internal classification of the group. Furthermore, the internal classifications of WT constructed in these papers do not correlate well with the geography of the languages or with speakers' judgments about their similarity and degree of mutual intelligibility.

In this essay, I employ phonological and morphological evidence from Eastern and Western Tukanoan languages to reevaluate the reconstructed phonemic inventories and internal classifications of WT advanced by Waltz and Wheeler (1972) and Chacón (forthcoming). I find that this evidence – drawn from dictionaries, grammars, wordlists, and my own fieldwork on Máíhĩki – is inconsistent with the conclusions of the previous research. I therefore propose a new reconstruction of the phonemic inventory of Proto-Western-Tukanoan and a new internal classification of the attested WT languages.

Acknowledgements

I am deeply grateful to the indigenous people, named and unnamed, who have shared their linguistic knowledge with researchers in the Amazon Basin over the last century. I owe a special debt to the Máíjùnà people who taught me some of their language in summer 2012, especially Alberto, Liberato, Jesusa, and Neida Morales Mogica; Luciano Tapullima Navarro; and Lizardo Gonzales Flores. *Ñúí chíbàyì!*

Conducting this research would have been impossible without the support of my advisor and mentor Claire Bowern. Claire deserves all of the credit (or depending on who you ask, the blame) for my decisions to major in linguistics and to do fieldwork. She has been a constant source of fresh ideas, constructive criticism, and excellent practical advice for me over the last two years. At Yale, this project has also benefited from conversations with Hannah Haynie, Raffaella Zanuttini, Maria Pinango, and the 2013 cohort of linguistics majors.

I also owe a great deal to Lev Michael, Christine Beier, and Stephanie Farmer in Berkeley. Lev and Chris especially welcomed me into the Máíhɨ̃ki Project team and taught me a remarkable quantity of linguistics in a short time. Over the last eighteen months, they and their student Stephanie have shown me greater kindness, patience, and generosity – both in the field and over the distance to New Haven – than anyone could have asked.

Finally, I cannot thank my family and friends enough for their companionship and encouragement throughout the preparation of this essay. Paul Skilton and Elizabeth Horan taught me language. Julia Calagiovanni, Caroline Mann, Gabe Murchison, Chamonix Adams Porter, and many others have put up with the results.

1 Introduction

1.1 The Tukanoan languages

Tukanoan is one of the six major language families of the Amazon Basin. Its twenty-five attested languages are spoken throughout northwestern Amazonia, with most clustering around the borders between Peru, Ecuador, Colombia, and Brazil. They cover a geographic area which ranges from the Vaupés-Japura river system in Brazil to the east, to the Napo River basin in Peru and Ecuador to the west. Yet this area, of some 90,625 square kilometers (Jackson 1983: 17), probably constitutes only a fraction of the territory where Tukanoan-speaking peoples lived before the arrival of Europeans in the region in the late seventeenth century. In the discussion which follows – as in most research on historical linguistics of the Americas – it is therefore essential to remember that the Tukanoan languages spoken today or attested in writing represent a subset of the diversity which existed before European contact. Languages do not fossilize, and it is inevitable that phylogenies will not show all the branches or taxa which once existed.

1.1.1 Tukanoan involvement in the Vaupés Linguistic Area (VLA)

Sixteen of the twenty-five Tukanoan languages are spoken primarily in the Vaupés Linguistic Area (VLA), a cultural and linguistic area which centers on the Vaupés River basin. In many respects, the language ecology of the Vaupes, which is located on the border between northwest Brazil and Colombia, resembles that of the Amazon Basin in general. The VLA displays high diversity of languages, small speaker populations for each language, and extensive language contact phenomena – all features common throughout Amazonia (Epps 2011: 8-9). Yet the

VLA differs from nearly all other linguistic areas, in the Amazon Basin and around the world, in that its languages have undergone extensive convergence in phonology and syntax despite near-zero rates of lexical borrowing.

This is the result of a social system which has fostered long-term intensive contact among unrelated languages from the Tukanoan, Arawak, Nadahup, and Kakua/Nukak families in the Vaupés (described by Sorensen 1967, Chernela 1989, and Jackson 1983, among many others). For at least seven centuries, the Tukanoan and Arawak peoples of the VLA have practiced linguistic exogamy within the region. Children acquire different languages from mother and father, belong to the father's ethnolinguistic group, and are taught not to mix languages (since language is a crucial marker of ethnic identity). The system of linguistic exogamy does not include the hunter-gatherer Nadahup and Kakua/Nukak peoples, whom the agriculturalist Tukanoan and Arawak groups regard as inferior (Epps 2008: 4).

1.1.2 Contact effects and reconstruction in Tukanoan

Despite the asymmetrical nature of the contact situation in the VLA, all of the area's languages have undergone grammatical convergence. VLA languages from all four families exhibit areal features such as contrastive tone, grammaticalized evidentiality, nasal assimilation or harmony, and extensive noun classifier systems (Epps 2011: 11, 15-16). The widespread presence of these features obscures their origins. For example, Aikhenvald and Dixon (1998: 251) claim that evidentiality in the VLA was innovated in a Tukanoan language and diffused into Arawak, Nadahup, and Kakua/Nukak languages in the region. Epps (2005: 640) disagrees, arguing that an evidential system existed in Proto-Nadahup, but was elaborated in the VLA language Hup due to contact with Tukanoan.

It is impossible to definitively answer such questions about the directionality and relative dates of contact-induced changes without information about the grammar of the proto-languages in contact (Thomason 2012). Reconstruction of Proto-Tukanoan (PT), whether of morphemes or of morphosyntactic categories, will therefore be essential to serious diachronic study of contact phenomena in the VLA. Yet comparative Tukanoan research has barely begun. There is no accepted reconstruction of the phonemic inventory of PT, and the internal classification of the family is at best unclear (Chacón forthcoming). Furthermore, while a handful of comparative articles on the Tukanoan languages of the VLA are available, only two published papers (Mason 1950 and the minimally different Waltz and Wheeler 1972) discuss the history of the Tukanoan languages spoken west of the Vaupés.

The language contact phenomena of the VLA complicate the reconstruction of Proto-Tukanoan, but do not prevent it completely. Many researchers – for instance, Aikhenvald (2002: 26) and Epps (2009: 589), writing on Amazonia, and Dixon (1980: 238) on Australia – have commented on the difficulty of determining which similarities between languages involved in a contact situation like that of the VLA arise from contact-induced diffusion, and which from genetic relationship. Where contact effects are so pervasive that rates of lexical borrowing in core vocabulary approach 50 percent, as in the Australian language Gurindji (McConvell 2009: 811), then it may become impossible to classify the languages involved on the basis of lexical cognates using the comparative method (Campbell and Poser 2008: 326). Yet such an extreme situation does not obtain anywhere in the Vaupés, where contact has caused extensive diffusion of grammatical categories and features, but very little lexical borrowing, between languages. Even Aikhenvald and Dixon (1998: 252), two prominent skeptics of the possibility of reconstructing genetic relationships between languages

subject to areal diffusion, have argued that the contact effects on Tukanoan languages in the VLA are not extensive enough to preclude reconstruction of Proto-Tukanoan phonemes and morphological categories. Patience Epps' work on the historical linguistics of the Nadahup family (Epps 2005, Epps 2009) further indicates that contact effects are also insufficient to prevent reconstruction of non-Tukanoan languages in the VLA.

1.2 Goals of this essay

This essay begins to fill the Tukanoan-shaped hole in Amazonian historical linguistics. In order to provide a starting point for reconstruction of PT, and thus for conclusions about the history of Tukanoan languages in the VLA, I examine the relationships among the languages spoken to the west of the Vaupés and previously described (by Mason 1950, Waltz and Wheeler 1972, and Chacón forthcoming) as "Western Tukanoan" (WT). I provide a review of the documentary literature on Tukanoan as a whole in §2.1 and §2.2, then evaluate the state of historical and comparative research on the family in §2.3. In §3, I discuss the linguistic comparative method (LCM), the principal method which this paper employs to reconstruct relationships between languages. §3 also describes some differences between the descriptive data used in this essay and that employed in previous historical-linguistic work on Tukanoan.

§4 and §5 contain the substance of the essay. In §4, I apply the LCM to lexical data from across the Tukanoan languages. Using the results of LCM analysis of a large cognate matrix described in §4.1, I reconstruct the consonant inventory of Proto-Western-Tukanoan (PWT) in §4.2 and extrapolate some phonological evidence for the phylogenetic unity of WT in §4.3. I then propose an internal classification of WT based exclusively on phonological evidence in §4.4. §5 examines the morphology of the WT languages within a similar framework. I first

describe my process for assembling a cognate matrix of bound morphemes from across the family, then identify some correspondences between morphemes which are unique to WT and offer reconstructions of the forms of these morphemes in PWT (5.3). While this analysis yields minimal evidence for the phylogenetic unity of WT, it does allow me to propose an internal classification of the languages based on morphological evidence (5.4). To conclude, I evaluate the sum of the phonological and morphological evidence presented in §4 and §5 for the existence of a WT clade and the internal classification of that clade (6).

2 Background

2.1 The Tukanoan languages of the Vaupés

Almost all documentary research on the Tukanoan languages, inside and outside the Vaupés, has been conducted by fieldworkers affiliated with the Summer Institute of Linguistics (SIL). SIL linguists produced the first or only comprehensive descriptive materials for 12 of the 13 documented Tukanoan languages spoken in or near the Vaupés: Alemán et al. (2000) and Miller (1999) on Desano; Jones and Jones (1991), Stolte and Stolte (1979), and Barasana Literacy Committee et al. (2009) on Barasana, Bará (Southern Barasana), and Eduria; Criswell and Brandrup (2000) on Siriano; Metzger (2010) on Karapana; Barnes and Tamayo (1988) on Tuyuka; West (1980) and West and Welch (2004) on Tukano; Strom (1992) on Retuara; Smothermon and Smothermon (1993) and Smothermon et al. (1995) on Makuna; Klumpp and Klumpp (1973) on Piratapuyo; and Waltz (2007) on Kotiria (also called Wanano).

Non-SIL linguists have also produced important documentation for several ET languages.

Their works include Stenzel (2004) on Kotiria; Gomez-Imbert (1988), Gomez-Imbert and Kenstowicz (2000), and Gomez-Imbert (2003) on Barasana and Tatuyo; and Ramirez (1997a, 1997b) on Tukano. According to the most recent survey of literature on Tukanoan (Chacón forthcoming: 3), there are an additional three Tukanoan languages spoken in or very close to the Vaupés – Pisamira, Tanimuka, and Yuruti – which have not yet been described with a published wordlist, dictionary, or grammar. With these three varieties, the number of living Tukanoan languages from the eastern part of the family's range comes to 16.

The volume and quality of the documentary data for these languages is highly variable. Waltz (2007), for example, is a comprehensive dictionary of Kotiria, listing more than 4000 headwords and including notes on phonology, a grammar sketch, and lists of minimal pairs demonstrating the contrastiveness of tone, aspiration, and nasality. Each entry includes a phonetic transcription, and lexical tone is consistently transcribed. Together, this work and Stenzel's (2004) reference grammar of Kotiria render it among the best-documented Amazonian languages. The documentation of Desano prior to the publication of Silva (2012) provides a much more typical example of the descriptive materials available for most Tukanoan languages. Although a SIL dictionary of Desano is available, it includes just 896 headwords and indicates tone only in minimal pairs (Alemán et al. 2000: 5). The SIL linguists who worked on this language also produced a grammar (Miller 1999), but it is short (180 pages), discusses very few topics in any depth, and does not include formal or theoretical explanations for the phenomena it describes.[1] Most SIL grammars and dictionaries of Tukanoan languages from the Vaupés share these characteristics.

1 Silva (2012) is a full-length descriptive grammar of Desano, informed by the SIL materials but many times more comprehensive than they are.

2.2 Tukanoan languages outside the Vaupés

SIL fieldworkers are also responsible for most documentary materials on Tukanoan languages spoken outside of the Vaupés. Here, the descriptive materials by SIL linguists are Wheeler (1987) on Siona; Cook and Gralow (2001) and Cook and Criswell (1993) on Koreguaje; Johnson et al. (1990) and Piaguaje et al. (1992) on Secoya; Velie (1975) and Velie and Velie (1981) on Máíhɨ̃ki (also called Orejón); and Morse and Maxwell (1999) and Morse et al. (1999) on Kubeo. Non-SIL linguists have also conducted research on these languages. Chacón (2012) describes the phonology and morphology of Kubeo in much greater depth than Morse and Maxwell (1999), and an ongoing documentation project on Máíhɨ̃ki led by Lev Michael has produced a 2000-headword dictionary (Michael et al. 2012b) and a 7,000-word text corpus (Michael et al. 2012a).

The SIL documentation of Koreguaje and Kubeo is of about the same quality as the typical SIL description of a Tukanoan language from the Vaupés. Wheeler's (1987) SIL grammar and dictionary of Siona is much more comprehensive and theoretically informed, approaching the quality of Waltz (2007); this work builds on Wheeler's (1970) Berkeley dissertation. On the other hand, the dictionaries of Máíhɨ̃ki and Sekoya are much less complete than almost any other SIL publication on a Tukanoan language, and the lexical data in Velie and Velie (1981) is sometimes inconsistent with field data, from the same speakers, recently gathered by Michael et al. (2012b). In this essay, I therefore use Chacón (2012) for Kubeo and Michael et al. (2012b) for Máíhɨ̃ki in place of the SIL sources where possible.

2.3 Historical and comparative Tukanoan research

Although Brinton (1892) gave Tukanoan its name, Mason (1950) was the first to propose it as a family-level group and to offer an internal classification of the Tukanoan languages. Based on application of the linguistic comparative method (LCM) to a very small sample of lexical data, he classified Máíhɨ̃ki, Koreguaje, Siona, and Sekoya (together with three very poorly attested dead languages) as forming a "Western Tukanoan" clade, while the languages in the family formed an "Eastern Tukanoan" clade. Waltz and Wheeler (1972) updated Mason's (1950) classification on the basis of new lexical data gathered by their SIL colleagues. They made minimal changes to his internal classification of Western Tukanoan, which claimed that Máíhɨ̃ki was the first-diverging language within the group, with Koreguaje diverging second, and Siona and Sekoya forming a subgroup. Waltz and Wheeler (1972) did depart from Mason (1950) in that they posited the existence of a Central Tukanoan clade consisting only of Kubeo, and in some details of their internal classification of Eastern Tukanoan, which is beyond the scope of this essay.

