

Subordination strategies in Tupian languages Estratégias de subordinação em línguas Tupi

Rik van Gijn^I, Ana Vilacy Galucio^{II}, Antonia Fernanda Nogueira^{III}

^IUniversity of Zürich, Zürich, Switzerland

^{II}Museu Paraense Emílio Goeldi, Belém, Pará, Brasil

^{III}Universidade Federal do Pará, Breves, Pará, Brasil

Abstract: Assessing the internal coherence and constituency of language families often centers either around comparing certain form-meaning correspondences, or around identifying the presence or absence of linguistic features across the members of the family. The former approach is generally restricted to the lexicon. The latter approach focuses mostly on structural characteristics of language. In this paper we present an alternative approach to comparing grammatical systems between languages within a language family, which aims at bringing these two approaches and their results closer to each other. We look at subordination strategies in a sample of Tupian languages, taking constructions as the basic unit of comparison, treating them as form-meaning correspondences. The Tupian family offers an especially intriguing case for studying subordination strategies in the South American context, given its enormous geographical spread and the variety of contact situations involving its member languages. Major patterns of subordination strategies can be discerned across the family, e.g. strategies involving nominalization, verbal incorporation and other subtypes of verbal serialization, but there is also a great degree of variability between the different languages. By mapping the structural diversity onto the known genealogy and geographic distribution, we hope to shed more light on the history of the Tupian family and on the diffusability of subordination strategies.

Keywords: Subordination. Tupian. Constructions. Comparative linguistics.

Resumo: Avaliar a coerência interna e a composição de famílias de línguas, frequentemente, gira ou em torno de comparar certas correspondências de forma-significado, ou em torno de identificar a presença ou ausência de características linguísticas entre os membros da família. A primeira abordagem é geralmente restrita ao léxico. A última abordagem concentra-se principalmente nas características estruturais da língua. Neste artigo, apresentamos uma abordagem alternativa para comparar sistemas gramaticais entre línguas dentro de uma família linguística, que visa a aproximar essas duas abordagens e seus resultados. Nós investigamos estratégias de subordinação em uma amostra de línguas Tupi, tendo construções como a unidade básica de comparação e tratando-as como correspondências de forma-significado. A família Tupi oferece um caso especialmente interessante para estudar estratégias de subordinação no contexto sul-americano, dada a sua enorme dispersão geográfica e a variedade de situações de contato envolvendo as línguas da família. Padrões gerais de estratégias de subordinação podem ser identificados para a família, por exemplo, estratégias envolvendo nominalização, incorporação verbal e outros subtipos de serialização verbal, mas há também um grande grau de variabilidade entre as diferentes línguas. Ao mapear a diversidade estrutural na genealogia e distribuição geográfica conhecidas, esperamos lançar mais luz sobre a história da família Tupi e a difusibilidade de estratégias de subordinação.

Palavras chave: Subordinação. Tupi. Construções. Linguística comparativa.

GIJN, Rik van; GALUCIO, Ana Vilacy; NOGUEIRA, Antonia Fernanda. Subordination strategies in Tupian languages. **Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas**, v. 10, n. 2, p. 297-324 maio-ago. 2015. DOI: <http://dx.doi.org/10.1590/1981-81222015000200006>.
Autor para correspondência: Plattenstrasse 54 ,CH-8032 Zürich, Switzerland. E-mail: erik.vangijn@uzh.ch.

Recebido em 11/02/2014

Aprovado em 07/05/2015



INTRODUCTION

Comparative historical linguistics has a long and respected history, in particular in relation to the lexicon and the reconstruction of sound systems. The Comparative Method involves the systematic comparison of form-meaning pairs, usually restricted to the lexicon and individual bound morphemes (McMahon; McMahon, 2005). There is much less established tradition in the comparison of grammatical features. More often than not, comparative (especially typological) research that targets grammatical (sub)systems within families does so by identifying the presence or absence of grammatical linguistic features, or subsystems across languages. Since these traditions tend to be of a rather different nature, it is often rather hard to compare the results they yield.

The approach to the comparison of structural features we introduce in this paper is intended to bring the two traditions sketched above closer together, hopefully allowing for a more natural intertwining of their results and a fruitful interaction between them. The idea behind the approach is to regard constructions as the basic unit of comparison, treating them as form-meaning pairings, much in the same way as one would do with lexical items, in line with Construction Grammar approaches to language (Goldberg, 2006). The chief difficulty is to ensure that constructions across languages are comparable. We do this by breaking down the formal component of constructions into several smaller variables, thus avoiding imposing predefined theoretical linguistic categories onto languages. At the same time, this enables us to take both a broad and reasonably fine-grained perspective on grammatical subsystems, making them maximally comparable. The main usefulness of the proposed method lies in presenting the bigger picture, and raising potential hypotheses regarding genealogical development and/or diffusion through contact. It also marks a starting point for more detailed studies by language or family specialists. Moreover, it has the potential of allowing for a more fluid distinction between grammar and lexicon.

More specifically we look at ‘subordination strategies’ in a genealogically and geographically representative sample of Tupian languages. The distribution of recurring subordination strategies across families in South America suggests patterns of diffusion, in spite of the alleged conservatism of subordinate structures (Van Gijn *et al.*, 2011). Given its huge geographical spread and consequently the many different situations of contact speakers of Tupian languages find themselves in, the Tupian stock offers one of the richest bases for doing this type of comparative within-family research in South America. We offer a brief overview of the major patterns of subordination strategies found in Tupian languages.

Nominalization has been reported as a common subordination strategy among Tupian languages (Brandon; Seki, 1984). In a brief overview of Tupian subordinate constructions, Rodrigues and Cabral (2012) mention three types of subordinate events: adverbial (temporal/conditional) subordination, complementation and relativization, highlighting the role played by nominalization in expressing these subordinate relations. In languages of different Tupian branches, nominalization strategies, combined or not with other structures, are used to denote subordinate events, expressing various semantic relations, as for instance Mekens adverbial constructions (Galucio, 2011), Kamaiurá relative constructions (Seki, 2000a), and adverbial constructions in various language branches, including Tupi-Guarani (TG), Mundurucu, Mawé, and Awetí (Rodrigues; Cabral, 2003)¹. The widespread occurrence of these constructions in a number of distinct syntactic contexts in the family and the identification of cognate derivational morphemes has been seen as evidence that nominalization is an old strategy already employed in Proto-Tupi to express sentential complements, nominal, and adverbial modification at the clause level (Rodrigues; Cabral, 2012). Another view is presented in Storto (2012), who provides an in-depth discussion of

¹ The authors discuss two types of dependent clauses, which they term gerundive clauses and subjunctive clauses. The former are described as coreferent dependent clauses expressing final, simultaneous or sequential events, and the latter either as coreferent or as non-coreferent dependent clauses, also expressing temporal/conditional (simultaneous, successive and conditional) events.

subordination in the Karitiana language, arguing for an analysis of embedded (adverbial, complement and relative) clauses in this language as truncated version of independent clauses, rather than nominalized constructions.

Multiple occurrence of verb phrases in a single clause is another type of dependent clause structure reported as a common Tupian feature, as discussed extensively by Jensen (1990, 1998, 1999) for TG languages, and by Moore (1994) for the whole family. For TG languages, subordination is usually discussed in terms of dependent/independent verb forms and their indexation system. Jensen (1990, 1998) reconstructs three major types of dependent constructions for Proto-Tupi-Guarani (PTG), characterized by formal subordinating marking and distinct indexation system: oblique-topicalized², temporal/conditional subordinate, and gerundive or dependent serial verbs. The indexation system in these dependent verb constructions is absolutive, which differs from the split intransitive and hierarchical indexation system of independent verbs (Jensen, 1998, 1999). The gerundive or serial verb construction involves two or more verbs with coreferent subjects in a single clause, and expresses actions perceived as part of a single event (Jensen, 1990, 1998). As in the case of nominalization, the widespread occurrence of this construction in the Tupian family and the possibility of grammatical and phonological reconstruction of the (gerundive or serial) verb suffix (cf. Jensen, 1990) are seen as evidence that this construction is an old strategy in the family. Considering especially non-Tupi-Guarani languages, Moore (1994, p. 155-157) also postulates the occurrence of multiple independent VPs in a single clause as an old feature already present in Proto-Tupi. In addition to being consistent with OV order, another original Proto-Tupi feature, multiple VPs are very frequent in the Tupian languages. However, Moore argues against an analysis of multiple VPs in terms of serial verb construction on the basis of their structural properties (scope of adverbial modification, possibility of discontinuity or non-adjacency). The finiteness status of these serial or multiple verb constructions is another relevant topic. Moore (1994) describes two main structures for clauses with multiple VPs, one in which there is a finite VP followed by any number of non-finite VPs with or without overt gerundive (or serial) marking, and another in which all the multiple VPs in the clause are non-finite. This latter case generally occurs in languages such as Gavião and Karo, where there is an obligatory finite auxiliary inflected for TAM in the clause. Rose (2013, p. 34-36) presents an analysis of PTG dependent clauses, including the serial verb constructions, as non-finite clauses. Based on the similarity of their morphosyntactic structure to nominal phrases (dependent marking similar to lexical nominalization markers, indexation system similar to nominal indexation system, negation and TAM coding shared with nominal phrases), she argues that PTG dependent clauses appear next to nominal phrases, thus closer to the non-finite end of the finiteness continuum.

The paper is structured as follows: first we present the composition of the language sample used and the comparative method we propose. Then we present the results of the comparison in terms of the similarity clusters we found for the different constructions. The following section contains a discussion of the results, as well as an alternative perspective on the constructions, taking into account only formal properties. The final section summarizes the main points of the paper.

