Argument Marking Patterns in

South American Languages

Argument Marking Patterns in South American Languages

een wetenschappelijke proeve op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus
prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
vrijdag 2 mei 2014
klokke 10.30 uur

door

Joshua Thomas Rigo Birchall

geboren 20 februari 1985
te Rockford, Verenigde Staten

Contents

Acknowledgements

Nobody writes a dissertation alone. While my name may be the one on the cover, this work is the product of many things: an idea, a place in time, a place in space, and most importantly, a group of outstanding people have been there for me along the way to offer their thoughts, help and encouragement. I'd like to take these opening pages to thank them and recognize all of their hard work.

First I'd like to thank my promotor, Pieter Muysken. He has been a constant source of ideas and support throughout these last four years. The topic of the thesis itself, argument marking in South American languages, was his own creation. He planted a seed in my mind—an idea—and helped it grow. He watered the seed by sharing the funds from his KNAW professorship and ERC grant to ensure that I had the time and resources to do the job right. He pruned my ideas as they grew, always with a soft touch, making sure that I wasn't just producing an intellectual thicket. If it weren't for Pieter, this project would have never happened.

Many thanks are also owed to my co-promotors Mily Crevels and Hein van der Voort. They were always willing to share their wealth of experience and insight with me when I needed it most. Hein was there for me from the moment I arrived in the Netherlands, he helped me keep my thoughts straight and made sure I kept going down the right track. Mily really came through for me when it counted. She tirelessly read through multiple versions of all of my chapters and it was her keen attention to detail that helped my ideas stay afloat. She always seemed to know what I needed, whether it was a little shove forward, some reining in, or just a friendly chat over some coffee.

This research is a testament to the time it was written in. Such an ambitous project would not have been possible just a decade ago. I am so grateful for the careful and difficult work put in by all of the field linguists who produced the grammars that I used in this study. Without them, and the patient and intelligent indigenous teachers they have collaborated with, we'd still be in the dark ages of language research in South America. Many have surely endured hardship in the name of science, leaving behind their families for extended

periods, battling exotic illnesses and traveling far outside of their comfort zones. I'd especially like to thank the scholars who took the time to discuss their languages of study with me: Laércio Bacelar, Ana Paula Brandão, Matt Coler, Aline da Cruz, Swintha Danielsen, Connie Dickinson, Sebastian Drude, Pattie Epps, Sidi Facundes, Helder Ferreira, Paul Frank, Vilacy Galucio, Rik van Gijn, Lucia Golluscio, Hebe Gonzales, Antoine Guillaume, Katharina Haude, Simon van de Kerke, Suzi Lima, Sérgio Meira, Denny Moore, Simon Overall, Gessiane Picanço, Françoise Rose, Léia Silva, Wilson Silva, Luciana Storto, Stella Telles and Rosa Vallejos. I am surely forgetting a few, and the fact that this list is so long just shows how great a group of linguists we have working in South America today.

This thesis is also a product of the place where it was written. My gratitude goes out to my colleagues working on the continent who have helped along the way. I thank Michael Dunn, Love Eriksen, Simeon Floyd, Harald Hammarström, Martin Haspelmath, Helen de Hoop, Olaf Koeneman, Loretta O'Connor, Leon Stassen and Robert Van Valin for their thoughtful discussions and input. I also thank Michele Gubian for help with R and Christina Bergmann for showing me the ropes with LaTeX. Tina Gregor and Simon van de Kerke both contributed immensely to the project by helping with some of the data coding and pointing out some of the thornier issues with my questionnaire. My colleagues in the Languages in Contact research group at Radboud also helped shape my thinking throughout these last years. I especially thank Olga Krasnoukhova and Neele Müller for helping to pave the way in the South America project. While we all have diverse interests, it was really useful to be able to work together with the likes of Susanne Aalberse, Margot van den Berg, Bob Borges, Pablo Irizarri, Gerrit Jan Koostra, Francesca Moro, Cefas van Rossem, Robbert van Sluis and Kofi Yakpo. Nijmegen wouldn't have been the same without them and the rest of the crew at Radboud. The lunches, meetings, coffee breaks, and countless discussions on linguistics and Dutch living helped me to feel part of a larger community.

I am also indebted to the work that my colleagues have put into reading this thesis during its preparation and evaluation. Spike Gildea deserves special mention for really engaging with me on the material after the manuscript committee. I also whole heartedly thank Sérgio Meira for the careful and detailed comments he gave me on the manuscript. Glenn Shepard also deserves mention here, helping to add an anthropologist's flair to the introduction and providing the beautiful cover photo.

And of course, none of this could ever have happened without the love and support of my family and friends. My dad Steve was always there for me when I needed him, always reminding me that I could do it. It it wasn't for him and my sister Melissa, this journey would have been much more difficult. I'd also like to thank my friends and colleagues in the Netherlands and Brazil who made the last four years so much more enjoyable.

Finally, I'd like to thank you, the reader, for being interested in South American languages. I hope you find them as beautiful and exciting as I do.

Abbreviations

The abbreviations and conventions used in the examples cited in this thesis have been adopted from the Leipzig Glossing Rules whenever possible.[1] The abbreviations used that do not conform to those suggested in the Leipzig Glossing Rules are listed below.

ABSTR	abstract
AGT	agentive
AN	animate
ASR	assertative
ATR	attributive
AUG	augmentative
CERT	certainty
COMP	comparative
CONJ	conjunction
CONT	continuous
CSO	cosubordination
DESID	desiderative
DFR	different referent
DIM	diminutive
DIST	distal
DR	direct
DST.CNTR	distant past contrast
DYNM	dynamic
EMPH	emphatic
EV	eventive
EVI	evidential
FAC	factive
FAM	familiar
FILL	filler particle

[1] http://www.eva.mpg.de/lingua/resources/glossing-rules.php

GCM	general class marker
HAB	habitual
HIAF	highly affected aktionsart
ID	indirect
IDR.CAUS	indirect causation
INAN	inanimate
INESS	inessive
INT	intentional
INTEN	intensifier
INTR	intransitivizer
ITR	iterative
LIM	limitive
LV	linking vowel
MAL	malefactive
MOM	momentaneous
MSC	masculine bound noun
NF	non-feminine
NFNT	non-finite
NPF	noun prefix
NPR	non-promotional
NSG	non-singular
PAUC	paucal
PERT	pertensive
PL.DIS	plural discrete
PLT	politeness marker
PONCT	punctual
PRD	predicative marker
PROS	prospective
PS	predicate specifier
QUAL	quality
QUOT/FOLK	quotative folklore
R	relational
RCS	recessive
REST	restrictive
RPT	reportative
RR	realis
SCM	specific class marker
SIT	situational
SMBL	semblative
SUB	subordinate
VCL	verbal class marker

CHAPTER 1

Introduction

The South American continent offers many exciting opportunities for scholars concerned with the study of human language. Through a boom in the description of the indigenous languages of South America that started at the end of the last century, two crucial aspects of these languages are slowly becoming apparent in current academic research. First, South American languages are incredibly diverse, both in terms of the different linguistic structures that they display and the sheer number of different genealogical units into which they can be demonstrably classified. Yet in the face of this incredible typological and phylogenetic diversity, a second crucial aspect of South American languages emerges: there are a large number of recurrent phonological and grammatical patterns found in South America that transcend attested language families and whose distributions are too robust to be attributed to mere chance. Finding a plausible explanation for the recurrent distribution of a number of linguistic structures amidst overarching typological and phylogenetic diversity is what can be called the *South American puzzle*.

If chance and transmission through descent within known genealogical groupings can both be ruled out as explanations for the South American puzzle, a few additional explanations still must be accounted for. Perhaps the languages are related on a much deeper time scale than the traditional comparative method is able to attain due to the nature of the method itself (Kaufman, 1990). Another possible explanation is that some of these similarities can be attributed to structural dependencies between linguistic features, as explored in the pioneering work of Greenberg (1963). These dependencies could reflect cross-linguistic tendencies and universal biases that human cognition places on language due to functional principles such as learnability and communicative efficiency (Evans

and Levinson, 2009), or perhaps these dependencies even arise from constraints built into a hypothetical innate language faculty of the mind (Baker, 1996, 2001). A final explanation could be that some of these shared linguistic features are a result of structural convergence through extended periods of interaction among different ethnolinguistic groups, resulting in a number of sociolinguistic processes often subsumed under the label of language contact (Thomason and Kaufman, 1988; Aikhenvald, 2002; Winford, 2005, among many others).

Together with the continued documentation and description of the native languages of South America, a more urgent task must first be carried out before one can even begin to explore the South American puzzle and search out an explanation for these recurring patterns. This task is the development of a precise and transparent method for the cataloging these linguistic structures and its subsequent application to a representative sample of languages from this continent that can serve as an empirical basis for any such explanations. It is with this task that the present dissertation is concerned, focusing on an area of language that is both modest in scope yet fundamental to human communication: the ways that languages encode the participants of a linguistic utterance, or in other words, the way that a language expresses who did what to whom.

1.1   Objectives and motivation

This is the first comprehensive cross-linguistic study of argument marking patterns in South American languages. It looks specifically at how the relations between a predicate and its semantically obligatory participants—its arguments—are expressed in basic main clause constructions. Languages often treat different semantic types of arguments identically in one or more grammatical processes, forming syntactic argument roles based on the valency of the predicate. Argument roles can further be treated identically across predicate classes of different valencies, forming grammatical relations such as subject, direct object, indirect object, absolutive and ergative. These grammatical relations are central to most linguistic theories and feature prominently in the descriptions of most languages. The argument roles under consideration here are those formed by the overt morphological marking of clausal participants through indexation and case marking.

Argument marking has been chosen as the topic of this thesis since every (or almost every) language has a grammatical system to express argument relations, thus ensuring that relevant data can be gathered for all of the languages under consideration. Even if the strategies used in this system vary greatly from language to language, the application of an explicit procedure for their analysis and comparison has the ability to identify a number of commonalities and major points of typological variation. The universality of argument relations allows for the comparability of the data with other linguistic domains such as the lexicon and phonology, as well as other core grammatical systems such as noun phrase structure and the expression of temporal, aspectual and

modal relations. In fact, the project in which the current research has taken place has also focused on these latter two grammatical domains in addition to argument marking in order to give a holistic view of the structural profile of South American languages. Some preliminary findings from this project are presented in O'Connor and Muysken (2014), with additional results presented in two doctoral dissertations (Krasnoukhova, 2012; Mueller, 2013).

This study was not designed as an explicit typological investigation of argument marking, in the sense that it makes no attempt to map out all of the possible variation within the grammatical domains under consideration. Rather, it is a comparative study of language structure within a typologically-informed framework, restricted to a sample of South American languages and concerned with the morphosyntactic treatment of certain argument roles by a few specific morphosyntactic processes within a subset of all possible construction types. Within this restricted domain, the study has the following objectives:

- Outline a coherent approach for the cross-linguistic investigation of argument marking, primarily restricted to the morphosyntactic encoding of arguments in basic main clauses.

- Apply this approach to a sample of 74 South American languages, capturing the major typological distinctions observed into a structural questionnaire.

- Analyze the occurrence of specific argument marking features in terms of their geographical distributions.

The research presented in this thesis contributes to the comparative study of South American languages by not only providing an analysis of the geographic distribution of certain linguistic features across the continent, but by presenting the empirical basis of this study in a coherent and transparent manner. Summaries of portions of the dataset are provided at the end of each relevant chapter, with the full dataset given in Appendix C. The data can be downloaded in a machine-readable format from the project website, where it is also possible to access additional information on the dataset such as exact references, representative examples and further notes on coding.[1]

As the title of this thesis suggests, the research presented here is not only concerned with the ways that arguments are marked in the languages under consideration, but how these marking strategies pattern together. Patterns can be observed on a number of different levels. First, we can observe patterns of similarity and difference in the treatment of various argument roles by a specific argument marker, notably the neutralization of certain roles in specific contexts leading to ergativity, accusativity, etc...; these patterns are referred to here as ALIGNMENT PATTERNS. We can also observe the ways that different argument markers within a single grammatical domain like case marking or indexation treat different argument roles, notably how multiple markers combine

---

[1]See www.ru.nl/linc for more information.

together, their position relative to the marked constituent, and the interplay between argument marking on the predicate and the realization of nominal and pronominal arguments; these patterns are referred to here as MARKING PATTERNS. Finally, GEOGRAPHIC PATTERNS can be observed in the way that different structural features are distributed across the South American continent.

The recurrent geographic distribution of certain grammatical features across apparent genealogical boundaries has attracted the attention of linguists for more than a century. Lafone Quevedo (1896) is an early attempt at classifying the languages of South America according to the position of argument marking on the verb. Modern comparative research on South American languages has tended to shift away from grouping languages according to one major structural feature, and has instead paid attention to a wider variety of features. A recurrent theme in many of these studies is the identification of characteristic linguistic features of a particular region of the continent, such as Amazonia or the Andes (cf. Büttner, 1983; Derbyshire, 1987; Payne, 1990; Dixon and Aikhenvald, 2000b; Adelaar, 2008, among others). These studies have greatly increased our understanding of South American languages and have helped to highlight the incredible diversity of linguistic structures found across the continent. However, these studies have tended not to present the sample of languages used or much of the data from which their generalizations are drawn. In the few cases where an explicit language sample was given, as in Derbyshire (1987), languages from outside of the specific region under consideration were not included. It is difficult to make any meaningful claims about the geographic distribution of a linguistic feature within a particular region if that distribution cannot be contrasted with the distribution of the same feature outside of that particular region (cf. Sherzer, 1976; Campbell et al., 1986; Enfield, 2005; Bickel and Nichols, 2006). For this reason, the current study includes languages from all major regions of the continent.

This study is not meant to be the last word in the centuries long discussion on the typology of South American languages. Instead, it outlines a new way to approach the problem and applies it to a specific area of grammar. The approach adopted here holds that large-scale inquiries such as the comparative study of language structures across an entire continent are best explored through the use of a questionnaire that is applied to a representative sample of languages using an explicit coding procedure. The data are coded in such a way that they are accessible to other scholars who would like to use them for their own purposes, regardless of their research goals or theoretical backgrounds. It is hoped that future research on South American languages continues along the path of producing transparent and readily available datasets that can build off the cooperation of different researchers and be supplemented with additional data as it becomes available.

Before digging into the details of the study, the following section places this research within the broader linguistic, cultural and historical context of the South American continent. Section 1.3 then discusses the sample used in this

study. Section 1.4 outlines the structure of the remainder of this thesis.

1.2 The South American context

South America was the last of the continents to be settled by humans. Conservative archaeological estimates suggest an initial colonization around 11,000 years before present, in line with the Clovis First hypothesis. More recent estimates point towards an arrival date between 12,500 and 15,000 years before present, around the time of the retreat of the Pleistocene glaciers (Dillehay, 2008). However, there is at least one controversial archaeological site, Pedra Furada located in eastern Brazil, that suggests an arrival date as early as 32,000 years before present (Guidon and Delibrias, 1986). The sheer number of different linguistic families in North and South America has led some linguists, most notably Nichols (1990), to argue in favor of the earlier chronology for the settling of the New World, since the linguistic diversity of the continents in many ways mirrors that in areas such as Papua New Guinea and Australia, which have been colonized by humans for over 40,000 years.

Regardless of the initial date of entry, it was not until about 8,000 years ago that humans began to domesticate plant and animal species, resulting in a gradual shift from a primarily foraging economy to an agricultural economy in parts of the highlands and lowlands by around 3,500 years ago (Piperno and Pearsall, 1998; Pearsall, 2008). This shift towards sedentism, with corresponding technological advances in ceramics and crop domestication, initiated the formative periods in portions of the highlands (Initial Period/Early Horizon) and the lowlands (Browman, 2001; Arroyo-Kalin, 2010). The formative periods are marked with increased population densities in certain regions and coincide with many of the major language expansions on the continent (Heckenberger and Neves, 2009). The late formative periods led to greater regional integration and sociopolitical complexity across ethnic lines, especially in areas such as the southern Amazonian fringe, the Peruvian highlands and the lower Amazonian floodplains (Hornborg, 2005; Neves, 2006, 2008), as well as an increase in landscape modification through earthwork projects such as roads, irrigation and terracing (Denevan, 2001). These developments set the stage for the South America that was encountered by Europeans at the beginning of the 16th century.

1.2.1 Phylogenetic diversity and linguistic classification

The South American continent hosts an incredible diversity of different language families, ranging from the large Arawakan and Tupian families, with approximately 60 members each, to a plethora of small language families and isolates that cannot be reliably classified as belonging to any higher-order phylogenetic grouping. Campbell (2012a) provides an estimate of 108 different languages families in South America, of which 55 languages are considered isolates,

i.e. families that include only a single member. This accounts for approximately
a quarter of all attested language families in the entire world. Furthermore, it
is worth noting that a number of the families accepted by Campbell have not
been conclusively proven but are tentatively classified as such based on sugges-
tive evidence, such as the hypothetical Pano-Tacanan family composed of the
Panoan and Tacanan languages (Key, 1968; Girard, 1971; Kaufman, 1990).

   Attempts to link together the established language families into 'macro-
families' based on a mass comparison of a few lexical items, as in the work of
Greenberg (1987), has been met with great skepticism by traditional historical
linguists (cf. Campbell, 1991). Even the lower-level macro-families of Green-
berg, such as 'Macro-Carib', which groups languages of the Cariban, Witotoan
and Peba-Yaguan families together with a few isolates, have not held up under
close scrutiny by specialists in these languages (cf. Gildea and Payne, 2007).
Another long standing hypothesis put forth in Rodrigues (1985a) on a relation-
ship between the Tupian and Cariban languages, with the possible inclusion of
the Macro-Jêan family, has yet to be fully accepted among South Americanists.
In the highlands, noticeable similarities between the Quechuan and Aymaran
languages have resulted in the proposal of a hypothetical Quechumaran family
by Orr and Longacre (1968) and Büttner (1983), among others, but a number
of researchers have argued that these similarities are more appropriately at-
tributed to long term language contact rather than genealogical relatedness, at
least within a time depth recoverable by the traditional historical comparative
method (Torero, 2002; Cerrón-Palomino, 2008; Adelaar and Muysken, 2004;
Adelaar, 2012). However, that is not to say that our understanding of related-
ness among South American language families has not improved over the years.
For instance, recent comparative work by Ribeiro (2010) has been able to show
that the Jabutian languages, once thought to be a separate language family,
are indeed a branch of the Macro-Jêan family.[2]

   The internal classifications of many of the larger families of the continent
have been strengthened through recent comparative work on the intermediate
clades within these families, such as the Mawetí-Guaraní branch of the Tu-
pian family (Rodrigues and Dietrich, 1997; Corrêa-da Silva, 2010; Drude and
Meira, 2015), the Apurinã-Piro-Iñapari branch of the Arawakan family (Fa-
cundes, 2002), the Southern Jê branch of the Macro-Jêan family (Jolkesky,
2010) and the Taranoan branch of the Cariban family (Meira, 2000a). How-
ever, much work still remains in both the documentation and comparison of
the extant and historical languages in South America, and it is expected that
these classifications will be refined as further comparative work is carried out.

   The classifications of the languages used in this thesis are based on those
given in Campbell (2012a) and Lewis (2009). When there are discrepancies

---

[2]The Macro-Jêan and Tupian languages are often referred to as 'stocks' in the South
American literature. In this thesis, the highest node in a phylogenetic tree that can be
demonstrably proven is referred to as a 'family', with intermediate nodes referred to as
'branches' of that family. As such, the term 'stock' would only apply to unproven genealogical
relations that play no role in the current study.

between the two classifications, the more conservative one has been selected.[3]

1.2.2 Major regions

This section presents an overview of the major geographic regions of South America as distinguished in this study. In the linguistic study of South American languages, it is common to make a distinction between the languages of the lowlands and those of the highlands (e.g. Payne, 1990), or those of the Amazon basin and those of the Andes (e.g. Dixon and Aikhenvald, 1999), with an often ambiguous or intermediate status attributed to the languages of the eastern foothills of the Andes. This study takes a different approach by dividing the continent into seven distinct regions. While the regions may display a certain linguistic or cultural homogeneity, the boundaries of such regions are delineated using explicitly geographic criteria. The languages of the eastern foothills of the Andes are classified as Amazonian if their territory lies primarily on an Amazonian tributary, or as Andean if their territory is primarily in the highlands.

Beyond being delimited by physical obstacles that make natural impediments to the movement of peoples, such as mountain ranges and major rivers, the regions tend to roughly correspond to or include previously defined cultural and/or linguistic areas. Such definitions are used for the sake of comparability with other claims about the languages of the continent, as well as to inform any hypothesis that must account for the likely locus of interaction between different linguistic populations. The use of culture areas in determining the major regions of South America are important to the study of language contact since culture areas are delimited through shared cultural traits that diffuse across populations in social networks that foster multi-ethnic interaction in ways that can also result in the spread of linguistic features (Sherzer, 1973). The culture areas presented in Murdock (1951) are used to guide the configuration of the different regions of the continent.[4]

Each region is discussed in the following paragraphs in terms of its geography and ethnolinguistic composition. These regions are further used in Chapter 8 for the assessment of the geographic distribution of certain argument marking features.[5] A map of South America with the major regions is found in

---

[3]The one exception to this rule is the acceptance of the Macro-Jêan family, which Campbell (2012a) and Lewis (2009) do not consider sufficiently attested but Ribeiro (2010) argue in favor of based on a substantial body of evidence.

[4]According to Murdock (1951, 416) these culture areas are based on techniques for securing food; the presence of specific agricultural cultivars and domesticated animals; sex-based social roles in food collection; material cultural traditions such as pottery, weaving and metallurgy; the size and style of dwellings; degrees of social and political stratification; marriage practices; and kinship categorizations. Murdock also includes linguistic affiliation as one of the criteria for evaluating a culture area, but it is clear in his descriptions that each area is by no means linguistically homogeneous nor based primarily on the linguistic classifications of its members.

[5]This scheme for the division of South America into seven distinct regions is also used in Birchall (2014b). This section adopts the same geographic criteria for the regions and provides additional information not presented in the original article.

Appendix A.

Northern Andes: This region consists of the Andean highlands and foothills
stretching from the Isthmus of Panama along the Cordillera Real of Ecuador.
It extends west towards the Pacific Ocean, with the eastern boundary of the
region formed by the headwaters of the Orinoco River.

   The Northern Andes shows considerable linguistic diversity with languages
from the Chibchan, Chocoan, and Barbacoan families, as well as a number
of isolates such as Nasa Yuwe, Kamsá and Andoke. The Quechuan language
family also reached the Northern Andes by means of the Inca Empire around
the same time as the Spanish arrival in the New World (Adelaar and Muysken,
2004, 53, 165-167).

   Archaeologically, the Northern Andes is sometimes associated with the cul-
tures of southern Central America. In this context, it is often referred to as
part of the 'Intermediate Area' since it sits between the complex state societies
of the Central Andes and the Mayan and other civilizations of the Yucatan
peninsula and southern Mexico at the time of the arrival of the Spanish (Wil-
ley, 1971). This term has been adopted by some linguists, such as Torero (2002)
and Constenla Umaña (1991), but more recent research has tended towards the
use of the term 'Isthmo-Colombian area' to highlight its own distinct culture
and history (cf. O'Connor, 2014). Constenla Umaña (1991) identifies a number
of shared linguistic features of this region. This region comprises the Colom-
bian culture area in Murdock (1951), together with the South American portion
of Isthmian area and the western portion of the Caribbean area, particularly
through the inclusion of the Chibchan languages of the Santa Marta mountains.

Northern Amazonia: This region comprises the tropical forests, plateaus
and coastal areas stretching from the lower portions of Amazon River north to
the Caribbean Sea. The eastern banks of the Rio Negro and its tributaries forms
the southwestern boundary of this region. This region is primarily composed of
the Guiana culture area in Murdock (1951), together with the smaller Savanna
and Orinoco areas and the eastern portions of the Caribbean area.

   Northern Amazonia hosts a number of different languages, including mem-
bers from the Cariban, Arawakan, Yanomaman and Salivan language families,
as well as the isolate Warao. Tupian languages of the Tupí-Guaraní branch
migrated into Northern Amazonia around the time of the arrival of the Por-
tuguese on the coast of Brazil (Rodrigues and Cabral, 2012, 501). The northern
portions of this region also host a number of creole languages with linguistic
elements from a variety of European, African and Amazonian languages. Both
Migliazza (1985) and Constenla Umaña (1991) identify a number of shared
linguistic features for (parts of) this region.

Central Andes: This region stretches from the Peruvian highlands of the
Cordillera Central south until the Atacama Desert. The Cordillera Occidental

of Bolivia and Argentina forms the easternmost extent of this region.

The Central Andes is best known for the expansionist Quechuan and Aymaran language families, but it also hosts a number of smaller language families such as Uru-Chipayan and Hibito-Cholonan, as well as the isolate Leko. Many scholars have proposed a number of different linguistic features characteristic of the languages of the Central Andes. The proposals found in Dixon and Aikhenvald (1999) and Adelaar (2008) are further discussed in Chapter 8.

The Central Andes has a rich history as being one of the few regions in the New World where complex state societies developed before European contact. A number of different archaeological traditions in the Central Andes, traditionally known as 'horizons', have been linked with different language expansions. For example, the Aymaran language family is thought to have begun its spread from the Nazca region of Peru during the Early Intermediate Period (1800 - 1400 years ago) and expanded during the Middle Horizon (1400 - 1000 years ago), leading of the occupation of much of the southern Peruvian highlands (Torero, 1970). Eventually, the Aymaras reached the Bolivian highlands during the Late Intermediate Period (1000 - 525 years ago), displacing and assimilating groups speaking Uru-Chipayan languages and the now-extinct Puquina, and continued their expansion during Incan and colonial times (Cerrón-Palomino, 2000, 2012).

While there is an ongoing debate about the original homeland of the Quechuan languages, linguistic and archaeological evidence suggests that the family dispersed due to multiple migrations (see Beresford-Jones and Heggarty, 2012, 27-34 for a state of the art). Torero (2002, 125-135) posits an initial split-up of the Quechuan protolanguage during the late Early Horizon (his 'Classic Period'; 2200 - 1400 years ago) that led to the distinction between the major branches of the family, Quechua I and Quechua II, while further expansions during the Late Intermediate and Incan Periods led to the modern distribution of the languages.

Western Amazonia: This region is composed of the tropical forests stretching from the Rio Negro of Brazil and Colombia until the headwaters of the Amazonian tributaries in the foothills of the Andes. The northwest banks of the Madeira and Madre de Dios rivers form the southern boundary of this region.

Western Amazonia hosts a wide variety of different language families, including members of the Arawakan, Panoan, Tucanoan, Jivaroan, Arawan, Witotoan and Nadahupan languages. A number of smaller families such as Peba-Yaguan, Cawapanan and Zaparoan are also located in this region, as well as a number of isolates such as Munichi, Puinave, Urarina, Candoshi and Waorani. Based on a phylogeographical analysis of comparative Arawakan lexical data, Walker and Ribeiro (2011) propose that the original homeland of the Proto-Arawak people was in Western Amazonia before they spread across the rest of the continent.

There has been considerable interaction between the peoples of the foothills

of Western Amazonia and those of the Andean highlands, even to the extant that it has had considerable structural effects on some of these languages, as in the case documented for Quechuan influence on the Arawakan language Yanesha' (Adelaar, 2006). Western Amazonia is also well-known for the sociolinguistic situation among the Tucanoan and Arawakan peoples of the Vaupés River, a tributary of the Rio Negro, where the inhabitants practice a particular form of linguistic exogamy that requires that one marry a partner that speaks a language different than their own. This highly multilingual environment also places cultural restrictions against lexical borrowing, although there has been considerable convergence of a number of grammatical features due to this situation (cf. Aikhenvald, 2006b for an overview).

Western Amazonia is also culturally diverse. Murdock (1951) identifies five different culture areas within the region: Jurua-Purus, Amazon, Caqueta, Loreto and Montaña. The Montaña area in particular is noted for the apparent influence that the highland societies have had on cultures there, such as the presence of loom weaving, non-matrilocal residence patterns, potato cultivation and extensive trade networks, but it is also noted that it shares many cultural traits with the lowland peoples of different regions of the Western Amazon such as the Jurua-Purus.

Southern Amazonia: This region of tropical forests and savannas stretches from the eastern banks of the Madeira and Madre de Dios rivers until the eastern tributaries of the Xingú River. The headwaters of the southern Amazonian tributaries form the border of this region.

Members of all four of the large lowland language families—Tupian, Arawakan, Macro-Jêan and Cariban—are spoken in Southern Amazonia. This region also hosts a number of smaller language families such as Nambikwaran, Chapacuran, Tacanan and Muran, as well as the southern branch of Panoan languages. The upper basin of the Madeira River is renowned for its high concentration of linguistic isolates, such as Kwaza, Kanoê and Aikanã along the Brazilian Guaporé River, and Itonama, Movima, Yurakaré, Canichana and Cayuvava along the Bolivian Mamoré River. Crevels and van der Voort (2008) identify a number of linguistic features common to the languages of the Upper Madeira. This area of Southern Amazonia corresponds neatly to the Marico cultural complex identified by Meireles (1991), which reflects an agglomeration of three culture areas first identified by Lévi-Strauss (1948): the Tupian, Chapacuran and Mojo-Chiquito areas. Metraux (1928) first put forth the hypothesis that the original homeland of the Proto-Tupí people was in Southern Amazonia. Rodrigues (1964) argues in favor of this hypothesis on the basis of linguistic evidence since the Upper Madeira hosts the overwhelming majority of the most divergent branches of the language family.[6] This hypothesis was later supported by archaeological evidence published in Miller (1992).

Southern Amazonia also contains the Xingú culture area, a region of com-

---

[6]See Chapter 5 for a discussion on the current classification of Tupian languages.

plex interaction between multiple ethnicities that include speakers of languages from the Cariban, Tupian, Arawakan, Macro-Jêan families, as well as the isolate Trumai (cf. Fausto et al., 2008; Franchetto, 2011). Heckenberger et al. (2008) describes the emergence of a dense network of small-scale urban settlements connected by raised roadway in the upper Xingú beginning around 1200 years ago. The discovery of these 'garden cities' has greatly contributed to the revision of our understanding of prehistoric Amazonian urbanism, and provides additional support to the proposal put forth in Roosevelt (1994) that the Amazonian floodplains once hosted a much higher population density than that found in the modern-day indigenous cultures of the region.[7] Seki (1999) considers the Upper Xingú an 'incipient linguistic area' due to a number of pervasive lexical borrowings and parallel phonological and grammatical changes.

Chaco-Planalto: This region extends from the central plains and northeastern drylands of Brazil through the Gran Chaco until the Paraná River and its tributaries in Argentina. The headwaters of the southern Amazonian tributaries form the northern boundary. The Atlantic coast forms the eastern boundary.

The Chaco-Planalto region contains languages from the Macro-Jeân, Guaycuruan, Matacoan, Tupian, Arawakan, Lule-Villelan, Mascoyan and Zamucoan languages. The Tupian languages in this region are all members of the expansionist Tupí-Guaraní branch, while the only Arawakan language still spoken in this region is Terêna, a geographic outlier of the family that Aikhenvald (1999) identifies as most closely related to the Baure language of the Bolivian Amazon. Unfortunately, little grammatical information is available on the languages of the Atlantic coast, and as such, none have been included in this study. Campbell and Grondona (2012) provide a number of linguistic features that are shared by many of the languages of the Chaco. A degree of similarity between the Macro-Jêan language family of the Planalto and the Matacoan and Guaycuruan families of the Chaco has been noted in Nonato and Sandalo (2007) and Viegas Barros (2005a), where the first authors attribute this to areal influences while the latter author suggests that they are a result of deep genealogical relations.

Southern Cone: This region includes the semi-arid Patagonian highlands and the archipelago of the Tierra del Fuego. The Atacama Desert and the headwaters of the southern tributaries to the Paraná River form the northern boundary of the region.

The Southern Cone region includes languages from the Chonan family as well as a number of isolates such as Mapudungun, Kawésqar, and Yahgan. Unlike the Mapudungun language that spread from the southern Andean highlands around the time of the European conquest, the original inhabitants of

---

[7]Heckenberger et al. (2011) note that by 1200 AD the Upper Xingú area had a population size and density greater than London during the same period. This population was greatly reduced following European contact.

Tierra del Fuego were primarily nomadic maritime peoples with many shared cultural and linguistic elements (Emperaire, 1955; Viegas Barros, 2005b).

1.2.3   The current state of affairs

The linguistic diversity observable today in South America is only a fraction of the diversity present on the continent when the Europeans arrived at the beginning of the 16th century. Aikhenvald (2012, 21) estimates that up to 60% of the indigenous languages of Amazonia have disappeared since the arrival of the Europeans. Across the continent a number of entire language families have completely disappeared since the colonial period, such as the Charruan languages of Uruguay and Argentina. Furthermore, a large number of indigenous ethnonyms are attested in the historical record for which little to no linguistic data are available for a reliable classification (cf. Loukotka, 1968; Campbell, 2012a). We also know from historical sources and earlier linguistic documentation that a number of the smaller language families in South America were once composed of many more individual members.

The population sizes of indigenous groups and the degree of transmission of their traditional languages varies greatly all across the continent. Members of some language families are spoken by large numbers of people and are still being learned by children, such as Quechuan languages in the Andean highlands that are spoken by over 8 million people (Cerrón-Palomino, 2003). On the other hand, some languages are only remembered by a few elders, as in the case of Puruborá, a Tupian language of Brazil (Galucio, 2005). In general, the small population size of many indigenous groups and the rapid encroachment of national languages such as Spanish and Portuguese have left the vast majority of South American languages endangered or on the brink of extinction (see Crevels, 2012b for an overview).

In recent years there has been an acceleration in the rate and scope of the documentation and description of South American languages. What was once an endeavor primarily restricted to missionaries and foreign academics, a substantial portion of indigenous language research is now being carried out by South American scholars, sometimes even by members of the language communities themselves (e.g. Caesar-Fox, 2003), often in the form of doctoral and master's theses hosted by institutions at home or abroad. A number of institutions in South America now offer doctorate programs in linguistics, and the potential for further development and international cooperation is very high. These new materials have contributed greatly to this project by allowing a more comprehensive coverage of the different languages of the continent in the sample.

1.3   Language sample

The languages used in this study were selected based on three criteria: genealogy, geography and the quality of the available descriptive materials. The quality of the descriptive materials was selected as a basic criterion for the language sample in order to have a high reliability in the coded data. In essence, the sample is a convenience sample since no a priori restrictions were placed on the inclusion of a particular language except for the quality of the available materials. Phylogenetic diversity and geographic diversity were considered when selecting languages to provide a degree of stratification into the sample in order to minimize the effect that autocorrelation has in the resulting data and its analysis (cf. Dryer, 1989; Mace and Pagel, 1994; Cysouw, 2005). Given the nature of the statistical tests used to examine the geographic distribution of argument marking features in Chapter 8, a genealogically-stratified sample with a high degree of geographic spread was chosen instead of a fully randomized sample (cf. Janssen et al., 2006; Bickel and Nichols, 2006).

The sample was designed to include the core sample of South American languages designed by Mily Crevels for the Languages in Contact research group at Radboud University. This core sample was also used in two other comparative studies on South American languages within the same research group, namely Krasnoukhova (2012) on noun phrase structure and Mueller (2013) on tense, aspect, modality and evidentiality marking. While each study has included additional languages according to their respective authors, the core sample was maintained across the studies to allow for greater cohesion and comparability across the larger project.

A total of 74 languages from 40 different language families are used in the sample, including 13 linguistic isolates. For the large language families of South America that have more than 20 members—Arawakan, Cariban, Macro-Jêan, Panoan and Tupian—at least three members from each family were included.[8] For the smaller families of the continent, one or two representative languages were included based on the available materials. When multiple languages were selected as representative for a language family, an attempt was made to include members from different internal branches of the family in order to maximize the representation of the internal diversity of that family in the sample. For example, within the Arawakan family, each of the six languages included in the sample are members of different subgroups of the family following the classification in Aikhenvald (1999), albeit not every subgroup is represented. It is

---

[8]There is an ongoing debate on the number of mutually unintelligible Quechuan languages (cf. Torero, 2002; Adelaar and Muysken, 2004). For the sake of this study, only two Quechuan languages were included, each from different branches of the family. The Tucanoan languages are also on the borderline between what is considered a large family (>20 language), with some debate as to whether some members of the families are best considered distinct languages or dialects of each other. Campbell (2012a, 107-108) and Barnes (1999, 207-209) each identify 18 members of the family. Due to the lack of complete descriptive materials on members of the western branch of the family, only languages from the eastern and central branches of the family were included.

recognized that languages of the Tupian language family are proportionally overrepresented in the sample. However, due to their widespread geographic distribution, considerable family internal structural diversity and the presence of these languages in the structural database due to their inclusion in a diachronic case study (cf. Chapter 5), the languages remain in the sample for the sake of increasing sampling density.[9]

The languages used in this study are presented in Table 1.1, alphabetically organized by language family.

LANGUAGE	FAMILY	REGION	PRIMARY SOURCES
Apurinã	Arawakan	Western Amazonia	Facundes (2000)
Baure	Arawakan	S. Amazonia	Danielsen (2007)
Lokono	Arawakan	N. Amazonia	Pet (2011)
Paresi	Arawakan	S. Amazonia	Brandão (2010); Rowan and Burgess (2008)
Tariana	Arawakan	W. Amazonia	Aikhenvald (2003a)
Yanesha'	Arawakan	W. Amazonia	Duff-Tripp (1997); Wise (1986, 1990)
Jarawara	Arawan	W. Amazonia	Dixon (2004)
Aymara	Aymaran	C. Andes	Hardman (2001); Cerrón-Palomino and Carvajal Carvajal (2009)
Jaqaru	Aymaran	C. Andes	Hardman (2000)
Awa Pit	Barbacoan	N. Andes	Curnow (1997)
Tsafiki	Barbacoan	N. Andes	Dickinson (2002)
Apalaí	Cariban	N. Amazonia	Koehn and Koehn (1986)
Hixkaryana	Cariban	N. Amazonia	Derbyshire (1979, 1985)
Ikpeng	Cariban	S. Amazonia	Pachêco (2001); Chagas (2013)
Tiriyó	Cariban	N. Amazonia	Meira (1999)
Chayahuita	Cawapanan	W. Amazonia	Rojas Berscia (2013)
Wari'	Chapacuran	S. Amazonia	Everett and Kern (1997)

Language sample(*Continued on next page*)

---

[9]The sampled Tupian languages are distributed as following across the ten primary branches (or genera) of the family: Arikém (1), Awetí (1), Juruna (1), Mawé (1), Mundurukú (1), Ramarama (1), Tuparí (2) and Tupí-Guaraní (5). The Tupí-Guaraní languages were sampled from different subrgoupings within the branch, following the classification in Rodrigues (1999), and occur in four different geographic regions.

Table 1.1 – *Continued from previous page*

LANGUAGE	FAMILY	REGION	PRIMARY SOURCES
Chimila	Chibchan	N. Andes	Trillos Amaya (1996)
Ika	Chibchan	N. Andes	Frank (1985)
Embera	Chocoan	N. Andes	Mortensen (1999); Aguirre Licht (1998)
Tehuelche	Chonan	S. Cone	Fernández Garay (1998)
Mocoví	Guaycuruan	Chaco-Planalto	Gualdieri (1998); Guirardello (1999)
Pilagá	Guaycuruan	Chaco-Planalto	Vidal (2001)
Cholón	Hibito-Cholonan	C. Andes	Alexander-Bakkerus (2005)
Itonama	isolate	S. Amazonia	Crevels (2011, 2012a)
Kanoê	isolate	S. Amazonia	Bacelar (2004)
Kwaza	isolate	S. Amazonia	van der Voort (2004)
Leko	isolate	C. Andes	van de Kerke (2009)
Mapundungun	isolate	S. Cone	Smeets (2008); Zúñiga (2006b); Golluscio (2010)
Movima	isolate	S. Amazonia	Haude (2006)
Munichi	isolate	W. Amazonia	Gibson (1996)
Nasa Yuwe	isolate	N. Andes	Jung (2008)
Puinave	isolate	W. Amazonia	Girón (2008)
Trumai	isolate	S. Amazonia	Guirardello (1999)
Urarina	isolate	W. Amazonia	Olawsky (2006)
Warao	isolate	N. Amazonia	Romero-Figeroa (1997)
Yurakaré	isolate	S. Amazonia	van Gijn (2006)
Aguaruna	Jivaroan	W. Amazonia	Overall (2007)
Katukina-Kanamari	Katukinan	W. Amazonia	Queixalós (2010); dos Anjos (2011)
Bororo	Macro-Jêan	Chaco-Planalto	Nonato (2008); Crowell (1979)
Guató	Macro-Jêan	Chaco-Planalto	Palácio (1984, 2004)
Rikbaktsa	Macro-Jêan	S. Amazonia	Silva (2011)
Timbira	Macro-Jêan	Chaco-Planalto	Castro Alves (2004, 2010)

Language sample(*Continued on next page*)

Table 1.1 – *Continued from previous page*

LANGUAGE	FAMILY	REGION	PRIMARY SOURCES
Xavánte	Macro-Jêan	Chaco-Planalto	Estevam (2011); dos Santos (2008)
Wichí	Matacoan	Chaco-Planalto	Terraza (2009)
Moseten	Mosetenan	S. Amazonia	Sakel (2004, 2011)
Dâw	Nadahupan	W. Amazonia	Martins (2004)
Hup	Nadahupan	W. Amazonia	Epps (2008)
Latundê	Nambikwaran	S. Amazonia	Telles (2002)
Sabanê	Nambikwaran	S. Amazonia	Antunes de Araujo (2004)
Matses	Panoan	W. Amazonia	Fleck (2003)
Shipibo	Panoan	W. Amazonia	Valenzuela (1997, 2003)
Yaminahua	Panoan	W. Amazonia	Faust and Loos (2002)
Huallaga Quechua	Quechuan	C. Andes	Weber (1989, 1983)
Imbabura Quechua	Quechuan	N. Andes	Cole (1982)
Cavineña	Tacanan	S. Amazonia	Guillaume (2008)
Cubeo	Tucanoan	W. Amazonia	Morse and Maxwell (1999); Chacon (2012)
Desano	Tucanoan	W. Amazonia	Miller (1999)
Awetí	Tupian	S. Amazonia	Borella (2000)
Cocama-Cocamilla	Tupian	W. Amazonia	Vallejos Yopán (2010)
Emerillon	Tupian	N. Amazonia	Rose (2003)
Juruna	Tupian	S. Amazonia	Fargetti (2001); Lima (2008)
Kamayurá	Tupian	S. Amazonia	Seki (2000)
Karitiana	Tupian	S. Amazonia	Everett (2006)
Karo	Tupian	S. Amazonia	Gabas Jr. (1999)
Mekens	Tupian	S. Amazonia	Galucio (2001)
Mundurukú	Tupian	S. Amazonia	Gomes (2006)
Nheengatú	Tupian	W. Amazonia	Cruz (2011)
Sateré-Mawé	Tupian	S. Amazonia	Silva (2010)
Tapiete	Tupian	Chaco-Planalto	González (2005)
Wayoró	Tupian	S. Amazonia	Nogueira (2011)
Chipaya	Uru-Chipayan	C. Andes	Cerrón-Palomino (2009)
Miraña	Boran	W. Amazonia	Seifart (2005, 2014)
Yanam	Yanomaman	N. Amazonia	Goodwin Gómez (1990); Ferreira (2012)

1.4   Organization of the thesis

The remainder of the thesis has the following structure:

Chapter 2 begins the investigation into argument marking patterns by outlining the methodology used in this study, looking at both the basic concepts used as a standard of comparison across the languages in the sample and the way that the observed data is organized in the structural questionnaire.

Chapter 3 introduces the first grammatical domain of investigation—the marking of arguments on the predicate—and outlines the ways in which the languages in the sample vary according to a number of major structural parameters in this domain. This chapter discusses the marking of person, number and gender on the predicate and pays special attention to the alignment of bound person forms.

Chapter 4 continues the investigation into verbal argument marking through the examination of two commonly occurring patterns that often elude categorical classification and are thus treated as 'split' systems: hierarchical marking and split intransitivity. This chapter shows that a great degree of structural variation can be observed in each marking pattern and outlines an approach to capture the major distinction of these patterns while still allowing them to be compared with marking patterns that do not show patterns conditioned by predicate classes or argument configurations.

Chapter 5 looks at the development of verbal argument marking among related languages through a case study on the Tupian language family. This case study considers the different proposals for the historical development of the various person prefix sets across the family and explores the viability of reconstructing the development of these sets using a parsimony analysis applied over two competing classifications of the family.

Chapter 6 explores the second major domain of argument marking under consideration: the realization of overt case distinctions on arguments. A number of different issues are discussed in relation to case marking: the distinction of case marking from pragmatic role marking, case marking patterns in intransitive, transitive and ditransitive constructions, as well as the distinction between the case marking of noun phrase arguments and pronouns.

Chapter 7 gives an overview of major strategies used to alter the argument marking pattern of a particular predicate through the increase, decrease and rearrangement of valency and the promotion and demotion of different clausal participants. The construction types under consideration are causatives, applicatives, passives, antipassives, anticausatives, reflexives and reciprocals, with additional discussion on the category of middle voice. The major valency changing operations are presented and discussed with regard to their effects on argument marking across the different derivational strategies.

Chapter 8 analyzes a number of the structural features observed in the

sample for their geographic distribution. The chapter adopts an approach to establishing the areality of a feature based on the ability of a particular region to predict the observed typological distribution of that feature. Multiple previous claims are examined and tested for different geographic regions of the continent. A number of structural features are identified as being areal to specific regions, at times in patterns that necessitate a reevaluation of certain claims in earlier work.

Chapter 9 concluded the thesis by summarizing the major findings of this study and proposing new lines for further research.

Methodology

This chapter introduces the methodology adopted in this study to compare argument marking patterns across the sample of South American languages. Section 2.1 outlines the linguistic concepts that have been used to establish an independent standard of comparison among these languages. Section 2.2 gives an overview as to how the structural distinctions observed in these languages are systematically encoded in a questionnaire designed to facilitate quantitative analysis and comparison of the data. It is important to emphasize that the aim of the methodology is to facilitate the comparison and compilation of the structural facts related to the morphosyntactic encoding of arguments in these languages. No attempt is made in this study to interpret these structures within a particular theoretical framework nor to make any explicit statements about the nature of language in general. However, the discrete nature of the information encoded in the structural questionnaire allows for a number of observations to be made about the languages in the sample. To the extent that the language sample can be taken as representative of the distribution of South American languages as a whole, the data presented in this thesis is further used to describe the geographic distributions of specific argument marking patterns across the continent (see Chapter 8).

2.1 Basic comparative concepts

A major advance in modern linguistics came from the understanding that not all languages can be described by using the same grammatical categories. For instance, Boas (1911, 81) highlights the inadequacy of terminology adopted

from traditional Latin grammar to describe indigenous languages of the Americas, stating the need to rely on categories derived from "the inner form of each language". This structuralist approach has led to languages being described in terms of descriptive categories that arise from language-specific properties rather than a universal list of predefined categories. Therefore an essential task of the descriptive linguist is to identify the grammatical properties of a particular language that produce the relevant distinctions (categories) needed to adequately account for the range of different constructions that occur in language use. The role of the descriptivist is thus quite distinct from that of the language typologist. The primary role of the typologist is to establish an independent standard of comparison by which specific linguistic properties can be compared across a sample of languages with distinct grammatical structures. This standard of comparison can then be used to express major typological differences between languages, state cross-linguistic generalizations observed in the data, and also serve as a basis to define further comparative concepts (Haspelmath, 2011a, 555).

In grammar writing, labels for language-specific categories are often adopted from traditional categories such as 'adjective' and 'past tense' if they can be distinguished using language-internal criteria. Even in these cases it is assumed that the grammatical properties of a particular category in a specific language are not identical to the properties of the corresponding category in another language (cf. Haspelmath, 2010). In this sense, the traditional labels adopted in grammars do not attest to the universality of such categories, but rather, these labels represent a language-specific category that shows sufficient similarity with the comparative concept embodied by that label to warrant its use.[1]

The language-specific nature of grammatical categories necessitates an independent standard of comparison to carry out any cross-linguistic investigation of a particular linguistic property. In principle, all languages have the ability to express any meaning, and thus, the semantic component of language is shared among all languages even though grammatical categories are language-specific. Typologists working in the tradition of Greenberg (1963) have long used semantic properties of a language as an invariant *tertium comparationis* for cross-linguistic study of grammatical properties.[2] Following in the Greenbergian tradition, the comparative concepts used in this thesis are defined by their semantic and functional properties rather than grammatical properties.

---

[1]The application of different traditional labels to language-specific categories is a choice made by the respective authors in a language description. Sometimes there are disagreements among typologists and descriptivists regarding whether a particular category within a language is best considered as a distinct category or should be subsumed under a larger category. A recent example of this in the South American literature is the discussion of whether adjectives in Quechuan languages are best considered an independent word class or a subclass of verbs, as taken up in Haspelmath (2012) responding to Floyd (2011).

[2]The application of strictly semantic criteria for the identification of grammatical structures is more accurately described as a post-Greenbergian development in typology. However, Greenberg (1963, 74) did recognize that "in identifying such phenomena in languages of differing structure, one is basically employing semantic criteria".

Comparative concepts are used as a basis to compare morphosyntactic CON-STRUCTIONS, which are grammatical patterns within a language that "link together idiosyncratic or arbitrary phonological, syntactic and semantic information" (Croft, 2001, 16). At the clausal level, a construction is a generalization of a specific morphosyntactic pattern, independent of the meaning of any particular predicate and epiphenominal in nature (Goldberg, 1995, 1). An example of a construction at the clause level is the transitive construction discussed in section 2.1.2, while an example of a construction at the phrasal level is the possessive construction, such as *maʔwɨt at kaʔa* 'the man's house' in Karo (4). This thesis is primarily concerned with morphosyntactic constructions that are formed in independent main clauses.

Within a construction, various rules, constraints and processes can apply to a set of arguments. These are referred to in this thesis as ARGUMENT SE-LECTORS, adopting the term used in Witzlack-Makarevich (2010), since they select a sets of arguments to be treated in a specific way, such as being marked with a particular case suffix. There is a long tradition in language typology to distinguish between argument selectors that function through the overt morphological encoding of arguments, sometimes called 'coding properties' after Keenan (1976), and those that are restricted to syntactic processes, sometimes called 'behavior and control properties' (see also Comrie, 1981; Dixon, 1994). Some examples of the latter type of argument selectors are constituent order, argument co-reference across clauses and access to certain types of relativization. Argument selectors that encode arguments and the relations that they bear with their predicate through overt morphological distinctions are called ARGU-MENT MARKERS in this thesis. Two types of argument markers are the focus of this study: case markers and verbal argument markers. CASE MARKERS express the relationship that an argument holds with a predicate through overt morphological distinctions on the arguments themselves, generally through affixation, form alternation or adpositions. VERBAL ARGUMENT MARKERS express the relationship that an argument holds with a predicate through overt morphological distinctions on the predicate itself, generally through the attachment of affixes and clitics. The grammatical process by which verbal argument markers express these relationships with the predicate is called INDEXATION. Argument markers can select different sets of arguments in different morphosyntactic constructions. The focus of the rest of this chapter is on how the current study distinguishes which morphosyntactic constructions to investigate, and how to represent the treatment of sets of arguments by different argument markers. In the following subsections, a general framework for the cross-linguistic comparison of argument marking is presented that uses the expression of meaning as a standard of comparison across different morphosyntactic constructions in different languages.

2.1.1 Arguments and semantic roles

Central to any discussion of argument marking is the distinction between arguments and adjuncts. The expression of a state or event generally requires that it be in relation to some entity in the world, either real or conceptual. These entities are called the SEMANTIC PARTICIPANTS of a clause. An ARGUMENT is a semantic participant that must be expressed in relation to a predicate (usually a verb) in order to form a grammatical utterance. An ADJUNCT is a participant that does not need to be obligatorily expressed in relation to a predicate in order to form a grammatical utterance. In other words, arguments are "part of the semantic representation of the verb" (Van Valin Jr. and LaPolla, 1997, 26), while adjuncts "provide supplementary information and refer to the circumstantial features of the setting (including its manner, place, and time)" (Witzlack-Makarevich, 2010, 42).

The semantic interpretation of an argument is conditioned by its predicate, while the semantic interpretation of adjuncts is independent of the meaning of the predicate (Comrie, 1993, 907). Predicates generally require one, two or three arguments, and are thus referred to as one-, two- and three-place predicates, respectively.[3] The number of arguments that a predicate requires is referred to as its VALENCY, a term adopted from chemistry into linguistics by Hockett (1958) and Tesnière (1959). The distinction between arguments and adjuncts can be seen for Karo, a Tupian language of Brazil, in (1):

Karo (Tupian; Gabas Jr. 1999, 31,93,164, 166)

(1)   a. *o=kə-t*
           1SG=walk-IND
           'I walked.'

       b. *ip   cú kə-t   icɨ   pət*
           fish big walk-IND water INESS
           'The big fish swam in the water.'

       c. *õn   korét wĩ-t*
           1SG bird   kill-IND
           'I killed the bird.'

       d. *pẽŋ   aʔ=wĩ-t   tágip mã*
           white.man 3SG=kill-IND bow   INS
           'The white man killed it with a bow.'

---

[3]In the current study, the number of arguments that a predicate requires is based on the descriptive materials available for that language. No attempts have been made to independently establish the valency frame of each individual predicate since this would require an in-depth lexico-semantic analysis of every language in the sample. Such a task falls well outside of the scope of this study, but this an issue that deserves further attention in both descriptive and typological studies. See Comrie (1993) and Witzlack-Makarevich and Bickel (2013) for further discussion on the inherent difficulties in distinguishing between arguments and adjuncts in modern linguistic theory.

The verb *kə* 'walk/swim' in Karo is a one-place predicate, i.e. it is monovalent, thus requiring only a single argument. In (1a) the argument of the predicate is expressed by the proclitic *o=* 'I', while in (1b) the sole argument of the predicate is expressed by the noun phrase *ip cú* 'big fish'. An additional semantic participant occurs in (1b), *icɨ* 'water'. This entity is not required in the clause for the utterance to be grammatical, as (1a) shows, and is thus an adjunct.

The verb *wĩ* 'kill' is a two-place predicate in Karo, i.e. it is bivalent, thus requiring two arguments to be expressed for grammaticality. In (1c) the arguments of the predicate are expressed by a free pronoun *õn* 'I' and the noun *korét* 'bird'. In (1d) the arguments of the predicate are expressed by the noun *pẽŋ* 'white man' and the proclitic *aʔ=* 'he/it'. Like in (1b), an additional entity occurs in (1d), *tágip* 'bow', which is not required by the predicate in order to form a grammatical utterance, and is therefore an adjunct. In Karo, adjuncts tend to be expressed post-verbally and are marked by postpositions.

Semantically, the single argument of the predicate *kə* 'walk/swim' must be a 'walker' (1a) or a 'swimmer' (1b). The two arguments of the predicate *wĩ* 'kill' must be an entity that carries out the event, a 'killer', and an entity that is affected by the event, a 'killed'. The adjuncts also hold a semantic relationship with the predicate: the water in (1b) is the location where the swimming takes place, while the bow in (1b) is the instrument used to carry out the killing. The semantic relationship that an argument or adjunct holds with its predicate is called a SEMANTIC ROLE.[4] While it is sometimes useful to discuss the arguments of a specific predicate with predicate-specific semantic roles, such as a swimmer, a killer and a killed, it is also possible to group semantic roles into more general categories.

The two most relevant general semantic roles for any discussion on argument marking is that of AGENT and PATIENT. Andrews (2007, 137) defines agent as "a participant which the meaning of the verb describes as doing something, or causing something to happen", and defines patient as "a participant which the verb describes as having something happen to it, and as being affected by what happens to it". At least since Gruber (1965), volition and intentionality has also been considered a properties of agent. In this sense, the walker, swimmer and killer in (1) are all agents, while the killed participants in (1c-1d) are patients. The precise definition of agent and patient, and the different properties associated with these different roles, has been the subject of intense discussion in modern linguistics (such as Jackendoff, 1976; Lakoff, 1977; Dowty, 1991; Primus, 1999, among many others). For this thesis, the definition of the agent as a volitional performer of an event and the patient as an entity affected by an event is adopted.

It is important to note that the underlying semantic structure of a proposition is independent of its syntactic structure (Van Valin Jr. and LaPolla, 1997;

---

[4]A variety of different labels have been given to the concept of semantic roles: thematic roles (Dowty, 1991), theta roles (Chomsky, 1981), thematic relations (Van Valin Jr. and LaPolla, 1997), participant roles (Croft, 2001) and case roles (Fillmore, 1968).

Croft, 2001). Throughout this thesis, the underlying semantic representation of an utterance is referred to as a SITUATION TYPE, following the use of the term in other typological studies such as Talmy (1976) and Kemmer (1993). The components of a situation type that are especially relevant to this study are the semantic roles of the participants in an utterance and the lexical aspect of the state of affairs expressed by the predicate. Additional semantic components of a proposition are discussed when relevant. The concept of a situation type is used throughout this thesis as a heuristic to constrain the semantic properties of a proposition in order to compare diverse morphosyntactic structures across languages.

2.1.2 Transitivity and argument roles

In contrast to the concept of valency, which relates the semantic properties of arguments to a particular predicate, TRANSITIVITY is a semantic property of a clause that describes the extent to which "an activity is 'carried over' or 'transferred' from an agent to a patient" (Hopper and Thompson, 1980, 251). In this thesis, the notion of transitivity is used to constrain the area of investigation concerning two-place predicates to those that express a transitive situation type. A TRANSITIVE SITUATION TYPE is "an effective volitional discrete action performed by a controlling agent and actually affecting a well individuated patient" (Lazard, 2002, 152).[5] Following Givón (2001) and Lazard (2002), the transitive situation type has semantic properties related both to the event itself and to the obligatory participants of the predicate that expresses the event. The event properties are given in (2) and the argument properties are given in (3):

(2)   Event properties of a transitive situation type per Lazard (2002)

    a. The event occurs in the real world, i.e. it is not prospective or imagined.

    b. The event is discrete, i.e. perfective or completive, not progressive, conative, habitual. etc., or somehow incomplete.[6]

(3)   Argument properties of a transitive situation type per Lazard (2002)

    a. The agent is a human that voluntarily performs the event.

    b. The agent controls the event, i.e. the event is not a natural process nor non-intentional or uncontrolled.

---

[5]Lazard (2002, 152) refers to the transitive situation type as the 'prototypical action'.

[6]In the few cases where there are multiple marking patterns conditioned by tense, mood and/or aspectual values that still satisfy the perfectivity criterion, the construction that expresses the past perfective, also commonly called the aorist, preterite or simple past, is selected as the basic construction for comparison. Multiple perfective transitive constructions are primarily restricted to Cariban languages in the sample. See Gildea (1998) for an in-depth discussion on these construction types in Cariban languages and their historical development.

    c. The patient is a well individuated, animate entity.

    d. The patient is actually affected, i.e. it undergoes a change as a consequence of the event.

Haspelmath (2011a) labels the class of verbs most likely to present these semantic properties as 'physical effect verbs', with typical members of this class being verbs like 'kill' and 'break'. This transitive situation type has the prototypical semantic properties of a TRANSITIVE CONSTRUCTION, which is defined as the morphosyntactic construction used to express a transitive situation type. Following this definition, the Karo examples (1c) and (1d) are clear instantiations of transitive constructions with the transitive verb *kə* 'kill' as its predicate.

In addition to the transitive construction for bivalent predicates, corresponding constructions are used for monovalent and trivalent predicates. An INTRANSITIVE CONSTRUCTION uses an intransitive situation type as its prototype, which can be characterized for the time being as a monovalent predicate expressing an event (but see section 4.2 for a discussion on how this characterization of an intransitive construction can be problematic for cross-linguistic comparison). A DITRANSITIVE CONSTRUCTION uses a ditransitive situation type as its prototype, which shares event properties with the prototypical action described in (2). Unlike a transitive situation type, a ditransitive situation type has two non-agent participants in addition to an agent: the recipient and the theme. In verbs that denote a physical transfer of possession such as 'give' or 'send', the recipient argument is the participant who acquires possession of the entity whose possession of which is being transferred, the theme. Since the seminal work in Dryer (1986), it has become common for language typologists to consider the non-agent arguments of ditransitive verbs separately from, and in comparison with, the patient argument of a transitive verb. Intransitive, transitive and ditransitive constructions identified using the semantic criteria discussed above are referred to in this thesis as BASIC CONSTRUCTIONS.

The transitive construction and the morphosyntactic treatment of the agent and patient arguments form a standard to compare the syntactic function of different arguments cross-linguistically. These different syntactic functions will be referred to in this thesis as ARGUMENT ROLES, adopting the term from Bickel and Nichols (2009). Using the classic schema popularized in Dixon (1979) and Comrie (1981), the different argument roles can be defined and referred to with the terms given in Table 2.1.

As with all morphosyntactic constructions, not only the semantically prototypical cases are considered examples of the construction. Additionally, all other instances that show identical morphosyntactic treatment of their arguments as that in clauses expressing a transitive situation type are considered as forming a transitive construction. With this in mind, compare the morphosyntactic treatment of the agent of the verb 'kill' in (1c) with that of the experiencer of the verb 'see' in (4):

Role	Definition
S	Sole argument of an intransitive clause
A	Argument treated like the agent of a (di)transitive clause
P	Argument treated like the patient of a transitive clause
T	Argument treated like the theme of a ditransitive clause
R	Argument treated like the recipient of a ditransitive clause

Table 2.1: Morphosyntactic argument roles in intransitive, transitive and ditransitive clauses

Karo (Tupian; Gabas Jr. 1999, 74)

(4) **õn** *maʔwɨt at   kaʔa top-t*
     1SG man   POSS house see-IND

     'I saw the man's house.'

The experiencer participant of the verb *top* 'see' in (4) is morphosyntactically treated the same as the agent participant of the verb *kə* 'kill' in (1c); both arguments are realized as the free pronoun *õn*, occur in a clause initial position, are not case marked with any overt morphology and are not indexed on the verb. Due to these morphosyntactic properties, the verb 'see' in Karo forms a transitive construction since its arguments are treated identically to the arguments of a transitive situation type, i.e. it is a transitive verb. While the argument *õn* in (1c) and (4) hold different semantic roles in relation to the predicate, they can both be considered A arguments since they receive the same morphosyntactic treatment within a transitive clause. Likewise, the non-A argument in (4) *maʔwɨt at kaʔa* 'the man's house' holds the semantic role of stimulus with the predicate and is not semantically a patient since it is not affected by the event. However, the stimulus participant in (4) is also the P argument since it shows identical morphosyntactic properties with the patient participant in (1c): it occurs clause medially between the A argument and the verb, it is not case marked, and had it not been expressed as a full noun phrase (NP), it would have been indexed as a clitic attached to the verb as in (1d).

Some linguistic approaches account for the identical morphosyntactic treatment of different semantic roles through aggregate semantic relations, such as 'macroroles' (Foley and Van Valin Jr., 1984) or 'proto-roles' (Dowty, 1991). For example, in the Role and Reference Grammar framework outlined in Foley and Van Valin Jr. (1984) and Van Valin Jr. and LaPolla (1997), the identical treatment of the Karo experiencer participant in (4) and the agent participant in (1c) is accounted for by considering both of these semantic roles as subsumed under the macrorole of Actor. While approaches such as these often provide useful insights into the way that different semantic roles are expressed in a language, they do not play an explicit role in the coding of the structural questionnaire used in this thesis. However, to capture this generalization, argu-

ments of bivalent predicates that are treated identically to agent participants in transitive situation types are referred to as AGENTIVE ARGUMENTS, while those that are treated identically to patient participants in transitive situation types are referred to as PATIENTIVE ARGUMENTS. While in most cases, the terms A argument and agentive argument are synonymous, such a distinction becomes useful when analyzing valency-changing derivations, as discussed in Chapter 7.

The arguments in a clause that are marked identically to the S, A and P argument roles of transitive and intransitive constructions are CORE ARGUMENTS. When discussing in terms of morphosyntactic treatment rather than semantic relations, adjuncts are referred to as OBLIQUES. Arguments that are required by the predicate but show a morphosyntactic treatment distinct from core arguments and identical to obliques are OBLIQUE ARGUMENTS.

Contrary to the position taken in Dixon (1994), where the argument roles S, A, P are innate "primative" categories in every language, this thesis adopts the position these terms are merely comparative concepts that function as analytic tools to compare the syntactic functions of different arguments by using the semantic roles of these arguments as a standard of comparison. This position is in line with the use of these terms in Comrie (1981).[7] It is worth noting that by restricting the analysis to a subset of constructions based around clauses that show semantic properties that allow them to be classified as (in-/di-)transitive, only a portion of the full diversity of argument marking patterns within a language are observed and discussed. However, this restriction also allows for a more precise formulation of the domain under comparison. With these analytic tools in hand, it is now possible to discuss how different argument types can be organized within the grammar of a language to express different grammatical relations.

2.1.3   Grammatical relations

Distinct from semantic roles, GRAMMATICAL RELATIONS are the "morphosyntactic properties that relate an argument to a clause" (Bickel, 2011, 399).[8] Grammatical relations are traditionally discussed in terms of the concepts of subject, object, indirect object and so forth. In this thesis, different grammatical relations are defined as a set composed of the morphosyntactic argument roles that are treated identically by an argument selector within a specific construction. The labels for different grammatical relations used in this thesis are given in Table 2.2. For a more exhaustive list of the terms used in the literature, see Bickel (2011, 404).

---

[7]For further discussion on the distinction between the Dixonian and the Comrian use of the SAPTR (sometimes SAOTG) terms, and the implications this has on comparative analyses, see Haspelmath (2011a).

[8] A variety of different labels have been given to the concept of grammatical relations: syntactic function (Dik, 1997), syntactic role (Croft, 2001) and grammatical function (Bresnan, 2001) are among the most common.

