

## A comparison of verbal person marking across Tupian languages Uma comparação da marcação verbal de pessoa nas línguas Tupí

Joshua Birchall

Museu Paraense Emílio Goeldi/MCTI. Belém, Pará, Brasil

**Abstract:** This paper explores the diachrony of the verbal person marking system across the large and structurally diverse Tupian language family. I argue that the historical development of these different patterns are best informed by analyzing their synchronic distributions with regard to the current evolutionary hypotheses on the family. I apply a parsimony reconstruction model across the topology of two different classifications and compare the results with what is known from traditional historical linguistic work. This study is able to provide support for previous claims about the family and also generates a number of additional hypotheses about the intermediate stages of development of these patterns.

**Keywords:** Morphosyntactic reconstruction. Phylogeny. Maximum Parsimony. Tupian. Indigenous Languages.

**Resumo:** Este artigo explora a diacronia dos sistemas de marcação verbal de pessoa entre as línguas da grande e estruturalmente diversa família linguística Tupí. Argumento que o melhor jeito para entender o desenvolvimento histórico dos padrões diferentes de marcação de pessoa é analisar suas distribuições sincrônicas em relação às hipóteses classificatórias desta família. Aplico um modelo de reconstrução com base em parcimônia à topologia de duas classificações diferentes e comparo os resultados com o que já sabemos através de trabalhos que utilizam o método histórico-comparativo tradicional. Os resultados deste estudo apoiam várias hipóteses anteriores sobre a família, e também geram algumas hipóteses adicionais sobre os estágios intermediários do desenvolvimento destes padrões.

**Palavras-chave:** Reconstrução Morfossintática. Filogenia. Máxima Parcimônia. Tupí. Línguas Indígenas.

---

BIRCHALL, Joshua. A comparison of verbal person marking across Tupian languages. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas*, v. 10, n. 2, p. 325-345, maio-ago. 2015. DOI: <http://dx.doi.org/10.1590/1981-81222015000200007>.

Autor para correspondência: Joshua Birchall Museu Paraense Emílio Goeldi. Avenida Perimetral, 1901- Terra Firme. CEP 66077-830. Belém, PA, Brasil. E-mail: [jtbirchall@gmail.com](mailto:jtbirchall@gmail.com).

Recebido em 20/02/2014

Aprovado em 07/05/2015



## INTRODUCTION

Just as phonological systems and lexical inventories evolve over time in ways that can be indicative of the history that related languages share, so do grammatical structures. Historical morphosyntax examines both the development of a particular form in a language and the changes in its function. The phonological form of a morpheme can be reconstructed using reoccurring sound correspondences that occur systematically in the languages under comparison. The function of a morpheme can change over time through processes of extension and reanalysis (Harris; Campbell, 1995; Gildea, 1998) or through contact-induced structural change (Heath, 1984). With this in mind, the reconstruction of the function of a morpheme should be based on the distribution of structural patterns across a language family and informed by the presence of cognate forms in the various functions.

The development of divergent pattern in the ways that different languages within a linguistic family express grammatical relations is one area that can especially benefit from a comparative approach to morphosyntactic diachrony. This paper explores the development of the verbal person markers across the Tupian language family from a typological and historical perspective in order to provide insights into how the different forms and functions of these grammatical systems arose over time.

The Tupian language family shows a high degree of internal variation regarding the ways in which the obligatory participants of a clause, i.e. the arguments, are expressed on the predicate. Many of the Tupian languages spoken in Rondônia only index the absolutive (S/O) argument on independent main clause predicates. However, the branches primarily spoken outside of Rondônia show a number of divergent patterns, such as the hierarchical marking pattern where only a single argument in transitive clauses are indexed depending on whether it is a speech act participant (SAPs; first and second persons) or not. Some of these languages also display different classes of intransitive verbs that index their sole argument using different marker sets, and/or have developed portmanteau morphemes that index certain situations where SAPs act on one another. Furthermore, other languages show marking patterns such as the indexation of only accusative arguments, as in Juruna, or the indexation of only nominative arguments, as in Nheengatú.

This paper first provides an overview of the Tupian languages by presenting two classificatory proposals for the family and then examining the typological diversity that the languages display in their verbal person marking systems. This is followed by a brief outline of previous proposals on the historical development of different person marking patterns in the family. The changes that took place in earlier stages of development of the Tupian family are first analyzed using a parsimony reconstruction model applied to the two classificatory proposals for the family. This preliminary analysis is then reconciled with additional information coming from the presence of cognate forms in the various verbal person marker sets. It is argued that the development of both the form and function of the verbal person markers can gain considerable insight from being examined with respect to hypotheses on the diversification of the family tree into its respective branches.

## THE TUPIAN FAMILY: AN OVERVIEW

### ON THE CLASSIFICATION OF THE TUPIAN FAMILY

Seven distinct branches of the Tupian family were first identified in Rodrigues (1958), whose classification was later expanded to ten branches by considering Mundurukú, Awetí and Mawé as separate clades distinct from Tupí-Guaraní (Rodrigues 1984, 1985). More recent work has focused on joining the Tupí-Guaraní branch into a single subgroup with Mawé and Awetí, which is referred to in this paper as Mawetí-Guaraní (Rodrigues; Dietrich 1997, Corrêa-da-Silva, 2010, Drude; Meira, 2015). Based on his reconstruction of the Proto-Tupí phonological system, Rodrigues (2007) proposes a binary division of the family into an Eastern branch and a Western branch. While the author admits that work is still

needed on the reconstruction of the intermediate stages of the family, this proposal provides a fully resolved hypothesis on how the branches developed with relation to one another<sup>1</sup>. This classification is given in Figure 1, where the tip labels of each branch of the tree correspond to the name of the respective branch of the family rather than a specific language.

Since this classification draws on the specialized knowledge of Tupian experts who have worked on the family for over half a century, it will be referred to in this paper as the 'expert classification'. While a number of systematic sound correspondences are presented in Rodrigues (2007), the specific motivation for the ordering of the splits is not discussed in detail. As such, it should only be considered a working hypothesis until the details of the classification can be more fully developed. Cabral (2002) notes that the distinction between the languages that show a primarily absolute indexation pattern and those with more divergent patterns was used "as a basis for a first division of the Tupian stock into two principal branches". In the absence of more explicit motivation for the ordering of splits within the family, it could prove problematic to use a tree whose topology was influenced by argument marking patterns to discuss the development of these patterns since the two are not independent. Argument marking patterns are also problematic to use as evidence for the Western branch of the family since the presence of absolute indexation could be a shared retention rather than a shared innovation. For these reasons, an additional classification is also used for the parsimony analysis.

Another classification of the Tupian family was published in Walker *et al.*, (2012) based on a computational analysis of 40 basic vocabulary items taken from the Automated Similarity Judgment Program (AJSP) database. This proposal is based

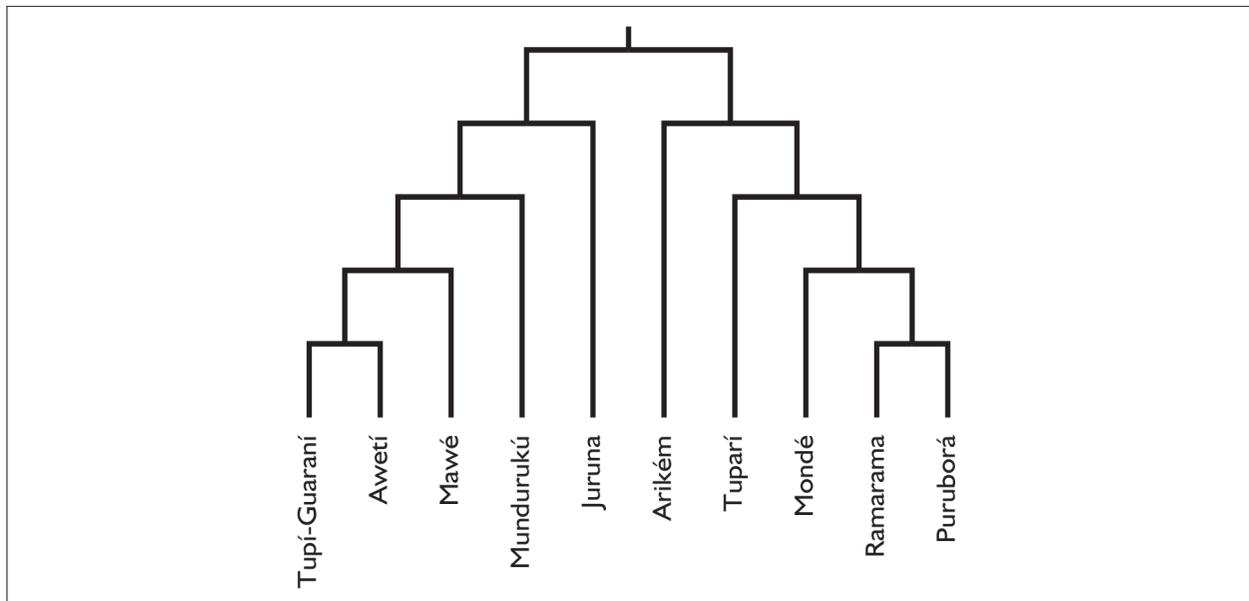


Figure 1. Expert classification of the Tupian family in Rodrigues (2007).