PRELIMINARIES

THE TUPIAN LINGUISTIC FAMILY AND THE SAMPLE

The Tupian linguistic family is one of the largest and most widely distributed indigenous groups in Lowland South America. It is traditionally divided in ten branches (Rodrigues, 1984, 1985) that together comprise about 40-45 languages, not counting

² The oblique-topicalized constructions are not directly relevant for the scope of this paper. It is not a proper subordinate construction, but rather a non-canonical type of main clause, conditioned by the fronting of an oblique, and showing the formal properties of dependent clauses (Rose, 2013, p. 32).

the dialectal differences between distinct ethnic groups (Moore *et al.*, 2008): Arikém, Mondé, Puruborá, Ramarama, Tupari, Mundurukú, Jurúna, Awetí, Mawé, and Tupi-Guarani. Some recent studies have proposed intermediary stages in the derivation from Proto-Tupi. There has been a proposal of Karo and Puruborá as being closer to each other than to any of the other branches, thus forming a subgroup (Galucio; Gabas, 2002), and also the grouping of Mawé, Awetí and Tupi-Guarani into a larger Maweti-Guarani branch (Rodrigues; Dietrich, 1997; Drude, 2006; Correa da Silva, 2011; Drude; Meira, this volume). The genetic relationship and internal classification of the Tupian family is shown schematically in Figure 1, which reflects the results of the proposed intermediary stages in the derivation from Proto-Tupi, and internal classification of the Tupari (Galucio; Nogueira, 2012) and Mondé branches (Moore, 2005)³.

The current geographical distribution of the Tupian languages is given in figure 2 (courtesy of Love Eriksen).

The Mawe-Aweti-Tupi-Guarani, Mundurukú and Jurúna branches are the more widespread branches, especially the Tupi-Guarani branch which is at the same time the largest and most widespread group. It contains approximately 22 languages and 40 dialectal variants, corresponding to half of the Tupian languages, distributed in a vast geographic area in lowland South America. In Brazil, the Tupi-Guarani languages are spoken from north to south in the states of Amapá, Amazonas, Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Pará, Paraná, Rondônia, Rio de Janeiro, Santa Catarina, and São Paulo. They are also found in adjacent areas of Brazil, in Argentina, Paraguay, Peru, Bolivia, and French Guiana.

The other five branches (Arikém, Mondé, Puruborá, Ramarama, Tupari) are restricted to the Brazilian state of Rondônia, near the Brazilian-Bolivian border. Based on the estimated time depth of the Tupian diversity in that region it has been considered as the most probable homeland for the Tupian languages and the origin of the Tupian expansion (Métraux, 1928; Rodrigues, 1964).

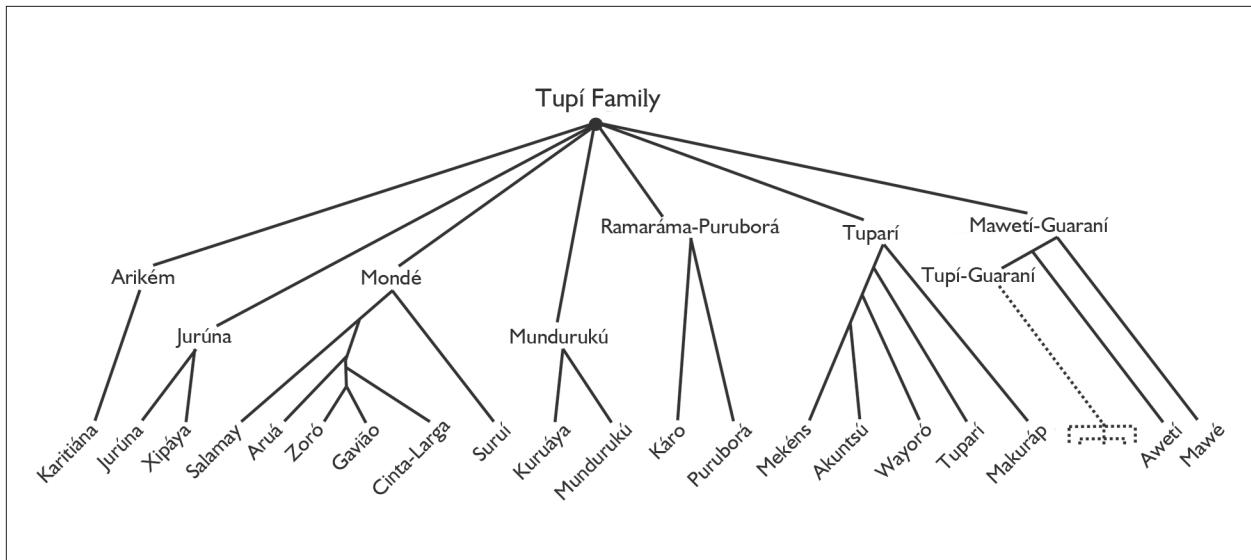


Figure 1. Internal classification of the Tupi family.

³ The dotted lines under the Tupi-Guarani node indicate that its several languages and subbranchings are not shown in the diagram by lack of space.



Figure 2. The distribution of the extant Tupian languages (L. Eriksen).



For the present paper, we have chosen a genetically and geographically diverse sample of Tupian languages, in order to investigate the different subordination strategies found across the family. We have coded the patterns of subordination strategies in ten languages, representing seven branches. Four of the languages/branches are spoken in the state of Rondônia, Brazil: Mekens (Tupari), Karo (Ramarama), Karitiana (Arikém), and Gavião (Mondé), two languages belong to branches spoken outside Rondônia: Mundurukú (Mundurukú) and Sateré-Mawé (Mawé). Tupi-Guarani, the largest branch of the family, is represented by four languages (Emérillon, Kamaiurá, Kokama and Tapiete)⁴ that are both geographically and genetically distant, i.e., belonging to distinct subgroupings in the TG branch. Emerillon is spoken in inland French Guiana, Kamaiurá in the Brazilian Amazon (Parque Indígena do Xingu), Kokama⁵ in the Amazon region with population in three countries (Brazil, Peru and Colombia), and Tapiete in the Argentinean Chaco.

SUBORDINATION STRATEGIES

For this paper we look at subordination strategies from a semantic point of view. This means that we are less concerned about whether a given construction constitutes a proper example of subordination *stricto sensu*, and more with how languages express the relation between two events, or states of affairs, that are often expressed by subordinate constructions. This potentially yields constructions of a very different nature⁶. Therefore, in order to maintain comparability between constructions, a questionnaire has been designed. We will briefly outline this questionnaire here, but space limitations do not permit us to go into too much detail. For a more extensive discussion the reader is referred to van Gijn and Hammarström, (forthc.); for an application to a different dataset see van Gijn (2014).

The basic premise for the questionnaire, following approaches in Construction Grammar (Goldberg, 2006), is that constructions are considered as form-meaning pairs, comparable to words. Thus, each relevant construction (i.e. those that can encode one of the semantic relation types defined below) has a form component and a meaning component. The questionnaire connects a number of semantically defined fields to the morphosyntactic profiles of their expression. The choice of semantic fields or independent variables is basically determined by three factors: i) they are an adapted subset of the categories used by Cristofaro (2003), which allows for a comparison of the South American patterns with the global patterns found by Cristofaro; ii) they should yield a reasonably good representation of the subordination strategies a language employs, meaning that semantic relation types are chosen that are expected to yield to different results, iii) information should be available from grammars for the majority of them, restricting the categories to the most common types. These considerations have led to the semantic types as given in Table 1⁷.

⁴ Since there is more than one proposed classification of Tupi-Guarani languages into subgroupings (e.g. Rodrigues, 1984, 1985; Schleicher, 1998; Mello 2000; Rodrigues; Cabral, 2002), and no agreement among specialists regarding the actual position of some of the languages, we will not refer to any specific subclassification in the paper, except to note that in all the available classifications the TG languages in our sample belong to different subgroupings within the TG branch.

⁵ The variety considered in this study is Kokama-Kokamilla, as described by Vallejos (2010) spoken mainly in Peru. Earlier classifications had classified it as a Tupi-Guaranian language (Loutkotka, 1968; Rodrigues; 1958, 1984, 1985). More recently, it has been analyzed as an abrupt-creole or mixed language (Cabral, 1995). However, the variety described by Vallejos, though showing features of a contact situation and traces of distinct (not all identified) languages that contributed to its current grammatical pattern, retains several features that justify its classification as a Tupi-Guaranian language, albeit one that has changed gradually over the centuries (Vallejos, 2010, p. 753-58).

⁶ We have excluded coordinate constructions, i.e. those constructions that show a (morphosyntactically and pragmatically) symmetrical relation between the event-denoting units; co-subordinate constructions (Van Valin; LaPolla 1997) are included. For more information see van Gijn & Hammarström, forthc.

⁷ SoA stands for State of Affairs, defined as entities that “can be located in relative time and can be evaluated in terms of their reality status” (Hengeveld; MacKenzie, 2008, p. 166). The subscripts ‘M’ and ‘S’ stand for ‘main’ and ‘subordinate’, respectively.

Table 1. Semantic relation types for subordination strategies.

Relation type	Definition
Temporal simultaneous	SoA _S places SoA _M in a temporal perspective, indicating that SoA _M takes place at a moment that overlaps in time with SoA _S .
Temporal successive	SoA _S places SoA _M in a temporal perspective, indicating that SoA _M takes place at a moment that follows in time to SoA _S .
Reason	SoA _S places SoA _M in a perspective of a cause-result chain, providing a motivation for the occurrence of SoA _M .
Location	SoA _M is carried out at a place or with the goal to reach the place where SoA _S occurs/occurred/will occur
Purpose of motion ⁸	SoA _M describes a motion event, which is carried out in order to bring about SoA _S .
Purpose general	SoA _M is carried out in order to bring about SoA _S .
Avertive	SoA _M is carried out in order to prevent SoA _S from occurring.
Potential condition	SoA _S describes an unrealized but possible situation that needs to be fulfilled in order for the SoA _M to occur.
Counterfactual condition	SoA _S describes an unrealized and no longer possible situation that needs to be fulfilled in order for the SoA _M to occur.
Phasal (terminative)	SoA _M indicates that some entity discontinues the temporal development of SoA _S in which s/he is involved as an agent
Modal (ability)	SoA _M indicates that some entity is capable of performing SoA _S .
Desiderative (want)	SoA _M expresses a wish on the part of an experiencer that a potential SoA _S will be realized; at the time of SoA _M , SoA _S is unrealized and possible.
Direct manipulation	SoA _M expresses a manipulative act by which an agent purposefully and successfully brings some other entity to carry out an action or to be in a state described by SoA _S .
Knowledge	SoA _M expresses a state of knowledge of some entity about a proposition, described by SoA _S .
Perception	SoA _M expresses an act of perception (we focus on visual perception); SoA _S expresses the state or action which is perceived.
Indirect utterance	SoA _M expresses an act of information transfer (we focus on proxies of 'say'); SoA _S expresses the information transferred, in the form of the propositional content of someone's utterance, from the perspective of the agent of SoA _M .
Equative evaluation	SoA _M expresses an evaluation (we focus on equative '(is) good/bad') of a proposition, which is expressed by SoA _S .
S-relative	SoA _S restricts the reference of some entity that is involved in SoA _M by describing a situation in which this same entity is involved as the only argument (S).
A-relative	SoA _S restricts the reference of some entity that is involved in SoA _M by describing a situation in which this same entity is involved as the initiating or controlling argument (A).
O-relative	SoA _S restricts the reference of some entity that is involved in SoA _M by describing a situation in which this same entity is involved as the patientive, or affected argument (P).