ARGUMENT ROLES	GRAMMATICAL RELATION
S	intransitive subject
S, A	subject, nominative
A	transitive subject, ergative
S, P	absolutive
P, T	direct object
P, R	primary object
T	secondary object
R	indirect object, dative

Table 2.2: Grammatical relations as a set of argument roles

The properties associated with these grammatical relations are often discussed with regard to the way that the arguments are marked in the clause, such as how NPs themselves are marked for grammatical relations (case marking) and how the arguments are marked on the predicate (indexation). When discussing the morphosyntactic properties of an argument or set of arguments in this thesis, these arguments will be referred to either by their argument role (e.g. the A argument) or by the grammatical relation that they form (e.g. the absolutive argument). Aside from a few illustrative examples presented here for the purpose of discussion, a more exhaustive set of examples of the different grammatical relations formed by different argument selectors in the languages of the sample are given in the chapters that follow.

Grammatical relations are often discussed with regards to the alignment that they display. Closely related to the concept of grammatical relations, ALIGNMENT is defined as a set of arguments that are treated identically in contrast to other argument roles. In this sense, alignment is the "neutralization of valence-specific argument roles in particular morphological and syntactic contexts" (Bickel and Nichols, 2009, 305). A list of the most common alignment types are given in Table 2.3.

SETS OF ARGUMENT ROLES	ALIGNMENT TYPE
S=A≠P	accusative
S=P≠A	ergative
S=A=P	neutral
S≠A≠P	tripartite
P=T≠R	indirective
P=R≠T	secundative
P=T=R	neutral

Table 2.3: Alignment types as sets of identically and differentially treated argument roles

Other often discussed alignment types in the literature such as 'active-

stative' (split intransitivity) and 'hierarchical' are treated in this thesis as marking patterns composed of multiple different alignments (see Chapter 4).

To illustrate the terminological distinction between grammatical relations and alignment, reconsider the Karo example in (1). In (1a) the S argument is indexed on the verb in the same way as the P argument in (1c), thus the absolutive argument is indexed in Karo, i.e. Karo has an absolutive grammatical relation that is marked through indexation. Furthermore, the S and P arguments are treated differently than the A argument, thus Karo shows ergative alignment in indexation. Correspondingly, the Katukina-Kanamari language of Brazil only indexes the ergative (A) argument, as shown in (5), but not the S or P arguments. Just like Karo, Katukina-Kanamari also shows ergative alignment in indexation even though a different grammatical relation is indexed.

Katukina-Kanamari (Katukinan; Queixalós, 2010, 238-239)

(5)   a.   *ki:tan idi:k*
           sleep   you
           'You slept.'

   b.   ***no**-ti   paiko*
           2sg-kill grandfather
           'You killed grandfather.'

Traditionally, languages have been described as being primarily 'ergative' or primarily 'accusative' based on the alignment that they display in case marking and indexation.[9] However, such a rough characterization of a language can be problematic since it is not always the case that these two argument markers display the same alignment. To illustrate this, consider some examples from Chayahuita in (6), a Cawapanan language spoken in Peru:

Chayahuita (Cawapanan; Rojas Berscia, 2013, 58-59, 61)

(6)   a.   *balsapuerto-ke   ka   sahkat-**awe***
           Balspuerto-loc 1sg work-1sg
           'I work in Balsapuerto.'

   b.   *ka-**ri**   ina tehpar-**awe***
           1sg-erg 3sg kill-1sg
           'I killed him.'

---

[9]Much modern research on grammatical relations, especially since Anderson (1976) and Keenan (1976), has focused a variety of argument selectors besides just case marking and indexation, such as argument co-reference across clauses and access to particular valency changing operations and other derivations such as the formation of relative clauses. Since this thesis only focuses on argument marking within independent clauses, different grammatical relations established through co-reference across clauses or in the formation of relative clauses will not be discussed. Valency-changing operations are discussed in Chapter 7.

c. *ina-**ri***   *tehpar-**in**-**ku***
   3SG-ERG kill-3SG-1SG
   'He killed me.'

In Chayahuita, the person of the nominative argument is indexed on the verb through a set of suffixes, in (6a) and (6b) through the 1st person marker -*awe*, while the 1st person P argument is indexed with the suffix -*ku* as seen in (6c), showing accusative alignment. However, Chayahuita also marks the ergative argument with the case suffix -*ri*, as seen in (6b) and (6c), showing ergative alignment. For this reason, any discussion of alignment and grammatical relations in this thesis is always in reference to a particular argument marker. No attempt is made here to classify entire languages as 'ergative' or 'accusative' based solely on argument marking properties since not all possible argument selectors that could form grammatical relations are examined for every language.

2.1.4   Distinguishing pronouns and verbal indexation

The major distinction between verbal argument markers and case markers is the locus of marking in the clause (Nichols, 1986). Languages with verbal argument markers are often called 'head-marking' since argument marking is expressed through morphological distinctions on the head of the clause, the predicate. Languages with case markers are often called 'dependent marking' since argument marking is expressed through morphological distinctions on the dependents of the clause, the arguments themselves. However, when dealing with PERSON FORMS, closed classes of grammatical morphemes that distinguish between the speaker, addressee and other participants outside of the speech act (first, second and third persons, respectively), the distinction between head-marking and dependent-marking is not always straightforward without explicit defining criteria (Siewierska, 2004; Haspelmath, 2013). This subsection outlines a number of properties that can be used to distinguish between PRONOUNS, person forms that function as clausal substitutes for nouns, and verbal argument markers.

The distinction between pronouns and verbal argument markers is not always entirely straightforward. Two different approaches to making this distinction are commonly used in the typological literature: co-occurrence of a verbal argument marker with a coreferential nominal in the same clause, the CONOMINAL, and the boundedness of the marker on the predicate. As discussed in the following paragraphs, both of these criteria can be somewhat problematic when used to distinguish between verbal argument markers and pronouns.

For Bickel and Nichols (2007) and Witzlack-Makarevich (2010), the main distinction between pronouns and verbal argument markers rests on whether the marker can co-occur with the conominal in the same clause. Such a distinction has a long history in linguistic theory and is well discussed in works

such as Jelinek (1984), Bresnan and Mchombo (1987) and Siewierska (2004). Traditionally, verbal argument markers that can occur with a conominal are called 'agreement' or 'cross-reference' markers, while those that cannot are called 'bound pronouns'. However, for the aims of this study, it is a reasonable question to ask whether person markers bound to the predicate that can occur with a conominal are fundamentally different from those that cannot occur with a conominal, i.e. whether the former case should be considered as an instance of indexation while the latter case should not be. Consider the example from Mekens, a Tupian language from Brazil, in (7):

Mekens (Tupian; Galucio 2001, 78-79)

(7) a. (*kise*)   ***ki****-er-a-t*
         1PL.INCL 1PL.INCL-sleep-THEM-PST
         'We (incl) slept.'

     b. (*sete*) ***ki****-so-a-t*
         3SG   1PL.INCL-see-THEM-PST
         'He saw us (incl).'

     c. *isɨɨ so-a-t*   *õt*
         deer see-THEM-PST 1SG
         'I saw the deer.'

     d. * *isɨɨ i-so-a-t*   *õt*
         deer 3SG-see-THEM-PST 1SG
         'I saw the deer.'

As shown in (7a), the marker *ki-* can co-occur with the free pronoun *kise*, which fits the criteria that verbal argument markers must be able to co-occur with a conominal (optionality of expression is indicated with parentheses). As (7b) shows, the same marker set that indexes S also indexes P, showing ergative alignment. However, in transitive constructions the marker cannot co-occur with its conominal, as shown in (7c) and (7d). Calling the marker *ki-* an 'agreement marker' in (7a) but a 'bound pronoun' in (7b) is an inelegant solution that has repercussions on the typological classification of the language. For example, if the distinction between bound pronouns and agreement markers were adopted, Mekens would be described as showing verbal agreement only for S but not for A. This disregards the fact that these markers are identical and they are both attached to the verb. Instead, Mekens can be described as showing ergative alignment for the indexation of the absolutive argument, albeit the co-occurrence behavior of the marker is different with regard to the indexation of S and P.

Bickel (2008, 194) justifies the use of the co-occurrence criterion to distinguish between 'agreement markers' and 'cliticized pronouns' because the latter are "expected to behave like pronouns". However, that is clearly not the case

in Mekens since the "cliticized pronoun" in (7b) behaves identically to the "agreement marker" in (7a) except regarding co-occurrence with a conominal. The "cliticized pronoun" in (7b) behaves no more like an independent pronoun than the agreement marker in (7a). For the purpose of coding the structural variation observed in the languages of the sample, it is problematic to state that Mekens indexes S but not P, especially considering that both arguments are marked using the same set of markers. For this reason, co-occurrence with a conominal is not adopted as the primary criterion to distinguish between verbal argument markers and pronouns. Co-occurrence of an argument marker with its conominal is further discussed in section 3.3.6 of the following chapter.

The primary criterion for distinguishing verbal argument markers from pronouns in Haspelmath (2013) is whether or not the person form is bound to the predicate. To distinguish syntactically dependent from free forms, he adopts a single criterion from Bloomfield (1933, 160): "Free forms are forms that can occur on their own, i.e. in a complete (possibly elliptical) utterance". Verbal argument markers can show varying degrees of syntactic and phonological dependence on the predicate. In many languages, indexation occurs by way of affixation, whereby the markers are both syntactically and phonologically bound to the predicate. This is in contrast to independent pronouns, which are neither syntactically nor phonologically bound to the predicate and function as a free word within the nominal domain. A third category, the 'clitic', is often invoked to refer to a morpheme that does not neatly conform to the affix vs. independent word distinction (see Zwicky, 1985; Bickel and Nichols, 2007). However, as Haspelmath (2011b) notes, there are no good cross-linguistic criteria to distinguish between affixes and clitics. For the purpose of this study, morphemes that show syntactic dependence on the predicate are treated as verbal argument markers regardless of their phonological status.

Just as verbal argument markers are person forms that are syntactically dependent on the predicate, a pronoun is a person form that is syntactically independent of the predicate. A number of different subclasses of pronouns have been identified across languages, including emphatic pronouns, demonstrative pronouns and reflexive pronouns, but what is of primary concern in this thesis are personal pronouns. A PERSONAL PRONOUN is a person form that is syntactically unbound, i.e. an independent grammatical word, and displays the following properties: a) it can be substituted for an NP participant, i.e. occur in the same syntactic contexts as an NP, and b) they cannot co-occur with a conominal (Schachter and Shopen, 2007; Haspelmath, 2013). Since pronouns can morphologically encode distinctions between different argument roles as dependents of the predicate, variations in the form of pronouns are considered here as a type of case marking.

One strategy for the case marking of pronouns is the addition of special segmentable morphology to a pronominal stem, as seen for the Aymara pronouns in (8):

Aymara (Aymaran; Cerrón-Palomino and Carvajal Carvajal, 2009, 207; Hardman, 2001, 150)

(8)   a. *naya juma-r uñj-sma*
        1SG 2SG-IL see-1→2
        'I see you.'

      b. *naya-t   may-t'a-si-ni-way-itu*
        1SG-ABL borrow-MOM-REFL-PROX-DIST-3→1
        'He borrowed it from me in passing.'

The 1st person singular pronoun in Aymara is *naya* (8a), and a 1st person oblique argument taking the ablative case is formed by adding the case marker *-ta* to the pronominal stem, forming *nayat* (8b). Notice that Aymara marks arguments both through case marking and indexation.

In some languages the pronouns are not marked with segmentable case markers, rather, they are composed of different paradigms that only reference a subset of argument roles, as shown for Warao, a language isolate of Venezuela, in (9):

Warao (isolate; Romero-Figeroa 1997, 5, 64, 66, 71)

(9)   a. *ma-rahe   haya-te*
        1SG-brother run-NPST
        'My brother runs.'

      b. *erike   hube abu-a-e*
        Enrique snake bite-PUNC-PST
        'The snake bit Enrique.'

      c. *ma-ha   eku   **ine**   yak-era   tane   uba-te*
        1SG-hammock inside 1SG.NOM good-AUG MANNER sleep-NPST
        'I sleep well inside of my hammock.'

      d. *warao **ine**   namina-te*
        Warao 1SG.NOM know-NPST
        'I know (speak) Warao.'

      e. ***ma**   airamo deri-n-a-e*
        1SG.ACC chief   advise-SG-PUNC-PST
        'The chief advised me.'

Notice in (9) that no NP arguments are marked for case, while the 1st person singular pronoun *ine* expresses the S and A arguments (9c-9d), forming a nominative grammatical relation, while the pronoun *ma* expresses the P argument (9e), forming an accusative grammatical relation. Unlike in Aymara, Warao pronouns show a case distinction, but this distinction is expressed using

different sets of pronouns rather than the addition of special (segmentable) morphology.

Unlike NP arguments, not all arguments of a predicate can be expressed as a personal pronoun in every language. Thus, when discussing case alignment in pronouns, one needs to consider not only contrasts and similarities of form, but also whether an argument can be expressed as a personal pronoun at all. This restriction on the use of personal pronouns is especially common in languages that index one or more arguments on the verb. To illustrate this point, let us return to argument marking in Mekens, with examples given in (10) partially repeated from (7) for convenience. Remember that Mekens indexes the absolutive argument on the verb with a single marker set, does not allow for the co-occurrence of an P argument with the P verbal argument marker, and does not mark any core NP arguments for case (Galucio, 2001). Additionally, the personal pronouns in Mekens can only be used to express the S or A arguments of a clause, as shown in (10):

Mekens (Tupian; Galucio, 2001, 78-79, 81)

(10) a. **kise**   **ki**-*er-a-t*
          1PL.INCL 1PL.INCL-sleep-THEM-PST
          'We (incl) slept.'

      b. **kise**   **i**-*so-a-t*
          1PL.INCL 3-see-THEM-PST
          'We (incl) saw it.'

      c. **sete ki**-*so-a-t*
          3SG   1PL.INCL-see-THEM-PST
          'He saw us (incl).'

      d. *ɨsɨɨ*   *so-a-t*   *õt*
          deer see-THEM-PST 1SG
          'I saw the deer.'

Notice how in (10), P is never expressed as a personal pronoun, while A and S can be expressed by the same set of pronouns (10a-10b). Since S and A are treated identically in terms of case marking, with P being treated differently, i.e. not expressed as a pronoun at all (only an index), Mekens is considered to show accusative case alignment in pronouns even though it has neutral alignment in the case marking of NPs.

To summarize, the main distinction between pronouns and indexation is the whether the argument marker occurs attached to the predicate or as a free word that can substitute for an NP argument. Alignment in both pronouns and indexation can be construed as a set of arguments that are treated identically for that particular argument marker, whether it is a set of verbal argument markers (indexes) or a set pronoun. When a particular argument cannot be

marked with a particular verbal argument marker set or pronoun set, it is considered to be treated differently. It is through the identical and different treatment of arguments by a particular argument marker, together with which arguments are marked with that argument marker, that it is possible to describe the argument marking pattern of a language for a particular argument marker. The way that an argument marking pattern can be construed as a number of different structural features and represented discretely in a typological database for further analysis is the primary focus of this investigation.

2.1.5 Applying the comparative concepts: a summary

To summarize, grammatical relations are identified through the morphosyntactic treatment of argument roles by different argument selectors in basic constructions. The method by which a basic construction is identified can be represented as a multistage process, as shown in (11). The transitive construction is used as an example since it is the most widely discussed in the descriptive literature and is often used as a basic of comparison for discussions of alignment with both intransitive and ditransitive clauses.

(11) Summary of the method for the identification of basic constructions, argument roles and grammatical relations in transitive clauses

   a. Identify the class of bivalent predicates in a language that requires a volitional performer (agent) and an affected entity (patient) as its arguments.

   b. Within this class of predicates, identify the transitive situation type, controlling both for the semantic properties of the event (2) and the semantic properties of the arguments (3).

   c. Identify the transitive construction based on the morphosyntactic treatment of the arguments in the transitive situation type. Other situation types whose predicates treat their arguments identically to the treatment of arguments in the transitive situation type are also considered as forming a transitive construction.

   d. Identify the argument role A by the morphosyntactic treatment of the agent in a transitive construction. The argument role P is identified through the treatment of the patient in a transitive construction.

   e. Grammatical relations are identified by the the identical treatment of argument roles by a specific argument selector, e.g. a case marker or a verbal argument marker. Transitive argument roles are often discussed in relation to the treatment of the sole argument of an intransitive clause S. Grammatical relations in ditransitive clauses are most often discussed in relation to the treatment of P in transitive clauses.

2.2 Structural questionnaire

In order to make any meaningful statement about observable grammatical patterns in a set of languages, there must be an empirical basis for that statement, i.e. there must be available data to base that statement on. How data is analyzed and encoded for future analysis is of utmost concern for the rigorous scientific investigation of language structure. In the previous section, a framework was introduced to establish a baseline of comparison across structurally different languages using a few key semantic notions such as agent, patient, and transitive situation type, which then allows for generalizations to be made regarding how these concepts are treated in morphosyntactic constructions. This section discusses how these generalizations are encoded into a structural questionnaire to form a database for investigation and analysis.

2.2.1 Databases in typological research

Typological databases aim to discretely encode the variation of a linguistic property (or properties) across a sample of languages. A few basic considerations must be taken into account when designing a typological database. The most important consideration is to ensure that the data are treated in a systematic and transparent manner in order to ensure their comparability. This requires that the areas of investigation be defined explicitly and that the same criteria are applied equally to all of the languages under investigation. The degree of detail to be encoded in the database is also an important consideration. At times the available descriptive literature may constrain the level of detail available to the typologist. For the current study, it was assumed that the level of description for each language in the sample was that of a grammar sketch of the scope and breadth that is expected of a doctoral dissertation. This constrains the available methodological tools available for the analysis of these languages since not every language in the sample will have a large corpus of texts or an extensive dictionary available. The database should also be designed in a modular fashion such that specific domains of the language can be explored independently of each other. This helps to ensure not only that the database can be easily used by others, but also that the database can be more easily integrated into or compared with other typological databases.

Typological databases can be organized in a variety of different manners. The most widely known typological database, *The World Atlas of Language Structures Online* (WALS; Dryer and Haspelmath 2011), defines a linguistic feature such as "alignment of verbal person marking" (Siewierska, 2011a) or "inclusive/exclusive distinction in verbal inflection" (Cysouw, 2011), and then gives a number of possible values for the expression of this feature. Another prominent typological database, the AUTOTYP database (Bickel and Nichols, 2002), provides a more flexible architecture that allows for the encoding of not just generalizations about a specific construction, but also a more detailed treatment of specific markers used within a construction as well as the definition

of language-specific constructions. The database used in Dunn et al. (2008) is somewhat different from the previous databases in two crucial ways: it was designed for use in historical linguistics for a specific region of the world rather than to make an global inventory of structural features, thus it focused on features that the authors judged as being the most historically relevant for their purposes; also, the database is designed as a questionnaire with a binary yes/no response. Each of these three databases provided inspiration for the development of the database used in this thesis.

2.2.2 Basic organization of the structural questionnaire

The database used here is composed of three levels of organization: domains, features and values. The database is organized as a questionnaire, and each question is assigned a two or three number identifier label.

Domains: The first level of organization used in the questionnaire is the domain. The domains considered are: (Q1) constituent order, (Q2) verbal argument marking, (Q3) inversion, (Q4) split intransitivity, (Q5) case marking, (Q6) applicatives, (Q7) causatives, (Q8) passives, (Q9) antipassives, (Q10) reflexives, (Q11) reciprocals and anticausatives (Q12). Each domain is represented in the first number of the identifier label of each question, such that any question that begins with Q1 is about constituent order or Q2 is about verbal argument marking. Each domain is first defined according to cross-linguistically applicable criteria, primarily based on functional or semantic properties, often similarly to the definitions used in the WALS database. The precise definition for each domain is given in their respective chapters, with the domains (Q3-Q4) being treated together in Chapter 4 on complex marking patterns and domains (Q6-Q12) being treated together in Chapter 7 on valency change.

Features: Once the domains under consideration are delineated and defined, a number of different features of within these domains can be examined. The entries (questions) can be either formulated as independent questions or dependent questions. Independent questions do not require a value in any other entries to provide a possible answer. Dependent questions require that a positive value is given for its head question in order for it to be applicable. Independent questions are expressed in the questionnaire with two number identifier labels, such as *Q2.1 Person can be marked for S:*. Dependent questions are expressed with a three number identifier with the second number in the series reflecting its head question, such as *Q2.1.1 Person marking for S when conominal is expressed in clause is:*. It would be impossible to diagnose where the behavior of the person marker for S (Q2.1.1) is there is no person marking for S, thus Q2.1.1 is dependent on Q2.1 As can be seen in these example questions, each entry is expressed in the form of a statement about the feature under investigation.

Values: Each question takes either a binary value [1=true, 0=false] or a multi-state value [a,b,c] as an answer. An additional value 'not applicable' [NA] is used under two circumstances: in independent questions when the language does not display the feature under consideration (e.g. there is no grammatical gender category, as in Q2.6), and in dependent questions when the head question has a negative value (e.g. one cannot talk about the treatment of the demoted A argument in a passive construction if there is no passive construction). A number of different values are delineated for the possible expression of a feature. These values were initially chosen based on previous knowledge and experience, but values were also be added to the initial list of possibilities when the language data under investigation did not conform to the initial values (cf. Bickel and Nichols, 2002). Flexibility in the list of possible values for a certain property allows for the database to more accurately represent the observed data, but it is also necessary to ensure that the final list of values are applied evenly across the sample. In other words, as new values were added, previous codings were reexamined for whether they are more accurately represented using the new value. This coding and re-coding process was initially quite tedious, but after examining a few dozen languages, less than half of the sample in this study, the final list of values tended to emerge from the data. When certain values from the list were unattested in the sample, they were removed from the list, leaving only the relevant values.

2.2.3   Coding example

To illustrate how a language is coded in the questionnaire, first consider the features related to the marking of the S argument within the domain of verbal argument marking:

Q2.1   *Person can be marked for S: [a=prefix, b=suffix, c=variable, d=unmarked]*

Q2.1.1 *Person marking for S when conominal is expressed in clause is: [a=obligatory, b=variable, c=unmarked]*

Q2.2   *Number can be marked for S: [a=independent prefix, b=independent suffix, c=fused w/person, d=unmarked]*

Q2.3   *Gender can be marked for S: [a=independent prefix, b=independent suffix, c=fused w/person, d=unmarked]*

Now, examine the examples given in (12) of intransitive clauses in Aymara, an Aymaran language spoken in Boliva and neighboring countries.

Aymara (Aymaran; Cerrón-Palomino and Carvajal Carvajal, 2009, 195, 206)

(12)   a. *achachila-xa jiwa-:-**n**-wa*
           elder-TOP   die-PST-3-EVI

           'The grandfather died.'

    b. *anata-**pxa**-:-**na***
           play-PL-PST-3

           'They played.'

As can be seen in (12), Aymara indexes the person of the S argument with the suffix *-n(a)*, thus for Q2.1 it is coded with the value [b]. It is obligatory to mark the person of S in Aymara regardless of whether the conominal is realized in the clause (12a) or not (12b), thus for Q2.1.1 it is assigned the value [a]. Furthermore, by comparing (12a) and (12b) it is clear that Aymara uses a distinct suffix *-pxa* to mark that the S argument in (12b) is plural in number, thus for Q2.2 it is given the value [b]. Gender is not a grammatical category of nouns in Aymara, thus for Q2.3 it is given the value [NA] 'not applicable'.

Now let us consider the coding of case marking in languages that show variation in case marking across different main clause constructions. The features related to case marking in intransitive and transitive constructions are as follows:

Q5.1   *S is case marked*: [1, 0]

Q5.2   *A is case marked*: [1 ,0]

Q5.3   *P is case marked*: [1 ,0]

Q5.6   *Nouns show the following alignment pattern*: [accusative, ergative, tripartite, neutral]

Let us consider case marking in Timbira, a Jêan language of Brazil. There are three main clause constructions in Timbira: what Castro Alves (2010) descriptively labels the 'agent-patient construction', the 'ergative-absolutive construction' and the 'nominative-absolutive' construction. Let us first examine the transitive marking pattern for each of these constructions, shown in (13) following the order of the constructions as listed above.

Timbira (Macro-Jêan; Castro Alves, 2004, 70, 98; Castro Alves, 2010, 452)

(13)   a. *wa a-pupu*
           1   2-see

           'I see you.'

b. *kahãj* **tɛ** *iʔ-kra pupun*
woman ERG 3-son   see.NFNT
'The woman saw his son.'

c. *kahãj   a-pən   narɛ*
woman 2-carry.NFNT NEG
'The woman doesn't carry you.'

In all three construction types shown in (13), the P argument is expressed on the verb with a prefix. The agent-patient construction shown in (13a) expresses events with an imperfective aspect, such as in present, future or the remote past tenses.[10] The ergative-absolutive construction shown in (13b) expresses events that are in the recent past, which is also refered to in Castro Alves (2004, 68) as the 'past perfective'. Out of the three constructions displayed in the examples, only this construction shows that A participant *kahãj* 'woman' marked by the ergative case marker *tɛ*. The nominative-absolutive construction shown in (13c) is used when an additional clause-final aspectual or modal operator is used, such as the negator *narɛ*. In the construction agent-patient construction, the etymologically finite verb form is used, while in the ergative-absolutive and nominative-absolutive constructions the etymologically non-finite verb form is used. In none of the constructions is the S argument marked with a case marker (Castro Alves, 2010). Of the three constructions shown in (13), it is the ergative-absolutive construction in (13b) that is considered the basic construction for coding into the structural questionnaire because it fulfills the event properties of a transitive situation type as listed in (2), specifically it displays perfectivity. Thus, for Q5.2 Timbira is assigned the value [1] since it shows case marking on the A argument in the basic construction. Since neither the S argument (Q5.1) nor the P argument (Q5.3) are marked for case, Timbira is assigned the value [ergative] for the alignment pattern shown in nouns given in Q5.6.

For each linguistic feature treated for the different domains covered in the questionnaire, specific criteria are given in their respective chapters for how these features were coded. The full structural questionnaire is given in Appendix B, with the values coded for each language given in Appendix C.

---

[10]The agent-patient construction is the only of the three main clause constructions to shown a marking distinction in indexation between active and descriptive intransitive verbs. See Castro Alves (2004).

Verbal argument marking: Simple patterns

This chapter examines the various strategies used by the languages in the sample to index arguments in the main clause basic constructions. Section 3.1 defines the domain of verbal argument marking and introduces a number of useful concepts and terms related to this phenomenon. Then, section 3.2 presents the different parameters in which verbal argument marking can vary from language to language, and looks at the strategies used to express different grammatical features such as person, number and gender within verbal argument marking. The final section explores the different major verbal argument marking patterns observed in the sample based on the presence, locus and alignment of bound person forms.

3.1 Definitions and important concepts

As discussed briefly in Chapter 2, argument marking refers to processes by which languages express the obligatory participants of an utterance, as well as the strategies used to distinguish the semantic and grammatical roles that such participants hold in the clause, whether on the arguments themselves (case marking) or on the predicate (verbal marking). INDEXATION is a more specific term, adopted from Lazard (1998), which refers to the process by which grammatical features of participants of the clause, such as person, number and gender, occur on the predicate by means of bound argument markers. Indexation is thus analogous to case marking, further discussed in Chapter 6, but whose morphological realization is restricted to inflection of the predicate. As a label for this process, 'indexation' is preferred here over other commonly used

terms, such as 'cross-reference' or 'agreement', as to avoid the terminological confusion associated with the latter terms and to treat the process of verbal argument marking as a phenomenon *sui generis* (Haspelmath, 2013). The terminological confusion regarding the use of 'agreement', 'cross-reference' and 'bound pronouns' is further discussed in section 3.3.6. Throughout this chapter, the term 'verbal argument marker', or at times just 'marker', is used to refer to the morphemes used to index a particular argument.

The discussion of verbal argument markers presented here is centered around two often-used descriptive tools that serve as useful comparative concepts: marker sets and marker slots. A MARKER SET is an inflectional paradigm of verbal argument markers that consistently indexes a specific grammatical relation in a particular construction. A MARKER SLOT is the position in which a particular marker set can occur (see section 3.3.1).

In some languages, the verbal argument marker is bound to the verb through affixes, such as the subject suffix *-y* and the object prefix *ma-* in Yurakaré shown in (1):

Yurakaré (isolate; van Gijn, 2006, 186)

(14) **ma**-*bëbë*-**y**   *petche=w*
      3PL-search-1SG fish=PL

      'I searched for fishes.'

In other languages, the verbal argument markers are not phonologically bound to the verb (or other predicating element) but are still considered to be part of the predicate. For example, the Wari' language of Brazil indexes the nominative argument on the predicate through what Everett and Kern (1997) have labeled the 'verbal inflectional clitic' (VIC), as shown in (15):

Wari' (Chapacuran; Everett and Kern, 1997, 149)[1]

(15) *'on=**nana***   *hwijima' ma'*
      whistle=3PL:RP/P children   that:PROX:HEARER

      'The children whistled.'

This clitic is an obligatory element of the clause and always occurs immediately following the verb and its modifiers. However, this clitic is not phonologically bound to the predicate, i.e. it does not receive stress like suffixes in the language, nor does it participate in the vowel harmony system. This mismatch between the traditional 'phonologically bound' definition of clitics (cf. Zwicky, 1985) and the various uses of the term 'clitic' in the writing of descriptive grammars highlights the need to exclude phonological boundedness when considering the status of a morpheme as a verbal argument marker, as discussed in section 2.1.4 of the previous chapter. The property of syntactic boundedness is adopted as

---

[1]The VIC is not marked with a clitic boundary ⟨=⟩ in the original text following the orthographic conventions developed by the New Tribes Mission and adopted by the Wari'.

the primary criterion to distinguish between pronouns and verbal argument markers in this thesis.

The paradigmatic structure of a set of verbal argument markers entails that it must be a closed set of elements that occur in complementary distribution with each other (Cysouw, 2003, 8). For the coding in the structural question-naire, a maximum of one discrepancy is allowed between forms for different argument roles within a paradigm. An example of a single discrepancy within a marker paradigm can be seen for Wayoró, a Tupian language of Brazil, in Table 3.1 based on data in Nogueira (2011).

PERSON	S	P
1SG	m-/o-	m-/o-
2SG	e-	e-
3	te-	ɲ-
1PL.INCL	tʃi-	tʃi-
1PL.EXCL	ote-	ote-
2PL	ɲdʒat-	ɲdʒat-

Table 3.1: Verbal argument markers in Wayoró (Tupian)

As can be seen in Table 3.1 the two sets of markers are identical for all person/number combinations except for the 3rd person.[2] Wayoró is considered to have only a single paradigm of verbal argument markers that index the absolutive argument.

To summarize, a verbal argument marker is the morphological realization of indexation, and as such, must fulfill three criteria: (i) the marker must morpho-logically distinguish between grammatical features of the indexed argument, (ii) the set of available markers must be paradigmatic in structure, and (iii) the marker must be syntactically bound to the predicate. The categories of differ-ent grammatical features such as person, number and gender, and the various distinctions observed within them are explored in the following sections.

3.2 Features and parameters

It is useful to distinguish between the grammatical features that are realized through indexation, such as person, number and gender, and the parameters through which these features are realized. The languages in the sample have been evaluated for a number of parameters relating to the way that arguments are indexed on the verb, including:

---

[2]The 3rd person marker ɲ- in Wayoró has different allomorphs for 3rd person singular *ndeke-* and 3rd person plural *ndeat-* (Nogueira, 2011, 78-81). These forms are identical to the independent pronoun forms *ndeke* and *ndeat*, respectively, and can only be distinguished by position rather than syntactic boundedness. They are not considered verbal argument markers in this study.

1. *Presence of markers*: which arguments are indexed

2. *Position of markers*: whether argument markers occur as a prefix or a suffix, and if two markers can be present, how the markers occur with relation to each other

3. *Alignment of markers*: whether the same set of markers can be used to index different argument roles across transitive and intransitive clauses

4. *Fusion of arguments*: whether multiple arguments within the same clause are indexed through a single marker

5. *Fusion of features*: whether person, number and gender are conflated into a single morpheme or marked separately

6. *Optionality of markers*: whether the occurrence of the argument marker is obligatory, variable or never present when the corresponding nominal argument is realized in the clause.

When taken together, these parameters describe a number of the key aspects of the verbal argument marking system in a language. The ways in which these various parameters are realized for the arguments indexed in intransitive and transitive clauses form the basis for describing the verbal argument marking pattern of that language.

Before turning to the various observed marking patterns in the data (see section 3.3), let us first examine the marking of the relevant grammatical features: person (section 3.2.1), number (section 3.2.2) and gender (section 3.2.3). Section 3.3 begins by examining parameters 1-3, which together form the key aspects of a simple verb marking pattern, with parameters 4-6 discussed in sections 3.3.4-3.3.6. Certain conditions that affect the realization of verbal argument markers in basic constructions, namely split intransitivity and hierarchical marking are discussed in Chapter 4 on complex verbal marking patterns.

3.2.1   Person marking

Q2.1   *Person can be marked for S:* [prefix, suffix, variable, not marked]

Q2.4   *Person can be marked for A:* [prefix, suffix, variable, not marked]

Q2.7   *Person can be marked for P:* [prefix, suffix, variable, not marked]

Q2.13   *First person inclusive is a distinct morphological category in verbal argument markers:* [1, 0]

The most often discussed component of verbal argument marking is the indexation of grammatical person in the clause. Central to the identification

of person marking is the presence of a morphological distinction between different grammatical categories of person. A person system is generally organized around three salient cognitive categories: the speaker ('I'), the addressee ('you'), and the referents that are not part of the speech act ('he/she/it/they') (see Siewierska, 2004; Cysouw, 2003). These categories are often referred to as the first, second and third persons, and can be represented as 1, 2 and 3, respectively.

The presence of a full person-based verbal argument marking system of at least one argument role in basic main clause constructions occurs in 66 of the 74 languages in the sample, with 42 of these languages displaying only indexation of arguments and no core case marking (see the maps in Appendix A). This makes person marking especially important for a discussion of argument marking strategies in South American languages. The presence of verbal markers for the different argument roles, and how these markers align, is subject to considerable variation across the different language families and regions of the continent (sections 3.3.2 - 3.3.4). Languages also vary as to how person marking is morphologically conflated with other grammatical features such as number and gender (section 3.3.5)

Distinctions within first person forms: In addition to the three person categories just introduced, person marking systems can make a number of further distinctions. The ability to distinguish whether first person pronominal forms include or exclude the addressee in a speech act (second person) is a commonly found feature in the languages of the world. In fact, the earliest attested description of such a distinction comes from a 16th century grammar of Quechua written by Domingo Santo Tomás (1560). Cysouw (2003, 80-90) provides a useful typology for examining the various formulations of the first person category in person marking paradigms based on the number and types of morphological distinction made within this category:

(16) Types of first person distinctions per Cysouw (2003)

    a. *'Unified-we'*: languages with morphologically distinct forms for only 'I' and 'we'.

    b. *'No-we'*: languages with no morphologically distinct form for 'we'.

    c. *'Inclusive-only'*: languages with only an inclusive 'we' as a morphologically distinct form. The exclusive form is often expressed with the first person singular form (sometimes with additional number marking).

    d. *'Inclusive/exclusive'*: languages with morphologically distinct forms for inclusive and exclusive 'we'.

    e. *'Minimal/augmented'*: languages with morphologically distinct forms for 'minimal inclusive', 'augmented inclusive' and 'exclusive'.

The first type, UNIFIED-WE, commonly occurs in the sample, and is indeed the most commonly occurring pattern in the world's languages (Cysouw, 2011).

In the lowlands, this pattern is observed in languages of the Arawakan, Nambikwaran, Matacoan, Guaycuruan and Tucanoan language families as well as in a number of isolates and members of other smaller families. In the Northern Andes, the Chocoan languages and the isolate Nasa Yuwe also show a UNIFIED-WE, as well as Imbabura Quechua (while most other Quechuan languages display an inclusive distinction). Muysken (1977) attributes the loss of the inclusive distinction in Ecuadorian Quechua varieties (to which Imbabura belongs) to paradigm simplification in a scenario of pidginization, but substrate effects from neighboring languages in a scenario of language shift could also have played a role in this development. The future tense paradigm of the subject marking suffixes in Imbabura Quechua are shown in Table (3.2) based on data in Cole (1982, 145):

PERSON	SINGULAR	PLURAL
1	-sha	-shun
2	-ngui	-nguichi
3	-nga	-nga

Table 3.2: Future tense subject markers in Imbabura Quechua

The second type of first person plural distinction, NO-WE, occurs in the person marking system of the Barbacoan language Awa Pit, where first person singular as well as first person plural forms are all indexed with -w or -s in the past tense and -s in the non-past tense in declarative main clauses (Curnow, 1997).[3] Awa Pit is the only example of a language that shows a no-we formulation of first person verbal marking in the sample.

Awa Pit (Barbacoan; Curnow 1997, 271)

(17) a. *ayna-tpa=na   au=na   kwa-ta-**w***
         cook-after=TOP 1PL=TOP eat-PST-LOCUT:SUBJ

      'After cooking, we ate.'

   b. *camisa pat-tawa=na   piikam-ta-**w***
         shirt   wash-after=TOP swim-PST-LOCUT:SUBJ

      'After washing (my) shirt, I went for a swim.'

The third type, INCLUSIVE-ONLY, is often found in the Quechuan and Aymaran families in the highlands and some languages from the Cariban and

---

[3]The 1st person subject marker -w is glossed in (17) as 'Locutor Subject' rather than simply 1st person S/A due to the mismatch between the hierarchical configurations functioning in indicative and interrogative clauses. For further discussion on person marking in Awa Pit, see section 4.1.4.

Macro-Jêan families in the lowlands.[4] In the Andean tradition, the first person inclusive form is sometimes referred to as the 'fourth person'. An example of this system can be seen for Timbira, a Macro-Jêan language of Brazil, in Table 3.3 based on data in Castro Alves (2004):

PERSON	SINGULAR	PLURAL
1	i-	mɛ̃=i-
2	a-	mɛ̃=a-
3	i(ʔ)- / h-	mɛ̃=i(ʔ)- / mɛ̃=h-
1DU.INCL	pa(ʔ)-	mɛ̃=pa(ʔ)-

Table 3.3: Inclusive-only person system in Timbira verbal argument markers

As can be seen in Table (3.3), the marker *mɛ̃=* attaches to the base person forms to produce the pluralized (non-dual) forms. The pluralized first person form *mɛ̃=i-* expresses the category of first person plural exclusive. The forms *pa(ʔ)-* and *mɛ̃=pa(ʔ)-* correspond to the categories 'first person dual inclusive' and 'first person plural inclusive', or rather, 'me and you' and 'me, you and someone else', respectively. Thus, the positing of a 'fourth person' category in Timbira allows for the simplification of the paradigm into two morphologically distinct sets (base forms and pluralized forms), thereby avoiding the insertion of a dual number category that is only relevant to the first person.

The fourth type, INCLUSIVE-EXCLUSIVE, is commonly found in the Tupian languages as well as a number of languages from smaller families such as the Chapacuran language Wari'. The pattern can be seen in the paradigm of subject markers in Kwaza, an isolate spoken in Brazil, as shown in Table (3.4) based on data in van der Voort (2004, 235):

PERSON	SINGULAR	PLURAL	DISTINCTION
1	-da	-a	inclusive
		-axa	exclusive
2	-xa	-xa(xa)	
3	-	-	

Table 3.4: Kwaza subject suffixes

It is important to note that the first person plural inclusive form *-a* and exclusive form *-axa* in Kwaza are both distinct from the first person singular form *-da*.

---

[4]Note that some languages may show different formulations of first person plural distinctions in verbal argument marker sets and free pronoun sets. For example, Tiriyó (Cariban) displays an inclusive-only type system in indexation but an inclusive-exclusive system in free pronouns (Meira, 1999). The types assigned to different languages in this section refer specifically to the distinctions made in the verbal argument marker sets.

The fifth type, MINIMAL-AUGMENTED, is not clearly attested in the sample. However, the system presented in Table (3.3) for Timbira is reminiscent of the minimal-augmented system reported in the Ilocano language of the Phillipines since the first person dual inclusive is a distinct morphological category within the paradigm (cf. Conklin, 1962; Greenberg, 1988; McGregor, 1989; Cysouw, 2003 for further discussion). However, since the distinction between the first person dual inclusive and first person plural inclusive is produced by the inclusion of an optional plural marker (which can also occur on nouns), rather than through its own distinct morpheme, Timbira is considered to be of the inclusive-only type (cf. Cysouw, 2003, 71, 256 for his analysis of the closely related language Northern Jê language Canela-Kraho).[5]

Q2.8 in the structural questionnaire only codes for the presence of a morphological distinction between first person plural inclusive and other possible first person categories. Thus, languages of the unified-we and the no-we type were coded as lacking this feature, with all remaining types coded as displaying it since they have a morphologically distinct first person inclusive form.

Partial systems of person marking: The presence of unmarked arguments in paradigms, especially for third person arguments, is widely attested within verbal argument marking paradigms cross-linguistically (Siewierska, 2011c). Additionally, languages often conflate the morphological form of different person categories, resulting in what Cysouw (2003) calls 'homophony' within a paradigm. Languages that show either of these patterns individually are considered to display person marking as long as there remains two morphologically realized person distinctions within the paradigm. However, languages that do not have at least two morphologically realized person distinctions and only index a single (homophonous) category for a particular argument show only partial person marking. This can be seen in a language like Imbabura Quechua, where only 1st person singular objects are indexed on the verb through the suffix -wa and all other persons are unmarked (Cole, 1982, 129). As such, Imbabura Quechua is not considered to display verbal argument marking of P in this study, even though other varieties of Quechua do have more morphologically rich systems of P marking on the verb (cf. Weber, 1989 for Huallaga Quechua). Another example of this can be seen in Trumai, an isolate spoken in Brazil, where only 3rd person absolutive arguments (no number distinction) are indexed on the verb through the enclitic =e/=n with all other persons unmarked (Guirardello, 1999, 116). While both of these languages indeed index a single specific argument on the verb, they do not display a full system of person marking since there are not two morphologically distinct persons indexed on

---

[5]The Urarina language also shows a system with a morphologically distinct first person inclusive dual (Olawsky, 2006). Like Timbira, segmentable number markers are used to distinguish first person dual inclusive and first person plural inclusive forms, thus diverging from the minimal-augmented prototype. However, the number markers vary across the different person forms, and the first person plural exclusive form does not show segmentable number marking, i.e. is its own distinct morpheme. See Table (3.6).

the verb. Only arguments that are indexed with full systems of verbal person marking are included in the discussion on verbal argument marking patterns insSection 3.3 and treated as displaying verbal argument marking for those argument roles in the structural questionnaire.

3.2.2 Number marking

Q2.2 *Number can be marked for S:* [independent prefix, independent suffix, fused with person, not marked]

Q2.5 *Number can be marked for A:* [independent prefix, independent suffix, fused with person, not marked]

Q2.8 *Number can be marked for P:* [independent prefix, independent suffix, fused with person, not marked]

Number marking is the morphological realization of the grammatical category of number. Number, as a feature restricted to the nominal domain, encodes the quantification of entities expressed by nominal elements (Kibort and Corbett, 2008). The most common distinction between categories of grammatical number is between singular and plural (non-singular). Corbett (2000, 20) considers the singular-plural opposition in agreement systems to be most basic, with this distinction forming the basis for all other number systems.[6] Nominal number relates to entities within a clause, while verbal number relates to the semantics of the event (Corbett, 2000, 243).

Number marking often distinguishes only singular and plural, but some languages morphologically encode additional categories such as 'dual' or 'paucal'. Ninam, a Yanomaman language also known as Yanam, is one of the few languages in the sample that contain a dual distinction in its verbal argument marking paradigms, as seen in Table 3.5 based on data in Goodwin Gómez (1990) and Ferreira (2012).

| | Set 1 (S/A/P) | | | Set 2 (S/P) | | |
	Singular	Dual	Plural	Singular	Dual	Plural
1	ʧa=	ʧehe=	ʧama=	-	k=	k=
2	wa=	wehe=	wama=	-	k=	k=
3	-	-	-	-	kɨp=	pɨk=

Table 3.5: Verbal argument markers in Ninam

As can be seen in Table 3.5 above, Ninam displays a singular-dual-plural opposition in the Set 1 paradigm across all marked person distinctions. Set 2

---

[6]It is important to note that the grammatical feature under investigation is *nominal number* and not *verbal number*.

indexes all non-singular absolute SAP arguments with $k=$, while still maintaining the dual-plural opposition in 3rd person marking.

It is also common for languages to conflate person and number categories within the same set of argument markers in such a way that there is no segmentable form that only refers to the grammatical number of the argument. This is also evident in the argument marker paradigms for Ninam shown in Table 3.5.

In some languages, the category of number is indexed through morphology segmentable and distinct from person marking, as shown in the verbal argument marking paradigm for Urarina, an isolate spoken in Peru, in Table 3.6 based on data in Olawsky (2006).

	Set 1 (S/A)	Set 2 (S/A)
1sg	-anɨ	-ɨ̃
2sg	-i	-i
3sg	-a	-a
1du	-aka	-aka
1pl.incl	-aka-**tɕe**	-aka-**tɕe**
1pl.excl	-akaanɨ	-akaanɨ
2pl	-i-**tɕe**	-i-**tɕe**
3pl	**-ɨrɨ**-a	**-ɨr**-e

Table 3.6: Verbal argument markers in Urarina affirmative main clauses

Set 1 markers in Urarina (what Olawsky calls 'A-forms') are used in pragmatically neutral clauses, while the Set 2 markers ('E-forms') are used in clauses where a focus marker precedes the verb, or if emphasis or focus is implied. Notice that the morpheme *-tɕe* is segmentable within the first person plural inclusive and second person plural values in the paradigm, while the morpheme *-ɨr(ɨ)* is segmentable from a general third person morpheme *-a/-e* in the third person plural value in the paradigm.

Some languages solely index number properties of their core arguments without any person distinctions. This can be seen in the Northern Embera, a Chocan language of Colombia, in (18), where plural subjects are indexed with the suffix *-da*:

Northern Embera (Chocoan; Mortensen, 1999, 17, 62)

(18)   a. *iči-ra   khã̃ĩ-shi-a*
            3S-abs sleep-pst-decl
            'He slept.'

        b. *tai-ra   mẽã   wã-shi-**da**-a*
            1pl-abs jungle go-pst-pl-decl
            'We went into the jungle.'

c. *hã̃ʉ̃ čapa-ra   wã-pi-shi-**da**-a*
    that brother-ABS go-CAUS-PST-PL-DECL
    'They made that guy leave.'

In the structural questionnaire, Q2.2, Q2.5 and Q2.8 encode the presence or absence of a morphological number distinction for the argument roles S, A and P. These questions also encode whether these markers are realized as forms that are morphologically distinct from the person markers used to index the corresponding arguments, or whether the expression of person and number features are fused into single marker.

3.2.3 Gender marking

Q2.3   *Gender can be marked for S:* [independent prefix, independent suffix, fused with person, not marked]

Q2.6   *Gender can be marked for A:* [independent prefix, independent suffix, fused with person, not marked]

Q2.9   *Gender can be marked for P:* [independent prefix, independent suffix, fused with person, not marked]

Q2.12   *Gender can only be marked in the singular.* [1, 0]

Just like person and number, gender is another grammatical feature of arguments that is frequently distinguished in indexation. Gender generally refers to categorization of the nouns in a language into different classes, and these class distinctions are "reflected in the the behavior of associated words"(Hockett, 1958, 231). As such, gender is often characterized as an agreement phenomenon between a 'controller', i.e. the noun itself, and a 'target', the element whose phonological form is conditioned by the class of the target (Corbett, 1991, 2006). An essential component of gender systems is that all nouns in the language are classified as belonging to a particular noun class.

In this thesis, the terms 'gender' and 'noun class' are treated as synonymous and can be used interchangeably, although certain authors try to restrict use of the term gender to a noun class system with only two or three classes (Dixon, 1986; Regúnaga, 2012). Languages vary with regard to the degree of semantic transparency in the assignment of nouns to different genders, but semantic features such as sex, animacy and humanness can often play a role in class composition (Aikhenvald, 2000). Furthermore, it is common practice in descriptive linguistics to label different genders using the categorization of different human sexes as a prototype. For example, the 'masculine' gender refers to the class of nouns in a language that trigger agreement in the same way (reflected in the phonological form of the controller) as human males, while the 'feminine'

gender refers to the noun class that includes human females and all nouns that trigger agreement in that same way. The 'neuter' gender is often used similarly for the noun class that does not pattern with either human males or human females. However, other labels such as animate or inanimate are also used when they more accurately reflect the semantic composition of class members.

Gender (or noun class) is often discussed in opposition to noun classifiers—morphemes that categorize nouns based on different semantic features such as animacy, shape, function or material (Grinevald, 2004, 1016). However, the essential difference between gender and noun classifiers is that classifiers do not trigger agreement (Seifart, 2005, 10). Noun classifiers in the context of South American languages is not explored in this thesis, but it is worth noting that South American languages, especially in the lowlands, display a wide variety of different types of nominal classifiers (Aikhenvald, 2012, Ch.10). However, since noun classifiers are sometimes marked on the verb, it is important to be able to distinguish these from gender marking. Grinevald (2000, 62) provides a number of useful criteria for distinguishing noun class (gender) from noun classifiers, as seen in Table (3.7).

	GENDER	CLASSIFIER
a.	Classify all nouns	Do not classify all nouns
b.	"Smallish" number of classes	"Largish" number of classes
c.	Of a closed system	Of an open system
d.	Fuse with other grammatical categories	Independent constituent
e.	Can be marked on noun	Not affixed to noun
f.	Realized in agreement patterns	Marked once
g.	No speaker variance to assignment	Speakers can vary at will

Table 3.7: Distinctions between genders and classifiers per Grinevald (2000)

As gender is an inherent feature of a noun, it can be reflected in many types of agreement relations between a noun and other constituents of the clause, such as in possessive constructions, adjectival modification or demonstratives.[7] However, the present discussion is restricted to the realization of gender marking within the domain of verbal argument marking and how the category of gender is expressed with regard to the other relevant grammatical features person and number.

Almost every language in the sample that has gender as a grammatical feature of nouns index these features for at least one argument in a transitive clause.[8] In the vast majority of cases where gender is indexed on the verb, the expression of gender marking is fused with person and number marking (60-77% depending on the argument role being indexed). This can be seen for

---

[7]See Krasnoukhova (2012) for a discussion of gender agreement in the nominal domain within a South American context using a sample comparable to the one used in this study. See also Regúnaga (2012) for a more in-depth survey of gender in South American languages.

[8]The notable exceptions are Yurakaré, Tehuelche and Rikbaktsa.

Lokono, an Arawakan language of Suriname, in (19):

Lokono (Arawakan; Pet, 2011, 13, 218)

(19) a. **ly**-*simaka bo*
          3SG.M-call 2SG
          'He called you.'

      b. **thy**-*simaka je*
          3SG.F-call 3PL
          'She called them.'

      c. **th**-*amon-bo*   *to*   *firobero*
          3SG.F-accompany-CONT ART:F tapir
          'It (the bird) accompanies the tapir.'

   The gender system in Lokono makes a distinction in the singular between masculine nouns (human males, small infants and animals/objects/spirits considered to be "good") and feminine nouns (human females, certain human males not of the speaker's ethnic group, all other animal/objects/spirits) in the indexation of third person arguments (Pet, 2011, 14). In (19a) the third person masculine subject is indexed with the prefix *ly-* while the third person feminine subjects in (19b) and (19c) are indexed with the prefix *th(y)-*.
   While it is most common for languages to index both person and gender features within a single marker set, there are a few attested examples in the sample where gender marking is indexed separately from person marking.[9] In his groundbreaking work on the classifier and noun class systems in Miraña, a Boran language of Colombia, Seifart (2005) provides an in-depth discussion of the distinctions made within nominal classification in the language. Gender (or noun class) in Miraña shows a high degree of semantic transparency and is organized into three distinct classes: masculine (animates), feminine (animates) and inanimate. These different genders are indexed on the verb by suffixes ("general class markers") that are in complementary distribution with classifiers ("specific class markers"), as shown in (20a) and (20b) respectively.[10] While nouns are able to be indexed on the verb by a number of different classifiers depending on their salient physical properties, a noun can only be indexed by a single general class marker (gender). The person of first and second person arguments is also indexed in basic main clauses through a series of proclitics, as seen in (20c):[11]

---

[9]A further example can be seen in gender marking through vowel alternations in tense-modal and mood suffixes in Jarawara (Arawan) as described is Dixon (2004, 100-102, 420).

[10]Seifart (2005, 136) notes that classifiers are rarely marked on the verb since the classifiers are generally restricted to inanimate entities (which tend not to be subjects). The one example identified in the texts of this work is given in (20b).

[11]The clitics are referred to in Seifart (2005, 52-53) as the series of "monosyllabic personal pronouns". Seifart (p.c.) affirms that these clitics can only occur syntactically bound to the verb.

Miraña (Boran; Seifart, 2005, 55, 66-67)

(20) a. *gwatá-**hɨ́**-gwɯɪ́ɯ́-dɯ́-rɛ*   *né:-**nɛ***
          cover.NMLZ-SCM.2D.round-DIM-COMP-REST seem-GCM.INAN
          'It looks just like a little lid.'

      b. *ɯ́mɛ́-ʔɛ-tɯɪ*   *gwáboʔhɯ́kɯɪnɯ́-**hɨ*** *ɯ́ɡwa:-hɨ*
          wood-SCM.tree-ABL stuck-SCM.2D.round metal-SCM.2D.round
          'The ax is stuck in the tree.'

      c. *mɯ́ʔtsɪ́*   ***mɛ**=tsá:-ʔi*
          1DL.EXCL.M 1PL=come-PRD
          'We (two, at least one male, exclusive) came.'

      In the examples shown in (20a) and (20b), the arguments *gwatáhɨ́gwɯɪ́ɯ́dɯ́rɛ* 'little lid' and *ɯ́ɡwa:hɨ* 'ax' can both be classified as two dimensional round objects and can be marked by the classifier *-hɨ*. This classifier occurs on the noun itself in (20a) and on the verb in (20b). However, both of these arguments are also members of the inanimate gender, and in (20a) the gender of the subject is indexed through the general class marker *-nɛ*. These examples serve to show that gender marking in Miraña is distinct from person marking.

3.3 Simple marking patterns

Languages that consistently index the same set of argument roles using the same set of argument markers across different classes of verbs in basic constructions display a simple marking pattern. Simple marking patterns are referred to by the number of morphological slots available for indexation and the grammatical relations that are indexed in these slots. The following discussion is restricted to the realization of person marking, together with any number or gender distinctions that are morphologically conflated with the person forms.

3.3.1 Position of marking

As mentioned in Chapter 1, the position of marking has often played a significant role in the categorization of South American language types at least since Lafone Quevedo (1896). This is not without good reason, since there is a significant geographic distribution between the use of prefixes and suffixes as verbal argument markers across the continent (see Chapter 8). The parameter of position of marking for S, A and P is encoded in the structural questionnaire by questions Q2.1, Q2.4 and Q2.7 already introduced in section 3.2.1.

      The position in which a verbal argument marker appears with relation to the main predicating element of the clause is called the position of marking. Verbal argument markers can attach before the predicate (prefixes), after the

predicate (suffixes) or can be composed of multiple parts that occur both before and after the predicate (circumfix). Languages can also vary as to position of marking within a single marker set. For reasons already explored in section 3.1, clitics and affixes are not distinguished as separate categories in this study, and as such, the use of the term 'prefix' in the context of position of marking can refer to either a prefix or a proclitic.[12]

Prefix sets: The most common position for the occurrence of verbal argument markers in the languages of the core sample is the prefix position across all argument roles S, A and P. Most Tupian and Cariban languages are good exemplars of languages with prefixing verbal argument markers. This can be seen in Juruna in (21) where the P argument is indexed by the proclitic *ese=*:

Juruna (Tupian; Fargetti, 2001, 147)

(21)   *ulu'udí*   ***ese****=dépú*
         1PL.EXCL 2PL=push
         'We (excl) pushed you (pl).'

Suffix sets: Verbal argument markers occurring in the suffix position are also commonly observed in South American languages, especially in many languages of the Andes and Western Amazon (see Chapter 8). This can be seen in (22) for Chipaya, an Uru-Chipayan language of Bolivia, where the subject is indexed by the suffix *-am*:

Chipaya (Uru-Chipayan; Cerrón-Palomino, 2009, 55)

(22)   *am-ki*   *waskir-kama oqh-ch-**am**-tra*
         2SG-TOP Escara-LIM   go-PST-2-DECL
         'You went until Escara.'

Circumfix sets: Verbal argument markers that occur as circumfixes are only observed in the languages of the Guaycuruan family within the sample. Grondona (1998) analyzes the verbal argument marking system in Mocoví as composed of two elements: a proclitic that indexes the person of the referential argument and a suffix that indexes primarily number, but also person, of the same referential argument. This can be seen for the intransitive subjects of the 'inactive' verb *awig* 'get.burned' in (23):

Mocoví (Guaycuruan; Grondona, 1998, 110)

(23)   a.   ***ir****=awig*
               1SG.II=get.burned

---

[12]However, an attempt has been made to maintain the morphological distinctions made by authors in the glosses of the examples presented.

'I got burned.'

b. **r=awig-i**
2.II=get.burned-2SG.FAM
'You got burned.'

c. **i=awig**
3.II-get.burned
'He got burned.'

d. **qar=awig**
1PL.II=get.burned
'We got burned.'

e. **r=awig-iri**
2.II=get.burned=2PAUC
'You (paucal) got burned.'

f. **r=awig-i:**
2.II=get.burned=2PL
'You (pl) got burned.'

g. **i=awig-er**
3.II=get.burned-3PL
'They got burned.'

While not all person/number combinations are marked with a circumfix in the Set II marker paradigm, the majority of instances in (23) show that verbal argument marking for a single argument is realized with both a proclitic and a suffix, with a few cases where indexation is realized solely through proclitics. However, for the Set I markers, not shown here, the alternation is between a proclitic, a suffix and a proclitic+suffix circumfix, as shown in Grondona (1998, 97-98).[13] Pilagá, the other Guaycuruan language in the sample, shows even fewer attestations of circumfixes in its subject marker sets (I/II), with all singular person distinctions indexed solely by a prefix and the plural forms by a prefix and a suffix (Vidal, 2001, 135-137). As such, there are no attested languages in the sample that show a strictly circumfix position in verbal argument marking for all person distinctions.

Mixed sets: While the Mocoví examples in (23) show the use of a marker set that requires both a prefix and a suffix simultaneously to index certain S arguments, languages can also have a marker set composed of both prefixes and suffixes that do not co-occur. This is shown in (24) using the subject marker

---

[13]An alternate analysis, the one adopted by Grondona (1998), is that each prefix or suffix component can have zero marking for specific person and/or number distinctions, thus all argument markers have both a proclitic and suffix component, even if one or both of those components are zero-marked.

set in Ika, a Chibchan language of Colombia:

Ika (Chibchan; Frank, 1985, 69)

(24)   a. *ʧua-na-**rua***
          see-DIST-1SG.I
          'I saw it.'

        b. *pablo-sin   gou-**kuära**   ni*
          pablo-COM make-1PL.EXCL.I CERT
          'Pablo and I (we) made it.'

        c. ***a**-ʧua-na*
          1PL.INCL.I-see-DIST
          'We saw it.'

        d. ***win**-ʧua-na*
          3PL.I-see-DIST
          'They saw it.'

Notice that in (24a) and (24b), the transitive subjects are marked with a suffix, while in (24c) and (24d) they are marked with a prefix. Since the subject marker set cannot be characterized as either prefixing, suffixing, or circumfixing, it is considered to be of the mixed set type. As there are no attested examples of languages with only circumfixes for verbal argument markers, the category [variable] in Q2.1, Q2.4, Q2.7 in the structural questionnaire is used to include both Mocoví-like variation and Ika-like variation in the position of person indexation.

3.3.2 Alignment of person markers

Q2.15   *Verbal person marker sets show the following alignment pattern:*
          [accusative, ergative, tripartite, neutral]

When discussing the alignment of verbal argument markers, it is important to distinguish between the argument roles that are marked, the different sets of markers available to mark each of them, and the overall alignment pattern of the marking. Siewierska (2003, 342) proposes four different factors that must be accounted for when diagnosing alignment patterns in verbal argument marking: (i) which arguments are and are not indexed, (ii) the phonological form of the markers, (iii) the location that argument markers occur relative to the verb stem and relative to each other (if more than one is present), and (iv) the conditions under which indexation occurs for each argument. For the purposes of this thesis, alignment of verbal argument marking is more narrowly defined as the alignment within a particular marker set, rather than a characterization of the all indexation occurring in the construction (which is considered here to

be part of the 'marking pattern'). The conditions that affect indexation are not explored in the current section. However, two especially relevant conditions on verbal argument marking in basic constructions, namely scenario (hierarchical marking) and intransitive predicate class (split intransitivity), are explored in Chapter 4.

Simple verbal argument marking patterns can be roughly divided into those displaying accusative, ergative, tripartite or neutral alignments based on the opposition of equivalence sets of argument roles that are indexed (and not indexed) by different sets of markers, e.g. languages that index S and A arguments in the same way as opposed to the indexation of P display an accusative alignment pattern. However, it is also important to distinguish between the marked and unmarked arguments of the predicate when referring to a particular argument marking pattern. This allows for a more specific characterization of the verbal argument marking pattern than simply stating that it is 'accusative' or 'ergative'.

Patterns within accusative alignment: Languages that index both the S and A arguments with the same set of markers show a nominative pattern (referred to as such since the nominative grammatical relation is the one marked). This pattern has also been called 'marked-nominative' (Dixon, 1994), and is considered a subtype of verbal argument marking patterns with accusative alignment. An example of strictly nominative verbal marking can be seen for Nasa Yuwe in (25):

Nasa Yuwe (isolate; Jung, 2008, 100)

(25) a. *nega-b*   *uʔx-we-ts-t^hu*
          Belalcázar-LOC go-IMP-PROG-DECL.1SG
          'I am going to Belalcázar.'

      b. *kutj ex-aʔs*   *vis-t^hu*
          corn field-OBJ.SG weed-DECL.1SG
          'I weed the cornfield.'

Languages that index both S and A arguments with the same set of markers and additionally index P on the verb with a separate set of markers in a different slot show a nominative-accusative pattern. This is the most widely attested pattern found in the sample, occurring in 30 of the 66 languages that display some form of person indexation of core arguments. This pattern can be seen for Munichi in (26):

Munichi (isolate; Gibson, 1996, 61-2)

(26) a. *miɲṣa tʃuʔ-maʔa matə-me-**ɲɯ***
          little   good-VERB improve-PFV-1SG
          'I have improved a bit.'

    b. *pjeɲca-**ɲɯ-pɯ***
       love-1SG-2SG
       'I love you.'

Languages that only index P arguments, with no indexation of S or A, show an accusative marking pattern. This pattern is only attested in two languages in the sample: Sabanê (Nambikwaran) and Juruna (Tupian).[14] An example of this verbal marking pattern can be seen for Sabanê in (27):

Sabanê (Nambikwaran; Antunes de Araujo, 2004, 173-174)

(27)   a. *wayulu **t**-ip-i-datinan*
         dog   1SG-see-VC-PRET.EV
         'The dog saw me.'

    b. *towali ay-i-datinan*
        1SG   go-VC-PRET.EV
        'I left.'

All of the various marking patterns seen in examples (25)-(27) show an accusative alignment pattern since they index the two arguments S and A in a similar fashion that is distinct from the indexation of P, regardless of which grammatical relations are and are not marked.

Patterns within ergative alignment: Languages that index both S and P arguments with the same set of markers show an absolute pattern. This pattern is attested in four languages across the sample, and it is frequently encountered in the non-Mawetí-Guaraní Tupian languages of Southern Amazonia, particularly those in upper basin of the Madeira River. This pattern can be seen for Mekens in (28).

Mekens (Tupian; Galucio, 2001, 78-79)

(28)   a. ***o**-er-a-t*
        1SG-sleep-THEME-PST
        'I slept.'

    b. *sete **o**-so-a-t*
        3SG 1SG-see-THEME-PST
        'He saw me.'

---

[14]Interestingly, both of these languages also display split intransitivity where a minor class of intransitie verbs can indeed be marked on the verb. For further examples and discussion, see section 4.2.3 for Sabanê and section 5.1.3 for Juruna.

Correspondingly, languages that index only the A argument, without in-dexation of S or P, show an ergative marking pattern. This pattern appears to be quite rare in South America. Only one language in the sample, Katukina-Kanamari spoken along the Biá River in Western Amazonia, displays such a pattern, as seen in (29):

Katukina-Kanamari (Katukinan; Queixalós, 2010, 238-239)

(29) a. *ki:tan idi:k*
          sleep   you
          'You slept.'

      b. ***no**-ti   paiko*
          2SG-kill grandfather
          'You killed grandfather.'

No languages in the sample present a further logical possibility for an ergative-aligned verbal marking patterns: ergative-absolutive. In Xavánte, a Macro-Jêan language of central Brazil, the absolutive argument is indexed on the verb in basic constructions for most person configurations (Estevam, 2011, 65-67). However, an ergative-absolutive indexation pattern does occurs in two restricted instances: when A is 2nd person in the basic construction (30c), or marked on the auxiliary *te* in imperative, negated or subordinate clauses (30d), as shown in (30):

Xavánte (Macro-Jêan; Estevam, 2011, 65, 210, 233, 424)

(30) a. *ãne wa **ĩ**-höjmana*
          well 1SG 1SG-live
          'I live well.'

      b. *te   za   **wa**-uprosi*
          2/3 PROSP 1PL-destroy
          'He will destroy us.'

      c. *te   **ĩ-wa**-ʔrujwapari zaʔra*
          2/3 2-1PL-detest   PL.DIS
          'You detest us.'

      d. *... dawedezé ñiti   **te**-te   **wa**-hözu mono wa*
              medicine far.from 3PL-AUX 1PL-prick ITR   SUB
          '... because they vaccinate us with expired medicine.'

Since A is only indexed for 2nd person in the basic construction, constitut-ing a partial system of person marking, or in certain non-basic constructions, Xavánte is considered to only show an absolutive verb marking pattern.

Patterns with tripartite alignment: Languages that index S, A and P with distinct marker sets for each argument role display tripartite alignment. A clear example of tripartite alignment in person indexation can be seen in (31) for Wichí, a Matacoan language spoken in the Argentinian Chaco and neighboring areas:

Wichí (Matacoan; Terraza, 2009, 125-126)

(31)   a. *hinu wex^w **ta**-k^{y}em-ɫi*
         man much 3.I-work-ITER
         'The man works a lot.'

      b. *xwan **i**-yahin-**nu***
         Juan 3.II-see-1
         'Juan sees me/us(excl.).'