<sup>1</sup> Aside from the tentative grouping of Ramarama and Puruborá into a single clade, Drude (2011) presents a classification of the Tupian family that uses only these subgroups, representing the comparative work carried out by the Tupi Comparative Project hosted at the Museu Goeldi in Belém. Since the classification only includes subgroupings already reconstructed through the comparative method, higher order relationships between the major branches are not yet included. This makes the classification more difficult to use for historical reconstruction and is therefore not utilized in the historical analyses.

on a distance matrix computed using normalized Levenshtein distances (edit distances) between pairs of corresponding lexical entries that have been transcribed with a simplified orthography (Holman *et al.*, 2008). The topology of the tree is then calculated by applying the Neighbor Joining algorithm to this distance matrix (Saitou; Nei 1987). The tree was rooted using Proto-Carib as an outgroup since the two protolanguages have been postulated as descending from a common ancestor (Rodrigues, 1985). A cladogram of the ASJP tree is given in Figure 2. Note that the tip labels of the tree correspond to the names of the languages from which the data were gathered, rather than the names of the branches (subfamilies) as in Figure 1. For ease of visualization, only the languages used in this study are included<sup>2</sup>.

Unlike the expert classification of the Tupian family presented in Figure 1, the ASJP classification has the benefit of being fully transparent and replicable, with all data and sources freely available online for inspection<sup>3</sup>. It is notable that the ASJP correctly identifies all of the ten branches that are traditionally distinguished in Tupian studies, as well as a higher order relationship between Mawé, Awetí and Tupí-Guaraní. But unlike the East-West division in the Rodrigues (2007) classification, the ASJP classification shows a more ladder-like model of diversification for the family, without the Rondônia groups forming a distinct clade. Both the expert classification and the ASJP classification include Mundurukú and Juruna as the nearest phylogenetic relatives to Mawetí-Guaraní, although ASJP first groups the two former branches into an intermediate clade before forming a clade containing all of the languages in Rodrigues' Eastern branch. Since both classifications agree on a higher order relationship between Mawetí-Guaraní, Mundurukú and Juruna, and since these families are primarily located outside of Rondônia, which is often considered the homeland for the Tupian peoples (Noelli, 1996), they are referred to in this paper as the 'expansionist' subgroup of the family.

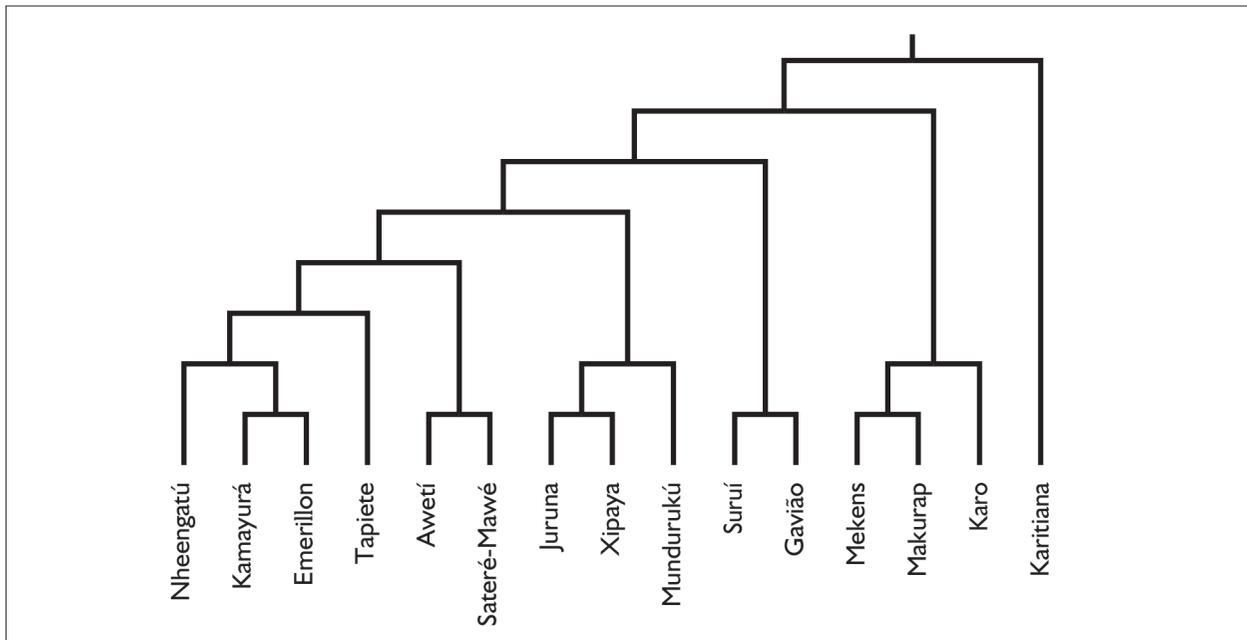


Figure 2. ASJP classification of the Tupian family adapted from Walker *et al.*, (2012).

<sup>2</sup> The Emerillon language was not included in the original ASJP tree in Walker *et al.*, 2012. Its position in the tree is based on its closest phylogenetic relative in their analysis, Wayampi. Both languages belong to subgroup 8 of Tupí-Guaraní (Jensen, 1999).

<sup>3</sup> <http://email.eva.mpg.de/~wichmann/languages.htm>

## VERBAL PERSON MARKING PATTERNS

As mentioned above, the Tupian languages show a wide array of different verbal person marking patterns. In this paper, the Tupian languages are approached from a typological perspective that aims to explore the ways in which the different languages express arguments on the predicate within independent main clauses. The following terms are adopted to represent three important comparative concepts that are used in the coding of different structural properties, following standard use in modern typology (Comrie, 1989; Dixon, 1994, among many others): S represents the sole argument of an intransitive predicate; A represents the Agent-like argument of the predicate in a prototypical transitive construction; and, O represents the Patient-like argument of the predicate in a prototypical transitive construction. Transitivity is used here as a semantic concept to identify a specific morphosyntactic construction within a language for comparison across languages, following Lazard (2002).

For the sake of comparison of different alignment patterns in the sets of verbal argument markers, no terminological distinction is made between pronominal agreement, cross-reference and bound pronouns, in the sense that the former two can co-occur with a nominal or a free pronoun while the latter cannot (Haspelmath, 2013). Imposing such a distinction on the data would obscure any diachronic discussion of alignment. For example, some of the Tupian languages with an absolutive marking pattern do not allow for the verbal marker to co-occur with a realized O nominal but only a S nominal, such that these languages would therefore be described as having an unusual marked-S system (Galucio, 2001, p. 77-79 for a discussion on the person markers in Mekens). These different characterizations of verbal argument marking are included under the term 'indexation' in this paper. As in many Tupian studies, the term 'person markers' is also used here to include both pronominal clitics and affixes that are bound to the verb.

An absolutive marking pattern, where S and O are indexed with the same morphological paradigm, is observed in main clause verbal inflection in languages of the Tuparí, Ramarama, Mondé and Arikém subfamilies<sup>4</sup>. An example of a language with ergative alignment through the indexation of the absolutive argument (S/O) is Wayoró in (1):

(1) Wayoró (Tuparí; Nogueira 2011:69-70)<sup>5</sup>

a.	<b>tʃi-pi:to-kar-a-t</b>	<i>tʃire</i>	b.	<b>tʃi-po-kw-a-t</b>	<i>agopkap</i>
	1PL.INCL-rest-VBLZ-TV-PST	1PL.INCL		1PL.INCL-burn-VBLZ-TV-PST	fire
	'We (incl.) rested'			'The fire burned us (incl.)'	

While many of the Tupian languages that reside in Rondônia show an absolutive indexation pattern like Wayoró, others in the family use distinct marker sets to index A and O arguments in the clause. In most languages, these marker sets are in complementary distribution with each other and occur in the same morphological slot, generally as a prefix or proclitic<sup>6</sup>. In languages like Sateré-Mawé, Mundurukú and some members of the Tupí-Guaraní branch, the basic main clause construction shows multiple alignment patterns where S of one class of intransitive verbs is indexed with

<sup>4</sup> Across the family, the set of markers that indexes O is often used to mark the object of a postposition and the possessor of a noun.

<sup>5</sup> The original transcription of the authors is replicated as closely as possible. Due to varying usage of the grapheme <j>, the transcriptions been altered to be consistent across the examples such that <j> represents the palatal glide /j/. Other uses of <j> have been converted into their IPA equivalent for ease of comparison.

<sup>6</sup> These different sets of markers are referred to in the paper using roman numerals, such that the marker set that indexes A is Set I and the set that indexes O is Set II. These labels are not to be confused with the labels given to different marker sets in other Tupian studies, such as in Jensen (1990), although their labels are equivalent for most cases within the Tupí-Guaraní branch.

the same set of markers as O, showing ergative alignment, while S of a different class of intransitive verbs is indexed with the same set of markers as A, showing accusative alignment<sup>7</sup>. Furthermore, the indexation of A and O in transitive clauses is conditioned by which argument is ranked higher on a person hierarchy that privileges speech act participants (1<sup>st</sup> and 2<sup>nd</sup> persons) over non-speech act participants (3<sup>rd</sup> persons), resulting in a person hierarchy that can be roughly characterized as SAPs > 3. As such, these languages show two different factors that condition indexation: the scenario between co-arguments in transitive clauses and the lexical class of intransitive predicates<sup>8</sup>.

First, let us consider the hierarchical indexation pattern in transitive clauses. Take for example the use of different sets of markers to index A and O in Mundurukú, as shown in example (2):

(2) Mundurukú transitive indexation (Gomes, 2006, p. 48-49, 52, 74)

- |  |   |
|--|---|
| <p>a. <i>bio</i>    <b>o=j-aoka</b><br/>         tapir    1SG.I=R1-kill<br/>         'I killed the tapir'</p>              | <p>c.    <i>õn</i>    <i>bio</i>    <i>aoka-m</i><br/>         1SG    tapir    kill-IPFV<br/>         'I will kill the tapir'</p>                   |
| <p>b. <i>ixe</i>    <b>wuj=dobuxikxiğ</b><br/>         3SG    1PL.INCL.II=find<br/>         'He is finding us (incl.)'</p> | <p>d.    <i>ğasũ</i>    <i>õn</i>    <b>e=nomuwãwã-m</b><br/>         now    1SG    2SG.II=R1.call.DUR-IPFV<br/>         'Now I am calling you'</p> |

Notice how in example (2a), the A argument is indexed with the Set I proclitic *o=*, while in (2b), the O argument is indexed with the Set II proclitic *wuj=*. This is due to the fact that the A argument in (2a) outranks the O co-argument on the person hierarchy, while the O argument in (2b) outranks the A co-argument since the indexed arguments are speech act participants<sup>9</sup>. However, verbs that are marked for imperfective aspect with the suffix *-m* never index the A argument (2c), while SAP O is indexed with a Set II proclitic (2d).