Recently, Chacón (forthcoming) has attempted to revise Waltz and Wheeler's (1972) classification on the basis of new (post-1970) documentary data made available by the advent of computer-based fieldwork. Like Waltz and Wheeler (1972), Chacón's (forthcoming) analysis relies entirely on application of the LCM to lexical data. While his classification of Waltz and Wheeler's (1972) Eastern and Central Tukanoan languages differs from their classification in several ways, his proposals for the membership and internal classification of the WT clade are identical to Waltz and Wheeler's (1972). Specifically, like the earlier work, Chacón's (forthcoming) classification relies crucially on the existence of a series of ejective consonants

in Máíhɨ̃ki to classify that language as the first-diverging in the WT clade, and Koreguaje as the second-diverging. However, recent field data casts serious doubt on the existence of an ejective series in Máíhɨ̃ki. Although Velie (1975) listed a handful of lexical items with ejective consonants in his first description of this language, his later dictionary of Máíhɨ̃ki (Velie and Velie 1981) does not posit any ejective-pulmonic contrast. Michael et al. (2012b), working with the same speakers consulted by Velie, also found no evidence for ejective consonants in Máíhɨ̃ki. As such, absent replication of Velie's early observations, field data best supports that the claims about ejectives in Velie (1975) arose from transcription errors or idiolectal variation, not a systematic ejective-pulmonic contrast in Máíhɨ̃ki. This necessitates reappraisal of Chacón's (forthcoming) reconstruction of the PT phonemic inventory, which includes a series of glottalized stops; of the sound changes which he uses to define the WT clade and individual WT languages; and of the genetic relationships which he proposes based on those changes.

3 Methods

3.1 The linguistic comparative method

This essay employs the linguistic comparative method (LCM) as its primary method for data analysis. The LCM is a method in historical linguistics which uses systematic phonological correspondences to identify genetic relationships between languages. Linguists begin an LCM analysis by compiling sets of morphemes which are similar in form and meaning (candidate cognates) from languages which may be genetically related. They then examine the set of candidate cognates for systematic phonological correspondences. Consistent sound

correspondences between languages in semantically close, non-borrowed items support that the languages are genetically related – that is, that they developed from a single speech variety (proto-language) and belong to the same phylogenetic unit (clade). After identifying language families through the LCM in this way, workers classify languages into genetic units based on shared innovations in lexicon, phonology and/or morphology relative to the reconstructed proto-language. Languages which belong to different subgroups may retain features from their most recent common ancestor. This entails that in historical linguistics, as in biology, shared retained characteristics cannot be used as evidence for clades.

Many linguists and non-linguists have argued that historical linguists using the LCM overstate the consistency of intergenerational transmission, and understate the role of contact, in language change. Researchers interested in proposing "long-distance" language relationships at greater time depths (greater than 6,000 to 8,000 years) have been especially critical of the limitations of the LCM (Campbell and Poser 2008: 297). Yet the LCM has delivered provably correct conclusions about the relationships between many hundreds of languages around the world. Methods of language classification which compete with the LCM, on the other hand, are largely the province of individual linguists, have not been tested on large sets of descriptive data, and have not produced classifications widely accepted in the field (Campbell and Poser 2008: 328). This paper therefore takes the view that – while models of language change should account for both horizontal and vertical transmission of information between grammars – the LCM is a valid and indeed essential tool for historical linguistics.

3.2 Input and output of the linguistic comparative method

The LCM is essentially a function for converting sets of candidate cognates into sets of sound correspondences, which form the basis for reconstruction of proto-sounds, sound changes, and genetic relationships. This essay uses both free lexical items and bound morphemes as input to its LCM analysis. The lexical items are gathered from the published and manuscript dictionaries cited in §4.1. The morphological data comes from published grammatical descriptions, theoretical articles, and unpublished fieldnotes on Máíhɨ̃ki. These sources are fully described at §5.1.

All data in this essay is presented in tables using a phonemicized orthography approximating the IPA. Nasality and tone are contrastive in most Tukanoan languages. Where nasality is a feature of the morpheme or syllable rather than of the segment, it is indicated in this orthography with a tilde [∼] before the nasalized material (following standard Tukanoanist practice). For example, because Kotiria has morpheme-level nasality, [nũmĩã] "woman" is phonemicized as /∼dubia/ (Stenzel 2004: 69). Symbols representing nasal stops and nasalized vowels (e.g. /m/, /ã/) will be used only for languages which have a segmental contrast between nasal and oral stops and vowels. Presentation of lexical data will include as much information about tone as available, and presentation of morphological data will include description of the tonal behavior of affixes where known. Yet because most documentary materials on Tukanoan contain little information about tone, comparative discussion of tone issues will not generally be possible.

The output of the LCM analysis will consist of sets of cognates and sound correspondences between Tukanoan languages, reconstructions of some items in PT and PWT, and hypothe-

sized sound changes from PT and PWT to attested languages. Cognate sets and the analyses based on them will rest on my judgments of semantic similarity and cognacy. Reconstructions and hypotheses about sound changes from the proto-languages will be informed both by the cognate sets and by cross-linguistic generalizations about the probability of transition between the relevant sounds. In reconstructing segments, the analysis will not select segments that are synchronically attested in Tukanoan over unattested proto-segments of equal phonological plausibility. Caution on this point is necessary because of the known phonological convergence effects in the Vaupés, which have almost certainly introduced new segments and distinctive features into ET (Epps 2011).

4 Proto-Western Tukanoan phonology

4.1 Lexical data

The input to the phonological analysis was a matrix listing lexical items for 134 basic-vocabulary concepts in each of 11 Tukanoan languages: Kubeo (data gathered from Morse and Maxwell 1999, Morse et al. 1999, Chacón 2012), Desano (Alemán et al. 2000), Barasana (Barasana Literacy Committee et al. 2009), Makuna (Smothermon and Smothermon 1993), Kotiria (Waltz 2007), Karapana (Metzger 2010), Tuyuka (Barnes and Tamayo 1988), Máíhɨ̃ki (Michael et al. 2012b), Koreguaje (Cook and Gralow 2001), Sekoya (Piaguaje et al. 1992), and Siona (Wheeler 1987). The list of concepts was drawn from Chacón (forthcoming), and differs from Chacón's list only in that it does not include person- or TAM-marking bound morphemes. 115 (85.8%) of the 134 lexical items in the rows were nominals (nouns, pronouns,

and free adjectives) or nominal elements (such as bound adjectival prefixes or classifiers).[2] The remaining 19 (14.2%) of the 134 lexical items in the rows were verbs.

Because of incompleteness in the data sources, the matrix was not complete. 228 (15.5%) of the 1474 cells in the matrix did not contain a lexical item. Table 1 presents the percentages of data missing in each of the 11 languages in the columns. Following the procedure employed by Chacón (forthcoming), cells were left blank if the data source either did not list a lexical item for the given meaning or listed a lexical item which was clearly not cognate with any other item in the same row of the matrix. As such, the proportions of missing data shown for each language in Table 1 are inversely related to both the completeness of the source(s) consulted and the degree of lexical innovation in the language relative to the others under study. Even in the languages with the greatest proportion of missing data, the data set is still fairly complete relative to the Tukanoan correspondence sets used by Mason (1950) and Waltz and Wheeler (1972).

Language	Missing Data	Language	Missing Data
Kubeo	23.1%	Desano	12.7%
Máíhɨ̃ki	22.4 %	Tuyuka	26.9%
Barasana	16.4 %	Siona	11.9%
Sekoya	15.7 %	Karapana	11.2%
Makuna	12.7%	Koreguaje	5.2%
		Kotiria	4.5%

Table 1 Missing lexical data by language

The matrix was compiled from the same sources cited by Chacón (forthcoming), with the exception that data on Máíhɨ̃ki came from Michael et al. (2012b) rather than Velie and

2 All of the bound forms included in the matrix are content elements which appear to have grammaticalized relatively recently from free words. They display the same sound correspondences as the free words in the data set, and there is no evidence that the bound elements in the matrix has undergone analogical changes which the free words have not.

Velie (1981). It differs from the correspondence set presented in the appendix to Chacón (forthcoming) in four main ways. First, Chacón stripped all tone information from the items in his matrix, while I entered tone information into this matrix when it was available. Second, this matrix does not include data from Retuara (or the closely related Tanimuka), Tukano, or Kueretu. I was not able to obtain the data sources used by Chacón for Tukano and Kueretu, and I did not find any lexical data in the source (Strom 1992) cited by Chacón for his "Retuara/Tanimuka" data. Third, in some cases I did not find the form given in Chacón's correspondence set anywhere in the data source which I consulted, and in others I found a cognate form which Chacón did not list. Discrepancies of the former kind occurred 67 times (in 4.5% of the cells) in the matrix, and discrepancies of the latter kind occurred 143 times (in 9.7% of the cells).

Fourth, and most significant for the analysis, some items appeared with different segmental or suprasegmental forms in Chacón's data than in the data sources. For example, Kotiria contrasts plain and aspirated voiceless oral stops in word-initial position. Chacón's cognate set lists several words with initial aspirated /p^{h}/ which are written with /p/ in Waltz (2007), and several others with initial plain /p/ which have /p^{h}/ in Waltz (2007). Discrepancies of this type, involving differences in contrastive segments or features between the appendix to Chacón (forthcoming) and the data sources, appear 72 times in the matrix (affecting 4.9% of the cells and 5.8% of the non-empty cells).

4.2 Western Tukanoan sound correspondences

Applying the LCM to the data contained in the matrix yielded 18 sound correspondences occurring in two or more items in the four languages (Máíhɨ̃ki, Koreguaje, Sekoya, and

Siona) previously classified as forming the "Western Tukanoan" clade. Tables 2, 3, and 4 present these correspondences, the environments in which they occur, and an example of a row in the matrix in which each correspondence is found. Each row of the tables also includes a reconstruction of the proto-consonant which is hypothesized to have yielded the correspondence set shown in that row. The column which contains the reconstruction in each table is labeled "Reconstruction" rather than "Proto-Western-Tukanoan," since – absent evidence from the other Tukanoan languages – the availability of reconstructions cannot prove that the languages under study form a clade.

Notation in the tables is phonemic and IPA, with V representing a vowel underspecified for nasality and $\tilde{V}$ representing a nasal vowel. The notation also does not differentiate [r] and [d]. Although most Tukanoan orthographies distinguish between [r] and [d], the sounds are allophones of a single phoneme, represented here as /ɾ/, in all Tukanoan languages.

Reconstruction	Máí	Kor	Sek	Sio	Environment	Example
*b	b	p	p	b, p	#_V	CHILI, DOVE.SP, LARVA
*t	t	t, t^h	t	t	#_V, V_V	TOBACCO, GRASS, STUMP
*k	k	k^h	k	k	#_V, V_V	HEAVY, LARVA, MANIOC
*g, ∅[3]	g	k	k	g	#_V	HOLE, TOOTH, TORTOISE
*k^w	k	k^h	k^w	k	$\tilde{V}$_$\tilde{V}$	NOSE, TO STOP/STAND
*k^w	k,∅	k, k^h	k	∅	V_V	ANT, PALM WEEVIL
*ʔ	following low tone	ʔ	ʔ	ʔ	V_V, V_C	FISH, HOUSE

Table 2 Sound correspondences in the "Western Tukanoan" languages: Oral stops

The reconstructions in Table 2 require some explanation. *b rather than *p is reconstructed for the correspondence exemplified by CHILI because of strong evidence for a *p > h sound change in Western Tukanoan (discussed in §4.3). The reflex of this segment in Koreguaje and Sekoya is [p] rather than [b] because those languages do not have a voicing distinction

3 Word-initial /g/ is regularly deleted in the Southern dialect of Máíhɨ̈ki.

for stops. Similarly, *t is the only possible reconstruction for the correspondence exemplified by TOBACCO and shown in (1).

(1) Reflexes of PT *[mɨto] (/∼bɨ-to/)[4] "tobacco"

i. Mai [mɨ́to] "tobacco"

ii. Kor [mɨto] "tobacco"

iii. Sek [mɨto] "tobacco"

iv. Sio [muto] "tobacco"

Reconstructing *t for the correspondence exemplified by (1) does force the conclusion that language-internal developments in Koreguaje led to a split of *t into /t/ and a new phoneme /t^{h}/. It is not possible to identify a conditioning environment for this split. Koreguaje /t/ and /t^{h}/ both consistently correspond to /t/ in the other languages in the matrix, and the two phonemes are not in complementary distribution – both occur word-initally and word-medially in overlapping vocalic contexts.[5]

All else being equal, these facts would generally lead one to reconstruct a phonemic contrast between two oral alveolar stops to at least PWT, if not PT. Yet reconstructing such a contrast would create two problems. First, /d/ is not a contrastive phoneme in any Tukanoan language (it is usually an allophone of /ɾ/). This means that it is implausible to reconstruct a voicing distinction for oral alveolar stops at any stage in the development of WT. It would instead be necessary to reconstruct some contrast other than voicing – perhaps a laryngeal contrast, perhaps some secondary articulation – between the alveolar

4 See the discussion of Table 3 for notes on nasality in Tukanoan.

5 Cook and Gralow (2001: 9) state that their orthographic [t] is "almost always aspirated," but this is not borne out by the phonetic transcriptions in their work.

proto-stops. Given the absence of such contrasts in the attested Tukanoan languages other than Koreguaje, that reconstruction would rest on very scanty evidence.

Second, reconstructing a contrast for alveolar stops in PWT entails reconstructing mergers between those two stops in all WT languages except Koreguaje, as well as in all Eastern Tukanoan languages. Yet Koreguaje is clearly a phonologically innovative language, with several phonemes – among them /f/, /v/, and three voiceless nasals – that are unknown in other Tukanoan languages and rare in Amazonian languages generally. It is therefore more plausible to characterize the double reflex in Koreguaje as the outcome of language-internal phonological innovation than as a feature inherited from PWT in Koreguaje but lost in other WT languages and all ET languages.

Due to data inconsistencies, the velar stops present a more difficult problem than the alveolar stop. Sekoya and Koreguaje do not contrast stops for voicing, but Koreguaje contrasts velar stops for aspiration and Sekoya contrasts them for labialization. Labialization is also contrastive for voiced and unvoiced velars in Máíhɨ̃ki. Additionally, many of the SIL-produced data sources on these languages use [cu] or [qu] to represent both [k] and [k^{w}] before front vowels, and no data source distinguishes orthographically between the sequence [kw] and the segment [k^{w}]. This distinction is even collapsed in the practical orthography employed by Michael et al. (2012b) for Máíhɨ̃ki, as (2) shows.