As can be seen in (31), Wichí indexes S and A in a prefix slot using distinct marker sets. P is indexed with a suffix marker set (31b). As is discussed further in section 4.2.2, Wichí has an additional marker set that indexes stative verbs differently than the intransitive event verb shown in (31a).[15] The different marker sets in Wichí are given in Table 3.8 based on data in Terraza (2009, 124).

PERSON	A	S$_{state}$	S$_{event}$	P
1SG/1PL.EXCL	n-	n-	nt-	-nu
2SG	la-	a-	lata-	-am
3	i-	-	ta-	-
1PL.INCL	ya-	ya-	yat-	-nam

Table 3.8: Verbal argument marker sets in Wichí (Matacoan)

Patterns with neutral alignment: Languages that index S, A and P with the same marker set display a neutral alignment. A clear example of neutral alignment within a marker set can be seen in (32) for Puinave, a language isolate of Colombia:

Puinave (isolate; Girón 2008, 318, 320)

(32)   a. ***ka**-si-kút-ya*
         3PL-PROS-flee-FUT
         'They are going to escape.'

---

[15]Depending on whether one wishes to include minor classes of intransitive verbs in a discussion of alignment, Wichí could be described as showing quadripartite alignment since A, P, and two different classes of S are marked with distinct marker sets. A similar alignment pattern is reported for Bésɨro (also known as Chiquitano), a Macro-Jêan language spoken in Bolivia (Sans, 2012).

b. *káwa-bíkdik **ka**-sek-dí   **ka**-wõk-á*
   chicken-egg 3PL-steal-PST 3PL-eat-ALL
   'They stole a chicken egg to eat.'

c. *tín-da   **ka**-'úíʔɤ-m-pɤn-di*
   bat-ASR 3PL-kill-AGT-RES-PST
   'The bat killed them.'

d. ***ka**-ja-péw-at*
   3PL-3SG-load-COMPL
   'He loads them.'

As can be seen in (32, the 3rd person plural marker *ka-* can index the S argument (32a), the A argument (32b) or the P argument (32c-32d). While the marker set used to index the different arguments is identical, the relative position of the markers to the verb distinguishes A from P in transitive clauses if both arguments are marked, with P being marked furthest from the verb (32d). For an example of a language with neutral marker alignment and only a single (hierarchical) marker slot, see the discussion on Mapudungun in section 4.1.1 of the following chapter.

3.3.3   Indexation in ditransitive constructions

Q2.10   *Ditransitive verbs can mark R:* [1, 0]

Q2.10.1 *Ditransitive verbs can index both R and T simultaneously:* [1, 0]

A final pattern to be considered concerning the alignment of person markers is whether the R argument role can be indexed on the verb in ditransitive constructions. The strategies used to mark the semantic theme (T) or recipient in clauses that typically include three arguments, such as those with the verbs 'give' or 'send', show considerable variation across the worlds languages (Dryer, 1986; Malchukov et al., 2010). Languages that mark T the same as P in transitive clauses show indirective alignment, with T and P considered as direct objects, while languages that mark R the same as P in transitive clauses show secundative alignment, with R and P considered primary objects. An example of a language with indirective alignment (T marked like P) in verbal marking is Puinave in (33), where the T argument of the ditransitive verb *bɤ́k* 'give' is indexed on the verb (like P in (33b)) and the R argument is not verbally indexed but marked with the oblique case suffix *-at*:

Puinave (isolate; Girón, 2008, 320, 351)

(33)   a. *ót   **ja**-ka-búk-ni   padatá bɤ̂i-at*
          3PL 3SG-3PL-give-REC.PST money 3SG.INDEF-OBL
          'They gave the money to the other one.'

b. ***ja***-*bi-bɨ́k-at*
3SG-1PL-bury-CPLT
'We already buried him.'

A language with secundative alignment in ditransitives (R marked like P) can be seen in Aguaruna in (34), where R is indexed on the verb (34a), identically to P in (34b):

Aguaruna (Jivaroan; Overall, 2007, 320, 456)

(34) a. *mi-na*   *su-**hu**-sa-ta*
          1SG-OBJ give-1SG-ATT-IMP
          'Give it to me!'
      b. *au*   *tsupi-**hu**-ka-mĩ*
          DST cut-1SG-INTS-REC.PST:3:DECL
          'He cut me.'

The languages that allow for R to be indexed on the verb show an interesting geographic distribution (see Chapter 8 for further discussion). For example, all sampled languages of the Central Andes that index P in transitive clauses also allow R to be indexed in ditransitive clauses.

Some languages that can index R in ditransitive constructions also allow for T to be indexed simultaneously on the same verb. Due to a high occurrence of third person zero marking in verbal argument marking paradigms (cf. Siewierska, 2011c), this pattern is often difficult to observe since both the R and T arguments are rarely both non-3rd persons. However, an example of this pattern can be seen in the Baure, an Arawakan language of lowland Bolivia, in (35):

Baure (Arawakan; Danielsen, 2007, 177)

(35) *pi=pa=**ni**=**ro***
      2SG=give=1SG=3SG.M
      'You give it to me.'

Since P, T and R are all indexed with the same set of enclitics, it can be stated that Baure shows neutral alignment with regards to its marker set. However, Danielsen (2007, 177) notes that the indexation of R argument always precedes the T argument, i.e. it appears closer to the verb stem.

3.3.4   Fusion of arguments

Q2.11   *In transitive clauses that can index A and P, marking is*: [A before P, P before A, hierarchical, portmanteau]

Languages that display verbal marking for A and P conflated into a single portmanteau morpheme in transitive clauses show a fused pattern. An essential characteristic of this pattern is that it is not possible to segment distinct morphemes that refer to only one of the indexed argument types in the majority of cases across the set of argument markers. Languages from two families in the Central Andes, Quechuan and Aymaran, often display a degree of fusion in their verbal argument marker sets for transitive clauses. Both language families index arguments through suffixes attached to the verb, and in cases where these markers are segmentable, the P suffix precedes the A suffix. In these cases, the A markers show nominative alignment. There has been some debate within Quechuan and Aymaran linguistics as to whether the 'transitions', the verbal argument marking cluster, can be analyzed as a single fused morpheme or a complex of multiple markers (see Adelaar and Muysken, 2004, Ch.3). Since it is possible to segment a distinct marker for P and S/A in the majority of cases in Huallaga Quechua, it is considered to display a nominative-accusative pattern, following the analysis presented in (Weber, 1989, 96-97). The Aymaran languages Jaqaru and Aymara also show a degree of fusion across their person marking forms, but the majority of cases do not allow for the segmentation of distinct morphemes referring to each argument individually (see Hardman, 2001, 101 for Aymara). In the Jaqaru example in (36), notice the segmentable nature of 2sg.A *-ta*, 1sg.P *-u*, and 1pl.P *-ush* markers in the present tense verb paradigm, while markers for the other arguments cannot be consistently identified.

Jaqaru (Aymaran; Hardman, 2000, 57)

(36) a. *ill-k-**ima*** 'I see you'

b. *ill-k-**uta*** 'You see me'

c. *ill-k-**ushta*** 'You see us'

d. *ill-k-**utu*** 'She sees me'

e. *ill-**ushtu*** 'She sees us'

f. *ill-k-**tma*** 'She sees us'

g. *ill-k-**ta*** 'You see him'

h. *ill-k-**t^ha*** 'I see him'

i. *ill-k-**tna*** 'We see him'

j. *ill-k-**i*** 'She sees him'

The only other example of a fused verbal marking strategy in the sample can be found in the Moseten, a Mosetenan language of Bolivia described in Sakel (2004). While spoken along the upper stretches of the Beni River in Southern Amazonia, the Moseten territory lies in the Bolivian foothills region where both Quechua and Aymara are dominant cultural forces. As such, fused marking

strategies have only been observed in the Central Andes and surrounding areas within the sample.[16]

3.3.5 Fusion of features

It is possible for the grammatical features indexed through verbal argument marking to be conflated into a single morphological form or marked separately through specialized morphology for each feature. As already discussed in section (3.2.2), it is possible for a language to index the number of a certain argument through specialized morphology, as shown for Timbira in Table (3.3) and Urarina in Table (3.6; alternatively, a language can index both person and number features of the argument with a single non-segmentable morpheme, as can be seen for Kwaza in Table (3.4). Similarly, gender can be indexed through an argument marker that fuses gender with person (and number), as seen for Lokono in (19), or gender can be indexed independently of other argument features with a specialized marker, as seen for Miraña in (20). No cases have been observed in the sample where a language conflates gender and number marking into a single morpheme with person marked separately.

3.3.6 Optionality of markers

Q2.1.1   *Person marking for S when conominal is expressed in the clause is:* [obligatory, variable, never]

Q2.4.1   *Person marking for A when conominal is expressed in the clause is::* [obligatory, variable, never]

Q2.7.1   *Person marking for P when conominal is expressed in the clause is::* [obligatory, variable, never]

As discussed in Section 3.1, the expression of grammatical features of an argument through morphology bound to the predicate is treated here as a unitary grammatical phenomenon: indexation. Previous work has often attempted to describe indexation in particular languages as either an agreement phenomenon or as bound pronouns (or sometimes even both; Bresnan and Mchombo (1987)). A central component in the different approaches to indexation revolves around the observation that verbal argument markers vary cross-linguistically with regard to their occurrence when a corresponding nominal argument, what is called the CONOMINAL, is also expressed in the clause (Haspelmath, 2013). Distinguishing verbal argument markers based on the behavior of the conominal

---

[16]However, it has been observed that some languages in the sample show one (or a few) portmanteau morphemes within a person marking system without displaying an overall fused pattern. These portmanteaus tend to reflect situations in which a speech act participant is acting on another speech act participant. For example, see the discussion on the person marking prefixes in Tupian languages in Chapter 5.

in clauses with marked indexation, the following distinctions can be made:

(37)   Verbal argument marker types based on conominal occurrence
   a. *Type 1*: markers occur with an obligatory conominal.
   b. *Type 2*: markers occur with optional conominals.
   c. *Type 3*: markers never co-occur with a conominal.

Such an approach facilitates comparison with previous claims about the nature of indexation and allows a direct correlation with categories already discussed extensively in the theoretical and typological literature (Jelinek, 1984; Bresnan and Mchombo, 1987; Siewierska, 2004; Corbett, 2006). As such, Type 1 markers are analogous to what has often been called 'agreement' markers, Type 2 markers are analogous to 'cross-reference' markers, and Type 3 markers are analogous to 'bound pronouns' or 'anaphorical agreement' markers.

The first type of verbal argument markers is not encountered in any of the languages in the sample. While common in Germanic languages, such as in English ***she** run-**s***, these types of verbal argument markers are rare in the world's languages (Siewierska, 1999). Thus all indexation in the languages in the sample are of either Type 2 or Type 3.

Type 2 markers are found in the majority of marker sets across the languages that index arguments on the verb (70-90%, depending on the argument role). However, even within these types of indexation systems there is an additional parameter of variation observable in the data that is not fully discussed in the typological literature: whether the verbal argument markers obligatorily or optionally occur with a conominal. Since the Type 3 markers never occur with a conominal, i.e. markers are in complementary distribution with the nominal argument, the observed variation in the sample can be encapsulated by whether the occurrence of verbal argument markers when a conominal is realized in the clause in obligatory, optional or unattested.

Obligatory markers: In some language, the occurrence of argument markers is obligatory irrespective of whether the conominal is present or not. This can been seen for the Macro-Jêan language Rikbaktsa, where both the subject and object verbal argument markers are obligatory regardless of whether the nominal argument is realized as a full nominal (38a), an independent pronoun (38b-38c), or not realized at all (38d):

Rikbaktsa (Macro-Jêan; Silva, 2011, 167, 239, 241-242)

(38)   a. *ka-ʃte   piknu ɾ-i-do*
           1SG.POSS-daughter fish 3SG-3SG-grill
           'My daughter grilled the fish.'
   b. *katʃa piknu **tʃi**-mi-ɾ-i-do=naha*
           1PL   fish 1PL-NPST-DIR-3PL-ascend=PL
           'We will catch fish.'

    c. *paɾahei-tʃa   katʃa **ni-mɨ**-do-wə=hik=naha*
       peccary-NF.PL 1PL   3SG-1PL-ascend-CAUS-PONCT=PL

       'The peccaries made us climb (the tree).'

    d. *ɾ-i-beɾe*
       3SG-3SG-kill.SG

       'He killed it.'

Variable markers: Some languages in the sample do not require the realization of a verbal argument marker when a conominal is present in the clause. Siewierska (1999, 228-230) notes that there are a number of factors that can contribute to the variation of the occurrence of verbal argument markers. One of these factors is the location of a lexical argument in the clause. This type of variation can be seen in Apurinã, an Arawakan language of Brazil, in (39):

Apurinã (Arawakan; Facundes, 2000, 384, 467)

(39)   a. *aõtu **o**-apa-nanu-ta-ru   suto*
         umari 3F-fetch-PROG-VBLZ-3M woman

         'The woman was gathering umari fruit.'

      b. *hãtako-ro apa-nanu-ta-ru   aõtu*
         youth-3F fetch-PROG-VBLZ-3M uxi

         'The girl is gathering uxi fruit.'

As can be seen in example (39), the subject prefix only occurs when the conominal argument (S/A) is post-verbal (39a). When the conominal subject occurs pre-verbally, no subject marker can occur on the verb (39b).[17]

While the variability in the occurrence of verbal argument markers in Apurinã is dependent on the positional realization of nominal arguments in the clause, this is not always the case. An example of this can be seen in Latundê, a Nambikwaran language of Brazil, where Telles (2002, 220) notes that when arguments occur in the clause as expressed nominals, indexation becomes optional. Due to zero marking in the third person, this can only seen through the co-occurrence of a first or second person free pronoun conominal with nominative marker set, as seen in (40):

Latundê (Nambikwaran; Telles, 2002, 153, 277-278)

(40)   a. *nũh kejn-**wahi**-tãn*
         1PL split.firewood-1PL-IMPF

         'We split firewood.'

---

[17]For further discussion on the interplay between word order and the realization of verbal argument markers in Apurinã, see Facundes (2000, 554-558)

    b. *nũh ạjh-tãn*
       1PL walk-IMPF

       'We went.'

    c. *hajn-ka-**wahi**-tãn*
       talk-BEN-1PL-IMPF

       'We told him.'

It is not clear in the description as to the exact circumstances in which the markers are realized or not, except that they are optional. Perhaps the distinction is purely stylistic or there may be additional motivations such as those in Apurinã.

Variable markers are distinguished from obligatory markers to register the behavior of the markers seen in examples (39-40). Given the three way classification of verbal argument markers presented in (37), markers such as the subject prefix in (39a) and the subject suffix in (40a) would clearly be labeled as cross-indexes (or cross-referencing markers). However, the absence of markers with a conominal in examples (39b) and (40b) would lead them to be considered as pro-indexes (or bound pronouns). These are clearly the same markers and a classification of verbal argument marking should be inclusive enough to account for this type of variation, even if the variation is not understood in its entirety based on the available descriptive materials. It could be argued that the condition of a post-verbal subject in Apurinã triggers the different observed patterns, but no such conditions are evident for Latundê. Haspelmath (2013) suggests that such languages can be considered an intermediate type on the continuum between cross-reference markers (Type 2 above) and pro-indexes (Type 3 above), but due to the high occurrence of variable markers in the sample, this distinction has been captured in the structural questionnaire. The precise nature of this variation in the behavior of argument markers is a topic that requires further investigation, both in the description of individual languages and in the typology of indexation systems.

Complementary markers: Some languages in the sample allow for either the realization of a nominal argument or a verbal argument marker, but never both for the same argument. This can be seen for the Warao language of Venezuela in (41), where pre-verbal independent subject pronouns (41a) are in complementary distribution with the suffixes that index the nominative argument (41b):

Warao (isolate; Romero-Figeroa, 1997, 64-65)

(41)   a. *warao ine namina-te*
        Warao 1SG know-NPST

        'I know (speak) Warao.'

b. *wauta   omi   naru-ki-tia-**ine***
   Wauta without go-INT-HAB-1SG
   'I am going to see Wauta as usual.'

Given the examples presented in (38-41), it can be seen that verbal argument markers show different patterns in terms of their realization with a conominal argument. The typology outlined by Haspelmath (2013) and summarized in (37) is useful in distinguishing between the type of verbal argument markers seen in Warao (41) from those seen in Rikbaktsa (38), Apurinã (39) and Latundê (40), based on the criterion of the obligatoriness of the realization of arguments in the clause. However, the same typology fails to distinguish between the differential behavior of the markers within the cross-referencing type of markers (38-40). In order to adequately capture the different properties of the verbal argument marker, this thesis also considers whether a verbal argument marker is obligatory, optional or never expressed when a conominal is present. As such, it is possible to split the category of cross-referencing markers into two subtypes: obligatory cross-reference markers and variable cross-reference markers, as summarized in Table (3.9).

MARKER TYPE	CONOMINAL	OCCURRENCE	LANGUAGE
Agreement	obligatory	obligatory	English
Obligatory cross-reference	optional	obligatory	Rikbaktsa
Variable cross-reference	optional	optional	Latundê
Bound pronominal	optional	complementary	Warao

Table 3.9: Revised types of verbal argument markers

3.4   Conclusions

This chapter has outlined a number of basic comparative concepts and parameters related to verbal argument marking. After introducing the basic approach to verbal argument marking used in this thesis, a small discussion of the major facets of person, number and gender marking was presented. Using the indexation of person for arguments as a starting point, the concept of a verbal marking pattern was introduced, which is a more specific characterization of indexation in a language than a general statement about overall alignment patterns. Now that the major types of simple verb marking patterns have been introduced, the following chapter looks at conditions that affect indexation such that it can be difficult to give an overall characterization of the verbal marking pattern in a language.

Verbal argument marking: Complex patterns

Languages often show multiple patterns in the ways that arguments are marked on the verb. These patterns form different constructions, which can be conditioned by a number of factors such as predicate class, tense-mood-aspect (TMA) inflection, negation, and the referential properties of the arguments.[1] For each of these different conditions, a language can display a specific marking pattern. A language with different (simple) marking patterns across different conditions are referred to as displaying a COMPLEX MARKING PATTERN. It is common for languages to display some form of variation in their argument marking strategies, especially if non-basic constructions found in subordinate clauses, copular clauses, clauses with focus marking, negated clauses, constructions that require an auxiliary verb, and so forth are considered. However, the discussion presented here is restricted to constructions in main clauses where the distinction between basic and non-basic constructions is not so straightforward. Within this more restricted domain, there are two prominently recurring complex patterns within the data: those conditioned by the referential scenario between arguments of a transitive clause (hierarchical marking) and those conditioned by the lexical class of monovalent predicates (split intransitivity).

Hierarchical marking and split intransitivity are of special interest to a discussion on verbal argument marking for a number of reasons. First, both

---

[1]Different marking patterns conditioned by polarity were not coded for since the affirmative is considered basic in the present work. Different verbal marking patterns conditioned by tense-mood-aspect inflection were controlled for by the application of the event properties of the transitive situation type discussed in section 2.1.2. The perfective-imperfective split in Mundurukú is discussed in section 5.1.3 and the multiple transitive constructions in Timbira are discussed in section 2.2.3.

hierarchical marking and split intransitivity have been the subject of intense debate in the typological and descriptive literature. There is considerable discussion within the field of linguistic typology as to whether languages that display components of these complex patterns are best treated as belonging to a distinct alignment type, as in Siewierska (2011a) and Nichols (1992), or as being composed of multiple alignment types, as argued by Dixon (1994). Second, these types of complex patterns are especially common in South American languages, being found across a wide range of linguistic families and geographic regions. As demonstrated below, there is much structural variation in the patterns observed in the sample, but these different patterns can be classified more generally into discrete types based on an examination of the realization of different marker sets and the positions that these sets occur in across the different construction conditions.

This chapter first explores hierarchical marking patterns and then turns to split intransitive marking patterns. After this structural overview, a case study is presented using the Cariban language Hixkaryana, which serves to illustrate a number of the topics discussed in the first two sections and how they can be applied to a language whose typological categorization is not immediately clear.

4.1   Hierarchical marking

Q3.1   *Language has a scenario-based split in verbal argument marking:*
          [1, 0]

Q3.1.1   *A distinct morpheme marks the inverse construction on the verb:*
          [1, 0]

Q3.1.2   *Referential arguments in inverse constructions are treated with*
          a marker set distinct from that used to index the referential ar-
          gument in direct constructions: [1, 0]

Q3.1.3   *Inverse construction types can be used to distinguish between 3rd*
          person arguments in non-local scenarios: [1, 0]

Languages with a hierarchical pattern in verbal marking select an argument to index in a particular marker slot according to which of two transitive arguments is most referential, i.e. the argument that is ranked higher on the person hierarchy. This leads to two different morphosyntactic constructions within the transitive situation type based on the relative referential properties of the arguments. This mismatch between different marking patterns in languages, especially those labeled as 'ergative' languages, was first discussed by Silverstein (1976), with the concept further explored in Dixon (1979) and Comrie (1981). While there have been various formulations of the referential

hierarchy, and various names applied to it such as the animacy, salience, nominal or indexability hierarchy, a key aspect of this system to be considered here with regards to verbal argument marking is that speech act participants (henceforth SAPs; 1st and 2nd person participants) are considered more referential than non-speech act participants (3rd person participants). The portion of the referential hierarchy concerned with the status of different person forms is referred to in the thesis as the PERSON HIERARCHY. The exact manifestation of the hierarchy varies from language to language, especially with regard to the treatment of SAP arguments that act on other SAP arguments. Languages can also show hierarchical marking patterns conditioned by the pragmatic status of different 3rd person arguments, as has been well described for Algonquian languages in North America (cf. Zúñiga, 2006a).[2] While the latter type of system often functions in conjunction with hierarchical marking conditioned on the person of the transitive arguments, the following section focuses primarily on the person-based split, with the pragmatic-based split discussed in Section 4.1.3.

Important to a discussion on hierarchical marking is that the rules governing the indexation of a particular argument do not depend solely on the referential properties of that argument, but also the referentiality of that argument relative to the other argument in the clause, i.e. the CO-ARGUMENT. The relationship between the co-arguments of a construction are referred to here as the SCENARIO. This chapter is primarily concerned with scenario between co-arguments in transitive constructions.[3] Scenarios can be generally organized around whether A and P are SAP arguments or 3rd person arguments. Scenarios where both A and P are SAPs are called LOCAL SCENARIOS. Scenarios where both A and P are 3rd person arguments are called NON-LOCAL SCENARIOS. Scenarios where one transitive argument is a SAP and the other is a 3rd person are called MIXED SCENARIOS. Mixed scenarios can be further divided into two types: transitive clauses where a SAP A acts upon a 3rd person P are DIRECT SCENARIOS, and clauses where a 3rd person A acts upon a SAP P are INVERSE SCENARIOS. Scenarios are composed of multiple person-based argument CONFIGURATIONS, such that an inverse scenario is typically composed of configurations such as 3→1sg, 3→2sg, 3→1pl.incl, etc. As is shown below, scenarios provide a general framework to discuss hierarchical marking patterns, but languages can show divergent patterns at the level of specific configurations.

A terminological distinction is also maintained in this thesis between sce-

---

[2] Gildea (1994) distinguishes between these two systems by labeling the person-based split in hierarchical marking as 'inverse alignment' and the pragmatics-based split between 3rd person arguments as 'inverse voice'. The term 'inverse alignment' is not adopted here since it does not fit with the more restricted concept of alignment used throughout this thesis.

[3] It is also possible to discuss the scenario between co-arguments in ditransitive constructions. Especially relevant is the scenario between the T and R arguments, which can have an effect on which argument is indexed. The effects that referential properties of co-arguments have on ditransitive indexation remains a topic for future investigation and is not discussed in this thesis.

narios and constructions. An INVERSE CONSTRUCTION is the construction type used in inverse scenarios when a language displays a scenario-based split. However, as is shown in the following discussion, some languages also extend this construction type to certain local and/or non-local scenarios. Languages that explicitly mark inverse constructions with verbal morphology distinct from their argument markers have INVERSE MARKERS. Correspondingly, a DIRECT CONSTRUCTION is the construction type used in direct scenarios when a language displays a scenario-based split. Both direct and inverse scenarios are considered subtypes of the more general transitive construction, and are treated as such in the structural questionnaire. Among the languages in the sample that display hierarchical verb marking patterns, two general groups can be identified: those where scenario conditions the selection and reference of verbal indexation as well as inverse markers, and those without any separate inverse markers.

4.1.1   Structural variation in hierarchical marking patterns

South America hosts many languages with hierarchical verb marking patterns, both with and without inverse markers. A total of 16 languages in the sample show a clear hierarchical verb marking pattern. Hierarchical patterns together with an inverse marker is clearly attested in three non-related languages in the sample: Mapudungun, Movima and Itonama.[4] Mapudungun, a language isolate spoken in Chile and Argentina, has a single set of suffixes that indexes the most referential argument, two suffixes that index either 3rd person A (42c) or 3rd person P (42b), and an explicit inverse marker (42b), as can be seen in (42).

Mapudungun (Araucanian; Smeets, 2008, 152, 415, Zúñiga, 2006b, 215)

(42)   a. *iñché uwaw-tu-**n***
          1SG   sleep-VERB-1SG.IND
          'I slept.'

       b. *kim-la-**fi-n***   *ti   añchümalleñ*
          know-NEG-3O-1SG.IND ART midget
          'I don't know this midget.'

       c. *mütrüm-**e-n-ew***   *chi kalku*
          call-INV-1SG.IND-3A ART warlock
          'The warlock called me.'

Itonama, a language isolate from Bolivia, shows a different pattern in its hierarchical marking. It has two different sets of argument marking prefixes.

---

[4]Yanam displays an interesting case that can only controversially be considered an inverse marker, as discussed in section 4.1.4.

One prefix set is used to index the nominative argument of independent clauses in direct scenarios (43a) and local scenarios (SAP → SAP; 43b), both of which comprise the direct construction. Another prefix set indexes P in independent clauses in inverse scenarios (43c), forming the inverse construction. The marker set that indexes P in inverse constructions is also used to index the S and A arguments in dependent clauses (43d). The morpheme *k'i-* marks inverse constructions (43c), and the suffix *-mo* indexes 1st person P in local scenarios (43b).

Itonama (isolate; Crevels, 2011, 580, 583; Crevels, 2012a, 264-265)

(43) a. *wase'wa* ***e'****-yaka-ne*   *machiriri*
       yesterday 2SG.I-sing-NEUT paper
       'Yesterday you were reading the book.'

    b. ***e'****-kamo'-ke-****mo***
       2SG.I-hit-PL-1
       'You hit us (in the face).'

    c. ***a'****-****k'i****-kamo'-ke*   *ihwana*
       2SG.II-INV-hit-PL Juan
       'Juan was hitting you.'

    d. ***a'****-may-sewa-na*
       2SG.II-SUB-see-NEUT
       'When you see him/her/it.'

Mapudungun and Itonama share a number of structural properties in their hierarchical marking patterns: each have one marker slot that indexes the most referential argument of the transitive clause, a specialized marker is used to indicate that the arguments of the predicate are configured in such a way to form the inverse construction, and each use a limited set of markers to index the non-referential co-argument in certain configurations. However, a major difference between the two languages is that Mapudungun only has a single set of markers to index the referential argument, displaying neutral alignment since these same markers are also used to index S, whereas Itonama has two marker sets that occur in complementary distribution in the same marker slot: one that indexes S/A and another that indexes P, with both marker sets displaying accusative alignment.

Similar patterns where the most referential transitive argument is indexed with one of two different sets of markers that occur in the same slot can be seen in other languages that do not have an explicit inverse marker. These types of patterns are encountered in a number of Cariban and Tupian languages, as well as certain Guaycuruan languages and some languages of the Northern Andes (see Table 4.3). An example of a language with a hierarchical marking pattern without an inverse marker is Ikpeng, a Cariban language of Brazil, in (44):

Ikpeng (Cariban; Pachêco, 2001, 70-71)

(44)   a. **w**-*aginum-lɨ*
            2SG.II-cry-REC.PST
            'You cried.'

        b. **m**-*eneŋ-lɨ*
            2SG.I-see-REC.PST
            'You saw him.'

        c. *eneŋ-lɨ*
            see-REC.PST
            'He saw him.'

        d. **o**-*eneŋ-lɨ*
            2SG.II-see-REC.PST
            'He saw you.'

        e. **k**-*ineŋ-lɨ*
            1SG.I-see-REC.PST
            'I saw you.'

        f. **ugw**-*ineŋ-lɨ*
            1PL.II-see-REC.PST
            'He saw us.'

        g. **ugw**-*ineŋ-lɨ*
            1PL.II-see-REC.PST
            'You saw me.'

Notice in example (44) that the 2nd person S of the intransitive verb *aginum* 'cry' (44a) is indexed in the same way as P of the transitive verb *ene* 'see' (44d).[5] In (44c), the 3rd person A and P are unmarked, while the 2nd person A in (44b) is indexed through a different set of markers than that used for S and P in (44a) and (44d). Just as in Itonama, different marker sets are used in Ikpeng depending on whether A or P is most referential on the person hierarchy. However, unlike in Itonama, the inverse construction in Ikpeng does not employ an inverse marker and is only expressed through the use of a different marker set than the one used in the direct construction.

4.1.2   Hierarchical marking in local scenarios

In local scenarios, the languages surveyed present a number of different marking patterns. In Ikpeng (44a), when a 1st person A is acting on 2nd person P, the 1st person set I (S/A) marker *k*- is used (44e), suggesting a hierarchical configuration that shows a relative rank between SAPs, namely 1 > 2 > 3.

---

[5]Note that *w*- and *o*- are allomorphs for second person singular within Set II. See Pachêco (2001, 63-69) for further details and examples.

However, when a 2nd person A is acting on a 1st person P, the marker *ugw-* is used (44g), which is also used to index 1st person plural P (44f).[6]

Itonama, on the other hand, always indexes A in the hierarchical prefix slot when it is an SAP, with SAP P expressed by suffixes in local configurations, as shown in (43b). So whereas Ikpeng partially maintains a hierarchical distinction between different SAP arguments in the local scenario, thus expanding the use of the inverse construction type beyond inverse scenarios, Itonama restricts the use of inverse constructions solely to inverse scenarios. The different strategies used to index arguments in local configurations in languages with a hierarchical marking pattern was not coded in the structural questionnaire. It is an interesting prospect for further detailed investigation.

4.1.3   Hierarchical marking in non-local scenarios

Just as the referential hierarchy can or cannot condition the indexation of different arguments in specific configurations within the local scenario for different languages, languages also vary as to whether the strategies used to distinguish between direct and inverse constructions are additionally used in configurations where both arguments are 3rd persons, i.e. in the non-local scenario. Adopting terms from the Algonquian literature, the more referential 3rd person argument is the PROXIMATE argument, while the less referential 3rd person argument is the OBVIATIVE argument. The distinction between a proximate and obviative 3rd person argument tends to be conditioned by semantic and/or pragmatic factor such as the animacy and topicality of the arguments (cf. Zúñiga, 2012). One language that makes such a distinction is Movima, a language isolate from Bolivia, as shown in (45).

Movima (isolate; Haude, 2006, 566; Haude, 2009, 519)

(45)   a.   *aya-**na**=**y'ɬi***
            wait-DR=1PL

            'We waited for (him).'

       b.   *vel-**kay**-a=**y'ɬi***
            look-INV-LV=1PL

            '(They) looked after us.'

       c.   *ena'   kon-**na**=**'ne**   is   empana:da*
            DUR.STD drain-DR=3F ART.PL empanada

            'She is taking the empanadas out (of the oil).'

---

[6]We can be reasonably confident that these are indeed the same marker rather than distinct yet homophonous forms since they both show the same *ugw-* / *wi-* allomorphy. This leads Pachêco (2001, 70) to analyze this marker as expressing both 1st and 2nd persons. The use of markers homophonous with 1st person plural forms to index both A and P of certain local scenarios in languages with a hierarchical marking pattern is also found in some Tupí-Guaraní languages (see ex. 66 for Kamayurá in the following chapter).

  d. *ew-**kay**-a-**'ne***   *os*   *alamre*
      hold-INV-LV=3F ART.N.PST wire

      'The wire held her back.'

Haude (2009, 520) describes Movima as having a referential hierarchy the makes the following distinctions: 1 > 2 > 3 human > 3 non-human animate > 3 inanimate. SAP arguments in direct (45a) and inverse (45b) scenarios are indexed on the verb through a set of enclitics, in these two cases it is with the marker =*y'łi* 'we/us'. When the clause is a direct construction (45a), the direct marker -*na* is used. When the clause is an inverse construction (45b), the inverse marker -*kay* is used. Similarly, in non-local scenarios when a higher ranked 3rd person A (e.g. human) acts on a lower ranked 3rd person P (e.g. inanimate), the direct marker -*na* is used (45c). Additionally, the enclitic indexes the A argument in this example, indicating that this is a direct construction. When the 3rd person A argument is ranked lower than the 3rd person P argument on the person hierarchy, the inverse marker -*kay* is used and the enclitic indexes the P argument, as seen in (45d). When both arguments are of equivalent ranking in the hierarchy, a distinction between topic and non-topic also conditions the occurrence of a direct or inverse construction (Haude, 2009, 520). Within the sample, this pragmatic distinction has only been observed in languages with an explicit inverse marker.[7]

4.1.4   Further cases of hierarchical marking patterns

Two languages in the sample present hierarchical marking patterns that diverge somewhat from the patterns presented in the previous sections: Yanam and Awa Pit. These two languages are discussed in this section in relation to the current typological formulation of hierarchical marking systems.[8]

In Yanam, a Yanomaman language of Brazil also known as Ninam, there are two preverbal slots to index transitive and intransitive arguments. The different marker sets in Yanam have already been introduced in Table 3.5 of section

---

[7]Furthermore, Zúñiga (2012) suggests that a similar distinction is also found in Jarawara, an Arawan language of Brazil described in Dixon (2004). Jarawara shows two different transitive construction types conditioned on the the discourse function of the transitive arguments: the A-construction is used when A is a "pivot argument", i.e. when the argument is shared with the preceding and/or subsequent clause; the O-construction is used when P is the pivot argument. The choice of pivot seems to be related to the status of the argument as a topic in the discourse (cf. Aikhenvald, 2012, 403), although Dixon (2004, 422) merely states that they are "determined by discourse considerations". But since the distinction between the A-construction and the O-construction is primarily restricted to word order and gender marking rather than person marking, it has not been considered here as an example of hierarchical marking.

[8]The Guató language of the Macro-Jêan family also presents a transitive marking pattern that is reminiscent of a scenario-conditioned system. However, this remains a topic for future investigation and it is not considered to show hierarchical marking in the current study, in accordance with the available descriptive materials in Palácio (1984, 2004).

3. The slot closest to the verb stem primary indexes a three-way number distinction for the absolutive argument, but also makes a SAP versus non-SAP person distinction (Set 2 in Table 3.5), with singular SAP arguments being unmarked. The slot furthest from the verb stem (Set 1 in Table 3.5) shows a clear hierarchical pattern: when A is a SAP in direct scenarios, it is indexed in the slot (46a); in inverse scenarios, P is indexed (46b). In local scenarios, P is also indexed (46c).

Yanam (Yanomaman; Ferreira, 2012, 13, p.c.)

(46)   a. *kamatʃə=n* **tʃa=kɨp=ʃə-ɨ**
           1SG=ERG   1SG=3DU=beat-DYN
           'I beat them (dual).'

       b. *ɨh=pɨk=n*   **tʃa=e=ʃə-ɨ**
           DEM=PL=ERG 1SG=INV=beat=DYN
           'They beat me.'

       c. *kami=tʃə=n* **wa=e=ʃə-ɨ**
           1=SG=ERG   2SG=INV=beat=DYN
           'I beat you.'

However, there is a mismatch between the inverse scenario and the use of the 'inverse' marker *e=*: when P is a SAP and singular, the 'inverse' marker occurs in only certain configurations, such as in 3pl→1sg configurations (46b) but not 3sg→2sg configurations (47c). When SAP P is non-singular (dual or plural), *e=* does not occur, as shown in (47a)(47b):

Yanam (Yanomaman; Ferreira, 2012, 13, p.c.)

(47)   a. *ɨh=pɨk=n*   **tʃama=k=ʃə-ɨ**
           DEM=PL=ERG 1PL=1/2.NSG=beat=DYN
           'They beat us.'

       b. *okoro=n* **wam=k=ɨsiwərem**
           dog=ERG 2PL=1/2.NSG=bite
           'The dog bit you (pl.).'

       c. *okoro=n* **wa=siwərem**
           dog=ERG 2SG=bite
           'The dog bit you.'

One possible analysis is that the 'inverse' marker *e=* is actually the Set 2 marker for a singular SAP P since they are in complementary distribution and occur in the same position between the Set 1 proclitics and the verb. However, if this were the case, one would expect *e=* to also occur in (47c), but it does

not. On the other hand, a configuration of 3→2sg would be expected to fall within the inverse construction since it is a configuration within the inverse scenario. Thus, the 3→2sg configuration is problematic for either analysis.

Based on a comparison with other Yanomaman languages, Ferreira (2012) shows that the marker $e=$ is likely cognate with the inverse markers in the other languages that indeed co-occur with the Set 2 proclitics and show a more prototypical distribution of the inverse marker.[9] Furthermore, Ferreira (2012) distinguishes between the marker slot closest to the verb and the morphological slot where the 'inverse' marker occurs. While the two slots are never filled at the same time in Ninam, certain morphophological processes occur between the hierarchical slot and the inverse marker slot that do not occur between the hierarchical slot and the slot where the Set 2 markers occur, e.g. $wa=$ 'you' + $e=$ 'inverse' are realized as $weh=$, which Ferreira considers as '2sg.inverse', such as in the 1sg→2sg configuration in (46c).

The use of the different Set 1 argument markers and 'inverse' marker in direct, inverse and local scenarios is presented in Table 4.1 based on information in Ferreira (2012). The most referential argument is referred to as the 'referent' in the table. Since the Set 2 markers always index P in transitive clauses, these are not included.[10]

While the hierarchical marking pattern for Yanam is clear, always indexing the most referential argument in mixed scenarios, it presents a borderline case for what constitutes an inverse marker. Since there is a mismatch between the inverse scenario and the use of the 'inverse' marker, the marker has not been considered an inverse marker in the structural questionnaire, even though there is comparative evidence to suggest that it likely developed from an inverse marker.

Awa Pit, a Barbacoan language spoken in Ecuador and Colombia, shows a different hierarchical marking pattern that diverges from the prototypical pattern discussed in section 4.1.1. Rather than a distinction between SAP arguments and 3rd person arguments, Awa Pit makes a hierarchical distinction between 1st person arguments and non-first person arguments in main indicative clauses, thus following a hierarchy of $1 > 2/3$. In interrogative clauses, the same markers are used to distinguish between 2nd person arguments and non-2nd person arguments (48d), thus following a $2 > 1/3$ hierarchy, leading Curnow (1997) to call this person category (1st person in statements, 2nd person in questions) the 'Locutor' since it reflects the participant in the utterance that is the source of knowledge, i.e. the 'epistemic authority'.[11] In the past tense declarative, if A is 1st person, it is indexed on the verb with the marker $-w$ (48a); if P is first person, it is indexed on the verb with the marker $-s$ (48b);

---

[9]In his description of Yanomamɨ, Ramirez (1994, 113-116) presents an early proposal for an inverse analysis of person marking in a Yanomaman language.

[10]This table was inspired by those used to describe hierarchical marking patterns in Witzlack-Makarevich (2010).

[11]This system shows many similarities with the conjunct/disjunct marking system in some Tibeto-Burman languages as described by Hale (1980) for Kathmandu Newari.

SCENARIO	MARKER	'INVERSE'	REFERENT		CO-ARGUMENT	
			ROLE	PERSON	ROLE	PERSON
Direct	tʃa=	-	A	1SG	P	3
	tʃehe=	-	A	1DU	P	3
	tʃama=	-	A	1PL	P	3
	wa=	-	A	2SG	P	3
	wehe=	-	A	2DU	P	3
	wama=	-	A	2PL	P	3
Inverse	tʃa=	e=	P	1SG	A	3
	tʃehe=	-	P	1DU	A	3
	tʃama=	-	P	1PL	A	3
	wa=	-	P	2SG	A	3
	wehe=	-	P	2DU	A	3
	wama=	-	P	2PL	A	3
Local	tʃa=	e=	P	1SG	A	2
	tʃehe=	-	P	1DU	A	2
	tʃama=	-	P	1PL	A	2
	wa=	e=	P	2SG	A	1
	wehe=	-	P	2DU	A	1
	wama=	-	P	2PL	A	1

Table 4.1: Overview of the hierarchical marking pattern in Yanam

if neither A nor P are 1st persons, this configuration is indexed on the verb with the marker -*zi* (48c):

Awa Pit (Barbacoan; Curnow 1997, 193-194, 199)

(48)   a. *kɨn-ka=na   na=na   Santos=ta izh-ta-**w***
             dawn-when=TOP 1SG=TOP Santos=ACC see-PST-LOCUT:SUBJ
             'At dawn I saw Santos.'

      b. *Libardo na-wa   pyan-tɨ-**s***
             Libardo 1SG-ACC hit-PST-LOCUT:UNDER
             'Libardo hit me.'

      c. *nu=na   Juan=ta   pyan-ti-**zi***
             2SG=TOP Juan=ACC hit-PST-NONLOCUT
             'You cut Juan.'

      d. *shi   ayuk=ta=ma   libro ta-ta-**w***
             what inside=in=INTER book put-PST-LOCUT:SUBJ
             'Under what did you put the book?'

      e. *nu-wa=na   mɨn=ma   pyan-tɨ-**s***
             2SG-ACC=TOP who=INTER hit-PST-LOCUT:UNDER

'Who hit you?'

While all the examples in (48) show the same hierarchical marking pattern, with referential A or P indexed depending on the argument scenario, the status of which argument is considered most referential is conditioned by whether the clause is in the indicative or declarative mood. The distribution of the different argument markers in Awa Pit transitive clauses is summarized in Table 4.2.

STATUS	PERSON	SET 1 (A)	SET 2 (P)	MOOD
Locutor	1	-w	-s	
Non-locutor	2	-zi	-	Declarative
	3	-zi	-	
Locutor	2	-w	-s	
Non-locutor	1	-zi	-	Interrogative
	3	-zi	-	

Table 4.2: Past tense verbal argument markers in Awa Pit

The Awa Pit verbal argument marking pattern is interesting in that it developed from a system of evidential marking rather than a pronominal source (Curnow, 1997, 211-212). This helps to explain the disjunction between the different person hierarchies functioning in declarative and interrogative clauses based around the category of epistemic authority. The grammaticalization of the evidential markers into a person marking system has taken place to different degrees across the Barbacoan languages. In Tsafiki, a Barbacoan language spoken in Ecuador, the use of the Locutor/Non-locutor ("Conjunct"/"Disjunct") distinction more closely corresponds to the evidential category of mirativity and does not reflect a clear person distinction.[12] Dickinson (2002, 90-102) puts forth a number of arguments for why the mirative marking in Tsafiki cannot be considered a person marking system, e.g. Locutor marking is not used in reported speech constructions even when the participant in the embedded clause is 1st person.

4.1.5   Summary of hierarchical marking patterns

A summary of the different structural parameters related to hierarchical marking patterns is presented in Table 4.3. Alignment here refers specifically to the marker set(s) used in the marker slot that is sensitive to the scenario between co-arguments.[13] The column labeled 'Inverse' refers to whether the language

---

[12]See DeLancey (1997) for an overview of mirativity.

[13]For languages with split intransitivity within a hierarchical marking pattern, alignment in the table refers to how the major class of S aligns with the transitive arguments. See section 4.2.1 for a definition of both major and minor intransitive verb classes.

has a distinct inverse marker.[14] The column labeled '3 > 3' refers to whether the inverse construction in that language distinguishes between different 3rd person arguments.

LANGUAGE	FAMILY	ALIGNMENT	INVERSE	3 > 3
Mapudungun	Araucanian	Neutral	+	+
Awa Pit	Barbacoan	Accusative	-	-
Apalaí	Cariban	Accusative	-	-
Hixkaryana	Cariban	Tripartite	-	-
Ikpeng	Cariban	Ergative	-	-
Tiriyó	Cariban	Ergative	-	-
Chimila	Chibchan	Accusative	-	-
Mocoví	Guaycuruan	Accusative	-	-
Itonama	isolate	Accusative	+	-
Movima	isolate	Neutral[15]	+	+
Awetí	Tupian	Tripartite	-	-
Emerillon	Tupian	Accusative	-	-
Kamayurá	Tupian	Accusative	-	-
Mundurukú	Tupian	Accusative	-	-
Sateré-Mawé	Tupian	Accusative	-	-
Tapiete	Tupian	Accusative	-	-
Yanam	Yanomaman	Neutral	(-)	-

Table 4.3: Summary of hierarchical indexation patterns

Based on the data included in this table and the information presented within section 4.1, a number of general observations can be made about the hierarchical verbal marking patterns in the languages examined:

- Languages with hierarchical marking patterns show all attested major alignment patterns within the referential marker sets (accusative, ergative, neutral, and tripartite).

- Hierarchical marking patterns without an explicit inverse marker are more commonly found than those with an inverse marker.

---

[14]Payne (1994) puts forth a hypothesis that the 'relational prefix' found in many Tupian languages, especially of the Tupí-Guaraní branch, functions as an inverse marker since it occurs in inverse constructions. However, these markers also attach to possessed nouns, postpositions and minor class intransitive verbs (cf. Jensen, 1998; Meira and Drude, 2013). The distribution of the relational prefix in transitive clauses can be attributed to the phonological status of the different marker sets, since it occurs with the proclitic set (S_{min}/P) and not the prefix set (S_{maj}/A). It is not considered an inverse marker in this study.

[15]While either A or P can be indexed with the same set of enclitics in Movima, S cannot. S can only be indexed with an additional set of enclitics that are also used to index A and P when non-referential (obviative in Haude's terms). For further details see Haude (2009) and Haude (2011).

- All of the languages with an explicit inverse marker are linguistic isolates. The small Yanomaman family is an exception to this, but the Yanam language does not have an inverse marker.

- None of the languages with an explicit inverse marker also display split intransitivity (see Table 4.5).

At present it is not possible to directly compare the continent-wide distribution of hierarchical patterns in South American languages with regard to the global distribution using the sample from *World Atlas of Language Structure* presented in Siewierska (2011a). This is due to the inability of a language to show both split intransitivity and hierarchical marking in the coding scheme adopted in the WALS database and the imprecise nature of the value [split] in this dataset. Impressionistically however, South American languages shows a much higher proportion of hierarchical marking patterns than the global sample based on a cursory examination of the available data. Other hierarchical patterns are also found in North America, Australia and the Himalayas.

4.2   Split intransitivity

Q4.1   *Language has two classes of intransitive verbs with separate formal marking of their arguments*: [1, 0]

Q4.1.4   *Minor class intransitive verbs show alignment with*: [A, P, neither]

Languages with split intransitivity display multiple argument marking patterns for different intransitive predicate classes. Within the sample, 22 languages display one of the split intransitive marking patterns discussed below.

Some authors have conceived of languages displaying split intransitivity as being composed of multiple basic alignment patterns, namely a nominative-accusative and an ergative-absolutive pattern (Dixon, 1994), while others have considered split intransitivity to be a distinct alignment pattern (Donohue and Wichmann, 2008; Siewierska, 2011a). However, there is no need to posit split intransitivity as a distinct alignment type for comparative purposes if one applies the same semantic considerations used to distinguish intransitive arguments across predicate classes as those used to distinguish transitive arguments within a specific construction. Before exploring the diversity of the split intransitive systems encountered in the sample, a few useful notions must first be introduced.

Subjecthood is a topic that has frequently been at the center of discussions on linguistic typology and language description. As early as Sapir (1917), it has been recognized that many languages in the Americas show variation in their treatment of intransitive subjects such that for one class of intransitive verbs the subject is marked like transitive subjects in that particular language, and

for another class of intransitive verbs, subjects are marked like transitive direct objects. Such splits in alignment patterns across different intransitive predicate classes has led to the languages displaying such patterns to be characterized using a number of different names in the literature, such as 'unergatives' and 'unaccusatives' (Perlmutter, 1978; Levin and Rappaport Hovav, 1995), 'active' and 'inactive' (Klimov, 1974), 'active-stative' (Mithun, 1991), 'split-S' (Dixon, 1994), 'semantic alignment' (Donohue and Wichmann, 2008) and 'split intransitivity' (Merlan, 1985; Van Valin Jr. and LaPolla, 1997, among others). The term split intransitivity is adopted here since many of the other terms make semantic or syntactic presuppositions about class membership which may not be cross-linguistically valid for all languages displaying such a split, as discussed in the following section.

4.2.1 Semantic considerations

Q4.1.1   *S of event verbs without control (e.g. hiccup, die, slip) are treated like minor class*: [1, 0]

Q4.1.2   *S of stative verbs with control (e.g. reside, be patient, be prudent) are treated like minor class*: [1, 0]

A topic that has received considerable attention in the typological literature on split intransitivity is the semantic basis for the division of intransitive verbs into different marking classes (cf. Donohue and Wichmann, 2008). Mithun (1991) shows that the division between different intransitive verb classes can be based on semantics features of both the argument and/or the clause, such as lexical aspect, affectedness, control and performance/effect/instigation, with the latter two features generally associated with agency. However, it is common for one of the classes of verbs to display certain semantic and pragmatic features typically associated with transitive agents, such as high animacy, volition and topicality. Due to its considerable influence and widespread adoption in the descriptive literature on South American languages, this discussion on the semantics of split intransitivity begins by quoting the characterization of such systems in Dixon (1994, 70):

> "We noted that there is a semantic basis to the assignment of A and O to semantic roles in a transitive clause. S, in contrast, simply marks the sole core NP in an intransitive clause. Since each grammar must include semantically contrastive marking for A and O, this can usefully be applied also to S - those S which are semantically similar to A (exerting control over the activity) will be S_a, marked like A, and those S which are semantically similar to O (being affected by the activity) will be S_o, marked like O."

Characterized as such, the approach to split intransitivity outlined above makes two basic assumptions about the mapping of semantic features of the

arguments onto their morphosyntactic realization: a) S semantically similar
to A [+control] will be marked like A, and b) S semantically similar to P [-
control] (Dixon's O) will be marked like P. [16] However, as discussed extensively
in Mithun (1991) for Guaraní, Lakhota and Central Pomo (the latter two being
North American languages), the semantic feature of [+/- control] is a major
typological parameter of variation in split intransitivity and is not sufficient to
predict class membership alone. To illustrate this point, let us first examine
the different marking patterns of intransitive and transitive verbs in Apurinã,
an Arawakan language of Brazil:

Apurinã (Arawakan; Facundes, 2000, 277-279, 282, 347)

(49)   a.   **p**-*etama-ta-**no***   *nota*
            2SG-see-VBLZ-1SG 1SG
            'You saw me.'

       b.   **o**-*serena suto*
            3F-dance   woman
            'The woman danced.'

       c.   **nh**-*irika nota*
            1SG-fall   1SG
            'I fell down.'

       d.   **nu**-*sutũka-pe nota*
            1SG-stink-PRF 1SG
            'I smell bad.'

       e.   *pooma-**no** nota*
            be.hot-1SG 1SG
            'I am hot.'

As can be seen in (49a), transitive verbs in Apurinã index the A argument
with a prefix and the P argument with a suffix. Facundes (2000) divides in-
transitive verbs into three classes: 'standard' intransitives index S with a prefix
and includes verbs with a semantically agentive subject (49b) or a semantically
patientive subject (49c); 'subjective descriptive' verbs index S with a prefix and
includes verbs that express physical and psychological properties of their sub-
jects (49d); and, 'objective descriptive' verbs index S with a suffix and includes
verbs that express the physical properties of their subjects (49e). Following a
strict reading of the characteristics of split intransitivity from Dixon above, it
would be expected that the subject of the intransitive verb *irika* 'fall down'
in (49c) would be marked with the suffix marker set used for transitive direct

---

[16]In later work, such as Dixon (2010), this characterization is portrayed as a general ten-
dency and not as an fixed rule. However, it is still unclear whether the notation S$_o$ is used to
represent an intransitive argument that is marked like O (i.e. P) or whether it is the argument
that shows patient-like semantics.

objects since the argument does not have control over the event and is directly affected by its outcome. Even within the descriptive verbs that express physical properties of their subjects, the different marking patterns are conditioned by the lexical class of the predicate rather than strictly semantic criteria.

As the Apurinã data show, it is problematic to assume an *a priori* link between the semantics of a verb and the argument marking pattern of a particular lexical class. However, an approach is still needed to compare the different marking patterns of different classes of intransitive verbs. Such an approach would ideally allow for a distinction between the marking pattern of a lexical class of intransitive verbs, the alignment of these verbs with transitive arguments and the semantics of the class members.

Since early work by Merlan (1985), it has been recognized that in languages with split intransitivity, the different classes of verbs are not always equally large, frequent or productive. After examining a number of different criteria to select which of multiple classes can be considered most basic, Witzlack-Makarevich (2010) proposes that the best criterion for this is type frequency, i.e. the number of distinct predicates that belong to a particular marking pattern. At first glance, this is an appealing proposal since it provides an objective criterion for comparing different verb classes independently of their semantic composition. However, when dealing with this type of quantitative criteria, how can we be sure that what we are counting is either exhaustive or even proportionally representative of all the intransitive verbs in the language? Manually counting dictionary entries may be prohibitively labor intensive for any comparative study that includes more than a handful of languages, and such materials are not always available for South American languages. What is lacking here is an approach that can be applied using only the information available in a typical grammatical sketch of the size and coverage of a doctoral dissertation, since this is the most common source of information available for the languages used in this study.

As an approach to facilitate cross-linguistic comparison, Haspelmath (2011a, 561) suggests comparing the coding properties of only a subclass of intransitive verbs, namely "uncontrolled change of state verbs like 'die', 'rust', 'get lost', 'rot', 'grow'..." If there is further variation even within this subclass of verbs, he suggests looking exclusively at the verb 'die'. However, it is unclear how useful this approach would be in languages that treat the verb 'die' in the same way as event verbs with a volitional performer that is distinct from the treatment of another class of verbs, such as in Apurinã (cf. Facundes, 2000, 278). While this approach is much easier to apply across a wider sample of languages, it seems counterproductive to use a semantic class of verbs whose treatment is known to vary considerably across languages as a standard of comparison. Rather, the standard of comparison should be a semantic class of verbs that show a more uniform morphosyntactic treatment across languages.

An alternative approach, the one adopted in this thesis, builds off of the framework already introduced in Chapter 2 to distinguish between the two arguments of a transitive clause. As a standard of comparison, the class of

intransitive verbs whose sole argument displays prototypical agent properties are referred to here as belonging to the MAJOR CLASS of intransitive verbs. In a survey of Austronesian languages with split intransitivity, Foley (2005) finds that of all the semantic properties attributed to subjecthood examined, that of being an agent, i.e. a volitional performer of an event, shows the most uniform morphosyntactic treatment across the languages surveyed. Thus, the major class of intransitive verbs in a language is composed of those members whose sole argument is a volitional performer, such as 'swim', 'run', and 'shout', and all the other verbs in that language that treat S in the same way. The classes of intransitive verbs whose subjects diverge from this prototype are referred to as belonging to a MINOR CLASS of intransitive verbs.

Adopting this framework for the categorization of intransitive predicates into major and minor classes, it is possible to compare split intransitive systems across the languages in the sample based on the strategies used to express these splits in marking as well as the alignment of the different predicate classes with A and P. Such an approach also avoids making presuppositions about the semantic properties of the minor class of intransitive verbs since these are known to vary considerably cross-linguistically.

4.2.2   Marker-based split

In many languages of South America with split intransitivity, S is indexed with different sets of markers that occur in complementary distribution within a particular marker slot. The most common pattern observed in the sample is where subjects of major class intransitive verbs are indexed like A, and subjects of minor class intransitive verbs are indexed like P, as is shown in (50) for Sateré-Mawé:

Sateré-Mawé (Tupian; Silva, 2010, 120-128)

(50)   a.   *eipe **ewe**-i-put*
            2PL 2PL.I-REL-run

            'You (pl.) ran.'

        b.   ***ewe**-i-ma'at*
            2PL.I-REL-trick

            'You (pl.) tricked him.'

        c.   ***e**-he-haɨ*
            2PL.II-REL-speak

            'You (pl.) speak.'

        d.   ***e**-h-enoi*
            2PL.II-REL-teach

            'He taught you (pl.)'

As can be seen above in examples (50b) and (50d), the transitive marking pattern in Sateré-Mawé shows a hierarchical pattern. The two marker sets used in for transitive verbs are also used to distinguish the different classes of intransitive verbs. Notice how S in (50a) is indexed by the same prefix as A in (50b), while S in (50c) is indexed by the same prefix as the referential P in the inverse construction in (50d). Such a pattern is typical of many Tupian languages, especially of the Mawetí-Guaraní branch, and is also attested in some Cariban and Guaycuruan languages.

It is not the case that all languages show alignment of markers between S of the major class of intransitive verbs and A of transitive verbs. While only clearly attested in Cariban languages in the current sample, such as Tiriyó shown in example (51), it possible for S of the major class of intransitive verbs (51a) to align with P of transitive verbs (51b):[17]

Tiriyó (Cariban; Meira, 1999, 107, 471, 501, 637)

(51)   a. ***ji**-këhtun-ta-e*
        1SG.I-scream-FUT.IMPF-CTY
        'I will scream.'

    b. *noosinpë=rëken   **j**-arimika-ne*
        grandmother:1SG=only 1SG.I-raise-PST.PFV
        'My grandmother raised me alone.'

    c. *irëme **w**-ëe-ne   karaiwa   noonoo=pona*
        so   1SG.II-come-PST.PFV Brazilian land=DIR
        'So I came to the land of the Brazilians.'

    d. *irë=mao=rëken papa   **w**-ene-ne*
        that=time=only father:2SG 1SG.II-see-PST.PFV
        'At that moment I saw your father.'

Furthermore, the major class of intransitive verbs need not align with either of the transitive argument roles, resulting in a tripartite alignment pattern as

---

[17]See Meira (2000b) for a discussion on the "accidental" origins of split intransitivity in Tiriyó and other Cariban languages. He argues that semantic considerations alone are not sufficient to divide intransitive verbs into different classes. Instead, his analysis shows that the minor class of intransitive verbs are either synchronically or historically derived from detransitivized transitive verb stems, often having a reflexive or medio-passive meaning. Additionally, Cariban languages tend to have a small number of verbs that also belong to this class but do not show any trace of detransitivizing morphology, such as 'say', 'go', 'defecate' and 'sleep'. Due to the medio-passive and reflexive semantics of the majority of the intransitive verbs in this primarily derived class, it is treated as the minor class in the study.

The fact that the major class of intransitive verbs in Tiriyó shows alignment with P within a hierarchical marking pattern serves as a potential counterexample to the claim in Nichols (1992, 68) that hierarchical patterns do not occur with an ergative base alignment, i.e. an alignment between the major class of intransitive verbs and the transitive core arguments.

seen in Wichí:[18]

Wichí (Matacoan; Terraza, 2009, 125-126)

(52) a. *am* **la**-*x^{w}et*   *la-qatnat*
         2SG 2.I-sharpen 3.POS-knife

         'You sharpen the knife.'

      b. *am* **lata**-*qatin*
         2SG 2.II-dance

         'You dance.'

      c. *k^{y}e a-qu-x^{w}ax*   *y-iɬ* **a**-*palitsax-ila*
         if   2.POS-mother-DIM 3-die 2.III-suffer-FUT

         'If your mother dies, you will suffer.'

    While Wichí does not display a hierarchical marking pattern for transitive
clauses, it uses the same prefix slot with multiple sets of markers to index the
subject of the different intransitive predicate classes as well as A.[19] The previous
three languages shown in examples (50-52) distinguish between multiple classes
of intransitive verbs through the use of multiple sets of argument markers that
can occur in a single morphological slot. Coincidentally, all of the languages in
the sample that show a marker-based split for S indexation, with the exception
of Awa Pit, only employ prefixes to do so.

4.2.3   Position-based split

Languages that index the sole argument of different classes of intransitive verbs
through separate marker sets that occur in different positions relative to the
verb stem, such as indexing S$_{major}$ with a prefix and S$_{minor}$ with a suffix, dis-
play what is referred to here as a position-based split in split intransitivity.
This pattern can be seen in (53d) for the treatment of intransitive verbs in Ika:

Ika (Chibchan; Frank, 1985, 9, 11, 29)

(53) a. *zoʒa-na-**rua***
         go-DIST-1SG

         'I went.'

---

[18]It is worth noting that the set of markers that index A on the verb (52a) in Wichí share
a number of forms with the set of markers that index the subjects of the stative minor class
verbs (52c), most notably 1st person singular/1st person plural exclusive *n-* and 1st person
plural inclusive *ya-*. See section 3.3.2 and Terraza (2009, 124-138) for further details.

[19]Wichí serves as a good counterexample to the statement in Dixon (1994, 110) that
"there is no example [of a split] where a special marking is used just for some S and not for
A or O." The Tupian languages Awetí and Nheengatú also serve as counterexamples to this
generalization. See section 5.1.3 in the following chapter for further examples and discussion.

b. *mi-ʧua-na-**rua***
2SG-see-DIST-1SG

'I saw you.'

c. ***nä**-ʧua-na*
1SG-see-DIST

'He saw me.'

d. ***nä**-kusein-u'*   *gui ni*
1SG-recover-NEG also CTY

'I still have not gotten better.'

The position-based split in the marking of intransitive subjects occurs in eight languages in the sample and is attested in at least one language of every major regione, with the exception of the Central Andes and the Southern Cone. Beyond Ika, as seen above, this pattern is also attested in some Arawakan and Nambikwaran languages, as well as in a small class of verbs in the isolate Yurakaré.[20]

Languages with splits manifested primarily through case marking and not verbal marking, as is the case for Tariana (Arawakan; Aikhenvald, 2003a), are not considered in this section, nor are languages with a minor class of verbs that do not index S. However, languages with a minor class of intransitive verbs that show a split in verbal marking are indeed considered here, even in cases where the major class of intransitive verbs does not index S. This can be seen in (54) for Sabanê:

Sabanê (Nambikwaran; Antunes de Araujo, 2004, 173-174, 178)

(54)   a. *wayulu **t**-ip-i-datinan*
dog   1SG-see-VC-PRET.EV

'The dog saw me.'

b. *towali ay-i-datinan*
1SG   go-VC-PRET.EV

'I left.'

c. ***t**-ilup-i-dana*
1SG-vomit-VC-PRES.EV

'I vomit.'

---

[20]There is considerable difficulty in establishing whether the small class of verbs in Yurakaré with notional subjects marked like P is indeed composed of intransitive verbs. Some of the verbs such as 'want' and 'know' seem to be best analyzed as minor class bivalent verbs, while others such as 'be.cold' and 'be.hungry' are more easily classified as intransitives (van Gijn, 2006, 162-166). While a split intransitive system is coded for Yurakaré in the structural questionnaire, the minor class is clearly restricted to only a handful of verbs and is quite marginal in the grammar.

4.2.4   Fluid intransitivity

Q4.1.3   *Verbs can alternate between classes depending on semantic considerations*: [1, 0]

In some languages, intransitive predicates can alternate between different marking patterns depending on the semantic property of the event encoded by the verb or that of its sole argument. This type of marking pattern has been called 'fluid-S' (Dixon, 1994) or 'fluid intransitivity' (Creissels, 2008; Witzlack-Makarevich, 2010), with the later term being adopted here.[21] In Latundê, a Nambikwaran language of Brazil, it is possible to use the marker set generally reserved for the minor class stative verbs on major class intransitive verbs if the subject is inactive, following the characterization of inactive participants put forth in Klimov (1974):

Latundê (Nambikwaran; Telles, 2002, 226)

(55)   a.   *kḭh-**tã̰***
             itch-1SG.I
             'I am scratching.'

       b.   *kḭh-**ta**-tã̰n*
             itch-1SG.II-IPFV
             'I feel itchy.'

According to Telles (2002, 224-225), the distinction between marking patterns does not rely on the control of the participant over the event, but rather of the lexical aspect (aktionsart) of the verb itself, with state (55b) or non-state ('dynamic'; 55a) being the relevant distinction.

Kamayurá, a Tupí-Guaraní language of Brazil, shows fluid intransitivity conditioned by the control that the sole participant exerts over the event. As Seki (1990, 371) describes it, "a contrast between participant in control vs. participant not in control is manifested in the use of different sets of person markers...". While few clear examples are provided to illustrate this claim in the descriptive materials, this distinction can be seen in the nominalization construction with the negative attribute marker *uma'e* in (56).

Kamayurá (Tupian; Seki, 1990, 372)

(56)   a.   ***i**-je'eŋ*   *uma'e*
             3.II-talk   NMLZ

---

[21]Witzlack-Makarevich (2010) considers fluid intransitivity distinct from split intransitivity. However, in the sample, all languages that display fluid intransitivity also display split intransitivity, with the variation of the former pattern restricted to a subclass of verbs. For this reason, the ability to select class membership based on semantic considerations (fluid intransitivity) is treated as a parameter within split intransitivity.

'the one who does not talk'

b. ***o**-je'eŋ uma'e*
3.I-talk NMLZ

'the one who doesn't talk'

In example (56a), where the subject is indexed using the minor class marker set, the participant does not have control over the fact that they cannot talk, i.e. they are mute. In example (56a), the participant has control over the act of talking, i.e. is able to talk but chooses not to.

In addition to lexical aspect and control, additional factors may affect the choice of different marking classes in languages with fluid intransitivity. Pilagá, a Guaycuruan language spoken in Argentina, has two marker sets that can index S. Set I typically refers to a subject that is a performer or source of an event, while Set II typically refers to a subject that is affected by an event (Vidal, 2008, 413). Some intransitive verbs, particularly those where S is a performer of an activity (but also others), exclusively index S with the Set I markers (Vidal's Set A; 57a). A small class of intransitive verbs exclusively index S with the Set II markers (Vidal's Set B; 57b). This minor class includes postural verbs such as 'sit' and 'stand', spontaneous bodily process verbs like 'sneeze' and 'menstruate', as well as mental process verbs such as 'think' and 'remember'.