Additionally, Mundurukú has two classes of intransitive verbs that each take a different set of markers to index the sole argument of the predicate<sup>10</sup>. For the major class of verbs, the subject is indexed with the Set I markers, while the subject of a minor class verb is indexed with the Set II markers, as seen in example (3):

<sup>7</sup> Many Tupian languages, particularly of the expansionist group, have two distinct classes of intransitive verbs that index their subjects with different marker sets. The class of verbs composed of members that displays semantic properties associated with the Agent argument, namely volition and control (cf. Foley, 2005), are referred to as the major class of intransitive verbs, abbreviated as  $S_{maj}$ . As discussed below, Tupian specialists have employed a number of different labels to describe this class of verbs such as 'active', 'agentive', 'procedural' and so forth. The class of verbs whose S is indexed differently from the major class intransitive verbs will be referred to as the minor class, whose argument is abbreviated as  $S_{min}$ . In Tupian languages, the verbs of the minor class often displays semantic properties typically associated with the Patient argument, namely affectedness and lack of control. However, semantic differences alone are not sufficient to posit a split intransitive system unless these semantic differences are reflected in the way that the S argument is indexed on the predicate.

<sup>8</sup> See Zúñiga (2006) for the use of the term 'scenario' in reference to the typology of hierarchical marking. In this paper, the term local scenario is used when SAP A acts on SAP O, direct scenario when SAP A acts on 3<sup>rd</sup> person O, inverse scenario when 3<sup>rd</sup> person A acts on SAP O, and non-local scenario when 3<sup>rd</sup> person A acts on 3<sup>rd</sup> person O.

<sup>9</sup> In non-local scenarios A is indexed on the verb. Unlike many Tupí-Guaraní languages that use a special portmanteau morpheme to index both A and O in local scenarios (see ex.5 below and Rose, this volume). Mundurukú always indexes O when it is a SAP (Gomes, 2006, p. 48). Furthermore, as can be seen in Table 2, there are a number of homophonies across the paradigms such that 1<sup>st</sup> and 2<sup>nd</sup> persons singular and 1<sup>st</sup> person plural exclusive arguments are indexed using the same form whether  $S_{maj}$ ,  $S_{min}$ , A or O.

<sup>10</sup> The major class of verbs, what Gomes (2006) calls 'procedural' intransitive verbs, are those that denote a dynamic event involving an agent that exerts control over the event. The minor class of intransitive verbs, the 'stative' verbs, denote a state or quality of the subject (Gomes, 2006, p. 63). Picanço (pers. comm. 2011) notes that certain verbs denoting events that involve a participant that does not exert control over the event, such as 'drown' or 'cry', are marked in the same way as the stative verbs.

(3) Mundurukú intransitive indexation (Gomes, 2006 p. 50-51, 141)

- a. *kuj*      **o=adʒok**  
 already      1SG.I=bathe  
 'I have already bathed'
- b. *wujdʒu*      **wuj=direm**  
 1PL.INCL      1PL.INCL.II=be.wet  
 'We (incl.) are wet'
- c. *ǰaxĩn*      *ma*      *wujdʒu*      *cu-m*  
 soon      really      1PL.INCL      leave-IPFV  
 'We are leaving really soon'

Given the examples shown in (2) and (3), Mundurukú major class intransitive verbs show accusative alignment (S=A), while the minor class verbs show ergative alignment (S=O). Indexation patterns with splits in intransitive marking like that shown for Mundurukú in (3a-b) have often led to these languages being described as 'split ergative'. This split only occurs with verbs in the unmarked perfective aspect in Mundurukú, while verbs in the marked imperfective aspect show an entirely different verbal marking pattern: accusative alignment through the indexation of O (3c).

The Awetí language shows a split in the indexation of intransitive subjects that diverges somewhat from the pattern seen for perfective Mundurukú verbs. Awetí has three sets of argument markers: Set I is used to mark A, Set II is used to mark  $S_{min}$  and O, and Set III is used to mark  $S_{maj}$ <sup>11</sup>.

(4) Awetí (Borella, 2000 p. 136, 140, 148, 155)

- a. *kujta-za*      **wejt-apit**      *naʔ*  
 DEM(M)-COL      3.I-burn      3SG(M)  
 'Those ones burned him'
- b. *kujta-za*      **o-azũ-eju**  
 DEM(M)-COL      3.III-dance-CONT  
 'Those ones are dancing'
- c. *ito*      **it-akup-eju**  
 1SG(M)      1SG.II-have.fever-CONT  
 'I have a fever'
- d. *kujtã*      **i-tup**  
 DEM(M)      1SG.II-see  
 'That one saw me'

Notice how in (4a) the 3<sup>rd</sup> person A is indexed by a prefix *wejt-*, while the 3<sup>rd</sup> person intransitive subject ( $S_{maj}$ ) in (4b) is indexed by the prefix *o-*. Furthermore, the 1<sup>st</sup> person singular  $S_{min}$  in (4c) is indexed with the same prefix *i(t)-* as the 1<sup>st</sup> person singular O in (4d). While there are certain homophonies within the prefix paradigm not presented above, these examples show that the class of active intransitive verbs in Awetí is not indexed using the same prefix paradigm as A, while still showing the same pattern of hierarchical marking with split intransitivity found elsewhere in the Mawetí-Guaraní and Mundurukú subgroups<sup>12</sup>.

Many Tupí-Guaraní languages show hierarchical indexation patterns similar to that seen for Mundurukú perfective verbs in (2) and (3). But unlike Mundurukú, these languages often have an additional marker set of portmanteau prefixes (Jensen's Set 4) that index a 1<sup>st</sup> person A with a 2<sup>nd</sup> person O co-argument in transitive clauses, as in (5) for Kamayurá<sup>13</sup>.

<sup>11</sup> Borella (2000, p.131) labels the  $S_{min}$  intransitive verbs as 'descriptive verbs' and the  $S_{maj}$  class as 'active intransitive verbs'.

<sup>12</sup> The indexation pattern in Awetí is problematic for the often-repeated assumption that "those S which are semantically similar to A (exerting control over the activity) will be  $S_A$ , marked like A..." (Dixon, p. 1994). This is one of the primary reasons that the  $S_{maj}/S_{min}$  notation is adopted here, so that a distinction can be made between the alignment of the intransitive class of predicates and the semantics of the lexical class.

<sup>13</sup> While no similar portmanteau forms occur in Awetí, Sateré-Mawé uses the prefix *moro-* to index local scenarios where 1<sup>st</sup> person A acts on a 2<sup>nd</sup> person O co-argument. Corrêa-da-Silva (2010, p. 248) suggests that the form was probably borrowed from a Tupí-Guaraní language.

(5) Kamayurá transitive indexation (Seki, 2000, p.140)

- |  |  |
|--|--|
| <p>a. <b>ere-etsak</b><br/>2SG.I-see<br/>'You (sg.) saw him'</p>         | <p>c. <b>oro-etsak</b><br/>1A.2SG.O-see<br/>'We (excl.) / I see you (sg.)'</p> |
| <p>b. <b>ne=r-etsak</b><br/>2SG.II=R-see<br/>'He/they saw you (sg.)'</p> | <p>d. <b>opo-etsak</b><br/>1A.2PL.O-see<br/>'We (excl.) / I see you (pl.)'</p> |

The Set II markers in Kamayurá show different attachment to the verb depending on the person of the marker. The SAP forms are analyzed as proclitics that attach after the inclusion of the 'relational prefix' (5b), a phonological class marker that attaches to certain nouns and verbs depending on their inflection (Meira; Drude, 2013). The 3<sup>rd</sup> person form *i-* is analyzed differently in various Tupí-Guaraní languages, but Seki (2000, p. 66) analyzes it as a 3<sup>rd</sup> person form of the relational prefix. Since the referential hierarchy in Kamayurá does not allow for the indexation of 3<sup>rd</sup> person O in transitive clauses, the 3<sup>rd</sup> person Set II form is best observed in the person inflection of minor class intransitive verbs, as seen in (6)<sup>14</sup>.

(6) Kamayurá minor class intransitive indexation (Seki, 2000, p.157)

- |  |   |
|--|---|
| <p>a. <b>i-'ajura i-huku</b><br/>3.II-neck 3.II-be.long<br/>'His neck is long'</p> | <p>b. <b>ene n=r-oryp</b><br/>2SG 2SG.II=R-be.happy<br/>'You are happy'</p> |
|--|---|

Jensen (1998) reconstructs an additional marker set for Proto-Tupí-Guaraní that was used to index the absolutive argument of dependent verbs when it is coreferential with the subject (S/A) of the matrix independent clause<sup>15</sup>. This can be seen in the Kamayurá 'gerund' constructions shown in (7).