⁸ Purpose of motion and avertive have been selected as separate categories because cross-linguistically they are often encoded by different constructions (cf. Schmidtke-Bode, 2009).

If a language in the sample has a construction that can be used to encode one or more of these semantic fields, it is taken into consideration. By taking this approach, we go well beyond the classic or canonical conception of subordination, including clause combinations, nominalizations, non-finite clauses, serial verb constructions, auxiliary verb constructions, verb-verb compounds and derivational affixes (e.g. desiderative or causative markers). For this reason, we use the term subordination strategies (the entire range of morphosyntactic expressions of the semantic types, not pre-defined but rather suggested by the language data) rather than subordination.

With respect to the formal component of each construction, based on previous typological work (especially Lehmann, 1988; Cristofaro, 2003; Malchukov, 2006; and Bickel, 2010), five major concerns guide the questions in the questionnaire.

- Finiteness/deverbalization: it is often the case that subordinate or dependent predicates have less inflectional possibilities than independent or superordinate predicates. The questions that relate to this parameter are meant to determine what verbal categories can be marked independently on the subordinate event-denoting unit (EDU). Chosen variables for this version of the questionnaire are subject and object agreement, aspect, tense, event modality, and epistemic modality/evidentiality.
- Nominalization: questions relating to this parameter determine how noun-like the EDU is. Chosen variables: the (im) possibility to take case markers/adpositions, to encode the subject or object as a possessor, to be modified by a determiner or attributive demonstrative, to be marked for nominal number, and the ability to trigger agreement on other words.
- Flagging: subordinate EDUs may or may not have a dependency marker associated with them (a marker that is added to the verbal or nominal inflection of the EDU mentioned above, and that does not occur on independent verbal predications). Questions that relate to dependency marking concern the morphological status of the marker (bound or free) and how they are positioned with respect to the subordinate EDU. Other questions concern the additional possibilities of these dependency markers: whether they single out a participant (in the case of participant nominalization) or whether they can indicate switch reference.
- Linearization: this pertains to relative relations, and more specifically to the position of relativized NPs with respect to their restricting EDU, but also to the relative order of main and subordinate EDUs and, as mentioned above, to the position of dependency markers with respect to the subordinate EDU.
- Integration: some of the semantic fields may be encoded by tighter constructions such as serial verb constructions, auxiliary verb constructions, verb-verb compounds and even derivational affixes. Three levels of integration are discerned: (i) combinations of independent EDUs, (ii) constructions where the subordinate and superordinate EDUs are separate, but obligatorily contiguous words, and (iii) morphologically bound combinations of EDUs (affixes and V-V compounds).

Just as word forms can (and often are) broken down into smaller pieces for the purposes of comparison (e.g. into morphemes, syllables, phonemes, phonetic features), so are the morphosyntactic profiles of the constructions in this study. Rather than saying that a particular construction is finite or not, we code for the different morphosyntactic features that can or cannot be present on a subordinate EDU. Similarly, rather than saying a construction is a nominalization, we code for the nominal features that can be present on the subordinate EDU. Of course, languages differ in their inventory of morphosyntactic features. Therefore, the answer possibilities for each question are yes, no, do not know, and does not apply for most questions, some questions have more options.

A final methodological point that should be mentioned here is that a particular formal encoding may be associated with several meanings, and a single meaning may be connected to more than one morphosyntactic profile. In order to deal with these many-to-many relations, each construction can have more than one potential meaning. The semantic component of the

questionnaire consists of questions whether the construction in question can express a given semantic relation type (asked for each of the semantic types in Table 1). Moreover, more than one construction can have a 'yes' in answer to the question whether they can express a particular semantic type (e.g. simultaneous temporal). In the end, what counts as a construction is a pairing of an abstract (i.e. without reference to any specific phonetic realizations) morphosyntactic profile coupled with one or more meanings.

SUBORDINATION STRATEGIES IN TUPIAN LANGUAGES

The total number of constructions identified for the ten languages of the sample is 75. These constructions show particular differences with respect to each other, which can be expressed in a distance measure, on the basis of the questionnaire discussed above. These distances can be visualized by the Neighbor Net representation given in figure 3, which takes into consideration both formal and functional (semantic) aspects of constructions.

Each end node in the network represents a construction. The names of the constructions consist of an ISO code for the language it belongs to, and two arbitrary elements (a number and a name) to uniquely identify the construction. Since figure 3 is rather hard to read, we divide the discussion of subordination strategies in Tupian languages into separate parts. Seven major branches of varying cohesiveness can be distinguished in figure 3. The names in Figure 3 approximately identify the nature of the group of constructions, and at the same time refer to the section in which the group of constructions is discussed. These subordination strategies (i.e. the clusters) have descriptive labels and are identified on the basis of the data and not pre-defined theoretical constructs.

CLAUSE-LIKE TEMPORAL STRUCTURES

The first group of constructions is represented to the right of figure 3, indicated by the label 'temporal'⁹. These temporal clauses seem to form a branch-like structure with the group of constructions to be discussed in the next subsection, but they are sufficiently distant from that group to warrant a separate discussion.

The common denominator for the constructions in this group is that they encode temporal relations, most of them exclusively, some of them in addition to either perception or purpose relations, and in one case (Gavião) to purpose relations and desiderative relations. Morphosyntactically, the subordinate predicates vary in their degree of finiteness. In most constructions, there is some verbal morphology possible on the subordinate predicate, for some constructions the exact extent of verbal potential is unknown. The following example from Gavião shows a verb inflected for tense and aspect¹⁰.

- (1) Gavião; Mondé (Moore, 1984, p. 86)¹¹

<i>eé</i>	<i>bó</i>	<i>tá-mága</i>	<i>[aa-ja-ii</i>	<i>a-já-bóc]</i>
then	TZ	3P-NSJV+ASSERT	3C-face-enter(=grow)	3C-PST-PSF
<i>ê-abí</i>	<i>a-volò-á</i>			
that-after	3C-come-SBM			
'Then they will arrive from there when they have already grown up.'				

⁹ This is not to be thought of as representing all temporal clauses in the sample, but rather similar morphosyntactic strategies that have the additional characteristic of encoding temporal relations. This is also true for the other 'semantic' labels.

¹⁰ We have adapted the glossing of the original examples in some cases for reasons of consistency and transparency. Examples from sources that have been written in Spanish, Portuguese, or French are rendered in English, our translations.

¹¹ Moore (1984) has a complex glossing system for auxiliaries, including 'unmarked' categories. We adopt only those values that either have a positive or negative value.



Figure 3. NeighborNet representation of the distances between the constructions.

In terms of nominal morphology, most constructions do not allow any, apart from one construction in Karo (arr-5) which allows for postposition marking, and one construction in Kamaiurá (kay-4) which allows for possession. Example (2) shows the construction in Karo where the nominalizer *ko* appears at the end of the subordinate clause, but pro-cliticizes to the next word, which may be a postposition. That is the case in (2), where the postposition *kəy* takes the entire nominalized subordinate clause as its object.

(2) Karo, Ramarama (Gabas Jr., 1999, p. 113)

owirup tēy kokəy okap ocagəpto cára koām

o=wirup tēp-t ko=kəy o=kap-ap

1SG=food boil-IND1 NLZ=DAT 1SG=AUX.FUT-IND2

o=cagəp=to? cá-t-a koām

1SG=dish=PL wash-GER also

'While my food boils I will go wash my dishes too.'

Table 2 shows the constructions, their verbal inflectional potential, and the semantic relations they can encode¹².

The constructions in this group are surrounded by many uncertainties, which warrants caution when talking about a general construction type. They have in common that they represent reduced verbal structures, few if any nominalization characteristics, often a dependency marker, and a common semantic denominator of encoding temporal relations. It seems to be present in 'Rondônian' Tupi, but also in Kokama, Sateré-Mawé and Kamaiurá.

Table 2. Characteristics of 'temporal clauses'.

Name	Language	V-features	N-features	Flagging	Function(s)
cod-8	Kokama	?	-	Suffix	Temporal
skf-5	Mekens	Tense, Aspect	-	-	Temporal, perception
mav-6	Sateré-Mawé	Person, Tense, Aspect	-	free (post)	Temporal
cod-6	Kokama	?, epistemic modality	-	free (post)	Temporal
arr-5	Karo	Person	Case	free (post)	Temporal, perception
cod-7	Kokama	?	-	free (post)	Temporal
arr-3	Karo	Person	-	Suffix	Temporal, Purpose
gvo-1	Gavião	Person, Tense, Aspect	-	-	Temporal, Purpose, Desiderative
arr-0	Karo	Person	-	Suffix	Temporal
arr-1	Karo	Person	-	Suffix, free (post)	Temporal
kay-4	Kamaiurá	Person	Possession	Suffix	Temporal
skf-8	Mekens	-	-	Suffix	Temporal, perception

¹² Only the functions that are attested in examples and/or mentioned explicitly are taken into account. Flagging morphemes include nominalizers. Clitics are classified within the group of (semi-) free morphemes.