(2) Contrastive labialization for velars in Máíhɨ̃ki (Farmer et al. forthcoming)

i. *gwɨ́yí* [g^{w}ɨ̃ʤí] "I am gathering"

ii. *gwɨ́yì* [gʉ̃́ɨ̃́ʤí] "I am digging (earth, a canoe)"

The non-labialized velar stops are relatively easy to reconstruct. *k is securely recon-

structed for the correspondence exemplified by HEAVY and illustrated in 3. This segment was is inherited as /k/ in Máíhɨ̃ki, Sekoya, and Siona, and as /k^{h}/ in Koreguaje. Similarly, we can securely reconstruct *g for the correspondence exemplified by HOLE and shown in (4). This segment was inherited as /g/ in Máíhɨ̃ki and Siona and /k/ in Koreguaje and Sekoya.

(3) Reflexes of PT *[ɾɨkɨ] "heavy"

i. MAI [dɨ́kɨ́-] "be heavy"

ii. KOR [rɨk^{h}ɨ] "heavy"

iii. SEK [dɨkɨ-] "heavy"

iv. SIO [rɨkɨ-] "be heavy"

(4) Reflexes of PT *[gohe] "hole"

i. MAI [góhé] "hole"

ii. KOR [kohe] "hole"

iii. SEK [kohe] "hole"

iv. SIO [gohe] "hole"

While *k and *g merged in Sekoya with the loss of voicing, Koreguaje contrasts the reflexes of these segments for aspiration, inheriting *k as /k^{h}/ and *g as /k/. The only exceptions to this regular sound change in Koreguaje are two items in which PWT *k appears to have been inherited as /k/ rather than /g/: [jãki-] "to chew," which is probably best reconstructed to PWT as *ʤãki, and [sũkiɲɨ], which is reconstructed to PWT as *sũki-. These two items are also the only two instances of a velar stop followed by the high front vowel /i/ in the Koreguaje data in the matrix, suggesting that the /i/ blocks aspiration of a

preceding velar in this language (possibly due to phonetic palatalization). If this is correct, then the Koreguaje sound changes can be parsimoniously stated as PWT *k > k / _i, *k > k^h elsewhere, and *g > k.

Turning to the labialized velars, *k^w is reconstructed for the two correspondences exemplified by NOSE and ANT. The former of these correspondences occurs in overlapping distribution with the correspondences exemplified by HEAVY, but yields [k^w] rather than [k] in Sekoya. In the latter correspondence, velar stops present in some items from Sekoya and Koreguaje are absent from the cognate items in Máíhɨ̃ki and Siona. (5) and (6) illustrate these patterns.

(5) Reflexes of PT *[ũ-k^we-] (/∼u-k^we-/) “nose”

i. MAI [ű́kebɨ] “nose”

ii. KOR [ɨ̃k^he] “noses”

iii. SEK [õk^we] “nose”

iv. SIO [ũkuebɨ] “nose”

(6) Reflexes of PT *[mek^{w}a] (/∼be-k^wa/) “ant sp.”

i. MAI [méa] “ant (generic)”

ii. KOR [mek^{h}a] “*arriera* ant, ant sp.”

iii. SEK [meka] “ant sp.”

iv. SIO [mea] “ant”

This suggests that the correspondence exemplified by (6) is the outcome of progressive lenition, eventually resulting in deletion, of some proto-stop with a velar place of articulation.

Although the stop must have contrasted with *k, it has only unvoiced reflexes, even in languages which contrast velar stops for voicing. This forces the conclusion that the (definitely velar) proto-segment which underlies the correspondence in (6) contrasted with *k on the basis of some feature other than voicing. Given the presence of /kʷ/ in the inventories of several of the attested languages, it is therefore most parsimonious to reconstruct *kʷ for this segment. Since the correspondences exemplified by NOSE and ANT occur in complementary distribution, this segment is reconstructed for both correspondence sets. This reconstruction is very tentative, due to the small number of tokens for both correspondences and to the problems with transcription of labialized segments discussed above.

The correspondence exemplified by FISH is clearer, and we can confidently reconstruct a glottal stop. This segment was inherited without change in Koreguaje, Siona, and Sekoya, and inherited as low tone on the vowel which followed the glottal stop (with no segmental reflex) in Máíhĩki. The loss of /*ʔ/ likely played a significant role in the genesis of underlying low tones in Máíhĩki (Farmer 2012), and – with PWT *w > Máíhĩki b, which will be discussed in the analysis of Table 4 – represents important evidence for phonological innovation in Máíhĩki relative to Siona and Sekoya.

Reconstruction	Máí	Kor	Sek	Sio	Environment	Example
*m (/*b/)	m	m	m	m	#_V, V_V	DEER, MACAW, MAN
m (/b/)	m	p, m	p	b[6]	#_$\tilde{V}$	GUAMA.FRUIT, PEOPLE
n (/ɾ/)	n, ɾ	n, ɾ	n, ɾ	n	#_V	CHARCOAL, PEACH PALM, WIFE
*ɲ (/*j/)	ɲ	ɲ	ɲ	ɲ	#_V, V_V	YAM, SNAKE, GRANDFATHER

Table 3 Sound correspondences in the "Western Tukanoan" languages: Nasal stops

*m is securely reconstructed for the correspondence exemplified by DEER, and *ɲ for

6 The orthography used for Siona in Wheeler (1987) indicates syllable-level nasality with a coda nasal stop. I suspect that some tokens of *b* in this source are phonetically [m].

that exemplified by YAM. These reconstructions are phonetic, and as discussed below, they reflect the underlying representations */b/ and */j/. The reconstructions of *m and *n for the corrrespondences exemplified by GUAMA.FRUIT and CHARCOAL are equally secure, but require some further explanation of nasality in Tukanoan.

Most Tukanoan languages have phonological processes which involve spreading of nasality within the morpheme or across morpheme boundaries. These processes, which are analyzed sometimes as spreading of [+nasal] from a nasalized vowel and sometimes as docking of a suprasegmental [+nasal] feature, typically target all voiced segments in the syllable. Targeting by nasal spreading causes approximants to become nasalized and voiced stops (the only voiced obstruents in these languages) to become nasal stops. That is, if the onset is a nasal stop, then the vowel is usually also phonetically nasal. Conversely, if the nucleus is nasalized, then the onset may be either nasal (if it is voiced) or oral (if unvoiced). For example, the first syllables of the Máíhĩki roots [tɬ́tɨ-] "burn (intrans.)" and [t̃́tɬ́] "bird sp., *Psophia crepitans*" can contrast for nasality because the onset is unvoiced. No such contrast is possible between [mɬ́mɬ̃] "butterfly" and a hypothetical form [*bɬ́bɬ̃], because phonetic nasality must spread from the nasal nucleus of a syllable (or the suprasegmental [+nasal] specification, depending on one's phonological analysis) to the voiced onset.

As such, when a segment in a Tukanoan word changes in nasality, it is often more accurate to regard the change as a morpheme-level change in the value of [nasal] than as a segmental sound change. The correspondences exemplified by GUAMA.FRUIT and CHARCOAL require just this analysis. In the proto-language, these items had /*b/ and /*ɾ/ in the underlying representations, but were realized with [*m] and [*n] due to the nasal context. The morpheme-level specifications of nasality then idiosyncratically changed in certain items

from [+nasal] to [-nasal]. In the case of the correspondences exemplified by GUAMA.FRUIT, the changes to [-nasal] in Sekoya and Siona, and in certain items in Koreguaje, yielded the surface representation *b. *b then fell together with *p in Sekoya and Koreguaje when those languages lost the voicing distinction for stops, producing the [p] reflex attested here.

The same process took place in the correspondence exemplified by CHARCOAL. Here two accounts are possible: either Máíhɨ̃ki, Koreguaje, and Sekoya idiosyncratically lost nasality in some morphemes which were nasal in the proto-language, or Siona idiosyncratically gained nasality in some morphemes which were oral in the proto-language. Table 3 adopts the first account in the interest of consistency with the analysis of the correspondence exemplified by GUAMA.FRUIT. In either case, the underlying representation in the proto-language must have been /*ɾ/, yielding [*n] in nasal contexts and [*r]/[*d] in oral ones.

Reconstruction	Máí	Kor	Sek	Sio	Environment	Example
*h	h, ∅[7]	h	h	h	#_V, V_V	FATHER, FIRE, BIG
*s	s	s	s	s	#_V	COLD, LONG/FAR
*j	j	j	ts	sʔ	#_i	FACE, EGG, BLOOD
*j	j	j, ʧ	j	j	#_V	COTTON, JAGUAR, TERMITE
*j	j	j	j	j	V_V	BAT, DANCE/SONG
*w	b	β	w	w	#_V	PARROT, TAPIR, TO FISH
*w	w	β	w	w	#_V	HOUSE

Table 4 Sound correspondences in the "Western Tukanoan" languages: Glottals and non-obstruents

*h,*s, and *j are securely reconstructed for the correspondences exemplified, respectively, by FATHER, COLD, and BAT. The other correspondences which involve [j] reflexes are also best explained by reconstructing *j for the proto-form. In the correspondence exemplified by FACE, evidence from the other Tukanoan languages strongly suggests that *j is the outcome of palatalization of a sequence /*ɾi/, realized as [*di], in Proto-Tukanoan (see §4.3). The

7 There is a change in progress in the Northern dialect of Máíhɨ̃ki causing deletion of [h] intervocalically.

	Labial	Alveolar	Velar	Palatal	Glottal
Stop	b	t	k, g, k^w		ʔ
Nasal	[m] ∼ /b/	[n] ∼ /ɾ/		[ɲ] ∼ /j/	
Tap		ɾ			
Fricative		s			h
Approximant	w			j	

Table 5 Reconstructed phonemic inventory for most recent common ancestor

tap was palatalized to /*j/ before the front vowel [i] in PWT, then underwent fortition and depalatalization to [ts] in Sekoya and [sʔ] in Siona. /*j/ also underwent fortition to [ʧ] before the back vowels [o] and [u], and in the item meaning "cotton," in Koreguaje.

The correspondences involving [w] are more difficult. All of the items involved in the correspondence exemplified by PARROT correspond to items with [w] in the Tukanoan languages not shown here, and all Tukanoan languages also have a form meaning "house" with initial [w]. Although Máíhɨ̃ki has [b] corresponding to [w] in other Tukanoan languages for PARROT, and [w] corresponding to [w] for HOUSE, it is difficult to argue that these two correspondences are reflexes of contrastive segments in the proto-language. The most parsimonious analysis is rather that one segment, *w, underlies both the PARROT and the HOUSE correspondence. On this account, *w occluded to [b] in Máíhɨ̃ki in the great majority of cases, but was inherited as [w] in the item meaning "house," perhaps because of its grammatical use as a repeater classifier for nouns referring to buildings (Neely 2012).

Table 5 summarizes the phoneme inventory reconstructed for the most recent common ancestor of Máíhɨ̃ki, Koreguaje, Siona, and Sekoya. This paper will now take up the question of whether these four languages are more closely related to each other than they are to the other Tukanoan languages – that is, of whether we can characterize the phonemic inventory in Table 5 as a reconstruction of "Proto-Western Tukanoan."

4.3 Phonological evidence for a Western Tukanoan clade

Several of the sound correspondences discussed in §4.2 strongly support the existence of a clade within Tukanoan consisting only of Máíhɨ̃ki, Koreguaje, Sekoya, and Siona. All of these languages show evidence of at least four shared phonological innovations not attested together in any other language in the sample.

The correspondence exemplified by FATHER represents the clearest of these innovations. Where Máíhɨ̃ki, Koreguaje, Sekoya, and Siona have [h], either word-initially or intervocalically, five of the other seven languages in the sample regularly have [p], as illustrated in Table 6. Sound changes from [p] to [h] are common and well-attested cross-linguistically, while the opposite change is phonetically unnatural and not well-attested. The most parsimonious analysis of the data in Table 6 is therefore that Proto-Tukanoan had *p in the items listed here, and that a *p > h sound change took place in Barasana, Makuna, Máíhɨ̃ki, Koreguaje, Siona, and Sekoya. No other language in the sample shows evidence of this change. This entails that PT *p > h represents a phonological innovation shared by and unique to these six languages, which suggests that they should be subgrouped together.

Kub	Des	Bar	Mak	Kot	Kar	Tuy	Máí	Kor	Sek	Sio	Meaning
p	p	h	h	p	p	p	h	h	h	h	FATHER
p	p	h	h	p	p	p	-	h	h	h	ARMADILLO
p	p	h	h	p	p	p	h	h	h	h	FIRE/FIREWOOD
p	p	h	-	p	p	-	h	h	h	h	PALM.WEEVIL
p	p	h	h	p	p	p	h	-	h	h	THUNDER
p	p	h	h	p	p	p	h	h	h	h	TO.BLOW

Table 6 Correspondence of [p] and [h]

Another sound change shared by and unique to Máíhɨ̃ki, Koreguaje, Siona, and Secoya appears in the correspondence exemplified by FACE. In this correspondence, the "Western

Tukanoan" languages have palatal glides or alveolar affricates where Kubeo has [h] and the other languages in the sample have [ɾ] before [i], as shown in Table 7.

Kub	Des	Bar	Mak	Kot	Kar	Tuy	Máí	Kor	Sek	Sio	Meaning
h	ɾ	ɾ	ɾ	ɾ	ɾ	ɾ	j	j	ts	-	FACE
h	ɾ	ɾ	ɾ	ɾ	ɾ	ɾ	j	j	ts	sʔ	EGG
h	ɾ	ɾ	ɾ	ɾ	ɾ	ɾ	j	j	ts	sʔ	RIVER
h	ɾ	ɾ	ɾ	ɾ	ɾ	ɾ	j	j	ts	sʔ	BLOOD

Table 7 Correspondence of [ɾ] to [j] and affricates before [i]

PT *ɾ should be reconstructed for all items in Table 7. Since word-initial /ɾ/ is realized as a voiced alveolar stop or tap in oral contexts in all languages in this sample, it is parsimonious to reconstruct the same allophony between [ɾ] and [d] to PT. The sequence */ɾi/ would therefore have been realized as [di] for all of the items in Table 7. [d] is cross-linguistically prone to palatalization before front vowels, and exactly such a change has occurred here. PT phonetic *[d] must have lenited to [ʤ] before [i] in Máíhɨ̃ki, Koreguaje, Siona, and Sekoya. The affricate then devoiced in Sekoya and Siona, yielding the attested [ts] and [sʔ] reflexes. In Koreguaje and Máíhɨ̃ki, on the other hand, the voiced affricate remained in the phonetic inventory. It is in free variation with [j] in Máíhɨ̃ki (Farmer et al. forthcoming), and with [j] and [ʧ] in Koreguaje Cook and Gralow (2001: 9). We therefore reconstruct *j for this segment in PWT. Lenition of PT *[d] before [i] also occurred in Kubeo, but yielded the fricative [h] rather than a glide or affricate.