(7) Kamayurá gerund constructions (Seki, 2000, p. 315-316)

- |   |  |
|---|--|
| <p>a. <b>a-ke-potat we-pytu'u-me=ran</b><br/>1SG.I-sleep-DESID 1SG.DEP-relax-GER=CONJ<br/>'I want to sleep and relax'</p> | <p>b. <b>ja-jemo'ypy jere-karu-m</b><br/>1PL.INCL.I-begin 1PL.INCL.DEP-eat-GER<br/>'We began to eat'</p> |
|---|--|

Not all Tupí-Guaraní languages show the same pattern of hierarchical indexation as Kamayurá. Nheengatú appears to have lost the hierarchical marking pattern found across much of the rest of the branch while still utilizing two distinct marker sets to index the subjects of different classes of intransitive verbs. Nheengatú additionally lost the inclusive-exclusive distinction in its person markers and does not have portmanteau forms for 1<sup>st</sup> person A acting on 2<sup>nd</sup> person O (Cruz, 2011, p. 132-3).

<sup>14</sup> The use of *i-* as a 3<sup>rd</sup> person O marker is well attested in the now extinct Tupí-Guaraní language Tupinambá, such as in the example **a-i-kutúk** 1SG.I-3.II-pierce 'I pierced it' (Jensen, 1990, p.121). This pattern of marking 3<sup>rd</sup> person O together with the Set I A marker in transitive clauses is reconstructed for PTG in Jensen (1998). This helps to explain the difference between the S<sub>min</sub> and A marker sets in Awetí for 1st person plural exclusive, where the A form **oʒoj-** arose from the S form **oʒo-** attached to *i-* in the transitive. See also Monserrat (1976).

<sup>15</sup> In some Tupí-Guaraní languages, the same set of prefixes is used to indicate coreference between the subject of a transitive clause and the possessor of O, such as in Tapirapé **ã-ma-pen we-pa** 1SG.I-CAUS-break 1SG.COREF-hand 'I broke my hand' (Leite 1989 *apud* Jensen 1998, p. 504). In most other instances, the S<sub>min</sub>/O markers are also used to express possession.

A final interesting pattern can be observed in the Juruna language. Juruna was initially described as only indexing O in Fargetti (2001), similar to its sister language Xipaya described in Rodrigues (1995). However, a more recent analysis of Juruna verb structure in Lima (2008) shows that there is a class of intransitive verbs, which she calls ‘unaccusatives’, that indeed allows for S (8b-c) to be indexed with the same marker set as that used for O (8a), while the other class of intransitive verbs (‘unergatives’) do not allow for S to be indexed (8d).

(8) Juruna (Fargetti, 2001, p. 178, 191; Lima, 2008, p. 176, 180)

- |   |   |
|---|---|
| <p>a. <b>una e=djídaku e=bé</b><br/>         1SG 2SG=hit 2SG=DAT<br/>         ‘I hit you’</p> | <p>c. <b>es=abéáta i=ʔúru anu</b><br/>         2PL=clothes 3SG=be.wet ASP<br/>         ‘Your (pl.) clothes are wet’</p> |
| <p>b. <b>ena e=kũãũ</b><br/>         2SG 2SG=faint<br/>         ‘You fainted’</p>             | <p>d. <b>ena tahu</b><br/>         2SG run<br/>         ‘You ran’</p>   |

While absolutive marking and the hierarchical/split intransitive marking are the two dominant patterns observed across the Tupian family, the divergent patterns in languages such as Juruna, Awetí and Mundurukú help to inform an analysis of how the dominant patterns arose.

## ON THE DEVELOPMENT OF THE VERBAL PERSON MARKERS

Now that the different marking patterns across the Tupian languages have been introduced, the discussion turns to the historical development of these patterns. While a reconstruction of the marker sets in Proto-Tupí is well beyond the scope of this article, it is possible to identify cognate forms and compare their grammatical function across the family based on previous work. These differences are explored with relation to two hypotheses on the classification of the Tupian language family, and the changes discussed are modeled over these trees using a parsimony model for ancestral state reconstruction. Since considerable work already exists on the development of verbal person marking within the Tupí-Guaraní branch, the following sections focus on the development of the markers across the family with regard to the changes that must have occurred to produce the patterns reconstructed for Proto-Tupí-Guaraní (PTG) as well as those observed within the other branches of the family.

## PREVIOUS CLAIMS

The two most explicit proposals for the development of verbal argument marking across the Tupian family are Jensen (1998) and Gildea (2002). Based on her experience with Tupí-Guaraní languages, Jensen (1998, p. 565-573) proposes a five stage process for the development of the hierarchical indexation in PTG from a putative Pre-PTG, as shown in (9):

(9) The development of PTG indexation per Jensen (1998)

- a. Pre-PTG indexed only absolutive arguments on independent verbs. Ergative arguments were expressed with free pronouns or nominals.
- b. Agentive intransitive verbs developed a new set of prefixes.



- c. The agentive intransitive prefix set was extended to also index A of transitive verbs when O is 3<sup>rd</sup> person. In these direct scenarios, both A and O are indexed. In inverse and local scenarios, where O is a SAP, only O is indexed.
- d. A distinction between SAPs develops in the person hierarchy such that 1>2>3.
- e. Portmanteau prefixes *\*oro-* and *\*opo-* are developed that index 1<sup>st</sup> person A acting on a 2<sup>nd</sup> person O co-argument.

Thus in Jensen's analysis, the  $S_{\text{min}}/O$  marker set is considered inherited from an earlier stage in the family, and an additional marker set for indexing agentive intransitive verbs was innovated from an unidentified source. The person hierarchy only developed in transitive clauses after the extension of the innovated marker set from marking only the subject of major class (agentive) intransitive verbs to the subject of transitive clauses. Both of these key aspects of this model are further discussed below.

Drawing from basic tenets of grammaticalization theory and his reconstructive work with Cariban languages, Gildea (2002) proposes a completely different diachronic pathway for the development of PTG person marking. His argumentation is based on principles of historical syntax that hold that older morphemes tend to be phonetically smaller, semantically more opaque, closer to the stem and show more morpho-phonemic irregularity than relatively newer morphemes (following Givón, 2000, p.120-121). Using this logic, Gildea argues that it is unlikely that pre-PTG indexed the absolutive argument on the predicate since the  $S_{\text{min}}/O$  marker set are phonetically larger than the  $S_{\text{maj}}/A$  set, less bound to the predicate, and show a higher degree of similarity with the free pronoun set. He proposes that the *\*i-/c-* 3<sup>rd</sup> person  $S_{\text{min}}/O$  prefix is the oldest and that the system that developed for PTG arose from a predominately nominative-accusative indexation system as shown in (10).

(10) The development of PTG indexation per Gildea (2002)

- a. Predicates index 3<sup>rd</sup> person O with the *\*i-/\*c-* prefix. It is possible that a class of intransitive verbs also used these prefixes for subject indexation.
- b.  $S_{\text{maj}}/A$  prefix set develops from a set of free pronouns that were lost prior to the development of PTG. These prefixes attach further away from the stem than the *i-/c-* markers, resulting in a nominative-accusative pattern when O is 3<sup>rd</sup> person.
- c. The 1<sup>st</sup> person acting on 2<sup>nd</sup> person portmanteau prefixes develop.
- d. The  $S_{\text{min}}/O$  marker set developed from the free pronoun set that replaced the earlier pronoun set that formed the  $S/A$  marker set.

Central to Gildea's model is that the 3<sup>rd</sup> person  $S_{\text{min}}/O$  prefix *\*i-/\*c-* was inherited from an earlier stage of the family and that the proclitic forms that index the SAP arguments within that same marker set developed after the inclusion of an additional marker set that indexes  $S_{\text{maj}}/A$ .

Both Jensen (1998) and Gildea (2002) contribute many important insights into the development of PTG argument indexation. However, at the time that these proposals were being written, only a few studies had been published on the grammar of the non-Tupí-Guaraní languages. Since the turn of the century there has been a marked increase in the quantity and quality of the descriptive materials available on these languages. Furthermore, greater attention has since been paid to the classification and reconstruction of the various branches of the family. Both of these factors can now help to refine our understanding of the development of this system.



## QUANTITATIVE ANALYSIS OF VERBAL PERSON MARKING PATTERNS

Since Jensen (1998) and Gildea (2002) propose quite different indexation systems for the early stages of the Tupian family before the development of PTG, a logical starting place for a discussion on the development of this system would be to examine the observed distribution of verbal argument marking patterns in the modern languages. One straightforward technique for this is the application of the parsimony principle to discrete typological data in order to reconstruct the ancestral state of these features over a given phylogenetic tree. A parsimony-ased comparative analysis considers that the best model for evolutionary development is the one that requires the least amount of changes needed in order to fully account for the observed data (Fitch, 1971). While languages surely develop in ways that are not always the most parsimonious, the simplest possible model is a reasonable place to begin the discussion. From there, any developments within the family that diverge from the most parsimonious scenario should be supported by additional evidence and argumentation (Cysouw, 2009).

A comparative phylogenetic analysis of the development of verbal person markers was carried out over the two discussed classificatory hypotheses using a parsimony reconstruction model. The languages shown in Table 1 were sampled for this analysis<sup>16</sup>.

Table 1. Language Sample.

Language	Branch	Sources
Karitiana	Arikém	Everett 2006, Storto p.c.
Awetí	Awetí	Borella 2000, Drude p.c.
Juruna (Yúdja)	Juruna	Fargetti 2001, Lima 2008
Xipayá	Juruna	Rodrigues 1995
Sateré-Mawé	Mawé	Silva 2010
Gavião	Mondé	Moore 1984, p.c.
Suruí (Paite)	Mondé	Van der Meer 1985
Mundurukú	Mundurukú	Gomes 2006, Picanço p.c.
Karo (Arara)	Ramarama	Gabas 1999
Makurap	Tuparí	Braga 2005
Mekens (Sakurabiat)	Tuparí	Galucio 2001
Wayoró (Ajurú)	Tuparí	Nogueira 2011
Tapiete	Tupí-Guaraní	González 2005
Nheengatú	Tupí-Guaraní	Cruz 2011
Emerillon	Tupí-Guaraní	Rose 2011
Kamayurá	Tupí-Guaraní	Seki 2000

<sup>16</sup> Three important factors were considered when constructing the language sample: genealogical diversity, geographic spread and the availability of materials. Since the target sample is composed entirely of members from a single language family, representatives of as many different clades of that family tree as possible were selected. Given the ten branches of the Tupian family presented in Rodrigues (1999), at least one member of each branch is included in the sample with the exception of Puruborá, for which little grammatical data is available (Galucio, 2005; Monserrat, 2005). Since the development of the person markers within the Tupí-Guaraní branch of the family has been thoroughly studied and discussed, this branch has not been densely sampled, but rather, only a few representative languages from different geographically widespread subgroups of the branch have been included.