POLYFUNCTIONAL ADVERBIALS (PREDOMINANTLY CLAUSE-LIKE)

The second group of constructions is primarily connected by the fact that they can encode multifunctional adverbial relations, minimally the combination temporal, reason, and conditional, a few also including purpose relations. The construction in (3) from Tapiete is a case in point. The dependent predicate is marked with the suffix *-rã* and can take most verbal inflection. The same construction can also mark conditional relations, temporal relations and reason relations, the latter optionally in combination with the free marker *yawaiwe* (see 3c).

(3) Tapiete; Tupi-Guarani (González, 2005, p. 240)

- a. *[a-pisi-rã]* *hama* *shi-mba'erasi-kwi*
 1SG.AC-catch-SUB then 1SG.IN-illness-FUT
 'If I touch it, then I get sick.'

- b. *hama* *[we-rú-ha-rã]* *piro-ha* (González, 2005, p. 237)
 then 3AC-bring-IMPER-SUB peel.off-IMPER
 'When they brought it, (they) peeled it.'

- c. *[yawaiwe sanya'i kãñi-ma-rã] opa wi'ää-ha* (González, 2005, p. 237)
 because child lost-RES-SUB all be.sad-1EXCL
 'Because the boy is already lost we are sad.'

Table 3 presents the constructions, their verbal inflectional potential, and the semantic relations they encode. As was the case with the previous construction type, there is some bandwidth in the morphosyntax of the constructions. Most constructions allow for a subset of the verbal features found in independent clauses, but mostly this is restricted. Nominal features are not particularly common, and if they do occur, they are usually rather limited. In Mekens for instance, the construction allows for case marking.

(4) Mekens; Tupari (Galucio, 2011, p. 32)

- kirit* *se-ajt-kwa-t* *[se-akar-ab=ese]*
 child 3C-cry-TR-PST 3C-fall-NLZ=LOC
 'The child cried when he fell down.' [Literally: 'The child cried at his own falling']

Table 3. Characteristics of 'polyfunctional adverbials'.

Name	Language	V-features	N-features	Flagging	Function(s)
skf-0	Mekens	-	Case	Suffix	Temporal, reason, condition
skf-1	Mekens	TMA	Case	Free (post)	Temporal, reason, condition
ktn-0	Karitiana	Aspect	-	-	Temporal, reason, condition
eme-0	Emérillon	Person, TMA	-	Free (post)	Temporal, reason, condition
myu-1	Mundurukú	Person, TMA	?	Free (post)	Temporal, reason, condition, purpose
tpj-2	Tapiete	Person, TA	-	Free (pre, optional) + bound	Temporal, reason, condition, purpose
kay-3	Kamaiurá	Person	Possession	Suffix	Temporal, reason, conditional

For most constructions, there is some kind of dependency marking, either through a suffix, or through a free subordinator.

The group of languages that have this construction type includes members from different branches of the Tupian family: Tupari (Mekens), Arikem (Karitiana), Mundurukú (Mundurukú) and different branches of Tupi-Guarani (Emérillon, Tapiete, Kamaiurá).

REASON (AND CONDITIONAL) CLAUSES

These constructions are mostly employed to encode reason relations (in Sateré-Mawé they can also encode conditionals; in Kokama there is a construction within this group that encodes conditionals) and can be characterized by the fact that they do not take any nominal morphology, and in many cases take a considerable amount of verbal morphology. They are, moreover, all flagged, either by a suffix or by a postposed particle. These characteristics are summarized in Table 4, which gives the constructions, the verbal inflectional potential, the flagging strategy and semantic relations they encode.

- (5) Karo; Ramarama (Gabas Jr., 1999, p. 206)

o=yaʔwat-t iʔke ye [amân ya-t]
 1SG=leave-IND1 NEG REAS [rain fall-IND1]
 'Because it rained, I did not leave.'

- (6) Sateré-Mawé; Mawé (Corrêa da Silva, 2010, p. 211)

hawii a-ti-kat aru pai uito [kue a-ti-oto]
 then 1SG-REL-look.for POT MSP I since 1SG-REL-take.away
iara] Ø-Ø-'e
 canoe 3SG-REL-say
 "‘Since it was me who took the canoe, I will look for it’, he said.’

- (7) Kokama; Tupi-Guarani (Vallejos, 2010, p. 635)

ay ikian awa i tsi-kaka [ni ra=purara pej-ikua]
 already this person have.fear-REC NEG 3SG.M=find way-REAS
 'He is already scared [because he doesn't find the way].'

Table 4. Characteristics of 'reason clauses'.

Name	Language	V-features	N-features	Flagging	Function(s)
arr-2	Karo	Person	-	Suffix (+ free variable)	Reason
cod-4	Kokama	Aspect	-	free (post)	Reason
skf-2	Mekens	TMAE	-	free (post)	Reason
mav-3	Sateré-Mawé	Person, Aspect	-	free (pre)	Reason, Conditional
mav-4	Sateré-Mawé	Person, TA	-	free (variable)	Reason, Conditional
cod-5	Kokama	Tense, Epistemic M.	-	suffix	Conditional

This group of languages is rather small, but covers a relatively wide spectrum of branches. It includes members from different branches of the Tupian family: Ramarama (Karo), Tupi-Guarani (Kokama), Tupari (Mekens), and Mawé (Sateré-Mawé).

MULTI-VERB PURPOSE CLAUSES

Most same-subject purpose-of-motion relations as well as some complementation relations are encoded by a tight V-V construction. These constructions are characterized by the lack of inflection on one of the verbs, and no marker of dependency. Compare the constructions from Mundurukú, Emérillon and Mekens:

- (8) a. Mundurukú; Mundurukú (Crofts, 2004, p. 222)

ey³-bu³but² pu²pum² tu³ e³pe³-sot^{2-1?}
 2PL-thing.belonging.to buy INT 2PL-come
 'Did you (pl) come to buy things?'

- b. Emérillon; Tupi-Guarani (Rose, 2011, p. 324)

o-ʔu(r)-tar kʷi poro-mō-maʔam
 3.I-come-FUT one.day IMPER.II-CAU-stand.up
 'He will come someday to resuscitate (raise) the people.'

- c. Mekens; Tupari (Galucio, 2011, p. 40)

kārā ar-a eba kise-set
 brazil.nuts get-THV EVID 1PL.INCL-leave
 'We had gone to fetch Brazil nuts.'

The three constructions in (8) are functionally and morpho-syntactically similar: the element indicating the purpose event is - in terms of verbal categories - restricted to taking person markers¹³, there is no explicit marking of subordination or other type of connector¹⁴, and no evidence of nominal characteristics¹⁵.

In Mundurukú, this special type of serial verb construction can also be employed to encode phasal relations. This may point to the fact that serial verb constructions (SVCs) may be the source for (some of) the derivational affixes discussed in the section "Affixing and compounding".

- (9) Mundurukú; Mundurukú (Crofts, 2004, p. 209)

kaypat.pan ġuto o'-e
 jump again 3-start
 'He started to jump again.'

¹³ In the case of Mundurukú, aspectual distinctions can also be marked.

¹⁴ In Emérillon, subtraction of the final <r> is a trace of former dependency marking; we thank Françoise Rose for pointing out this fact to us.

¹⁵ Set II prefixes in Emérillon, although also found in independent transitive clauses marking the P argument, do pattern with possessives behaviorily, in that they allow for the use of the coreference marker for third persons (Françoise Rose p.c.).

A related construction to the ones in (8) is found in Kamaiurá, with the addition of a gerund marker on the verb indicating the purpose event:

(10) Kamaiurá; Tupi-Guarani (Seki, 2000b, p. 195)

o-ho ka'i-a juka-m
 3-go monkey-NUC kill-GER
 'He went to kill a monkey.'

A similar construction can be found in Mekens, Kokama and Emérillon (but in the latter marked with a prefix). Slightly more finite structures that also encode purpose clauses (not just of motion) are found in Kamaiurá and Kokama, where aspect can be marked on the verb encoding the purpose event.

(11) Kokama; Tupi-Guarani (Vallejos, 2010, p. 627)

ay kuashi iwati-n ra uyepe tuyuka=ri
 already sun get.up-NLZ 3SG.MSP go.down ground=DIF
tupapenan [r=utsu-ka-tsen]
 again 3SG.M=GO-REI-PURP
 'When the sun is already up, he again goes down to the ground [for him to keep going]'

Multi-verb purpose clauses are found in Mekens (Tupari), Mundurukú (Mundurukú), Sateré-Mawé (Mawé), and in different Tupi-Guaranian languages Kokama, Kamaiurá, and Emérillon.

NOMINALIZED RELATIVES

Contrary to the adverbial relations, the relative relations show much more coherence morpho-syntactically, and thus seem to represent a more stable Tupian construction type. Jensen (1999, p. 160) mentions that most TG languages have a relativizer, but they differ in the extent to which it is used. Some languages only use it for intransitives. Relative clauses in TG languages are generally characterized by a decrease of verbal inflectional potential, and an increase of nominal inflectional possibilities, and can therefore be termed nominalizations. These nominalizations are moreover overtly marked, often by a suffix or an enclitic, which indicates the role (S, A, or P) of the relativized NP in the relative event. Jensen (1998, p. 539-543) reconstructs six nominalizing markers and their respective allomorphs for PTG: **-a* 'action', **-ár* 'agent', **-ab* 'circumstantial', **-pyr* 'patient nominalizer without reference to A', **emi-* 'patient nominalizer in relation to A', and a general clause nominalizer with possibly two allomorphs **-ba'e*/**-ma'e*. Cognates of these nominalizing markers are also found in the other Tupian branches, supporting the reconstruction of four nominalizing markers for Proto-Tupi (Rodrigues; Cabral, 2012, p. 533-537): **-at* 'agent', **-ap* 'circumstantial', **-pit* 'patient', and **-mi* 'object'.

We will exemplify the system of nominalized relatives in Kamaiurá in more detail, and summarize on the basis of this pattern how the other languages in the sample relate to it.