The palatalization change represents another clear phonological innovation supporting subgrouping of Máíhɨ̃ki, Koreguaje, Siona, and Sekoya together. It could be argued that the PT segment underlying the correspondences in Table 7 was a glide, which was inherited as such or underwent minimal fortition in those four languages, lenited to [h] in Kubeo, and

occluded to [d] ∼ [ɾ] in Desano, Barasana, Makuna, Kotiria, Karapana, and Tuyuka. Yet the PT *[d] > [j, ts, sʔ] /_i account has much stronger phonetic motivation, is better-attested cross-linguistically, and requires reconstruction of a much smaller number of sound changes. As such, the best-supported account of the data in Table 7 is that /ɾ/ is the inherited segment, while the glides and affricates in Máíhɨ̃ki, Koreguaje, Siona, and Sekoya are innovative.

Only Máíhɨ̃ki, Koreguaje, Siona, and Sekoya exhibit the PT *p > h change illustrated in Table 6 as well as the palatalization changes shown in Table 7. Each of the changes occurs in at least one other language in the sample: Barasana and Makuna have *p > h, and Kubeo has lenition of */ɾ/ before [i]. Yet no language outside the "Western Tukanoan" group displays both of the changes. This strongly suggests that the two changes discussed so far are the outcome of shared phonological development of Máíhɨ̃ki, Koreguaje, Siona, and Sekoya in isolation from the other Tukanoan languages. In other words, *p > h and [ɾ]-palatalization are sufficient to define a Western Tukanoan clade.[8]

Two other correspondences, involving [s] and [j], further support that WT is a valid phylogenetic unit. These are illustrated by Table 8.

Kub	Des	Bar	Mak	Kot	Kar	Tuy	Máí	Kor	Sek	Sio	Meaning
j	j	s	s	s	s	s	j	j	j	j	BAT
∅	j	-	j	s	s	-	j	j	j	j	CHEEK
-	j	j	j	j	j	j	s	s	s	s	PECCARY.SP
h	j	-	-	j	j	j	s	s	s	s	COLD

Table 8 Correspondences between [s] and [j]

8 Kubeo has been classified (by Mason 1950 and Waltz and Wheeler 1972) as more closely related to the WT languages than to the other Tukanoan languages. Although it could be argued that [h] reflex of PT */ɾ/ in Kubeo represents a similarity between it and WT, there is no evidence for either PT *p > h or lenition of PT */ɾ/ to a glide or affricate before [i] in Kubeo. I therefore exclude Kubeo from the WT clade. Further analysis of Kubeo's relationship to WT is important to comparative Tukanoan studies, but outside the scope of this paper.

The correspondences shown in Table 8 and found elsewhere in the data make clear that there is a significant phonological affinity between [s] and [j] in Tukanoan. The WT languages frequently have [j] where Barasana, Makuna, Kotiria, Karapana, and Tuyuka have [s], and [s] where Kubeo has [h] and the rest of the sample has [j]. It is likely that the correspondences shown in Table 8 are reflexes of two different continuant segments in PT: one which lenited to [s], and another which lenited to [j], in WT. Yet neither of the correspondences is entirely regular and exceptionless, and the matrix contains relatively few tokens of the segments involved in them. As such, while the [s] ∼ [j] pattern lends further support to the hypothesis of shared, isolated phonological development of the WT languages, it does not provide the same level of robust evidence for the clade as *p > h and palatalization do.

4.4 Internal classification from phonological data

The sound correspondences discussed in §4.2 and 4.3 strongly support that the WT languages are closely related and form a discrete phylogenetic unit within Tukanoan. On the other hand, the phonological evidence sheds little light on the internal classification of the WT languages. The phonological point most important to subgrouping is that Koreguaje exhibits significant differences in phonemic inventory from the other WT languages. It has contrastive aspiration, phonemic voiceless nasals, and a pair of bilabial fricatives. None of these features are present in any other WT language, and only contrastive aspiration appears in any other language in the family. These features suggest that Koreguaje should be classified in a subgroup of one, diverging from PWT at some point between the split of PWT from PT and the divergence of Máíhɨ̃ki, Siona, and Sekoya from their most recent common ancestor. The absence of a voicing distinction in both Koreguaje and Sekoya presents a significant challenge

to this classification, since subgrouping Koreguaje away from the other languages requires reconstructing two separate losses of the voicing distinction in Koreguaje and Sekoya. Despite this difficulty, it is still more parsimonious to reconstruct two losses of voicing than to claim that Sekoya, which is very phonologically similar to Máíhɨ̃ki and Siona, shares an most recent common ancestor with Koreguaje that it does not share with those two languages.

Within the group of phonologically more conservative WT languages spoken along the Napo River – Máíhɨ̃ki, Sekoya, and Siona – Siona and Sekoya appear to be more closely related to each other than to Máíhɨ̃ki. Two facts support the classification of Máíhɨ̃ki as splitting from Proto-Napo before the divergence of Sekoya and Siona from their most recent common ancestor. First, Máíhɨ̃ki has [j] ∼ [ʤ] as the outcome of the palatalization change, while Sekoya and Siona have the unvoiced affricates [ts] and [sʔ], and do not have free variation between the palatal glide and the affricate which is the reflex of PT */ɾ/ before [i]. If we posit the sequence of sound changes *[d] > [ʤ] > [*ʧ] > [ts, sʔ], then the presence of an unvoiced affricate as the reflex of PWT *ɾ / ⊥i represents a phonological (or at least phonetic) innovation shared by Sekoya and Siona but not found in Máíhɨ̃ki. Second, Máíhɨ̃ki exhibits two regular sound changes, PWT *b > w word-initially and PWT *ʔ > ∅ with following low tone, which do not appear in Siona or Sekoya. These sound changes force the conclusion that Máíhɨ̃ki underwent a period of significant phonological innovation which Sekoya and Siona did not share – that is, that it belongs to a separate subgroup within the Napo clade. The tree in Figure 1 graphically represents this account, which classifies Koreguaje as the first-diverging language, Máíhɨ̃ki as the second-diverging language, and Sekoya and Siona as forming a phonologically conservative subgroup within WT.

This classification correlates well with geography. Koreguaje is spoken along the Caquetà

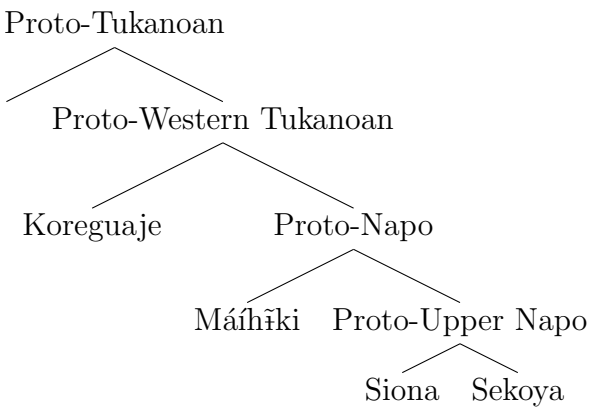

Figure 1 Internal classification of WT from phonological evidence

River in Colombia and is relatively geographically isolated from the three other WT languages, all of which are spoken along the Napo River and its tributaries (in Ecuador, Colombia, and northeastern Peru). Within the Napo group, Siona and Sekoya are spoken in contiguous areas on the upper Napo – Siona upriver, on the border between Colombia and Ecuador, and Sekoya on the border between Colombia and Peru – while Máíhɨ̃ki is spoken on the lower Napo in Peru. The most recent ancestor of Siona, Sekoya, and Máíhɨ̃ki is therefore labeled as "Proto-Napo," and the ancestor of Siona and Sekoya as "Proto-Upper-Napo," in Figure 1.

5 Proto-Western-Tukanoan morphology

5.1 Morphological data

Previous classifications of the Tukanoan languages by Mason (1950), Waltz and Wheeler (1972), and Chacón (forthcoming) have relied entirely on arguments for ordered sound changes, extracted from lexical data. Yet the linguistic comparative method does not operate exclusively on lexical data. It can also be used on bound morphemes to reconstruct the nature and marking of grammatical categories in a proto-language; to reconstruct paths of

regular or analogical morphological change; and to identify additional evidence for sound changes. Additionally, morphology may be less susceptible to borrowing and other forms of contact-induced change than lexicon (Bowern et al. 2008: 2, Meillet 1925). This entails that classifications constructed from morphological analysis, or joint morphological and phonological analysis, may reflect the phylogenetic signal more accurately than trees built exclusively from phonological data (Campbell and Poser 2008: 181).

I therefore compiled a second comparative matrix of morphological data from 11 Tukanoan languages. This matrix consisted of 663 bound morphemes gathered from grammars and articles describing Kubeo (Chacón 2012), Desano (Miller 1999), Barasana (Jones and Jones 1991), Makuna (Smothermon et al. 1995), Tukano (West 1980), Kotiria (Stenzel 2004), Tuyuka (Barnes 1984, Malone 1988), Máíhɨ̃ki (fieldnotes by Lev Michael, Stephanie Farmer, and me), Siona (Wheeler 1987), Sekoya (Johnson et al. 1990), and Koreguaje (Cook and Gralow 2001). Ten languages were included both in this matrix and in the lexical data matrix described at §4.1, while two languages, Karapana and Tukano, were included in only one of the matrices. Karapana was included in the lexical data matrix but excluded from the morphological analysis because I was not able to access a description of its morphology. Similarly, Tukano was included in the morphological analysis but not in the lexical data matrix because I was able to access a grammar of the language, but not a dictionary.

The matrix included only bound verbal morphemes. I focused collection of the data on finite inflectional morphemes. In Tukanoan, finite verbal inflections consist of tense-aspect-modality (TAM) markers, fusional morphemes marking both TAM and subject person and number, and tense-evidential markers (some but not all of which also mark subject person and number). However, I also added to the matrix some non-finite verbal inflectional markers

which were well-attested across the sample of languages. These included verbal subordinators, such as simultaneous subordinate clause markers and sequential subordinate clause markers, and nominalizers. I aimed to input as many morphemes belonging to the relevant categories into the database as possible, without regard for cognacy between the morphemes listed for each language. As a result, the matrix contains no empty cells; it simply lists all and only the morphemes described in the documentation for each language.

I selected verbal inflectional morphemes as the input to the morphological analysis primarily for reasons of data volume and quality. Most Tukanoan languages have relatively impoverished systems of nominal inflectional morphology, but display complex nominal derivational morphology and extensive noun classifier systems. The size of Tukanoan classifier systems – some languages in the sample have more than 100 classifiers – makes the history of classifiers too large a topic for an essay of this scope. Additionally, in many grammars of Tukanoan languages, the discussion of classifiers either is not exhaustive (not listing all of the classifiers which exist in the language) or does not fully describe the grammatical behavior of classifiers. Classifier systems are also subject to heavy analogical pressure and pressure from iconicity, which tend to accelerate the progress of free nouns down the grammaticalization cline (Aikhenvald 2000: 372). These factors make nominal morphology in Tukanoan diachronically unstable, and would introduce significant noise into any classification of the family based on it. No such problems exist in the domain of finite verbal inflections and subordinators, which are relatively few in number per language, appear to be fully described in the available grammars, and are less subject to the semantic pressures which induce change in classifiers.

5.2 Morphological profile

Before the results of the morphological analysis are presented, some background on verbal inflectional morphology Tukanoan and especially WT languages is necessary. In general, Tukanoan languages obligatorily mark verbs for subject person, number, and gender; TAM; and evidentiality. Marking of evidentiality (a grammatical category which conveys the speaker's source of information for an utterance) is typically obligatory and fusional with TAM marking. Almost all verbal morphology is suffixing, and many verbal suffixes, especially derivational suffixes, are transparently derived from free verbs. This pattern likely reflects a cycle of grammaticalization from free verb to verbal suffix via serial verb constructions of the form ROOT1-ROOT2-INFLECTION. Serial verbs of this form occur in all Tukanoan languages.

On inflected verbs, TAM or fusional tense-evidential markers always follow any derivational suffixes, and are followed by subject person-number-gender markers (if these are not fused with the morphemes marking TAM and evidentiality). Finite subject person-number-gender markers exhibit extensive syncretism, especially between first and second person paradigms, across the sample of languages. Syncretism of finite inflectional markers is somewhat more extensive in the languages which are classified (in 4.4) as WT. On non-finite verbs, evidentiality is not usually marked, and the verbal subordinator immediately follows any derivational suffixes. Verbal subordinators are almost always morphologically nominalizers, and usually carry TAM information in themselves. In some languages in the sample, the subordinating nominalizers fusion ally mark subject number and gender. In others, additional morphology is suffixed to the nominalized verb to mark subject number and gender. Figures 5.2 and 5.2 graphically represent the main types of affix ordering across the sample of languages.

Finite	Root-	(Srlzr-Root-)	Neg-	Deriv-	Neg-	TAM-Evidential	Subject.Agreement
Nonfinite	Root-	(Srlzr-Root-)	Neg-	Deriv-	Neg-	Tense-Nmlzr	Nominal.Number-Gender

Figure 2 Less fusional template for the Tukanoan verb

Finite	Root-	(Srlzr-Root-)	Neg-	Deriv-	Neg-	TAM-Evidential-Subject.Agreement
Nonfinite	Root-	(Srlzr-Root-)	Neg-	Deriv-	Neg-	Tense-Subject.Number.Gender-Nmlzr

Figure 3 More fusional template for the Tukanoan verb

There are also differences among Tukanoan languages in the TAM and evidential categories morphologically marked on the verb. These differences are especially significant within WT. Koreguaje is the only Tukanoan language which does not mark tense on finite verbs, and also the only one which marks evidentiality using an auxiliary verbal construction rather than with a verbal suffix. Similarly, Máíhɨ̃ki is unique within Tukanoan in that it has no grammaticalized evidentiality. It is also the only WT language with a distinct future tense.