The analysis was implemented using the Mesquite software package (Maddison; Maddison, 2011). Each language was given a general typological feature value based on the arguments that are indexed in independent transitive clauses and the alignment of the marker sets used with the major class of intransitive verbs. For the analysis using the expert classification shown in Figure 1, the values for each branch were given based on the majority consensus of the languages from that branch in the sample. A side by side comparison of the analysis applied to both classifications can be seen in Figure 3, including the reconstructed values for intermediate nodes of the tree.

The analysis applied to the ASJP classifications (Figure 3, right) clearly reconstructs an absolutive indexation pattern for Proto-Tupí. However, the topology of the expert classification (Figure 3, left) makes it impossible to resolve a single pattern using purely independent clause data. This may be partly due to the way in which the data are coded, since each language is given only a single typological characterization. This method of coding is often used in typological databases, but a number of details that may be historically informative are lost so that each language can fit into a pre-established category. The analysis also differs between classifications regarding the point in the family history that the hierarchical marking pattern developed, with the ASJP classification positing its development in the early stages of the expansionist group before the diversification of the different branches, with a subsequent loss in the Juruna branch. The analysis applied to the expert classification posits the development of the hierarchical marking pattern after the separation of the Juruna branch from the other expansionist languages (Monserrat; Soares, 1983).

An alternative approach to coding is to allow each language to be coded for the alignment of the different marker sets used in independent clauses. Under this coding scheme, a language like Mundurukú that has one set of markers that indexes A and  $S_{maj}$  and another set that indexes O and  $S_{min}$  is coded as having two observed states—an S/A marker set and a S/O marker set—rather than a single state ‘Hierarchical’. The Mundurukú imperfective construction that only indexes the O argument is not considered in this analysis since it uses the same marker set as the unmarked perfective construction. This alternative analysis is shown in Figure 4.

Using the alternative coding scheme based on marker sets, the ASJP classification again reconstructs an absolutive marker set for Proto-Tupí, while the analysis of the expert classification does not allow for a single state to be reconstructed for Proto-Tupí due to the tree topology. The analysis applied to both classifications reconstructs an ancestral marker set for the expansionist group that indexes O with the same markers as the minor class of intransitive verbs. As discussed in the following section, traditional comparative method evidence suggests that these two sets are related to each other rather than forming two completely independent sets, albeit having undergone a number of historical developments that obscure their true cognacy.

Some other aspects of the parsimony analysis are inconsistent with what we know from classic comparative method reconstruction, and as such, the analysis should only be considered a preliminary hypothesis on which to base more detailed work<sup>17</sup>. For instance, the marker set-based analysis does not reconstruct multiple marker sets for Proto-Mawetí-Guaraní, largely due to the three marker set system found in Awetí. The parsimony analyses propose different diachronic pathways for the development of the hierarchical marking pattern when the data are treated with the different coding schemes, with Figure 3 proposing a single development that was maintained in

<sup>17</sup> It is worth noting that this paper only explores two possibilities for the coding of discrete typological features for the parsimony analysis. A number of other possibilities remain to be examined that may shed further light on these issues, such as the application of these methods to model the evolution of a particular form across different functions.

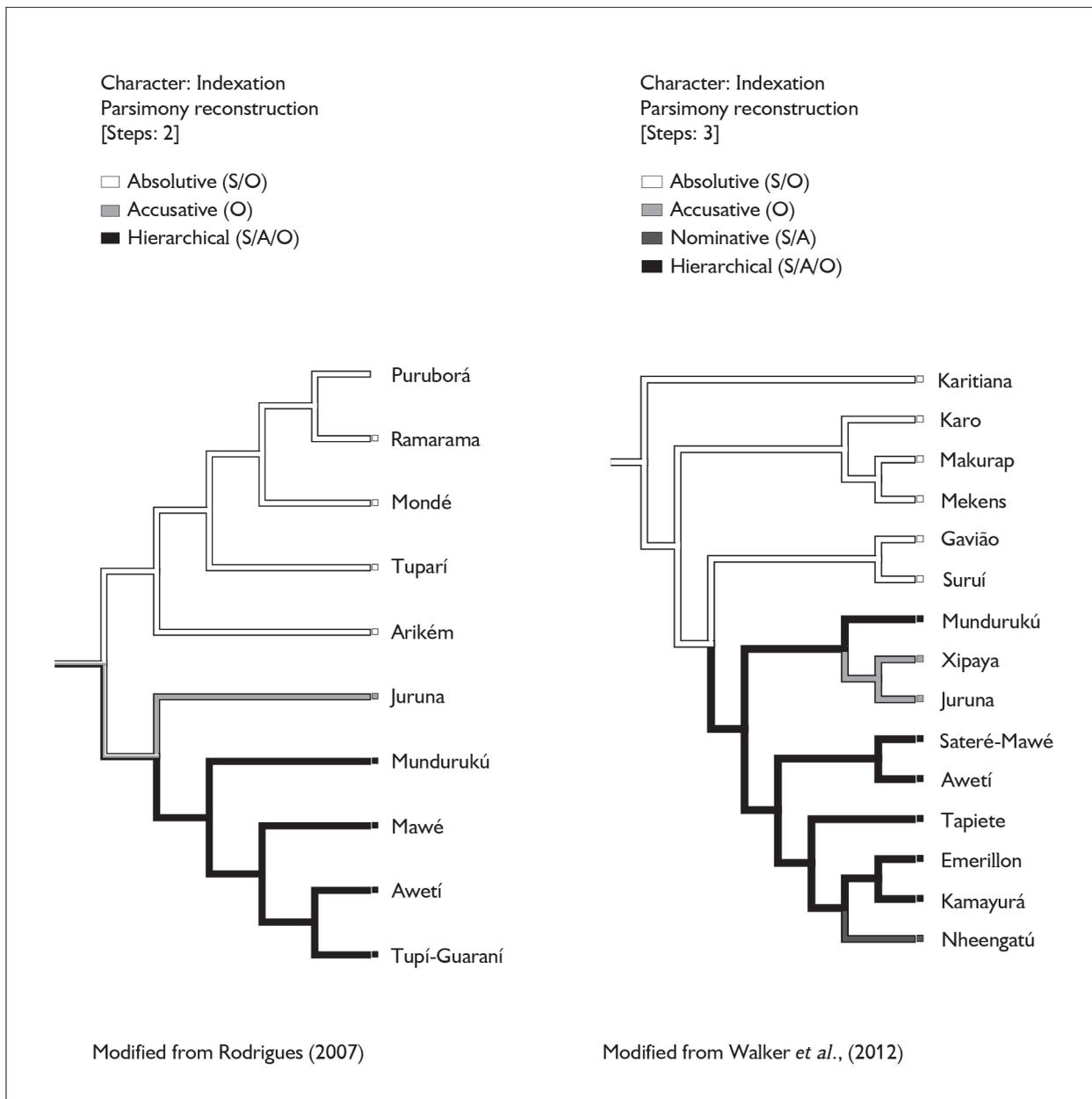


Figure 3. A reconstruction of typological patterns in Tupian argument marking modeled over the expert classification (left) and ASJP classification (right) using maximum parsimony.

Proto-Mawetí-Guaraní, whereas Figure 4 proposes a development of the system in Mundurukú independent of the development of hierarchical marking found elsewhere. This can be interpreted as showing that the hierarchical pattern of marking developed early on in the expansionist group of languages, but that the marker sets went through a number of different configurations between the ancestor of the expansionist languages and the formation of PTG, with the presence of a single  $S_{min}/O$  marker set being maintained throughout this history.