Kamaiurá encodes relativizations by nominalizing the predicate with suffixes. These nominalizing suffixes single out one of the participants in the main verb. For the relativization of an S participant, three nominalizers are used: *-ama'e* (attributive), *-uma'e* (negative attributive), and *-ipyt* (patient); the first two markers attach to active intransitive verbs as

well as descriptive verbs, the difference between these types of verbs is in the marking of the participant. It should be noted that the TG languages present various patterns of argument marking. Independent verbs show an active-stative system for intransitive verbs, and a more heterogeneous system governed by a person hierarchy for transitive verbs, in addition an ergative-absolutive system is found in other types of verbal constructions, mainly dependent constructions (Jensen, 1999, p. 155; Rose, 2013, see also Birchall, this volume, for a more detailed account)¹⁶.

- (12) Kamaiura; Tupi-Guarani (Seki, 2000b, p. 179)

a-mo-y'u rak akwama'e-a i-'ywej-ama'e-her-a
 1SG-CAU-drink at man-NUC 3-be.thirsty-NLZ-PST-NUC
 'I made the man who was thirsty drink.'

The patientive suffix *-ipyt* functions as a de-agentivizer, or passivizer; it really seems to be encoding a P-relativizer. The agent of the underlying transitive verb can be expressed as a dative¹⁷.

- (13) Kamaiura; Tupi-Guarani (Seki, 2000b, p. 179)

o-yk akwama'e-a i-mono-pyr-er-a morerekwar-a upe
 3-arrive man-NUC 3-send-NLZ-PST-NUC chief-NUC DAT
 'The man arrived who was sent by the chief.'

A arguments are relativized by a nominalized verb that carries the marker *-tat*.

- (14) Kamaiura; Tupi-Guarani (Seki, 2000b, p. 179)

akwama'e-a o-juka wyrapy-a kunu'um-a pyhyk-ar-er-a
 man-NUC 3-kill hawk-NUC boy-NUC hit-NLZ-PST-NUC
 'The man killed the hawk that hit the boy.'

The relativized argument can both precede and follow the relative clause, and the RC can also appear on either side of the main verb (Seki, 2000b, p. 183). RCs can also occur without a head. It is unclear what verbal categories can still be marked on the nominalized RC.

¹⁶ The questionnaire is set up such that it allows for the encoding of a wide range of languages (also non-Tupian). In terms of person indexing, it distinguishes first between presence and absence of person marking, and when present whether it is different or the same as in independent clauses. If it is different, it distinguishes between nominal (possessive) and non-nominal (alternative) indexing. The complex patterns of Tupian, especially TG, person indexing are richer than can be represented in the questionnaire, and requires more detailed study. The reader is referred to Birchall (this volume) for more details on person indexing in Tupian languages.

¹⁷ There is a cognate of this morpheme in Mekens, the suffix *-l/pit*, which is an adjectivizer/passivizer, and derives patientive adjective stems corresponding to (stative) intransitive verbs in TG languages, as in example (i) below from Galucio (2001, p. 102).

- (i) *o-ike otat poka-ap o-etobeka-pit ar-a-t*
 1SG-brother fire burn-NLZ 1-loose-ADJZR get-THV-PST
 'My brother found my lost cigarette lighter'

O arguments can be relativized by using the nominalizer *emi-*. The agent is expressed as a possessor, by person proclitics.

- (15) Kamaiura; Tupi-Guarani (Seki, 2000b, p. 179)

je=r-yke'yr-a *w-etsak* *kye'i-a*
 1SG=REL-brother-NUC 3-see knife-NUC
ne=r-emi-ekar-er-a
 2SG=REL-NLZ-search-PST-NUC
 'My brother saw the knife that you were looking for.'

Table 5 gives the dependency markers that the different languages use in their relativization strategies. It can be observed that there is some overlap between the forms of the markers, which can be divided into five groups that can all be related to the proposed reconstructed forms for Proto-Tupi (cf. Rodrigues; Cabral, 2012) and PTG (cf. Jensen, 1998), listed above: a group of markers consisting of a bilabial sound and the vowel /-(w)a(p)/ (Tapiete, Karo, Mundurukú); a group of markers containing the sequence /-(t)at/ (Kamaiurá, Mundurukú, Gavião); a group of the form /-(V)ma'e/ (Kamaiurá, Emérillon); a group of the form /-(V)pyt/ or /-(V)pit/; an arguably more contentious group of markers that are prefixed and containing the vowel /i/ (Kamaiurá, Mekens, Karitiana), related to an old object nominalizer of subordinate clauses morpheme, **mi*, reconstructed for PT by Rodrigues and Cabral (2012)¹⁸.

Considering the meaning and form of the reconstructed markers, the current systems found in the sample languages is clearly direct derived from an old system, but they have been subject to a number of subsequent developments that have changed the individual systems of the languages in question. The clause nominalizer /-(V)ma'e is used for S, A, and P

Table 5. Nominalizers used for relativization.

	<i>S-rel</i>	<i>A-rel</i>	<i>P-rel</i>
Kamaiura	-ama'e (attr.) -uma'e (neg. attr) -ipyt (patient)	-tat	emi-
Tapiete	-(w)a	-(w)a	-(w)a
Emérillon	-ma'ë	-ma'ë	-ma'ë
Kokama	-n	-n	-n
Mekens	-ø (-Vpit)	-ø	i-
Karo	-(a)p (~m)	-(a)p (~m)	?
Karitiana	-ø	-ø	-ø (i-; ti-)
Mundurukú	-(i)at	-(i)at	-(i)ap
Gavião	?	mát	mát

¹⁸ In Karitiana (cf. Storto, 2014), there is a prefix *i-*, analyzed as a participial marker, which nominalizes the complement of the copular verb in cleft constructions, and an object focus morpheme *ti-*, which occurs in main clause non-declarative object focus constructions, object *wh*-clefts and object relatives. The latter morpheme *ti-* is the one considered cognate with the object focus marker *i-* in Mekens, employed in O-relativizations. Despite the distinct synchronic status of the Karitiana morpheme, its cognate status and potential common source as object nominalization in an earlier stage in the history of Tupi warrants its inclusion in this section.

relativization in Eméillon, but appears only in S relativization in Kamaiurá. Nonetheless, the general picture for these markers, then, is one of old shared markers, already in use in the proto-language, and well in place before the branches split off. Moreover, the constructions have rather homogeneous morphosyntax. All constructions are nominalizations in that the marked form acquires certain nominal characteristics, like the ability to take case markers, determiners, nominal number markers, or in some cases the expression of one of the arguments as a possessor.

POLYFUNCTIONAL COMPLEMENTATION STRATEGIES

This “branch” does not have a very strong signal, as it only contains five constructions (the fifth, skf-7, being somewhat peripheral to the branch), and there is a lot of conflicting signal as indicated by the webbing of the branch. Nevertheless, the five constructions come from five different languages (three of which are in the TG branch), and further research may reveal that there are more Tupian languages with similar constructions. The construction is in the first place characterized by the fact that it can perform the function of several types of complementation.

Table 6 indicates, for each of the five constructions, the functions that they can perform, as well as the verbal and nominal features that can be found on the subordinate EDU and the means used to flag the EDU as subordinate.

All of the constructions can encode perception relations, and most of them also cognition and desiderative relations; phasals and ability relations are slightly less widespread. As can be seen in Table 6, in Kamaiurá and Karitiana, the construction also extends to adverbial relation types. The functions of these constructions may reach further than we indicate, to functions that we have not coded for.

In terms of morphosyntax, there is more variation between the constructions, contributing to the diffuse nature of the branch. In general terms, the constructions tend to maintain a number of verbal characteristics, though in a reduced form in some of the cases (e.g. Karitiana, Tapiete, and Kokama). Nominal characteristics are minimal, though present in the constructions in Kamaiurá and Karitiana. Example 16 shows the use of the oblique case marker for Karitiana.

(16) Karitiana; Arikém (Storto, 2012, p. 229)

y-py-sondyp-yn yn [Inacio 'ep opi]-ty
 1-ASSERT-KNOW-NFUT I Inácio tree cut-OBL
 'I know that Inácio cut the tree'

The construction is unmarked for most constructions in this group; i.e. it often just figures a reduced verbal structure, possibly with some nominal characteristics in some languages, and no special flagging as in example (17).

Table 6. Characteristics of ‘polyfunctional complementation strategies’.

Name	Language	V-features	N-features	Flagging	Function(s)
kay-0	Kamaiurá	TMA	Case, Poss (S/O)	Suffix	Perception, Equative, Phasal, Desiderative, Reason, Purpose
ktn-1	Karitiana	Aspect	Case	-	Desiderative, Cognition, Perception, Purpose
tpj-1	Tapiete	Person, Aspect	-	-	Phasal, Ability, Desiderative, Perception, Utterance, Cognition
cod-0	Kokama	Person, Aspect	-	-	Phasal, Ability, Desiderative, Perception, Cognition
skf-7	Mekens	TMA, Evid	-	-	Manipulation, Perception, Cognition

- (17) Tapiete; Tupi-Guarani (González, 2005, p. 373)

pi-yasaka ai-poko-po a-iyeta
 2PL-silent 1SG.AC-begin-FUT 1SG.AC-talk
 'Shut up, I am going to start talking.'

The construction in Kamaiurá is marked with the nominalizer *-tap/-taw*.

- (18) Kamaiurá; Tupi-Guarani (Seki, 2000b, p. 171)

1-potar=ete i-jo-taw-a
 1SG-want=really 3-go-NLZ-NUC
 'I want him to go.'

AFFIXING AND COMPOUNDING (PHASALS, DESIDERATIVES, MANIPULATIVES)

A more coherent branch of figure 3 (i.e. with less reticulation) is the branch called 'Derivational affixes'. These structures are mainly found in semantically tight constructions, in particular in the following three types of complement relations: same-subject desiderative, direct manipulation, and phasal (start/finish). Major subtypes, causing the reticulation in the branch in figure 3, are prefixed versus suffixed derivations. An example of each type of construction comes from Emérillon (desiderative suffix) and Gavião (causative prefix).

- (19) a. Emérillon; Tupi-Guarani (Rose, 2011, p. 351)

apam-a-ně nōde-apisi-taně
 stranger-a-CONTR 1INCL.II-massacre-DES
 'The strangers wanted to massacre us.'