5.3 Morpheme correspondences

Applying the LCM to the WT languages in the morphological data matrix yielded one correspondence between nominalizing morphemes, and five sets of correspondences between paradigms of fusional morphemes marking TAM and subject person, number, and gender. These correspondences involve the simultaneous-subordinate-clause marker; the finite past and present declarative and interrogative markers for the default verb conjugation class; and the finite past declarative markers for a smaller verb conjugation class. Tables 9 - 14, and the associated discussion, describe these relations among the WT languages. Except in the case of the same-subject subordinate clause markers and the very similar present interrogative markers, no morphemes corresponding to these items appear in any ET language in the sample. As in §4.2, each row in the tables includes the candidate cognate morphemes from each WT language (where they exist), a gloss for this set of morphemes, and a reconstruction of the form of the morpheme in PWT. The reconstruction column is headed "PWT" rather than "Reconstruction" in order to reflect the results of §4.4.

PWT	Máí	Sio	Sek	Kor	Gloss
*-kɨ	-kɨ	-gɨ	-ɨ	-k^{h}ɨ	M.SG
*-ko	-ko	-go	-o	-k^{h}o	F.SG
*-hɨ̃	-hɨ̃	-hɨ̃	-hɨ̃	-hɨ	PL

Table 9 Simultaneous subordinate clause markers in WT

5.3.1 Simultaneous-clause nominalizers

The clearest correspondences exist in the set of nominalizers which mark the verb in a same-subject simultaneous clause. A same-subject simultaneous clause is a subordinate clause which denotes an action simultaneous with the action of the main clause, and shares the subject of the main clause. (7) provides an example of this construction in narrative speech in Máíhɨ̃ki.

(7) *gónó ű́kúkɨ á̋ʧi*

gónó ű́kú -kɨ á̋ -ʧi
manioc.beer drink -M.SG.SIMUL eat -M.SG.FUT.NI

"Tomando masato, (lo) voy a comer."
"I will eat (it) while drinking manioc beer."
(Michael et al. 2012a, *El antepasado que tomó masato de San Juan*, line 15)

In this paradigm, the nominalizers used with masculine singular and feminine singular subjects are extremely similar acoss the WT languages. The reflexes in Máíhɨ̃ki, Koreguaje, and Siona differ only in the laryngeal features of the initial consonant, and the reflexes in Secoya differ from those in the other three languages only in that the initial consonant of the marker is deleted. Furthermore, the nominalizer used with a plural subject is exactly the same in all four languages, except that the vowel is oral in Koreguaje and nasal in Máíhɨ̃ki, Siona, and Sekoya. The form of this nominalizer in PWT must therefore have been either *-hɨ or *-hɨ̃. I have listed *-hɨ̃ in Table 9 because it is more frequent in the data, but it is impossible to determine whether the Napo languages inherited a nasal vowel in this morpheme

from PWT, or whether they innovated the nasality after diverging from Koreguaje. In either case, the contrast in nasality offers the first piece of morphological evidence for the status of Máíhɨ̃ki, Siona, and Sekoya as a subgroup.

Reconstruction of the proto-morphemes for masculine-singular and feminine-singular subject markers is more difficult. Since the cognate sets for these two markers exhibit the same sound correspondences, it is almost certain that the masculine and feminine forms began with the same consonant in PWT. Yet the correspondence between /k/ in Máíhɨ̃ki, /g/ in Siona, /k^h/ in Koreguaje, and ∅ in Sekoya is not present in any of the phonological correspondence sets discussed in §4.2. This requires us to reconstruct the proto-morphemes underlying the correspondence using the morphological data presented in Table 9 and subsequent tables, together with circumstantial evidence from the phonological analysis.

Based on that evidence, I tentatively reconstruct *k as the initial consonant in the two singular simultaneous-clause markers. This consonant was clearly a velar stop, which under this paper's analysis means that it must be /k/, /g/, or /k^w/. For morphology-internal reasons, /g/ is not an appealing reconstruction here. Tables 11 and 14 indicate a correspondence between Máíhɨ̃ki /g/, Siona /g/, and Sekoya /k/ in bound morphemes, which this data does not exhibit. Additionally, reconstructing *g in these morphemes would entail reconstructing three sound changes which are not attested in the phonological analysis: deletion of *g in Sekoya, devoicing in Máíhɨ̃ki, and *g > *k^h in Koreguaje. Similarly, Table 12 in this section, and the correspondence exemplified by HEAVY in Table 2 in §4.2, indicate that *k in the proto-language was generally inherited as /k/ in the Napo languages. Reconstructing *k in these morphemes would therefore require us to posit two exceptional sound changes, voicing in Siona and deletion in Sekoya, and would follow two regular sound changes, *k > /k/ in

Máíhɨ̃ki and *k > /k^{h}/ in Koreguaje.

This makes *k^{w} a tempting reconstruction for the initial consonant in the masculine-singular and feminine-singular simultaneous clause markers. Reconstructing *k^{w} requires reconstructing at least one sound change in each of the attested languages. Specifically, this analysis entails the claims that the proto-segment delabialized in Máíhɨ̃ki, voiced and delabialized in Siona, lenited to zero in Sekoya, and underwent the regular sound change to /k^{h}/ in Koreguaje. Although this is a large number of sound changes, all of the changes in question are phonetically natural and appear for some velar stop, if not specifically for *k^{w}, somewhere in this analysis. Additionally, reconstructing *k^{w} as underlying these two correspondence sets would not require us to posit that the same exceptional sound change from *k or *g yielded both the correspondence found in this table and those in Table 13.

On the other hand, the simultaneous-clause markers listed above have clear cognates in nominalizing and subordinating morphemes in the non-WT Tukanoan languages (Farmer 2011: 58). These cognate morphemes all have the shape *-gɨ̃*, *-go* (in Desano and Barasana) or *-kɨ̃*, *-ko* (in Kubeo, another paradigm in Barasana, Kotiria, and Tukano). Therefore, the most parsimonious choice is to reconstruct a non-labialized velar stop as the initial consonant of the proto-form underlying the masculine and feminine markers shown in Table 9. As discussed above, we achieve the minimum number of exceptional sound changes, and achieve the maximum similarity between the PWT reconstruction and the most likely reconstruction for Proto-Eastern Tukanoan, by reconstructing *k. Yet given the dissimilarity of the proposed morphological correspondences here with the regular sound correspondences for *k shown in Table 2, we should regard the reconstruction of *k in these morphemes as fairly tentative.

5.3.2 Present declarative inflections

PWT	Máí	Sio	Sek	Gloss
*-hi	-hɨ̃	-hi	-hi	3SG.M.PRES.DECL
*-ko	-ko	-go	-ko	3SG.F.PRES.DECL
*-jɨ	-ji	-jɨ	-jɨ	3PL+NON3.PRES.DECL

Table 10 Finite present declarative TAM-subject agreement markers in WT

We now turn to the five finite verbal inflectional paradigms which exhibit robust correspondences across the WT languages. Table 10 lists the fusional morphemes which mark present tense, declarative modality, and subject person, number, and gender on regular verbs in three WT languages. (Koreguaje does not appear in this data set because it does not mark tense on verbs, and its subject agreement markers do not appear to be cognate with those in any other Tukanoan language.) As with the data in Table 9, the morphemes shown here are clearly cognate. Each correspondence set displays the same (or phonetically extremely similar) onset across the row, and the vowels are also relatively similar.

The proto-form for the third-person masculine singular marker in Table 10 can be confidently reconstructed with an initial *h, but the choice between *ɨ̃ and *i for the vowel is arbitrary. I reconstruct *i because the Upper Napo languages are generally more phonologically conservative than Máíhɨ̃ki. For the third-person feminine singular marker, I reconstruct *k rather than *g on the basis of evidence that *g in the proto-language uniformly yielded /g/ in Máíhɨ̃ki (cf. the discussion of *k^{w} above). This reconstruction does entail reconstructing an exceptional change of *k to /g/ for the third-person feminine singular marker in Siona. However, analogy due to the use of *-go* as a nominalizer and classifier (for nouns denoting feminine beings) in Siona could easily explain this change. Finally,

we find strong evidence for the consonant but inconclusive evidence for the vowel in the proto-form of the marker which is inherited as *-ji* in Máíhɨ̃ki and *-jɨ* in the Upper Napo languages. I reconstruct *-jɨ for this morpheme on the grounds that the non-third-person subject agreement marker may have grammaticalized from the first-person singular pronoun early in the development of WT (Stephanie Farmer, p.c.). The first-person pronoun is [jɨ] or [jɨʔɨ] in most ET languages and all WT languages except Máíhɨ̃ki. This suggests that both the pronoun [ji] and the verbal suffix [-ji] may be innovations internal to Máíhɨ̃ki.

5.3.3 Past declarative inflections

PWT	Máí	Sio	Sek	Gloss
*-bi	-gɨ	-bi	-pi	3SG.M.PST.DECL
*-go	-go	-go	-ko	3SG.F.PST.DECL
*-wɨ	-bɨ	-wɨ	-wɨ	3PL+NON3.PST.DECL

Table 11 Finite past declarative TAM-subject agreement markers in WT

Continuing with finite declaratives, Table 11 lists the morphemes which mark declarative past TAM and subject agreement for the default verb conjugation class. As in Table 10, the paradigm has two distinct forms marking agreement with third-person-singular feminine and masculine subjects, and a single syncretic form for all other combinations of subject person, number, and gender. All ET languages except Kotiria display this same alignment of subject marking, contrasting two third-person-singular markers with one non-third-person marker in the finite declarative paradigm.[9] Morphological marking of subjects on declarative verbs, with a 3-NON3 contrast, can therefore be reconstructed at least to PWT and perhaps to PT.

These forms are clearly cognate, and reconstructing the proto-morphemes underlying them

9 Kotiria does not exhibit this alignment for finite declarative verbs because it has morphological marking of subjects only on nonfinite and irrealis verbs.

is a relatively straightforward matter. The proto-form for the third-person masculine singular marker was likely *-bi. In Siona, this marker was inherited without change; in Sekoya, the initial *b merged with /p/ due to the loss of the voicing distinction. The Máíhɨ̃ki marker *-gɨ* here is not cognate with the equivalent morphemes in Sekoya and Siona. Rather, *-gɨ* is probably the outcome of analogy between this paradigm and non-finite paradigms in Máíhɨ̃ki, where *-gɨ* and *-go* are frequently used as masculine and feminine nominalizers and noun classifiers. The alternative to this account is to posit that Máíhɨ̃ki preserved the inherited third-person masculine singular marker, while Proto-Upper Napo innovated *-bi. While it is possible that Siona and Sekoya are innovative here, reconstructing a regularity-decreasing change in these languages is less well-motivated than reconstructing a paradigm-leveling change in Máíhɨ̃ki for this item. Siona and Sekoya also show no other signs of innovation in the paradigm, lending further support to the hypothesis that the *-bi/*-pi form is archaic. In the other forms, the only differences between the three languages' reflexes of the third-person singular feminine marker and the non-third person marker are instances of regular sound change. The third-person singular feminine form was clearly *-go, which was inherited without change in Siona and Máíhɨ̃ki and underwent only devoicing, due to the loss of the voicing distinction, in Sekoya. Similarly, the non-third person singular marker must be reconstructed as *-wɨ. This marker was inherited without change in Siona and Sekoya and underwent the regular PWT *w > b change in Máíhɨ̃ki.

5.3.4 Past declarative inflections, *ni*-class verbs

All of the WT languages have two verb conjugation classes. The "default" class contains the great majority of verbs, and displays no root allomorphy, while the "*ni*-class" contains

PWT	Máí	Sio	Sek	Gloss
*-hiʔi	-kɨ̀	-híʔí	-híʔí	3SG.M.PST.DECL-NI
*-koʔi	-kò	-kóʔí	-koʔi	3SG.F.PST.DECL-NI
*-ʔɨ	-hɨ̀	-ʔɨ́	-ʔɨ	3PL+NON3.PST.DECL-NI

Table 12 Finite past declarative TAM-subject agreement markers for *ni*-class in WT

a small number of verbs which have roots of the form CVi- in citation form. These verbs display extensive root allomorphy – in Máíhɨ̃ki, between CVi-, CV-, CVV-, and CVni- forms (from which the conjugation class takes its name) – and also display a different pattern of subject marking for certain values of TAM, including the finite past declarative. Table 12 lists the subject agreement markers which appear on *ni*-class verbs in this paradigm.

The paradigm in Table 12 presents greater analytical challenges than those previously discussed. The first issue for reconstruction of this paradigm is the presence of /ʔ/ in the Siona and Sekoya forms. As Table 2 indicates and Farmer (2012) argues, PWT */ʔ/ was preserved without change in Siona and Sekoya, but inherited as low tone on the following vowel (with no segmental reflex) in Máíhɨ̃ki. If we accept this account, then the third-person feminine singular form here can be reconstructed to PWT as *-koʔi. This reconstruction implies that Siona and Sekoya have the archaic form, while the Máíhɨ̃ki form reflects the outcome of deletion of the glottal stop and the related low tone tonogenesis (*-koʔi > *-koì) followed by vowel coalescence (*-koì > *-kò).

By the same reasoning, we can also reconstruct a form ending in *-ʔV for the third-person masculine singular marker in this paradigm. This item could plausibly have been either *-hiʔi, if the Siona and Sekoya forms are archaic, or *-kɨʔi, if the Máíhɨ̃ki form is archaic. Here as in the regular finite past declarative paradigm, analogy with nominalizers and noun classifiers of the forms *-kɨ* and *-ko* can cleanly explain the innovation of *-kɨ* in this paradigm in

Máíhɨ̃ki, while no such explanation would be available for supposed innovation of *-hi* in Siona and Sekoya. I therefore reconstruct *-hiʔi rather than *-kɨʔi for the third-person masculine singular form here, and suggest that Máíhɨ̃ki innovated *-kɨ́*, under analogical pressure from the nominal paradigm, after the feminine equivalent of this form emerged as *-kò*.