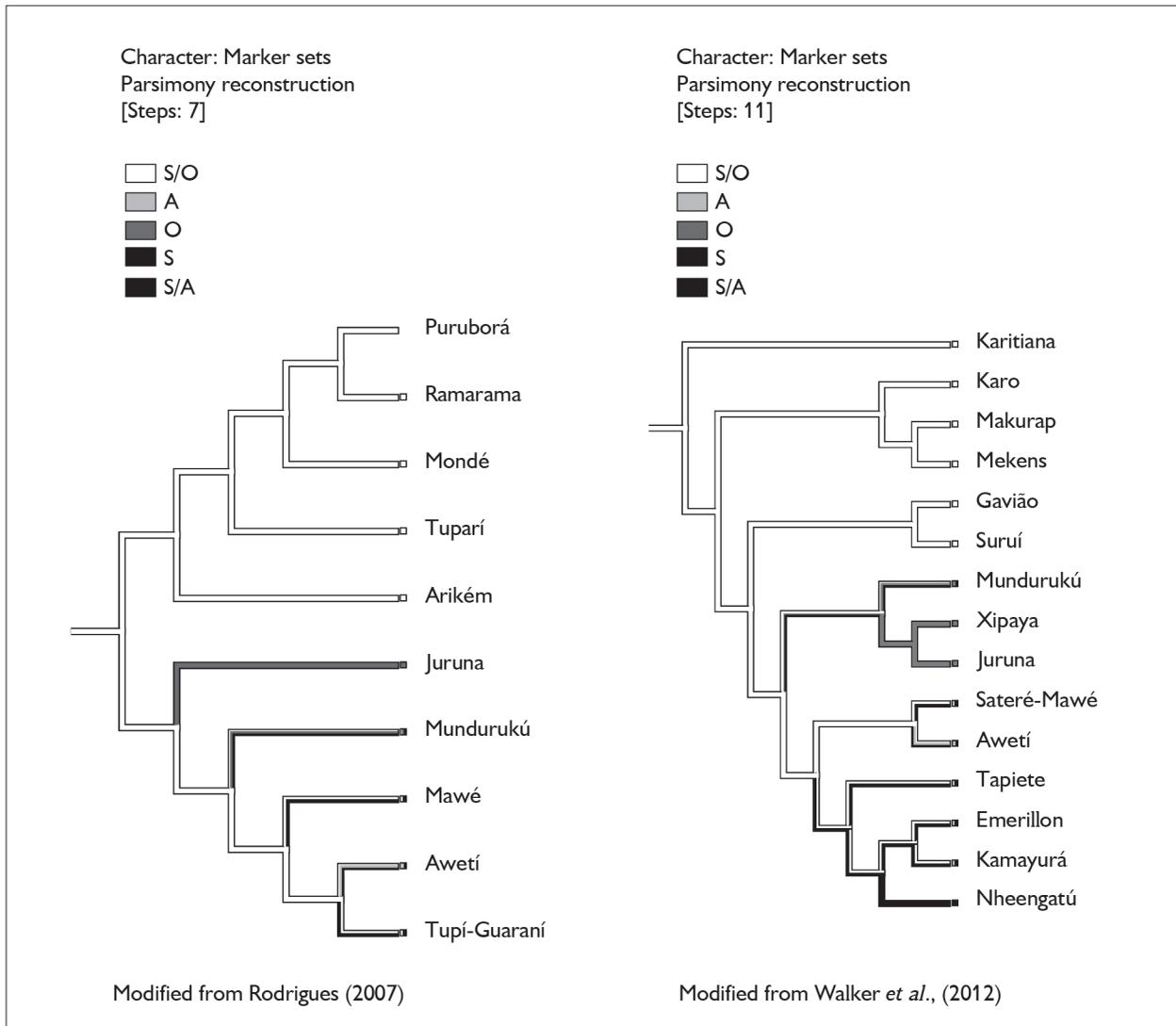


Figure 4. A reconstruction of marker sets used in Tupian argument marking modeled over the expert classification (left) and ASJP classification (right) using maximum parsimony.

## ON THE COGNACY OF FORMS WITHIN MARKER SETS

The different marker sets that occur in the languages of the sample are presented in Table 2. The arguments indexed by each marker set are indicated. Certain allomorphies are simplified for ease of comparison, and reconstructed forms from Jensen (1998) are given for Proto-Tupí-Guaraní as well as those for Proto-Tupí from Rodrigues and Cabral (2012). The reconstruction of PTG marker sets in Schleicher (1998) are largely congruent with those in Jensen (1998). The 1>2 portmanteau morphemes present in many Mawetí-Guaraní languages are not included for the sake of space. The markers that are likely cognates with the reconstructed Proto-Tupí forms are indicated in **bold**, based on the forms identified in Rodrigues and Cabral (2012, p.543). Additional provisional cognate judgments are made by the author following systematic sound correspondences identified in Rodrigues (2007) when those languages were not included in the former study.

Table 2. A summary of Tupian verbal person marker sets.

Language	Branch	Set	1sg	2sg	3sg	1pl.incl	1pl.excl	2pl	3pl
Proto-Tupí		S/O	*o=	*e=	?	*Vj=	*or'e=	*ej=	?
Karítiana	Arikém	S/O	i-	a-	-	ij-	ita-	aj-	-
Gavião	Mondé	S/O	o-	e-	ci- / saa-	pa-	to-	me-	ta-
Suruí	Mondé	S/O	o-	e-	a-	pa(m)=	toj=	mej=	táʔ=
Karo	Ramarama	S/O	o=	e=	aʔ= / ʔa=	ɪʔ=	té=	karo=	tap=
Makurap	Tuparí	S/O	o=	e=	e=	ki=	te=	eki=	e= / -
Mekens	Tuparí	S/O	o-	e-	i- / s-	ki-	ose-	ejat-	tejat-
Wayoró	Tuparí	S/O	o- / m-	e-	te- / ʔ- / ndeke-	ʔi-	ote-	ñdzat-	te- / ʔ- / ndeat-
Juruna	Juruna	S <sub>min</sub> /O	u=	e= / l=	i=	s(e)=	ul(u)=	es(se)=	(!)se= / abidai=
Xipaya	Juruna	O	u-	e-	-	se-	use-	ese-	-
Mundurukú	Mundurukú	S <sub>mag</sub> /A	o=	e=	oʔ=	a=	oʔe=	epe=	oʔ=
Sateré-Mawé	Mawé	S <sub>min</sub> /O	o=	e=	-	wuj=	oʔe=	ej=	-
		S <sub>mag</sub> /A	a-	e-	i-	wa-	uru-	ewe-	-
		S <sub>min</sub> /O	u-	e-	-	a-	uru-	e-	iʔatu-
Awetí	Awetí	A	a(t)-	ʔe(t)-	wej(t)-	tí(t)-	oʔoj(t)-	pej(t)-	wej(t)-
		S <sub>min</sub> /O	i(t)-	ʔe-	t- / i-	kaj-	oʔo-	ʔeʔi-	t- / i-
		S <sub>mag</sub>	a(j)-	ʔe-	o-	kaj-	oʔo-	ʔeʔi-	o-
PTG	Tupí-Guaraní	S <sub>mag</sub> /A	*a-	*ere-	*o-	*ja-	*oro-	*pe-	*o-
		S <sub>min</sub> /O	*ʔjé=	*né=	*i- / c- / t-	*jané=	*oré=	*pé=	*i- / c- / t-
		S/O <sub>dep</sub>	*wi-	*e-	*o-	*jere-	*oro-	*peje-	*o-
Tapiete	Tupí-Guaraní	S <sub>mag</sub> /A	a-	ndV-	o- / wV-	ja-	o- / -ha	pV-	o- / wV-
		S <sub>min</sub> /O	ʃ(V)-	ndV-	-	ñand(V)-	iʔ(i)...-ha	pV-	y(V)- / h(i)-
Kamayurá	Tupí-Guaraní	S <sub>mag</sub> /A	a-	ere-	o-	ja-	oro-	pe-	o-
		S <sub>min</sub> /O	je=	ne=	i- / t- / h-	jene=	ore=	pe=	i- / t- / h-
Emerillon	Tupí-Guaraní	S <sub>mag</sub> /A	a-	ere-	o-	si-	oro-	pe-	o-
		S <sub>min</sub> /O	e-	de-	i- / t-	node- / kóde-	ore-	pe-	i- / t-
Nheengatú	Tupí-Guaraní	S <sub>mag</sub> /A	a-	re-	u-	ja-	pe-	pe-	tau-
		S <sub>min</sub>	se-	ne-	i- / s-	jane-	pe-	pe-	o-



A number of important observations can be made about the retention and innovation of forms within the marker sets. First, of the expansionist subgroup of languages, Mundurukú is the only language outside of Mawetí-Guaraní to have distinct forms that index A in transitive clauses:  $o\mathcal{P}=$  3<sup>rd</sup> person,  $a=$  1<sup>st</sup> person plural inclusive and  $epe=$  2<sup>nd</sup> person plural. None of these forms appear cognate with the reconstructed Proto-Tupí forms in Table 2, suggesting that they are innovations. Likely cognates of these forms are maintained in the marker sets reconstructed for PTG  $S_{maj}/A$  set:  $o-$  corresponding to  $o\mathcal{P}=$  is retained in the 3<sup>rd</sup> person, and  $pe-$  corresponding to  $epe=$  is retained in the 2<sup>nd</sup> person plural. This suggests a continuous diachronic development of a distinction between A and O retained from a common ancestor. Interestingly, forms similar to  $epe=$  occur as 2<sup>nd</sup> person plural free pronouns in many Mawetí-Guaraní languages but not in Mundurukú, where the free pronoun is *ejdzu*.

Other forms in the Mundurukú Set I markers were extended from the older Set II markers, which are cognate with the Proto-Tupí forms. The reanalysis of these retained forms for 1<sup>st</sup> and 2<sup>nd</sup> persons singular and 1<sup>st</sup> person plural exclusive as  $S_{maj}/A$  markers suggests that the distinction between Set I and Set II forms developed gradually over multiple stages in the family history. An alternative hypothesis where all the innovative forms seen in Proto-Mawetí-Guaraní  $S_{maj}$  and A markers were also innovated in pre-Mundurukú and then partially replaced by reanalysis of the Set II forms does not appear to be supported by the available evidence.

The lack of clear cognacy between the Proto-Tupí  $S/O$  proclitics as reconstructed in Rodrigues and Cabral (2012) and the PTG  $S_{min}/O$  proclitics as reconstructed in Jensen (1998) can be attributed to the development of the free pronouns in PTG. The  $S/O$  markers in dependent clauses in PTG (Jensen's Set 3) are true reflexes of the reconstructed Proto-Tupí absolutive marker set, at least in the singular (Rodrigues, 1985, p.380; Jensen, 1998, p. 574). In main clauses, the indexing function of the Proto-Mawetí-Guaraní O set was replaced in PTG by attaching the independent pronouns to the verb, resulting in phonologically-reduced clitic forms. Drude (pers. comm. 2013) suggests that these clitic forms are not directly cognate with the earlier pronouns because they include an additional morpheme, a stress-bearing "formative element" \*e<sup>18</sup>. It is possible that after the addition of the \*e formative to the free pronouns in pre-PTG, the new stress pattern resulted in the loss of unstressed phonological material. A clear example of this proposed pathway can be seen for the 2<sup>nd</sup> person singular pronoun: Proto-Mawetí-Guaraní *en* became pre-PTG *ené*, resulting in the PTG  $S_{min}/O$  marker *né=*. The other non-3<sup>rd</sup> person pronouns developed along similar lines.