- b. Gavião; Mondé (Stute, 1985, p. 13)

èna áneh méne máh ó-mà-tí-á
 like.that 3SG=AUX.DYN Cmpl AUX.DYN 1SG-CAU-feel.ashamed-SBM
 'That he acted like that made me feel ashamed.'

Example (20) from Kamaiurá shows a third possibility: it has a cessative verb *pik*, which can also be used as a suffix, with the same meaning. The same is true for the desiderative verb *potat* in Kamaiurá. These constructions have been classified as V-V compounds, because both components also occur as free verbs, as in (20b).

- (20) a. Kamaiurá; Tupi-Guarani (Seki, 2000b, p. 133)

a-karu-pik
 1SG-eat-CES
 'I stopped eating.'

- b. Kamaiurá; Tupi-Guarani (Seki, 2000b, p. 199)

n=a-pik-ite we-porahaj-te korin

NEG=1SG-stop-NEG 1SG-dance-GER FUT

'He won't stop dancing.'

Table 7 lists the constructions that make use of a derivational affix, indicating, from left to right, the semantic field(s) that can be encoded by the construction, the languages in which it occurs with the branch of Tupi that the language belongs to between brackets, the category of the affix, and the form of the affix.

There are certainly cognates in the list of affixes in the sample languages: *potat* (Tapiete) and *pota* (Kamaiurá), and perhaps also *tañe* (Emérillon) for the desideratives. For causation, all the languages, except Kokama, have a causative prefix of the form *m(V)-* (Tapiete, Mundurukú, Emérillon, Gavião, Karitiana, Karo, Kamaiurá, Mekens, Mawé), and some have a cognate causative suffix: Tapiete *ka*, Kamaiurá *ukat*, and Emérillon *okal*. The causative prefix correspondence is much more widely shared, and seems to be particularly old. It has been reconstructed both

Table 7. Affixation and compounding as subordination strategies in Tupi languages.

Semantic field	Language (branch)	Cat affix	Form affix
Desiderative (SS)	Tapiete (TG)	suffix (cmp)	pota
	Kamaiurá (TG)	suffix (cmp)	potat
	Kamaiurá (TG)	suffix	wej
	Emérillon (TG)	suffix	tanê
	Sateré-Mawé (Sateré-Mawé)	suffix	
	Karitiana (Arikém)	suffix	wak
Manipulation	Tapiete (TG)	prefix	mbi
	Tapiete (TG)	suffix	ka
	Kamaiurá (TG)	suffix	ukat
	Emérillon (TG)	suffix	okar
	Emérillon (TG)	prefix	bo/mo
	Kokama (TG)	suffix	ta
	Kamaiurá (TG)	prefix	mo
	Mundurukú (Munduruku)	prefix	mu(y)
	Mundurukú (Munduruku)	prefix	duju
	Gavião (Mondé)	prefix	ma
	Mekens (Tupari)	prefix	mô
	Karitiana (Arikém)	prefix	m
	Karo (Ramarama)	prefix	ma
	Sateré-Mawé (Mawé)	prefix	mo
phasal (stop)	Kamaiurá (TG)	suffix (cmp)	pik
	Emérillon (TG)	suffix	pa
phasal (start)	Karitiana (Arikém)	suffix	'oot

for PTG and for Proto-Tupi as **mo-*, a derivational valence changing prefix deriving transitive verbs with causative function (Rodrigues; Cabral, 2012). The causative suffix is less widespread but it has been reconstructed up to the TG branch as a causative of transitive verbs **-ukar*. In short, The TG languages Tapiete, Kamaiurá, and Emérillon, in spite of their geographic separation, share all three of these correspondences.

There furthermore seems to be a shared grammaticalization path that leads to the morphological attachment of verbs in V-V compounds, and eventually leading to a derivational affix. We come back to this issue in the next section.

DISCUSSION

Based on a comparison of constructions found in ten Tupian languages from seven different branches discussed in the previous section, we can set up the following preliminary inventory of Tupian subordination strategies in Table 8. The second column lists the languages that are represented in each group (with their affiliation between brackets), and the last column the average distance between the constructions within each “branch” of Figure 1, measured in the degree of overlap between the answers to the questions in the questionnaire (the smaller the average distance, the more coherent the group).

All constructions are shared between languages from different branches, and also between TG and non-TG languages. Polyfunctional adverbials, multi-verb constructions, participant nominalizations, and affixation are particularly widespread in the oldest branches of the family, as well as in TG, suggesting that they are rather old and persistent in time.

Table 8. A preliminary inventory of Tupian subordination strategies.

Strategy		Languages (affiliation)	# branches	# constr	Avg distance
1	Temporal clauses	Mekens (Tupari), Karo (Ramarama), Gavião (Mondé), Sateré-Mawé (Mawé), Kokama (TG), Kamaiurá (TG)	5	12	0.156
2	Polyfunctional adverbials	Mekens (Tupari), Karitiana (Arikem), Mundurukú (Mundurukú), Tapiete (TG), Kamaiurá (TG), Emérillon (TG)	4	7	0.167
3	Reason/conditional	Mekens (Tupari), Karo (Ramarama), Sateré-Mawé (Mawé), Kokama (TG)	4	7	0.125
4	Multi-verb constructions	Mekens (Tupari), Karo (Ramarama), Mundurukú (Mundurukú), Sateré-Mawé (Mawé), Kokama (TG), Kamaiurá (TG), Emérillon (TG)	4	10	0.165
5	Participant nominalizations	Mekens (Tupari), Karo (Ramarama), Karitiana (Arikem), Gavião (Mondé), Sateré-Mawé (Mawé), Mundurukú (Mundurukú), Tapiete (TG), Kokama (TG), Kamaiurá (TG), Emérillon (TG)	7	13	0.260
6	Polyfunctional complements	Mekens (Tupari), Karitiana (Arikem), Tapiete (TG), Kokama (TG), Kamaiurá (TG)	3	4	0.280
7	Derivational affixes	Mekens (Tupari), Gavião (Mondé), Karitiana (Arikem), Karo (Ramarama), Sateré-Mawé (Mawé), Mundurukú (Mundurukú), Tapiete (TG), Kokama (TG), Kamaiurá (TG), Emérillon (TG)	7	20	0.089

The polyfunctional complements do not seem to be common in non-TG languages, so this might be a TG innovation. Conversely, the dedicated temporal and reason/conditional clauses seem to have been lost in the TG languages of the sample, except in Kokama and Kamaiurá. This can be explained as a reflex of our comparison method, since these EDUs in TG were grouped under the polyfunctional adverbials construction type.

Disregarding participant nominalization and derivational affixation which occur in some way or another in all ten languages of the sample, the subordination inventory in Table 8 does not reflect the proposed historical splits within the Tupian family. There is no corroboration, for instance, for an Eastern versus Western division, in terms of languages spoken outside Rondônia versus languages that are restricted to Rondônia (Rodrigues; Cabral, 2012). No subordination strategy occurred in all and/or only the four languages from Rondônia or vice versa. Rodrigues and Dietrich (1997) propose that the Mawé, Aweti and Tupi-Guaranian branches form a larger unit pointing to an earlier split from the other Tupian languages that only later separated into the current branches (see also Drude; Meira, this volume). Nevertheless, the subordination patterns found in the sample do not align with this historical split, since none of the patterns singles out either the Mawé and TG branches or the TG branch. Except for the two patterns found in all ten languages, no other pattern is present in all four TG languages. This confirms results of other studies on typological features of Tupian (e.g. Eriksen; Galucio, 2014, Birchall, this volume, on person indexing), where a similar mismatch with received internal classifications of Tupian can be observed. This suggests a relatively separate diachronic development for lexicon and grammar at least for Tupian, but this may be a more general pattern - see Danielsen *et al.*, (2011) for the case of the Arawakan family. Future research should make clear whether this separate development of grammar and lexicon is a reflection of the hypothesized difference in stability suggested by Dunn *et al.*, (2005) who claim that abstract grammatical structure is diachronically more stable than lexicon¹⁹.

Some strategies are more convincing as Tupian strategies than others: those strategies that are encountered in most branches, covering most constructions, and showing the greatest internal coherence can more plausibly be argued to be part of the “Tupian inventory” than the ones that score less on these parameters. Following this line of reasoning, the most convincing Tupian construction type is derivational affixation as a complementation strategy. The least convincing subordination strategy are the polyfunctional complement clauses, represented by only four construction types in three different branches, and with the lowest internal coherence (average distance). The participant nominalizations deserve a special mention, because they are represented by 13 constructions, and also found in all branches. Nevertheless, the internal coherence is rather low, with an average distance of 0.260²⁰. This is consistent with the picture sketched in Table 5, with a shared ancient system, but a lot of subsequent diachronic developments through which the subsystems of the different languages have diversified.

Figure 4 is a network displaying distances between constructions only taking into account morphosyntax, and disregarding semantics. The nominalized relatives and the affixation patterns are clearly recognizable on both ends of the network, although the former are much more diffuse. The patterns in the middle are much less clearly differentiated.

¹⁹ An alternative and in fact diametrically opposed possibility is that the grammatical signal, more than the lexical signal, is influenced by language contact. Since the distance matrices do not contain data from likely source languages, this question cannot be answered fully, but a case study on the Guaporé-Mamoré - a proposed linguistic area that includes a good number of Tupian languages (Crevels; van der Voort, 2008) in van Gijn and Hammarström (forthc.) does not suggest clear patterns of diffusion of subordination structures across language families.

²⁰ As can be seen in figure 3, this group of constructions contains four rather obvious outliers (ktn-2, gvo-0, mav-2, and skf-6). Disregarding these reduces the average distance to 0.208 - however, this also reduces the number of constructions represented in this group to 9, and the number of branches to 4.