Farmer's (2012) account of tonogenesis suggests that the proto-form of the non-third person marker ended with *-ʔV as well. The complete segmental form of this morpheme is more difficult to reconstruct, since the Siona and Sekoya forms suggest that it was *-ʔɨ, while Máíhɨ̃ki could imply *-hɨʔi (parallel with the singular markers) or *-hɨʔɨ. One plausible account for the development of the correspondence set shown here begins with the premise that the morpheme was *-ʔɨ in PWT. After the divergence of the Máíhɨ̃ki subgroup, the glottal stop in this form was lost and low tone tonogenesis occurred, yielding *-ɨ̀ for the non-third person marker, and *-hì and *-kò for the two third-person markers, in this paradigm. One of two changes then took place. Either analogy with the third-person masculine singular marker *-hì caused the non-third person marker to change from *-ɨ̀ to *-hɨ̀*, or the third-person masculine singular marker and the non-third person marker syncretized to *-hɨ̀*. After this change, the analogy discussed above caused the third-person masculine singular marker to become *-kɨ̀*, obscuring the path of change which resulted in *-hɨ̀*. Although this account involves a sequence of several analogical changes, it is more parsimonious than reconstructing *-hɨʔɨ as the proto-form, which would beg the question of what phonological or morphological pressure in Siona and Sekoya led to the loss of intervocalic /h/ in this context alone.

PWT	Máí	Sio	Sek	Kor	Gloss
*-kɨ	-kɨ	-gɨ	-ɨ	-k^{h}ɨ	NON1.SG.M.PRES.INT
*-ko	-ko	-go	-o	-k^{h}o	NON1.SG.F.PRES.INT
*-je	-je	-je	-je ∼ -ɲe	-je	PL+1SG.PRES.INT

Table 13 Interrogative present TAM-subject agreement markers in WT

5.3.5 Present interrogative inflections

We now turn to the interrogative paradigms. Table 13 lists the subject agreement markers used with default-class verbs in the interrogative modality. In all four WT languages considered here, the interrogative paradigm exhibits a split in subject marking between first-person and non-first person subject markers. This alignment is both different from that found in the declarative paradigm, which contrasts third- and non-third-person arguments, and from the typological norm for subject-marking on the verb. While 3/NON3 is a typologically common split in marking for languages with obligatory morphological marking of arguments on the verb, 1/NON1 is a much less common split. Additionally, no Eastern Tukanoan language in the morphological data matrix exhibits such a split in subject marking for the interrogative paradigm. The 1/NON1 split found in Tables 13 and 14, then, provides reasonably strong evidence for the shared development of the WT languages.[10]

The forms in Table 13 bear a strong resemblance to the simultaneous-clause markers listed in Table 9. The only difference between these paradigms is in the morpheme used to mark agreement with a plural subject, which is *-je* in the present interrogative paradigm but *-hɨ̃* in the paradigm for simultaneous-clause marking. Given the high degree of syncretism between these paradigms, any substantial explanation of the reconstructions in Table 13 would be

10 However, the 1/NON1 split cannot be adduced as evidence for the status of WT as a clade. Absent a reconstruction of the PT interrogative paradigm as lacking a 1/NON1 split, we do not know if the split is a shared retention from PT in PWT or a shared innovation in WT.

redundant. The forms and reconstructions for the masculine- and feminine-gendered markers are exactly the same here and in Table 10, and *-je is the only reasonable reconstruction for the default marker (for first-person singular and all plural subjects). The segmental similarity between this paradigm and the simultaneous-clause marker paradigm may itself reflect a shared diachronic origin for both suffixes in nominalizers and/or classifiers used with nominalizing function. Máíhɨ̃ki morphophonology provides some evidence for this theory, since in that language both the simultaneous-clause marker and the interrogative markers cause the verb to display the tonal behavior of a noun (in exactly the same way as dedicated nominalizers).

5.3.6 Past interrogative inflections

PWT	Máí	Sio	Sek	Gloss
*-gɨ	-gɨ	-gɨ	-kɨ	NON1.SG.M.PST.INT
*-go	-go	-go	-ko	NON1.SG.F.PST.INT
*-ɾe ∼ *-te	-ɾe	-ɾe ∼ -te	-ɾe ∼ -te	PL+1SG.PST.INT

Table 14 Interrogative past TAM-subject agreement markers in WT

Table 14 lists the subject agreement markers for past tense interrogative verbs, which comprise the final set of correspondences in TAM-subject marking that this essay will discuss. Here as in Table 13, there are no contrastive differences in segmental form between the three languages' reflexes of this paradigm. For the two non-first-person markers, *-gɨ and *-go are the only reasonable reconstructions. These forms are inherited without change in Máíhɨ̃ki and Siona, and undergo the regular devoicing due to merger in Sekoya. The first-person and plural marker could be reconstructed either as *-ɾe or as varying freely between *-ɾe and *-te, as in Siona and Sekoya. I prefer *-ɾe ∼ *-te on the principle that irregularity and

free variation in morphology should generally be reconstructed to the proto-language (first articulated by Meillet 1925).

5.4 Internal classification from morphological data

Koreguaje, Máíhɨ̃ki, Siona, and Sekoya share a number of morphological features which are not found in any other Tukanoan language. These include a verbal conjugation class displaying root allomorphy, a set of lexicalized causatives marked by vowel mutation, non-fusional marking or zero marking of evidentiality on the verb, and a split between first-person and non-first-person subjects in the subject-marking paradigm for interrogative verbs. Yet as mentioned in §2.3, there has not yet been a systematic attempt to reconstruct bound morphemes to PT. Morphological reconstruction of PT is beyond the scope of this study, and perhaps – given the low volume and quality of data on the ET languages – of any study. In the absence of such a reconstruction of PT, it is impossible to determine which features of WT are innovative and which were retained from PT. Since only shared innovations can prove the existence of a clade, this entails that morphological features shared by all and only the WT languages, no matter how robust, cannot provide evidence for the WT clade.

On the other hand, the results of §5.3 do strengthen the phonological analysis, displaying almost exactly the same sound changes observed in the lexical data. The data in Table 11 confirm that the PWT *w > Máíhɨ̃ki /b/ sound change also operated on bound morphemes, while Table 12 provides further evidence for the account of low tone tonogenesis proposed by Farmer (2012). The only correspondence set which appears only in the morphological data is the correspondence, in the present declarative and interrogative paradigms, between Máíhɨ̃ki /k/, Siona /g/, Koreguaje /k^h/, and ∅ in Secoya. PWT *k is reconstructed for this

correspondence for the reasons discussed in §5.3.1.

§5.3 also yields significant evidence for the internal classification of WT. The absence of morphological marking of TAM in Koreguaje means that the morphological analysis provides minimal positive evidence for the position of that language within the clade. On the other hand, the absence of marking also strongly suggests that Koreguaje is less closely related to any of the other WT languages than they are to one another – supporting the conclusion of §4.4 that Koreguaje is the first-diverging language in WT. Morphological reconstruction based on data from Siona, Sekoya, and Máíhɨ̃ki also produces an internal classification congruent with the phonology-based tree. Máíhɨ̃ki is clearly morphologically innovative relative to Siona and Sekoya. It was subject to paradigm leveling (§5.3.3) and analogy (§5.3.2) not found in Siona or Sekoya, and also experienced a unique process of low-tone tonogenesis and consequent analogy following the loss of PWT *ʔ (§5.3.4). Máíhɨ̃ki also differs from Siona and Sekoya in that it has a distinct future tense. Similarly, Siona and Sekoya differ from Máíhɨ̃ki in that they have morphological marking of evidentiality on the verb. Both languages contrast unmarked evidentiality with hearsay evidentials, and in both the hearsay evidential is marked with an agglutinative verbal suffix /-ɲa/.

In sum, Máíhɨ̃ki has numerous innovations – regular sound changes, analogical changes in paradigms, and the future tense – which are not shared by Siona and Sekoya. This confirms the result of §4.4 that Máíhɨ̃ki diverged early from the Napo subgroup, leaving Siona and Sekoya to form an Upper Napo clade. The morphology-based internal classification of WT is therefore entirely congruent with the phonology-based classification, and both can be graphically represented by Figure 4 (identical to Figure 1 in §4.4).

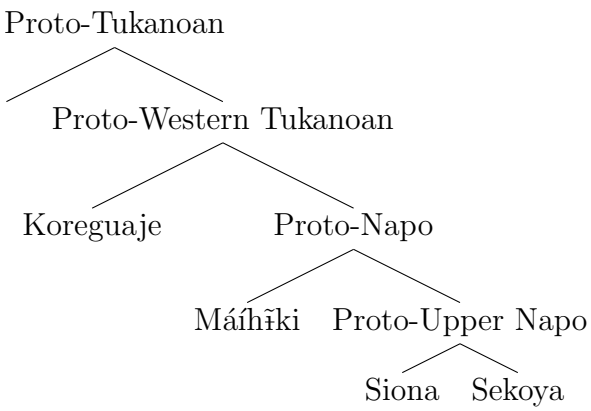

Figure 4 Internal classification of WT from morphological evidence

6 Conclusions

The phonological evidence discussed in §4 strongly supports the existence of a Western Tukanoan clade consisting only of Koreguaje, Máíhɨ̃ki, Siona, and Sekoya. This clade is defined by two sound changes: PT *p > PWT *h, and PT *ɾ ∼ *d > PWT *ɖʒ ∼ *j / _i. It is further characterized by correspondences relating WT /s/ to ET /j/ and WT /j/ to ET /s/. Significantly, this phonological definition excludes Kubeo, a VLA language which has sometimes been classified as WT, from the WT clade. The WT languages also display a number of shared morphological features which are not found in other Tukanoan languages. These WT-unique features include the six cognate paradigms for verbal inflection listed in §5.3; a second verb conjugation class (the *ni*-class); a set of lexicalized causatives marked by vowel mutation; and the absence of an elaborate system of grammaticalized evidentiality. The distribution of these features within Tukanoan is consistent with the existence of a WT clade. However, for the reasons discussed at §5.4, it does not provide evidence for a clade.

Phonological and morphological evidence yield congruent internal classifications of the languages within the WT clade. Both the phonological and morphological analyses indicate

that Koreguaje is the most divergent language within WT. In phonology, Koreguaje differs widely from the other languages in phonemic inventory. It is the only WT language with contrastive aspiration and exhibits several other contrastive phonemes – including a series of voiceless nasals, /f/, and /v/ – which are not found in any other Tukanoan language. Koreguaje also displays much less extensive verbal inflectional morphology than the other languages in the family, with zero morphological marking of TAM on finite verbs. Its main finite verbal inflection is a series of subject-agreement markers, none of which appear to be cognate with morphemes of the same function in any other Tukanoan language. These facts lead us to place Koreguaje high in the tree, as the first language to diverge from PWT.

Positing Koreguaje as the first-diverging lineage leads us to classify the three other WT languages under consideration into a subgroup. Máíhɨ̃ki shows clear evidence of phonological and morphological innovation relative to the other two languages in this Napo group. The primary phonological innovation is a sound change from PWT and Proto-Napo *w > Máíhɨ̃ki /b/, found in both free words and bound verbal morphemes. The evidence for morphological innovation in Máíhɨ̃ki is more extensive, including a number of analogical changes in verbal inflection and the innovation of a future tense. We therefore subgroup Máíhɨ̃ki away from Siona and Sekoya, and place those two languages in a distinct Upper Napo subgroup. Table 15 summarizes the evidence for each subgroup proposed here and represented in Figure 4.

Several questions remain for future historical research on Tukanoan and especially WT. One major issue for further research is that the absence of any reconstruction of PT morphology currently precludes us from identifying any shared morphological features of WT as evidence for a WT clade. Reconstruction of PT morphology, perhaps beginning with the finite verbal inflections discussed here, will be necessary for any more complete morphology-based internal

WT	Koreguaje	Máíhɨ̃ki	Upper Napo
PT *p > h sound change	Phonemic aspiration	PWT *b > w sound change	*-ɲa* hearsay evidential
Palatalization of PT *ɾ ∼ *d / _i	Phonemic bilabial fricatives	Low tone tonogenesis	
	Phonemic voiceless nasals	Analogy in present and past declarative paradigms	
	No morphological marking of TAM	Future tense	

Table 15 Evidence for the WT clade and its subgroups

classification of Tukanoan. Many phonological and morphological phenomena within WT also await historical-comparative analysis. Verbal derivational morphology and nominal morphology, not discussed in this essay, may yield new evidence for the internal classification of WT. More complete documentation of tone systems in WT languages other than Máíhɨ̃ki is also likely to produce data which can deepen our understanding of WT phonology within Tukanoan and improve the phonology-based internal classification of WT.

References

Aikhenvald, Alexandra. 2000. *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.

Aikhenvald, Alexandra and R.M.W. Dixon. 1998. Evidentials and areal typology: A case study from Amazonia. *Language Sciences* 20. 241–257.

Alemán, Tulio, Reinaldo Lopez and Marion Miller. 2000. *Wira ya, peamasa ya wererituri (diccionario desano-espanol)*. Bogotá: Editorial Alberto Lleras Camargo.

Barasana Literacy Committee, Paula Jones and Wendell Jones. 2009. *Diccionario bilingüe: Eduria and barasana-español, español-eduria and barasana*. Bogotá: Fundación para el Desarrollo de los Pueblos Marginados.

Barnes, Janet. 1984. Evidentials in the Tuyuca verb. *International Journal of American Linguistics* 50(3). 255–271.

Barnes, Janet and Cesar Tamayo. 1988. *Mi primer diccionario español-tuyuca, tuyuca-español*. Bogotá: Summer Institute of Linguistics.

Bowern, Claire, Bethwyn Evans, Grace Koch and Luisa Miceli. 2008. Introduction. In Claire Bowern, Bethwyn Evans, Grace Koch and Luisa Miceli (eds.), *Morphology and language history: in honour of Harold Koch*, 1–14. Amsterdam: John Benjamins Publishing Company.

Brinton, Daniel G. 1892. *Studies in South-American native languages*. Philadelphia: Proceedings of the American Philosophical Society.

Campbell, Lyle and William J. Poser. 2008. *Language classification: history and method*. Cambridge: Cambridge University Press.

Chacón, Thiago. 2012. *The phonology and morphology of Kubeo: The documentation, theory, and description of an Amazonian language*. Ph.D. thesis, University of Hawaii at Manoa.

Chacón, Thiago. forthcoming. A revised proposal of Proto-Tukanoan consonants and Tukanoan family classification. *International Journal of American Linguistics* forthcoming.

Chernela, Janet. 1989. Marriage, language, and history among Eastern Tukanoan speaking peoples of the Northwest Amazon. *Latin American Anthropology Review* 1(2). 36–41.

Cook, Dorothy M. and Linda L. Criswell. 1993. *El idioma koreguaje (tucano occidental)*. Bogotá: Asociación Instituto Lingüístico de Verano.

Cook, Dorothy M. and Frances L. Gralow. 2001. *Diccionario bilingue koreguaje-español*. Bogotá: Editorial Alberto Lleras Camargo.