The fact that the  $S_{min}/O$  marker set in PTG can be traced to a development after the separation of Mawé, Awetí and Tupí-Guaraní into their respective branches provides support for the claim in Gildea (2002) that these proclitics developed later than the 3<sup>rd</sup> person prefix *i-/c-* of the same set. When examining the marker sets across the whole family, the 3<sup>rd</sup> person  $S_{min}/O$  form \**j-* in PTG does indeed appear to be a retention from an earlier stage of development in the family. For example, 3<sup>rd</sup> person *i-* is found in the Tuparí language Mekens, and a similar proclitic form *i=* is used to index 3<sup>rd</sup> person arguments in certain focus constructions in Karo (Gabas 1999, p.122-125). This form is also found in Juruna.

## DISCUSSION

The analysis presented above allows for a refinement of our understanding of the developments that took place in the person marking system of Tupian verbs, particularly among the expansionist group of languages. The first step in the

<sup>18</sup> Schleicher (1998, p. 240-243) analyzes this *e* morpheme as an accusative case suffix. Drude; Meira (pers. comm. 2013) refer to this morpheme as 'Hemmauer's e' due to it first being recognized by Roland Hemmauer during an informal discussion.

development of the marking patterns in the expansionist languages was a restriction of the use of the originally absolutive marker set for indexing the subject of intransitive verbs. This restriction is retained in the Juruna branch of the family, where Xipaya does not index any intransitive subjects and Juruna only indexes those of a minor class of intransitive verbs. This restriction can also be seen in the Mundurukú imperfective construction.

Due to differing classifications regarding the relation of the Juruna and Mundurukú branches, it is difficult to ascertain whether this restriction on intransitive indexation predated the development of the person hierarchy. However, no forms in Juruna have been identified as cognate to the forms that distinguish the marker sets in Mundurukú, suggesting that the expert classification may more accurately portray the relationship of these groups to one another than the ASJP classification. Based solely on the analysis of the expert classification, it is clear that the restriction of intransitive indexation arose before the development of the person hierarchy as a condition on transitive indexation.

As mentioned above, some of the innovative forms in Mundurukú have cognate forms in the marker set that indexed  $S_{maj}/A$  in PTG (2<sup>nd</sup> person plural, 3<sup>rd</sup> person). It is still unclear whether there existed cognates in PTG of the innovative  $a=$  in Mundurukú 1st person plural inclusive  $Smaj/A$  marking (with the 1st person marker  $*a-$  being a possible candidate). As yet, there is no strong hypothesis for how these innovative forms arose. However, notice that the plural person markers reconstructed for Proto-Tupí by Rodrigues and Cabral (2012) shown in Table 2 all include the glide  $/j/$ , either as the onset of the last syllable, as in  $o'e=$  '1pl.excl', or as the coda of the monosyllable forms, as in  $vj=$  '1pl.incl' and  $ej=$  '2pl'. Also note that for the first person exclusive form and the second person plural form, the initial vowel of the marker corresponds to the singular form,  $o=$  and  $e=$ , respectively. This could suggest a four-term person marking system may have existed at some point in the history of the Tupian languages, much like that found in many Cariban, Northern Jê, Matacoan and Aymaran languages. It is possible that the glide element results from a putative plural marker that fused with the person marker to express the plural form. A similar argument is made for the development of the person markers and free pronouns in the Tuparí branch by Galucio and Nogueira (2012), where they reconstruct the plural marker  $*-jat$ , resulting in synchronic forms such as  $ejat-$  in Mekens for second person plural. If this is indeed the case for the family in general, it could be possible that the  $a=$  marker found in Mundurukú is a reflex of the heretofore unidentified 1<sup>st</sup> person plural inclusive person marker before the accretion of the additional plural element. Such a diachronic pathway can help to explain the difficulty in identifying clear reflexes of this reconstructed form in the daughter languages (Rodrigues; Cabral 2012, p. 544). While clearly speculative at this point, such a hypothesis deserves further examination.

Gildea (2002) suggests that innovative forms found in PTG arose through the cliticization of a set of free pronouns that were later lost before the formation of PTG. If at some point in the history of the family there did exist a set of free pronouns that developed into the  $S_{maj}/A$  prefix set in PTG, it would be expected that reflexes of this set could be found in the free pronouns of the languages that did not develop a set of  $S_{maj}/A$  marking prefixes, as in the Arikém, Tuparí, Mondé, Ramarama and Juruna branches. Free pronouns of selected members of these branches are given in Table 3.

As can be seen by comparing the free pronouns shown in Table 3 with the prefix/proclitic sets in Table 2 for the same languages, the bound sets appear to be phonologically reduced forms deriving from the free pronoun sets rather than representing a separate diachronic development, with 3<sup>rd</sup> person forms being notable exceptions. The free pronouns and proclitic set in Proto-Tupí as reconstructed in Rodrigues and Cabral (2012) also maintain the same distinction in the sense that the latter are phonologically reduced forms of the former. At present, there still does not seem to be a clearly identifiable source for the innovated forms that index  $S_{maj}$  and  $A$  in the expansionist languages that mark such arguments.

Table 3. Free pronouns in some non-TG Tupian languages.

Language	Branch	1sg	2sg	3sg	1pl.incl	1pl.excl	2pl	3pl
Karitiana	Arikém	in	ãn	i	itʃa	ita	atʃa	i
Mekens	Tuparí	õt	ět	te	kise	ose	ejat	tejat
Suruí	Mondé	oen	een	xien	paen	tojen	méjen	táen
Karo	Ramarama	õn	ẽn	at / ŋa	iʔtə	té	kaʔto	tap
Juruna	Juruna	(u)na	ena	amĩ	sĩ	(uluʔ)udĩ	esĩ	anĩdai

## CONCLUSIONS

The diversity of different verbal argument marking patterns across the Tupian language family provides an interesting case study for the application of both computational and traditional techniques to make inferences about the morphosyntactic diachrony of a language family. The parsimony analyses tend to support the claim that Proto-Tupí originally had a system of indexation that marked absolutive arguments on the predicate. From there, the system began to change in the languages that spread outwards from Rondônia, while the branches that remained in Rondônia maintained the absolutive pattern. Early on in the history of the expansionist group of languages there was a restriction on the indexation of intransitive subjects with the ancestral marker set, resulting in the indexation of only the O argument in Xipaya and Mundurukú imperfective constructions and the indexation of both O and  $S_{min}$  in Juruna. The hierarchical pattern found in the Mundurukú and Mawetí-Guaraní branches is the result of a single gradual development due to the retention of cognate forms across these groups that are not found elsewhere in the family. The PTG  $S_{min}/O$  markers then developed from a reanalysis of the free pronouns into verbal proclitics resulting from the addition of the formative element \*e.

While much work remains to be done on the reconstruction of Tupian phonology and grammar, this paper has confirmed some previous proposals on the earlier stages of development of the verbal argument marking system while putting forth some novel ideas on when and how these changes developed. These proposals highlight the importance of including both classificatory hypotheses and data from a wide range of languages in future work on the morphosyntactic reconstruction within the Tupian language family.

## ACKNOWLEDGEMENTS

I would like to acknowledge the support of the Languages in Contact research group in Nijmegen funded by grants from the European Research Council and Dutch Royal Academy of Sciences (KNAW) awarded to Pieter Muysken. Thanks go to the Tupian experts who enthusiastically discussed their languages of study with me, notably Aline da Cruz, Sebastian Drude, Vilacy Galucio, Hebe González, Suzi Lima, Denny Moore, Gessiane Picango, Françoise Rose and Luciana Storto. The paper was substantially improved thanks to comments from the editors and one anonymous reviewer, as well as Mily Crevels, Michael Cysouw, Rik van Gijn, Harald Hammarström, Sérgio Meira and Pieter Muysken. However, those mentioned do not necessarily share the views adopted in this paper, and all errors are my own.

## REFERENCES

- BORELLA, Cristina de Cássia. *Aspectos morfosintáticos da língua Awetí (tupi)*. Universidade Estadual de Campinas, 2000.
- BRAGA, Alzerinda de Oliveira. *Aspects morphosyntaxiques de la langue Makurap/Tupi*. Université de Toulouse -- Le Mirail, 2005.

CABRAL, Ana Suelly Arruda Câmara. Natureza e direções das mudanças de alinhamento ocorridas no tronco Tupí. In: QUEIXALÓS, Francesc (Org.). **Ergatividade na Amazônia**. Paris and Brasília: Centre d'études des langues indigènes d'Amérique and the Laboratório de Línguas Indígenas, 2002, v. 1, p. 5-7.

COMRIE, Bernard. **Language Universals and Linguistic Typology: Syntax and Morphology**. 2 ed. Oxford: Blackwell, 1989.

CORREIA-DA-SILVA, Beatriz Carretta. **Mawé/Awetí/Tupí-Guaraní: Relações linguísticas e implicações históricas**. Universidade de Brasília, 2010.

CRUZ, Aline da. **Fonologia e gramática do Nheengatú: A língua geral falada pelos povos Baré, Warekena e Baniwa**. Amsterdam: Vrije Universiteit, 2011.

CYSOUW, Michael. Reconstruction of morphosyntactic function: Nonspatial usage of spatial case marking in Tsezic. **Language**, v. 85, n. 3, p. 588-617, 2009.

DIXON, Robert M W. **Ergativity**. Cambridge: Cambridge University Press, 1994.