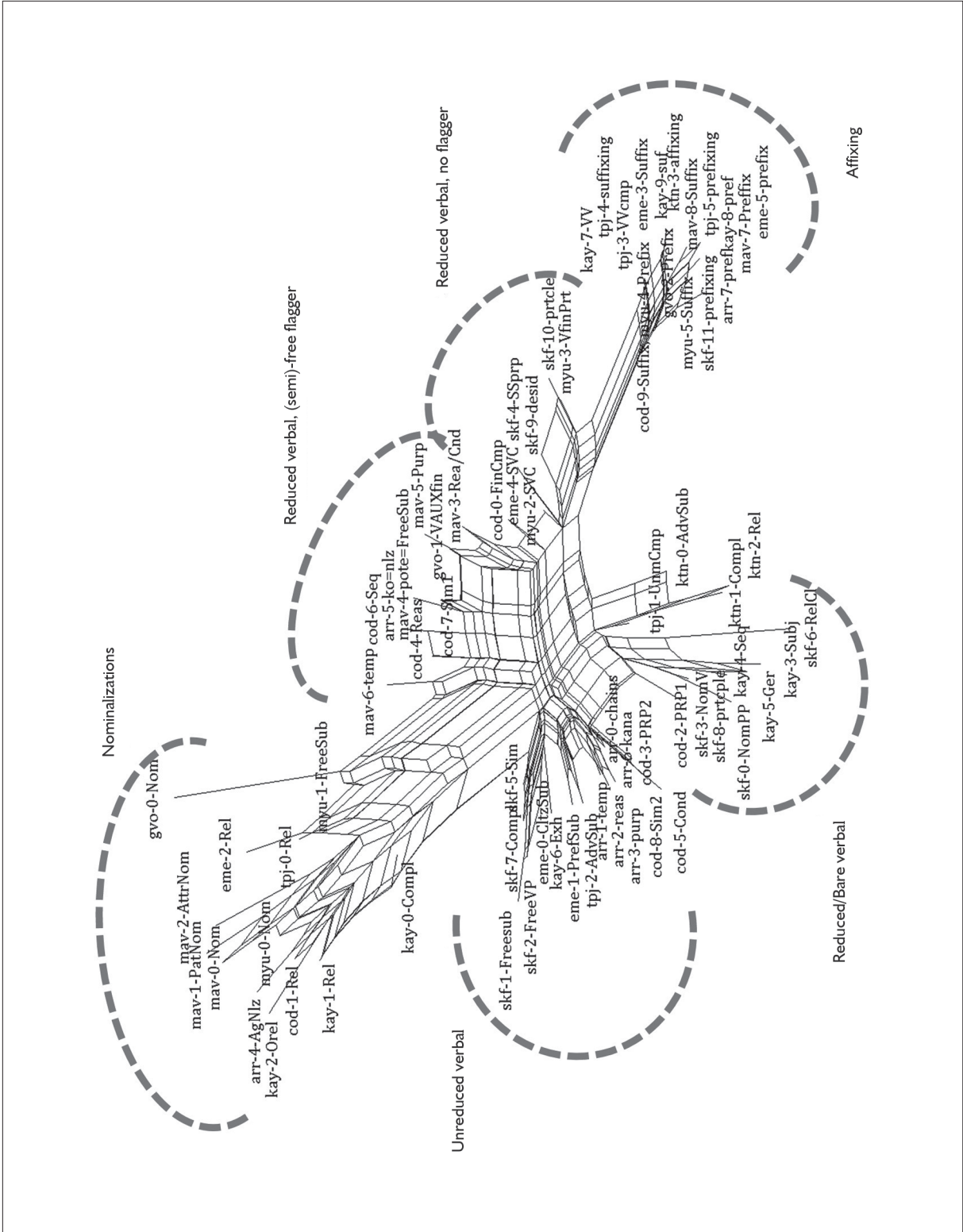


Figure 4. Distance between morphosyntactic profiles of the constructions.

Their relative groupings are mostly based on what we know about the potential to encode verbal categories on the dependent EDU. In the group labeled “unreduced verbal”, to the left of the graph, there is no evidence for the fact that the subordinate EDU’s cannot take verbal categories. This is not to say that there is positive evidence that they can take all verbal categories, just that we have not found evidence to the contrary. For the constructions in the group labeled “reduced/ bare verbal” there is negative evidence for the fact that many verbal categories cannot be realized for the dependent EDU. For most of these constructions, aspect and, for some, also person agreement are the only categories allowed on the subordinate EDU. Some of the constructions also score positively for allowing for nominal categories. The constructions termed “reduced verbal, (semi-)free flagger” usually allow for some verbal categories, restrictions mostly concern modality and evidentiality. In addition, most of the constructions are flagged by a free morpheme, usually following the subordinate clause. The group called “reduced verbal, no flagger” includes constructions that are often restricted in terms of the verbal categories they can take, and in addition they are not marked by any dependency marker. Moreover, they show some characteristics (though in diverse constellations) of syntactically tight constructions - which causes them to appear near the “affixing” group - such as the inability to negate both EDUs independently, obligatory contiguity, fixed order, and the impossibility for both EDUs to be used independently.

All in all, looking at the morphosyntactic picture, there seems to be clear evidence across the family for nominalized structures, especially for relative relations, and affixing, especially for certain complement relations, like direct manipulation, desideratives, and phasals. We might tentatively hypothesize, given the more or less continuous nature of the construction types in Figure 4, and the semantics of the different structures discussed on the basis of figure 3 above, that there is a grammaticalization path for complex constructions from biclausal, flagged structures with a certain degree of verbal potential, through unmarked juxtaposition with more restricted independent verbal potential, to affixes.

This is in line with one of the major grammaticalization paths for subordinate clauses suggested by Heine and Kuteva (2007), which they term “integration” and a further path from complex clauses to compounds and affixation mentioned by e.g. Givón (2009). Heine and Kuteva (Ibid.) mention another development path through which subordinate clauses arise diachronically, termed “expansion”, whereby a nominal referent acquires verbal properties over time. There seems to be less evidence for this channel in Tupi, since the nominalizations are rather cut off from the rest of the constructions. In case of the evolutionary process of expansion, one would expect a more smooth transition. However, see Rose 2013 for a defense of ‘finitization’ as a diachronic process shaping some of the complex sentences of Emérillon²¹.

CONCLUSION

In this exploratory paper, we hope to have given an incentive for the comparative study of subordination strategies in Tupi. We have tried to identify a number of hypotheses about the development of subordination strategies within the Tupian language family. Further, more in-depth, research should confirm or reject these.

- A few strategies seem to be old and persistent in time: the use of participant nominalizations to form relative relations, the use of derivational affixes as a complementation strategy, the use of multi-verb constructions, in particular for purpose-of-motion, and the use of a single construction to encode temporal, conditional, and reason relations.

²¹ Rose (2013) mainly talks about formerly nominal constructions that have become more verbal, which fits the proposed channel by Heine and Kuteva (2007).

- The existence of dedicated temporal clauses with reduced verbal potential may also have been present at the early stages of the family, but seems to have been lost in some of the TG languages.
- There is a general grammaticalization path from biclausal structures, via tight V-V constructions, to affixes. There is less evidence for a path between nominal and verbal constructions.
- Comparable to other case studies on comparative typological features of the Tupian family, the patterns found for this study do not converge on received internal classifications arrived at by the application of the comparative method. This may point towards relatively separate developments for lexicon and grammar.

ABBREVIATIONS

1	first person	FUT	future tense	OBL	oblique
2	second person	GER	gerund	PL	plural
3	third person	I	first series person markers	POT	potential
AC	active	II	second series person markers	PSF	postfactive
ADJZR	adjectivizer	IMPER	impersonal	PST	past
ASSERT	assertive	IN	inactive	PURP	purpose
AUX	auxiliary	INCL	inclusive	REAS	reason
C	coreferent	IND1	indicative 1	REC	reciprocal
CAU	causative	IND2	indicative 2	REI	reiterativ
CES	cessative	INT	interrogative	REL	relational element
CMPL	complementizer	LOC	locative	RES	resultative
CONTR	contrastive	M	male speech	SBM	syntactic boundary marker
DAT	dative	MSP	male speech	SG	singular
DES	desiderative	NEG	negation	SUB	subordinator
DIF	diffusive locative	NFUT	non-future	THV	thematic vowel
DYN	dynamic	NLZ	nominalizer	TR	transitivizer
EVID	evidential	NSJV	non-subjective	TZ	topicalization
EXCL	exclusive	NUC	nuclear case		

REFERENCES

- BICKEL, Balthasar. Capturing particulars and universals in clause linkage. In: BRIL, Isabelle; (Ed.). **Clause Linking and Clause Hierarchy**. Amsterdam: John Benjamins, 2010. p. 51-101.
- BRANDON, Frank R.; SEKI, Lucy. Moving Interrogatives Without An Initial +Wh Node In Tupi. In: COOK, Eung-Do; GERDTS, Donna B. (Ed.). **The Syntax of native American languages**. New York: Academic Press, 1984. p. 72-102. (Syntax and Semantics, n. 16).
- CABRAL, Ana Suelly. **Contact-induced language change in the Western Amazon: the non-genetic origin of the Kokama language**. 1995. 415 f. Ph. D. dissertation - University of Pittsburgh, Pittsburgh, 1995.
- CORRÊA-DA-SILVA, Beatriz. **Mawé/Awetí/Tupí-Guaraní: relações linguísticas e implicações históricas**. 2010. 423 f. Tese (Ph. D. dissertation) - Universidade de Brasília, 2010.
- CREVELS, Mily; VOORT, Hein van der. The Guaporé-Mamoré region as a linguistic area. In: MUYSKEN, Pieter (ed.) **From Linguistic Areas to Areal Linguistics**. Amsterdam/Philadelphia: John Benjamins, 2008.

CRISTOFARO, Sonia. **Subordination**. Oxford: Oxford University Press, 2003.

CROFTS, Marjorie. **Aspectos da língua Mundurukú**. Cuiabá: Sociedade Internacional de Linguística, 2004.

DANIELSEN, Swintha; DUNN, Michael; MUYSKEN, Pieter. The spread of the Arawakan languages: a view from structural phylogenetics. In: HORNBERG, A.; HILL, J. D. (Ed.). **Ethnicity in ancient Amazonia: reconstructing past identities from archaeology, linguistics, and ethnohistory**. Boulder: University Press of Colorado, 2011. p. 173-196.