Criswell, Linda L. and Beverly Ann Brandrup. 2000. Un bosquejo fonológico y gramatical del siriano. In María Stella González de Pérez and María Luisa Rodríguez de Montes (eds.), *Lenguas indígenas de Colombia: una visión descriptiva*, 395–417. Bogotá: Instituto Caro y Cuervo.

Dixon, R.M.W. 1980. *The Australian languages*. Cambridge: Cambridge University Press.

Epps, Patience. 2005. Areal diffusion and the development of evidentiality: Evidence from Hup. *Studies in Language* 29. 617–50.

Epps, Patience. 2008. Grammatical borrowing in Hup. In Yaron Matras and Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 551–565. Berlin: Mouton de Gruyter.

Epps, Patience. 2009. Language classification, language contact, and Amazonian prehistory. *Language and Linguistics Compass* 3(2). 581–606.

Epps, Patience. 2011. Phonological diffusion in the Amazonian Vaupés. Presentation at 2011 CUNY Phonology Forum Conference on Phonology of Endangered Languages.

Farmer, Stephanie. 2011. Tukanoan verbal morphological categories. Unpublished dissertation prospectus, University of California - Berkeley.

Farmer, Stephanie. 2012. Low tone tonogenesis in Máíhɨ̃ki. Talk given at the 2012 Winter Meeting of the Society for the Study of the Indigenous Languages of the Americas.

Farmer, Stephanie, Christine Beier, John Sylak and Lev Michael. forthcoming. Máíhɨ̃ki phonology in comparative Tukanoan perspective. *International Journal of American Linguistics* forthcoming.

Gomez-Imbert, Elsa. 1988. Construcción verbal en barasana y tatuyo. *Amerindia* 13. 97–108.

Gomez-Imbert, Elsa. 2003. Une langue de nord-ouest Amazonien: le barasana. *Faits de langues* 21.

Gomez-Imbert, Elsa and Michael Kenstowicz. 2000. Barasana tone and accent. *International Journal of American Linguistics* 66. 419–463.

Jackson, Jean. 1983. *The fish people: linguistic exogamy and Tukanoan identity in Northwest Amazonia*. Cambridge: Cambridge University Press.

Johnson, Orville, Stephen Levinsohn and Alva Wheeler. 1990. *Gramatica secoya*. Quito: Instituto Lingüístico de Verano.

Jones, Wendell and Paula Jones. 1991. *Barasano syntax*, Studies in the Languages of Colombia. Dallas: Summer Institute of Linguistics.

Klumpp, James and Deloris Klumpp. 1973. Sistema fonológico de piratapuyo. In *Sistemas*

fonológicos de idiomas colombianos, 107–120. Bogotá: Ministerio de Gobierno.

Malone, Terrell. 1988. The origin and development of Tuyuca evidentials. *International Journal of American Linguistics* 54(2). 119–140.

Mason, J. Alden. 1950. The languages of South American Indians. In *Handbook of South American Indians, vol 6*, 157–317. Washington DC: Government Printing Office.

McConvell, Patrick. 2009. Loanwords in Gurindji, a Pama-Nyungan language of Australia. In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the world's languages*, 790–822. Berlin: De Gruyter Mouton.

Meillet, Antoine. 1925. *La méthode comparative en linguistique historique*. Paris: Champion.

Metzger, Ronald G. 2010. *Diccionario carapana-español*. Bogotá: Editorial Alberto Lleras Camargo.

Michael, Lev, Christine Beier, Stephanie Farmer and Amalia Skilton. 2012a. Northern and western Máíhɨ̃ki narrative text database. Máíhɨ̃ki Project internal document.

Michael, Lev, Christine Beier, Stephanie Farmer, John Sylak, Greg Finley, Kelsey Neely, Amalia Skilton and Grace Neveu. 2012b. Diccionario bilingue máíjɨ̱ki-castellano, castellano-máíjɨ̱ki, borrador de agosto 2012. Máíhɨ̃ki Project internal document.

Miller, Marion. 1999. *Desano grammar*, Studies in the Languages of Colombia. Dallas: Summer Institute of Linguistics.

Morse, Nancy, Jay K. Salser Jr. and Neva de Salser. 1999. *Diccionario cubeo-español*. Editorial Alberto Lleras Camargo.

Morse, Nancy and Michael Maxwell. 1999. *Cubeo grammar*, Studies in the Languages of Colombia. Dallas: Summer Institute of Linguistics.

Neely, Kelsey. 2012. Máíhɨ̃ki classifiers and bound part terms in comparison. Máíhɨ̃ki Project

internal document.

Piaguaje, Ramon, Elias Piaguaje, Orville Johnson and Maria Johnson. 1992. *Vocabulario secoya*. Quito: Instituto Lingüístico de Verano.

Ramirez, Henri. 1997a. *A fala Tukano dos Ye'pa-Masa (diccionario tukano-portugues)*, vol. tomo II. Manaus: Inspetoria Salesiana Missionária da Amazonia.

Ramirez, Henri. 1997b. *A fala Tukano dos Ye'pa-Masa (gramática)*, vol. tomo I. Manaus: Inspetoria Salesiana Missionária da Amazonia.

Silva, Wilson de Lima. 2012. *A descriptive grammar of Desano*. Ph.D. thesis, University of Utah.

Smothermon, Jeffrey R. and Josephine H. Smothermon. 1993. *Macuna-español diccionario de 850 palabras*. Bogotá: Editorial Alberto Lleras Camargo.

Smothermon, Jeffrey R., Josephine H. Smothermon and Paul S. Frank. 1995. *Bosquejo del macuna: aspectos de la cultura material de los macunas, fonología, gramática*. Bogotá: Asociación Instituto Lingüístico de Verano.

Sorensen, Arthur. 1967. Multilingualism in the Northwest Amazon. *American Anthropologist* 69(6). 670–684.

Stenzel, Kristine. 2004. *A reference grammar of Kotiria (Wanano)*. Ph.D. thesis, University of Colorado.

Stolte, Joel and Nancy Stolte. 1979. Barasano del norte. In *Aspectos de la cultura material de grupos etnicos de Colombia*. Summer Institute of Linguistics.

Strom, Clay. 1992. *Retuara syntax*, Studies in the Languages of Colombia. Dallas: Summer Institute of Linguistics.

Thomason, Sarah. 2012. When is language contact the best explanation for a linguistic

change? Lecture at Yale University, October 11, 2012.

Velie, Daniel. 1975. *Bosquejo de la fonología y gramática del idioma orejón (coto)*. Lima: Instituto Lingüístico de Verano.

Velie, Daniel and Virginia Velie. 1981. *Vocabulario orejón*. Lima: Ministerio de Educación and Instituto Lingüístico de Verano.

Waltz, Nathan. 2007. *Diccionario bilingue wanano-español*. Bogotá: Fundación para el Desarrollo de los Pueblos Marginados.

Waltz, Nathan and Alva Wheeler. 1972. Proto-Tucanoan. In Esther Matteson (ed.), *Comparative studies in Amerindian languages*, 19–49. The Hague: Mouton.

West, Birdie. 1980. *Gramática popular de tucano*. Bogotá: Ministerio de Gobierno.

West, Birdie and Betty Welch. 2004. *Gramática pedagógica del tucano*. Bogotá: Fundación para el Desarrollo de los Pueblos Marginados.

Wheeler, Alva. 1970. *Grammar of the Siona language, Colombia*. Ph.D. thesis, University of California, Berkeley.

Wheeler, Alva. 1987. *Ganteya bain: El pueblo siona del río Putumayo, Colombia*. Bogotá: Editorial Townsend.

Appendix: Lexical Correspondence Sets

See §4.1 for a discussion of the orthography used in the following cognate set. Reconstructions of PWT use the phonemic inventory stated in §4.3, except that hypothesized reflexes of PWT *ɾ / _i are written as *ʤ rather than *j for clarity.

GLOSS	KUB	DES	BAR	MAK	KOT	KAR	TUY	MAI	KOR	SEC	SIO	PWT
1sg.pro	jɨ	jɨʔɨ	jɨ	jɨ	jɨʔɨ	-	jɨɨ	ji	jɨʔɨ	jɨʔɨ	jɨʔɨ	*jɨʔɨ
2pl.pro	mɨha	mɨa	mɨ́á	mɨa	mɨsá	-	mɨ́a	mɨ́sáhuna	mɨsa	mɨsaru	mɨsakʷa	*mɨsa
achiote	muha-	mosã	musá	musa	buusi-	musã	-	mósá, bósá	põsa	põsa	bõsa	*mosa
agouti	buɨ	bui	búú	bu	bú	buu	-	-	-	wɨ	wɨ	*wɨ
ant sp	mea-	megã	meká	meka	-	mekã	mekãsía	méa	mekʰa	meka	mea	*mekʷa
armadillo	pãmu-	pãmu	hãmo	hãmo-	pʰamó	pãmo	pãmo	-	hamu	hãmu	hãmu	*hamu
back	-	sũgũ	hɨdóhɨ́a	hɨdo-	sukũ-	hõkɨ̃	sukubíro	hété-	sõkɨɲɨ	hete	gʷɨrɨbɨ	*hete
bat (gen)	ojo-	ojo	osó	oso	só	oso	oso	ójo	ojo	ojo	ojo	*ojo
big	-	wɨarii	haí-	hai-	pʰíri	pai-	-	háí-	ɨha-	hai	hai	*hai
black	ɲemi-	ɲi-	ɲi-	ɲi	ɲína	ɲii-	ɲii	néá-	jihaʔ-	nea-	sʔi	*nea-
black ink	wei	-	-	-	-	-	-	-	-	-	-	-
blood	hiwe	di	ríí	ri	dí	rií	díi	jie	jie	tsie	sʔie	*ʤie
breast	ope	ãpɨ̃-	õhɛ́	õhe-	pénó	õpɛ̃-	õpɛ̃ri	óhé-	ohe	ohe-	ohe	*ohe
buriti palm	nei	-	rẽe	rẽ	naʔáó	nee	-	-	neʔe	deʔreo	neʔe	*neʔe
caimo - fruit sp	toa	-	-	-	tóá-	-	-	tóa	tʰoaa	toa	toa	*toa
capybara	-	-	-	-	diá wááʧɨ́	-	-	júára	kʰuẽso	kueso	gueso	*gueso
centipede	-	-	muní	-	ɲaʔkí	muni	-	-	sɨsɨri	kate	-	-
charcoal 1	-	nitɨ̃	-	rɨ̃ti	niití	nitɨ̃	nitɨ̃	-	-	-	-	-
charcoal 2	neo	-	-	-	-	-	-	néhõ	neo	neo	neo	*neo
cheek	wio	waju-	-	waju-	waʔsú-	waso-	-	bájo-	ajo-	wajo-	wajo	*wajo
chili	bia	bia	bíá	bia	biá	bia	bía	bía	pia	pia	bia	*bia
cold	hɨhɨ	jɨsa-	-	ɨsa	jɨsiá	jɨsi-	jɨsió	sísí-	sɨsɨ-	sɨsɨ	sɨsɨ	*sɨsɨ
cotton	jɨi	jura	-	-	juutá	-	busa	jɨ́i	ʧɨi	jɨi	jɨi	*jɨi
dance, song	baja-	baja-	basá	basa-	basáa	basa-	basa	úhá	uha	-	-	*uha
deer	ɲama-	ɲama	ɲámá	ɲama	ɲamá	ɲama	ɲama	ɲáma	ɲámá	ɲama	ɲama	*ɲama
dove (sp)	-	buha	buhá	buha	sirípí	bua	bua	-	siʔi	suʔte	ʧiʔwi	-

GLOSS	KUB	DES	BAR	MAK	KOT	KAR	TUY	MAI	KOR	SEC	SIO	PWT
duck (gen)	bede	diakara	-	-	diá pʰúrú	riapotaka	díakata	-	pete	pete	bete	*bete
ear	kãmu-	gãmi-	ŋamó	gãmo	kʰaʔmó-	ãmoo-	kãmopéro	gãho-	kãho	kãho-	gãho	*gãho
egg	hĩdɨ	diu	ria	ria	dié	riá	dije	jíá	jia	tsia	sʔia	*ʤia
elder	bɨkɨ	bɨgɨ	bɨkɨ́	bɨkɨ	bɨɨkɨ́	bɨkɨ	bɨkɨ	áí	áí-	ai-	-	*ai-
elevated structure	kaja	kaja	kasá-	kasa-	kasá-	kasa-	kasa	-	kʰaja-	-	-	-
feces	kɨda	gɨra	gɨdá	gɨda	tɨ́	ɨta	-	ɨ̃ta	kɨta	ɨ̃ta	gʷɨta	*gɨta
face	hiwa	dia-	rio-	rio	wɨʔdɨ́á	ria-	-	jía	jíá	tsia	tɨʔribɨ	*ʤia
father	pakɨ	pagɨ	hakɨ́	hakɨ	pɨkɨ́	pakɨ	pakɨ	hakɨ	haʔkʰɨ	haʔkɨ	hakɨ	*haʔkɨ
fire, firewood	peka	pea	héá	hea	pitʃá	pee	peka	héka	hẽkʰa	heka	hẽka	*hẽka
fish (n)	-	waʔi	wai	wai	waʔí	wai	wai	bai	aʔi	waʔi	waʔi	*waʔi
fish with net (v)	-	waaja-	-	waja-	waʔsá-	weri-	-	-	ʧóó-	-	johi	-
fishing net	papi	wẽhẽ-	bahí	jori-	wãŋɲíno	bapii	bapigɨ	-	jorɨ	-	jorɨ	*jorɨ
flower (gen)	ko	go-	goó	go-	kʰoʔó	oó-	kooro	-	koro-	horo	koro	*koro-
foot	kɨbo	guubu-	gɨbó	gɨbo	daʔpó-	rɨpo	dɨpo	gɨ̃õ	kũʔa-	kɨ̃o	gʷɨõ	*gɨo
fruit sp (guaba)	mene	merẽ	méné	mene	mené	mene	-	méne	pene-	pẽne	bene	*mene
garden	-	poe	wese	wese-	weesé	-	wese	bese	eʔse	weʔse	weʔse	*weʔse
grandfather	ŋeku	ŋekɨ	ŋikɨ́-	ŋikɨ	ŋɨɨtʃí	ŋikɨ̃	ŋekɨ̃	ŋiku	kuʔeɨ	ŋekɨ	gʷʔegʷɨ	*ŋikɨ
grape	ɨjei	igi	ɨje	ɨje	sé-	ɨse	ɨse	éré	ɨje	-	ɨje	*ɨje
grass	-	taa	taa	ta	tʰá	taa	táa	tájá	taja	taja	taja	*taja
green, blue	hɨ̃me-	jahsari	sɨ̃me	sɨ̃me-	jaʔsá-	hɨ̃me-	-	míɲá	ĩha-	wĩɲa	sɨʔɲo-	*biɲa
guaracu fish	-	bodeka	bodéka	bodeka	boʔtéá	boteka	botea	-	-	-	-	-
hand	pɨdɨ	moho-	ámto	ãmo	wãmó	wãmo	-	hɨ̃tɨ	hɨ̃tɨ	hɨ̃tɨ	ɨ̃tɨ	*hɨ̃tɨ
head	hipo	dipu-	riho-	riho-	daapú	rɨpo-	dɨpoa	ʧõ-	sĩho	sĩo-	sĩho	*sĩho
heavy	dɨkɨ	nɨkɨ̃-	rɨ̃kɨ	rɨ̃kɨ	ŋikɨ́ nina	nɨkɨ̃-	-	dɨ́kɨ́-	rɨkʰɨ	dɨkɨ	rɨkɨ-	*ɨkɨ
heron sp	jai	jahi	jehe	jahe	jahá	jee	jee	-	kaoro	jai	-	-
hole	kobe	gobe	gohé	-	kopá	ope	kope	góhé	kohe	kohe	gohe	*gohe
heat	-	kũŋu-	ási-	asi	-	asi-	asi	atʃú	asú-	-	haʔsu-	*asu
house	-	wiʔi	wii	wi	wɨʔɨ́	wii	wii	we	ɨʔe	wɨʔe	wɨʔe	*wɨʔe