Pilagá (Guaycuruan; Vidal, 2008, 416, 418)

(57) a. ***se**-taka-taq*
1.I-speak-PROG

'I am speaking.'

b. ***ni**-onaɣak*
1.II-be.happy

'I am happy.'

In addition to these classes of verbs that are restricted to indexing S with a particular marker set, the vast majority of verbs allow for either Set I or Set II to be used depending on the semantics of the event and its participants.[22] The distinction between marker set choice can depend on the distinctions between induced and spontaneous events, non-reflexive and reflexive events, events with high versus low intentionality, and motion events that encode movement away from or towards a reference point. The first of these pairs of situation types take Set I markers and the second of the pairs take Set II markers, as shown in Table 4.4.

---

[22]Vidal (2008, 420) emphasizes that control of the participant over the event is not a relevant semantic parameter.

	SET I MARKERS	SET II MARKERS
	Induced	Spontaneous
Situation types	Non-reflexive	Reflexive
	High intentionality	Low intentionality
	Motion away	Motion towards

Table 4.4: Fluid intransitivity in Pilagá

The distinction between motion events is illustrated in (103), but see Vidal (2008, 420-428) for further examples.

Pilagá (Guaycuruan; Vidal, 2008, 426)

(58) a. ***aw**-yelaq*
          2.I-go.back

          'You go back there.'

      b. ***an**-yelaq*
          2.II-go.back

          'You come back here.'

A number of the situation types expressed by Set II markers in Pilagá, such as reflexive situations and spontaneous events, correspond to those typically expressed by middle voice constructions (cf. Kemmer, 1993).[23] Interestingly, both sets of markers can also be used with transitive subjects. There is an additional limited set of markers that index 1st person singular and 2nd person singular P.

Pilagá (Guaycuruan; Vidal, 2008, 415)

(59) a. *yi-**aw**-eʔet*
          1-2.I-fix

          'You dress me.'

      b. *yi-**an**-qotoʕon*
          1-2.II-awaken

          'You wake me up.'

While Pilagá certainly represents a typologically interesting case due to the fact that it has multiple marker sets to index both intransitive and transitive subjects, it serves as an excellent example to illustrate how the different semantic properties of an event and its participants can factor into a fluid intransitive marking pattern.

---

[23]See section 7.3 for further discussion on middle voice and valency decreasing constructions.

4.2.5   Summary of split intransitive marking patterns

A summary of the different structural parameters related to split intransitivity is presented in Table 4.5. The table presents the alignment of the major and minor classes with transitive argument roles, and the strategy used to express the split between classes. In the column labelled S[-cont.], Table 4.5 also presents whether the members of the minor class include verbs whose sole argument typically does not exert control over an event such as 'die' or 'slip', identified as one of the major parameters in the typology of split intransitivity in Mithun (1991). The column labeled 'Split' refers to whether the marking distinction between intransitive classes is primarily realized through multiple marker sets occurring in a single marker slot, or if the markers occur in different marker slots depending on the predicate class.

LANGUAGE	FAMILY	S_{maj}	S_{min}	S[- cont.]	SPLIT
Apurinã	Arawakan	A	P	Major	Position
Baure	Arawakan	A	P	Major	Position
Lokono	Arawakan	A	P	Major	Position
Paresi	Arawakan	A	none	Minor	Marker
Yanesha'	Arawakan	A	P	?	Position
Awa Pit	Barbacoan	A	P	Minor	Marker
Tiriyó	Cariban	P	A	Major	Marker
Ikpeng	Cariban	P	A	Minor	Marker
Ika	Chibchan	A	P	Minor	Position
Pilagá	Guaycuruan	A	none	Variable	Marker
Mocoví	Guaycuruan	A	P	Minor	Marker
Yurakaré	isolate	A	P	Major	Position
Wichí	Matacoan	none	none	Major	Marker
Latundê	Nambikwaran	A	P	Major	Position
Sabanê	Nambikwaran	A	P	Minor	Position
Awetí	Tupian	none	P	Major	Marker
Emerillon	Tupian	A	P	Major	Marker
Kamayurá	Tupian	A	P	Minor	Marker
Mundurukú	Tupian	A	P	Minor	Marker
Tapiete	Tupian	A	P	Major	Marker
Nheengatú	Tupian	A	none	Minor	Marker
Sateré-Mawé	Tupian	A	P	?	Marker

Table 4.5: Summary of split intransitive verbal marking patterns

Based on the data included in this table and the information presented within section 4.2, a number of general observations can be made about the patterns of verbal marking in languages with split intransitivity:

- There is a strong tendency for the major class of intransitive verbs to show alignment with A (81.8%).

- There is a strong tendency for the minor class of intransitive verbs to show alignment with P (72.7%).

- Multiple cases attest to the fact that the major class of intransitive verbs does not always align with A, as well as multiple cases attesting to the fact that the minor class of intransitive verbs does not always align with P.

- In the cases where the minor class of intransitive verbs does not align with P, the distinction between the intransitive classes is primarily expressed through the use of different marker sets that occur in complementary distribution with the same marker slot.

- In cases where the distinction between intransitive classes is primarily expressed through the realization of argument markers in different marker slots (position-based split), these languages either index both transitive arguments in these different slots or they only index P.

- Verbs expressing an event whose sole participant does exert control over such an event vary considerably as to whether they belong to the major or minor class within specific languages.

4.3 Case study: Complex marking pattern in Hixkaryana

Hixkaryana, a Cariban language spoken in Brazil, helped to draw the attention of linguists all across the globe towards South American languages after the publication of Derbyshire (1979), one of the first descriptions of an Amazonian language in a modern structuralist framework. Much interest was given to the fact that it seems to display a number of typological rarities, most notably a dominant object-verb-subject constituent order. However, Hixkaryana also presents an interesting complex marking pattern for the verbal indexation of arguments. The original description of the use of the different argument markers in Derbyshire (1985, 188-190) is reformatted and reproduced in Table 4.6.

A\P	1	2	3	1PL.INCL	S
1		ki-	i-		ki-
2	mɨ-		mɨ-		mɨ-/o(w)-
3	ro-	o-	y-/nɨ-	kɨ-	nɨ-
1PL.INCL			tɨ-		tɨ-
1PL.EXCL		o-	nɨ-		nɨ-

Table 4.6: Original description of Hixkaryana verbal argument markers

Given this description of the argument markers, it is not immediately clear how to best characterize the marking pattern for Hixkaryana.[24] However, certain patterns in the original description point to the fact that Hixkaryana has a hierarchical marking pattern. Knowledge of other Cariban languages such as Ikpeng (44) and Tiriyó (51) helps to draw attention to the fact that Hixkaryana could also show a hierarchical marking pattern with two distinct marker sets for A and P. The primary indicator of this pattern in Hixkaryana is that whenever A or P is 3rd person, the argument marker varies for each person value of the co-argument. If one rearranges the different marker sets based on the referential status of the marked argument, as suggested in Gildea (2012, 461-462) for Proto-Carib, the marker sets presented in Table 4.7 can be identified.

Person	Set I (A)	Set II (P)	Set III (S)
1	i-	ro-	kɨ-
2	mi-	o-	o-/ow-/mɨ-
3	y-/nɨ-	-	nɨ-
1pl.incl	tɨ-	kɨ-	tɨ-
1pl.excl	nɨ-	-	nɨ-

Table 4.7: Verbal argument marker sets in Hixkaryana

A number of different examples of transitive clauses in Hixkaryana are given in (60).

Hixkaryana (Cariban; Derbyshire 1985, 191)

(60) a. ***tɨ***-*nyahma-yatxhe*
          1pl.incl.I-feed-npst.coll
          'We (incl.) will feed them.'

     b. ***kɨ***-*hananɨhɨ-yatxkonɨ*
          1pl.incl.II-teach-dist.pst
          'He used to teach us (incl.).'

     c. ***mɨ***-*onkuhto-txowɨ*
          2.I-deceive-imm.pst.coll
          'You deceived them.'

     d. ***o***-*momokɨ-yaha*
          2.II-wait.for-npst
          'He is waiting for you.'

---

[24] Derbyshire (1985, 190) notes that 1st person plural exclusive P cannot be indexed on the verb and must be realized as the free pronoun *amna*. For this reason it was excluded from Table 4.6. However, no examples of a 1st person plural exclusive P is given in the grammar, and the description in the text is unclear whether the 3rd person A prefix occurs on verb in 3→1pl.excl configurations.

e. ***kɨ****-tayma-no*
   1PL.INCL.II-push-IMM.PST
   'I pushed you.'

f. *uro* ***mɨ****-onytxa-no*
   1SG 2.I-hear-IMM.PST
   'You heard me.'

g. *woto* ***y****-ono-no*
   meat 3.I-eat-IMM.PST
   'He ate the meat.'

h. ***nɨ****-yweronɨhyama-yatxhe*
   3.I-teach.traditions-NPST.COLL
   'He teaches them the traditions.'

Looking solely at the indexation of arguments in direct and inverse scenarios in transitive clauses, two different constructions can be identified: the Set I markers are used when A is a SAP and P is 3rd person, as in (60a) and (60c), forming the direct construction; the Set II markers are used when P is a SAP and A is 3rd person, as in (60b) and (60d), forming the inverse construction. For this reason, Hixkaryana shows a clear hierarchical marking pattern in transitive constructions and is included in Table 4.3.

In local scenarios, a 1→2 configuration is indexed with the Set I marker *kɨ*- for 1st person plural inclusive, as in (60e).[25]. In 2→1 configurations, the A argument is indexed with the Set I marker *mɨ*- and the 1st person singular pronoun *uro* is obligatorily expressed, as seen in (60f). In non-local scenarios, the marker *y*- is used when P is expressed as a conominal (60g), while *nɨ*- is used when no conominal is expressed (60h).

The use of the different argument markers to index the referential argument in transitive clauses is summarized in Table 4.8.

Given the description presented in Table 4.8, the SAP>3 hierarchy has a very clear effect on the selection of which argument is indexed in the prefix slot. In local scenarios, there appears to a hierarchy of 2>1SG/1PL.EXCL, since A is indexed in 2→1SG configurations and P is indexed in 1PL.EXCL→2 configurations. However, the use of *kɨ*- '1PL.INCL.II' to index the 1SG→2 configuration, and the lack of examples for 2→1PL.EXCL configurations, obscures this pattern.

The indexation of S in intransitive clauses also shows an interesting pattern in Hixkaryana. Unlike the other Cariban languages shown above that make a paradigmatic distinction between the marking of major class and minor class intransitive verbs, Hixkaryana only distinguishes between these predicate classes in the 2nd person. The marker *mɨ*- occurs on detransitivized verbs and some motion verbs such as 'come', 'arrive' and 'go.out', as in (61b); *ow*- occurs with two verbs 'sleep' and 'weep', as in (61c); *o*- occurs on all other intransitive

_____

[25]This is a similar situation to the marking of 2→1 configurations in Ikpeng (44g).

SCENARIO	MARKER	REFERENT		CO-ARGUMENT	
		ROLE	PERSON	ROLE	PERSON
Direct	i-	A	1SG	P	3
	mi-	A	2SG	P	3
	ti-	A	1PL.INCL	P	3
	ni-	A	1PL.EXCL	P	3
Inverse	ro-	P	1SG	A	3
	o-	P	2	A	3
	ki-	P	1PL.INCL	A	3
	(amna)	P	1PL.EXCL	A	3
Local	ki-	A	1SG	P	2
	mi-	A	2	P	1SG
	?	A	2	P	1PL.EXCL
	o-	P	2	A	1PL.EXCL

Table 4.8: Hierarchical marking pattern in Hixkaryana transitive clauses

verbs, as in (61d). A number of different examples of intransitive constructions in Hixkaryana are given in (61).

Hixkaryana (Cariban; Derbyshire 1985, 191)

(61)   a. **kɨ**-*rata-no*
           1SG.III-weep-IMM.PST
           'I wept.'

       b. **mɨ**-*omokɨ-no*
           2.III-come-IMM.PST
           'You have come.'

       c. **ow**-*rata-yaha*
           2.III-weep-NPST
           'You are weeping.'

       d. **o**-*horohɨ-no*
           2.III-stop-IMM.PST
           'You stopped.'

In terms of alignment, the marker sets in Hixkaryana present an interesting case. Due to the semantic considerations discussed in section 4.2.1, the class of predicates that indexes 2nd person S with *o-* is considered the major class. The *ow-* class and the *mi-* class are considered minor classes of intransitive verbs. As can be seen in Table 4.7, all classes of S index the 1st person singular with the prefix *ki-*, which is not shared with the 1st person singular marker in either the A marking Set I nor the P marking Set II. Thus for 1st person singular S, Hixkaryana shows a tripartite alignment pattern (S≠A≠P). However, for

the 3rd person, 1st person plural inclusive and 1st person plural exclusive, S is marked identically to A, showing an accusative alignment pattern (S=A≠P). For the 2nd person, each class of intransitive verbs shows a different alignment pattern: the major class that uses *o-* shows ergative alignment since that marker also indexes 2nd person P (S=P≠A), the minor class that uses *mɨ-* shows accusative alignment since that marker also indexes 2nd person A, and the minor class that uses *ow-* shows tripartite alignment since neither 2nd person A nor 2nd person P are indexed with this same marker. The alignment across the different marker sets for the different predicate classes is summarized in Table 4.9.

PERSON	S_{major} 'stop'	S_{minor1} 'come'	S_{minor2} 'weep'
1	S≠A≠P	S≠A≠P	S≠A≠P
2	S=P≠A	S=A≠P	S≠A≠P
3	S=A≠P	S=A≠P	S=A≠P
1PL.INCL	S=A≠P	S=A≠P	S=A≠P
1PL.EXCL	S=A≠P	S=A≠P	S=A≠P

Table 4.9: Alignment of marker sets in Hixkaryana

In order for two argument roles to show alignment, they must be marked identically across the full paradigm of the marker sets. As stated in section 3.1, up to one divergent form is allowed in the coding between marker sets for them to still be considered as forming a single paradigm. Following this criterion, Hixkaryana does not display split intransitivity even though there are different classes of intransitive predicates that index 2nd person S differently due to the fact that there is only a single distinction between the classes in the marker set. This criterion also applies across different argument roles for the formation of a single marker set. As shown in Table 4.9, the marker set used to index S of major class intransitive verbs shows two distinctions with the marker set that indexes A (Set I). For this reason, the marker set that indexes the major class of intransitive verbs is considered distinct from the set that indexes A, and for the sake of the structural questionnaire and Table 4.3, Hixkaryana is coded as displaying tripartite alignment in verbal argument marking.[26]

The complex marking pattern in Hixkaryana is very similar to that found in another Cariban language of Brazil in the sample, Apalaí. Following the same analytic procedure as that applied to the Hixkaryana data, the person marker sets in Apalaí are summarized in Table 4.10) based on data in Koehn and Koehn (1986, 108-109).

Different from Hixkaryana, the S argument in Apalaí shows alignment with

---

[26]Witzlack-Makarevich (2010, section 9.7) presents an interesting proposal for the quantification of alignment in languages with a hierarchical marking pattern in indexation such as Hixkaryana and the other examples discussed in section 4.1. However, in this thesis, the one-distinction rule for paradigmaticity was applied together with the semantic considerations for argument roles as a way to diagnose alignment in a language.

Person	A	P	S	Alignment
1	y-	j-	y-	S=A≠P
2	m-	o-	o-/m-	S_{maj}≠A=P;S_{min}=A≠P
3	n-/kyn-	-	n-/kyn-	S=A≠P
1PL.INCL	s-	ky-	s-	S=A≠P
1PL.EXCL	ynan-	yna-	ynan-	S=A≠P

Table 4.10: Verbal argument marker sets in Apalaí

the A argument in at least four out of the possible five person distinctions, meeting the criterion for accusative alignment. There is also a 2nd person split in the indexation of the S argument in Apalaí, with one class of intransitive verbs (e.g. 'go', 'eat', 'come') showing alignment with A, while another class of intransitive verbs ('sleep', 'cry', 'climb') showing alignment with P. While semantic considerations are difficult to apply to distinguish intransitive verb classes in Cariban languages, as discussed in Meira (2000b), the class containing the verbs 'sleep', 'cry' and 'climb' is considered the major class, in parallel with the treatment of intransitive verbs in Hixkaryana and Tiriyó.

4.4 Conclusions

This chapter has explored two commonly occurring verbal marking patterns in South American languages that are difficult to characterize as strictly ergative or accusative: hierarchical marking and split intransitivity. As already discussed in Chapter 3, this thesis restricts the use of the term 'alignment' in verbal argument marking to the subset of transitive and intransitive arguments that are marked identically in contrast to the remaining arguments within a specific marker set, with up to one deviation. In languages where the sole argument of an intransitive clause is indexed with different marker sets and/or in different marker slots depending on the class of the predicate, alignment is coded (for the purpose of the structural questionnaire) as the subset of arguments treated identically to the major class of intransitive verbs, i.e. the class of intransitive verbs that includes those that express events with a volitional performer. The approach adopted here allows for a more refined cross-linguistic comparison between languages that show a general marking pattern that can be classified as hierarchical, split intransitive or both.

For hierarchical marking patterns, i.e. two different constructions conditioned by the scenario between co-arguments, a number of structural parameters of comparison were identified: the alignment of the marker sets that occur in the referential marker slot, the presence of a specific inverse marker, and whether the inverse construction type is also used to distinguish between 3rd person arguments in non-local scenarios. These different parameters highlight the major structural variation observable within these patterns. For example,

while Mapudungun (42) and Itonama (43) both have a hierarchical marking pattern that indicates the inverse construction with a special marker, they differ with regards to whether these same constructions distinguish between 3rd person arguments, where Mapudungun does while Itonama does not. Furthermore, they also differ with regard to the alignment of the argument marker set(s) that occur in the referential marking slot, where Mapudungun has a single marker set with a neutral alignment while Itonama has two marker sets that each show accusative alignment.

For split intransitivity, i.e. complex marking patterns conditioned by different intransitive predicate classes, a number of structural parameters for comparison were identified: the alignment of the major class of intransitive predicates, the alignment of the minor class of intransitive predicates, whether events without a volitional performer belong to the major or minor class, whether verbs can pattern with different marking classes depending on semantic considerations, and whether the differences in the indexation of the sole argument of each class of verbs is manifested primarily through separate marker sets in complementary distribution in the same marker slot or if different marker slots are utilized. The examples shown in this chapter highlight the fallacy in labeling such systems as instances of 'split ergativity' since not all languages with different marking patterns conditioned by intransitive predicate class display even a single marker set with ergative alignment, such as in Wichí or Nheengatú. While acknowledging that many languages with split intransitivity indeed divide their intransitive verbs classes along different semantic lines, the discussion focused on the structural variation between split intransitive systems rather than the semantic variation. The case of Hixkaryana in section 4.3 served to illustrate a borderline case of split intransitivity and how the methodological framework adopted in this thesis can be applied to ensure comparability across languages with these complex patterns.

As can be seen by comparing Table 4.3 and Table 4.5, many of the languages with hierarchical marking also show split intransitivity. Thus applying a single label to the language as either 'hierarchical' or 'split intransitive' is especially problematic. By decoupling the concept of alignment of a marker set from the concept of an argument marking pattern, it is possible to discuss the variation within each of these systems while also facilitating the comparison of these systems cross-linguistically.

Diachrony in verbal argument marking: A Tupian case study

Since languages evolve by descent with modification, when two languages are descended from a single ancestor, the two will inevitably share many common linguistic elements. Just as phonological systems and lexical inventories change 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 and Campbell, 1995; Gildea, 1998), or contact-induced structural change (Heath, 1984; Thomason and Kaufman, 1988). 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.

One area of grammar that can especially benefit from a comparative approach to both form and function is how a language family developed different patterns for the expression of grammatical relations. This chapter explores the development of the person markers across the Tupian language family from a typological and historical perspective to provide insights into the development of the form and function of morphemes within this grammatical system. The changes that took place in earlier stages of development of the Tupian family are analyzed using a parsimony reconstruction model applied over two classificatory proposals for the family, with additional information coming from the

presence of cognate forms in the various marker sets. These classifications are introduced and discussed in section 5.1.1.

The Tupian language family shows a high degree of internal variation regarding the ways in which arguments are expressed on the predicate. Many of the Tupian languages spoken in Rondônia only index the absolutive (S/P) arguments of independent main clause predicates. However, the branches primarily spoken outside of Rondônia show a number of different patterns, such as the indexation of the most referential argument in transitive clauses, the indexation of separate classes of intransitive verbs using different marker sets, and portmanteau morphemes that index certain argument configurations where speech act participants act on other speech act participants. Furthermore, some languages show marking patterns usually subsumed under accusative alignment, such as the indexation of only accusative arguments in Juruna or the split intransitivity without hierarchical marking and object indexation in Nheengatú.

Section 5.1 provides an overview of the Tupian languages by first presenting two classificatory proposals for the family and introducing the language sample used in the study. It then goes on to examine the typological diversity that the languages display in their verbal person marking systems.

Section 5.2 begins with 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 over 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. The discussion in section 5.3 offers a number of new proposals on the development of the verbal person marking system based on these two analyses. It is argued that the development of both the form and function of the verbal person markers, and morphosyntactic recosntruction more generally, can gain considerable insight from being examined with respect to hypotheses on the diversification of the family tree into its respective branches.

5.1 The Tupian languages: an overview

5.1.1 On the classification of Tupian languages

Seven distinct branches of the Tupian family were first identified in Rodrigues (1964), whose classification was later expanded to ten branches by considering Mundurukú, Awetí and Mawé as separate branches distinct from Tupí-Guaraní (Rodrigues, 1985b). More recent work has focused on joining the Tupí-Guaraní branch into a single clade with Mawé and Awetí, which is referred to in this chapter as Mawetí-Guaraní (Rodrigues and Dietrich, 1997; Corrêa-da Silva, 2010). 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 thus not utilized in the historical analyses in section 5.2.

Based on his reconstruction of the Proto-Tupí phonological system, Rodrigues (2007) proposes a two-way split of the family into the Eastern Branch and the 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. This classification is given in Figure 5.1, 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.
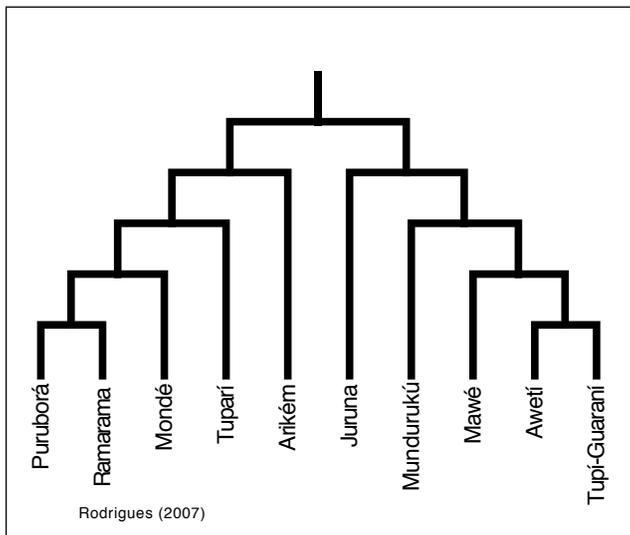
Figure 5.1: Expert classification of the Tupian family in Rodrigues (2007)

Since this classification draws on the specialized knowledge of Tupian experts who have worked on the family for half a century, it is referred to in this chapter 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 absolutive 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". Without 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. For this reason, an additional classification is also used for the parsimony analysis presented in section 5.2.

A recent classification of the Tupian family is 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. The distance matrix used as a basis for the tree is computed using the normalized Levenshtein distance (edit distance) between 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 (cf. Saitou and 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, 1985a). A cladogram of the tree is given in Figure 5.2, and the tip labels of the tree correspond to the names of the languages from which the data were gathered, rather than the name of the branch (subfamily) as in Figure 5.1. For ease of visualization, only the languages used in this study are presented.[1]
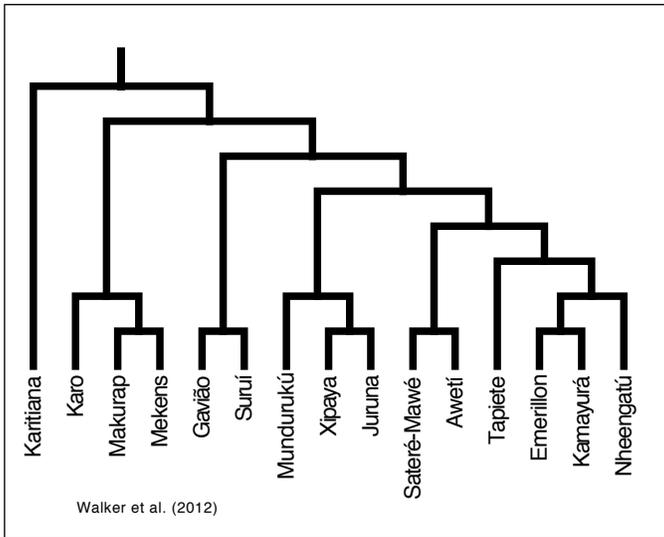
Figure 5.2: ASJP classification of the Tupian family adapted from Walker et al. (2012)

---

[1]The Emerillon language was not included in the original ASJP tree in Walker et al. (2012). Its position in the tree was inferred through its closest phylogenetic relative included in their analysis, Wayampi. Both languages belong to subgroup 8 of Tupí-Guaraní, according to the classification given in Jensen (1999).

Unlike the expert classification of the Tupian family presented in Figure 5.1, the ASJP classification has the benefit of being fully transparent and replicable, with all data and sources freely available online for inspection.[2] 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 expert classification, the ASJP classification shows a more ladder-like model of diversification of the family, without the Rondônian groups forming a distinct clade together. Both the expert classification and the ASJP classification include Mundurukú and Juruna as the nearest phylogenetic sisters 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 analyses recognize an intermediate relationship between the Mawetí-Guaraní, Mundurukú and Juruna branches, and since these branches are primarily located outside of Rondônia, the postulated homeland of the Proto-Tupí people (cf. Noelli, 1996), these branches together are referred to as comprising the 'expansionist' group of languages throughout this study. The branches of the family that primarily reside in Rondônia—Mondé, Tuparí, Arikém, Ramarama and Puruborá—are referred to as the 'Rondônian' group of languages throughout the study. Neither of these terms are meant to imply a fully resolved classification of the family and are only adopted as conventions to help explain some of the recurring patterns in the data.

5.1.2 Tupian language sample

Along with the Tupian languages in the core sample, a number of additional languages were included in this study to provide denser sampling of the language family.[3] 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 have been 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.

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 are included.

---

[2]`http://email.eva.mpg.de/~wichmann/languages.htm`

[3]Note that the Cocama-Cocamilla language included in the general sample of the thesis has not been included in this case study. This is due to the considerable debate about the origins of the language and the fact that it has undergone considerable grammatical restructuring due to contact-induced change, presumably as a result of substrate influence and other areal effects from speakers of Arawakan languages (cf. Cabral, 1995; Vallejos Yopán, 2010). Most classifications of the language family group this language with Nheengatú, both of which are considered descendants of the historical Tupinambá language spoken along the Brazilian coast at the time of the arrival of the Portuguese (Jensen, 1999, 129).

LANGUAGE	BRANCH	SOURCE [4]
Karitiana	Arikém	Everett (2006), Storto p.c.
Awetí	Awetí	Borella (2000), Drude p.c.
Juruna (Yudja)	Juruna	Fargetti (2001); Lima (2008)
Xipaya	Juruna	Rodrigues (1995)
Sateré-Mawé	Mawé	Silva (2010)
Gavião	Mondé	Moore (1984), p.c.
Suruí (Paiter)	Mondé	van der Meer (1985)
Mundurukú	Mundurukú	Gomes (2006)
Karo (Arara)	Ramarama	Gabas Jr. (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 (2003)
Kamayurá	Tupí-Guaraní	Seki (2000)

Table 5.1: Tupian language sample

5.1.3   Verbal person marking patterns

As mentioned above, the Tupian languages show a wide array of different verbal argument marking patterns. For the sake of comparison of the different alignment patterns in verbal argument marking, no distinction is made between pronominal agreement and bound pronouns, in the sense that the former can co-occur with a nominal or free pronoun while the latter cannot (Haspelmath, 2013). Bound pronominal clitics and affixes are often subsumed under the term 'person markers' in many Tupian studies. Imposing such a distinction on the data would obscure any diachronic discussion of alignment since some of the Tupian languages with an absolutive marking pattern do not allow for the verbal marker to co-occur with a realized P conominal, such that these languages would therefore be described as having an unusual marked-S system.[5]

It has been suggested that Proto-Tupí, or some stage between Proto-Tupí and Proto-Tupí-Guaraní (PTG), displayed ergative alignment in verbal argument marking through indexation of the absolutive argument with a set of prefixes (Jensen, 1998; Rodrigues and Cabral, 2012). An absolutive marking pattern, where S and P are indexed with the same marker set, is observed in main clause verbal inflection in languages of the Tuparí, Ramarama, Mondé and Arikém subfamilies.[6] Interestingly, all of these branches belong to the Western Branch of the Tupian family as proposed in Rodrigues (2007). An example of

---

[5]See section 2.1.4 and Galucio (2001, 77-79) for a discussion on the person markers in Mekens.

[6]Across the family, the set of markers that indexes P is often used to also index the object of a postposition and the possessor of a noun.

a language with ergative alignment in verbal argument marking through the indexation of the absolute grammatical relation is Wayoró, shown in (62):

Wayoró (Tuparí; Nogueira, 2011, 69-70)[7]

(62) a. ***tʃi**-pi:to-kar-a-t*   *tʃire*
       1PL.INCL-rest-VBLZ-TV-PST 1PL.INCL
       'We (incl.) rested.'

     b. ***tʃi**-po-kw-a-t*   *agopkap*
        1PL.INCL-burn-VBLZ-TV-PST fire
        'The fire burned us (incl.).'

While many of the Tupian branches show a clear ergative alignment pattern like Wayoró, languages like Sateré-Mawé, Mundurukú and many members of the Tupí-Guaraní branch show a more complex marking pattern. In these languages, S of the major class of intransitive verbs is indexed with the same paradigm of markers as A, showing accusative alignment, while S of a different class of intransitive verbs, the minor class, is indexed with the same paradigm of markers as P, showing ergative alignment. Furthermore, the indexation of A and P in transitive clauses is conditioned by which argument is ranked higher on a person hierarchy that selects speech act participants for indexation over non-speech act participants, resulting in a person hierarchy that can be roughly characterized as SAP > 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. 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 P in Mundurukú, as shown in (63):

Mundurukú transitive indexation (Mundurukú; Gomes, 2006, 48-49, 52, 74)

(63) a. *bio* ***o**=j-aoka*
        tapir 1SG.I=R1-kill
        'I killed the tapir.'

     b. *ixe* ***wuj**=dobuxikxiɡ̃*
        3SG 1PL.INCL.II=find
        'He is finding us.'

     c. *õn bio aoka-m*
        1SG tapir kill-IPFV
        'I will kill the tapir.'

---

[7]The original transcription of the authors is replicated as closely as possible. Due to varying usage of the grapheme <j>, the transcriptions have 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.

    d. *g̃asũ õn   **e**=nomuwãwã-m*
       now 1SG 2SG.II=R1.call.DUR-IPFV
       'Now I am calling you.'

Notice how in example (63a), the A argument is indexed with the Set I proclitic *o=*, while in (63b), the P argument is indexed with the Set II proclitic *wuj=*. This is due to the fact that the A argument in (63a) outranks the P co-argument on the person hierarchy, while the P argument in (63b) outranks the A co-argument since the indexed arguments are speech act participants.[8] However, verbs that are marked for imperfective aspect with the suffix *-m* never index the A argument (63c), while SAP P argument is indexed with a Set II proclitic (63d).

Additionally, Mundurukú has two classes of intransitive verbs that each take a different set of markers to index the sole argument of the clause. 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, 63).[9] For the procedural verbs, the subject is indexed with the Set I markers (S_{maj}), while the subject of a stative verb is indexed with the Set II markers (S_{min}), as seen in example (64):

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

(64)   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. *g̃axĩn ma   wujdʒu   cu-m*
          soon   really 1PL.INCL leave-IPFV
          'We are leaving really soon.'

Given the examples shown in (63) and (64), Mundurukú procedural verbs show accusative alignment (S marked like A), while the stative verbs show

---

    [8]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 P in certain local scenarios, Mundurukú always indexes P when it is a SAP (Gomes, 2006, 48). Furthermore, as shown in Table 5.2, there are a number of homophonies across the paradigms such that 1st and 2nd persons singular and 1st person plural exclusive arguments are indexed using the same form whether S_{maj}, S_{min}, A or P.

    [9]Picanço (p.c.) notes that certain verbs denoting events that involve a participant that does not exert control over the event, such as 'drown' and 'cry', are marked in the same way as the stative verbs, i.e. also belong to the minor class of intransitive verbs.

ergative alignment (S marked like P). However, S is not indexed when intransitive verbs are marked for imperfective aspect, as in (64c). Thus Mundurukú shows a split in its argument marking pattern conditioned by whether a verb is marked for the imperfective aspect or not.

Indexation patterns with splits in intransitive indexation like that shown for Mundurukú in (64a-64b) has led some scholars such as Dixon (1994) to describe alignment in these languages as 'split ergative' since one class of intransitive verbs show ergative alignment while the other shows accusative alignment. 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 P.

The Awetí language shows a split in the indexation of intransitive subjects that diverges somewhat from the pattern seen in perfective Mundurukú verbs and elsewhere in the Mawetí-Guaraní subgroup. It is also 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, 1994, 70). Awetí has three sets of argument markers: Set I is used to mark A, Set II is used to mark non-agentive S (S_{min}) and P, and Set III is used to mark agentive S (S_{maj}). Borella (2000, 131) labels the S_{min} intransitive verbs as 'descriptive verbs' and the S_{maj} class as 'active intransitive verbs'.

Awetí (Awetí; Borella, 2000, 136,140,148,155)

(65) a. *kujta-z̧a*   ***wejt**-apɨt naʔ*
          DEM(M)-COL 3.I-burn   3SG(M)
          'Those ones burned him.'

      b. *kujta-z̧a*   ***o**-az̧ũ-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 (65a) the 3rd person A argument is indexed with the Set I prefix *wejt-*, while the 3rd person intransitive subject (S_{maj}) in (65b) is indexed with the Set III prefix *o-*. Furthermore, the 1st person singular subject (S_{min}) in (65c) is indexed with the same Set II prefix *i(t)-* as the 1st person singular P in (65d). While there are considerable 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.

Many Tupí-Guaraní languages show a hierarchical indexation pattern similar to that seen for the Mundurukú perfective construction in (63) and (64). But unlike Mundurukú, these languages often have an additional marker set of portmanteau prefixes that index a 1st person A with a 2nd person P co-argument in transitive clauses, as seen for Kamayurá in (66).

Kamayurá transitive indexation (Tupí-Guaraní; Seki, 2000, 140)

(66)   a. **ere**-*etsak*
            2SG.I-see
            'You saw him.'

        b. **ne**=*r-etsak*
            2SG.II=R-see
            'He saw you.'

        c. **oro**-*etsak*
            1→2SG-see
            'I/we(excl.) saw you.'

        d. **opo**-*etsak*
            1→2PL-see
            'I/we(excl.) saw you (pl.).'

While no similar portmanteau forms occur in Awetí, Sateré-Mawé uses the prefix *moro*- to index local scenarios where 1st person A acts on a 2nd person P co-argument. Sebastian Drude (p.c. 2013) states that this morpheme is not reconstructable in Proto-Mawetí-Guaraní even though it clearly reconstructs for Proto-Tupí-Guaraní. (Corrêa-da Silva, 2010, 248) suggests that the form was probably borrowed from a Tupí-Guaraní language.

The Set II markers in Kamayurá show different types of attachment to the verb depending on the person of the marker. The SAP forms are analyzed as clitics that attach after the inclusion of the 'relational prefix', a phonological class marker that attaches to certain nouns and verbs depending on their inflection, as in (66b).[10] The 3rd person form *i*- is analyzed differently in various Tupí-Guaraní languages, but Seki (2000, 66) analyzes it as a 3rd person form of the relational prefix. Since the person hierarchy in Kamayurá does not allow for the indexation of 3rd person P arguments in transitive clauses, the 3rd person Set II form is best observed in the person inflection of minor class intransitive verbs, as seen in (118).[11]

---

[10]See Meira and Drude (2013) for a recent study on the development of the relational prefixes in Tupian languages.

[11]The use of *i*- as a 3rd person P 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,

Kamayurá minor class intransitive indexation (Seki, 2000, 157)

(67)   a. *i-'ajura*   ***i**-huku*
            3.II-neck 3.II-be.long

            'His neck is long.'

        b. *ene* ***ne**=r-oryp*
            2SG 2SG.II=R-happy

            'You are happy.'

Jensen (1998) reconstructs an additional marker set for Proto-Tupí-Guaraní that was used to index the absolute argument of dependent verbs when it is coreferential with the subject (S/A) of the matrix independent clause.[12]. In most other instances, the S_{min}/P markers are also used to express possession. This can be seen in the Kamayurá 'gerund' construction as shown in (68).

Kamayurá gerund construction (Tupí-Guaraní; Seki, 2000, 315-316)

(68)   a. *a-ke-potat*   ***we**-pytu'u-me=ran*
            1SG.I-sleep-DESID 1SG.DEP-relax-GER=CONJ

            'I want to sleep and relax.'

        b. *ja-jemo'ypy*   ***jere**-karu-m*
            1PL.INCL.I-begin 1PL.INCL.DEP-eat-GER

            'We began to eat.'

The Tupí-Guaraní language 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, as shown in (69).

Nheengatú (Tupí-Guaraní; Cruz, 2011, 142, 144, 185, 428)

(69)   a. *jande ti=**ja**-kuntai*   *portugues*
            1PL   NEG=1PL.I-speak portuguese

            'We don't speak Portuguese.'

1990, 121). This pattern of marking 3rd person P 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_{maj} and A marker sets in Awetí for 1st person plural exclusive, where the A form *oʐoj-* arose from the S form *oʐo-* attached to *i-* in the transitive. See also Monserrat (1976).

[12]In some Tupí-Guaraní languages, the same set of prefixes are used to indicate coreference between the subject of a transitive clause and the possessor of P, 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, 504))

b. *ae **u**-istragai jande*
   3SG 3SG.I-ruin 1PL

   'She ruined us.'

c. ***ja**-purasi*
   1PL.I-dance

   'We danced.'

d. *jande **jane**-akanhemu*
   1PL   1PL.II-be.scared

   'We got scared.'

Nheengatú additionally lost the inclusive-exclusive distinction in its argument markers and does not have portmanteau forms for 1st person A acting on 2nd person P, as discussed in Cruz (2011, 132-133).

A final interesting pattern can be seen in the Juruna language. Juruna was initially described as only indexing P 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 minor class of intransitive verbs, which she calls 'unaccusatives', that indeed allows for S (70b-70c) to be indexed with the same marker set as that used for P (70a), while the major class of intransitive verbs ('unergatives') do not allow for S to be indexed (70d).

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

(70)   a. *una **e**=djídaku e=bé*
           1SG 2SG=hit   2SG=DAT

           'I hit you.'

        b. *ena **e**=kũ̃ãũ̃*
           2SG 2SG=faint

           'You fainted.'

        c. *es=abéáta   **i**=ʔúrú   anu*
           2PL=clothes 3SG=be.wet ASP

           'Your clothes are wet.'

        d. *ena tahu*
           2SG run

           'You ran.'

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 can help to inform us about how the dominant patterns developed.

5.2 On the development of verbal person markers

Now that the different marking patterns across the Tupian languages have been introduced, it is possible to discuss the 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. These differences are explored with relation to the two hypotheses on the classification of the Tupian language family presented earlier, and the changes discussed are modeled over these trees using the principle of maximum parsimony. Since considerable work already exists on the development of verbal argument 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 PTG as well as those observed within the other branches of the family.

5.2.1 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, 565-573) proposes a five stage process for the development of the hierarchical indexation pattern in PTG from a putative Pre-PTG, as shown in (71):

The development of PTG indexation per Jensen (1998)

(71) 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 P is 3rd person. In these direct scenarios, both A and P are indexed. In inverse and local scenarios, where P is a SAP, only P is indexed.

d. A distinction between SAPs developed in the person hierarchy such that 1>2>3. However, it is unclear whether 1st person A was indexed with a 2nd person co-argument.

e. Portmanteau prefixes *oro-* and *opo-* are developed that index 1st person A acting on a 2nd person P co-argument.

Thus in Jensen's analysis, the S_{min}/P marker set is considered a retention 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 constructions after the extension of

the innovated marker set from marking only the subject of major class intransitive verbs to the subject of transitive clauses. Both of these key aspects of this model are further discussed in section 5.3.

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 morphophonemic irregularity than relatively newer morphemes (following Givón, 2000, 120-121). Using this logic, Gildea argues that it is unlikely that pre-PTG indexed the absolute argument on the predicate since the forms of the S/P marker set are phonetically larger than the S/A set, less bound to the predicate, and show a higher degree of cognacy with the free pronoun set. He proposes that the *i-/*c- 3rd person S_{min}/P prefix is the oldest retention, and that the system that developed for PTG arose from a predominately nominative-accusative indexation system, as shown in (72).

The development of PTG indexation per Gildea (2002)

(72) a. Predicates were indexed for 3rd person P with the *i-/c-* prefix. It is possible that that a class of intransitive verbs also used these prefixes for subject indexation.

   b. S_{maj}/A prefix set developed from a set of free pronouns that were lost prior to the development of PTG. These prefixes attach further from the stem than the *i-/*c- markers, resulting in a nominative-accusative pattern when P is 3rd person.

   c. The 1st person acting on 2nd person portmanteau prefixes developed.

   d. The S_{min}/P 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 3rd person S_{min}/P prefix *i-/*c- was retained from an earlier stage of the family history and that the proclitic forms that index the SAP arguments within the same marker set developed after the inclusion of an additional marker set that indexes S_{maj}/A.

Both Jensen (1998) and Gildea (2002) contribute many important insights into the development of PTG argument indexation. However, at the time that Jensen and Gildea were discussing the development of PTG person marking, 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.

5.2.2 On the development of different argument marking patterns

Since Jensen (1998) and Gildea (2002) propose quite different indexation systems for the early stages of Tupian family before the development of PTG, a logical starting place for this discussion on the development of this system would be to examine the observed distribution of verbal argument marking patterns in the modern languages with regard to the two hypotheses on the classification of the family. 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.[13] A parsimony analysis holds that the best model for evolutionary development is the one that requires the least amount of changes 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 (cf. Cysouw, 2009).

A comparative phylogenetic analysis of the development of verbal argument markers was carried out over the two discussed classificatory proposals using a parsimony reconstruction model. The analysis was implemented using the Mesquite software package (Maddison and 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, of the sort discussed in section 5.1.3. For the analysis using the expert classification shown in Figure 5.1, 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 5.3, including the reconstructed value for intermediate nodes of the tree.

---

[13]See Dunn et al. (2005) and Bowern (2009) for further discussion on the uses of maximum parsimony in historical linguistics.
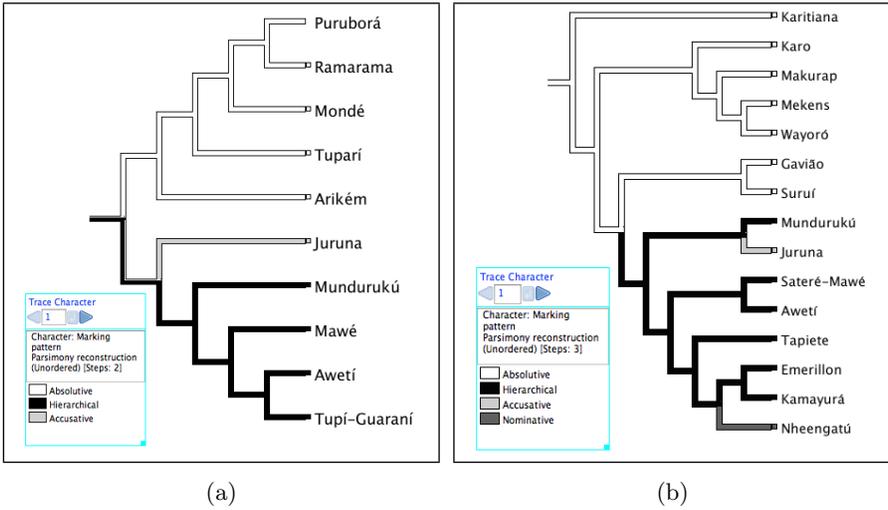
Figure 5.3: A parsimony reconstruction of marking patterns in Tupian person marking modeled over the expert classification (5.3a) and the ASJP classification (5.3b)

The analysis applied to the ASJP classifications (Figure 5.3b) clearly reconstructs an absolutive indexation pattern for Proto-Tupí. However, the topology of the expert classification (Figure 5.3a) makes it impossible to resolve a single pattern using the input data alone. 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, such as the World Atlas of Language Structures (Dryer and Haspelmath, 2011), 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 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 that the hierarchical marking pattern was retained in the expansionist languages and lost in the Rondônian branch (cf. Monserrat and Soares, 1983; Gildea, 2002).

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 P and S_{min} is coded as having two observed states: an S_{maj}/A marker set and a S_{min}/P marker set, rather than a single state 'Hierarchical'. The imperfective construction that only indexes the P argument is not considered in this analysis since it uses the same marker set as the unmarked imperfective construction, which is the more

basic of the two construction types given the criteria used for establishing the basic transitive construction in Chapter 2. This alternative analysis with the modified coding scheme is shown in Figure 5.4.

| (a) | (b) |
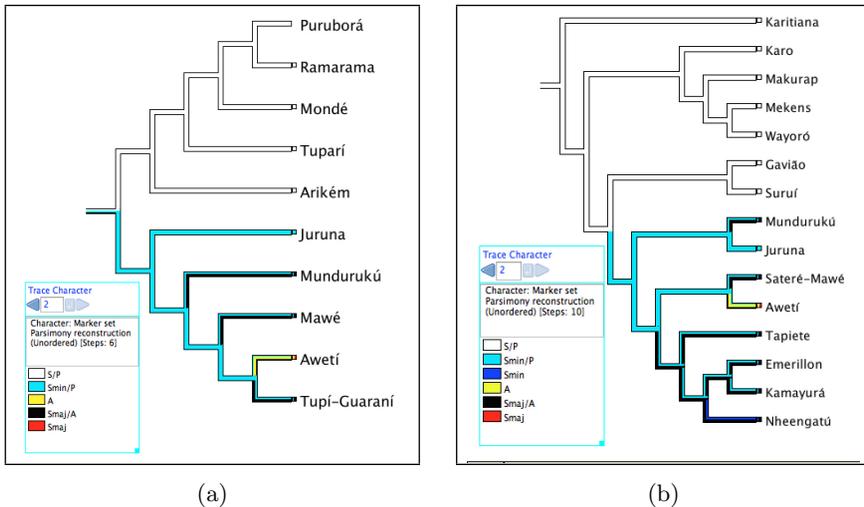
Figure 5.4: A parsimony reconstruction of marker sets used in Tupian argument marking modeled over the expert classification (5.4a) and the ASJP classification (5.4b)

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 using both classifications reconstructs an ancestral marker set for the expansionist group that indexes P with the same markers as the minor class of intransitive verbs. As is shown in section 5.2.3, traditional comparative method evidence suggests that these two sets are primarily cognate with each other rather than forming two completely independent sets, even though they have undergone a number of historical developments that obscure their true cognacy.

Some other aspects of the analysis are inconsistent with what we know from traditional comparative method reconstruction, and as such, the analyses should only be considered preliminary hypotheses from which to base more detailed work.[14] For instance, the marker set-based analysis does not reconstruct multiple marker sets for Proto-Mawetí-Guaraní, which is largely due to the three marker set configuration found in Awetí. The parsimony analyses

---

[14]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 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.

propose different diachronic pathways when the data are treated with the different coding schemes, with Figure 5.3 proposing a single development of the hierarchical pattern that was maintained in Proto-Mawetí-Guaraní, whereas Figure 5.4 proposes a development of the hierarchical system in Mundurukú independently from the development of this pattern elsewhere in the family. 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 S_{min}/P marker set being maintained throughout this history.

5.2.3   Comparison of marker sets

The different marker sets that occur in the languages of the sample are presented in Table 5.2. The argument roles indexed by each marker set are given. 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 reconstructed form for the 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 cognate with the reconstructed Proto-Tupí forms are indicated in **bold**, based on the cognate forms identified in Rodrigues and Cabral (2012, 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.[15]

---

[15]In Awetí the 1st person plural exclusive marker in all three sets *oz̥o-* is clearly cognate with the reconstructed **orʲe=* since it shows a systematic z̥:r correspondence within the rest of the Mawetí-Guaraní branch, e.g. 'bring' Awetí *ez̥ut*, Mawetí *eruut*, PTG **rur*, or the comitative causative prefix Awetí e**z̥**o-, Mawé *ere-*, PTG *ero-* (Drude, p.c.).

LANGUAGE	BRANCH	SET	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
Proto-Tupí		S/P	*o=	*e=	?	*Vj=	*orʲe=	*ej=	?
Karitiana	Arikém	S/P	i-	a-	-	ij-	ita-	aj-	-
Gavião	Mondé	S/P	o-	e-	ci-/saa-	pa-	to-	me-	ta-
Suruí	Mondé	S/P	o-	e-	a-	pa(n)=	toj=	mej=	tá?=
Karo	Ramarama	S/P	o=	e=	a?=/ŋa=	i?=	té=	karo=	tap=
Makurap	Tuparí	S/P	o=	e=	e=	ki=	te=	eki=	e=/-
Mekens	Tuparí	S/P	o-	e-	i- / s-	ki-	ose-	ejat-	tejat-
Wayoró	Tuparí	S/P	o- / m-	e-	te- / ɲ-	tʃi-	ote-	ñdʒat-	te-/ɲ-
Juruna	Juruna	S_{min}/P	u=	e=/l=	i=	s(e)=	ul(u)=	es(se)=	(i)se=/abïdai=
Xipaya	Juruna	P	u-	e-	-	se-	use-	ese-	-
Mundurukú	Mundurukú	S_{maj}/A	o=	e=	o?=	a=	otʃe=	epe=	o?=
		S_{min}/P	o=	e=	-	wuj=	otʃe=	ej=	-
Sateré-Mawé	Mawé	S_{maj}/A	a-	e-	i-	wa-	uru-	ewe-	-
		S_{min}/P	u-	e-	-	a-	uru-	e-	i?atu-
Awetí	Awetí	A	a(t)-	?e(t)-	wej(t)-	ti(t)-	ozoj(t)-	pej(t)-	wej(t)-
		S_{min}/P	i(t)-	?e-	t- / i-	kaj-	ozo-	?e?i-	t- / i-
		S_{maj}	a(j)-	?e-	o-	kaj-	ozo-	?e?i-	o-
Proto-TG	TG	S_{maj}/A	*a-	*ere-	*o-	*ja-	*oro-	*pe-	*o-
		S_{min}/P	*tʃé-	né-	*i- / c- / t-	*jané-	*oré-	*pé-	*i- / c- / t-
		S/P_{dep}	*wi-	*e-	*o-	*jere-	*oro-	*peje-	*o-
Tapiete	TG	S_{maj}	a-	ndV-	o- / wV-	ja-	o- / -ha	pV-	o- / wV-
		S_{min}/P	ʃ(V)-	ndV-	-	ñand(V)-	ir(ɨ)-...-ha	pV-	y(V)- / h(ɨ)-
Kamayurá	TG	S_{maj}/A	a-	ere-	o-	ja-	oro-	pe-	o-
		S_{maj}/P	je=	ne=	i- / t- / h-	jene-	ore-	pe=	i- / t- / h-
Emerillon	TG	S_{maj}/A	a-	ele-	o-	si-	olo-	pe-	o-
		S_{min}/P	e-	de-	i- / t-	node-/kõde-	ole-	pe-	i- / t-
Nheengatú	TG	S_{maj}/A	a-	ne-	i- / s-	ja-	ja-	pe-	tau-
		S_{min}	se-	ne-	i- / s-	jane-	jane-	pe-	o-

Table 5.2: A summary of Tupian verbal argument marker sets

A number of important observations can be made about the retention and innovation of forms within the marker sets based on the cognate forms given in Table 5.2. First, within the expansionist group of languages, Mundurukú is the only language outside of Mawetí-Guaraní to have forms that index A in transitive clauses: *oʔ=* 3rd person, *a=* 1st person plural inclusive and *epe=* 2nd person plural. None of these forms appear cognate with the reconstructed Proto-Tupí forms in Table 5.2, suggesting that they are innovations. Likely cognates of these forms are maintained in the reconstructed S_{maj}/A marker set in PTG: **o-* corresponding to *oʔ=* is retained in the 3rd person singular, and **pe-* corresponding to *epe=* is retained only in the 2nd person plural. This suggests a continuous diachronic development of a distinction in indexation between A and P that was retained from a common ancestor. Interestingly, forms similar to *epe=* occur as 2nd person plural free pronouns in many Mawetí-Guaraní languages but not in Mundurukú, where the free pronoun is *ejdʒu*.

Other forms in the Mundurukú Set I markers appear to be extended from the Set II markers, which are cognate with the Proto-Tupí forms. The reanalysis of these retained forms for 1st and 2nd persons singular and 1st 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 PTG S_{maj}/A markers were also innovated in pre-Munduruku and then partially replaced by reanalysis of the Set II forms does not appear to be supported by the available evidence since one would not expect the retention of cognate forms in the other person categories if that were the case.

The lack of clear cognacy between the Proto-Tupí S/P proclitics as reconstructed in Rodrigues and Cabral (2012) and the PTG S_{min}/P proclitics as reconstructed in Jensen (1998) can be attributed to the development of the free pronouns in PTG. The S/P markers in dependent clauses in PTG (Jensens Set 3) are true reflexes of the reconstructed Proto-Tupí absolutive marker set, at least in the singular (cf. Rodrigues, 1985a, 380;Jensen, 1998, 574). In main clauses, the indexing function of the Mawetí-Guaraní S_{min}/P set was replaced in PTG by attaching the independent pronouns to the verb, resulting in phonologically reduced clitic forms. Drude (p.c.) 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.[16] 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 2nd person singular pronoun: Proto-Mawetí-Guaraní *en* became pre-PTG *ené*, resulting in the PTG S_{min}/P marker *né=*. The other non-3rd person S_{min}/P person markers developed along similar lines.

The fact that the S_{min}/P marker set in PTG can be traced to a development

[16]Schleicher (1998, 240-243) analyzes this *e* morpheme as an accusative case suffix. Drude and Meira (p.c.) refer to this morpheme as 'Hemmauer's *e*' due to it first being recognized by Roland Hemmauer during an informal discussion.

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 3rd person prefix *i-/*c- of the same set. When examining the marker sets across the whole family, the 3rd person S_{min}/P form *i- in PTG does indeed appear to be a retention from an earlier stage of development in the family. For example, 3rd person i- is found in the Tuparí language Mekens, and a similar proclitic form $i=$ is used to index 3rd person arguments in certain focus constructions in Karo (Gabas Jr., 1999, 122-125). This form is also found in Juruna.

5.3   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 within the expansionist group of languages. The first step in the development of the marking patterns in the expansionist languages was a restriction in 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, Juruna only indexes those of the minor class of intransitive verbs. This restriction can also be seen in the Mundurukú marked 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 with 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 can be suggested that the class restrictions on intransitive indexation arose before the development of the person hierarchy as a condition on transitive indexation.[17]

As mentioned above, some of the innovative forms in Mundurukú have reflexes in both the marker set that indexed S_{maj}/A in PTG (2nd person plural, 3rd person). It is still unclear whether there existed forms in PTG that were cognate with the innovative $a=$ in Mundurukú 1st person plural inclusive S_{maj}/A marking (with the 1st person singular marker *a- being a possible candidate). As of yet, there is no strong hypothesis for how these innovative forms arose.[18] Gildea (2002) suggests that they arose through the cliticization

---

[17]It is possible that additional support for this proposal could be found through a comparative lexical analysis of the members of the different intransitive predicate classes across the expansionist languages. Such an analysis is beyond the scope of the present study and remains as a topic for future investigation.

[18]Note that the plural person markers reconstructed for Proto-Tupí by Rodrigues and Cabral (2012) shown in Table 5.2 all include the glide /j/, either as the onset of the last syllable, as in *$orje=$ '1pl.excl', or the coda of the monosyllable forms, as in *$Vj=$ '1pl.incl'

of a set of free pronouns that was 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 5.3 along with the free forms reconstructed for Proto-Tupí in Rodrigues and Cabral (2012, 549).

LANGUAGE	1SG	2SG	3SG	1PL.INCL	1PL.EXCL	2PL	3PL
Proto-Tupí	*on	*en	?	?	*or^{j}e	*ej	?
Karitiana	ɨn	ãn	i	ɨtʃa	ɨta	atʃa	i
Mekens	õt	ẽt	te	kise	ose	ejat	tejat
Suruí	oen	een	xien	paen	tojen	méjen	táen
Karo	õn	ẽn	at/ŋa	iʔtə	té	kaəto	tap
Juruna	(u)na	ena	amï	sí	(uluʔ)udí	esí	anïdai

Table 5.3: Free pronouns in some non-Tupí-Guaraní languages

As can be seen by comparing the free pronouns shown in Table 5.3 with the prefix/proclitic sets in Table 5.2 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 3rd 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 most of the innovated forms that index S_{maj}/A in the expansionist languages that mark such arguments.

A final possibility is that the marker sets identified here as innovations are not innovations at all. Rather, they could be retentions of ancestral forms that were present in the protolanguage. As Meira (p.c.) rightfully points out, if the

---

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 (cf. section 3.2.1). It is possible that the glide element results from a putative plural marker that was joined 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. Such a diachronic pathway can help to explain the difficulty in identifying clear reflexes of this reconstructed form in the daughter languages (cf. Rodrigues and Cabral, 2012, 544). If this is indeed the case for the family in general, it could be possible that the a= found in Mundurukú is a reflex of the heretofore unidentified 1st person plural inclusive person marker before the accretion of the additional plural element. While clearly speculative at this point, such a hypothesis deserves further examination.

source of a given form cannot be ascertained, then it is not methodologically clear whether or not one should prefer to see it as an innovation or a retention: both possibilities remain *a priori* equally plausible. Because of this, historical linguists tend to posit the retention of a form over its innovation in cases when no source material for the forms in question can be identified. While this conservative analysis may be methodologically safer, there is evidence that points against it in this case, namely the synchronic distribution of the argument marking patterns across the Tupian family given the topology of the family tree. Assuming the expert classification given in Figure 5.1, the presence of multiple marker sets in the root of the tree would require two losses of the second (S_{maj}/A) marker set: once in the Rondônian group of languages after the split off of the expansionist group, and once more in the Juruna branch after it split off from the remaining expansionist languages. Assuming the ASJP classification in Figure 5.2, the presence of multiple marker sets in the root of the tree would require four independent losses of the second (S_{maj}/A) marker set: once after the split off of the Arikém branch from the remaining cluster languages, once more after the split off of the Karo-Tuparí branch from the remaining cluster of languages, again after the split off of the Mondé branch from the expansionist languages and finally in the Juruna branch after the split its split off from the Mundurukú branch. But if it is so easy to lose such a system, one would expect that this would also have occurred independently in at least one of the branches of the large Mawetí-Guaraní subgroup where the relatedness of these markers is most clear. However, this is not the case, as shown in Figures 5.3 and 5.4. Since there is no additional evidence for this more complex scenario involving multiple independent losses of the marker set, such as the demonstrable retention of a fossilized cognate form in the pronominal system of one of the languages with an exclusively absolutive indexing pattern, the most parsimonious scenario is taken here to be the most plausible explanation for the development of the marker sets across the Tupian family.

5.4 Conclusions

The diversity of different verbal argument marking patterns across the Tupian language family provides an interesting case study for combining both computational and traditional techniques to make inferences about the morphosyntactic diachrony of a language family. The parsimony analysis 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 tended to maintain 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 the P argument in Xipaya and Mundurukú imperfective constructions and the indexation of both P and S_{min} in Juruna. The hierarchical

pattern found in the Mundurukú and Mawetí-Guaraní branches is the result of a single development due to the retention of cognate forms across the these groups that are not found elsewhere in the family. The PTG S_{min}/P 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 helped to confirm some previous proposals on the earlier stages of development of the person 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 comparative work on the morphosyntactic reconstruction.

CHAPTER 6

Case marking patterns

The way that languages express relations between a predicate and its arguments
through overt morphological distinctions on the arguments themselves has been
a topic of linguistic inquiry for thousands of years. The Ancient Greeks are cred-
ited with beginning the Western tradition of investigation into case, which was
later developed further by the Romans and the Arabs. Much of the terminology
used today to discuss case such as 'nominative', 'accusative', 'genitive' and 'da-
tive' were adopted from the labels for cases in Latin. An independent tradition
of thinking about case developed at the end of the Vedic period in India through
the publication of Pāṇini's grammar of Sanskrit, the *Astādhyāyī*, which focused
on the semantic relationship (*kāraka*) between a predicate and its arguments
(Blake, 2009). These different traditions and approaches highlight the fact the
case marking is a grammatical phenomenon closely linked with the semantic
relationship that an argument holds with its predicate.

In South American linguistics, case marking has featured prominently in the
description of Andean languages, especially those of the Quechuan, Aymaran,
Barbacoan and Chocoan language families. Case marking is less widespread in
the lowlands (see Chapter 8), but a number of languages in prominent families
such as Panoan, Tucanoan, Nadahupan and Macro-Jêan display case marking
of core arguments, as well as a large number of smaller families and language
isolates. This chapter begins with a discussion of a number of important con-
cepts related to case as an argument marker (section 6.1). Section 6.2 presents
an overview of the different argument marking patterns formed through case
marking. Section 6.2.3 examines some splits in the case marking of arguments
conditioned by whether the argument is an NP or a pronoun. Section 6.2.4
discusses variation in the marking of P arguments based on their animacy. A

brief discussion on the case marking of oblique grammatical roles is given in section 6.3.

6.1   Definitions and important concepts

Case marking is often referred to as a type of dependent marking since it expresses the relationship between a predicate (the head) and its argument (the dependent) on the argument itself through overt morphological distinctions (Nichols, 1986). These overt morphological distinctions on nouns can be expressed through affixation, adpositions, tone alternations or stem alternations (such as ablaut).

   In order to compare the case marking patterns of different languages, it is important to be able to identify the use of case marking in basic constructions. Following from the methodology introduced in Chapter 1 to identify the basic transitive construction from other constructions formed by bivalent predicates, the semantic relationship between the predicate and its arguments must be taken into account. In some languages there are multiple case markers used to index similar arguments and it is not always immediately clear which one should be considered the most basic. In the Altiplano variety of Aymara spoken in Bolivia, bivalent predicates show three different treatments of the non-subject argument: either the final vowel of the nominal is deleted (73a-73b), the argument is marked with *-ru* 'illative' (73c), or the argument is marked with *-ta* 'ablative' (73d), as discussed in Hardman (2001, 148-159) and Cerrón-Palomino and Carvajal Carvajal (2009, 189-191).

Aymara (Aymaran; Cerrón-Palomino and Carvajal Carvajal, 2009, 206-208; Hardman, 2001, 150)

(73)   a.   *qamaqi-xa **qawr[a]***   *jiwa-ya-:-na*
             fox-TOP   llama.ACC die-CAUS-PST-3→3
             'The fox killed a llama.'

       b.   *Luwisu **juph[a]***   *sata-sh-i*
             Luis   quinoa.ACC plant-PROG-3→3
             'Luis is planting quinoa.'

       c.   *naya juma-**r** uñj-sma*
             1SG   2SG-IL   see-1→2
             'I see you.'

       d.   *naya-**t**   may-t'a-si-ni-way-itu*
             1SG-ABL borrow-MOM-REFL-PROX-DIST-3→1
             'He borrowed it from me in passing.'

   The different treatment of a non-subject argument in Aymara depends on the predicate and the arguments it requires. For comparative purposes, the stem

alternation strategy shown in (73a-73b) is considered the basic case marking pattern in Aymara since it is the strategy used to mark the P argument in transitive constructions, i.e. the pattern shown by predicates that require an agent and a patient in constructions that display the additional event and argument properties associated with the transitive event type. However, it is important to note that the other non-subject arguments in (73c) and (73d) are also indexed on the verb, highlighting the need to identify grammatical relations independently for each argument selector based on semantic criteria. According to (Hardman, 2001, 148-151), the illative marker *-ru* and the ablative marker *-ta* can mark both argument ('complements') and adjuncts ('relationals') in Aymara. For this reason, arguments that are case marked with either of these suffixes are considered oblique arguments.

6.1.1   Case marking of argument and adjuncts

There is a long tradition in modern linguistics to distinguish between the case marking of participants that are required by the predicate, i.e. arguments, and the case marking of optional participants, i.e. adjuncts. This distinction is sometimes referred to as grammatical case versus semantic case (Blake, 2001, 31-32), or structural case versus inherent case (Chomsky, 1981). In this thesis, the case used to mark core argument of the predicate are referred to as CORE CASES, while the case assigned to adjuncts of the clause are referred to as OBLIQUE CASES. In this sense, any case that is used to mark the S, A, P, T or R argument in basic intransitive, transitive and ditransitive clauses is a core case. Any cases that are used to mark adjunct participants, or the oblique arguments of non-basic one-place, two-place and three-place predicates, insofar as they are distinct from the case marking of core arguments, are considered oblique cases.

   The distinction between core and oblique case marking can be illustrated with examples from Cavineña, a Tacanan language of Bolivia, in (74):

Cavineña (Tacanan; Guillaume, 2008, 123, 509)

(74)   a.   *iba=**ra**=tu*   *iye-chine*   *takure*
             jaguar=ERG=3SG kill-REC.PST chicken
             'The jaguar killed the chicken.'

       b.   *i-ke=bakwe*   *e-kwe*   *e-wane=**tsewe**   kanajara-kware*
             1SG-FM=CONTR 1SG-GEN 1SG-wife=ASSOC rest-REM.PST
             e-kwe   *tujuri=**ju***
             1SG-GEN mosquito.net=LOC
             'I was resting with my wife in my mosquito net.'