DRUDE, Sebastian. On the position of the Awetí language in the Tupí family. In: DIETRICH, Wolf (Ed.). **Guaraní y 'Mawetí-Tupí-Guaraní': estudios históricos y descriptivos sobre una familia lingüística de América del Sur**. Berlin: LIT, 2006. p. 11-46.

DUNN, Michael *et al.* Structural phylogenetics and the reconstruction of ancient language history. **Science**, v. 309, n. 5743, 2005. p. 2072-2075.

ERIKSEN, Love; GALUCIO, Ana Vilacy. The Tupi expansion. In: O'CONNOR, Loretta; MUYSKEN, Pieter (Ed.). **The native languages of South America: origins, development, typology**. Cambridge: Cambridge University, 2014. p. 177-199.

GABAS JR., Nilson. **A Grammar of Karo, Tupi (Brazil)**. 1999. Tese (Ph. D. dissertation) - University of California, Santa Barbara, 1999.

GALUCIO, Ana Vilacy. **The Morphosyntax of Mekens (Tupi)**. 2001. Tese (Ph. D. dissertation) - University of Chicago, Chicago, 2001.

GALUCIO, Ana Vilacy; GABAS Jr., Nilson. Evidências de agrupamento genético Karo-Puruborá, tronco Tupi. 2002. Paper presented at 17^o Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística, Gramado. 2002.

GALUCIO, Ana Vilacy. Subordinate adverbial constructions in Mekens. In: GIJN, Rik van; HAUDE, Katharina; MUYSKEN, Pieter (Ed.). **Subordination in native South American languages**. Amsterdam: John Benjamins, 2011. p. 25-44.

GALUCIO, Ana Vilacy; NOGUEIRA, Fernanda. Comparative study of the Tupari branch of the Tupi family: contributions to understanding its historical development and internal classification. In: CONFERENCE ON INDIGENOUS LANGUAGES OF LATIN AMERICA, 5., 2011, Austin. **Proceedings of the...** Austin: Center for Indigenous Languages of Latin America (CILLA). 2012. Disponível em: <<http://www.ailla.utexas.org/site/events.html>>. Acesso em: 15/10/2013.

GIJN, Rik van. Subordination strategies in South American languages: nominalization. In: O'CONNOR, Loretta; MUYSKEN, Pieter (Ed.). **The native languages of South America: origins, development, typology**. Cambridge: Cambridge University, 2014. p. 274-296.

GIJN, Rik van; HAMMARSTRÖM, Harald. Forthc. A construction-based approach to measuring distances between languages: subordination strategies in South America. Forthcoming.

GIJN, Rik van; HAUDE Katharina; MUYSKEN Pieter. Subordination in South America: an overview. In: GIJN, R. van.; HAUDE, K.; MUYSKEN, P. (Ed.) **Subordination in native South American languages**. Amsterdam/Philadelphia: John Benjamins, 2011. p. 1-24.

GIVÓN, Talmy. **The genesis of syntactic complexity**. Amsterdam/Philadelphia: John Benjamins. 2009.

GOLDBERG, Adele E. **Constructions at work: the nature of generalization in language**. Oxford: Oxford University, 2006.

GONZÁLEZ, Hebe. **A grammar of Tapiete (Tupi-Guaraní)**. 2005. Ph. D. dissertation - University of Pittsburgh, Pittsburgh, 2005.

HEINE, Bernd; KUTEVA, Tania. **The genesis of grammar: a reconstruction**. Oxford: Oxford University, 2007.

HENGVELD, Kees; Mackenzie, J. Lachlan. **Functional Discourse Grammar**. Oxford: Oxford University Press, 2008.

JENSEN, Cheryl. Cross-referencing changes in some Tupí-Guaraní languages. In: PAYNE, Doris L (Ed.). **Amazonian Linguistics: studies in Lowland South American Languages**. Austin: University of Texas, p. 117-158, 1990.

JENSEN, Cheryl. Comparative Tupí-Guaraní morphosyntax. In: DERBYSHIRE, Desmond C.; PULLUM, Geoffrey K. (ed.). **Handbook of Amazonian languages**. Berlin, New York: Mouton de Gruyter, 1998. p. 489-618.4 v.

JENSEN, Cheryl. Tupí-Guaraní. In: DIXON, R.M.W.; AIKHENVALD, Alexandra Y. (Ed.). **The Amazonian languages**. Cambridge: Cambridge University, 1999. p. 125-164.

LEHMANN, Christian. Towards a typology of clause linkage. In: HAIMAN, J.; THOMPSON, S.A. (Ed.). **Clause combining in grammar and discourse**. Amsterdam, Philadelphia: John Benjamins, 1988. p. 181-225.

LOUKOTKA, Čestmír. **Classification of South American Indian languages**. Los Angeles: University of California, 1968.

MALCHUKOV, Andrej. Constraining nominalization: form/function competition. **Linguistics**, v. 44 n. 5, p. 973-1009, 2006.

McMAHON, April; McMAHON, Robert. **Language classification by numbers**. Oxford: Oxford University Press. 2005.

MELLO, Antonio. **Estudo histórico da família linguística Tupí-Guaraní: aspectos fonológicos e lexicais**. Ph. D. dissertation. Florianópolis: Universidade Federal de Santa Catarina, 2000.

MÉTRAUX, A. **La civilisation matérielle des tribus Tupí-Guaraní**. Paris: Paul Geuthner. 1928.

MOORE, Denny. **Syntax of the language of the Gavião Indians of Rondônia (Brazil)**. 1984. 200 f. Ph. D. dissertation - University of New York, Nova York, 1984.

MOORE, Denny. A Few Aspects of Comparative Tupi Syntax. **Revista Latinoamericana de Estudios Etnolingüísticos**, Lima, v. 8, 1994.

MOORE, Denny. Classificação interna da família linguística Mondé. **Estudos Lingüísticos**, v. 36, p. 515-520, 2005.

MOORE, Denny; GALUCIO, Ana Vilacy; GABAS JR., Nilson. O desafio de documentar e preservar as línguas amazônicas. **Scientific American Brasil**, Edição especial Amazônia - Destinos, n. 3, p. 36-43, 2008.

RODRIGUES, Aryon D. Classification of Tupí-Guaraní. **International Journal of American Linguistics**, v. 24, p. 231-234, 1958.

RODRIGUES, Aryon D. A classificação do tronco linguístico Tupí. **Antropologia**, v.12, p. 99-104, 1964.

RODRIGUES, Aryon D. Relações internas na família linguística Tupí-Guaraní. **Antropologia**, v. 27 e 28, 33-53, 1984/85.

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

RODRIGUES, Aryon D.; CABRAL, Ana Suelly. Revendo a classificação interna da família Tupí-Guaraní. **Línguas indígenas brasileiras: fonologia, gramática, história**. Belém: EDUFPA, 2002. p. 327-337. 1 v.

RODRIGUES, Aryon D.; CABRAL, Ana Suelly. Sobre o desenvolvimento de padrões absolutivos em famílias orientais do tronco Tupi. In: QUEIXALÓS, F. (Ed.). **Ergatividade na Amazônia II**. Brasília: [s. n.], 2003.

RODRIGUES, Aryon D.; CABRAL, Ana Suelly; CORRÊA DA SILVA, Beatriz. Evidências linguísticas para a reconstrução de um nominalizador de objeto ****mi-** em Proto-Tupí. **Estudos da Língua**, v. 4. n. 2, p. 21-39, 2006.

RODRIGUES, Aryon D.; CABRAL, Ana Suelly. Tupían. In: CAMPBELL, Lyle; GRONDONA, Verónica (eds.). **The Languages of South America: a comprehensive guide**. Berlin/Boston: Mouton de Gruyter, 2012. p. 495-574.

ROSE, Françoise. Grammaire de l'émérillon teko, une langue tupi-guarani de Guyane française. **Langues et Sociétés d'Amérique traditionnelle**, Louvain, v. 10, 2011.

ROSE, Françoise. Finitization: a shift of dependency-coding strategy from Proto-Tupí-Guaraní to Emérillon. **Diachronica**, v. 30, n. 1, p. 27-60, 2013.

SCHLEICHER, Charles. **Comparative and Internal Reconstruction of the Tupi-Guarani Language Family**. 1998. 372 f. Ph. D. dissertation - University of Wisconsin-Madison, Madison, 1998.

SCHMIDTKE-BODE, Karsten. **A Typology of Purpose Clauses**. Amsterdam/Philadelphia: John Benjamins, 2009.

SEKI, Lucy. Estratégias de Relativização em Kamaiura. In: VOORT, Hein van der; KERKE, Simon van de. (Ed.). **Indigenous Languages of Lowland South America**. Leiden: Centre for Non-Western Studies, 2000a. p. 309-324.

SEKI, Lucy. **Gramática do Kamaiurá, língua Tupí-Guaraní do Alto Xingú**. Campinas:Unicamp, 2000b.

SILVA, Raynice da. **Estudo morfossintático da língua Sateré-Mawé**. Ph. D. dissertation. Campinas: Universidade Estadual de Campinas, 2010.

STORTO, Luciana. Subordination in Karitiana. **Amérindia** (Paris), v. 35, p. 219-237, 2012.

STORTO, Luciana. Constituent order and information structure in Karitiana. In: Rik van GIJN, Jeremy HAMMOND, Dejan MATIC, Saskia Van PUTTEN, Ana Vilacy GALUCIO. (Eds.). **Information Structure and Reference Tracking in Complex Sentences**. 1ed. Amsterdam/Philadelphia: John Benjamins, 2014, v. 1, p. 163-192.

STUTE, Horst. Os auxiliares dinâmicos da língua Gavião. In: **Porto Velho Work Papers**. Porto Velho: Summer Institute of Linguistics, 1985. Disponível em: <<http://www.sil.org/americas/brasil/SILapub.html>> . Acesso em: 03 Abr. 2012.

VALLEJOS, Rosa. **A grammar of Kokama-Kokamilla**. 918 f. Ph. D. dissertation - University of Oregon, Oregon, 2010.

VAN VALIN, Robert D. Jr.; LAPOLLA, Randy J. **Syntax**. Cambridge: Cambridge University Press, 1997.