GLOSS	KUB	DES	BAR	MAK	KOT	KAR	TUY	MAI	KOR	SEC	SIO	PWT
humming-bird (gen)	mimi-	mimi	mímí	mimi	wĩmí	mimi	mimi	mími	mimi	mimi	mimi	*mimi
insect sp	huhi-	-	-	-	-	-	-	-	sũsi	-	susi	*sũsi
jaguar (gen)	jawi-	jee	jai	jai	jáí-	jai	jái	jáí	jai	jai	jai	*jai
kingfisher (gen)	hãha	sãrã	ɲasá	ɲasa	sãná	ɲasã	sãna	-	sãʔsa	sãʔsa	keʔre	*sãʔsa
lake	hita-	-	-	-	-	ɨtabɨkɨra	opataro	-	jiara	haira	sʔitara	*ʤiara
land	jeba	jeeba	-	sita	jaʔpá	jepa	dita	jíha	-	jeha	jiha	*jeha
larva	-	-	bekó	-	-	beko	-	-	pekʰo	peko	peko	*peko
leg	ɲɨa-	ɲigã-	ɲikɨ́-	ɲiki-	ɲɨɨʧíkɨ́	ɲikãã	ɲikã	gɨ̃só-	ɲukʰa	nɨka	gʷɨ̃so	*ɲɨka
long, far	hoa	-	-	-	joá-	joa-	joaro	so	soʔo	soʔo	sʔoa-	*so
macaw sp	ma	mahã	mahá	maha	mahá	maa	maa	má	maa	toa-ma	ma	*ma
man	ɨ̃mɨ̃	ɨ̃mɨ	-	ɨ̃mi-	míno	ɨ̃mɨ	ɨ̃mɨ	ímɨ	ɨmɨɨ	ɨ̃mɨ	ɨmɨ	*ɨ̃mɨ
manioc	kɨi	kɨ̃ĩ	kɨ̃	kɨ̃	kʰɨ́	kii	kii	-	kʰɨi	kɨi	kɨi	*kɨi
monkey (gen)	-	gaki	gaké	gake	ká	ake	ake	tãké	tãkʰe	tãke	tãke	*tãke
monkey sp	hihijo	-	isíkamɨ	-	sĩĩ	-	-	ʧíʧí	sisi-	-	sisi	*sisi-
mosquito	mɨre-	mɨrea-	mɨté	mɨte	mɨɨtéa	mɨtẽ	mɨtẽã	mɨ́te	mɨte	mɨte	mɨte	*mɨte
mouth	hihe	disi-	rise	rise	disé-	rise-	ɨsero	jo-	jɨʔopo	jɨʔo-	jiʔo	*ʤɨʔo
name	ãmi-	wãɨ̃	wãme	wãme	wãmá	wãme	wãme	mámi	mami	mami	mami	*mami
navel	hõmi-	-	-	-	sɨmɨ́ká	-	-	sóhõ-	sũhu-	sõho-	ʧõho	*sõho
nose	ũe-	ĩʔgɨ̃-	ɨ̃ŋẽã	ɨ̃gã	kenõ	ɨ̃kẽã	ẽkẽã	ũke-	ɨ̃kʰe	õkʷe-	ũkue	*ũkʷe-
paca sp	heme	sẽme	sẽme	seme	sãmá	hẽme	sẽme	seme	seme	sẽme	seme	*seme
pacu fish	-	uhu	uhu	uhu-	makóa	ɲimi	yuu	-	-	paku	pãku	-
palm weevil	pɨ̃ko-	pɨ̃gã-	hɨ̃kõ-	-	pɨʧõã	-	-	hɨ̃ko	hɨ̃ʔkʰo	hɨ̃ʔko	hɨ̃ʔko	*hɨ̃ko-
parrot (gen)	weko	weko	weko	weko	waaʧó	weko	weko	békó	ekʰo	weko	weko	*weko
path	ma	maʔa	máá	ma	maʔá	maa	maa	ma	maʔa	maʔa	maʔa	*maʔa
peccary sp	-	jese	jese	jese	jeesé	jese	-	-	sẽse	sẽse	sẽse	*sẽse
penis	noe-	-	ahéá	siti-	nunɨ́	nuni-	-	nóé-	hɨ̃ko-	kʷiri	-	-
people, 1pl.incl	maha	masɨ̃	masá	masɨ	maasá	masã-	mani	máí	mai	pãi	baɨ̃	*mai
poison	-	nima	-	rɨ̃ma	ɲimá	nima	nima	-	jima	tsɨ̃ma	-	*ʤima
pot, clay	hodo-	soro-	sotɨ	sotɨ	sitú	hotɨ	dii	-	toto	soto	soto	*soto

GLOSS	KUB	DES	BAR	MAK	KOT	KAR	TUY	MAI	KOR	SEC	SIO	PWT
potato	japi	ɲapĩ	ɲahíi	ɲahi	ɲapí	ɲapĩ	-	jáhí	jahi	jahi	-	*jahi
pupuna palm sp	ɨ̃re-	ɨ̃rĩ	ɨ̃ne	hota	ɨ̃néó	ɨ̃ne-	-	ɨ́ne	ɨne	ɨ̃ne	ɨne	*ɨ̃ne
red	hũa-	dia-	sũa-	sũa-	sõʔã-	hõã-	sõã	má-	maaha-	ma	ma	*ma
river	hia	dia	ria-	ria-	diá	ria	día	jíája	jiaja	tsiaja	sʔia	*ʤiaja
root	nɨo	nugũ	ɲéémá	ɲema	nɨʔkó	nɨkõ	-	séu	saiɨ	-	sita	-
snake (gen)	ãja	ãɲã	áɲa	ãɲa	ãɲá	ãɲa	ãɲa	áɲa	aɲa	aɲa	aɲa	*aɲa
spider (gen)	pɨ̃pɨ-	bɨpɨ	bɨhɨ́	bɨhɨ	wɨ̃pɨ́	bɨpɨ	bɨpɨ	hɨ̃hɨ̃	hɨ̃hio	hɨ̃hɨ	hɨ̃hɨ	*hɨ̃hɨ
spirit	-	wãtĩ	wãti	-	waatĩ-	wãtĩ	-	ãĩ-	ati	wati	wati	*wãti
stone	kɨda-	ɨ̃tã-	ŋɨtáa	gɨ̃ta-	tãkã	ɨ̃tã	ɨ̃tã	átáa	kata	kɨ̃nao	gata	*gata
stump, stick	tutu-	turu	tutu	tutu	tutú-	tutu	tuarígɨ	tú-	tuu	tu-	tu	*tu
tapir (gen)	wekɨ	wekɨ	wekɨ	wekɨ	waatʃɨ́	wekɨ	wekɨ	békɨ́	ek^{h}ɨ	wẽkɨ	wekɨ	*wekɨ
tarira fish	doje	doe	-	roe	daasápiõ	rose	-	-	roje	-	-	-
termite	-	-	butú-	-	buutú-	butu-	-	jújú	ʧuʧu	juju	juju	*juju
thorn	-	pora	hotá	hota-	potá-	pota	pota	mío	miu	miu	miu	*miu
three	-	-	idía	ɨdia-	tíá	-	itiá	-	-	-	samute	-
thunder (n)	õpõ	bupu	bɨhó	bɨho	wɨ̃pó	bɨpo	bɨpo	mɨ́há	karaɨ	muhuɨ	mɨho	*mɨha
to bite (v)	kũ-	kũrĩ-	kuní-	kũni	baʔká-	baka-	baka	kũĩ-	k^{h}ũʔi	kũʔi	kuʔĩ-	*kũĩ
to blow (v)	pu-	puri-	hútí-	huti-	putí-	púti-	-	húú-	huʔi-	hui	hui-	*hui
to break (vt)	pope	bohe-	bohé-	bohe-	p^{h}oʔá-	wati-	bee	húá-	hũʔhe-	hoʔa-	hũhe-	*huha-
to chew	hãhɨ	baʔga-	bage-	bage-	jaʔká-	baku-	jage	náki-	jãki-	tsãki	sãki-	*ʤãki
to cut	-	nuʔri-	-	hata-	dɨté-	ta-	tɨtɨ	títé-	ruʔte	teʔte-	tɨ̃ʔte-	*tɨte-
to end	-	peʔre-	hédi	hedi-	p^{h}iʔtí-	hãna-	yaponó	sáó-	píʔni-	-	-	-
to float	-	paaja-	haja-	haja-	p^{h}aʔsá-	pasa-	pasa	-	aa-	wawa-	wawa-	*wawa-
to gather, collect	hewa	tea-	-	kãi-	sáá-	hee-	ɲee	ʧíá-	sia-	sia-	ʧia-	*sia-
to know	mahi-	masi-	másí	masi	mããsí-	masĩ-	masĩ	-	masi	-	masi-	*masi
to plant	ote	ore-	oté-	ote-	tóá-	ote-	ote	óté-	ote-	-	-	*ote-
to rub	-		-	ware-	t^{h}ú saʔka	-	-	-	karo-	-	soʔõ-	-
to sit	-	doa-	eharũhũ	ruhi-	duhí	rui-	dui	ɲui-	ɲuʔi-	du-	ɲuʔi-	*ɲuʔi-

GLOSS	KUB	DES	BAR	MAK	KOT	KAR	TUY	MAI	KOR	SEC	SIO	PWT
to sleep	kã-	kãrĩ-	kání-	kãni	kʰãní-	kãni-	kãni	kãĩ-	kʰãi-	kãi-	kaĩ-	*kãi-
to smoke (meat)	hɨjowa-	siju-	seso-	siso-	sisó-	-	sɨ̃sõ	sɨ́jo-	sɨ́jo-	sɨjo-	sɨjo-	*sɨ̃jo
to squeeze	pipo-	bipi-	búhe-	bihe-	wiipé-	ripe-	bipe	-	kaɲo-	-	-	-
to stop, stand	nú-	nigu-	-	rɨ̃gõ	duukú-	nɨkã-	-	nɨ́ká-	nɨ́kʰa-	nɨka-	nɨka-	*nɨka
to swell	-	bihi-	míhí-	mihi	wiipí	bipi-	bipi	húhu-	kʰótó-	koto-	haʔhu-	*koto
to urinate	kõre-	gõʔrẽ-	ŋóne-	gõne-	kʰuʔnũ-	õne-	kõne	góne-	kone-	kõne-	gone	*gone
to wait	kode-	kore-	-	ju-	kʰoʔtá-	kote-, juri-	-	ũté-	ʧaʔá-	ute-	-	-
toad sp	popo	moma-	momó-	momo-	wãʔmúɨ́	momo-	-	hoho	hohoo	hoho	hoho	*hoho
tobacco	-	muru	mɨno	mɨno	mɨʔnó	mɨno	mɨno	mɨ́to	mɨto	mɨto	mɨto	*mɨto
tocandira ant (sp)	piará	-	hetá	-	pitá	peta	-	-	-	-	-	-
tongue, liver	heme-	nero	ɲémé-	ɲeme	ɲaʔmá-	ɲeme-	ɲemero	néme-	jeme-	tsẽme	sʔeme-	*ʤeme
tooth	kõpi-	gɨkɨ-	guhí	guhi-	píríáká	opi	upi	gúĩ-	kõhi	kũhi	gũhi	*gũhi
tortoise, turtle (gen)	kũi-	-	gúú	gu-	kʰú-	uu	kúu	góu	koɨ	kou	goɨ	*goɨ
toucan (gen)	hoewe	nasĩ	rase	rase	daasá	rase	dase	ɲásé	ɲase	ɲase	ɲase	*ɲase
tree	hokɨ-	jukɨgɨ	jukɨ́	jukɨ-	juukɨ́	jukɨ	jucɨ	sũkíɲi	sũkiɲɨ	sõkɨ	sũki	*sũki
wasp (gen)	uʧi-	uti	uti-	uti-	tíróá	uti-	utiá	utí-	uti	uti	uti	*uti
water	oko	deko	okó	ide	kó	oko	oko	ókó	okʰo	oko	oko	*oko
white	bo-	bore-	botí-	boti-	boʔtá	boti-	buti	bo-	po-	po	bo-	*bo-
wife	-	mara-	mánáhó	rõmio	namó	nmo	nɨmo	nɨ́hõ	rɨ̃ho	nɨho	rɨ̃ho	*nɨho
wind	-	miru-	minó	mino	wĩʔnó-	wĩno	wĩno	tútu	-	-	tutu	-
woman	nomió	noméo	rõmio	rõmio	numino	rõmio	numió	nómio	romio	nomio	romi	*nomio
woodpecker (gen)	kõre-	-	kóné	kõne	kʰõné	kõne	-	kóne	kʰone	kõne	kone	*kone
yam	ɲamu-	ɲamu	ɲamó	ɲamo	ɲahmú-	-	ɲumuko	-	ɲaho	-	ɲaho	*ɲaho