   As can be seen in (74), Cavineña marks case through the use of postpositions that cliticize to the last phonological word of their argument (Guillaume, 2008,

509). The postposition *=ra* in (74a) attaches to the A argument of transitive clauses and is considered the only core case marker in Cavineña. The postpositions *=tsewe* and *=ju* in (74b), labeled 'associative' and 'locative' respectively, are attached to adjuncts that are not required by the intransitive predicate *kanajara* 'rest', and are thus considered oblique case markers. For instances of oblique case markers that attach to arguments required by the predicate, i.e. oblique arguments, see examples (73c-73d) for Aymara in section 6.1.

6.1.2   Case marking and information structure

Languages can use a variety of strategies to organize information within an utterance in specific ways to make certain aspects of the information especially salient in relation to the rest of the utterance. It is a common strategy to draw attention to a particular entity in a discourse by marking it with special morphology, giving it a distinct intonational pattern, or by using a non-basic constituent order. This pragmatic marking of a participant in discourse can, at times, look very similar to the case marking of arguments in a clause, so it is important to be able to distinguish the two.

The pragmatic relation of a participant in a discourse is its PRAGMATIC ROLE, analogous to the semantic role or morphosyntactic argument role of a participant. The most relevant pragmatic roles for a discussion on case marking are topic and focus. Grundel (1988, 210) defines TOPIC as a pragmatic role for an entity that "the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to". The topic is often already introduced into the discourse and recoverable from context, forming the information that is already presupposed in the utterance. Since subject grammatical relations tend to express arguments that are the topic of the sentence or discourse (Keenan, 1976, 318-319), one needs to be especially careful not to conflate topic marking with nominative case marking. In contrast to the topic is the pragmatic role of FOCUS, which is the "semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition" (Lambrecht, 1994, 213). The focus role is also called the 'comment' with respect to the topic, especially in cases where the focus is not marked. In other words, the topic is the entity being discussed, i.e. "the matter of current concern" (Lambrecht, 1994, 150), while the focus is new information that comments on the topic. Of concern here are participants that treated in a distinct way because of their specific pragmatic role.

The Aymara example in (73) serves to illustrate the need to disambiguate pragmatic marking of participants (arguments or adjuncts) from case marking. Notice how in (73a), the A argument *qamaqi* 'fox' is marked with the suffix *-xa*. At first glance, the marker *-xa* is similar to the different case markers used to mark the object-like oblique arguments in (73c) and (73d). According to Hardman (2001, 170-171), the marker *-xa* does not express the relation that the argument holds with the predicate, but rather, to "link the form in stip-

ulated ways to the sentence as a whole, or, in some cases, to the discourse"
by functioning as a topic marker or an "attenuator". In Aymara, it is possible
to exclude the *-xa* marker as expressing nominative (or any other) case even
though it commonly occurs on S/A arguments due to the fact that it can also
occur on a number of different entities in the same clause, including time ad-
juncts (75a). The *-xa* marker can even be attached to the main verb (75b):

Aymara (Aymaran; Hardman, 2001, 150, 171)

(75) a. *ch^{h}armanti-**x***   *juma-**xa**-y*   *iskulya-r jut-ta-**xa***
          this.morning-TOP 2SG-TOP-PLT school-IL come-2→3-TOP

          'This morning (it was) you (who) came to school (right?).'

      b. *qawq^{h}a-ru-s*   *aych k^{h}ar-ja-:-**xa***
          how.many-IL-? meat butcher-PART-1→3.FUT-TOP

          'How many pieces should I cut the meat into?'

In Aymara, the marking of information structure through the topic marker
-xa can be distinguished from case marking of argument based on its distri-
bution. However, information structure can also affect case marking in differ-
ent ways, such as through the use of different pragmatically marked clausal
constructions. In Puinave, a language isolate of Colombia, four major clausal
constructions are identified in Girón (2008): the 'finite predicate' construc-
tion (76a) , the 'event-perspective' construction (76b), the 'object-perspective'
construction (76c) and the 'agent-perspective' construction (76d). Each con-
struction type has specific word order, case marking and indexation properties,
and entails a different information structure, with a particular argument or the
predicate itself in focus (Girón, 2008, 317). Examples of each construction are
given in (76):

Puinave (isolate; Girón, 2008, 321-322, 325, 331)

(76) a. *bɤ́n*   *a-péwãi-**at***   *ja-'ɤnók*   *ja-t-wɯɯn-jée*
          before 1SG-brother-ERG 3SG-bring 3SG-RCS-cure-PURP

          'Then my brother brought her so they could cure her.'

      b. *ói-ma*   *ja-tep-bék-dik*   *jɤ́'ɤ*   *bi-'an-jín*
          this-RPT 3SG-EMPH-finish-EV PRO.DEM 1PL-grandmother-PST

          'Like this supposedly our late grandmother finished it.'

      c. *ka-'in-**at***   *i-r̂n-pɤn-ot*   *nát*   *tɯijútot*
          3PL-mother-ERG ATR-scold-RES-PL DEM.PL children

          'It was their mother that scolded those children.'

      d. *ót-da*   *ja-mãp-a-t*   *jéde yãmsí'*
          3PL.INDEF-ASR 3SG-catch-AGT-PL DEM bocón

          'They (are the ones who) caught this *bocón* (fish).'

While the details of each construction are too complex to be discussed here at length (cf. Girón, 2008, 317-346), the condition relevant for the present discussion is how each pragmatic construction type affects the case marking of arguments. Notice how the agent participants of (76a) and (76c) are marked with the ergative case marker -*at*, while the corresponding agents in (76b) and (76d) are not marked for case. According to Girón (2008, 319) the 'finite predicate' construction shown in (76a) is the most pragmatically neutral, without drawing special focus to neither the arguments nor the event, and is thus considered the basic transitive construction for the coding of Puinave in the structural questionnaire.

These examples from Aymara and Puinave serve to highlight the need to take information structure into account when considering the case marking pattern in a language. The Aymara examples illustrate the need to consider the distribution and pragmatic function of a marker before considering it a case marker. The Puinave examples show that information structure can have a drastic effect on the realization of case markers in a language, and in order for languages to be compared systematically, the most pragmatically-neutral construction must be selected as the most basic.

6.2   Case marking patterns

Q5.1   *S is case marked*: [1, 0]

Q5.2   *A is case marked*: [1 ,0]

Q5.3   *P is case marked*: [1 ,0]

Q5.6   *Nouns show the following alignment pattern*: [accusative, ergative, tripartite, neutral]

The case marking pattern in a language refers to which arguments are marked for case and the alignment that is shown in the case marking, much like the simple marking patterns discussed for verbal argument marking in Chapter 3.

6.2.1   Intransitive and transitive case marking patterns

This section examines different case marking patterns in intransitive and transitive constructions, organized around whether the case marking shows accusative or ergative alignment. Neutral alignment patterns, where S, A and P are all treated in the same way for case marking, are not discussed here. No examples of tripartite case marking of NP arguments were observed in the sample, so this pattern is not considered in this section either (but see section 6.2.3 on case alignment in pronouns). As has been noted in Dixon (1979),

the cross-linguistic tendency is for languages with accusative alignment to case mark the P argument and those with ergative alignment to case mark the A argument. While the languages in the sample conform to this tendency, they also show a number of divergent patterns that are of typological interest.[1]

Patterns with accusative alignment: Accusative case alignment is most often found in languages that case mark the P argument in transitive clauses, in what is called a marked accusative pattern. This pattern can be seen in Tsafiki, a Barbacoan language of Ecuador, in (77):

Tsafiki (Barbacoan; Dickinson, 2002, 42, 227)

(77) a. *junni kebi i-na-sa   na   ja-na-nu-ti-e*
        then dark become-PROG-DFR child come-PROG-EVI-RPT-DECL
        'They say when it was getting dark a child was coming.'

      b. *jaa-tsan-ke   matu=te=ri   lu-ba-n   oko*
        3DIST-SMBL-do:VCL old.days=LOC=FOC red-QUAL-ST spirit
        *tsachi=la=**ka**   fi-na-man-ti-e*
        person=PL=ACC eat-PROG-SIT-RPT-DECL
        'They say in the old days the red demon was eating the people in this way.'

Notice how the P argument *tsachila* 'people' in (77b) is marked with the accusative case enclitic *=ka*, while the A argument *luban oko* 'red spirit' is unmarked for case, just like the S argument *na* 'child' in (77a). Case marking of the P argument with accusative alignment is the most common case marking pattern in the sample (18 of 32 languages with a case-marked transitive argument).

A different case marking pattern that shows accusative alignment is where both S and A are marked for case, while the P argument is unmarked. This pattern is called a marked nominative pattern. A marked nominative case marking pattern is found in Tehuelche, a Chonan language of Argentina, as shown in (78):

Tehuelche (Chonan; Fernández Garay, 1998, 347, 349)

(78) a. *ma:ger **š**   pe-k'*
        Ma:ger NOM be.seated-RR
        'Ma:ger is seated (over there).'

      b. *j-ank'o **š**   e-mta:we-k'e*
        1-father NOM 1-raise-RR
        'My father raised me.'

---

[1]See Creissels (2009) for an overview of typologically uncommon case marking patterns. Curiously, Creissels does not include any South American languages in his discussion.

    c.   *k-a:ren-š-k'n e-torpen*
         3-fetch-PS-RR 1-herd

         '(I) went to fetch my herd.'

Notice how the S argument *Ma:ger* in (78a) and the A argument *jank'o* 'my father' in (78b) are both marked by the nominative postposition *š*, while the P argument *etorpen* 'my herd' in (78c) is unmarked for case. Tehuelche is the only example of a marked nominative case marking pattern in the sample.

Patterns with ergative alignment: Ergative case alignment is found in many South American languages, especially those of Amazonia and the Northern Andes. The high number of languages with ergative alignment in case marking and indexation has led Aikhenvald (2012, 230) to call Amazonia the "most ergative area in the world". The most common case marking pattern with ergative alignment is where the A argument is marked for case, with the absolutive argument unmarked. This can be seen for Trumai, an isolate language of Brazil, in (79):

Trumai (isolate; Guirardello, 1999, 72, 256-257, 259, 261)

(79)   a. *pet'ew achĭkida*
        frog   jump

        'The frog jumps.'

    b. *fe'de   disi   kodechĭch-**ek***
        jaguar hit/kill snake-ERG

        'The snake killed the jaguar.'

    c. *ine-**k** atlat mapa*
        3-ERG pan   break

        'He broke the pan.'

    d. *kiki fa   ine-**tl***
        man hit/kill 3-DAT

        'The man beat/killed him.'

    e. *kiki-**k**   atlat kĭtĭ hai-**tl***
        man-ERG pan   give 1-DAT

        'The man gave the pan to me.'

In Trumai, the A argument is marked with an ergative case suffix *-ek* when it is an NP (79b) or a free pronoun (79c). The S and P arguments are not marked for case. However, in (79d), the patientive argument of the verb *fa* 'hit/kill' is case marked with the suffix *-tl*, which is also used to mark the R argument of basic ditransitive constructions. The fact that Trumai has two verbs that express the notion of 'hit' and 'kill' highlights the need to look at a large class of bivalent predicates when considering which argument marking

pattern is used in basic transitive constructions (cf. Haspelmath, 2011a). The verb *disi* 'hit/kill' in (79b) patterns like most other prototypically transitive 'physical effect verbs' such as *mapa* 'break' (79c) or *tako* 'bite', while *fa* 'hit/kill' patterns with the class of bivalent verbs whose patientive argument is conceived of more as "a kind of location (or location/goal) where the action is performed or the contact with the agent is created" (Guirardello, 1999, 264). This latter class of verbs tends to include events whose second participants are often poorly individuated and quite predictable due to the nature of the activity itself, such at 'eat', 'drink' and 'sew'. For this reason, the case marking pattern seen in (79b-79c) is considered the basic case marking pattern for transitive constructions in Trumai.

A different case marking pattern with ergative alignment can be seen for Northern Embera, a Chocoan language of Colombia, in (80):

Northern Embera (Chocoan; Mortensen, 1999, 49-50, 162, 165)

(80) a. *mãẽpᵾrᵾ wẽrã-ra   kʰãĩ b-e-sʰi-a*
          then   woman-ABS sleep be-PRF-PST-DECL
          'Then the woman fell asleep.'

      b. *ũmakʰĩrã-ta   mãᵾ̃-ne čʰiru-da   wã-sʰi-a*
          man-ABS.FOC this-in   brush-GOAL go-PST-DECL
          'Then the *man* went into the brush.' (emphasis in original)

      c. *tidu hũẽ-pᵾrᵾ-para mãᵾ̃-ne wẽrã-pa   kʰᵾda-ra   hãpa-de*
          inside arrive-PRS-time this-in   woman-ERG baggage-ABS canoe-in
          eda pué eru-b-a-sʰi-a
          into load have-be-IPFV-PST-DECL
          'By the time he got there (to the hunting house), the woman had all the baggage loaded into the canoe.'

      d. *wãpᵾ-ta-ta   mãᵾ̃-ne tai-pa   yerrehe-ta*
          go-PRES-PL-SUB this-in   1PL-ERG chimpanzee-ABS.FOC
          ũnũ-sʰi-da-a   ũmẽ
          see-PST-PL-DECL two
          'Going along we then saw chimpanzees, two of them.'

All core arguments S, A and P are case marked with a clearly segmentable suffix in Northern Embera. Argument roles S and P are marked with the same case suffix *-ra* (80a, 80c), while the A argument is marked with the case suffix *-pa* (80c-80d), resulting in a marked ergative-absolutive pattern. Interestingly, Northern Embera has an additional case suffix for absolutive arguments when they occupy a focused pragmatic role in the clause, the marker *-ta*. According to Mortensen (1999, 49), the marker *-ta* is used for the first time an entity is mentioned in a discourse or when it is "recalled from the background". The example in (80b) comes from a stretch of discourse between a husband and

wife about how she had been kidnapped by a demon. In the clauses preceding the example, the discourse is about the reaction of the wife to the conversation they were having about how to get revenge on the demon. Once the discourse refocuses on the husband, the NP referring to him is marked with the absolute focus suffix -*ta*.

6.2.2   Ditransitive case marking patterns

Q5.4   *R is case marked*: [1 ,0]

Q5.8   *Ditransitive constructions show the following ditransitive align-
         ment pattern*: [indirective, secundative, neutral]

This section examines the different case marking patterns observed in di-transitive constructions by comparing which arguments are case marked and the alignment that the markers display. Alignment is discussed with regard to how T and R align with transitive P, including indirective, secundative and marked neutral alignments.

Patterns with indirective alignment: Many South American languages show indirective case alignment in basic ditransitive constructions by marking the R argument, the indirect object, differently from the T and P arguments, which together are treated identically and form a direct object grammatical relation. In some cases, only the R argument is case marked while the T and P arguments are not marked for case at all. Case markers that only mark the R argument are often called dative case markers. A marked dative ditransitive case marking pattern can be seen for Leko, a language isolate of Bolivia, in (81):

Leko (isolate; van de Kerke, 2009, 298, 316)

(81)   a.   *ch'eka   min-ate   ber chika lais   choswai kulew-ra*
            yesterday see-1.PST one much good woman   Candelaria-LOC
            'Yesterday I saw a really beautiful woman in Candelaria.'

       b.   *seneng-**ki**   hu-ku-ate   dulsi*
            everyone-DAT 3PL-give-1.PST candy
            'I gave everyone candy.'

Notice how in (81b) the R argument *seneng* 'everyone' is marked with the dative case marker -*ki*, while the T argument *dulsi* 'candy' is unmarked for case, just like the P argument in (81a).

A few languages in the sample with a marked dative ditransitive case mark-ing pattern use the same case marker for R as that used for A. This can be seen for Ika, a Chibchan language of Colombia, in (82):

Ika (Chibchan; Frank, 1985, 28, 42)

(82) a. *in'gui tʃeirua-**se'**-ri wakuma-ri guako-u-na*
          one   man-ERG-TOP skunk-TOP kill-AUX-DIST
          'A man killed a skunk.'

      b. *abran-di   juan-**se'** kafé a'be   uʒ-in*
          Abran-TOP Juan-DAT coffee deliver AUX-MED-WIT
          'Abran delivered coffee to Juan.'

   Notice how the A argument *in'gui tʃeirua* 'a man' in (82a) is marked with
the case suffix *-se'*, which is the same marker used to mark the R argument *Juan*
in (82b). Since there appears to be no established tradition in the typological
literature to refer to a grammatical relation that treats A of transitive clauses
the same with R of ditransitive clauses, the case is marked according to its spe-
cific function in each of the examples. Similar case syncretism between ergative
and dative case markers are found in another Chibchan language Chimila, as
well as the isolate Puinave, all languages of Colombia. Since ditransitive align-
ment is conceived of based on the alignment of T and R with P, this is not
considered a distinct marking pattern for the purpose of this study.

   Languages can also show indirective case alignment by marking T and P
with the same case marker and R with a distinct case marker, showing a dative-
direct object case marking pattern in ditransitive constructions. This pattern
can be seen for Tsafiki in (83), with example (83a) repeated from (77b) above:

Tsafiki (Barbacoan; Dickinson, 2002, 227, 245)

(83) a. *jaa-tsan-ke   matu=te=ri   lu-ba-n   oko*
          3DIST-SMBL-do:VCL old.days=LOC=FOC red-QUAL-ST spirit
          *tsachi=la=**ka**   fi-na-man-ti-e*
          person=PL=ACC eat-PROG-SIT-RPT-DECL
          'They say in the old days the red demon was eating the people in
          this way.'

      b. *junni aman sona=**ka**   ya=**chi***
          then   now   woman=ACC 3SG=DAT
          kuwa-la-ki-man-ti-e
          give-PL-do:VCL-SIT-RPT-DECL
          'Then they say they gave him a woman.'

   Notice how in (83b) the R argument *ya* 'him' is marked with the dative en-
clitic *=chi*, while the T argument *sona* 'woman' is marked with the accusative
enclitic *=ka*, just like the P argument in (83a). Interestingly, the dative marker
=ka is homophonous with the possessive marker *=ka* leading to ambiguity be-
tween the sentence meaning 'they gave him a woman' and 'they gave (someone)

his woman'. However, Dickinson (2002, 244-245) states that the two interpretations can be distinguished by word order when context alone does not suffice since possessive constructions show a rigid possessor-possessed order.

The final logical possibility for a ditransitive marking patterns with indirective alignment, one where P and T are case marked and R is unmarked, what can be called a marked direct object pattern, did not occur in the sample.

Patterns with secundative alignment: Another common ditransitive alignment type, the secundative, treats the P and R arguments identically, but differently from T. The most frequently occurring ditransitive marking pattern involving secundative alignment is a marked primary object pattern, where P and R are case marked and T is unmarked, as seen for Kwaza, a language isolate of Brazil, in (84):

Kwaza (isolate, van der Voort, 2004, 106,112)

(84)   a. *jere'xwa 'kay-ki   natau-**'wã***
           jaguar   scratch-DECL Natal-ACC
           'The jaguar scratched Natal.'

       b. *zjwãu-**'wã** ãwã'txi-da-ki*
           João-ACC   show-1SG-DECL
           'I am showing it to João.'

       c. *zjwãu ãwã'txi-da-ki*
           João   show-1SG-DECL
           'I am showing João to him.'

Notice how the R argument in (84b) *zjwãu* 'João' is marked with the case suffix -*'wã*, just like the P argument *natau* 'Natal' in (84a).[2] In (84c) the argument *zjwãu* 'João' is the T argument, the entity being shown to someone, and is unmarked for case.

A different ditransitive marking pattern with secundative alignment is where only the T argument is case marked while P and R are unmarked for case. This is called a marked secondary object pattern and can be seen for Karitiana, a Tupian language of Brazil, in (85):

Karitiana (Tupian; Everett, 2006, 393, 411)

(85)   a. *nelson naka-o:t   ipsõɲ*
           Nelson NSAP-catch piranha
           'Nelson caught the piranha.'

---

[2]In van der Voort (2004) the object case marker -*'wã* is glossed as AO 'animate object' due to the fact that it only occurs on animate P and R arguments. This can be considered an instance of differential object marking, see section 6.2.4, and is glossed as ACC 'accusative' here.

b. *i naka-hi:t h̃ı̃m   pisɨp-**itɨ** i*
    3 NSAP-give animal meat-OBL 3

'She gave the meat to him.'

This pattern is attested in the sample for Karitiana as well as the Chapacuran language Wari' and the Bolivian isolate Movima, all of which are located along in the basin of the Mamoré River.[3] The final logical possibility for ditransitive marking patterns with secundative alignment, where P and R are case marked identically, with T case marked with a distinct marker, is called a secondary object-primary object pattern. This can be seen for Miraña, a Boran language of Colombia, in (86):

Miraña (Boran; Seifart, 2005, 65)

(86) a. *dʒomái ɨ:té-ʔi   okáhi-**kɛ***
          Dʒomái see-PRD tapir-ACC

'Dʒomái saw a tapir.'

b. *o:-**kɛ**   áhkɯ-:bɛ   bájnɛ-húı-**βɯ***
    1SG-ACC give-GCM.M.SG tobacco-SCM.tube-ALL

'He gave me a cigarette.'

Notice how the P argument *okáhi* in (86a) and the R argument *o:* 'me' in (86b) are both marked with the case suffix *-kɛ*, while the T argument *bájnɛhúı* 'cigarette' is marked with the case suffix *-βɯ*. The marker *-βɯ* is glossed as 'allative' because the same marker is also used to express the Goal argument (or adjunct) of certain motion events (Seifart, 2005, 68, Seifart, 2014).

Patterns with neutral ditransitive alignment: A large number of languages in the sample do not distinguish between the case marking of P, T and R, thus displaying a neutral alignment pattern. These types of case marking patterns are sometimes called double object constructions. Neutral ditransitive alignment is most commonly seen for languages without case marking of any core arguments, resulting in an unmarked neutral pattern, as seen in (87) for Wichí, a Matacoan language of Argentina:

Wichí (Matacoan; Terraza, 2009, 224)

(87) a. *n-kʸ ox yelata-s*
          1-buy   horse-PL

'I buy horses.'

---

[3]Other Tupian languages tend to display a marked dative pattern in ditransitive constructions, with the exception of Juruna which displays a marked secondary object pattern, where the P case marker is historically derived from the Proto-Tupí dative/locative postposition **pe* (cf. Fargetti, 2001, 133-134; Rodrigues and Cabral, 2012, 517-519).

      b. *n-wen-hum-kʸe*   *kalamelus noqsas*
          1-have-APPL-DISTR caramels   children
          'I give caramels to the children.'

    Notice how neither the P argument *yelatas* 'horses' in (87a) nor the T or R arguments in (87b), *kalamelus* 'caramels' and *noqsas* 'children' respectively, are marked for case. Interestingly, there are no basic (underived) ditransitive verbs in Wichí (Terraza, 2009, 222). All trivalent verbs are derived from the addition of an applicative suffix, such as *-hu* attaching to the transitive verb stem *wen* 'have' in (87b) to derive the verb *wenhu(m)* 'give'. Similar patterns where ditransitive verbs are only formed through applicative constructions are found in Guaycuruan languages such as Pilagá and Mocoví. Other unmarked neutral ditransitive case marking patterns are found in some Panoan and Tacanan languages, as well as the isolates Itonama and Yurakaré.

    The other logical possibility of a case marking pattern with neutral alignment in ditransitive constructions is where all three relevant arguments P, T and R are case marked with the same marker. A marked neutral ditransitive case marking pattern can be seen for Huallaga Quechua, a Quechuan language of Peru, in (88):

Huallaga Quechua (Quechuan; Weber 1996, 254, 557)

(88)   a. *čawra-qa huk sunča tinri-**ta***   *urku-n-čaw*   *ulli-yku-n*
        then-TOP one wasp jaguar-OBJ forehead-3-LOC sting-AF-3
        'Then one of the wasps stung the jaguar on its forehead.'

      b. *mačka papa-**ta**-ši*   *qara-n wamra-**ta**-qa*
         starchy potato-OBJ-EVI.ID give-3 child-OBJ-TOP
         '(The witch) gives the girl starchy potatoes.'

    Notice how the P argument *tinri* 'jaguar' in (88a) is marked with the object case suffix *-ta*, just like the T argument *mačka papa* 'starchy potatoes' and the R argument *qara* 'girl' in (88b). According to Weber (1996, 248, ft.3), this pattern is a recent innovation in subfamily the Quechuan languages to which Huallaga belongs (Quechua I).[4]

    Interestingly, similar marked neutral ditransitive case marking patterns are found in many of the languages of the upper basin of the Rio Negro in Brazil and Colombia, including members of the Tucanoan, Arawakan and Nadahupan families, suggesting that this feature may have spread through through intense language contact.[5]

---

[4]In Imbabura Quechua spoken in Ecuador, the other Quechuan language in the sample, the T argument of ditransitives is marked with the *-ta* suffix, just like P, while the R argument is marked with the suffix *-man* (Cole, 1982, 70). In Huallaga Quechua, the suffix *-man* is still used to mark the goal participant of certain monovalent bivalent verbs of motion as well as some trivalent verbs such as *kača* 'send' (Weber, 1996, 268-273).

[5]As far as the author is aware, this has not been explicitly proposed as an areal feature

6.2.3 Case alignment in pronouns

Q5.7 *Pronouns show the following alignment pattern*: [accusative, ergative, neutral, tripartite]

Until this point, the discussion of case marking patterns has focused on the case marking of NP arguments as well as arguments that can be expressed as free pronouns that show the same marking pattern as the NP arguments. However, this is not always the case, as already shown for Warao and Mekens in section 2.1.4, which show both show a neutral case marking pattern for NP arguments but accusative alignment in the case alignment of pronouns. This section examines a few further cases where NP and pronoun arguments show non-neutral alignments that differ from each other. As briefly discussed in section 2.1.4, alignment in pronouns is evaluated with regard to three parameters: a) the occurrence/non-occurrence of clearly segmentable case morphology, b) the presence of different contrastive sets of pronoun paradigms, and c) the ability for a free pronoun to be used to express a particular argument role. In this sense, the form of bound pronouns attached to the predicate are not considered when evaluating the alignment of free pronouns, only the fact that they cannot function as a free pronoun. Due to the fact that pronouns can be organized into paradigms (sets), sometimes without a base form for comparison analogous to the citation form of NPs, the term 'case alignment' is preferred over the use of case marking in pronouns, except in instances where a clearly segmentable case marker can be identified across the paradigm.

An interesting mismatch between the case marking of NPs and the case alignment of pronouns is found in Yaminahua, a Panoan language spoken in Peru and Brazil. The different case treatment of NP and free pronoun arguments in Yaminahua can be seen in (89):

Yaminahua (Panoan; Faust and Loos 2002, 21, 26, 80, 101)

(89) a. *mã̃   oi   nese-a*
         already rain pass-COMPL
         'The rain already passed.'

      b. *anõ   noko   xiki pi-ã̃-ita*
         paca.ERG 1PL.POSS corn eat-MAL-REC.PST
         'The paca ate our corn.'

      c. *koka-fe   ẽ   ka-i*
         uncle-COM 1SG.NOM go-PROG
         'I am going with my uncle.'

---

in any of the surveys of the languages of the Upper Rio Negro, but similarities in object case marking among these languages has been frequently mentioned in the literature about languages of the area such as Aikhenvald (2002, 2006b), Epps (2006), Zúñiga (2007) and Stenzel (2008). See section 8.2.3 of Chapter 8 for a formal evaluation of the geographic distribution of this case marking pattern.

   d. ẽ   *kirika ane-tiro*
      1SG.NOM book   read-POT

     'I can read the book.'

   e. *mĩshitonõ* **ea**   *fexa-ita*
      cat.ERG   1SG.ACC scratch-REC.PST

     'The cat scratched me.'

According to Faust and Loos (2002, 79), if the A argument of a transitive clause is an NP, it is marked for ergative case with the suffix *-n* or through the nasalization of the final vowel of final word in the NP, as seen in (89b) and (89e). If the S argument is expressed as an NP, it is not marked for case (89a), nor are P arguments that are expressed as an NP (89b, 89d). When a core argument is expressed as a free pronoun, one of the pronoun forms from Table 6.1 is used. Different from NP arguments, the pronouns sets show different case alignments according to whether the argument is a subject (S/A), an object (P/T/R) or an oblique argument (or adjunct).

PERSON	SET 1 (S/A)	SET 2 (P/T/R)	SET 3 (Oblique)
1SG	ẽ	ea	e
2SG	mĩ	mia	mi
1PL	nõ	noko	no
2PL	mã	mato	ma/mato

Table 6.1: Free pronoun sets in Yaminahua from Faust and Loos (2002, 79)

As can be seen in Table 6.1, the nominative set of pronouns (Set 1) appears to be composed from the same stem as the oblique set (Set 3) with the additional nasalization of the vowel. The nominative pronouns can be seen for S in (89c) and A in (89d). In first and second persons singular, the object pronouns (Set 2) are formed through the addition of a suffix *-a* attached to the same stem as the oblique set. However, for the first person and second person plural, the suffix *-a* does not occur.[6] An example of the object pronouns can be seen in (89e). An example of the oblique pronouns can be seen marking the possessor in (89b). Interestingly, none of the other Panoan languages examined in this study show the same split between NP and free pronoun arguments, but another example can be found in the extinct Panoan language Wariapano, also known simply as Pano (Valenzuela, 2000, 2004). Another interesting split in Panoan languages between the case aligment of NP and pronoun arguments can be seen in Kashibo-Kakataibo, which shows ergative alignment in NP arguments and tripartite alignment in free pronouns (Zariquiey Biondi, 2011).[7]

---

  [6]In the related Panoan language Shipibo-Conibo, a similar segmentable suffix *-a* occurs on the first and second persons singular and the first person plural pronouns expressing an absolutive argument (Valenzuela, 2003, 185).

  [7] Fleck (2010, 40-41) proposes that Proto-Pano had ergative marking of NP arguments

However, neither of these two Panoan languages have been included in the sample used in this study.

A different case alignment pattern can be seen in the free pronouns of the Jê language Timbira, spoken in Central Brazil. Timbira has an ergative case marker, the postposition *tɛ*, that occurs immediately following the A argument of transitive clauses, as shown in (90):

Timbira (Macro-Jêan; Castro Alves, 2004, 81, 92, 98, 101-102, 116)

(90) a. *pe   hũmrɛ ma tɛ̃*
          REM.PST man   DIR go
          'The man traveled.'

      b. ***ka*** *mɛ ŋkre*
          2   PL sing
          'You (pl.) sing.'

      c. *rʉpti* ***tɛ*** *rʉp kuran*
          jaguar ERG dog kill.NFNT
          'The jaguar killed the dog.'

      d. ***wa*** *i-**tɛ**=mɛ̃*   *hĩ*   *tʃet*
          1   1-ERG=PL meat roast.NFNT
          'We roasted the meat.'

      e. *rʉpti* ***tɛ*** ***iʔ**-kuran*
          jaguar ERG 3-kill.NFNT
          'The jaguar killed it.'

      f. *paʔ-**tɛ***   ***h**-akɛp*
          1PL.INCL-ERG 3-cut.NFNT
          'We cut it.'

Notice how the A argument is followed by the ergative postposition *tɛ* whether it is an NP (90c, 90e) or a pronoun (90d, 90f). The ergative postposition also indexes the person of the A argument with a prefix when it is an SAP (third person is unmarked), as seen in (90d) and (90f). In fact, this indexation of the A argument on the ergative postposition makes the expression of A as a pronoun optional. The P argument of a transitive clause can be expressed as an NP occurring between A and the verb (90c-90d) or it can be indexed on the verb without a conominal (90e-90f). Regarding alignment, Timbira can be characterized as displaying ergative alignment for NP arguments but tripartite alignment for pronoun arguments, primarily do to the fact that P cannot

and accusative alignment in pronouns, similar to the situation found in Yaminahua. His reconstruction of the pronoun sets in Proto-Mayoruna (the branch of Panoan that includes Matses) proposes that these pronouns showed ergative alignment in the singular, neutral alignment in first person plural and accusative alignment in second and third persons plural.

be realized as free pronoun while A and S can, with A pronouns additionally marked by the ergative postposition.[8]

6.2.4   Differential object marking

Q5.5   *Case marking of indefinite inanimate P arguments is:* [obligatory, variable, not possible]

As shown in the previous sections of this chapter, case marking patterns can vary in a particular language due to a number of different factors. In section 6.1, case marking patterns conditioned by different predicate classes are discussed, as well as ways to distinguish between the marking of grammatical relations and pragmatic roles, in order to identify the basic case marking pattern in a language that is to be compared with the other languages in the sample. Section 6.2.3 examined different case marking patterns conditioned by whether the argument is expressed as a full noun phrase or a pronoun. One final pattern to be discussed here is differential case marking patterns conditioned by the referential properties of the arguments themselves.

The term DIFFERENTIAL OBJECT MARKING (DOM) was first introduced in Bossong (1984) to describe the effect that semantic properties, especially animacy, have on the realization of case marking and indexation of direct objects in transitive clauses.[9] A number of different factors have been described as

---

[8]It is worth noting that the ergative postposition in Timbira only occurs in the simple past tense. In the irrealis mood, which includes future tense, and in the remote past tense, a tense-aspect-mood particle occurs in the same clausal position that would otherwise be occupied by *tɛ* (See also section 2.2.3). It is unclear what properties distinguish the aspectual particles from the ergative postposition beyond the fact that the latter also indexes the person of A. Following the tradition of Jêan studies, the postposition *tɛ* is tentatively considered an ergative case marker, but further comparative research in the Macro-Jêan languages could further clarify the situation. See Castro Alves (2010) for a proposal on the diachronic origins of ergative alignment in some Northern Jê languages.

A similar construction can be seen in transitive clauses of the Macro-Jêan language Bororo, where a tense-negation-mood particle occurs in the position immediately following the A argument (Crowell, 1979; Nonato, 2008). This particle also indexes the A argument if it is not expressed as an NP (1b). In intransitive clauses, this same particle is suffixed to the verb (1a). The clearly verbal nature of these particles in Bororo, where each particle has a clear effect on the interpretation of event, shows it is part of the predicate (an auxiliary) and is thus not a case marker.

Bororo (Macro-Jêan; Crowell, 1979, 68, 88)

(1)   a.   *imedɨ maragodi-**re***
            man   work-NEUT

            'The man is working (or worked).'

      b.   *a-**re**   karo bɨ*
            2SG-NEUT fish put

            'You put the fish down.'

[9]The differential treatment of A and S arguments, what Bossong (1984) calls 'differential

contributing to DOM, often under the label of 'discourse prominence', which includes the properties of definiteness and specificity in addition to animacy (Aissen, 2003; de Swart, 2007). Definiteness and specificity are difficult concepts to identify cross-linguistically without relying solely on translation equivalents rather than language-internal properties, especially since almost no languages in the sample mark definiteness and specificity through distinct grammatical morphemes like the definite articles that are so common in Indo-European languages.[10] For this reason, animacy-based variation in case marking is the focus of the following discussion and it used as the primary comparative concept for the coding of DOM in the structural questionnaire. Definiteness, as defined using language-specific criteria by the respective authors of the descriptive materials, is also considered when animacy alone does not account for variation in the case marking of P.

The fact that there is a cross-linguistic tendency for inanimate P arguments to not receive case marking in comparison to animate P arguments has to do with the functional role of case marking in the clause. Since case marking expresses the relation of an argument to its predicate, the overt morphology functions to distinguish between the two arguments in a transitive clause (which falls out of the cross-linguistic tendency for S not to be case marked). As such, when there are two arguments in a clause, one animate and one inanimate, the most natural reading is that the animate argument is acting upon the inanimate one (Hopper and Thompson, 1980; Comrie, 1989, 122-129; Aissen, 2003). The high occurrence of DOM in the sample (and cross-linguistically) is one of the primary motivations for including the animacy of P as one of the criteria for evaluating the basic transitive construction type in section 2.1.2.

The high occurrence of differential object marking in the languages of the Upper Rio Negro region of Brazil and Colombia has received considerable attention in the comparative literature on this region (Aikhenvald, 2006b; Epps, 2006; Zúñiga, 2007; Stenzel, 2008). An example of differential object marking can be seen in the treatment of P in Hup, a Nadahupan languages spoken in Brazil, as shown in (91):

Hup (Nadahupan; Epps, 2008, 175-176)

(91) a. *tã̃ʔáy̆ tɨh=tǽh-ǎn   cúʔ-úy*
          woman 3SG=offspring-OBJ grab-DYNM
          'The woman grabs her son.'

      b. *yɔ̃ʔɔ́m=ʔı̃h   tɨh dʼəh-dʼəh-yé-éh*
          powerful=MSC 3SG send-send-enter-DECL
          'He picked out (someone to be) a leader.'

---

subject marking' is not included in this study, but see de Hoop and de Swart (2008) for an overview.

[10]See Krasnoukhova (2012, 54-55) for a discussion on the occurrence of definite and indefinite articles in a comparable sample of South American languages.

    c. *tăh-**ãn**=mah   j'ám   tɨh wɔn-máh-ãh*
       tapir-OBJ=RPT DST.CNTR 3SG follow-RPT-DECL
       'He followed the tapir, long ago, they say.'

    d. *hɔhɔ́h=mah tɨh ʔey-yɔhɔ́y-ɔ́h*
       frog=RPT   3SG call-search-DECL
       'He was calling and searching for the frog.'

    e. *tã̃ʔãy   tɨh=tã̃h   cúʔ-úy*
       woman 3SG=small grab-DYNM
       'The woman grabs the small (inanimate) thing.'

In Hup, the object case marker *-ăn* is obligatory when P is a human (88a), optional when P is an animal (91c-91d), and never occurs when P is inanimate (91e). Specificity and definiteness also play a role in case marking in Hup, where a human P argument can be unmarked for case if it is both non-specific and indefinite (91b); however, human arguments that are specific and indefinite are still case marked with *-ăn* (Epps, 2008, 174).[11]

    The Nasa Yuwe language of Colombia also shows variable treatment of the P argument based on its referential properties, as seen in (92):

Nasa Yuwe (isolate; Jung, 2008, 100)

(92)   a. *wakas-weʔs-**tʲi**   uj-na   ũsʰ aʔw*
         white-COL-OBJ.PL see-PRS.PTCP be-DECL.1PL
         'We are seeing the white men.'

      b. *kutj ex-**aʔs**   vis-tʰu*
         corn field-OBJ.SG weed-DEC.1SG
         'I weed the cornfield.'

      c. *adʲ   señora-ʔ jaʔxa um-u-ʔ-k*
         1SG.M lady-TOP bag   weave-IPFV-HAB-FAC.3SG
         'My lady weaves a bag.'

    Unlike in Hup, the primary condition for DOM in Nasa Yuwe is the definiteness of the P argument. Inanimate P arguments can be case marked if they are definite (92b), but not if they are indefinite (92c). Due to the fact that both animacy and definiteness often play a role in the case marking of P arguments, as seen in examples (91) and (92), the structural questionnaire uses an inanimate and indefinite P argument as the prototype to be compared across languages in the sample.

---

[11]See Epps (2008, 170-181) for further examples and a full account of the conditions for the use of *-ăn* in Hup. Zúñiga (2007, 225) proposes that DOM in Hup has developed through close contact with the Tucanoan speakers of the region, many of whose languages also show DOM. The related Nadahupan language Dâw spoken outside of this region does not appear to display DOM (Martins, 2004, 499).

6.2.5 Summary of case marking patterns

Table 6.2 presents a summary of the languages in the sample that display case marking of core arguments in transitive and intransitive constructions. Only languages with a marked alignment type are included in the table. The columns S, A and P refer specifically to the case marking of NP arguments. In the column P, the positive value is enclosed with parentheses for languages that display differential object marking. The columns Align_{NP} and Align_{Pro} refer to the alignment found in intransitive and transitive constructions for the case marking of NP and pronoun arguments, respectively. As throughout this study, alignment is calculated as subsets of identically and differently treated arguments. Some language names have been abbreviated for the sake of space.

LANGUAGE	FAMILY	S	A	P	ALIGN$_{NP}$	ALIGN$_{Pro}$
Tariana	Arawakan	-	-	(+)	Accusative	Accusative
Jarawara	Arawan	-	-	+	Accusative	Accusative
Aymara	Aymaran	-	-	+	Accusative	Accusative
Jaqaru	Aymaran	-	-	+	Accusative	Accusative
Awa Pit	Barbacoan	-	-	(+)	Accusative	Accusative
Tsafiki	Barbacoan	-	-	(+)	Accusative	Accusative
Chayahuita	Cawapanan	-	+	-	Ergative	Ergative
Chimila	Chibchan	+	-	-	Ergative	Ergative
Ika	Chibchan	+	-	-	Ergative	Neutral
N. Embera	Chocoan	+	+	+	Ergative	Ergative
Tehuelche	Chonan	+	+	-	Accusative	Accusative
Kwaza	isolate	-	-	(+)	Accusative	Accusative
Nasa Yuwe	isolate	-	-	(+)	Accusative	Accusative
Puinave	isolate	+	-	-	Ergative	Ergative
Trumai	isolate	+	-	-	Ergative	Ergative
Aguaruna	Jivaroan	-	-	+	Accusative	Accusative
Katukina	Katukinan	+	-	-	Ergative	Ergative
Timbira	Macro-Jêan	-	+	-	Ergative	Tripartite
Dâw	Nadahupan	-	-	+	Accusative	Accusative
Hup	Nadahupan	-	-	(+)	Accusative	Accusative
Sabanê	Nambikwaran	-	-	(+)	Accusative	Accusative
Matses	Panoan	+	-	-	Ergative	Ergative
Shipibo	Panoan	+	-	-	Ergative	Ergative
Yaminahua	Panoan	+	-	-	Ergative	Accusative
Huallaga Q.	Quechuan	-	-	+	Accusative	Accusative
Imbabura Q.	Quechuan	-	-	+	Accusative	Accusative
Cavineña	Tacanan	+	-	-	Ergative	Ergative
Cubeo	Tucanoan	-	-	(+)	Accusative	Accusative
Desano	Tucanoan	-	-	(+)	Accusative	Accusative
Juruna	Tupian	-	-	+	Accusative	Accusative
Miraña	Boran	-	-	(+)	Accusative	Accusative
Yanam	Yanomaman	+	-	-	Ergative	Ergative

Table 6.2: Summary of case marking patterns in transitive and intransitive constructions

Based on the data included in Table 6.2 and the information presented in the previous sections, a number of general observations can be made about the case marking patterns in the languages of the sample, including:

- For languages that case mark NP arguments, the South American languages used in this study show a slightly higher distribution of ergative alignment with regard to accusative alignment than the global sample in the WALS database (Comrie, 2011). However, this distribution does

not show a statistically significant difference between the South American languages used in this study and the non-South American languages in the WALS sample ($p=0.377$; one-tailed Fisher's Exact Test). Languages with neutral case alignment are not considered in this calculation, nor languages labeled as 'tripartite' and 'active-stative' in WALS.

- Only a single language in the sample, Northern Embera, goes against the generalization in Dixon (1979) that languages with ergative alignment in case marking do not overtly mark the absolutive argument.

- Within families, all languages either share the same case marking pattern with core NP arguments or some languages have an unmarked neutral pattern in contrast to a single marked pattern (such as Tariana in comparison with other Arawakan languages).

- Ditransitive case marking patterns are much more variable across language families than transitive case marking patterns, e.g. Karitiana and Huallaga Quechua.

- The presence of differential object marking is primarily confined to Western Amazonia and the Barbacoan languages, with Kwaza and Sabanê being notable exceptions in Southern Amazonia.

6.3   Oblique case marking

Q5.9   *Obliques are case marked*: [preposition/prefix, postposition/suffix, neither]

The South American languages in the sample present a diverse array of different strategies to case mark oblique participants of the clause. When discussing oblique cases, it is often useful to distinguish between spatial cases, e.g. those expressing semantic roles of location, goal or source, and those expressing non-spatial relations such as instrument or accompaniment (Haspelmath, 2009). Unlike core cases, which for all languages in the sample S, A and P are marked with suffixes or postpositions, oblique cases tend to vary as to whether they employ prefixes/prepositions or suffixes/postpositions. An example of a language that uses prepositions to mark oblique case is Mocoví, a Guaycuruan language of Argentina, in (93):

Mocoví (Guaycuruan; Grondona, 1998, 135, 139, 169)

(93)   a. *lwis i=ilew*
            Luis 3.II=die
            'Luis died.'

    b. *i=ilew **ke** ñi   n=atarenataganagaki*
        3.II=die OBL DEM.seated NPOSS=hospital

        'He died in the hospital.'

    c. *s=anat=ñi   **ke** da   qoʔpaq*
        1.I=fall=down OBL DEM:standing tree

        'I fell down from the tree.'

The predicate *ilew* 'die' in Mocoví only requires a single argument, a patient, that is not case marked and can be indexed on the verb with the Set II proclitics (also used for P) as seen in (93a). An additional adjunct participant, such as the location of the dying event (93b) can be included in the clause marked with the oblique preposition *ke*. The source of a motion event, such as in (93c) for the predicate *anatni* 'fall down', can also be marked with the oblique preposition *ke*. The use of prepositions to mark oblique cases is rare in the sample, occurring only in Mocoví, the Chapacuran language Wari' and the Macro-Jêan language Guató.

An additional feature of oblique case marking in Mocoví is that there is only a single preposition *ke* that marks all oblique relations that can be expressed as an adjunct.[12] The single generalized oblique marker in Mocoví can be contrasted with the rich set of postpositions found in Urarina, an isolate of Peru, as seen in (94):

Urarina (isolate; Olawsky, 2006, 225, 232, 242)

(94)   a. *nii lureri **asae** ni-akwa-e barꭡe itulere inio-kꭡrꭡ*
          that house under be-DISTR-3 masato all.kinds meat-PL

          'There were masato, all kinds of things, different types of meat in that house.'

    b. *lureri **kahe** hwa-ꭡ̃*
        house from descend-1SG

        'I have descended from the house.'

    c. *hwãa **ajɲa** manoel=te   nꭡkꭡe **kꭡane** kꭡ-re=ĩ*
        Juan with Manuel=FOC creek inside go-IRR.3=ASR

        'Manuel will go over the river with Juan.'

Urarina has a large number of different postpositions that can be used to express a number of oblique relations of adjuncts such as location (94a) and source (94a), both relations that were marked by the same preposition *ke* in Mocoví in (93). There are many other postpositions given in Olawsky (2006, 224-225),

---

[12]Instrument participants, for example, cannot be expressed as an adjunct of the clause in Mocoví. They are expressed using two juxtaposed clauses, such as for the example discussed in Grondona (1998, 168-169) translated as 'I rubbed Juan's chest with the liquid medicine' is in fact composed of two clauses: 'I use the liquid medicine' and 'I rub Juan's chest.'

of which the comitative postposition *ajna* 'with' and the locative/directional postposition *kʉane* 'inside' are shown in (94c).

A number of South American languages show oblique case marking systems that only include two case markers. When only two markers occur, one most often has a locative function, expression various semantic roles related to location and direction, and another marker that has an instrumental or comitative function (or both). A two marker system of oblique case marking can be seen for Desano, a Tucanoan language spoken in Brazil and Colombia, in (95):

Desano (Tucanoan; Silva, 2012, 160, 165-166)

(95) a. *igo   ba-di-soda-do-**ge**   doa-bõ*
          3SG.F eat-NMLZ.INAN-cook-NMLZ.ABSTR-LOC sit-3SG.F

          'She is sitting in the kitchen.'

      b. *adi-a   yʉʉ abe-wii-**ge***
          come-PRF 1SG sun-house-LOC

          'I came from the house of the sun.'

      c. *gahi   bũdũ-yʉkʉ-**bẽdã**   adi-yũ-bĩ*
          another tobacco-tree-COM/INS come-EVI.QUOT/FOLK-3SG.M.IPFV

          'He came with another tobacco plant.'

      d. *ĩgʉ̃   yee-gʉ-**bẽdã**   keo-gãdĩ-gʉ̃*
          3SG.M jaguar-CLF.TRUNK-COM/INS measure-advance-3SG.M
          i-yũ-bĩ
          do-EVI.QUOT/FOLK-3SG.M.IPFV

          'He came measuring with his sacred cane.'

As can be seen in (95), both location participants (95a) and source participants (95b) are marked with the case suffix *-ge*, while accompaniment (or comitative) participants (95c) and instrument participants (95d) are marked with the case suffix *-bẽdã*. Two-term oblique case marking systems are found in other Tucanoan languages, as well as Nadahupan and Nambikwaran langauges. While Arawakan languages tend to have multiple locative and directional case markers in addition to distinct markers for instrumental and comitative functions, Tariana, which has been in close contact with Eastern Tucanoan languages through involvement in the linguistic exogamy system practiced throughout the Upper Rio Negro region, has a two-term system functionally similar to the Tucanoan pattern (Aikhenvald, 2003a, 148-154). Aikhenvald (2006b) attributes this pattern in Tariana to structural diffusion under areal influence.

In summary, South American languages overwhelmingly mark oblique cases through suffixes or postpositions rather than prepositions. Within the sample, 64 languages employ either suffixes or postpositions, while only three languages—Wari', Mocoví and Guató—employ prepositions to mark obliques.

Interestingly, these languages are all spoken in the Chaco-Planalto or Southern Amazonia regions. Six languages do not mark obliques with any additional morphology.

6.4   Conclusions

Case marking is one of the primary and most well-studied aspects of argument marking in human language. This chapter has presented an overview of the different argument marking patterns found in the sample of South American languages and discussed a few important topics related to identifying the basic construction in a language for case marking, namely distinguishing between predicate-conditioned case marking patterns (section 6.1), between core and oblique case (section 6.3), and between case marking and information structure marking (section 6.1.2). Alignment in case marking was also discussed with regard to transitive and ditransitive alignment patterns (sections 6.2.1-6.2.2), possible mismatches between case alingment in NP arguments and pronouns (section 6.2.3), and the differential treatment of P arguments based on their referential properties (section 6.2.4). A few parameters specifically related to the case marking of oblique participants, i.e. adjuncts, were discussed in section 6.3.

A few topics that are of typological interest have not been discussed here, such as the effect of scenarios and argument configurations on case marking, as well as split intransitivity. These topics have been more thoroughly discussed with regards to verbal argument marking in Chapter 4 since they are only found in the case marking systems of a few languages in the sample, but the general conditions are the same.[13] A more in-depth investigation of these complex patterns remains a topic of further research.

---

[13]See Witzlack-Makarevich (2010, 150-154) for an interesting discussion on the effect of scenarios on the pronoun system in Aguaruna.

CHAPTER 7

Valency changing strategies

The previous chapters of this thesis have focused on the structural variation that languages exhibit with regard to the treatment of clausal participants by two types of argument markers: indexation and case marking. This chapter examines the strategies that the languages in the sample employ in order to alter the valency of a predicate. VALENCY CHANGING STRATEGIES can either increase the valency of a predicate, redefine the semantic role of an argument that the predicate requires, or decrease the valency of a predicate.

Just as South American languages display a wide variety of different argument marking patterns, they also employ a large number of different strategies to derive verb forms that display distinct valency frames from their underived counterparts. Section 7.1 introduces a number of important concepts used in the study of valency-changing strategies and outlines the approach used in this thesis to compare these structures across the sample. Section 7.2 provides an overview of the strategies used to increase the valency of a predicate, either by introducing a new participant into the A argument role (causatives; 7.2.1) or by introducing a new participant into the P argument role (applicatives; 7.2.2). Section 7.3 provides an overview of the strategies used to decrease the valency of a predicate, either by demoting the agentive participant to an oblique grammatical status and introducing the patientive participant as the S argument (passives; 7.3.1), or by demoting the patientive participant to an oblique status (antipassives; 7.3.2). Further valency decreasing strategies such as reflexive constructions, reciprocal constructions and middle voice are discussed in sections 7.3.3-7.3.4.

7.1 Definitions and important concepts

The VALENCY FRAME of a predicate is the relationship between the semantic roles of the arguments it requires and the way that these arguments are treated in the morphosyntax. Valency frames have also been called the diathesis of a predicate (Mel'cuk, 1994), and are similar to the concept of the subcategorization frame of a predicate in generative grammar (Chomsky, 1965). Certain predicates in some languages can alternate between valency frames without any marked morphosyntactic derivation of the predicate, such as in the classic ambitransitive alternation in English *John broke the stick* and *The stick broke* (Levin, 1993). Adopting the terminology from The Leipzig Valency Classes Projects Database Questionnaire Manual,[1] this type of alternation between valency frames is called an UNCODED ALTERNATION. Such alternations are not considered here since this distinction is not overtly marked. Of primary concern for this chapter and the structural questionnaire are CODED ALTERNATIONS, alternations between different valency frames of a predicate by means of overtly expressed morphosyntactic devices. A number of different morphosyntactic devices can be used to alter the valency frame of a predicate, such as the use of auxiliaries or verbal particles, but the focus of this chapter will be on bound morphology attached to the predicate, with other strategies discussed when relevant.

The primary means to change the valency frame of a predicate is through argument promotion and demotion. Van Valin Jr. (1980, 316) defines ARGUMENT PROMOTION as a "change in the syntactic status of an NP such that it becomes accessible to one or more grammatical processes which it could not otherwise undergo, for example, relativization." Argument promotion can affect the argument marking of both core and oblique participants in different ways: a) oblique participants can be promoted into core argument roles such that they can be selected by argument markers such as case and indexation, and b) there can be a change in the syntactic status of core prticipants such that they can be selected by argument markers associated with other argument roles than those typically assigned to their particular semantic roles in the basic construction. In the second case, the concept of promotion is not meant to imply an abstract hierarchy of grammatical relations, e.g. in the sense of the NP accessibility hierarchy (Keenan and Comrie, 1977), but rather, a distinction between the accessibility and non-accessibility to specific argument selectors. Correspondingly, ARGUMENT DEMOTION is a change in the syntactic status of a participant such that it is no longer accessible to argument selectors associated with the argument role typically assigned to its semantic role in the basic construction. Demoted arguments are generally only optionally included in the clause, often with a morphosyntactic treatment corresponding to that of an oblique participant. The semantic role of a demoted argument is sometimes still implied or recoverable in the utterance, as in agentless pas-

---

[1] http://www.eva.mpg.de/lingua/valency/pdf/DatabaseQuestionnaireManual.pdf

sive constructions (see section 7.3.1), even though it is no longer obligatorily expressed. However, since valency is defined in this thesis as the participants of a clause that must be expressed in order to form an acceptable utterance, argument demotion is assumed to result in a reduction of the valency of a predicate.[2] In addition to the change in syntactic status of a participant, argument promotion and demotion also affect the pragmatic status of the participants. Argument promotion and demotion are generally associated with an increase or decrease, respectively, of the relative topicality of the participants in the clause (Givón, 2001, 83). However, the use of these terms here refers specifically to their morphosyntactic effects on clause participants, with the pragmatic effects considered as a result of the change in their syntactic status.

Certain coded alternations of valency frames are sometimes referred to as difference types of VOICE in a language, such as passive or antipassive alternations, but this concept often excludes construction types that introduce new arguments into the valency frame of a predicate such as causative and applicative constructions (Mel'cuk, 1994, 11). It has been argued that voice is inflectional in nature since it does not alter the propositional meaning of the utterance, i.e. the expressed situation type (Mel'cuk, 1994; Shibatani, 2004). However, this criterion of propositional alternation is somewhat difficult to delimit and apply cross-linguistically. Furthermore, morphological polysemy between categories that are traditionally considered inflection (voice) and those that are traditionally considered derivation (valency change) further complicates this distinction (see section 7.3 for further discussion). Other scholarly work such as Dixon and Aikhenvald (1997, 2000a) has argued that all coded alternations in valency frames are syntactic derivations. In order to present a more inclusive look at the grammatical phenomenon of valency change, all different strategies that languages use for coded alternations of valency frames are treated together in this chapter.

The valency frame of a predicate has two components: the semantic relationship of the clausal participants to the predicate, i.e. the situation type of the construction, and the way that these different participants are treated in the morphosyntax, i.e. the argument marking pattern of the construction. In this chapter, the different valency-changing constructions used in different languages are defined according to the semantic situation type that they represent and the effects that they have on the valency of a predicate viz. the underived valency frame of a corresponding predicate in basic constructions. By using these semantic criteria as a standard of comparison, it is possible to investigate the diverse argument marking patterns that are produced by such

---

[2]As Comrie (1993, 906-908) points out, the obligatoriness of expression of a participant is sometimes inadequate as the sole criterion for argumenthood, especially in languages that allow for NPs to be freely omitted when recoverable from the discourse. However, in instances of argument demotion, the semantic role of the demoted argument is often recoverable from the nature of the event itself rather than the discourse, much like how in an utterance such as *He cut the bread* in English implies that an instrument was used to carry out the event even though no instrumental semantic role is an argument of the predicate.

constructions across languages with different morphosyntactic profiles.

7.2 Valency increasing strategies

This section explores major parameters of variation in two commonly occurring valency increasing constructions: applicatives and causatives. In CAUSATIVE CONSTRUCTIONS, a new participant is included in the valency frame of a predicate corresponding to the A argument role. In APPLICATIVE CONSTRUCTIONS, a new participant is included in the valency frame of a predicate corresponding to a non-subject argument role.

7.2.1 Causatives

Q6.1   *Causative constructions are marked by verbal morphology:* [1, 0]

Q6.1.1   *Causative constructions can only be derived from intransitive verbs:* [1, 0]

Q6.1.2   *The causee in transitive-derived causative constructions is treated as:* [Oblique, R, T, A]

Q6.1.3   *Direct and indirect causation are formally distinguished:* [1, 0]

Q6.1.4   *Direct and sociative causation are formally distinguished:* [1, 0]

The defining property of a monoclausal causative construction is that is promotes a new participant into the argument role of transitive subject (A), increasing the valency of the base verbal form by one (Dixon, 2000, 30). The causative situation type can be conceived of as involving two component situation types: a causing situation and a resulting (or 'caused') situation (Shibatani, 1976).[3] When discussing causative constructions, it is useful to use a number of specific semantic roles to refer to the different participants in the causative situation. The CAUSER is the initiator or controller of the causing situation, while the CAUSEE is the participant affected by the causing situation but is also a participant in the resulting situation. The role of the causee is primarily

---

[3]Some scholars such as Cole (1983) consider the resulting situation as a dependent clause, facilitating the application of rules for the assignment of case in a language. Such a biclausal structure is not assumed for any languages in this thesis, and it is worth noting that the empirical basis for assuming such a structure can be problematic in languages the form causative constructions through a single predicate with additional derivational morphology. As such, a more theory-neutral and cross-linguistically applicable approach is to use the semantic composition of the utterance involving notional situation types, in the sense of Talmy (1976), rather than impose an abstract syntactic structure on the language that may not be independently supported (and is curiously similar to English).

as a performer of the resulting situation, but it can also undergo a change of state as a participant in the resulting situation.

Causation can be expressed through a number of different strategies in a language, such as lexical pairs like *die/kill* ('cause to die'), biclausal (analytical) constructions, and through morphological devices where a predicate expressing a causative event is derived from a base verbal form through the addition of specialized morphology (Comrie, 1989, 159-164). This section is primarily concerned with the third causative strategy, which are often called either morphological or SYNTHETIC CAUSATIVE CONSTRUCTIONS, with the latter term adopted here. Within monoclausal causative constructions, those composed of multiple independent predicating elements, i.e. expressed solely through a complex predicate, were not considered as morphological causative constructions (cf. Song, 2011).[4]

Dixon (2000, 62) identifies nine different parameters of variation that are relevant when discussing causative constructions, ranging from the transitivity of the verbal base to the semantic properties of the causer and causee. Since this discussion focuses on causative derivations targeting basic transitive and intransitive verbs, the treatment of minor predicate classes are not considered, thus allowing these parameters to be condensed into three major issues that will be discussed for each causative construction: a) the transitivity of the verbal base form, b) the grammatical treatment of the causee participant, and c) the semantic properties of the causer in relation to the causee and the causation situation.

Let us first consider the causative constructions found in Urarina, an isolate language from Peru. In example (96), a causative construction is formed with an intransitive verbal base through the its use with the suffix -*a*:

Urarina (isolate; Olawsky, 2006, 618)

(96) a. ɨ-*a*
         come-3
         'He came.'

      b. *katɕa* ɨ-***a***-*a*
         man   come-CAUS-3
         'He (himself) made the man come.'

The causative construction formed with the marker -*a* implies that the causer participant was directly involved in the causation of the resulting situation, along the lines of what is typically referred to as DIRECT CAUSATION. Direct causation most often involves the physical manipulation of the causee by the causer, whereby the causee does not act as a volitional entity in the

---

[4]For examples of monoclausal causative constructions formed through complex predicates, see Epps (2008, 398-404) for Hup and Everett and Kern (1997, 317-318) for Wari'.

execution of the resulting situation (Shibatani and Pardeshi, 2001, 89). Direct causation is taken as the prototypical case of causation in the coding of causative constructions into the typological database, although other types of causative constructions are also considered.

Urarina has a second causative construction derived from either intransitive (97a) or transitive bases (97b) with the addition of the marker -*erate*, as seen in (97):

Urarina (isolate; Olawsky, 2006, 612, 614)

(97)   a. *kanaanaj-ɄrɄ amia-**rati**-a   katɕ*
           child-PL   work-CAUS2-3 man
           'The man made the children work.'

       b. *aka **ke**   kutia-**rate**-kore katɕa-ɄrɄ*
           3SG OBL call-CAUS2-3PL man-PL
           'The people made him call her (his wife).'

Notice that the causee participant in (97b) is treated as an oblique marked by the postposition *ke* in the causative construction formed with the marker -*rati* when it targets a transitive verbal base. Comparing the situation types expressed by the different causative constructions in (96) and (97), there is a semantic distinction between the causer participants in (96b) and (97) with regard to their involvement in the causation of the resulting situation. While the causer in (96b) is conceived of as being directly involved in the situation, i.e. by physically ensuring that the man arrives, the examples in (97) imply that the causer was only indirectly involved in carrying out resulting situation, such as through commanding or coercing the causee. Causation events where the causee remains a volitional performer of the resulting situation without the direct involvement of the causer is commonly known as INDIRECT CAUSATION. Shibatani and Pardeshi (2001, 89) consider direct and indirect causation as two prototypical cases on opposite ends of a continuum between more patient-like and and more agent-like causees. As seen in the difference between the -*a* and -*erate* causative constructions in Urarina, this distinction in the semantics of involvement is commonly grammaticalized into different construction types.[5]

Languages that have morphological causative constructions that target a transitive verbal base can manifest the distinction between direct and indirect causation through the treatment of the causee participant rather than with a specific indirect marker. This distinction can be seen for Cavineña in (98):

---

[5]The distinction in Urarina between direct and indirect causation is restricted to constructions derived from an intransitive verbal base, due to the fact that the -*a* construction cannot target transitive bases. For causative constructions derived from a transitive base through the addition of -*erate*, the interpretation of the degree of involvement of the causer in the resulting situation is determined from context (Olawsky, 2006, 620).

Cavineña (Tacanan; Guillaume, 2008, 292-293)

(98) a. *ebakwa=ra=tu   misi   ara-wa*
          child=ERG=3SG tamale eat-PRF
          'The child ate tamale.'

      b. *e-puna=ra=tu   ara-**mere**-wa misi   tu-ja ebakwa*
          NPF-female=ERG=3SG eat-CAUS-PRF tamale 3-GEN child
          'The woman fed the child tamale (i.e. she put the tamale in the child's mouth).'

      c. *e-puna=ra=tu   duju-**mere**-wa e-bakwa=ke=**keja***
          NPF-female=ERG=3SG take-CAUS-PRF 3-child=3=LOC.GNL
          misi   e-tare=ju
          tamale NPF-house=LOC
          'The woman had tamale taken to her house by her child.'

In the Cavineña direct causative construction in (98b), the causee of the derived verb *ara-mere* 'feed (cause to eat)' is treated like P in the clause because it is not case marked.[6] In the indirect causative construction in (98c), the causee of the derived verb *duju-mere* 'bring (cause to take)' is treated like an oblique due to the fact that it is marked with the general locative postpositional clitic =*keja* and is only optionally expressed in the clause.

In some languages, an additional construction type is reserved for causation events where the causer actively participates in the resulting situation. This distinction is often called a SOCIATIVE CAUSATIVE, which is the term adopted in this thesis, but it is also called a comitative causative in the Tupian descriptive tradition, where such a construction type is widespread. In Emerillon, a Tupí-Guaraní language of French Guiana, there are three different causative markers: *mo-* 'direct causative', *(e)lo-* 'sociative causative' and -*okal* 'indirect causative', as seen in (99):

Emerillon (Tupian; Rose, 2003, 358, 362-363, 366)

(99) a. *o-zaug*
          3.I-bathe
          'He bathed.'

      b. *wane idʒe a-**mo**-zaug*
          well   1SG 1SG.I-CAUS-bathe
          'I bathed it well.'

---

[6]Cavineña has an unmarked neutral ditransitive marking pattern, thus the causee is treated like P, T and R with regards to case marking. There is no verbal argument marking in Cavineña, but the pronominal enclitic =*tu* attached to the first constituent in the examples included in (98) can reference the P argument in the clause, in a process that Guillaume (2006) considers analogous to indexation. However, since these pronominals do not attach to the predicate and can be marked segmentable case morphology, they are not treated here as verbal argument markers.

    c. *de-**lo**-zaug*
       2SG.II-SOC.CAUS-bathe

       'He made you bathe with him.'

    d. *awak^wəl o-pihi-**okal**   o-aʔɨl*
       man   3.I-paint.with.achiote-IDR.CAUS 3.COREF-child

       *o-lek^w ala-**pe***
       3.COREF-wife-DAT

       'The man had his wife paint their child with achiote.'

The major distinction between the use of the direct causative marker *mo-* and the indirect causative marker *-okal* is that in the latter construction the causer does not directly manipulate the causee, but rather, causes the resulting situation to occur by means of coercion (Rose, 2003, 364).