DRUDE, Sebastian. "Derivational verbs" and other multi-verb constructions in Awetí and Tupí-Guaraní. In: AIKHENVALD, Alexandra Y; MUYSKEN, Pieter C (Orgs.). **Multi-verb Constructions: A View from the Americas**. Leiden: Brill, 2011, p. 213-254.

EVERETT, Caleb. **Gestural, perceptual, and conceptual patterns in Karitiana**. Rice University, 2006.

FARGETTI, Cristina Martins. **Estudo fonológico e morfossintático da língua Juruna**. Universidade Estadual de Campinas, 2001.

FOLEY, William A. Semantic Parameters and the Unaccusative Split in the Austronesian Language Family. **Studies in Language**, v. 20, n. 2, p.385-430, 2005.

FITCH, Walter M. Toward defining the course of evolution: minimum change for a specified tree topology. **Systematic Zoology**, v. 20, n. 4, p. 406-416, 1971.

GABAS JR., Nilson. **A grammar of Karo, Tupí (Brazil)**. University of California Santa Barbara: 1999.

GALUCIO, Ana Vilacy. **The morphosyntax of Mekens (Tupi)**. University of Chicago, 2001.

GALUCIO, Ana Vilacy. Puruborá: Notas etnográficas e linguísticas recentes. **Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas**, v. 1, n. 2, p. 159-192, 2005.

GALUCIO, Ana Vilacy. NOGUEIRA, Antônia Fernanda. Comparative study of the Tupari branch of the Tupian family: contributions to understanding its historical development and internal classification. **Memorias del V Congreso de Idiomas Indígenas de Latinoamérica, 6-8 de octubre 2011, Universidad de Texas**. Austin: University of Texas, 2012

GILDEA, Spike. **On Reconstructing Grammar: Comparative Cariban Morphosyntax**. Oxford: Oxford University Press, 1998.

GILDEA, Spike. Reconstructing Pre-Proto-Tupí-Guarani main clause grammar. In: **Atas do I Encontro Internacional do grupo de trabalho sobre línguas indígenas da ANPOLL**. Belém: Universidade Federal do Pará, 2002, v. I, p. 315-326.

GIVÓN, Talmy. Internal reconstruction: As method, as theory. In: GILDEA, Spike (Org.). **Reconstructing Grammar: Comparative Linguistics and Grammaticalization**. Amsterdam: John Benjamins, 2000, p. 107-159.

GOMES, Dionei Moreira. **Estudo morfológico e sintático da língua Mundurukú**. Universidade de Brasília, 2006.

GONZÁLEZ, Hebe A. **A grammar of Tapiete (Tupi-Guarani)**. University of Pittsburgh, 2005.

HARRIS, Alice C; CAMPBELL, Lyle. **Historical Syntax in a Cross-linguistic Perspective**. Cambridge: Cambridge University Press, 1995.

HASPELMATH, Martin. Argument indexing: A conceptual framework for the syntactic status of bound person forms. In: BAKKER, Dik; HASPELMATH, Martin (Orgs.). **Languages across Boundaries: Studies in Memory of Anna Siewierska**. Berlin: Mouton de Gruyter, 2013, p.197-225.



- HEATH, Jeffrey. Language contact and language change. **Annual Review of Anthropology**, v. 13, p. 367-384, 1984.
- HOLMAN, Eric; WICHMANN, Søren; BROWN, Cecil H; VELUPILLAI, Viveka; MULLER, André; BAKKER, Dik. Explorations in automated language classification. **Folia Linguistica**, v. 42, n.2, p. 331-354, 2008.
- JENSEN, Cheryl. Cross-referencing changes in some Tupí-Guaraní languages. In: PAYNE, Doris L (Org.). **Amazonian Linguistics: Studies in Lowland South American Languages**. Austin: University of Texas Press, 1990, p. 117-158.
- JENSEN, Cheryl. Comparative Tupí-Guaraní morphosyntax. In: DERBYSHIRE, Desmond C; PULLUM, Geoffrey K (Org.). **Handbook of Amazonian Languages**. Berlin: Mouton de Gruyter, 1998, v. 4, p. 487-618.
- JENSEN, Cheryl. Tupi-Guarani. In: DIXON, R M W; AIKHENVALD, Alexandra (Orgs.). **The Amazonian Languages**. Cambridge: Cambridge University Press, 1999, p. 125-164.
- LAZARD, Gilbert. Transitivity revisited as an example of a more strict approach in typological research. **Folia Linguistica**, v. 36, p. 141-190, 2002.
- LIMA, Suzi Oliveira de. **A estrutura argumental dos verbos na língua Juruna (Yudja)**. Universidade de São Paulo, 2008.
- MADDISON, Wayne P; MADDISON, David R. **Mesquite: a modular system for evolutionary analysis. Version 2.75**. Available at: <<http://mesquiteproject.org>>.
- MEIRA, Sérgio; DRUDE, Sebastian. Sobre a origem histórica dos "prefixos relacionais" das línguas Tupí-Guaraní. **Cadernos de Etnolinguística**, v.5, n.1, p. 1-30, 2013.
- MONSERRAT, Ruth Maria Fonini. Prefixos pessoais em Aweti. **Publicações do Museu Nacional, Série Linguística III**, 1976.
- MONSERRAT, Ruth Maria Fonini. Notícia sobre a língua Puruborá. In: RODRIGUES, Aryon Dall'Igna; CABRAL, Ana Suelly Arruda Câmara (Orgs.). **Novos Estudos sobre Línguas Indígenas**. Brasília: Editora UnB, p. 9-22, 2005.
- MONSERRAT, Ruth Maria Fonini; SOARES, Marília Facó. Hierarquia referencial em línguas Tupi. **Ensaio de Linguística**, v. 9, p. 164-187, 1983.
- MOORE, Denny. **Syntax of the language of the Gavião Indians of Rondônia, Brazil**. New York University, 1984.
- NOELLI, Francisco Silva. As hipóteses sobre o centro de origem e rotas de expansão dos Tupi. **Revista de Antropologia**, v. 39, p. 7-53, 1996.
- NOGUEIRA, Antonia Fernanda. **Wayoro êmêto: Fonologia segmental e morfossintaxe verbal**. Universidade de São Paulo, 2011.
- PAYNE, Doris. The Tupí-Guaraní inverse. In: FOX, Barbara; HOPPER, Paul J (Org.). **Voice: Form and Function**. Amsterdam: John Benjamins, 1994, p. 313-340.
- RODRIGUES, Aryon D. Morfologia do verbo Tupí. **Letras**, v. 1, p. 121-152, 1953.
- RODRIGUES, Aryon D. Classification of Tupi-Guarani. **International Journal of American Linguistics**. v. 24, v.3, p. 231-234, 1958.
- RODRIGUES, Aryon D. Evidence for Tupi-Carib Relationships. In: MANELIS KLEIN, Harriet E; STARK, Louisa (Orgs.). **South American Indian Languages: Retrospect and Prospect**. Austin: University of Texas Press, 1985, p. 371-404.
- RODRIGUES, Aryon D. You and I = neither You nor I: The personal system of Tupinambá. In: PAYNE, Doris L (Org.). **Amazonian Linguistics: Studies in Lowland South American Languages**. Austin: University of Texas Press, 1990, p. 393-405.
- RODRIGUES, Aryon D. Tupi. In: DIXON, R M W; AIKHENVALD, Alexandra (Orgs.). **The Amazonian Languages**. Cambridge: Cambridge University Press, 1999, p. 107-124.
- RODRIGUES, Aryon D. As consoantes do Proto-Tupí. In: CABRAL, Ana Suelly Arruda Câmara; RODRIGUES, Aryon D (Orgs.). **Línguas e Culturas Tupí**. Campinas: Editora Curt Nimuendajú, 2007, v. 1, p. 167-203.
- RODRIGUES, Aryon D; CABRAL, Ana Suelly Arruda Câmara. Tupian. In: CAMPBELL, Lyle; GRONDONA, Veronica (Orgs.). **The Indigenous Languages of South America**. Berlin: De Gruyter Mouton, 2012, p. 495-574.



RODRIGUES, Aryon D; DIETRICH, Wolf. On the linguistic relationship between Mawé and Tupí-Guaraní. **Diachronica**, v. 14, n. 2, p. 265-304, 1997.

RODRIGUES, Carmen Lúcia Reis. **Étude morphosyntaxique de la langue Xipaya (Brésil)**. Université Paris VII, 1995.

ROSE, Françoise. **Grammaire de l'Émérillon Teko, une Langue Tupi-Guarani de Guyane Française**. Louvain: Peeters, 2011.

SAITOU, N; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, v. 4, n. 4, p. 406-425, 1987.

SCHLEICHER, Charles O. **Comparative and internal reconstruction of the Tupi-Guarani language family**. University of Wisconsin, 1998.

SEKI, Lucy. **Gramática Kamaiurá: Língua Tupi-Guarani do Alto Xingu**. Campinas: Editora Unicamp, 2000.

SILVA, Raynice Geraldine Pereira da. **Estudo morfossintático da língua Sateré-Mawé**. Campinas: Universidade Estadual de Campinas, 2010.

VAN DER MEER, Tine. Case marking in Suruí. In: FORTUNE, David L (Org.). **Porto Velho Working Papers**. Brasília: Summer Institute of Linguistics, 1985, p. 208-230.

WALKER, Robert S; WICHMANN, Søren; MAILUND, Thomas; ATKISSON, Curtis J. Cultural phylogenetics of the Tupi language family in lowland South America. **PLoS ONE**, v. 7, n. 4, 2012.

ZÚÑIGA, Fernando. **Deixis and Alignment: Inverse systems in indigenous languages of the Americas**. Amsterdam: John Benjamins, 2006.