In examples (97) and (99), the causative constructions with a transitive base have treated the causee participants with different case morphology; in Urarina, the causee is treated as an oblique argument marked with the preposition *ke*, as seen in (97b), while for Emerillon in (99d), the causee is case marked with the suffix *-pe*, which is also used to mark the R argument of basic ditransitive constructions. Differently in Warao, the causee is treated like the direct object argument and can be expressed by an accusative pronoun without an accompanying postposition or additional case marker, as seen in (100):

Warao (isolate; Romero-Figeroa, 1997, 9-10, 94)

(100)   a. ***ka***   *hi-rakoi   teoriasi-te*
           1PL.ACC 2SG-sister disdain-NPST

           'Your sister disdains us.'

       b. *harako-ma   daukuaha tai kona-n-a-e*
           hunting-DAT fruit   3SG bring-SG-PUNC-PST

           'I brought fruit for the hunting journey.'

       c. *wahabu-ma bare-tire*   ***ka***   *e-nahoro-a-e*
           venison-DAT padre-female 1PL.ACC CAUS-eat-PUNC-PST

           'The nuns made us eat venison.'

The Warao example in (100c) is one of the few clear cases where the causee is unambiguously treated like a direct object rather than R since the language shows indirective alignment (P=T≠R) in ditransitive constructions.[7] In most cases where the causee participant is treated like P, the language shows secundative alignment in ditransitive constructions, as in Kwaza (van der Voort, 2004, 113), or neutral alignment, as in the direct causation construction in Cavineña in (98). The causative construction in Trumai is the only instance

---

[7]Another clear example of a causee treated like a direct object can be found in Karo (Gabas Jr., 1999, 81-82).

within the sample of the causee in a morphological causative construction being treated like A of a transitive construction, as shown in (101):

Trumai (isolate; Guirardello, 1999, 257, 303)

(101)   a. *ine-**k** atlat-**o** mapa*
      3-ERG pan-ABS break
      'He broke the pan.'

    b. *hai-**ts** chï in atawaka-**k** atlat mapa **ka***
       1SG-ERG FOC/TENSE Atawaka-ERG clay.pan break CAUS
       'I made Atawaka break the clay pan.'

    c. *amati-**k** chï in tata-**k** karakarako taf kïtï **ka** ha*
       Amati-ERG FOC/TENSE Tata-ERG chicken egg give CAUS 1
       wan-ki
       PL-DAT
       'Amati made Tata give us chicken eggs.'

As can be seen in (101), the causee of causative constructions formed with transitive (101b) or ditransitive bases (101c) is case marked with the ergative suffix *-k* (for NP arguments) just like the A argument in (101a). Guirardello (1999, 303) states that the only way to distinguish between the causer and causee within the derived clause is through word order. The treatment of the causee is coded into the structural questionnaire based on its treatment with regard to case marking and indexation, and following such a procedure, the causee is considered to be treated like A.

The morphosyntactic treatment of the causee participants of the causative constructions derived from transitive verbal bases from examples (97) - (101) are shown in Table 7.1. The corresponding argument role that is treated with each argument marked in corresponding basic constructions is given after the marker used for the treatment of the causee in the examples shown above.

LANGUAGE	CONSTRUCTION	CASE		INDEX	
		MARKER	ROLE	MARKER	ROLE
Urarina	Indirect	ke	Oblique	-	-
Cavineña	Direct	-	P/T/R	-	-
	Indirect	=keja	Oblique	-	-
Emerillon	Sociative	-	-	Set II	P/T
	Indirect	-pe	R	-	-
Warao	Direct	Set II pro.	P	-	-
Trumai	Direct	-k	A	-	-

Table 7.1: Causee treatment in causative constructions with transitive bases

This table highlights the varying grammatical treatment that is given to

causee participants in causative constructions derived from transitive bases, corresponding to Q6.1.2 in the structural questionnaire.

7.2.2 Applicatives

Q7.1   *Applicative constructions are marked by verbal morphology:* [1, 0]

Q7.1.1   *Applicative constructions only select benefactive/malefactive participants:* [1, 0]

Q7.1.2   *Applicative constructions can only select animate participants:* [1, 0]

Q7.1.3   *Applicative constructions can be derived from transitive verbs:* [1, 0]

Q7.1.4   *The promoted argument can be expressed as an oblique in corresponding non-derived clauses:* [1, 0]

The defining property of an applicative construction is that it promotes a new participant into an non-subject argument role in which it could not otherwise be expressed in this way by the base verbal form.[8] The promoted object is referred to as the APPLIED OBJECT. The grammatical treatment of the applied object can correspond to P, T or R depending on the valency of the base form and the different grammatical relations that the language holds for non-subject arguments. The construction most often results in an increase in valency of the predicate, deriving a transitive verb from an intransitive base and a ditransitive verb from a transitive base. The situation type that is expressed by an applicative construction varies depending on the semantic role of the applied object. The one commonality across the semantic roles of the applied objects is that they are not patients, hence the applicative construction is used to express these non-patient roles as core arguments.

In some languages, the applicative construction is the only grammatical means to express a specific semantic role within a particular construction, resulting in an OBLIGATORY APPLICATIVE construction (Creissels, 2010, 30). The Aguaruna language has two applicative markers *-hu* and *-tu* whose distributions are lexically conditioned and in complementary distribution (Overall, 2007, 306). The use of these applicative markers is primarily restricted to promoting a new object argument into the valency frame of the predicate with a benefactive or malefactive semantic role, as seen in (102).[9] It is also possible to

---

[8]The term 'non-subject' is used here in a generalize sense to include different grammatical relations that contrast with transitive subjects (A), such as direct/primary/secondary/indirect objects and even absolutive arguments.

[9]The general term 'object' argument is used here to discuss applicatives in Aguaruna

increase the valency of a ditransitive verb, resulting in a quadrivalent predicate, as shown in (102c):

Aguaruna (Jivaroan; Overall, 2007, 63, 306, 446)

(102) a. *paampa-na   yu-a-ma-ha-i*
           plantain-ACC eat-HIAF-REC.PST-1SG-DECL
           'I ate plantains.'

    b. *amitʃi mi-na   ataʃu-na   yu-**hu**-tu-a-ɨ̃*
           fox   1SG-ACC chicken-ACC eat-APPL-1SG-HIAF-3:PFV
           'The fox ate my chicken.' (lit. the fox ate the chicken to my detriment)

    c. *huhu   papi apa   su-**hu**-tu-sa-ta*
           DEM:PROX book father:PERT:2 give-APPL-1SG-ATT-IMP
           'Give this book to your father for me!'

Notice how in (102a) the verb *yu* takes two arguments: an agent and a patient. In (102b) the transitive verb *yu* is made into a trivalent verb through the addition of the suffix *-hu*, such that it takes two arguments marked with the accusative case suffix *-na*: a patient participant *ataʃu-na* 'the chicken' (the thing eaten) and a malefactive participant *mi-na* 'me' (the entity that is negatively affected by the eating event). This is identical to the ditransitive construction in Aguaruna that has a marked neutral case marking pattern, with the most referential of the two non-A arguments being indexed on the verb following a 1SG > 1PL/2 > 3 hierarchy (cf. Overall, 2007, 315). This is the only applicative construction found in Aguaruna and there are no additional means to express a benefactive or malefactive participant in a corresponding basic construction, hence it can be considered an obligatory applicative construction.[10]

In some languages, obligatory applicative constructions are the only grammatical means to express a certain type of semantic participant without increasing the valency of a transitive predicate. In Pilagá, a Guaycuruan language of Argentina, there are no adpositions or other types of oblique case markers, so the only way to express participants that are not selected by the base form of a verb, such as instrument, comitative or benefactive participant for many basic transitive and intransitive verbs, is through the use of its rich verbal applicative morphology (Vidal, 2001, 316). In (103a) the base form of the verb *čaʕa* 'cut' is used, selecting a patient participant as its direct object; in (103b), the base form is derived with the instrumental suffix *-sona*, resulting in an instrumental participant being selected as the direct object. The examples in (103) show

---

since P, T and R are treated identically with regard to argument marking. Overall (2007, 306) notes that the goal of intransitive motion verbs can also be applied objects.

[10]The benefactive semantic role is defined by Van Valin Jr. and LaPolla (1997, 85) as "the participant for whose benefit some action is performed", and conversely, the malefactive participant is the entity to whose detriment some action is performed.

that the instrumental applicative construction in Pilagá can only promote a
new P argument into the valency frame of the predicate, but cannot increase
the valency of the verb (unlike other applicative constructions in the language).
In order to express that the cutting event was carried out with the knife and
that the meat was the entity affected by the cutting, a biclausal construction
must be used, as in (103c):

Pilagá (Guaycuruan; Vidal, 2001, 326-328)

(103)   a. *yi-čaʕa-yi   so' lapat*
             1.I-cut-DIR   CLF meat

             'I cut the meat.'

         b. *di-čaʕa-**sona** so' ganaʕat*
             3.I-cut-INS   CLF knife

             'He cut with a knife.'

         c. *yi-čaʕa-yi   so' lapat   yi-do'ok so' ganaʕat*
             1.I-cut-DIR   CLF meat   1.I-poke CLF knife

             'I cut the meat; I poke the knife.'

    In a NON-OBLIGATORY APPLICATIVE construction, the underived verbal
base can express the semantic role of the applied object as an oblique in the
basic construction, allowing for a semantically similar paraphrase between the
base and derived constructions whose major distinction is the grammatical
treatment of the participant that is the applied object in the derived construc-
tion. In Shipibo-Conibo, a Panoan language of Peru, there are a number of
different applicative constructions marked on the verb through suffixes. The
associative applicative marker *-kiin* promotes a comitative participant into the
valency frame of the predicate. In (104), the associative applicative marker
-kiin is attached to a transitive verb *oro* 'clear', resulting in a ditransitive
verb that shows an unmarked neutral ditransitive marking pattern. Unlike the
other applicative constructions in Shipibo-Conibo, such as those formed with
the malefactive applicative marker *-xon* (see Valenzuela, 2003, 746-749), a se-
mantically similar paraphrase of the associative construction can be expressed
through the use of the underived base verb with the comitative participant
marked with the postposition *betan*, as in (104b), and serves as a clear example
of a non-obligatory applicative construction:

Shipibo-Conibo (Panoan; Valenzuela, 2003, 685, 764)

(104)   a. *tita-n-ra   papa wai   oro-**kiin**-ai*
             mother-ERG-EV father chacra clear-ASSOC-INC

             'Mother clears the chacra with father.'

         b. *tita-n-ra   wai   oro-ai   papa   **betan***
             mother-ERG-EV chacra clear-INC father COM

'Mother clears the chacra with father.'

  c. *yoxaman-ra   bake bachi-n   jini-**kin**-ke*
     old.woman.ERG-EV child mosquito.net-ALL enter-ASSOC-COMPL

    'The old woman made the child enter the mosquito net (by entering herself).'

  d. *yoxaman-ra   bake bachi-n   jini-**ma**-ke*
     old.woman.ERG-EV child mosquito.net-ALL enter-CAUS-COMPL

    'The old woman made the child enter the mosquito net (but she herself did not enter).'

Interestingly, the comitative applicative in Shipibo-Conibo can also be used to promote a new A argument into the valency frame of a predicate, resulting in the expression of a sociative causative situation, as seen in (104c). When the causative marker *-ma* is used, as in (104d), direct causation is expressed.[11]

Some languages have a restriction on the transitivity of the verbal base form that can be targeted for derivation in an applicative construction. In Jarawara, an Arawan language of Brazil, the applicative construction formed by the marker *ka-* attaching to an inflecting verb or an auxiliary can increase the valency of an intransitive verb (Dixon, 2004, 254-266). The construction is generalized in the sense that it can include a number of different semantic roles as applied objects, such as comitative participants (105b) or theme participants (105d), and in most cases the applied objects can be realized as obliques marked by the postposition *(ni)jaa* in a corresponding non-derived construction, as seen in (105):

Jarawara (Arawan; Dixon, 2004, 255-256, 258)

(105) a. *okobi   wine otaa   **nijaa***
       father:1SG live:M 1PL.EXCL OBL:HUM

      'My father lived with us.'

    b. *okobi   otara   **ka**-wine   otaa   fota-ra*
       father:1SG 1PL.EXCL.ACC APPL-live:M 1PL.EXCL be.big:PL-NEG:F
      otaa
      1PL.EXCL

      'My father lived with us when we were small.'

---

[11] As noted in Shibatani and Pardeshi (2001, 116-122), syncretism between applicative and causative constructions is found in a number of languages all over the world. Peterson (2007, 65) notes that although syncretisms between comitative applicative and sociative causative constructions are attested in only a few cases cross-linguistically, there is evidence that these two constructions are often diachronically related in these instances. Discussing this point, Guillaume and Rose (2010, 392) suggest that this diachronic pathway may be underrepresented in the typological literature since benefactive and instrumental applicative constructions are often considered more prototypical cases of applicative constructions than comitative applicatives.

c. *jomee habo na-ka   owa **nijaa***
   dog   bark AUX-DECL:M 1SG OBL:HUM

   'The dog barks at me.'

d. *jomee owa   habo **ka**-na-ka*
   dog   1SG.ACC bark APPL-AUX-DEC:M

   'The dog barks at me.'

Example (105b) shows an applicative construction that promotes a comitative participant into the P argument role. The same comitative participant can be realized as an oblique marked with the postposition *nijaa* in a corresponding non-derived construction formed with an intransitive verb (105a). Example (105d) shows an applicative construction that promotes a theme participant, also with an intransitive verbal base, with the corresponding paraphrase shown in (105c).

7.2.3   Summary of valency increasing strategies

These valency increasing strategies are schematically represented in Table 7.2. In the valency frames listed below, the first participant in the set corresponds to the S or A argument role depending on the valency of the predicate. In constructions with a bivalent predicate, the second argument in the valency frame corresponds to the P argument role. In constructions with a trivalent predicate, the second participant in the set corresponds to the ditransitive argument role that aligns with P, i.e. either T or R depending on the ditransitive alignment in the language for the relevant argument markers. The third participant in a construction with a trivalent predicate, either the causee in causative constructions derived from a transitive base or the base patient in an applicative construction derived from a transitive base, show variable morphosyntactic treatments, as discussed in the previous sections.

CONSTRUCTION	VALENCY FRAME	PROM.	DEMO.	VAL.
Basic intr.	[agent]	-	-	1
Basic tr.	[agent, patient]	-	-	2
Causative intr.	[causer, causee]	causer	-	2
Causative tr.	[causer, patient, causee]	causer	(causee)	3
Applicative intr.	[agent, oblique]	oblique	-	2
Applicative tr.	[agent, oblique, (patient)]	oblique	(patient)	2/3

Table 7.2: Overview of valency increasing strategies

Given the examples shown in the chapter and the data coded in the structural questionnaire, a number of preliminary observations can be made about the valency-increasing strategies used by the languages in the sample:

Causatives: Within the sample, 62 languages were identified as displaying a morphological causative construction.

- Among the languages that were identified as having a morphological causative construction, the majority allowed at least one construction to target transitive bases (72.6%).

- At least 15 languages in the sample show a formal grammatical distinction between sociative and direct causation. A large number of Tupian languages in the sample show a morphological distinction for sociative causation restricted to intransitive bases, which Rodrigues and Cabral (2012, 530) reconstruct for Proto-Tupí as **mo-*. After examining mentions of distinct sociative causative markers in the descriptive literature of languages from all over the world, Guillaume and Rose (2010, 388-391) propose that this category is a possible areal feature of South America as a whole. They also note that the distribution of sociative causative markers is primarily restricted to Southern and Western Amazonia and suggest that this typological pattern may have spread as a result of contact with Tupian languages. The geographical distribution of sociative causatives within the continent itself is further explored in section 8.2.4 of the following chapter.

- A total of 11 languages were identified as displaying a formal grammatical distinction between direct and indirect causation. It is interesting that this distinction is less commonly marked in the sample languages than the sociative-direct distinction. In some languages, such as for Urarina in (96-97), this distinction is only formalized for a specific transitivity value of the verbal base. Unlike sociative causatives, the indirect causative distinction was manifest through a wider range of morphosyntactic strategies such as the differential treatment of the causee participant, as in Cavineña in (98), or through serial verb constructions, as in Tariana (see Aikhenvald, 2003a, 275).

- As discussed already in section 7.2.1 and summarized in Table 7.1, the languages in the sample that have morphological causative constructions that target transitive bases display a variety of different morphosyntactic treatments for the causee participant, including argument marking corresponding to the treatment of A, T, R and obliques. In addition to the treatment of the causee with regard to which argument role it is marked as, it is also useful to consider the grammatical relation that the causee forms with other argument roles. Strictly speaking in terms of case marking, 26 languages treated the causee as an objective argument, i.e. forming a grammatical relation that includes P together with T and/or R depending on the ditransitive alignment pattern in the language. A total of ten languages in the sample treat the causee like a dative argument, i.e. corresponding to the treatment of R as distinct from the treatment

of P and T. Six languages treat the causee like an oblique with marking distinct from P, T and R, while only a single language (Trumai) treats the causee as an ergative argument.

Applicatives: Within the core sample of languages used in the study, only 29 languages were identified as having a morphological applicative construction based on the available descriptive materials. Unlike causative constructions, a considerable portion of the descriptive materials do not provide much information on valency increasing strategies that promote new object arguments into the valency frame of a predicate. Nonetheless, some observations can be made about the occurrence of applicative constructions in the sample:

- A total of 12 languages were identified in the sample as displaying a single morphological causative construction that was restricted to promoting benefactive and/or malefactive participants. Interestingly, these languages are almost all located in the Central Andes or Western Amazonia regions. When compared to a global sample of applicative constructions in Polinsky (2011b), South American languages show a higher distribution of languages that have an applicative construction restricted to promoting benefactive/malefactive participants when compared to constructions with more generalized semantic targets, but this difference does not show strong statistical significance when the sample used here is compared to the non-South American languages in WALS ($p = 0.371$, one-tailed Fisher's Exact Test).[12]

- For the 29 languages identified as having a morphological applicative construction, only 12 allowed for the semantic participant corresponding to the applied object to be realized as an oblique participant in corresponding basic constructions.

- In almost all cases, the applicative construction in a language can be applied to both transitive and intransitive verbal bases. Only in two cases within the core sample was there a restriction on the transitivity of the base: the Jarawara construction in (105) and the benefactive construction in Cocama-Cocamilla are both restricted to intransitive bases (Vallejos Yopán, 2010, 381-382). See also Braga (2005, 165-167) for the applicative construction restricted to intransitive bases in Makurap, which is included in the additional sample of Tupian languages used in Chapter 5.

---

[12]There are some inconsistencies in the coding of applicative constructions for some South American languages between the two maps given in Polinsky (2011b). For example Map 109A, Shipibo-Conibo (Panoan) is coded as having 'Benefactive and other roles' as applied objects, whereas in Map 109B the same language is treated as having the value 'No other roles (= Only benefactive)'. In the database used in this thesis, following the description in Valenzuela (2003), Shipibo-Conibo is considered to have multiple applicative constructions including the benefactive but also others, such as the comitative applicative construction shown in (104). The values in Map 109A were used for the significance test above.

7.3 Valency decreasing strategies

This section explores major parameters of variation in four commonly occurring valency decreasing constructions: passives, antipassives, reflexives and reciprocals. Valency decreasing strategies involve the demotion of a core argument to a non-core function, reducing the valency of the predicate by one. This section primarily discusses the derivation of monovalent predicates from bivalent predicates, with a focus on transitive constructions. In PASSIVE CONSTRUCTIONS derived from transitive bases, the base A argument is demoted, resulting in the promotion of the base P argument to argument role of S. In ANTIPASSIVE CONSTRUCTIONS, the P argument is demoted, resulting in the base A argument being treated like the S argument in the derived valency frame.

Reflexive and reciprocal constructions also generally result in the decrease of valency of a bivalent or trivalent verbal base and are further defined by the specific situation type that they express. In REFLEXIVE CONSTRUCTIONS, the semantic participants corresponding to the base A and P arguments in a transitive clause are considered the same participant, resulting in the realization of only a single argument S that is interpreted as expressing the semantic role of both the base A and P arguments. In other words, the participants in a reflexive situation do something to themselves. In RECIPROCAL CONSTRUCTIONS, the participants corresponding to the base A and P arguments in a transitive clause are considered to hold multiple semantic roles in relation to the predicate, prototypically being both agent and patient simultaneously, resulting in the realization of a single plural argument S that is interpreted as expressing the semantic roles of both agent and patient. In other words, the participants in a reciprocal situation do something to each other. While the primary strategy to express these constructions in many languages is through valency decrease, some languages instead rely on a special class of reflexive or reciprocal pronouns to express these situation types. In these cases, the base valency of the predicate is maintained. Both strategies will be discussed in sections 7.3.3-7.3.4.

7.3.1 Passives and anticausatives

Q8.1   *Passive constructions are marked by:* [verbal morphology, auxiliary, both, none]

Q8.1.1   *The demoted A of a passive construction can be expressed in the clause:* [1, 0]

Q12.1   *Anticausative constructions are marked with verbal morphology:* [1, 0]

Q12.1.1   *The marking of anticausative constructions is conflated with the marking of reciprocal and/or reflexive constructions:* [1, 0]

A passive construction can be defined by the situation type that the construction expresses and the effects that such a construction has on the valency of the verbal base. In a passive construction, the situation type corresponds to a basic transitive construction, prototypically including both an agent and a patient participant. However, the construction reduces the valency of the base predicate by one through the demotion of the agent participant. This results in the patient participant being treated as the S argument in a passive construction derived from a transitive predicate. Since the situation type of the passive construction obligatorily includes an agent participant whose syntactic status is demoted, the morphosyntactic realization of the agent either corresponds that of an oblique, or the participant is omitted entirely from the utterance but still implied.

Passives are one of the most widely discussed valency decreasing strategies in the typological literature. There are two general strategies used in languages to form a passive construction: PERIPHRASTIC PASSIVES are those expressed through the use of an auxiliary verb, as in English (cf. Keenan and Dryer, 2007, 336-339); SYNTHETIC PASSIVES are those formed with only a lexical verb and the addition of specialized verbal morphology (cf. Siewierska, 2011b).

In periphrastic passive constructions, the additional morphology attached to the lexical verbal base can result in the loss of some of its verbal properties and a change in its function outside of the passive construction, often as a nominal modifier (a participle) or a nominal itself (a nominalization). Since the distinction between a participle and a nominalization primarily rests on the presence of an adjective-noun distinction in a particular language (cf. Haspelmath, 2010), either type of construction has been included as a periphrastic passive construction if it meets the definitional criteria given above.

In Nasa Yuwe, a language isolate from Colombia, the periphrastic passive construction is formed with the perfect past participle marker -$n^j i$ and the copula *ũs*, as shown in (106):

Nasa Yuwe (isolate; Jung, 2008, 73)

(106) a. *ũʔkwe-ʔ*   *nasa*   *yuwe-ʔs*   *pija-ts-t^{h}u*
          1SG.F-TOP people language-OBJ.SG learn-PROG-DECL.1SG
          'I am learning Nasa Yuwe.'

       b. *nasa*   *yuwe-ʔ*   *pija-$n^j i$*   *ũs-aʔ*
          people language-TOP learn-PRF.PTCP be-DECL.3SG
          'Nasa Yuwe was learned.'

According to Jung (2008, 73), it is possible to realize the demoted agent as an 'attributive genitive', corresponding to the treatment of a possessor in a possessive construction, but no examples of this were encountered in the description.[13]

---

[13]Based on this characterization, the agent can be realized without case marking in a

Notice how the topic marker *-ʔ* is attached to the A argument of the base construction in (106a), while it is attached to the derived S argument in (106b). This is in line with the general pragmatic function of a passive construction, which is to topicalize the patient participant while demoting the agent, such that the patient is the only topical argument of the derived clause (Givón, 2001, 122).

In Bororo, a Macro-Jêan language of Brazil, a synthetic passive construction is formed with the verbal suffix *-dɨ*.[14] Crowell (1979, 54, 60) notes that the passive construction in Bororo is used only infrequently, often to avoid mention of an agentive argument. Nonetheless, the Agent of a passive construction can still be realized as the object of the preposition *koia* 'by, because', as in (107c):

Bororo (Macro-Jêan; Crowell, 1979, 54, 57, 61)

(107)   a. *e-re   bola barigu*
            3PL-NEUT ball throw

            'They threw the ball.'

        b. *bola barigu-**dɨ**-re*
            ball throw-PASS-NEUT

            'The ball was thrown.' (also: 'There was a throwing of the ball.')

        c. *bola barigu-**dɨ**-re   a-**igoia***
            ball throw-PASS-NEUT 2SG-OBL

            'The ball was thrown by you.'

According to Keenan and Dryer (2007, 328-329) the most common cross-linguistic type of passive construction is that shown in (107b), where the event expresses an action that requires an agent and a patient, and where the agent is not expressed in the clause. These passive constructions without a lexically expressed agent are called AGENTLESS PASSIVES, while those with an agent

---

position preposed to the participle, as in other attributive possessive constructions in Nasa Yuwe (cf. Jung, 2008, 121-122). Thus, the passive construction in Nasa Yuwe is similar to the passive construction formed with *ka-* 'be' in Quechuan languages. In some varieties of Quechua, such as Cuzco Quechua discussed in Muysken (1986), the demoted agent is realized with an overt genitive marker *pa-*, suggesting that the participle is indeed a noun in an existential construction. In Imbabura Quechua, there is no genitive marking on the demoted agent (cf. Cole, 1982, 133-134), thus appearing unmarked for case much like a nominative argument, except that it cannot be indexed on the verb. See Witzlack-Makarevich (2010, 209-212) for further discussion on the Imbabura passive construction.

[14] Crowell (1979, 54-59) glosses this suffix as 'nominal' since the same marker is used in existential clauses. In fact, he considers the passive construction in Bororo to be a type of existential construction, similar to the analysis of the English passive in Langacker and Munro (1975). Nonato (2008, 68-69, 168-172) argues that the marker is best considered a general detransitivizer and emphasizes its function in deriving intransitive verbs with an inchoative reading. However, Nonato does not account for the use of the marker *-dɨ* with telic action verbs such as that shown in (107).

expressed as an oblique, such as in (107c), are called PERSONAL PASSIVES.[15]

Unlike Bororo, some languages have a passive construction that does not allow the agent to be expressed as an oblique under any circumstances. This can be seen for Baure in (108):

Baure (Arawakan; Danielsen, 2007, 246)

(108) a. *ni=papa   to   čičirop*
           1SG=harvest.beans art bean
           'I harvest beans.'

       b. *to   čičirop ver ro=ko=pa-**si***
           ART   bean   PFV 3SG.M=ATTR-harvest.beans-PASS
           'The beans have been harvested.'

In Baure, the passivization of transitive verbs requires the use of two derivational morphemes: the attributive marker *ko-* and the passive marker *-si*, as seen in (108b).[16]

There are two commonly occurring constructions in South American languages that are structurally and functionally similar to passive constructions but fail to show at least one of the defining properties given above. The first construction is what is commonly called an ANTICAUSATIVE CONSTRUCTION, where an intransitive verb is derived from a transitive base resulting in the treatment of the patient participant as the derived S argument (Comrie, 1985, 325-326). The anticausative construction is considered distinct from the passive construction due to the fact that the agent in the anticausative derivation is not implied in the event expressed by the construction, i.e. the anticausative situation type include no agent participant. In this sense, anticausative derivations are restricted to situations that can occur spontaneously "without an initiating actor", such as by means of natural processes like gravity, growth, decay and combustion (Haspelmath, 1987, 15).

In Wayoró, a Tupian languages of Brazil, the intransitivizing prefix *e-* can express an anticausative situation when attached to a transitive base that expresses an event that occurs spontaneously, as shown in (109a-109b). When the same marker *e-* is attached to a verb that express an event that does arise out of a natural process and requires an external force or entity for the event

---

[15]The term 'impersonal passive' used in Siewierska (2011b), which corresponds to the use of agentless passive in this thesis, is not adopted due to its frequent use to label passive constructions formed with intransitive bases, such as in Dutch *Er werd opnieuw gevochten (door de voetbalsupporters bij de wedstrijd)* 'Again there was fighting (by the soccer fans at the match).' For further discussion of passives derived from intransitive bases, see Keenan and Dryer (2007, 345-348).

[16]See Danielsen (2007, 238-240) for further information of derivation of stative verbs from active verbs with the attributive marker *ko-*. According to Danielsen (p.c.), there are no examples of the passive construction in her corpus using one of the few underived stative verbs in Baure. This suggests that the use of the attributive marker has grammaticalized as part of the construction.

to occur, as with *piriga* 'pierce' in (109c), the utterance expresses a reflexive situations, as shown in (109d):

Wayoró (Tupian; Nogueira, 2011, 187-188, 203)

(109) a. *õn   agopkap eβaka*
       1SG firewood light
       'I light the firewood.'

    b. *agopkap te-e-eβaka-t*
       firewood 3-INTR-light-PST
       'The firewood lit.'

    c. *ãrãmĩrã iko-ɲẽrã   piriga-t*
       woman   game-meat pierce-PST
       'The woman pierced the game meat.'

    d. *ŋgwajkip te-e-piriga-t*
       man   3-INTR-pierce-PST
       'The man pierced himself.'

The fact that the predicates in examples (109b) and (109d) are indeed monovalent is clear from the use of the 3rd person index *te-*, which is reserved for S arguments (cf. Nogueira, 2011, 68). Constructions such as that shown in (109) for Wayoró, which express reflexive situations together with other situation types that result from a decrease in transitivity are commonly labeled MIDDLE VOICE CONSTRUCTIONS. Kemmer (1993) presents a typological overview of the concept of middle voice and lists the following situation types that are sometimes subsumed under this label: grooming and body care ('wash'), nontranslational motion ('turn'), change in body posture ('kneel'), self-benefactives ('acquire'), naturally reciprocal events ('wrestle'), translational motion ('ascend'), emotional responses ('be.frightened'), further mental states and processes ('believe', 'consider'), and spontaneous events (such as 'ignite' in (109b)), among others including a passive situation. However, middle voice constructions are not treated as a distinct type in the structural quesitonnaire due to its nonspecific nature and the variety of different situation types expressed in different languages under this label (but see 7.3.5 below).

The construction in Wayoró marked by the intransitivizer *e-* is not considered a passive since it cannot express a canonical passive situation in which the demoted agent is implied in the event. However, in some languages a single morpheme may express a variety of different situation types, including the passive situation but also others, leading to ambiguity in how this language should be treated with regard to the structural questionnaire. In instances where at least one of the situation types expressed by a particular morpheme includes the passive situation, the language was treated as having a passive construction. Tiriyó, a Cariban language of Brazil and Suriname, can derive monovalent

verbs from bivalent verbal bases with the detransitivizer *ë(t)-*. The different situation types express by this marker in Tiriyó are shown in (110):

Tiriyó (Cariban; Meira, 1999, 259)

(110)   a. Passive situations: *ë-enepo* 'be seen'
        b. Grooming and body care: *e-pontë* 'get.dressed'
        c. Change in body posture: *e-tahpaka* 'sit.down'
        d. Naturally reciprocal events: *ë-epo(rï)* 'meet'
        e. Emotional response: *e-meneka* 'be.surprised'
        f. Spontaneous events: *e-tohka* 'burst'
        g. Translational motion: *et-ainka* 'run.away'
        h. Mental states and processes: *ët-amorehtë* 'dream'

In addition to the situation types expressed above, the detransitivizer *ë(t)-* in Tiriyó can also express reflexive and reciprocal situations, as well as result in an antipassive derivation (Meira, 1999, 254-258). While the construction formed by the detransitivizer expresses a multitude of different situation types beyond the passive, the ability to express a passive situation, along with the structural criteria of patient promotion and agent demotion of a transitive verbal base, Tiriyó is considered to have a passive construction. A similar polysemous construction labeled 'middle voice' that includes passive situations is also found in Yurakaré (see van Gijn, 2010 for details). As a general note, verbal bases that express prototypical transitive events tend to result in either a reflexive, reciprocal or passive interpretation, while bivalent verbs whose participants do not neatly conform to agent and patient semantic roles entailed in a transitive event are those that are interpreted as expressing the other situation types commonly subsumed under the label 'middle' after undergoing A demotion.

A second construction type that is functionally similar to a passive construction but fails to meet at least one of its defining properties is the UNSPECIFIED SUBJECT CONSTRUCTION. In an unspecified subject construction, the predicate is marked to indicate that the A argument of the clause is irrelevant to the utterance, resulting in a meaning that can be paraphrased in English as 'someone' or 'people' (Keenan and Dryer, 2007, 354-356). This inherently results in an increase in the topicality of the P argument, which leads Givón (2001, 127-128) to label such a construction a 'non-promotional passive'. However, unlike a passive construction, there is no change in the valency of the verbal base in an unspecified subject construction. Keenan and Dryer (2007) states that the most common means to form an unspecified subject construction is through the use of a specialized argument index, as seen Kwaza in (111):

Kwaza (isolate; van der Voort, 2004, 259-260, 569)

(111)   a. *ywynwỹ-simjã'ku 'dy=asa-xa-xa-ki*
           tree-CLF:stump   cut=leave-2-AS-DECL

'You're cutting tree stumps away.'

b. *ywy'nwỹ* *'dy=asa-**wa**-ki*
    tree   cut=leave-IS-DECL

'Someone cut the log and cleared it out of the way.'

c. *a'wỹi-da-ki*   *a'rũi-wã* *mawdy'ne-**wa**-ta*
    see-1SG-DECL tapir-OBJ call.to.come-IS-CSO

'I saw them (someone) call tapir to come (to them).'

In Kwaza, the unspecified subject construction is formed with the indefinite subject marker -*wa*, as seen in (111b). This marker occurs in the morphological slot that indexes the nominative argument, such as the 2nd person marker -*xa* in (111a). Crucially, the use of the indefinite subject marker does not affect the valency of the predicate, as can be seen in (111c) where the P argument of the unspecified subject construction is marked with the case marker -*wã*, which is reserved for animate P and R arguments (see section 6.2.2).

An intermediary case between a passive construction and a prototypical undefined subject construction is found in Wichí, where the 'non-promotional passive' construction is marked with the prefix *t*- in a morphological slot distinct from the person index. This construction often results in a free translation that is equivalent to a passive construction in English, as shown in (112):

Wichí (Matacoan; Terraza, 2009, 192-193)

(112) a. *atsinha i-tsoy*   *inot*
          woman 3.I-spill water

'The woman spilled the water.'

b. *inot*   ***t**-i-tsoy*
    water NPR.PASS-3.I-spill

'The water was spilled (by someone).'

However, there are two properties of this construction that demonstrate that this construction is indeed not a passive in the terms described above. First, as described by Terraza (2009, 124), the third person prefix *i*- is only used for A arguments in transitive clauses, and never S, for which third person arguments are indexed with *ta*- for major class intransitive verbs and unmarked for the minor class of stative verbs. However, as seen in (112), the marker *i*- is prefixed to the predicate in the construction to index a third person A argument. An additional property of the 'non-promotional passive' in Wichí is that when the patientive argument is a speech act participant, it is indexed by the suffix marker set reserved for P arguments, as seen in (113b):[17]

_____

[17]The verb stem *tsex^w en* 'blame' obligatorily occurs with the applicative marker -*a* (Terraza, p.c.). A number of bivalent verb stems in Wichí obligatorily occur with an applicative suffix to form transitive verbs, as discussed in Terraza (2009, 219-220).

Wichí (Matacoan; Terraza, 2009, 190)

(113) a. *i-tsex^{w}en-a   eł   hop maq to   toy*
           3.I-blame-APPL someone FOC thing SUB disappear
           'They blame someone for the thing that disappeared.'
       b. **t**-*i-tsex^{w}en*-**am**-*a*
           NPR.PASS-3.I-blame-2-APPL
           'They blame you for something.'

While the undefined subject construction in Wichí is pragmatically similar to a passive construction, resulting in the topicalization of P as evidenced by the ability for this argument to occur preverbally as in (112b) and (113b), it is not considered a passive construction here because it does not result in a decrease of the valency of the verbal base.

7.3.2   Antipassives

Q9.1   *Antipassive constructions are marked by verbal morphology:* [1, 0]

Q9.1.1 *The demoted P argument can be expressed as an oblique:* [1, 0]

Q9.1.2 *The marking of antipassive constructions is conflated with the marking of reciprocal and/or reflexive constructions:* [1, 0]

Antipassive constructions are similar to passive constructions in the sense that the situation type expressed by the construction is identical to its transitive counterpart, requiring both an agent and a patient participant.[18] The valency effects of an antipassive construction are quite different than those of a passive construction. In antipassive constructions the valency of the verbal base is reduced by one through the demotion of the patient. This results in the base P argument being unexpressed or realized as an oblique, albeit still implied in the event. It is often the case that the change in argument role of the A argument of the base predicate to S of the derived predicate results in the promotion of this argument, allowing it to be targeted by certain argument selectors such as case marking and indexation, but also intraclausal argument selectors such as coreference in coordination, as seen for Katukina-Kanamari in (114):[19]

---

[18]This is why passives and antipassives are commonly called 'voice', while other valency changing strategies that express a situation type different from a transitive construction are less commonly labeled as voice phenomena.

[19]Argument selectors outside of case and indexation are not discussed in this thesis, but see Witzlack-Makarevich (2010) for an overview.

Katukina-Kanamari (Katukinan; Queixalós, 2010, 243, 257-258)

(114) a. *anyan hinuk-na   toman wiri*
            3SG   group-ERG shoot peccary
            'They shot a peccary.'

      b. ***wa**-toman   adu wiri **katu**   wa*
            ANTIP-shoot 1SG peccary SOC.INS PROSP
            'I am going to shoot a peccary.'

      c. *nodia-na   pikan owi **wa**-tohik tu   niama*
            Nodia-ERG hear   Owi ANTIP-see NEG then
            'Nadia$_i$ heard Owi$_j$ but she$_j$ did not see (him$_i$).'

The clause in (114a) displays all of the typical properties of a transitive construction in Katukina-Kanamari: the A argument is case marked with the suffix -*na* and there is a AVP constituent order. In (114b), the bivalent predicate *toman* 'shoot' is derived as monovalent through the use of the antipassive prefix *wa-*. The resulting clause shows all of the typical properties of an intransitive construction in the language: there is no case marking of the sole obligatory participant, the S argument, and the constituent order is VS. Additionally, the base agent participant is demoted and marked by an oblique postposition *katu*. However, as (114c) shows, the demoted agent does not need to be expressed within the clause for it to be grammatical.

In Katukina-Kanamari, the absolutive argument is coreferential across paratactically coordinated clauses (Queixalós, 2010, 244). Since the second clause in the coordinated construction in (114c) is intransitive, the S argument is coreferential with the P argument of the transitive clause that occurred before it, as indicated in the free translation with subscripts. In other words, the absolutive argument is the 'syntactic pivot', to use the term popularized in Dixon (1979). In this sense, one of the functional motivations for the use of the antipassive construction is to allow the agent participant of the clause to be accessible to certain argument selectors that are restricted to absolutive arguments.[20] An antipassive construction that results in the promotion of the agent participant, allowing it access to a number of argument selectors that it could not otherwise access in the base construction, are referred to in Foley and Van Valin Jr. (1984, 170) as 'foregrounding antipassives'. This type of construction is thus restricted to languages that show ergative alignment in one or more of the argument selectors under consideration, as in Katukina-Kanamari, since if the same argument selectors showed accusative alignment, there would be no change in the accessibility of an argument to those argument selectors since S and A form a single grammatical relation, the nominative. At the same time,

---

[20]In addition to coreference across coordinated clauses, Queixalós (2010, 258) lists a number of argument selectors restricted to absolutive arguments in Katukina-Kanamari, including contrastive focus, ostention, coordination of participants within a clause, interrogation, relativization and clause subordination.

note that while the change of the agent participant from the A argument role
to the S role in the antipassive construction increased its accessibility to the ar-
gument selectors mentioned above, it also is demoted from its access to another
argument selector relative to the basic construction—the ergative case marker
-na. However, for Foley and Van Valin Jr. (1984) the important characteristic
of a foregrounding antipassive construction is that an argument is promoted
into an argument role that has access to the syntactic pivot.[21]

   If 'foregrounding' were the only function of an antipassive construction, it
would be expected that these constructions would be restricted to languages
with ergative alignment in one of more argument selectors. However, there are
also languages that show predominantly accusative alignment in argument se-
lectors such as case and indexation that also have antipassive constructions.
This can be seen for Lokono, an Arawakan language of Suriname, in (115):

Lokono (Arawakan; Pet, 2011, 25-27)

(115)   a. *li   **fary**-fa aba kabadaro*
             3SG kill-FUT one jaguar

             'He will kill a jaguar.'

         b. *li   **fara**-fa*
             3SG fight.ANTIP-FUT

             'He will fight.'

         c. *li   **fara**-fa   to   kabadaro **oma***
             3SG kill.ANTIP-FUT ART jaguar   with

             'He will fight with the jaguar.'

         d. *to   hiaro **kanaby**-fa to   kodibio-be*
             ART woman hear-FUT   ART bird-PL

             'The woman will hear the birds.'

         e. *to   hiaro **kanaba**-fa   to   kodibio-be **khonan***
             ART woman hear.ANTIP-FUT ART bird-PL   about

             'The woman will listen to the birds.'

   In Lokono, the antipassive construction is formed by means of a stem al-
ternation where the vowel of the final syllable, which is obligatorily either *e*,
i, *o* or *y* (but never *a*), is changed to *a*, forming what Pet (2011, 25) calls the
'*a*-stem'. This stem alternation derives a monovalent predicate from a bivalent
predicate. The demoted patient is either omitted in the derived construction

---

[21]Foley and Van Valin Jr. (1984, 114-115) prefer the term 'pragmatic pivot' since the choice
between the basic construction and an antipassive or passive construction is often motivated
by discourse factors such as topicality or intraclausal argument coreference. They later note
that pivots are indeed syntactic in nature, but that they are a "*syntacticization* of certain
discourse relations, one of which is topicality" (Foley and Van Valin Jr., 1984, 134, emphasis
in the original).

(115b), or it can be expressed as an oblique, either marked by the postposition *oma* 'with' (115c) or the postposition *knonan* 'about/on/of' (115e). The use of the antipassive construction affects the telicity of the event expressed by the predicate, giving an interpretation of the event being less bounded in time and more general in nature, and can often lead to an iterative reading, as seen in (115c) where *fara* 'fight', the antipassive derivation of *fary* 'hit/kill' can be construed as hitting repeatedly (Pet, 2011, 27). However, the antipassive construction in Lokono does not affect the accessibility of the agent participant to different argument selectors, as in Katukina-Kanamari, since both S and A are subsumed under the nominative grammatical relation in Lokono. Thus, the primary function of the Lokono antipassive construction is to demote patientive participant rather than to promote the agentive participants to a special syntactic status, corresponding to what Foley and Van Valin Jr. (1984, 172) call a 'backgrounding antipassive'.

The two sets of examples in (114) and (115) show antipassive constructions that allow for the demoted patient to be expressed as an oblique, forming semantically similar paraphrase of the base construction. However, just as seen in section 7.3.1 for passive constructions, a structural parameter of variation within antipassive constructions is whether or not the demoted patient can be expressed as an oblique.[22] Adopting corresponding terminology, antipassive constructions that allows the demoted P argument to be expressed by an oblique are PERSONAL ANTIPASSIVES, while those that do not allow for this participant to be expressed, albeit still implied by the situation type of the predicate, are PATIENTLESS ANTIPASSIVES. An example of a patientless antipassive construction can be seen for Matses in (116):

Matses (Panoan; Fleck, 2003, 931)

(116)   a.   *aid*   *opa-n*   *matses pe-e-c*
             that.one dog-ERG people bite-NPST-IND
             'That dog bites people.'

        b.   *aid*   *opa pe-**an**-e-c*
             that.one dog bite-ANTIP-NPST-IND
             'That dog bites.' (also: 'That dog always bites me / is biting me.')

The antipassive construction in Matses is formed through the use of the suffix -*an*. In these constructions, the patient can be either unknown and/or indefinite, or it can imply an (unmarked) first person patient participant. However, in neither interpretation can the patient be realized as a participant in the

---

[22]In a slightly different formulation of the defining properties of an antipassive construction, Dixon and Aikhenvald (2000b, 9) not only include the demotion of the base P argument, but require that it is possible for it to be express as an oblique. This additional criterion is not adopted in this thesis, but similarly to passive constructions, the presence of some patientive argument must be implied in the situation type expressed by the predicate in the antipassive construction.

clause (Fleck, 2003, 932). The reduction in valency of the derived construction can be most clearly seen through case marking; in (116a) the A argument is marked with the ergative case marker *-n*, while in (116b), the corresponding derived S argument (also an agent) is unmarked for case, showing that it is an absolutive argument.

Languages with antipassive constructions sometimes also express reflexive constructions with the same marker. In Cavineña, the patientless antipassive and reflexive constructions are both marked with the circumfix *k(a)- ... -ti*, glossed as 'reflexive' in both instances in (117):

Cavineña (Tacanan; Guillaume, 2008, 270, 275)

(117)   a. *amena tume chapa=dya=di=pa*
             FILL   then dog=FOC=EMPH=REP
             ***ka**-rikwa-**ti**-aje-ya=dya*
             REFL-bark.at-REFL-GO.DISTR-IPFV=FOC
             'And the dog was barking too.'

         b. ... *tujuri   siri=tsewe **ka**-rama-**ti***
             mosquito.net old=ASSOC REFL-cover-REFL

             '(I didn't have any warm clothes so) I covered myself with my old mosquito net.'

Guillaume (2008, 274) notes that the patientless antipassive construction in Cavineña is typically used to leave an underlying patient participant unspecified as an argument in order for the activity expressed by the predicate to be focused, as seen in (117a). The same circumfix is also used to form a reflexive construction, as shown in (117b).

7.3.3   Reflexives

Q10.1   *Reflexive constructions are marked with verbal morphology:* [1, 0]

Q10.2   *Language has independent reflexive pronouns:* [obligatory, optional, none]

The situation type expressed by a reflexive construction is where the semantic participants, the agent and patient of a prototypical transitive clause, are the same entity or entities acting on themselves. In this sense, the reflexive construction does not always entail a reduction in valency to be expressed, just as in the English reflexive construction *He shot himself in the foot*. Strategies that form a reflexive construction through the use of a specialized pronominal, or through special nominal morphology attached to free pronouns, are called

PRONOMINAL REFLEXIVE CONSTRUCTIONS, just as the English reflexive pronouns *myself* or *himself*. However, there is a common cross-linguistic strategy to form reflexive constructions by means of specialized morphology attached to the verb, which reduces the valency of the predicate by one. For derived monovalent predicates from transitive bases, this reduction in valency results in the interpretation of the sole argument S as being both the agent and patient participant of the event expressed by the predicate. This reflexive strategy is called a SYNTHETIC REFLEXIVE CONSTRUCTION. There are thus two general strategies for the express of reflexivity in languages, pronominal and synthetic, where only the latter results in a reduction of the valency of the verbal base.

The Kamayurá language of Brazil serves as a classic example of a synthetic reflexive construction. Through the addition of the reflexive prefix *je-*, a major class intransitive verb is derived from a transitive verbal base, as seen in (118):

Kamayurá (Tupian; Seki, 2000, 279)

(118)   a.   *ka'ahera o-kytsi kye'ia pupe*
            paper   3.I-cut knife INS
            'He cut the paper with a knife.'

        b.   *ene ere-**je**-kytsi   kye'ia pupe*
            2SG 2SG.I-REFL-cut knife INS
            'You cut yourself with a knife.'

According to Seki (2000, 280), reflexive constructions formed with the prefix *je-* are clearly intransitive since they do not allow for a P argument to occur anywhere in the clause and they index the sole argument of the predicate with the set of prefixes that index S of major class intransitive verbs and A (cf. 5.2 in section 5.2.3).

A different type of synthetic reflexive construction can be seen for Wari', a Chapacuran language of Brazil, in (119):

Wari' (Chapacuran; Everett and Kern, 1997, 189, 191, 240, 339)

(119)   a.   *wac=na-pa*   *quit*
            cut=3SG.NFUT-1SG knife
            'The knife cut me.'

        b.   *wac=**xucun**   pije'*
            cut=REFL.3SG.M child
            'The child cut himself.'

        c.   *quep=ina-in   temem'*
            do=1SG.NFUT-3N bow
            'I made a bow.'

        d.   *quep mi'=**xije**   pain   temem'*
            do   BEN=REFL.1SG OBL.3N bow

'I made myself a bow.'

Unlike Kamayurá where the reflexive is formed with a segmentable reflexive marker, in Wari' the reflexive construction is formed through the use of a special paradigm of enclitic argument markers. In the basic transitive construction in (119a), the third person singular A argument is indexed with the enclitic *=na* while the first person singular P argument is indexed with the suffix *-pa* that attaches to the subject enclitic. The example in (119b) shows that a special subject enclitic is used to index a third person singular S argument in a reflexive construction. The fact that the reflexive construction is intransitive can be seen by comparing the examples in (119c) and (119d). In (119c), the patient participant of the transitive verb *quep* 'do/make' is indexed on the predicate with the 3rd person neuter suffix *-in*. However, in the reflexive construction in (119c), the patient participant is realized as an oblique marked with the preposition *pain*. The modifier use of *mi'* 'give' in the example results in coreference between the agent and the benefactive participant in the reflexive construction.[23]

The pronominal reflexive construction in Timbira is different than the synthetic reflexive constructions shown for Kamayurá (118) and Wari' (119) since it does not result in a decrease of the valency of the verbal base, as shown in (120):

Timbira (Macro-Jêan; Castro Alves, 2004, 69-70)

(120) a. *hũmrɛ tɛ karӠ kuran*
    man ERG deer kill.*nfnt*
    'The man killed the deer.'

   b. *hũmrɛ tɛ **amjĩ** kuran*
     man ERG REFL kill.NFNT
     'The man killed himself.'

   c. *i-tɛ **amjĩ** tɛ kwĩn*
     1-ERG REFL leg break.NFNT
     'I broke my leg.'

Just like the basic transitive construction in (120a), the reflexive construction in (120b) displays a number of properties that indicate that no decrease in valency has occurred: the agent *hũmrɛ* 'man' occurs clause initially and is followed by the ergative postposition *tɛ*, while the patient *amjĩ* 'himself' occurs immediately preceding the predicate. Since *amjĩ* 'himself' occurs in a clausal position reserved for NP arguments in transitive clauses, Timbira has a pronominal reflexive construction. However, unlike the free pronoun set in Timbira, the reflexive pronoun is only a single form for all different persons

---

[23]See Birchall (2014a) for further information on the benefactive construction in Wari'.

(unlike English), and can be used to indicate a reflexive relationship between the A argument and a possessor of the P argument, as seen in (120c), similar to the Wari' construction in (119c).

7.3.4 Reciprocals

Q11.1 *Reciprocal constructions are marked with verbal morphology:* [1, 0]

Q11.2 *Reciprocal constructions are marked with a specialized particle:* [1, 0]

Q11.3 *Language has independent reciprocal pronouns:* [obligatory, optional, none]

Q11.4 *Marking of reciprocal and reflexive constructions is morphologically conflated:* [1, 0]

The defining property of a reciprocal construction is that there are multiple semantic participants of the situation type expressed by the predicate, and these participants are simultaneously agents acting upon the other participants and patients being acted upon by the other participants. In other words, a reciprocal situation expresses the notion of reciprocality, as in 'they see each other' in English, where participant 1 sees participant 2 and participant 2 also sees participant 1. Just like reflexive constructions discussed in the previous section, reciprocal constructions can be expressed using specialized pronouns, forming PRONOMINAL RECIPROCAL CONSTRUCTIONS, or can be expressed using verbal morphology that decreases the valency of the predicate, forming SYNTHETIC RECIPROCAL CONSTRUCTIONS. The nature of the situation expressed by a reciprocal construction requires that there be multiple participants in the clause, generally resulting in explicit plural marking on the NP argument or the expression of plurality in verbal argument marking. Due to similarity in the situation types expressed by both reciprocal and reflexive constructions, these two constructions are often expressed using similar or identical grammatical strategies (cf. Maslova and Nedjalkov, 2011).

Some languages such as Wari' shown above in (119) express reciprocal and reflexive constructions using the same verbal morphology, as shown in (121):

Wari' (Chapacuran; Everett and Kern, 1997, 191-192)

(121) a. *to'=**xucucun*** *hwijima'*
  hit=REFL.3PL.M children

  'The children are hitting each other (or themselves).'

  b. *mana'* ***caracan=xequequem***
  be.angry RECP=REFL.3PL.F

'They (the women) are angry with each other.'

Both the reciprocal and reflexive constructions are minimally formed by using a special set of reflexive enclitic verbal argument markers, resulting in reciprocal constructions and reflexive construction with multiple participants being formally identical, as seen in (121). However, when the construction must be explicit that the situation expressed by the predicate is reciprocal and not reflexive, the postverbal modifier particle *caracan* 'each other' is used.

In Leko, an isolate language spoken in Bolivia, synthetic reciprocal constructions are marked by the suffix *-mo*, resulting in a decrease of the valency of the verbal base, as seen in (122):

Leko (isolate; van de Kerke, 2009, 312)

(122)   a.   *yobas-aya yanapas-**mo**-no-aya-te   dihwo bal-ich-ki*
             man-PL   help-RECP-CONT-PL-DECL peanut plant-INF-DAT

             'The men help each other to plant peanuts.'

        b.   *on   choswai sutih-cha-no-te*
             this woman   wash-PRS-CONT-DECL

             'This woman is washing herself.'

Unlike the Wari' example in (121), Leko has a formal distinction between reciprocal and reflexive constructions. As shown in (122a), the reciprocal construction is marked on the verb with specialized morphology while reflexive constructions are expressed without any specialized morphology, as shown in (122b), and are interpreted as such as a result of no P argument being realized in the clause or indexed on the predicate, as discussed in van de Kerke (2009, 312).

Languages with pronominal reciprocal constructions can distinguish between reciprocal and reflexive constructions through the use of distinct pronominal forms, as seen for Timbira in (123):

Timbira (Macro-Jêan; Castro Alves, 2004, 71)

(123)   a.   *mɛ̃=paʔ-tɛ   **ajpen** pupun*
             PL=1PL.INCL-ERG RECP   see.NFNT

             'They saw each other.'

        b.   *mɛ̃=paʔ-tɛ   **amjĩ** pupun*
             PL=1PL.INCL-ERG REFL see.NFNT

             'They saw themselves.'

In Desano, a Tucanoan language of Brazil and Colombia, reciprocal and reflexive constructions can be formed through the use of a free pronoun together

with the particle *basi* 'self', forming a reflexive/reciprocal pronoun, as seen in (124a) and (124b) for the respective construction type. However, Desano also has an additional strategy to form reciprocal constructions, what can be called an ANALYTIC RECIPROCAL CONSTRUCTION, where the particle *gãbẽ* occurs in the clause primarily before the verb, as shown in (124c. It is also possible for both strategies to be used in the same clause, as seen in (124d):

Desano (Tucanoan; Miller, 1999, 118-119)

(124)   a. *were-tari-ro   gãʔbẽ-a   **bãrĩ**   **basi***
           say-INTEN-NEUT want-1/2.PRS 1PL.INCL REFL
           'We need to confess to one another.'

       b. *tabe-bɨ   **yɨ** **basi***
           cut-1/2.PST 1SG REFL
           'I cut myself.'

       c. *iri bãʔã ohogoro-ge **gãbẽ** bokatĩri-bɨ   wapikɨ-rã-ge*
           this trail end-LOC   RECP meet-1/2.PST four-AN.PL-LOC
           'At the end of this trail, we four met up with each other.'

       d. *gɨa   **basi gãbẽ** weretabũ ii-bɨ*
           1PL.EXCL REFL RECP discuss   do-1/2.PST
           'We (excl.) conversed amongst ourselves with one another.'

While comparable analytic strategies are logically possible to form reflexive constructions, these strategies were only observed as a strategy to form reciprocal constructions in the language sample.

7.3.5 Summary of valency decreasing strategies

Putting aside the pronominal strategies sometimes used for reflexive and reciprocal constructions for the time being, the valency decreasing strategies discussed in this section are schematically represented in Table 7.3. The same convention used in Table 7.2 for the representation of the syntactic functions of the semantic participants. However, as noted in section 7.3.1 the status of the agent participant in passive constructions is akin to the treatment of an oblique even though its implication, either though event semantics or argument realization, is semantically obligatory. For this reason, the agent of a passive is represented with curly brackets {} in the table.

Passives: Within the language sample, a total of 34 languages are identified as having a passive construction based on the descriptive materials available.

- The synthetic passive strategy is by far the most common in South American languages, being identified in 31 of the languages surveyed. Only three

Construction	Valency frame	Promoted	Demoted	Val.
Basic intr.	[agent]	-	-	1
Basic tr.	[agent, patient]	-	-	2
Passive	[patient]{agent}	-	agent	1
Anticausative	[patient]	-	agent	1
Antipassive	[agent]	-	patient	1
Reflexive	[agent/patient]	-	-	1/2
Reciprocal	[agent/patient]	-	-	1/2

Table 7.3: Overview of valency decreasing strategies

languages are identified as having exclusively periphrastic passive constructions: Nasa Yuwe, Imbabura Quechua and Cubeo. Huallaga Quechua and Urarina have both a synthetic and a periphrastic passive construction.

- The languages in the sample show an equal distribution between those with a personal passive construction and those with an agentless passive construction.

- A total of six languages in the sample have passive constructions that are conflated with reflexive constructions. All of these languages use a synthetic strategy to form passive constructions. In these languages, the different situation type expressed by the construction is highly dependent on the semantics of the event expressed by the predicate and the semantic properties of its arguments. When the derived S argument is inanimate, thus unable to affect change upon itself, the interpretation tends towards a passive reading, while when the derived S argument is human or animate, the interpretation tends towards a reflexive or ambiguous reading.

Antipassives: An antipassive construction is identified in 11 languages in the sample based on the available descriptive materials.

- Within the 11 languages with an antipassive construction, only five languages show ergative alignment in case marking and/or indexation, restricted to the major class of intransitive verbs: Cavineña, Tiriyó, Timbira, Matses and Katukina-Kanamari.

- Among the languages without ergative alignment between major class intransitive verbs and transitive verbs, five languages show split intransitivity in indexation. None of the two remaining languages display case marking of core arguments, and Movima has a hierarchical indexation system with neutral alignment (see section 4.1) and Moseten has a fused indexation system that only indexes person for A and P but not S (Sakel, 2011).

- Antipassive constructions are conflated with reflexives constructions in only four languages in the sample: Cavineña, Tiriyó, Moseten and Tapiete. While much work is still to be done on the historical development of antipassive constructions, one common pathway for their development is through functional extension of a reflexive construction, as Terrill (1997) shows for a number of Pama-Nyungan languages of Australia. The large proportion of South American languages that do not show syncretism between antipassive and reflexive constructions suggests that diachronic sources for antipassive markers be searched for elsewhere in the grammar as well, such as through the extension of aspectual markers such as iterative, durative and habitual, which are often syncretic with antipassive constructions (cf. Polinsky, 2011a).

Reflexives: Within the language sample, a total of 58 languages are identified as having a reflexive construction based on the descriptive materials available.

- A large majority of languages in the sample rely primarily on synthetic constructions to express reflexive situations (86%). A total of six languages allow for the use of an emphatic reflexive pronoun together with the synthetic construction.

- Only eight languages in the sample rely exclusively on a pronominal strategy to form reflexive constructions.

Reciprocals: A total of 64 languages in the sample are identified as having reciprocal constructions based on the descriptive materials available.

- A large majority of the languages in the sample rely primarily on synthetic constructions to express reciprocal situations (72.9%).

- A total of 32 languages have reciprocal constructions that are identical to the their respective reflexive constructions.

- Six languages in the sample have a reciprocal construction formed exclusively with a pronominal strategy. An additional two languages have emphatic reciprocal pronouns. Except for Desano, shown in (124), no language uses identical pronouns for both reflexive and reciprocal constructions.

- Unlike reflexive constructions where analytic constructions are unattested in the sample, a total of four languages can express reciprocal constructions solely with an analytic strategy. As shown in (119b), Wari' is the only language in the sample that has an additional emphatic reciprocal particle even though the reciprocal construction primarily uses a synthetic strategy.

On middle voice: The term middle voice is not used in this chapter as a
distinct construction type due to the wide variety of different situation types
included under such a term. However, it is still possible to examine the presence
of a category that corresponds to an aspect of what is commonly called middle
voice, but in a more narrow sense. Using the data in the structural question-
naire, it is possible to identify languages that conflate reflexive marking with
either anticausative or antipassive constructions, following the suggestion in
Kemmer (1993) that most descriptive uses of the term 'middle voice' imply the
conflation of a reflexive situation type with one or more of her 'middle' situa-
tion types, as listed in section 7.3.1. The distribution of a such a construction
type is further discussed in section 8.2.4 of the following chapter.

7.4   Conclusions

South American languages use a wide variety of different strategies to change
the valency of a predicate. This chapter has examined a number of major va-
lency changing operations that result in either the increase of valency of the
verbal base, such as in causative constructions, or the decrease in the valency
of the verbal base, such as in passive constructions. The overall picture is that
South American languages rely heavily on synthetic construction types to ex-
press the situation types discussed above through a coded alternation of the
valency of the predicate. Especially with valency decreasing constructions, but
also with valency increasing constructions, there is a tendency for many lan-
guages to conflate multiple situation types into a single morpological construc-
tion, forming polysemous categories that are often labelled in the literature
as 'intransitivizer', 'transitivizer' or 'middle'. This chapter has shown the ap-
plicability of defining these different valency-changing strategies through the
semantic situation types that they express and the effects that they have on
the valency of the predicate. The identification of these prototypical situation
types helps to control the standard of comparison across a diverse sample of
languages, thus aiding in the evaluation of the structural variation that these
different constructions present.

CHAPTER 8

Geographic patterns

The distribution of linguistic features across the globe is a result of the interaction of a number of different processes, including inheritance from a common ancestor, contact-induced change and language-internal change over time. One aspect of South American languages that has puzzled researchers for over a century is the striking geographic distribution of certain structural features across apparent phylogenetic groupings. Early ethnographic work in South America resulted in the proposal by d'Orbigny (1839) that the indigenous population of the continent can be roughly grouped into three major 'races' based primarily on their physiological characteristics: the Ando-Peruvian, the Pampean and the Brasilo-Guaranian races (cf. Brinton, 1891). Lafone Quevedo (1896) builds off of this misguided cultural taxonomy by noticing a striking difference in the locus of person indexation across these three proposed groups and suggests that the development of the mixed locus marking in Matacoan and Guaycuruan languages is in part due to its geographic position between the prefixing Guaranian language type of eastern Brazil and the suffixing Quechuan language type of the Andes. While this observation does not make a strong claim for structural convergence, it is notable in the sense that it includes the role of language contact and geographic proximity in explaining the development of language structure long before the birth of the modern field of areal linguistics.[1]

While Tovar and Tovar (1984) continued with the tradition of Lafone Quevedo by grouping languages into broad types according to their structural profile,

---

[1]It is worth noting that during this time period the field of anthropology in general was exploring the applicability of diffusion as an explanation for the observed distribution of different cultural traits in the New World, as exemplified in works such as Boas (1911) and Nordenskiöld (1920, 1924).

much of the more recent typological work on South American languages has involved the identification of specific structural features associated with different geographic regions of the continent. Derbyshire (1987) is a pioneering study in this regard. Using an impressively large sample given the state of description at the time, Derbyshire analyzed 40 Amazonian languages from a number of different families for typical Greenbergian typological parameters, with an additional emphasis placed on identifying ergative patterns in case marking, indexation and constituent orders. A number of further studies have focused on regional features of Amazonian or lowland languages, such as Dixon and Aikhenvald (1999) and Payne (1990). There is also a body of work that examines the languages of the Andes, such as Torero (2002), Adelaar and Muysken (2004) and Adelaar (2008), among others, while Campbell and Grondona (2012) look at the structural features of the languages of the Gran Chaco and the Southern Cone. A useful overview of these different works is given in Campbell (2012b).

This chapter examines the geographic distribution of a selection of features from the three main grammatical domains covered in this thesis—indexation, case marking and valency change. It also looks at the distribution of different constituent order patterns, a topic that has not been discussed in much detail in the thesis but is nonetheless included in the structural questionnaire. To test the distribution of these different linguistic properties, a definition of an areal linguistic feature is operationalized and tested across the geographic regions discussed in section 1.2.2 of the introduction. These features are further tested across larger geographic regions that aggregate multiple smaller regions into macroregions.

8.1 Definitions and important concepts

A wealth of proposals have been made regarding the distribution of linguistic features across the South American continent, often with the aim of identifying those that are 'areal', 'typical' or 'characteristic' of specific regions, with a number of notable studies mentioned in the previous section. Most discussions on the geographic distribution of particular linguistic features revolve around the identification of linguistic areas. A LINGUISTIC AREA, also known as a Sprachbund, is defined by Thomason (2001, 99) in the following way:

> "[A] linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor."

A linguistic area is postulated based on a cluster of shared linguistic features, what are referred to as AREAL FEATURES in this study. Most definitions of a linguistic area are also explicit that some of the three or more languages of the area must come from different language families, or at the very least, from different branches within a family (cf. Campbell et al., 1986; Enfield, 2005).

Based on the above definition, a linguistic area has two essential properties: a) areal features are shared by three or more languages within this region, and b) these features must have resulted from language contact rather than accident or inheritance. The ways in which these two properties are established in this study are discussed in the following sections.

8.1.1 Identifying potential linguistic areas

Before one can begin to explore the distribution of linguistic features within or across particular geographic regions, a number of testable hypotheses must first be established. The most common procedure for this is through the consideration of extralinguistic criteria such as ethnography, archaeology, demography, genetics, ecology, geography and economics (Bickel and Nichols, 2006). For such a hypothesis to be plausible, it should reflect what is known about the history of different populations and the ways that they interact across linguistic boundaries.

In this study, the 74 members of the language sample are divided across seven distinct geographic regions. These regions are grounded in the geographic reality of the continent and are delineated using natural features that can act as impediments to cross-regional interactions, such as major rivers, mountain ranges or distinct ecological zones. These regions also reflect the sociocultural reality of the continent, roughly corresponding to, or including, culture areas in the ethnological literature, especially those proposed in Murdock (1951), as discussed in section 1.2.2. Culture areas are good indicators of interactions among different populations since they are primarily identified on the basis of shared cultural traits that themselves have diffused through multiethnic social relations (cf. Sherzer, 1973).

Table 8.1 presents the number of different languages and language families included in each region together with the three-letter identifier codes that are used in tables to refer to each region throughout the rest of this chapter.

REGION	IDENTIFIER	LS	FAMILIES
Northern Andes	NAd	7	5
Northern Amazonia	NAz	7	5
Central Andes	CAd	6	5
Western Amazonia	WAz	20	13
Southern Amazonia	SAz	24	14
Chaco-Planalto	ChP	8	4
Southern Cone	SCo	2	2

Table 8.1: Regional distribution of sampled languages

As can be seen in Table 8.1, Southern and Western Amazonia together contain the majority of languages included in the sample. This is partially a result of historical reasons. Many languages of the coastal areas of the continent

disappeared before any systematic linguistic documentation could be carried out on them. Large migrations of Quechuan and Aymaran speaking populations around the time of Spanish arrival also had a dampening effect on the linguistic diversity of the Central Andes (see section 1.2.2). This means that the ethnolinguistic diversity of the continent was best maintained in the less accessible portions of the continent such as the headwaters and hinterlands of the Amazonian tributaries. This sampling distribution is also based on the synchronic ethnolinguistic composition of the continent. Campbell (2012a) states that there are around 420 South American languages spoken today and Dixon and Aikhenvald (1999) count approximately 300 languages within the Amazon Basin, suggesting that an estimated seven out of ten of all languages in South America are spoken in the Amazonia. The proportion of Amazonian languages in this sample is thus largely representative of this distribution (68.9%).

In addition to the regions as shown in Table 8.1, the languages of the continent can be further grouped into MACROREGIONS. Macroregions are composed of two or more of the previously-defined regions and subject to the same criteria for establishing the areality of a linguistic feature. The Andean and Amazonian macroregions are included as good candidates for linguistic areas since a large body of previous work has identified a number of different grammatical features that appear characteristic to these regions. They have already described them as linguistic areas or potential linguistic areas in previous work. In this study, the three Amazonian regions—Northern, Western and Southern—are aggregated to form the Amazonian macroregion. Additionally, the two Andean regions, Northern and Central are aggregated to form the Andean macroregion, with the addition of the Southern Cone language Mapudungun since it is spoken in the southern Andean highlands.

Two additional macroregions are also included as possible candidates for linguistic areas: Eastern South America and Western South America, which together comprise the entire language sample. In an early evaluation of the morphological characteristics of South American languages, particularly those of the lowlands, Payne (1990, 214) observes that the a number of typological features "have a roughly eastern versus a western geographic distribution". In a comparative study of noun phrase structure with a sample of 55 South American languages, Krasnoukhova (2012) identifies a number of noun phrase features that show a skewed East-West distribution on the continent. For the western portion of the continent, she identifies the following features: pre-head position for all nominal modifiers; absence of gender and classifiers; property concepts expressed as nominals; and a lack on inalienably possessed noun classes. For the eastern portion of the continent, she identifies the following features: pre-head position for demonstratives; lexical possessors and numerals; post-head position for property words; presence of gender and classifiers; property concepts expressed verbally; and the presence of inalienably possessed noun classes.[2] This information on the noun phrase in Krasnoukhova (2012) as well as the

---

[2]See Krasnoukhova (2012, 261-266) for further discussion of these findings.

preliminary study in Payne (1990) suggests that in addition to the Amazonian and Andean macroregions, promising results may also be obtained by considering the distributions of argument marking features across the Eastern and Western macroregions. A summary of the members of each macroregion are shown in Table 8.2, together with information on the number of languages (Ls) and families in each region.

Macroregion	Identifier	Ls	Families	Members
Amazonia	AMZ	51	27	NAz, WAz, SAz
Andes	AND	14	10	NAd, CAd, Mapudungun
Eastern S. A.	ESA	39	18	NAz, SAz, ChP
Western S. A.	WSA	35	24	NAd, CAd, WAz, SCo

Table 8.2: The macroregions used in this study

It is worth noting that these are not the only possible configurations of geographic regions for the subdivision of South American languages. For example, regions based primarily on elevation are not considered in the final analysis, but could make for a fruitful line of investigation in future work, assuming one had evidence that elevation played a significant role in the interactions between populations speaking different languages. For this study, major ecological boundaries are coupled with previous proposals of cultural and linguistic areas to generate good candidates for regions that potentially show an areal distribution of linguistic features.

8.1.2 The size and scope of potential linguistic areas

An important issue to discuss with regard to identifying linguistic areas is the size and scope of a potential area. Areal features are sometimes distinguished from 'macro-areal features' or 'regional traits' in that areal features are shared by languages that are currently (or historically) in contact with each other, while the latter types occur "across a scattering of languages within a certain geographical area" (Aikhenvald and Dixon, 2010). No such terminological distinction is adopted in this chapter, but rather, the areality of specific linguistic features is discussed in terms of the scope of the region under consideration, whether it be a non-aggregated region such as Northern Amazonia or an aggregated macroregion such as the Andes. The true difficulty lies in deciding whether geographic proximity alone is sufficient to posit a degree of social interaction that could have resulted in contact-induced linguistic change.[3]

---

[3]The sociolinguistic processes at work in these interactions often include bilingualism (or multilingualism), codeswitching and language shift. These processes can result in a number of linguistic outcomes such as the transfer of phonological, lexical and grammatical material and the convergence of different morphosyntactic and discursive patterns (cf. Winford, 2005; Sakel, 2007; Muysken, 2008, among many others). As Thomason and Kaufman (1988, 35) duly note "[...] it is the sociolinguistic history of the speakers, not the structure of their

The unique linguistic situation of the Vaupés River, the most widely-discussed example of a linguistic area in Amazonia, serves as a key example to highlight some inherent difficulties with drawing a distinction between areal and 'macro-areal' features for South American languages. One of the often-discussed areal features of the Vaupés is the restructuring of the case marking systems of the non-Tucanoan languages of the area towards the Tucanoan profile of accusative marking with a neutral ditransitive case alignment (cf. Aikhenvald, 2006a,b, 2011; Epps, 2006; Zúñiga, 2007; Stenzel, 2008). Indeed, the changes that have taken place in both Tariana, an Arawakan language, and Hup, a Nadahupan language, are very suggestive of structural convergence under language contact when these languages are compared to related languages spoken outside of the area. However, a marked neutral ditransitive case marking pattern shows a significant distribution over the whole region of Western Amazonia ($p < 0.001$, see 8.2.3 below), i.e. it is also found in other languages of Western Amazonia such as Aguaruna (Jivaroan) and Jarawara (Arawan), and occurs in a significantly higher proportion of languages in this region than in the rest of the continent. While there is additional evidence that suggests that the non-Tucanoan languages of the Vaupés did indeed change towards the Tucanoan profile, such as the number of oblique case distinctions and the use of differential object marking, the main motivation behind calling the marked neutral ditransitive alignment pattern an areal feature, rather than a 'macro-areal' feature in the terms of Aikhenvald (2011), is the fact that it is shared by a number of languages that have been in historical with each other. This shows that the use of the term 'linguistic area' in this context is not a description of the geographic distribution of a particular linguistic feature or set of features, but rather, it is an explanation for this distribution. Since the occurrence of this specific feature cannot be easily explained by historically-attested social interactions by languages outside of the Vaupés with those inside of the area, it would not be considered an areal feature by those scholars who adhere to what Campbell et al. (1986, 533-534) call the 'historicist approach', i.e. those that adopt a strict requirement for documented social interaction in diagnosing linguistic areas.

In South America, the difference between areal and 'macro-areal' features is not a very clear-cut distinction. A major complicating factor of this is the time depth of language contact. The firsthand accounts of early explorers and the available archaeological evidence suggest that the ethnolinguistic distribution of South American peoples is quite different today than it was from the formative periods until the early colonial period (Roosevelt, 1994; Whitehead, 2003;

---

language, that is the primary determinant of the linguistic outcome of language contact." In this sense, certain linguistic features are thought of as easier to borrow, such as lexical nouns, while other features such as inflectional morphology are more difficult to borrow. A useful summary of different proposals on borrowing hierarchies is given in Wilkins (1996). However, the well-known case of structural change in the Arawakan languages of the Vaupés region due to contact with Eastern Tucanoan languages is a prime example of how cultural constraints such as taboos on lexical borrowing and code-switching can influence the linguistic outcomes of language contact (Aikhenvald, 2003b, 2006b).

Neves, 2008; Heggarty and Beresford-Jones, 2012). Denevan (1992) estimates that the indigenous population of South America could have been as high as 24-25 million people at the time of European arrival. Exogenous epidemics and violent conflicts resulted in a drastic demographic decline in subsequent generations, such that the indigenous population of Amazonia alone was reduced to a low point of an estimated 200,000 people during the mid-twentieth century from an original population of two million people or more (Aikhenvald, 2012, 5). Such a dramatic shift in population density together with large-scale territorial encroachment undoubtedly altered the sociolinguistic landscape of South America considerably before a reliable historical record could be established. In certain cases, archaeological evidence can be used to infer prehistoric interactions among different ethnic groups (e.g. Neves, 1998), but this type of data is not always available for every region of the continent. Since the historical period for South America is maximally five hundred years, and in practice for many regions it is considerably less, any social interaction before this period must be inferred from either the present day situation of indigenous groups, which was likely altered as a result the European conquest, or archaeological data for which our territorial coverage is limited.[4] For this reason, the regions used in this study are not restricted to indigenous societies that are in present contact with each other, but rather, based on geographical and cultural conditions that are suggestive that interaction between different populations may likely have occurred in the past. These more inclusive criteria for establishing potential linguistic areas not only help to minimize the potential effects of being reliant on the incomplete historical record in South America but also account for the fact that the geographic location of populations of speakers can vary over time, with different groups coming into and out of contact with each other as social, economic and environmental conditions change.

8.1.3   Establishing areality

Which linguistic features to consider and how to establish that these features indeed display an areal distribution is of central concern when attempting to diagnose a linguistic area. Sufficiently common features that occur in most or all languages, such as the presence of plosive consonants, will be of little use for diagnosing a linguistic area since it will be difficult to distinguish whether it could have resulted from chance, inheritance or contact. Rather, the shared feature must be of low enough frequency in the languages outside of the area in order to show a marked geographic distribution within it.

   The precise way to define the marked geographic distribution of a feature, what is referred to here as the AREALITY of the feature, differs among scholars. In her discussion on the geographic distribution of features belonging to a linguistic area, Aikhenvald (2006a, 11) states that a feature shared among

---

[4]Recent work combining archaeological, ethnohistorical and linguistic data into geographic information systems (GIS) shows considerable promise in helping to identify some of these prehistoric interactions (cf. Eriksen, 2011), but much work in this area is still needed.

members of different language families within a linguistic area should not be found in "languages from these families or subgroups spoken outside the area." Such a characterization can prove useful when comparing languages from large families with a wide geographic distribution, such as the Tupian or Arawakan families of South America, but for the numerous smaller families that are principally confined to a specific geographic region, this criterion is difficult to apply.

Three components are needed in an approach to define the areality of a linguistic feature: a) it must maintain the basic notion that when a language occurs within a particular linguistic area, it is more likely to present a particular linguistic feature than a language that occurs outside of that area, b) such an approach must be independent of the size and distribution of any single language family, and c) it must be formulated in such a way to be testable and explicit. The approach adopted in this study stems from the work in Bickel and Nichols (2006), where they define the areality of a linguistic feature by evaluating how good of a predictor a particular geographic region is for explaining the observable typological distribution of the feature. If a geographic region is a good predictor variable for whether a specific feature appears inside of the region rather than outside of it, that feature shows areality for that region. Put a different way, if a feature has a greater probability that it occurs in a randomly selected language within a particular region than in a randomly selected language outside of it, it shows areality for that region. The use of the term 'areal feature' in this chapter is not meant to imply that the every areal feature occurs within a linguistic area, since, as discussed above, a linguistic area is generally conceived of as a socio-historic explanation for the distribution of areal features but not an observation of the distributions themselves.

Since the areality of a linguistic feature does not necessarily attribute its observed distribution to language contact, other possible explanations must also be accounted for. The possibility that the areality of a feature is merely a result of coincidence or chance is eliminated through the standard statistical method of significance testing. In this study, significance is tested by means of a Fisher's Exact Test using a 2x2 contingency table (presence/absence x region/non-region) for each feature examined, with $p < 0.05$ interpreted as significant (cf. Janssen et al., 2006). It is important to note that the distribution of a particular linguistic feature within a region is tested against the distribution of that features across the rest of the continent, not including that region. For example, a test of the areality of indexation only as an argument marking strategy for Amazonian languages compares the proportion of this feature among languages within the Amazonian macroregion to the proportion of this feature among all languages of the continent that fall outside of the Amazonian macroregion.

The possibility that the areality of a feature is a result of shared inheritance through descent from a common ancestor is eliminated through the use of a genealogically-stratified sample (see section 1.3). Additionally, features that occur among languages from only a single language family within a particular

region are further excluded from analysis.

The final possible explanation that the areality of a feature is not due to contact-induced change is that the observed feature distribution is a result of universal dependencies or preferences among correlated linguistic structures, such as the well-known tendency of languages with object-verb constituent orders to use postpositions to mark oblique participants of the clause (Greenberg, 1963; Hawkins, 1983; Dryer, 1992). This is eliminated as a possible explanation by simply not considering features that are known to be associated with each other. If these possible explanations can be accounted for, then contact-induced change can help to explain the areality of a feature within a region.

8.2   Geographic patterns in argument marking

This section examines the geographic patterns observable in the distributions of the argument markers examined in the previous chapters of this thesis. Before discussing the geographic distribution of argument marking features in the sample, it is important to note that many previous studies on the distribution of language structures in South America have focused on a wide variety of different features from many different domains of grammar. For example, in the introduction to their well-known overview volume *The Amazonian Languages*, Dixon and Aikhenvald (1999, 8-10) consider the following grammatical features as indicative of an Amazonian linguistic area and "shared by all (or most) languages in the area":[5]

(125)   Areal grammatical features of Amazonian languages per Dixon and Aikhenvald (1999)

    a. Polysynthetic and *head marking*

    b. Classifiers and/or grammatical gender

    c. Few oblique cases

    d. Possession marked on possessed noun

    e. *One core argument marked on the verb*

    f. *Complex verbal marking patterns, often involving ergative alignments*

    g. Marker sets marking possessor shares forms with verbal indexes

    h. Most languages have prefixes

    i. Prefix verbal indexes occur further from the root than valency changing markers

    j. Verbal categories such as tense, mood and aspect are expressed as optional suffixes

    k. Subordinate clauses involve nominalized verbs

---

[5]Those features listed in *italics* are further discussed in the remainder of this chapter.

l. Adverbs and adpositions may be incorporated into verb, often pre-
ceding the root

The Amazonian features listed in (125) are then contrasted with what Dixon
and Aikhenvald (1999) identify as grammatical features of the Andean linguistic
area, based primarily on those found in the dominant Quechuan and Aymaran
families of the region:

(126)   Areal grammatical features of Andean languages per Dixon and Aikhen-
vald (1999)

a. Synthetic and *combine head and dependent marking*

b. No classifier or gender system

c. Extensive set of case markers

d. Possession marked on both possessor and possessed

e. *Nominative-accusative indexation pattern*

f. Similar but not identical sets for verbal indexes and possession

g. *No verbal prefixes*

h. Tense and aspect marked on verb with obligatory suffixes

i. No incorporation of nouns, adverbs or adpositions

As can be seen in (125) and (126), an impressive list of distinguishing gram-
matical features have been identified in Dixon and Aikhenvald (1999) for their
proposed Amazonian and Andean linguistic areas. While the authors clearly
demonstrate extensive knowledge of Amazonian languages, with both having
carried out personal fieldwork in Amazonia, the empirical basis for these claims
are never demonstrated beyond a few illustrative examples. Some of these pro-
posed features are problematic. For example, as Payne (2001) rightly notes,
feature (125i) above regarding the order of derivational and inflectional affixes
is in fact a well-attested statistical universal. Furthermore, the concept of the
Andean linguistic area in Dixon and Aikhenvald (1999) does not take into ac-
count the wealth of information available on the language families and isolates
of the region beyond the Quechuan and Aymaran languages. Payne (2001)
recognizes that a number of the languages of the Northern Andes also show
a number of the proposed Amazonian areal features. Adelaar (2008) presents
a more careful evaluation of possible Andean areal features that considers a
wide variety of different languages from the region, but he cautiously concludes
that there is "still very little evidence for recognizing and delimiting linguistic
typological areas, let alone, an Andean linguistic area".

The following sections examine the distribution of a number of structural
argument marking features across the language sample, organized according
to the domain of marking that they pertain to. Previous claims such as those
listed above are included in the discussion whenever relevant. For each feature
under consideration, a summary of the data is provided, followed by an analysis

of its areality using the procedure outlined in the previous section. Since the Southern Cone region has only two representative members in the sample, features in this area are not analyzed for their areality. However, data from the Southern Cone languages are included as members of the general population of South American languages to be contrasted with the specific region under investigation.

The tests for areality in this section only consider the presence of a particular argument marking feature and do not consider the absence of a feature. For this reason, a one-tailed Fisher's Exact Test is used. Results that show that the feature in question shows areality for the region ($p < 0.05$) are indicated in the analysis tables in **boldface**. Results that do not meet the threshold of areality but nonetheless are suggestive for further investigation are indicated in the analysis tables in *italics*. When a region or macroregion shows a lower proportional presence for a feature than the corresponding non-region, this is indicated in the tables with a hash mark '-'. P-values that are less than or equal to 0.001 are represented as 0.001 in the results tables.

8.2.1 Geographic patterns in head and dependent marking

Before going into the specific domains of argument marking, the first analysis presented here examines the overall argument marking typology of South American languages, focusing on the head versus dependent marking distinction. All languages in the sample employ indexation and/or case marking as an argument marking strategy. As mentioned in the previous section, Dixon and Aikhenvald (1999, 8-10) identified exclusively head marking as an areal feature of Amazonia, and head marking together with dependent marking as an areal feature of the Andes. A summary of the distribution of indexation (head marking) and case marking (dependent marking) as argument marking strategies in the languages of the sample across the different regions is given in Table 8.3.

FEATURE	NAD	NAZ	CAD	WAZ	SAZ	CHP	SCO
Only indexation	0	6	3	6	19	7	1
Only case	2	0	0	4	2	0	0
Indexation and case	5	1	3	10	3	1	1

Table 8.3: Argument marking strategies

A map showing this distribution is given in Appendix A. The areality of the different argument marking strategies can be tested using the procedure described in section 8.1.3. The results of this analysis are presented in Table 8.4.

The results in Table 8.4 show that a combination of indexation and case marking is an areal feature of both the Northern Andes and Western Amazo-

Region	Indexation	Case	Both
NAd	-	0.163	**0.033**
NAz	*0.108*	-	-
CAd	-	-	0.297
WAz	-	*0.131*	**0.048**
SAz	**0.006**	-	-
ChP	*0.065*	-	-
AMZ	0.215	0.521	-
AND	-	0.471	**0.033**
ESA	**0.001**	-	-
WSA	-	*0.099*	**0.001**

Table 8.4: The areality of argument marking strategies

nia at the regional level. Indexation without case marking is an areal feature of Southern Amazonia. At the macroregional level, the results show that indexation without case marking is an areal feature of Eastern South America but not Amazonia. Argument marking strategies that include both case marking and indexation is an areal feature for the languages of the Andes as well as for Western South America, in accordance with the observation shown in (126) from Dixon and Aikhenvald (1999, 10).

8.2.2   Geographic patterns in verbal marking

This section examines the distribution of a number of structural parameters related to the indexation of arguments on the verb, including: the locus (position) of marking, alignment patterns, marking patterns and the expression of clusivity.[6]

Locus of marking: Since Lafone Quevedo (1896), whether arguments are indexed through prefixes, suffixes or both has been discussed in terms of its geographic distribution for South American languages. For example, in her discussion of morphological features of lowland South American languages, Payne (1990, 221) mentions prefixing as a 'characteristic feature' of verbal argument marking. Dixon and Aikhenvald (1999, 10) consider a lack of prefixes an areal feature of Andean languages. To explore the distribution of this feature, languages were scored for whether each argument type (major class S, A and P) is marked by a prefix or a suffix. A summary of the distribution of the locus of indexation is shown in Table 8.5.

---

[6]In a different study, Birchall (2014b) examined the distribution of verbal argument marking features in a sample of 65 South American languages. This section presents a reevaluation of some of the claims made through the use of a larger language sample. Portions of the text in this section draw directly from this article.

Feature	NAd	NAz	CAd	WAz	SAz	ChP	SCo
S prefix	0	6	1	7	13	5	1
A prefix	0	6	1	9	9	4	1
P prefix	1	6	2	3	13	5	1
S suffix	4	1	5	7	5	0	1
A suffix	4	1	5	7	6	0	1
P suffix	2	1	3	6	6	1	1
S variable	1	0	0	0	1	2	0
A variable	1	0	0	0	1	2	0
P variable	0	0	0	0	2	2	0
Any prefix	1	7	2	10	17	8	1
Any suffix	5	2	5	10	9	3	1

Table 8.5: Locus of marking

A map showing this distribution is given in Appendix A. The areality of these features can be tested using the procedure described in section 8.1. If a language indexes any core argument S, A or P as a prefix, it is considered to have prefixes as argument markers. If a language indexes any core argument S, A or P as a suffix, it is considered to have suffixes as argument markers. Languages that mark some core arguments as prefixes and others as suffixes, such as Yurakaré that marks S and A as suffixes and P as prefixes (cf. ex.3.1), are considered to have both prefixes and suffixes as argument markers. Languages with markers split between prefixes and suffixes for a particular argument role, as in Itonama where P can be indexed either by a suffix or a prefix (cf. ex.43), have been treated as having both a prefix and a suffix to index that particular argument. Table 8.6 presents the results of an analysis for whether the languages of each region show areality for expressing core arguments as either prefixes or a suffixes.

Region	Prefix	Suffix
NAd	-	0.173
NAz	**0.030**	-
CAd	-	*0.077*
WAz	-	0.491
SAz	0.210	-
ChP	**0.017**	-
AMZ	0.176	-
AND	-	**0.010**
ESA	**0.001**	-
WSA	-	**0.033**

Table 8.6: The areality of the locus of marking

The results in Table 8.6 show that prefixes as verbal argument markers is an areal feature of Northern Amazonia and the Chaco-Planalto at the regional level, and it is areal for Eastern South America at the macroregional level. The results also show that suffixes as verbal argument markers is an areal feature of both Western South America and the Andes.

Alignment: Amazonian languages are well-known for their diversity of alignment types and complex marking patterns. Earlier work on Amazonian languages noted that "ergatively organized systems, in whole or part, are quite common" (Derbyshire, 1987, 316). This conception of Amazonian languages continues today such that Aikhenvald (2012, 213) calls Amazonia "the most ergative area in the world". In contrast, Andean languages are often noted for the frequent occurrence of accusative alignment in their argument marking (cf. feature 126e above). A summary of the different alignment patterns found in the languages of the sample is given in Table 8.7:[7]

FEATURE	NAD	NAZ	CAD	WAZ	SAZ	CHP	SCO
Accusative	5	5 (4)	5	13	14	4	0
Ergative	2 (0)	3 (1)	0	4 (1)	14 (4)	5 (2)	0
Tripartite	0	1	0	1 (0)	2 (1)	3 (1)	0
Neutral	0	1	1	2	2	0	2

Table 8.7: The alignment of verbal argument marker sets

Table 8.8 on the following page presents the results from the analysis on the areal distribution of alignment in verbal argument marker sets for the indexation of A, P, major class S and minor class S. It is worth noting that languages with split intransitivity are treated as displaying two different alignment types for this calculation: the alignment of major class S with transitive core arguments and the alignment of minor class S with transitive core arguments. Languages that do not index any core arguments were not considered in the following analysis.

The results in Table 8.8 show that ergative alignment in verbal argument marker sets is an areal feature for Southern Amazonia at the regional level and for Eastern South America at the macroregional level.

Transitive marking pattern: As discussed in Chapters 3 and 4, South American languages display a wide variety of different verbal marking patterns. Dixon and Aikhenvald (1999, 8) consider marking patterns that index only a single argument on transitive verbs as an areal feature of the languages of the Amazon. While this is undoubtedly the case for a number of prominent

---

[7]The value in parentheses indicates the count of different alignment patterns excluding minor classes of intransitive verbs. The number without parentheses indicates the alignment pattern between major class S and transitive core arguments and minor class S and transitive core arguments.

Region	Accusative	Ergative
NAd	0.180	-
NAz	-	-
CAd	0.454	-
WAz	0.266	-
SAz	-	**0.010**
ChP	-	*0.123*
AMZ	0.523	0.249
AND	0.274	-
ESA	-	**0.001**
WSA	0.178	-

Table 8.8: The areality of accusative and ergative alignment in indexation

Amazonian language families like Tupian, Cariban, and Tucanoan, a question still remains as to whether this distribution is significant if the languages of the rest of the continent are taken into account. Table 8.9 presents a subset of features related to the identification of the transitive verbal argument marking pattern, as discussed in section 3.2, including the presence of markers, the fusion of arguments and the relative order of markers.

Feature	NAd	NAz	CAd	WAz	SAz	ChP	SCo
A before P	1	2	2	5	5	4	1
P before A	1	1	1	4	3	1	0
Portmanteau	0	0	2	0	1	0	0
Hierarchical	2	4 (5)	0	0	6 (7)	2	1
Only A	1	0	1	7	1	0	0
Only P	0	0	0	0	6	1	0
2 arguments	2	3	5	9	9	5	1
1 argument	3	4	1	7	13	3	1
0 arguments	2	0	0	4	2	0	0

Table 8.9: The presence, fusion and relative order of person indexes on transitive verbs

As can be seen in Table 8.9, South America as a whole is almost evenly split between languages that index only a single transitive argument (n=32) and those that index both transitive arguments (n=34), with a smaller proportion not indexing any transitive arguments at all (n=8). To explore the distribution of the presence of single or double argument indexation, the languages in the sample were divided into two groups. The first group is composed of languages that index only a single transitive argument on the verb, including those that show hierarchical marking patterns that only allow for the indexation of a single verbal argument. The second group is composed of languages that can index

two transitive arguments on the verb, including languages with fused argument marking patterns. Languages with hierarchical marking in a specific slot that allow two transitive arguments to be indexed on the verb in direct scenarios, specifically Itonama and Yanam, are treated as members of the two argument group. However, the number of languages with hierarchical marking in any marker slot, irregardless of whether an additional argument can be indexed in a transitive clause, is indicated in parentheses in the corresponding row. Languages that do not index any arguments on the verb, such as Northern Embera and Trumai, are included in the analysis. The areality of the single and double argument indexation patterns is given in Table 8.10:

Region	1 arg.	2 args.	0 args.	Hierarchical
NAd	-	-	*0.139*	0.510
NAz	0.349	-	-	**0.006**
CAd	-	*0.067*	-	-
WAz	-	-	*0.102*	-
SAz	*0.144*	-	-	0.276
ChP	0.280	0.267	-	0.510
AMZ	0.233	-	0.521	0.559
AND	-	0.220	-	-
ESA	*0.107*	-	-	**0.005**
WSA	-	0.442	*0.099*	-

Table 8.10: The areality of single and double argument transitive indexation patterns

The results in Table 8.10 show that none of the geographic regions show areality for either single, double or no argument indexation on transitive verbs, nor do any of the macroregions. However, a few results are suggestive ($p < 0.15$): single argument indexation in Southern Amazonia, a lack of indexation in the Northern Andes, and double argument indexation in the Central Andes. The last observation is very close to meeting the threshold for areality. The Eastern macroregion shows a suggestive distribution for single argument indexation and the Western macroregion shows a suggestive distribution for a lack of indexation. It is possible that these suggestively-distributed features could indeed display areality with a modified language sample.

A final feature considered here is the presence of hierarchical marking in indexation. Both languages that have only a single hierarchical marker slot as well as those with a hierarchical marker slot and additional marker slots are considered as displaying hierarchical marking in this analysis, i.e. the computation is based on the parenthetical values in Table 8.9. The results show that languages of Northern Amazonia show areality for hierarchical marking, as do the languages of the Eastern South America macroregion in general.

Ditransitive marking: As discussed in section 3.3.3, the languages in the sample vary with regard to the indexation of non-subject arguments in ditransitive constructions. In his survey on the typological characteristics of the languages of the Andes, Adelaar (2008, 30) states that in ditransitive clauses "the encoded object often represents an indirect (human) object, rather than a direct object", or in other words, the languages of the Andes tend to show primary object alignment in indexation for ditransitive clauses. For the data used in this analysis, see Q2.10 in Appendix C. The results of an analysis for whether the R argument can be indexed on the verb in ditransitive clauses is given in Table 8.11. For this calculation, only languages that can also index P in transitive clauses are considered (cf. Q2.7 in Appendix C).

Region	R indexation
NAd	0.154
NAz	-
CAd	**0.041**
WAz	0.336
SAz	-
ChP	-
AMZ	-
AND	**0.002**
ESA	-
WSA	**0.001**

Table 8.11: The areality of R indexation in ditransitive constructions

As the results in Table 8.11 show, R indexation is indeed an areal feature of the Andes, in accordance with Adelaar (2008). R indexation is also an areal feature of Western South America.

Clusivity: As discussed in section 3.2.1 a grammatical distinction between 1st person plural pronominal forms that either include or exclude the addressee in a speech act (2nd person) is a commonly found in South American languages (cf. Filimonova, 2005; Crevels and Muysken, 2005). In fact, the earliest attested description of clusivity comes from a 16th century grammar of Quechua by Domingo de Santo Tomás in 1560. While the 1st person plural inclusive/exclusive distinction occurs in languages of various regions of South America, Adelaar (2008, 31) proposes that such a distinction in verbal argument marking is a characteristic feature of Andean languages. Table 8.12 presents the results of an analysis for whether a language distinguishes clusivity in person indexation. For the data used in this analysis, see Q2.13 in Appendix C. Note that data on whether a language displays a clusivity distinction in free pronouns has not been included in this study.

Surprisingly, the languages of the Andes show a lower distribution of clusivity in indexation than the rest of the continent. The results in Table 8.12

Region	Clusivity
NAd	-
NAz	0.483
CAd	0.322
WAz	-
SAz	*0.083*
ChP	0.337
AMZ	0.431
AND	-
ESA	**0.017**
WSA	-

Table 8.12: The areality of clusivity in indexation

show that the one region for which clusivity in indexation shows a statistically significant areal distribution is Eastern South America.

8.2.3   Geographic patterns in case marking

As discussed in Chapter 6, South American languages display a wide variety of different case marking patterns, with a number of languages showing common patterns such as accusative marking in Tsafiki (77) or ergative marking in Cavineña (74), as well as cross-linguistically rare patterns such as marked nominative case in Tehuelche (78) and a marked ergative-absolutive pattern in Northern Embera (80). The case marking of core arguments has not featured prominently in any of the previous work on the areal distribution of grammatical features in South America. Derbyshire (1987) does note that "ergatively organized systems, in whole or part, are quite common" in Amazonia, but highlights the fact that many languages tend to show splits in the alignment of case marking conditioned by clause type, tense-mood-aspect and pronoun vs. NP distinctions. Adelaar (2008, 29) also states that case marking on NP arguments is quite common in Andean languages, but notes that a number different alignment patterns in case marking can be observed beyond the often-discussed accusative marking in Quechuan, Aymaran and Barbacoan languages. This section explores the distribution of alignment in case marking for NP arguments and pronominal arguments in intransitive, transitive and ditransitive constructions.

Alignment in NPs: The first analysis presented in this section investigates the distribution of case marking in South American languages by examining the use of case marking on full noun phrase arguments in basic main clause constructions. A summary of the data in the sample is given in Table 8.13.

A map showing this distribution is given in Appendix A. Table 8.14 presents the results of an analysis for the areality of ergative and accusative alignment

Feature	NAd	NAz	CAd	WAz	SAz	ChP	SCo
Accusative	4	0	3	8	3	0	1
Ergative	3	1	0	6	2	1	0
Neutral	0	6	3	6	19	7	1

Table 8.13: The distribution of case alignment patterns in NP arguments

in case marking. The analysis also considers whether the presence or absence of any case marking on NP arguments is areal for the regions under consideration.

Region	Accusative	Ergative	Any case	No case
NAd	*0.067*	0.099	**0.002**	-
NAz	-	-	-	*0.108*
CAd	0.172	-	0.525	-
WAz	*0.081*	*0.089*	**0.005**	-
SAz	-	-	-	**0.006**
ChP	-	-	-	*0.065*
AMZ	-	0.628	-	0.215
AND	**0.028**	0.465	**0.020**	-
ESA	-	-	-	**0.001**
WSA	**0.001**	*0.075*	**0.001**	-

Table 8.14: The areality of case alignment in NP arguments

The results in Table 8.14 show that none of the non-aggregated geographic regions display an areal distribution for a particular alignment type in case marking. However, the presence of any case marking on NP arguments is an areal feature of the languages of the Northern Andes and Western Amazonia, while the absence of case marking (represented as neutral alignment in Table 8.13) is an areal feature for Southern Amazonia. At the macroregional level, accusative alignment in the case marking of NP arguments and the presence of case marking in general are areal features of both the Andes and Western South America. The absence of case marking on NP arguments is an areal feature of Eastern South America.

Alignment in pronouns: Let us now turn to the case alignment in free pronoun arguments in the languages of the sample. For an overview of the data used in this analysis, see Q5.7 in Appendix C. Table 8.15 presents the results for an analysis of the areality of case alignment in free pronoun arguments, examining the occurrence of accusative, ergative and neutral alignments. Tri-partite alignment in pronouns are not considered here since this pattern is too rarely attested in South American languages to yield informative results. It is worth noting that the characterization of pronominal alignment used in this study refers exclusively to alternations in the form of a pronoun and the ability

for an argument role to be expressed as a pronoun, but not the treatment of pronouns in other argument selectors (see section 6.2.3).

REGION	ACCUSATIVE	ERGATIVE	NEUTRAL
NAd	0.169	0.238	-
NAz	-	0.655	0.173
CAd	-	-	0.287
WAz	0.167	*0.088*	-
SAz	-	-	*0.058*
ChP	-	-	0.296
AMZ	-	0.339	-
AND	0.309	0.607	-
ESA	-	-	**0.002**
WSA	**0.035**	*0.114*	-

Table 8.15: The areality of case alignment in free pronoun arguments

As the results in Table 8.15 show, no particular case alignment shows areality for any of the non-aggregated regions. However, Western South America displays areality for accusative case alingment in pronouns, while Eastern South America displays areality for neutral case alignment in pronouns.

Case marking patterns in ditransitive constructions: As discussed in section 6.2.2, South American languages display a wide variety of different case marking patterns in ditransitive constructions. A summary of the alignments displayed in these patterns for the languages used in the sample is given in Table 8.16. This analysis is based on the case marking of NP arguments in transitive and ditransitive constructions and makes a distinction in neutral alignment between languages that case mark the object arguments P, T and R (marked neutral) and those that do not case mark any object arguments (unmarked neutral).

FEATURE	NAD	NAZ	CAD	WAZ	SAZ	CHP	SCO
Indirective	5	7	4	6	11	3	0
Secundative	2	0	0	2	6	0	0
Marked neutral	0	0	1	6	0	0	0
Unmarked neutral	0	0	0	5	7	5	2

Table 8.16: The distribution of ditransitive case alignment patterns

Table 8.16 shows that indirective alignment is most common in South America, being observed in all regions except the Southern Cone. The additional ditransitive marking patterns show distributions that are confined to certain regions. Table 8.17 presents the results of an analysis of the areality of different case alignment patterns in ditransitive constructions, making a further

distinction between marked and unmarked neutral alignments.

Region	Indirect	Secundative	Neutral$_{mark}$	Neutral$_{unmark}$
NAd	0.193	0.238	-	-
NAz	**0.005**	-	-	-
CAd	0.311	-	0.410	-
WAz	-	-	**0.001**	0.624
SAz	-	*0.054*	-	0.456
ChP	-	-	-	**0.026**
AMZ	-	0.339	0.305	-
AND	0.158	0.607	-	-
ESA	0.239	0.440	-	0.259
WSA	-	-	**0.003**	-

Table 8.17: The areality of ditransitive case alignment patterns

The results in Table 8.17 show that indirective alignment in ditransitive constructions is an areal feature of Northern Amazonia and an unmarked neutral pattern is an areal feature of the Chaco-Planalto region. The marked neutral pattern shows a strong areal distribution for the languages of Western Amazonia, and more generally, for the languages of Western South America. It is worth noting that the distribution of secundative alignment patterns in Southern Amazonia only barely falls short of the threshold to be considered an areal feature of this region. Additionally, in a supplemental analysis not shown above, the results show that a marked secondary object pattern where T is case marked and P and R are unmarked for case, displays areality for Southern Amazonia ($p = 0.031$), in accordance with the discussion in section 6.2.2.

8.2.4 Geographic patterns in valency changing strategies

Valency changing strategies have not played a prominent role in previous comparative studies on South American languages. This section explores the distribution of different valency changing strategies across the languages in the sample, with a focus on the distribution of verbally-marked derivations. In the present analysis, languages for which a specific construction type is not identified in the available descriptive materials are treated as not possessing that construction type.

Valency increasing strategies: As discussed in section 7.2, two main valency changing derivations are considered in this study: causative constructions and applicative constructions. The different semantic types of causative constructions are also included in Table 8.18 (see section 7.2.1).

The results from an analysis on the areality of these different valency increasing strategies are presented in Table 8.19. In the calculations for the are-

FEATURE	NAD	NAZ	CAD	WAZ	SAZ	CHP	SCO
Causative	6	7	6	16	21	5	1
Indirect causation	1	1	1	6	2	0	0
Sociative causation	2	1	1	3	8	0	0
Applicative	2	1	5	10	8	2	1

Table 8.18: The distribution of valency increasing strategies

ality of the indirect and sociative causative constructions, only languages that present a verbally marked causative construction are considered.

REGION	CAUSATIVE	CAUSATIVE$_{ind.}$	CAUSATIVE$_{soc.}$	APPLICATIVE
NAd	0.469	-	0.482	-
NAz	0.410	-	-	-
CAd	0.469	-	-	**0.017**
WAz	-	**0.023**	-	0.170
SAz	0.474	-	*0.090*	-
ChP	-	-	-	-
AMZ	0.450	0.321	0.300	-
AND	0.175	-	-	*0.102*
ESA	-	-	0.357	-
WSA	0.457	*0.063*	-	**0.031**

Table 8.19: The areality of different valency increasing strategies

The results in Table 8.19 show that the presence of an indirect causative construction is an areal feature of the languages of Western Amazonia. They also show that a verbally marked applicative construction is an areal feature of the Central Andes at the regional level and of Western South America at the macroregional level.

The occurrence of verbally marked causative constructions across the continent is so high that no region shows an areal distribution for a greater presence of this feature. For example, even in a region such as the Andes where all languages with relevant available data show a verbally marked causative construction, the distribution is not areal given the high concentration of these same constructions outside of the region.

Valency decreasing strategies: A number of the different valency decreasing strategies observed in the South American languages in the sample are presented in section 7.3. The distribution of these features across the regions used in this study is shown in Table 8.20. Following the definition used in section 7.3, the category of 'middle' represents a valency decreasing derivation that conflates reflexive and/or reciprocal marking with other valency decreasing operations such as antipassives (Q9.1.2) or anticausatives (Q12.1.1). Only

verbally marked reflexive and reciprocal constructions are included in the table below, while the passive constructions include both those marked directly on the verb as well as those expressed by an auxiliary (cf. Q8.1).

FEATURE	NAd	NAz	CAd	WAz	SAz	ChP	SCo
Passive	2	5	3	12	9	2	1
Antipassive	0	2	0	3	4	2	0
Reflexive	3	6	5	13	17	4	2
Reciprocal	4	6	6	16	15	5	2
Middle	0	2	1	4	10	3	0

Table 8.20: The distribution of valency decreasing strategies

Table 8.21 presents the results of an analysis on the areality of the different valency decreasing strategies for the regions used in this study based on the data presented in Table 8.20 .

REGION	PASSIVE	ANTIPASS.	REFLEXIVE	RECIPROCAL	MIDDLE
NAd	-	-	-	-	-
NAz	0.153	0.230	*0.138*	0.344	0.615
CAd	0.582	-	0.566	0.344	-
WAz	*0.113*	-	-	-	-
SAz	-	0.554	0.585	-	**0.048**
ChP	-	0.230	-	-	0.370
AMZ	*0.149*	0.321	0.358	-	0.166
AND	-	-	-	0.396	-
ESA	-	*0.100*	0.289	-	**0.018**
WSA	0.254	-	-	0.366	-

Table 8.21: The areality of different valency decreasing strategies

The results in Table 8.21 show that the presence of a valency decreasing derivation that conflates reflexive and/or reciprocal markers with anticausative and/or antipassive constructions, what is labeled here as a 'middle', is an areal feature of Southern Amazonia at the regional level and of Eastern South America at the macroregional level. This finding is somewhat congruent with the claim in Aikhenvald (2012, 229) that a 'typical feature' of Amazonian languages is that they tend to have only a single multifunctional valency decreasing device. However, in this study, the middle construction is found to be areal for Eastern South America rather than Amazonia proper.

8.2.5   Geographic patterns in constituent order

South America is well-known for its diversity of different main clause constituent orders. Within the sample, all logically possible constituent order types

are attested in transitive clauses, although two of the constituent orders, VPA and PAV, are only attested in a single language each. For each language in the sample, the basic constituent order has been identified based on the available descriptive materials that discuss the surface realization of NP arguments in main clauses. When multiple constituent orders are identified as well attested in the language, a basic order is only included if it occurs in the majority of clauses with two expressed NP arguments, otherwise the language is treated as having no dominant order (granted that quantitative data are available). No attempts at independent text counts have been carried out, and as such, these data should be considered only preliminary. A summary of the different constituent orders of core arguments with relation to the predicate for the different regions considered in this study is shown in Table 8.22.

FEATURE	NAD	NAZ	CAD	WAZ	SAZ	CHP	SCO
SV	6	4	6	13	15	6	1
VS	1	2	0	5	6	2	1
none	0	1	0	2	3	0	0
PV	6	6	6	12	12	4	1
VP	1	1	0	7	11	4	1
none	0	0	0	1	1	0	0
APV	6	2	6	9	11	4	1
AVP	1	1	0	5	5	3	1
VPA	0	0	0	0	1	0	0
VAP	0	0	0	2	4	1	0
PVA	0	3	0	2	0	0	0
PAV	0	1	0	0	0	0	0
none	0	0	0	2	3	0	0

Table 8.22: The distribution of main clause constituent orders

Table 8.23 presents the results of an analysis on the areality of the different constituent orders for the regions used in this study based on the data presented in Table 8.22. The column headings VX and PX refer to 'verb initial' and 'P initial', respectively.

The results in Table 8.23 show that the APV constituent order, more commonly known as SOV, is an areal feature of the Central Andes at the regional level and the Andes in general at a macroregional level. The PV constituent order type is also an areal feature of the Andes. Interestingly, P initial constituent orders are an areal feature of Northern Amazonian languages, which fits with the claim in Migliazza (1985, 20) that almost all languages of the 'Orinoco-Amazon' region, except for a few Arawakan languages, have object before verb constituent orders.

REGION	SV	VS	PV	VP	APV	AVP	VX	PX
NAd	0.295	-	0.196	-	*0.072*	-	-	-
NAz	-	0.510	0.196	-	-	-	-	**0.001**
CAd	*0.097*	-	*0.058*	-	**0.018**	-	-	-
WAz	-	0.513	-	0.551	-	0.445	-	0.519
SAz	-	0.495	-	*0.105*	-	-	*0.067*	-
ChP	0.521	0.592	-	0.258	-	0.231	0.619	-
AMZ	-	0.327	-	0.252	-	-	0.219	*0.097*
AND	0.115	-	**0.049**	-	**0.006**	-	-	-
ESA	-	0.384	-	*0.126*	-	0.486	0.169	0.391
WSA	0.245	-	*0.136*	-	*0.077*	-	-	-

Table 8.23: The areality of main clause constituent orders

8.3 Summary and discussion

This chapter has explored the distribution of over 40 different argument marking features for the sampled languages. Through the assessment of the areality of the these features, a number of them have been identified as showing an areal distribution for certain regions of the continent. Organized by region, these features are:

- *Northern Andes*: Both case and indexation together as argument marking strategies

- *Northern Amazonia*: Prefixes as verbal person markers, hierarchical marking patterns in indexation, indirective case alignment in ditransitive clauses and object initial constituent orders

- *Central Andes*: The ability to index R in ditransitive clauses, verbally marked applicative constructions and subject-object-verb constituent orders

- *Western Amazonia*: Both case and indexation together as argument marking strategies, marked neutral case marking patterns in ditransitive clauses and morphologically distinct indirect causative constructions

- *Southern Amazonia*: Indexation alone as an argument marking strategy, ergative alignment in indexation and middle type valency decreasing constructions

- *Chaco-Planalto*: Prefixes as verbal person markers and unmarked neutral case marking patterns in ditransitive clauses

These results show that the regions identified through major geographic features and culture areas indeed show an areal distribution for certain argument

marking features. However, a note of caution is warranted for the interpretation of the results from these regions results since the sample sizes for many regions are so small that a single change in coding can affect the outcome considerably. In many cases, the areal features of a particular region reflect larger areal patterns across macroregions.

Of the four macroregions considered in this study, a number of areal argument marking features have been identified for three of them. A summary of the areal features identified for Eastern South America is given in (127).

(127)   Areal argument marking features in Eastern South America
   a. Use of indexation alone as an argument marking strategy
   b. Prefixes as verbal person markers
   c. Hierarchical marking patterns in indexation
   d. Ergative alignment in verbal argument marker sets
   e. 1st person plural inclusive/exclusive distinction in verbal argument marker sets
   f. Neutral (unmarked) case alignment in NP and pronouns
   g. Middle type valency decreasing construction

A number of these features are not surprising given the ethnolinguistic makeup of this region. For example, Eastern South America contains all of the Cariban and Macro-Jêan languages, as well as the majority of Tupian languages used in this study. These languages typically use prefixes for argument indexation. Many Tupian, Cariban and Macro-Jêan languages additionally lack case marking of core arguments.

One cannot immediately discount the possibility that a relationship temporally deeper than the current application of the comparative method between the Cariban and Tupian families, and possibly even Macro-Jêan languages, may have played a role in the distribution of these features (cf. Rodrigues, 1985a). However, even after almost thirty years of investigation, solid evidence for such a relationship has yet to be produced and it is currently not accepted among most scholars of South American languages. The diffusion of these features through language contact is further supported by the fact that this region also contains a number of language isolates and smaller language families that also share many of these identified features, such as a lack of case marking in the Matacoan and Guaycuruan languages of the Chaco or hierarchical indexation in the Bolivian isolates Itonama and Movima and the Yanomaman language Yanam. Even Paresi, an Arawakan language spoken on the border between Southern Amazonia and the Planalto, has lost the object suffixes found all across the family, retaining only the prefixes that index intransitive and transitive subjects.

There are a number of social and demographic factors that could have contributed to the diffusion of these features, such as the expansion of Tupí-Guaranian peoples out of Amazonia and into the Chaco, Planalto and Atlantic

coast regions (Eriksen and Galucio, 2014); the interaction of different ethnic groups in the missions of lowland Bolivia (Cosme Bueno, 1770; Meireles, 1989; Crevels and van der Voort, 2008); and the chiefdoms along the headwaters of the southern Amazonian tributaries such as those of the Paresi and the multi-ethnic Xinguanos (Métraux, 1948; Heckenberger et al., 2008), as well as the large Tupinambá chiefdoms of the lower Amazon (Roosevelt, 1991). Unlike Western Amazonia, Southern and Northern Amazonia show evidence for long-term occupation by relatively large sedentary groups (McMichael et al., 2012), and it is known from historical and archaeological evidence that these groups often interacted in ways that would have fostered contact-induced language change, either through conquest and language shift, as in the case of the Tupian expansion, or through multilingualism resulting from more peaceful relations between different ethnic groups, as in the case of the upper Xingú and the Arawakan expansion south of the Amazon in general (Eriksen, 2011; Franchetto, 2011). Traditional models on the prehistoric distributions of peoples in the Amazon have often assumed that the floodplains of Amazonia have been continuously occupied for many thousands of years, so much so that population pressure along these rivers provided the impetus for the large-scale language migrations that occurred during the formative period (Lathrap, 1970). However, more recent archaeological research shows that even in the eastern portion of the continent, prehistoric populations were not evenly distributed along these floodplains, but rather, they were concentrated in specific regions such as in the lower basin of the Amazon River and the upper basin of the Madeira River (Neves, 2008, 363). The concentration of multiple ethnolinguistic groups into localized pockets that provided access to readily available resources would have served as ideal settings for an increase in the social interaction between these peoples, and by extension, the diffusion of linguistic traits.

The procedure used for the diagnosis of the areality of linguistic features also identifies a number of areal features for the Andes, as shown in (128).

(128) Areal argument marking features in the Andes

    a. Use of both case and indexation as argument marking strategies

    b. Suffixes as verbal person markers

    c. The R argument role can be indexed in ditransitive constructions

    d. Accusative case alignment for NP arguments

    e. Subject-object-verb constituent order

It is remarkable that all of these features are found in Quechuan and Aymaran languages even though only four of the fourteen Andean languages included in the study come from these families. This underscores the prominent position that speakers of these languages have played in Andean societies and the impact that their respective expansions must have had on the neighboring populations.

The Andes and Western South America share a number of features. With the exception of subject-object-verb constituent order, all areal features of the Andes also show an areal distribution for Western South America. The areal features identified for Western South America are shown in (129).

(129)   Areal argument marking features in Western South America
        a. Use of both case and indexation as argument marking strategies
        b. Suffixes as verbal person markers
        c. The R argument role can be indexed in ditransitive constructions
        d. Accusative case alignment for NP and pronoun arguments
        e. Marked neutral case marking patterns in ditransitive constructions
        f. Verbally marked applicative constructions

The overlap between the areal features of these two macroregions is striking. Western South America presents a number of areal traits in addition to those shared with the Andean languages, namely accusative case alignment in free pronouns, verbally marked applicative constructions and marked neutral case marking in ditransitive constructions. A number of these features are also found in the Andean languages, but since they are also found in many non-Andean languages of Western South America, particularly those of Western Amazonia, they could not be considered areal for the Andes macroregion proper.

Perhaps one of the most surprising results of the analysis is that not a single feature considered shows an areal distribution for Amazonia. This shows that Amazonia is not a good candidate for a linguistic area given the domains of investigation used in this study. Only two features show a suggestive distribution for Amazonia—the presence of a passive construction and object-initial constituent orders. Given that a multitude of areal features have been identified for the other macroregions that were posited as potential linguistic areas, the fact that no features were identified for Amazonia should be accounted for.

One possible explanation of these results lies in the methodology used in this study. Since the Amazonian languages comprise the vast majority of the languages of South America, the control group composed of non-Amazonian languages that was used to test against the Amazonian distributions is relatively small (n=23). It is possible that through the addition of further non-Amazonian languages, the Amazon as a linguistic area could become more salient in the data.[8] However, given the fact that the Chaco-Planalto languages share a number of features with the languages of Northern and Southern Amazonia, it is

---

[8]An additional possibility would be to include other languages from neighboring regions of the New World that lie outside of South America. The languages of Central America and the Caribbean would be good candidates for this since interaction between languages of these regions and those of South America is well attested in the historical record. A number of South American language families have distributions that extend into the regions, such as the Arawakan languages in the Caribbean and the Chibchan languages in Central America. Social interactions between speakers of different languages did not stop at the continental border, and the inclusion of such languages could yield potentially informative results. However, such

unclear whether this reformulation of the language sample would have any considerable effect on the results.

An additional explanation for these results emerges when one considers what is known from the prehistoric ethnolinguistic distribution of Western Amazonia and the sociocultural dynamics of the Andes and the societies of the Amazonian fringe. For example, McMichael et al. (2012) find that there was extensive pre-Colombian occupation of Eastern and Central Amazonia, but that the interfluvial zones of Western Amazonia were considerably less affected by human impact, as evidenced by a lesser degree of charcoal and phytoliths of agricultural species in the archaeological record. They conclude that the lowland portions of Western Amazonia were "predominantly occupied by relatively small and shifting human populations during the pre-Columbian era".[9] At the same time, the societies located on the upper stretches of the Western Amazonian tributaries lack a number of typical Amazonian cultural traits such as bitter manioc production, the vertical loom, ceremonial trumpets and a clan-based social organization (Steward, 1948), further suggesting that there was little interaction between the western and eastern portions of Amazonia. In their discussion on the ethnolinguistic composition of the Andes, van de Kerke and Muysken (2014) note that during intermediate periods between the major Quechuan and Aymaran expansions, the pre-Incan regional societies of the Andes often participated in the integration of peoples from lower altitudes into a "vertical exchange system". This suggests that there was a greater degree of interaction between the peoples of Western Amazonia with those from the Andes rather than with the societies of the more eastern portions of Amazonia. Such a scenario fits nicely with the observed distribution of areal linguistic traits in South America.

8.4   Conclusions

This chapter has presented the results from an analysis of the areality of specific argument marking features across different regions of South America. A number of features have been identified as areal for both the regions and larger aggregated macroregions, as shown in section 8.3. One of the most surprising results of this analysis is that no argument marking features have been identified as showing an areal distribution for Amazonia, while a number of areal features

---

a broad investigation of areal linguistic features across the New World as a whole is beyond the scope of the current study.

[9]While Arawakan peoples generally occupied the more fertile floodplains of the major rivers (Eriksen, 2011; Walker and Ribeiro, 2011), the interior portions of Western Amazonia were occupied by peoples with less reliance on agriculture. The Tucanoan peoples are thought to have expanded into the lowlands from a homeland in the hinterlands of the upper Apaporis and Caquetá rivers, near the Andean foothills (Chacon, 2014). It is possible that the different groups of people commonly referred to as Makú, which were once thought to comprise a single language family but recent historical work has shown that these are best considered multiple families (cf. Bolaños and Epps, 2009; Jolkesky, 2009), are descendants of these earlier populations, as suggested in Nimuendajú (1982, 169) and Dixon and Aikhenvald (2000b, 17).

have been identified for the Andes, Eastern South America and Western South America. This suggests that the sociohistorical dynamics of prehistoric Amazonia were not conducive to the widespread diffusion of any of the linguistic features that were explored here, and that a more salient characterization of the continent arises in the linguistic data through the division of the continent into Eastern and Western macroregions.

There have been a number of proposals that attempt to account for the large amount of shared linguistic material across different genealogical groupings. Most prominently, Dixon and Aikhenvald (1999) build off of the punctuated equilibrium model presented in Dixon (1997) to explain the high degree of shared linguistic features across Australian languages. They conceptualize Amazonia and surrounding regions as having been marked by long periods of stable interaction across ethnic groups (equilibrium), resulting in the diffusion of certain linguistic features and the convergence of languages towards a common structural profile. This equilibrium was then interrupted by major cataclysmic events that rapidly altered the ethnolinguistic composition of the area (punctuations), such as the development of agriculture that resulted in the expansions of the major lowland language families, and the European invasion that resulted in a drastic reduction in the indigenous populations of the continent. The underlying mechanics of such a model indeed have the potential to help motivate the distribution of shared linguistic features of the continent, as well as the difficulties in recognizing genealogical groupings at the same time depth that has been possible in the Old World.

A possible critique of such a model is that the notion of continent-wide periods of equilibrium is too simplistic to provide enough explanatory power to account for the regional skewing of particular traits. Additionally, archaeological evidence suggests that the adoption of agriculture was a slow and gradual process across the continent, spanning over many thousands of years and restricted to certain regions of the continent during different time periods (Pearsall, 2008). Correspondingly, the language expansions that presumably resulted from the increased populations produced by such an economic shift occurred much later. This suggests that different regions of the continent went through different periods of equilibrium and punctuation at different time periods, at least before the arrival of Europeans. Future work on the prehistory of South American languages could benefit greatly from a stronger emphasis on the regional integration of different ethnolinguistic groups rather than larger scale processes of diffusion.

Future research on the areal patterning of language structures in South America could greatly benefit from further comparative work that explores genealogical relationships that link together the currently accepted language families. At present, it is difficult to completely discount the possibility of "deep" genealogical relationships playing a role in the distribution of linguistic features across the continent. But since there are so many different proposals on these relationships, and that time and time again these proposals have not held up to close scrutiny by specialists, it is difficult to evaluate which proposed deep

relationships are worth considering and which are not, given our current state of knowledge. Only the currently accepted family groupings have been included in this study. The renewed interest in the description and documentation of South American languages has the potential to provide the empirical basis for these investigations, and the results here may need to be revised in light of any future findings.

The results from the study presented in this chapter help to illustrate the point that certain linguistic features are shared within particular regions, and that these regions are part of larger macroregions. While the development of any linguistic feature in any language has its own complex history determined by the characteristics that it inherited from its parent language and other language-internal changes, the fact that the features identified above show such a striking distribution in these regions across many genealogical groupings suggests that contact-induced change has played a role in these developments. The diffusion of these linguistic elements mirrors the complex regional networks of social interaction that bound together the different societies of the continent. As scholarly inquiry continues to refine our understanding of these social networks and other historical processes, so too can the community of linguists working in South America refine our understanding of the complex linguistic puzzle that the continent presents to the field.

Conclusions

This thesis began by introducing the complex puzzle that South American languages present to scholars concerned with the genesis of linguistic diversity in the New World and the extent of possible structural variation in human language. It was proposed that a detailed study of the different strategies used in these languages to encode the participants of an utterance can serve to highlight the range of linguistic variation of the continent, especially in light of the wealth of new information available on indigenous languages in recent years. The typological and areal patterns in this data can help us to better understand how this diversity arose, and can provide insights into the complex relationships between the populations that spoke these languages.

The research presented in this thesis is the first comprehensive investigation of the strategies used by South American languages to mark the arguments of basic main clause constructions. It employs a geographically and genealogically stratified sample of 74 South American languages that represent a total of 40 different language families. Recognizing the language-specific nature of different linguistic categories, the study makes use of semantically-defined comparative concepts that serve as an independent standard by which the diverse morphosyntactic structures of these different languages can be compared. The typological approach used to analyze the linguistic data rests on the identification of different semantic participant roles and the way that these participants are treated in the morphosyntax of a language, forming distinct argument roles. The primary focus of this study is the treatment of these argument roles by different argument markers such as indexation and case marking, and how these argument roles pattern together to form grammatical relations in these different grammatical domains. The languages used in this study are further divided

into different geographic regions that correspond to or include previously proposed cultural and linguistic areas, with each region delimited by geographic features such as mountain ranges, major river and ecological zones. This provides a heuristic for the subsequent analysis of the geographic distributions of the linguistic features explored throughout this study.

The research presented here is notable in that it draws from the wealth of recent data available for languages across the entire continent and encodes the major observable typological distinctions into a structural questionnaire. This has allowed for a more holistic view of the different argument marking strategies on the continent to emerge while also allowing for the application of a number of different quantitative techniques to analyze the development and distribution of these features over time and space. The last few pages of this thesis are now used to reflect on the results of this research by summarizing the different typological issues and geographic patterns identified in the data. The chapter concludes by outlining a few possibilities to further refine and expand this line of research in future work.

9.1 Typological issues

The approach used for the cross-linguistic comparison of argument marking features across the sample of South American languages has provided a number of valuable insights into the typological variation found in these languages. The incredible diversity of language structures observable across the continent poses a number of different challenges for language description, linguistic typology and historical linguistics.

As a whole, South American languages most frequently use indexation, or head-marking, as an argument marking strategy for core arguments in transitive and intransitive constructions (see Appendix A). For this reason, a large portion of this thesis is dedicated to mapping out the different patterns found in verbal argument marking, with a special emphasis on the realization of bound person forms. Chapter 3 outlines a number of important concepts related to the identification of the verbal argument marking pattern in a language. Different features such as the presence, position, alignment and fusion of arguments within different marker sets, and how these different marker sets combine on the predicate to express these arguments, are all included as important components of the marking pattern.

Chapter 4 on complex verbal marking patterns explores two areas of verbal argument marking that are often treated as distinct alignment types in the typological literature: hierarchical marking patterns and split intransitivity. The approach outlined in this chapter is tailored to the incipient state of description encountered for many of these languages and does not rely on an in-depth lexico-semantic analysis for each and every language. Rather, it is argued that through the control of a few semantic conditions, these grammatical patterns can be readily broken down into a number of distinct structural parameters,

thereby facilitating a more inclusive comparison between these and the other languages of the continent that show different indexation patterns.

Chapter 5 concludes the investigation into verbal argument marking by exploring the development of the person marking system across the large and structurally diverse Tupian language family of the lowlands. It is argued that the historical development of these different patterns can be informed by analyzing their synchronic distribution with regard to the current evolutionary hypotheses on the family. Through the application of a parsimony reconstruction model across the topology of two competing classificatory proposals, the results of this study confirm a number of previous hypotheses about the family, such as the presence of an absolutive indexation pattern in Proto-Tupí. By combining ancestral state reconstruction with what is known from traditional historical linguistic work, namely the identification of cognates within the marker sets, this case study also generates support for a number of additional hypotheses about the intermediate stages of development of these patterns.

Chapter 6 explores the different patterns found in the case marking of NP and pronominal arguments in the languages of the sample. A number of criteria are established to control for the pragmatic and referential effects that different arguments can have on the realization of case marking in order to ensure maximum comparability. The languages surveyed display a number of typologically rare case marking patterns, such as marked nominative and marked ergative-absolutive, in addition to the more common patterns such as marked accusative and marked ergative. Special attention is also given to the case marking of the non-subject arguments in ditransitive constructions, as well as the effects that animacy has on the realization of case markers in transitive constructions.

Chapter 7 examines the different strategies that languages employ to alter the argument marking pattern of a particular predicate through valency-changing derivations. The different alternations are defined according to the effects that they have on the valency of the predicate and the semantic situation type that they express. The most common result of the alternation of valency is the promotion or demotion of particular argument roles in terms of their ability to be selected by different argument markers such as case and indexation. Beyond examining many of the commonly discussed valency changing constructions such as causatives, applicatives, passives, antipassives, reflexives and reciprocals, the chapter also discusses the inherent difficulties in the multiple different uses of the term 'middle voice'. These different construction types are defined in such a way to ensure a high degree of comparability across languages, and a number of different parameters of structural variation are explored.

The typological issues discussed throughout the thesis aid in capturing the major structural parameters of variations across these different grammatical domains into a structural questionnaire. They further help to clarify the choices made when assigning a particular feature value to a language. Since this research has aimed to be as inclusive and comprehensive as possible, it has at times been insufficient to uncritically adopt the parameters of evaluation from

other large-scale typological studies, and at times, a further refinement in the conceptual distinctions have been deemed necessary. For example, if one were to adopt the traditional terminology for discussing case marking systems, the typologically rare marked nominative pattern in Tehuelche would be treated the same as the cross-linguistically common marked accusative pattern in languages such as Hup and Tsafiki; both systems would be considered to show 'accusative case marking'. However, such a distinction conflates the alignment of the markers with the arguments that are being marked, and through this simplification, potentially informative facts about the structure of these languages are lost. For this reason, the term 'marking pattern' was adopted throughout the thesis as a way to characterize the different facets that are involved in the treatment of a set of arguments by an argument marker while highlighting the diversity of these different patterns beyond simple alignment.

The questions used in the structural questionnaire are given in Appendix B, with the codings of these features for the languages of the sample given in Appendix C. These coded features serve as the input for an analysis of the geographic distribution of argument marking features, as well as the empirical basis for any other claims made in the study. It is also hoped that by presenting the data in a transparent and explicit manner, this dataset can be adopted, adapted and otherwise utilized by other researcher working in the area.

9.2 Geographic patterns

A number of different studies have been produced over the last few decades that attempt to identify a number of linguistic features that the authors see as characteristic of specific regions of the South America. The spread of these features has often been attributed to diffusion through contact-induced language change, and multiple proposals have be made to identify specific linguistic areas of different size across the continent. The large collection of discrete data collected on argument marking features for the languages considered in this study serve as the ideal input to test some of these previous claims, as well as to explore the geographic distribution of these features more generally.

Building off of previous quantitative approaches to areal typology, the study presented in Chapter 8 outlines a procedure to diagnose the areality of a linguistic feature given a pre-defined region and the languages that it contains. Over 40 different linguistic features were tested for areality at two levels of geographic analysis. The first level corresponds to the major regions of the continent identified in Chapter 1. The second level of analysis explored larger macroregions that were formed through the aggregation of different configurations of the aforementioned major regions. The macroregions used—Amazonia, the Andes, Eastern South America and Western South America—are in accordance with previous research that suggests that these different macroregions are potential candidates to be identified as linguistic areas based on the sharing of multiple linguistic traits and due to providing the necessary sociocultural

environments for the spread of different linguistic features over time. The first step in the test for areality is the identification of the proportional representation of a particular argument marking feature in each region or macroregion. This distribution is then contrasted with the distribution of that feature in all of the South American languages outside of the region under investigation. If the region under consideration shows a higher proportional presence of this feature in question, the distributions are tested to see whether they are significantly different from one another, i.e. that the differences within these distributions are not merely a result of chance similarities. Such an analysis thus evaluates how good of a variable a particular region is for predicting the typological distribution of a certain feature. If a geographic region is a good predictor variable for the distribution of the feature, and the occurrence of this feature cannot be attributed to other possible explanations, then the diffusion of this linguistic feature through contact-induced language change remains a possible explanation for this distribution. The possibility of these distributions being attributed to shared inheritance from a known common ancestor is controlled for through the use of a genealogically-stratified language sample. The possibility that these distributions are a result of pure chance and independent language-internal changes is controlled for through the use of statistical inference techniques whereby only features that show over a 95% probability that they cannot be attributed to chance are accepted.

This analysis produced a number of interesting results at both the regional and macroregional levels. These results are summarized in section 8.3 of the previous chapter and will not be repeated here. The most striking outcome of this analysis is that, while a considerable number of features were identified as showing an areal distribution for the Andes, Western South America and Eastern South America, not a single feature showed an areal distribution for Amazonia as a macroregion. This suggest that Amazonia is not a good candidate for a linguistic area based on the features examined in this study.

A further intriguing result is that many of the areal features identified for the Andes are also shared by the languages of Western South America in general. This suggests that many of the sociohistorical processes that helped spread these linguistic features in the Andes also had an influence on the development of these features in Western Amazonia, especially among the languages located along the upper stretches of the Amazonian tributaries in the Andean foothills.

These results are not entirely surprising when one takes into account what is known about the distributions of the indigenous people of South America through other disciplines such as archaeology and ethnohistory. For example, large portions of the Amazon Basin remained relatively unpopulated until after the formative period language expansions. The pre-formative indigenous populations of the Amazon were thus concentrated in specific geographic regions that would have provided the ideal contexts from the diffusion of linguistic features and the convergence of the languages towards a common structural profile. The formative era language expansions would then have distributed these features more widely, resulting in a reformulation of the social networks

that foster multiethnic contact, leading to the distribution of linguistic features that are observable today.

9.3 Future research

This study is a preliminary attempt to gain a better understanding of the typological distribution of argument marking features in South America by using a sample of 74 indigenous languages from the continent. The approach used in this thesis has been shown to present intriguing results, and further insights can surely be gained through a refining of the comparative concepts and analytical techniques used.

An obvious next step is to increase the sample density used in future studies. A large number of the smaller language families in South America have not been represented in the current study, such as members of the Zaparoan, Peba-Yaguan, Guahiboan, Zamucoan, Mascoyan families as well as many linguistic isolates. A number of branches within the larger linguistic families have also not been sampled, such as members of the Southern Jê and certain other branches of the Macro-Jêan family, the Venezuela Cariban languages and the Campa Arawakan languages. As new descriptive materials are produced and the existing materials further examined, the inclusion of additional languages has the potential to greatly expand our perspective. Comparable data from adjacent areas outside of South America proper could also help to further accentuate the geographic patterns of the continent.

A better understanding of the genealogical diversity of indigenous languages and the deep time relationships between the established families will help to inform future studies. A number of these proposals currently remain suggestive yet still unproven, while other relations may not yet have been recognized. One of the major factors that limits our ability to attribute the distribution of certain linguistic features to language contact is the temporal depth of our current linguistic classifications and the possibility that some of these areal features were in fact inherited from deeper phylogenetic groupings of languages than those that are currently accepted by historical linguists. Further comparative work on the classification and reconstruction of language families should continue using both traditional and computational techniques.

As language typology further integrates quantitative techniques from other disciplines and develops new practices within the field itself, there is a great potential to increase the amount of information that can be gathered from the existing dataset. Additional techniques for sampling procedures, randomization tests, isogloss identification and so forth can play a role in further expanding the tools at our disposal.

The results that have been presented in this thesis suggest that the domain of inquiry should be expanded to additional grammatical systems and that the existing comparative data should be analyzed with similar quantitative techniques. It is possible, and even likely, that an analysis of additional data will lead

to the identification of further areal patterns in South America. The East/West distinction that is so salient in the argument marking data should be further tested with other linguistic features. As our typological understanding of South American languages increases, it will be possible to continue working towards a broader synthesis with the other disciplines concerned with the prehistory of the New World, such as archaeology, genetics, geography, ethnography and demography. Such a multi-pronged approach that integrates language, culture, population dynamics and ecology holds the greatest potential for unlocking new insights into the development of the indigenous societies of the continent, and with it, moving us gradually closer to a better understanding the South American puzzle.

Maps

Figure A.1: Major geographic regions
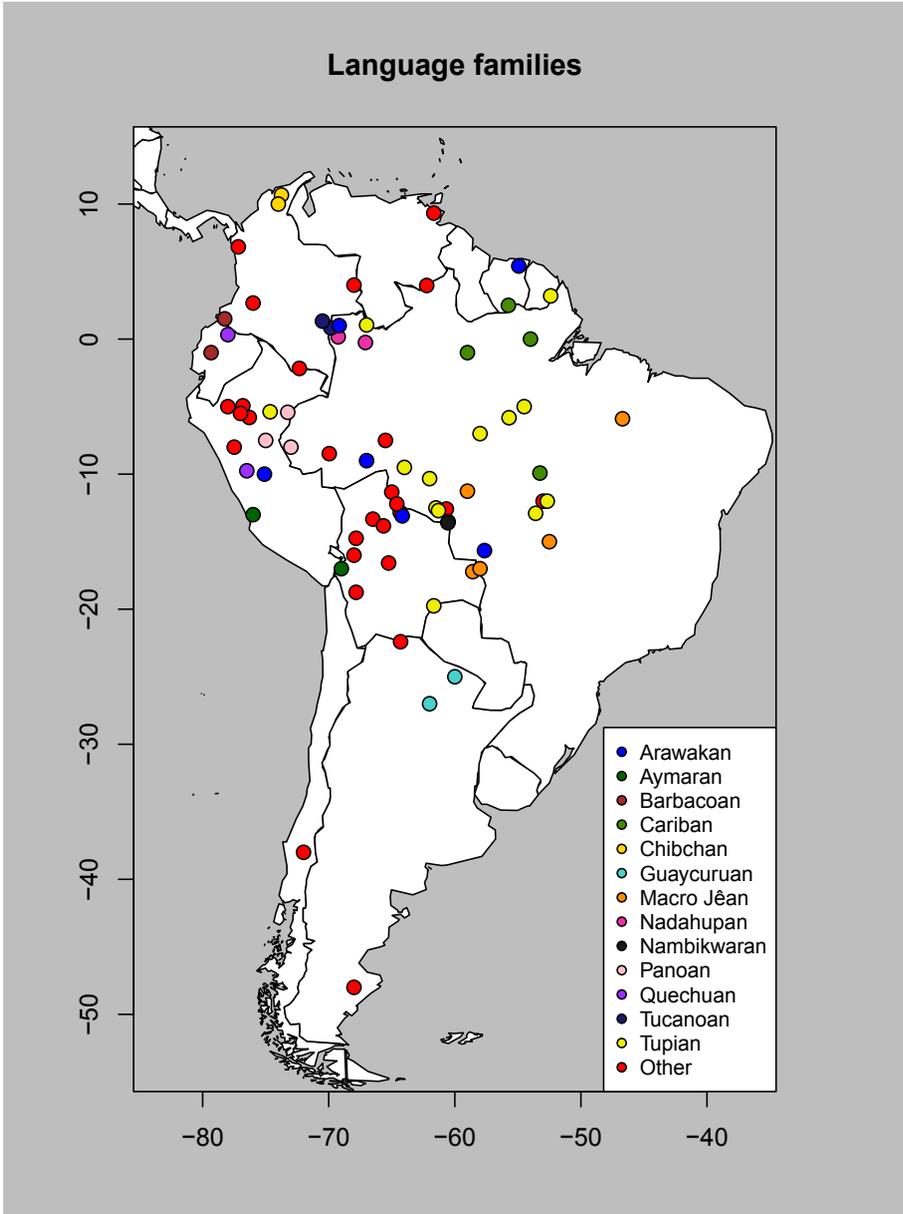
Language families

Figure A.2: Language families in the sample
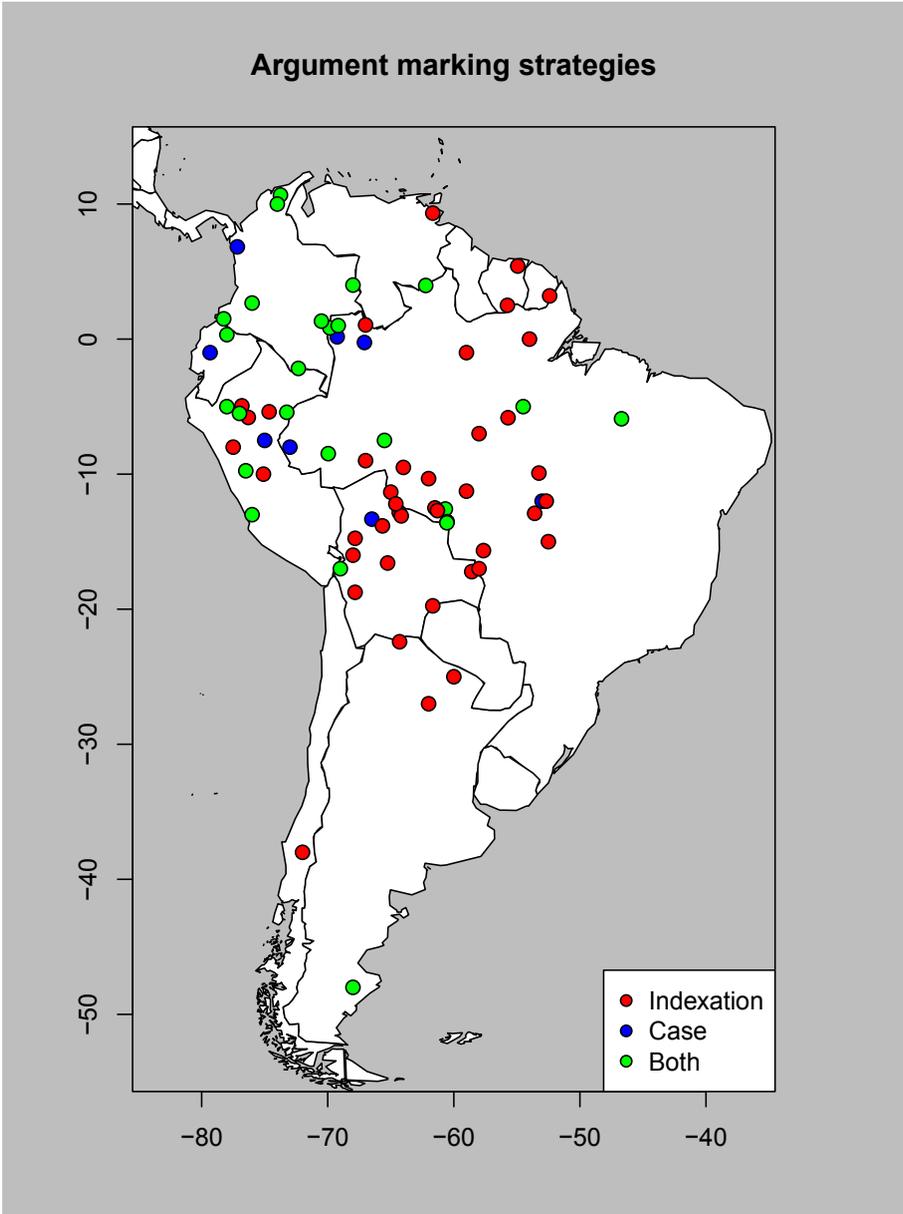
Figure A.3: Distribution of argument marking strategies
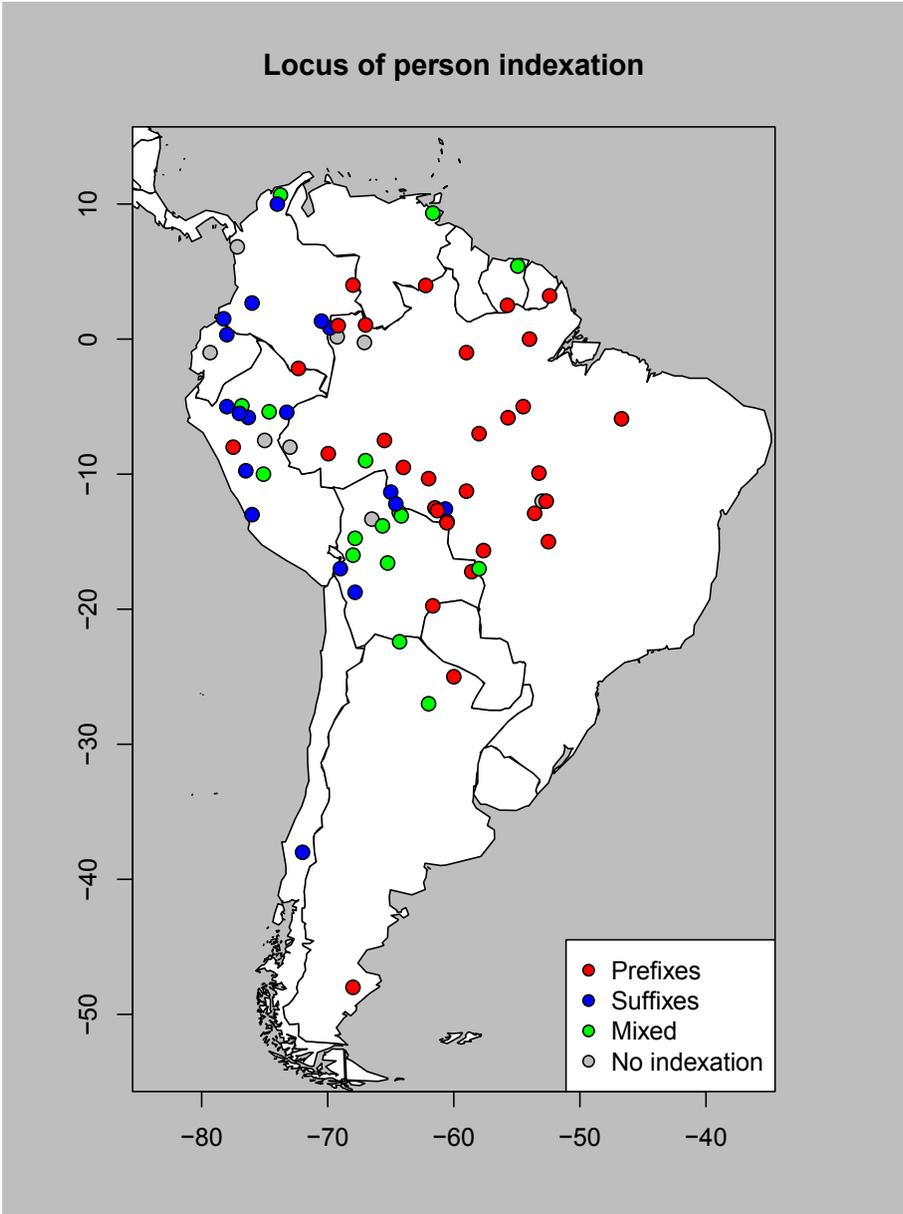
Figure A.4: Distribution of the locus of person indexation
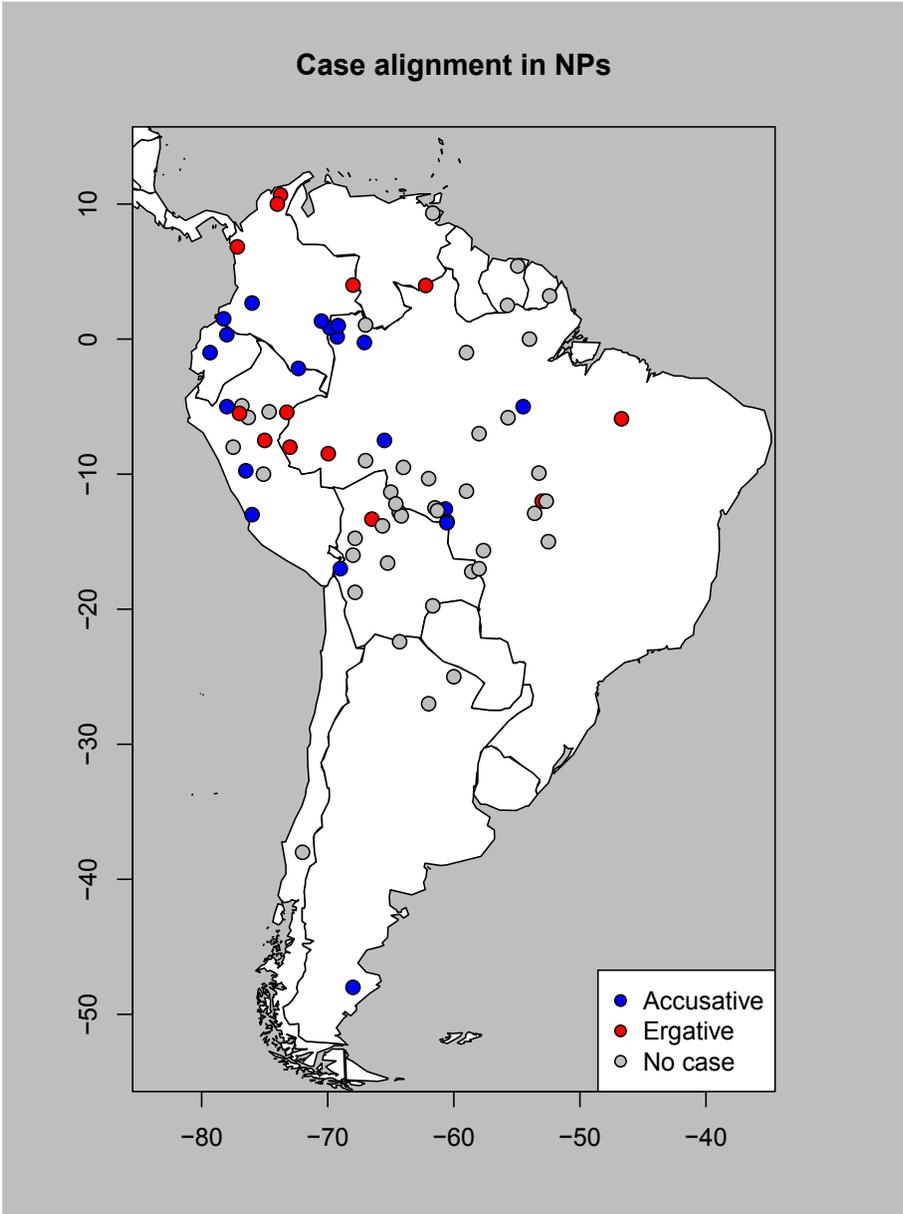
Figure A.5: Distribution of case alignment in NP arguments

Structural questionnaire

Table B.1: Structural questionnaire

DOMAIN	LABEL	QUESTION	VALUE
Constituent order	Q1.1	The basic constituent order in intransitive clauses is:	a=SV, b=VS, c=neither
	Q1.2	The basic constituent order in intransitive clauses is:	a=APV, b=AVP, c=VPA, d=VAP, e=PVA, f=PAV, g=none
	Q1.3	The basic constituent order for P and V in transitive clauses is:	a= PV, b=VP, c=neither
	Q1.4	The basic constituent order for R in ditransitive clauses is:	a=between A&V, b=between T&V, c=between A&T, d=clause final, e=clause initial, f=free
Verbal marking	Q2.1	Person can be marked for S:	a=prefix, b=suffix, c=variable, d=not marked
	Q2.1.1	Person marking for S when conominal is expressed in the clause is:	a=obligatory, b=variable, c=not marked
	Q2.2	Number can be marked for S:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked
	Q2.3	Gender can be marked for S:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked, e=no gender
	Q2.4	Person can be marked for A:	a=prefix, b=suffix, c=variable, d=not marked
	Q2.4.1	Person marking for S when conominal is expressed in the clause is:	a=obligatory, b=variable, c=not marked
	Q2.5	Number can be marked for A:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked

Structural questionnaire (*Continued on next page*)

Table B.1 – *Continued from previous page*

DOMAIN	LABEL	QUESTION	VALUE
	Q2.6	Gender can be marked for A:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked, e=no gender
	Q2.7	Person can be marked for P:	a=prefix, b=suffix, c=variable, d=none
	Q2.7.1	Person marking for P when conominal is expressed in the clause is:	a=obligatory, b=variable, c=not marked
	Q2.8	Number can be marked for P:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked
	Q2.9	Gender can be marked for P:	a=independent prefix, b=independent suffix, c=fused w/person, d=not marked, e=no gender
	Q2.10	Ditransitive verbs can mark R:	1, 0
	Q2.10.1	Ditransitive verbs can mark both R and T simultaneously:	1, 0
	Q2.11	In transitive clauses that can mark A and P, marking is:	a=A before P, b=P before A, c=hierarchical, d=portmanteau
	Q2.12	Gender can only be marked in the singular:	1, 0
	Q2.13	First person inclusive is a distinct morphological category in vebal argument markers:	1, 0
	Q2.14	Verbal argument markers are fused with TMA categories	1, 0
	Q2.15	Verbal person markers show the following alignment pattern:	a=accusative, b=ergative, c=tripartite, d=neutral
Inversion	Q3.1	Language has a scenario-based split in verbal argument marking:	1, 0
	Q3.1.1	A distinct morpheme marks the inverse construction on the verb:	1, 0
	Q3.1.2	Referential arguments in inverse constructions are treated with a marker set distinct from that used to index the referential argument in direct constructions:	1, 0

Table B.1 – *Continued from previous page*

DOMAIN	LABEL	QUESTION	VALUE
	Q3.1.3	Inverse constructions can be used to distinguish between third person participants:	1, 0
Split intransitivity	Q4.1	Language has two classes of intransitive verbs with separate formal marking of their arguments:	1, 0
	Q4.1.1	S of event verbs without control (e.g. hiccup, die, slip) are treated as:	a=major class, b=minor class, c=variable
	Q4.1.2	S of stative verbs with control (e.g. reside, be.patient, be.prudent) are treated as:	a=major class, b=minor class, c=variable
	Q4.1.3	Intransitive verbs can alternate between classes depending on semantic considerations:	1, 0
	Q4.1.4	Minor class of intransitive verbs show alignment with:	a=A, b=P, c=neither
Case marking	Q5.1	S is case marked:	1, 0
	Q5.2	A is case marked:	1, 0
	Q5.3	P is case marked:	1, 0
	Q5.4	R is case marked:	1, 0
	Q5.5	Case marking of inanimate arguments is:	a=obligatory, b=variable, c=not possible
	Q5.6	Nouns show the following alignment pattern:	a=accusative, b=ergative, c=neutral
	Q5.7	Pronouns show the following alignment pattern:	a=accusative, b=ergative, c=neutral, d=tripartite
	Q5.8	Ditransitive constructions show the following ditransitive alignment pattern:	a=indirective, b=secundative, c=neutral
	Q5.9	Obliques are case marked:	a=prefix/preposition, b=suffix/postposition, c=neither
Causatives	Q6.1	Causative constructions are marked by verbal morphology:	1, 0
	Q6.1.1	Causative constructions can only be derived from intransitive verbs:	1, 0
	Q6.1.2	The causee in transitive-derived causative constructions is treated as:	a=Oblique, b=R, c=T, d=A
	Q6.1.3	Direct and indirect causation are formally distinguished:	1, 0
	Q6.1.4	Direct and sociative causation are formally distinguished:	1, 0

Structural questionnaire (*Continued on next page*)

Table B.1 – Continued from previous page

DOMAIN	LABEL	QUESTION	VALUE
Applicatives	Q7.1	Applicative constructions are marked by verbal morphology:	1, 0
	Q7.1.1	Applicative constructions only select benefactive/malefactive participants:	1, 0
	Q7.1.2	Applicative constructions can only select animate participants:	1, 0
	Q7.1.3	Applicative constructions can be derived from transitive verbs:	1,0
	Q7.1.4	The promoted argument can be expressed as an oblique in corresponding non-derived clauses:	1, 0
Passives	Q8.1	Passive constructions are marked by:	a=verbal morphology, b=auxiliary, c=both, d=none
	Q8.1.1	The demoted A of a passive construction can be expressed in the clause:	1, 0
Antipassives	Q9.1	Antipassive constructions are marked by verbal morphology:	1, 0
	Q9.1.1	The demoted P argument can be expressed as an oblique:	1, 0
	Q9.1.2	The marking of antipassive constructions is conflated with the marking of reciprocal and/or reflexive constructions:	1, 0
Reflexives	Q10.1	Reflexive constructions are marked with verbal morphology	1, 0
	Q10.2	Language has independent reflexive pronouns:	obligatory, optional, none
Reciprocals	Q11.1	Reciprocal constructions are marked with verbal morphology:	1, 0
	Q11.2	Reciprocal constructions are marked with a specialized particle:	obligatory, optional, none
	Q11.3	Language has independent reciprocal pronouns:	1, 0
	Q11.4	Marking of reciprocal and reflexive constructions is morphologically conflated:	1, 0
Anticausatives	Q12.1	Anticausative constructions are marked with verbal morphology:	1, 0
	Q12.1.1	The marking of anticausative constructions is conflated with the marking of reciprocal and/or reflexive constructions:	1, 0

Questionnaire codings

Table C.1: Questionnaire codings: Questions 1.1-2.15

LANGUAGE	1.1	1.2	1.3	1.4	2.1	2.1.1	2.2	2.3	2.4	2.4.1	2.5	2.6	2.7	2.7.1	2.8	2.9	2.10	2.10.1	2.11	2.12	2.13	2.14	2.15
Apurinã	a	b	b	e	a	b	c	c	a	b	c	c	b	b	c	c	1	0	a	0	0	0	a
Baure	b	d	b	b	a	a	c	c	a	a	c	c	b	c	c	c	1	1	a	1	0	0	a
Lokono	a	b	b	d	a	c	c	c	a	c	c	c	b	c	c	c	0	0	a	0	0	0	a
Paresi	a	a	a	d	a	a	c	c	a	a	c	c	d	-	d	e	0	0	-	-	0	0	a
Tariana	c	g	a	e	a	a	c	c	a	a	c	c	d	-	d	e	0	-	-	1	0	0	a
Yanesha'	b	d	b	b	d	a	c	e	a	a	c	c	b	b	c	e	1	1	a	1	0	0	a
Jarawara	a	a	a	e	a	a	c	b	a	a	c	e	a	a	c	b	1	0	b	1	1	1	a
Aymara	a	a	a	a	b	a	b	e	b	a	b	e	b	a	b	e	1	0	b	-	1	1	a
Jaqaru	a	a	a	e	b	a	d	e	b	a	?	e	b	a	d	e	1	0	d	-	1	1	a
Awa Pit	a	a	a	c	b	a	b	e	b	a	b	e	b	?	b	e	1	0	c	-	0	0	a
Tsafiki	a	a	a	e	d	-	?	e	d	-	?	e	d	-	d	e	0	-	-	-	0	0	-
Apalaí	c	e	a	?	a	c	c	e	a	c	c	e	a	c	c	e	0	0	c	-	1	0	a
Hixkaryana	b	e	a	d	a	a	c	e	a	a	c	e	a	a	c	e	0	0	c	-	1	0	c
Ikpeng	b	d	b	d	a	a	b	e	a	a	b	a	b	a	b	e	0	0	c	-	0	0	b
Tiriyó	b	e	a	b	a	a	c	e	a	a	c	e	b	?	c	e	0	0	c	-	1	0	b
Chayahuita	a	a	a	e	b	a	c	e	b	a	c	e	b	a	c	e	1	0	a	-	1	1	a
Wari'	b	c	b	b	b	a	c	c	b	a	c	c	b	b	c	c	1	0	a	0	1	1	a
Chimila	b	b	b	d	b	?	c	e	b	?	c	?	b	?	c	b	1	0	c	-	0	0	a
Ika	a	a	a	c	c	b	c	e	c	b	c	e	a	b	c	e	1	0	a	-	1	0	a
Embera	a	a	a	c	d	b	b	e	d	b	b	d	d	-	d	e	0	-	-	-	-	0	-
Tehuelche	a	a	a	e	a	c	a	d	a	c	c	e	a	b	c	c	1	1	a	?	0	0	d
Mocoví	b	b	a	?	c	a	b	e	c	a	b	e	c	b	c	c	0	0	c	-	0	0	a
Pilagá	a	b	b	?	c	a	c	e	a	a	c	e	a	a	b	e	1	0	b	-	0	0	a
Cholón	a	a	a	?	a	a	c	e	a	?	c	e	a	a	c	e	1	?	a	-	0	0	d
Itonama	b	d	b	d	a	a	c	e	b	a	c	?	b	a	c	e	1	0	a	-	1	0	a
Kanoê	a	a	a	b	a	a	d	e	b	a	d	e	b	b	d	e	1	0	a	-	0	0	d
Kwaza	a	b	b	e	b	a	c	e	b	a	b	e	b	b	c	c	1	0	b	-	1	0	a
Leko	a	a	a	e	b	a	c	c	b	a	c	c	a	a	c	e	1	0	a	-	0	?	a
Mapudungun	b	b	b	d	b	b	b	e	b	b	b	c	b	b	b	c	1	0	c	-	0	1	d
Movima	b	d	b	c	c	b	c	c	c	c	c	c	c	c	b	c	1	0	c	-	1	0	d

Questionnaire codings (*Continued on next page*)

Table Questionnaire codings: Questions 1.1-2.15 (*Continued from previous page*)

LANGUAGE	1.1	1.2	1.3	1.4	2.1	2.1.1	2.2	2.3	2.4	2.4.1	2.5	2.6	2.7	2.7.1	2.8	2.9	2.10	2.10.1	2.11	2.12	2.13	2.14	2.15
Munichi	b	d	b	?	b	b	c	e	b	b	c	e	b	c	c	e	1	?	a	-	0	0	d
Nasa Yuwe	a	a	a	c	b	a	c	e	b	a	c	e	d	-	d	e	0	-	-	-	0	1	a
Puinave	c	g	c	d	a	b	c	e	a	b	c	e	a	b	c	e	0	0	b	-	0	0	d
Trumai	a	a	a	d	d	-	d	e	d	-	d	e	d	-	d	e	0	-	-	-	-	0	-
Urarina	b	e	a	f	b	a	b	e	b	a	b	e	a	c	b	e	0	0	b	-	1	0	a
Warao	a	f	a	f	b	c	e	e	b	c	b	e	a	c	d	e	0	0	b	-	0	0	a
Yurakaré	b	g	b	e	b	c	e	d	b	a	c	e	a	a	c	e	1	1	b	1	0	0	a
Aguaruna	a	a	a	c	b	a	c	e	b	a	c	d	b	a	c	c	1	1	b	-	0	0	b
Katukina	b	b	b	?	d	a	d	e	b	a	c	e	d	-	d	e	0	0	b	-	0	0	b
Bororo	a	a	a	d	a	a	c	e	a	c	c	e	a	a	c	e	0	0	-	-	1	0	b
Guató	b	d	b	d	c	a	c	e	c	a	?	?	c	?	c	e	?	?	a	-	1	0	?
Rikbaktsa	a	a	a	b	a	a	b	d	a	a	b	d	a	a	c	d	0	0	a	-	0	0	a
Timbira	a	a	a	c	a	c	a	e	d	-	d	e	a	c	a	e	0	0	-	-	1	0	b
Xavánte	a	a	a	?	a	b	b	e	d	-	b	e	a	b	a	e	0	0	a	-	0	0	b
Wichí	a	b	b	d	a	a	b	e	a	a	b	e	b	c	b	e	1	0	a	-	1	0	c
Moseten	a	b	b	b	d	-	d	b	b	a	b	c	b	a	c	c	1	0	d	0	1	0	?
Dâw	a	b	a	d	d	-	d	e	d	-	c	e	d	-	c	e	0	-	-	-	-	0	-
Hup	a	a	a	e	d	-	d	e	d	-	d	e	d	-	d	e	0	-	-	-	-	0	-
Latundê	a	a	a	b	b	b	c	e	b	b	c	e	b	b	c	e	1	1	b	-	0	?	a
Sabanê	a	a	a	c	d	-	d	e	d	-	d	e	a	c	c	e	1	0	-	-	0	?	a
Matses	a	a	b	b	b	a	d	e	b	-	b	e	d	-	d	e	1	-	-	-	0	0	a
Shipibo	a	a	a	c	d	a	d	e	d	a	d	e	d	-	d	e	1	-	-	-	0	0	-
Yaminahua	a	a	a	b	d	-	d	e	d	-	d	e	d	-	d	e	0	-	-	-	-	0	-
Huallaga Q.	a	a	a	a	b	a	c	e	b	a	c	e	b	a	c	e	1	0	b	-	1	1	a
Imbabura Q.	a	a	a	c	b	a	b	e	b	a	b	e	d	-	d	e	0	-	b	-	0	1	a
Cavineña	c	g	c	e	d	-	d	e	d	-	d	e	d	-	d	e	0	-	-	-	-	0	-
Cubeo	b	e	c	?	b	?	c	c	b	?	c	c	d	-	d	e	0	-	-	0	1	0	a
Desano	a	a	a	a	b	a	c	c	a	a	c	c	d	-	d	d	0	-	-	1	0	1	a
Awetí	a	b	b	d	a	a	c	e	a	a	c	e	a	a	c	e	0	0	c	-	1	0	c
Cocama	a	b	b	d	a	c	c	e	a	c	c	e	b	c	c	e	0	0	a	-	1	0	a
Emerillon	a	a	a	?	a	a	c	e	a	a	c	e	a	a	c	e	0	0	c	-	1	0	a

Table Questionnaire codings: Questions 1.1-2.15 *(Continued from previous page)*

LANGUAGE	1.1	1.2	1.3	1.4	2.1	2.1.1	2.2	2.3	2.4	2.4.1	2.5	2.6	2.7	2.7.1	2.8	2.9	2.10	2.10.1	2.11	2.12	2.13	2.14	2.15
Juruna	c	a	a	c	d	-	d	e	d	-	d	e	a	?	c	e	0	-	-	-	1	0	a
Kamayurá	a	a	a	d	a	a	c	e	a	a	c	e	a	b	c	e	0	0	c	-	1	0	a
Karitiana	a	b	b	e	a	b	c	e	d	-	d	e	a	a	c	e	1	0	-	-	1	0	b
Karo	a	a	a	d	a	c	c	c	d	-	d	e	a	c	c	e	0	?	-	-	1	0	b
Mekens	a	a	a	d	a	a	c	c	d	-	d	e	a	c	c	e	0	0	-	-	1	0	b
Munduruku	a	a	a	d	a	a	c	e	a	a	c	e	a	a	c	e	0	0	c	-	1	0	a
Nheengatu	a	b	b	b	a	a	c	e	a	a	c	e	d	-	-	e	0	-	-	-	0	0	a
Sateré	a	b	b	d	a	a	c	e	a	a	c	e	a	a	c	e	0	-	c	-	1	0	a
Tapiete	a	a	a	d	a	?	c	e	d	?	c	e	a	?	c	e	0	0	c	-	1	0	b
Wayoró	c	g	a	e	a	a	c	c	b	-	d	e	d	c	c	e	0	0	c	-	1	0	a
Chipaya	a	a	a	c	b	b	b	?	b	b	b	?	d	-	-	e	0	-	-	?	1	1	a
Miraña	a	a	a	f	a	b	b	b	a	b	b	b	d	-	d	e	0	0	-	0	0	0	a
Yanam	a	a	a	d	a	a	c	e	a	a	c	e	a	a	c	e	0	0	?	-	0	0	d

Table C.2: Questionnaire codings: Questions 3.1-6.1.4

LANGUAGE	3.1	3.1.1	3.1.2	3.1.3	4.1	4.1.1	4.1.2	4.1.3	4.1.4	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.1	6.1.1	6.1.2	6.1.3	6.1.4
Apurinã	0	-	-	-	1	0	0	?	b	0	0	0	1	-	c	c	a	b	1	0	c	0	0
Baure	0	-	-	-	1	0	0	?	b	0	0	0	0	-	c	c	c	b	1	0	b	0	0
Lokono	0	-	-	-	1	0	1	0	b	0	0	0	1	-	c	c	a	b	1	0	a	0	0
Paresi	0	-	-	-	1	1	?	0	c	0	0	0	1	-	c	c	a	b	1	0	c	0	0
Tariana	0	-	-	-	1	0	1	1	b	0	0	1	1	b	a	a	c	b	1	0	c	1	0
Yanesha'	0	-	-	-	1	?	?	?	b	0	0	0	0	-	c	c	c	b	?	?	?	?	?
Jarawara	0	-	-	-	0	-	-	-	-	0	0	1	1	a	a	a	c	b	1	0	c	1	0
Aymara	0	-	-	-	0	-	-	-	-	0	0	1	1	a	a	a	a	b	1	0	a	0	0
Jaqaru	0	0	1	0	0	-	-	-	-	0	0	1	?	?	a	c	?	b	1	0	b	0	0
Awa Pit	1	-	-	-	1	1	?	?	b	0	0	1	1	c	a	a	b	b	1	1	b	1	1
Tsafiki	0	0	1	0	0	-	-	-	-	0	0	0	1	b	c	c	a	b	1	0	-	1	1
Apalaí	1	0	1	0	0	-	-	-	-	0	0	0	1	-	c	c	a	b	1	0	b	?	?
Hixkaryana	1	0	1	0	0	-	-	-	-	0	0	0	1	-	c	c	a	c	1	0	b	0	0
Ikpeng	1	0	1	0	1	1	1	0	a	0	0	0	1	-	c	c	a	b	1	0	b	0	0
Tiriyó	1	0	1	0	1	0	0	?	a	0	0	0	1	-	c	c	a	b	1	0	b	0	0
Chayahuita	0	-	-	-	0	-	-	-	-	0	1	0	0	a	b	b	c	b	1	-	b	1	1
Wari'	0	-	-	0	0	-	-	-	-	0	0	0	0	-	c	b	b	a	0	-	-	-	-
Chimila	1	0	1	-	0	-	?	-	-	0	1	1	1	?	b	b	a	b	?	?	?	?	?
Ika	0	-	-	-	0	1	?	0	b	0	1	1	1	c	b	c	a	b	1	1	-	0	0
Embera	0	-	-	-	0	-	-	-	-	1	1	1	1	a	b	c	a	b	1	0	b	0	0
Tehuelche	0	-	-	-	0	-	-	-	-	1	1	0	0	b	a	a	c	b	?	?	?	?	?
Mocoví	1	0	1	0	1	1	1	0	b	0	0	0	0	-	c	a	c	a	1	1	-	0	0
Pilagá	0	-	-	-	1	0	0	1	c	0	0	0	0	-	c	c	c	c	0	-	-	-	-
Cholón	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	a	b	1	0	c	1	0
Itonama	1	1	1	0	0	-	-	-	-	0	0	0	0	-	c	c	c	c	1	1	-	0	0
Kanoê	0	-	-	-	0	-	-	-	-	0	0	0	0	-	c	c	c	b	0	-	-	-	-
Kwaza	0	-	-	-	0	-	-	-	-	0	0	1	1	c	a	a	a	b	1	0	b	0	0
Leko	0	-	-	-	0	-	-	-	-	0	0	1	1	b	c	c	c	b	1	0	a	0	0
Mapudungun	1	1	0	1	1	-	-	-	-	0	0	0	0	-	c	c	c	b	1	0	b	0	0
Movima	1	1	0	1	0	-	-	-	-	0	0	0	0	-	c	c	b	b	1	0	b	0	1

Questionnaire codings (*Continued on next page*)

Table Questionnaire codings: Questions 3.1-6.1.4 *(Continued from previous page)*

LANGUAGE	3.1	3.1.1	3.1.2	3.1.3	4.1	4.1.1	4.1.2	4.1.3	4.1.4	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.1	6.1.1	6.1.2	6.1.3	6.1.4
Munichi	0	-	-	-	0	-	-	-	-	0	0	0	?	-	c	?	?	b	1	0	b	0	0
Nasa Yuwe	0	-	-	-	0	-	-	-	-	0	1	1	?	a	a	a	b	b	1	0	b	0	0
Puinave	0	-	-	-	0	-	-	-	-	0	1	0	1	b	b	b	a	c	0	-	-	-	-
Trumai	0	-	-	-	0	-	-	-	-	0	1	0	1	a	b	b	a	b	1	0	d	0	0
Urarina	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	a	b	1	0	a	1	0
Warao	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	a	c	b	1	0	c	0	0
Yurakaré	0	-	-	-	1	0	0	0	b	0	0	0	0	-	c	c	c	b	1	1	-	0	0
Aguaruna	0	-	-	-	0	-	-	-	-	0	1	1	1	a	a	a	c	b	1	0	b	0	0
Katukina	0	-	-	-	0	-	-	-	-	1	0	0	1	?	b	b	a	b	1	1	-	0	0
Bororo	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	a	b	0	-	-	-	-
Guató	?	?	?	?	0	-	-	-	-	0	0	0	0	-	c	c	c	a	1	1	-	?	?
Rikbaktsa	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	c	a	1	0	a	0	0
Timbira	0	-	-	-	0	-	-	-	-	0	1	0	1	a	b	d	a	b	1	1	-	0	0
Xavánte	0	-	-	-	0	-	-	-	-	0	0	0	1	c	c	c	a	b	0	-	-	-	-
Wichí	0	-	-	-	1	0	0	0	c	0	0	0	0	-	c	c	c	c	1	1	-	0	0
Moseten	0	-	-	-	0	-	-	-	-	0	0	0	0	-	c	c	a	b	1	0	b	0	0
Dâw	0	-	-	-	0	-	-	-	-	0	0	0	1	c	a	a	b	b	0	-	-	-	-
Hup	0	-	-	-	0	-	-	-	-	0	0	0	1	c	a	a	c	b	0	-	-	-	-
Latundê	0	-	-	-	1	0	1	1	b	0	0	0	1	-	c	c	c	b	1	1	-	0	0
Sabanê	0	-	-	-	1	1	?	?	b	0	0	1	1	a	a	a	b	b	?	?	?	?	?
Matses	0	-	-	-	0	-	-	-	-	0	1	0	0	c	b	b	c	b	1	0	c	0	0
Shipibo	0	-	-	-	0	-	-	-	-	0	1	0	0	a	b	b	c	b	1	0	c	1	1
Yaminahua	0	-	-	-	0	-	-	-	-	0	1	0	0	a	b	a	c	b	1	0	b	0	0
Huallaga Q.	0	-	-	-	0	-	-	-	-	0	0	1	1	a	a	a	c	b	1	0	a	0	1
Imbabura Q.	0	-	-	-	0	-	-	-	-	0	0	1	1	a	a	a	c	b	1	0	c	0	0
Cavineña	0	-	-	-	0	-	-	-	-	0	1	0	0	a	b	b	c	b	1	0	c	1	1
Cubeo	0	-	-	-	0	-	-	-	-	0	0	0	1	b	c	a	b	b	1	?	c	?	?
Desano	0	-	-	-	0	-	-	-	-	0	0	1	1	b	c	a	a	b	1	0	c	0	1
Awetí	1	0	1	0	1	0	1	0	b	0	0	0	1	-	c	a	a	a	1	0	c	0	1
Cocama	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	a	b	1	0	c	1	0
Emerillon	1	0	1	0	1	0	?	0	b	0	0	0	1	-	c	c	a	b	1	0	b	1	1
Juruna	0	-	-	-	1	?	?	0	b	0	0	1	1	b	a	a	b	b	1	1	-	1	0

Questionnaire codings *(Continued on next page)*

Table Questionnaire codings: Questions 3.1-6.1.4 (Continued from previous page)

LANGUAGE	3.1	3.1.1	3.1.2	3.1.3	4.1	4.1.1	4.1.2	4.1.3	4.1.4	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.1	6.1.1	6.1.2	6.1.3	6.1.4
Kamayurá	1	0	1	0	1	1	0	1	b	0	0	0	1	?	c	c	a	b	1	0	b	0	1
Karitiana	0	-	-	-	0	-	-	-	-	0	0	0	0	-	c	c	b	b	1	1	-	0	0
Karo	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	c	a	b	1	0	c	0	1
Mekens	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	a	a	c	1	0	b	0	1
Munduruku	1	0	1	0	1	1	1	?	b	0	0	0	1	-	c	a	a	b	1	0	b	0	1
Nheengatu	0	-	-	-	1	?	?	0	c	0	0	0	1	-	c	d	a	b	1	1	-	0	0
Sateré	1	0	1	0	1	0	1	0	b	0	0	0	1	-	c	c	a	b	1	1	-	0	1
Tapiete	1	0	1	0	1	?	1	?	b	0	0	0	0	-	c	a	c	b	1	0	c	0	0
Wayoró	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	a	a	b	1	1	-	0	0
Chipaya	0	-	-	-	0	-	-	-	-	0	0	0	1	-	c	a	a	b	1	0	?	0	0
Miraña	0	-	-	-	0	-	-	-	-	0	0	1	1	c	a	a	b	b	1	0	b	0	0
Yanam	1	0	0	0	0	-	-	-	-	0	1	0	1	c	b	b	a	b	1	0	b	0	0

Table C.3: Questionnaire codings: Questions 7.1-12.1.1

LANGUAGE	7.1	7.1.1	7.1.2	7.1.3	7.1.4	8.1	8.1.1	9.1	9.1.1	9.1.2	10.1	10.2	11.1	11.2	11.3	11.4	12.1	12.1.1
Apurinã	0	-	-	-	-	a	1	1	1	0	1	c	1	0	c	0	?	?
Baure	1	0	0	1	0	a	0	1	0	0	1	c	1	0	c	0	?	?
Lokono	?	?	?	?	?	a	0	1	1	?	1	b	?	?	?	?	?	?
Paresi	1	0	0	1	1	a	1	0	-	-	1	c	1	0	c	?	1	1
Tariana	0	-	-	-	-	a	1	0	-	-	1	c	1	0	c	1	?	?
Yanesha'	?	?	?	?	?	?	?	?	?	?	1	c	1	0	c	0	?	?
Jarawara	1	0	0	0	1	d	-	0	-	-	0	c	0	1	c	0	?	?
Aymara	1	1	1	1	1	d	-	0	-	-	1	c	1	0	c	1	?	?
Jaqaru	?	?	?	?	?	d	-	?	?	?	0	b	1	0	c	1	?	?
Awa Pit	0	-	-	-	-	d	-	0	-	-	0	c	0	0	c	0	?	?
Tsafiki	0	-	-	-	-	d	-	0	-	-	0	a	1	0	c	0	?	1
Apalaí	0	-	-	-	-	a	1	?	?	?	1	?	1	0	?	1	?	1
Hixkaryana	0	-	-	-	-	a	0	0	-	-	1	c	1	0	?	?	1	1
Ikpeng	?	-	-	-	-	d	-	0	-	-	1	c	?	?	?	?	1	1
Tiriyó	0	-	-	-	-	a	?	1	0	1	1	c	1	0	c	1	1	1
Chayahuita	0	-	-	-	-	d	-	?	?	?	1	c	1	0	c	1	1	0
Wari'	0	-	-	-	-	a	0	0	-	-	1	c	1	1	c	1	1	0
Chimila	?	?	?	?	?	d	-	?	?	?	?	?	?	?	?	?	0	?
Ika	1	0	0	1	0	d	-	0	-	-	1	?	1	0	c	1	?	?
Embera	0	-	-	-	-	d	-	0	-	-	0	?	?	?	?	?	?	?
Tehuelche	0	-	-	-	-	d	-	0	-	-	1	c	?	0	c	?	?	?
Mocoví	1	0	0	-	0	a	1	1	0	0	1	c	1	0	c	1	1	1
Pilagá	1	0	0	0	0	d	-	0	-	-	1	c	1	0	c	0	?	-
Cholón	1	1	1	1	?	a	0	0	-	-	1	c	?	?	?	0	?	?
Itonama	1	1	1	1	0	d	-	0	-	-	1	c	1	0	c	0	?	-
Kanoê	0	-	-	-	-	d	-	0	-	-	?	?	?	?	c	1	0	1
Kwaza	1	0	1	1	0	a	0	0	-	-	1	c	1	0	c	?	?	?
Leko	1	1	1	1	1	a	0	0	-	-	0	c	1	0	c	0	1	?
Mapudungun	1	0	0	1	?	d	0	0	-	-	1	c	1	0	c	1	?	?
Movima	1	0	0	1	1	-	-	1	1	0	1	c	1	0	c	1	?	?

Questionnaire codings (*Continued on next page*)

Table Questionnaire codings: Questions 7.1-12.1.1 *(Continued from previous page)*

LANGUAGE	7.1	7.1.1	7.1.2	7.1.3	7.1.4	8.1	8.1.1	9.1	9.1.1	9.1.2	10.1	10.2	11.1	11.2	11.3	11.4	12.1	12.1.1
Munichi	0	-	-	-	-	a	0	0	-	-	?	?	1	0	c	?	?	?
Nasa Yuwe	1	0	0	1	1	b	1	0	-	-	1	c	1	?	?	1	?	?
Puinave	?	?	?	?	?	d	-	?	?	?	0	c	1	0	c	0	?	?
Trumai	0	-	-	-	-	d	1	0	-	-	0	c	0	0	c	0	0	-
Urarina	0	-	-	-	-	c	1	0	-	-	1	c	0	1	c	0	1	1
Warao	0	-	-	-	-	a	1	0	-	-	1	c	1	0	c	1	?	?
Yurakaré	1	0	0	1	1	a	0	0	-	-	1	b	1	0	c	1	1	1
Aguaruna	1	1	1	1	0	d	-	0	-	-	1	c	1	0	c	0	?	?
Katukina	1	0	?	1	1	d	-	1	-	?	1	c	1	0	b	1	1	?
Bororo	?	?	?	?	?	a	1	1	?	?	0	a	0	0	a	0	?	?
Guató	?	?	?	?	?	d	?	?	?	?	?	?	0	0	?	0	?	?
Rikbaktsa	0	-	-	-	-	?	?	d	-	0	1	0	c	1	0	c	1	1
Timbira	0	-	-	-	-	d	-	1	0	0	0	a	0	0	a	0	1	0
Xavánte	0	-	-	-	-	d	-	?	?	?	1	?	1	0	?	0	1	1
Wichí	1	0	0	1	0	d	0	0	-	-	1	c	1	0	c	1	0	-
Moseten	1	0	0	1	0	a	1	1	1	1	1	b	1	0	b	1	1	0
Dâw	?	?	?	?	?	a	0	0	-	-	0	a	1	0	c	0	0	0
Hup	1	1	1	1	0	a	1	0	-	-	1	c	1	0	c	0	?	?
Latundê	1	0	0	1	1	d	-	0	-	-	1	c	1	0	c	1	?	?
Sabanê	?	?	?	?	?	?	?	?	?	?	0	c	?	?	?	0	?	?
Matses	1	1	1	1	0	a	1	1	0	0	1	c	1	0	c	0	1	1
Shipibo	1	0	0	1	1	a	0	0	-	-	1	b	1	0	c	0	1	1
Yaminahua	1	0	0	?	?	d	-	0	-	-	1	c	1	0	c	0	?	?
Huallaga Q.	1	1	1	1	0	c	0	0	-	-	1	c	1	0	c	0	?	?
Imbabura Q.	1	1	1	-	-	b	1	1	-	1	1	c	1	0	c	1	?	?
Cavineña	0	-	-	-	-	a	0	1	0	-	1	c	1	0	c	1	?	?
Cubeo	1	1	1	-	-	b	1	0	-	1	?	c	1	0	c	1	1	?
Desano	1	1	0	1	0	a	1	1	-	-	0	a	0	1	b	1	?	?
Awetí	0	-	-	-	-	d	-	0	-	-	1	c	1	0	c	0	?	?
Cocama	1	1	1	0	1	d	-	0	?	?	0	?	1	?	?	1	?	1
Emerillon	0	-	-	-	-	d	-	0	-	-	1	c	1	0	c	?	?	?

Table Questionnaire codings: Questions 7.1-12.1.1 (Continued from previous page)

LANGUAGE	7.1	7.1.1	7.1.2	7.1.3	7.1.4	8.1	8.1.1	9.1	9.1.1	9.1.2	10.1	10.2	11.1	11.2	11.3	11.4	12.1	12.1.1
Juruma	0	-	-	-	-	?	?	0	?	?	1	c	0	1	c	0	?	?
Kamayurá	0	-	-	-	-	d	-	0	-	-	1	c	1	0	c	0	0	-
Karitiana	0	-	-	-	-	a	0	0	-	-	0	a	?	?	?	?	1	0
Karo	0	-	-	-	-	a	0	0	-	-	1	c	1	0	c	0	?	?
Mekens	0	-	-	-	-	d	-	0	-	-	0	a	0	0	a	1	?	?
Mundurukú	0	-	-	-	-	a	1	?	?	?	1	c	1	0	c	1	1	1
Nheengatu	0	-	-	-	-	a	0	0	-	-	1	c	1	0	c	1	?	?
Sateré	0	-	-	-	-	d	-	?	-	?	0	a	1	0	c	0	1	1
Tapiete	0	-	-	-	-	d	-	0	-	-	1	c	0	0	c	1	1	1
Wayoró	0	-	-	-	-	d	-	?	?	-	1	b	?	?	?	1	1	1
Chipaya	1	1	1	?	1	a	1	?	?	?	1	c	1	0	c	0	1	1
Miraña	0	-	-	-	-	a	0	0	-	-	1	c	1	0	c	0	?	?
Yanam	1	0	0	?	1	d	-	?	?	?	1	c	1	0	c	1	?	?

Bibliography

Adelaar, W. (2006). The Quechua impact in Amuesha, an Arawak language of the Peruvian Amazon. In A. Aikhenvald and R. M. W. Dixon (Eds.), *Grammars in Contact: A Cross-Linguistic Typology*, pp. 290–312. Oxford: Oxford University Press.

Adelaar, W. (2008). Towards a typological profile of the Andean languages. In A. Lubotsky, J. Schaeken, and J. Wiedenhof (Eds.), *Evidence and Counter-Evidence. Essays in Honour of Frederik Kortlandt: Volume 2: General linguistics*, Volume 32 of *Studies in Slavic and General Linguistics*, pp. 23–33. Amsterdam: Rodopi.

Adelaar, W. (2012). Languages of the Middle Andes in areal-typological perspective: Emphasis on Quechuan and Aymaran. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A Comprehensive Guide*, pp. 575–624. Berlin: De Gruyter Mouton.

Adelaar, W. F. H. and P. C. Muysken (2004). *The Languages of the Andes*. Cambridge: Cambridge University Press.

Aguirre Licht, D. (1998). *Fundamentos Morfosintácticos para una Gramática Emberá*. Bogotá: Centro Colombiano de Estudios de Lenguas Aborígenes, Universidad de los Andes.

Aikhenvald, A. (1999). Arawak. In R. M. W. Dixon and A. Aikhenvald (Eds.), *The Amazonian Languages*, pp. 65–106. Cambridge: Cambridge University Press.

Aikhenvald, A. (2002). *Language Contact in Amazonia*. Oxford: Oxford University Press.

Aikhenvald, A. (2003a). *A Grammar of Tariana, from Northwest Amazonia*. Cambridge: Cambridge University Press.

Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices.* Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Aikhenvald, A. Y. (2003b). Multilingualism and ethnic stereotypes: The Tariana of northwest Amazonia. *Language in Society 32*, 1–21.

Aikhenvald, A. Y. (2006a). Grammars in contact: A cross-linguistic perspective. In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Grammars in Contact: A cross-linguistic typology*, pp. 1–66. Oxford: Oxford University Press.

Aikhenvald, A. Y. (2006b). Semantics and pragmatics of grammatical relations in the Vaupés linguistic area. In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Grammars in Contact: A Cross-Linguistic Typology*, pp. 237–266. Oxford: Oxford University Press.

Aikhenvald, A. Y. (2011). Areal features and linguistic areas: Contact-induced change and geographical typology. In O. Hieda, C. König, and H. Nakagawa (Eds.), *Geographical Typology and Linguistic Areas: With special reference to Africa*, pp. 13–39. Amsterdam: John Benjamins Publishing Company.

Aikhenvald, A. Y. (2012). *Languages of the Amazon.* Oxford: Oxford University Press.

Aikhenvald, A. Y. and R. M. W. Dixon (2010). Areal features and regional traits. *Paper presented at the Max Planck Institute for Psycholinguistics, November 2010*.

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory 21*(3), 435–483.

Alexander-Bakkerus, A. (2005). *Eighteenth-century Cholón.* Doctoral dissertation, Universiteit Leiden.

Anderson, S. (1976). On the notion of subject in ergative languages. In C. N. Li (Ed.), *Subject and Topic*, pp. 1–25. Academic Press.

Andrews, A. D. (2007). The major functions of the noun phrase. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Volume 1: Clause Structure* (2nd ed.)., pp. 132–223. Cambridge: Cambridge University Press.

Antunes de Araujo, G. (2004). *A grammar of Sabanê: A Nambikwaran language.* Doctoral dissertation, Vrije Universiteit Amsterdam.

Arroyo-Kalin, M. (2010). The Amazonian formative: Crop domestication and anthropogenic soils. *Diversity 2*, 473–504.

Bacelar, L. N. (2004). *Gramática da língua Kanoê.* Doctoral dissertation, Katholieke Universiteit Nijmegen.

Baker, M. C. (1996). *The Polysynthesis Parameter*. Oxford: Oxford University Press.

Baker, M. C. (2001). *The Atoms of Language: The mind's hidden rules of grammar*. New York: Basic Books.

Barnes, J. (1999). Tucano. In R. M. W. Dixon and A. Aikhenvald (Eds.), *The Amazonian Languages*, pp. 207–226. Cambridge: Cambridge University Press.

Beresford-Jones, D. and P. Heggarty (2012). Archaeology, linguistics, and the Andean past: A much-needed conversation. In P. Heggarty and D. Beresford-Jones (Eds.), *Archaeology and Language in the Andes: A cross-disciplinary exploration of prehistory*, pp. 1–41. Oxford: Oxford University Press.

Bickel, B. (2008). On the scope of the referential hierarchy in the typology of grammatical relations. In G. C. Corbett and M. Noonan (Eds.), *Case and Grammatical Relations: Studies in honor of Bernard Comrie*, pp. 191–210. Amsterdam: John Benjamins Publishing Company.

Bickel, B. (2011). Grammatical relations typology. In J. J. Song (Ed.), *The Oxford Handbook of Language Typology*, pp. 399–444. Oxford: Oxford University Press.

Bickel, B. and J. Nichols (2002). Autotypologizing databases and their use in fieldwork. In P. Austin, H. Dry, and P. Wittenburg (Eds.), *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics, Las Palmas, 2627 May 2002*. Nijmegen: MPI for psycholinguistics.

Bickel, B. and J. Nichols (2006). Oceania, the Pacific Rim and the theory of linguistic areas. Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society.

Bickel, B. and J. Nichols (2007). Inflectional morphology. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Volume III: Grammatical categories and the lexicon* (2 ed.)., Language Typology and Syntactic Description, pp. 169–240. Cambridge: Cambridge University Press.

Bickel, B. and J. Nichols (2009). Case marking and alignment. In A. Malchukov and A. Spencer (Eds.), *The Oxford Handbook of Case*, pp. 304–321. Oxford: Oxford University Press.

Birchall, J. (2014a). The multi-verb benefactive construction in Wari' and Oro Win. In F. Queixalós, S. Telles, and A. C. Bruno (Eds.), *Valency changing strategies in South American languages*, pp. 101–119. Bogotá: Instituto Caro y Cuervo.

Birchall, J. (2014b). Verbal argument marking patterns in South American languages. In L. O'Connor and P. Muysken (Eds.), *The Native Languages of South America: Origins, development, typology*, pp. 223–249. Cambridge: Cambridge University Press.

Blake, B. J. (2001). *Case* (2 ed.). Cambridge: Cambridge University Press.

Blake, B. J. (2009). History of the research on case. In A. Malchukov and A. Spencer (Eds.), *The Oxford Handbook of Case*, pp. 13–26. Oxford: Oxford University Press.

Bloomfield, L. (1933). *Language*. New York: H. Holt and Company.

Boas, F. (1911). Introduction. In F. Boas (Ed.), *Handbook of American Indian Languages: Volume 1*, pp. 1–84. Washington: Government Printing Office.

Bolaños, K. and P. Epps (2009). Linguistic classification of Kakua, a language of northwest Amazonia. Paper presented at the Conference on Indigenous Languages of Latin America IV, October 29-31, 2009, University of Texas at Austin.

Borella, C. d. C. (2000). Aspectos morfossintáticos da língua Aweti (tupi). Master's thesis, Universidade Estadual de Campinas.

Bossong, G. (1984). Animacy and markedness in Universal Grammar. *Glossologia 39*(2-3), 7–20.

Bowern, C. (2009). Defining complexity: Historical reconstruction and Nyulnyulan subordination. *Rice Working Papers in Linguistics 1*, 1–17.

Braga, A. d. O. (2005). *Aspects morphosyntaxiques de la langue Makurap/Tupi*. Doctoral dissertation, Université de Toulouse – Le Mirail.

Brandão, A. P. B. (2010). Verb morphology in Paresi-Haliti (Arawak). Master's thesis, University of Texas at Austin.

Bresnan, J. (2001). *Lexical-Functional Syntax*. Oxford: Blackwell.

Bresnan, J. and S. A. Mchombo (1987). Topic, pronoun, and agreement in Chichêwa. *Language 63*, 741–782.

Brinton, D. G. (1891). *The American Race: A Linguistic classification and ethnographic description of the native tribes of North and South America*. New York: N.D.C. Hodges Publisher.

Browman, D. (2001). Highland Andean formative. In P. Peregrine and M. Ember (Eds.), *Encyclopedia of Prehistory, Volume 7: South America*, pp. 123–137. New York: Springer.

Büttner, T. T. (1983). *Las Lenguas de los Andes Centrales: Estudios sobre la clasificación genética, areal y tipológica.* Madrid: Cultura Hispánica del Instituto de cooperación iberoamericana.

Cabral, A. S. A. C. (2002). Natureza e direções das mudanças de alinhamento ocorridas no tronco Tupí. In F. Queixalós (Ed.), *Ergatividade na Amazônia*, Volume 1, pp. 5–7. Paris and Brasilia: Centre d'études des langues indigènes d'Amérique and the Laboratório de Línguas Indígenas.

Cabral, A. S. D. A. C. (1995). *Contact-Induced language change in the Western Amazon: The non-genetic origin of the Kokama language.* Doctoral dissertation, University of Pittsburgh.

Caesar-Fox, D. C. (2003). *Zauro'nödok Agawayo Yau: Variants of Akawaio spoken at Waramadong.* Doctoral dissertation, Rice University.

Campbell, L. (1991). On so-called pan-americanisms. *International Journal of American Linguistics 57*(3), 394–399.

Campbell, L. (2012a). Classification of the indigenous languages of South America. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 59–166. Berlin: De Gruyter Mouton.

Campbell, L. (2012b). Typological characteristics of South American indigenous languages. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 259–330. Berlin: De Gruyter Mouton.

Campbell, L. and V. Grondona (2012). Languages of the Chaco and Southern Cone. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 625–667. Berlin: De Gruyter Mouton.

Campbell, L. R., T. S. Kaufman, and T. C. Smith-Stark (1986). Meso-America as a linguistic area. *Language 62*(3), 530–570.

Castro Alves, F. d. (2004). *O Timbira falado pelos Canela Apãniekrá: Uma contribução aos estudos da morfossintaxe de uma língua Jê.* Doctoral dissertation, Universidade Estadual de Campinas.

Castro Alves, F. d. (2010). Evolution of alignment in Timbira. *International Journal of American Linguistics 76*(4), 439–475.

Cerrón-Palomino, R. (2000). *Lingüística Aimara.* Cusco: Centro de Estudios Regionales Andinos.

Cerrón-Palomino, R. (2003). *Lingüística Quechua* (2 ed.). Cusco: Centro de Estudios Rurales Andinos "Bartolomé de Las Casas".

Cerrón-Palomino, R. (2008). *Quechumara: Estructuras paralelas de las lenguas quechua y aymara*. La Paz: Plural editores.

Cerrón-Palomino, R. (2009). Chipaya. In M. Crevels and P. Muysken (Eds.), *Lenguas de Bolivia. Tomo I: Ámbito Andino*, Volume 1 of *Lenguas de Bolivia*, pp. 29–78. La Paz: Plural editores.

Cerrón-Palomino, R. (2012). Unravelling the enigma of the 'Particular Language' of the Incas. In P. Heggarty and D. Beresford-Jones (Eds.), *Archaeology and Language in the Andes: A cross-disciplinary exploration of prehistory*, pp. 265–294. Oxford: Oxford University Press.

Cerrón-Palomino, R. and J. Carvajal Carvajal (2009). Aimara. In M. Crevels and P. Muysken (Eds.), *Lenguas de Bolivia. Tomo I: Ámbito Andino*, pp. 169–213. La Paz: Plural editores.

Chacon, T. C. (2012). *The phonology and morphology of Kubeo: The documentation, theory, and description of an Amazonian language*. Doctoral dissertation, University of Hawai'i at Manoa.

Chacon, T. C. (2014). A revised proposal of proto-Tukanoan consonants and Tukanoan family classification. *International Journal of American Linguistics*, to appear.

Chagas, A. F. A. (2013). *O verbo Ikpeng: Estudo morfossintático e semântico-lexical*. Doctoral dissertation, Universidade Estadual de Campinas.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT press.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Cole, P. (1982). *Imbabura Quechua*. Amsterdam: North Holland Publishing Co.

Cole, P. (1983). The grammatical gole of the causee in Universal Grammar. *International Journal of American Linguistics 49*(2), 115–133.

Comrie, B. (1981). *Language Universals and Linguistic Typology: Syntax and morphology* (1 ed.). Oxford: Blackwell.

Comrie, B. (1985). Causative verb formation and other verb-deriving morphology. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Volume III: Grammatical categories and the lexicon* (1 ed.)., pp. 309–348. Cambridge: Cambridge University Press.

Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and morphology* (2 ed.). Chicago: University of Chicago Press.

Comrie, B. (1993). Argument structure. In J. Jacob, A. von Stechow, W. Sternefeld, and T. Vennemann (Eds.), *Syntax: An international handbook of contemporary research*, Volume 1, pp. 905–914. Berlin: Walter de Gruyter.

Comrie, B. (2011). Alignment of case marking of full noun phrases. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Conklin, H. C. (1962). Lexicographical treatment of folk taxonomies. In F. W. Householder and S. Saporta (Eds.), *Problems in Lexicography*, pp. 119–141. Bloomington: Indiana University Research Center in Anthropology, Folklore and Linguistics.

Constenla Umaña, A. (1991). *Las Lenguas del Área Intermedia: Introducción a su estudio areal*. San José: Universidad de Costa Rica.

Corbett, G. C. (1991). *Gender.* Cambridge: Cambridge University Press.

Corbett, G. C. (2000). *Number.* Cambridge: Cambridge University Press.

Corbett, G. C. (2006). *Agreement.* Cambridge: Cambridge University Press.

Corrêa-da Silva, B. C. (2010). *Mawé/Awetí/Tupí-Guaraní: Relações linguísticas e implicações históricas*. Doctoral dissertation, Universidade de Brasília.

Cosme Bueno, F. (1770). *Descripcion del Obispado de Santa Cruz.* El conocimiento de los tiempos, ephemeride del año de 1771. Lima: En la Imprenta Real; Calle de Palacio.

Creissels, D. (2008). Remarks on split intransitivity and fluid intransitivity. In O. Bonami and P. Cabredo Hofherr (Eds.), *Empirical Issues in Syntax and Semantics*, Volume 7, pp. 139–168.

Creissels, D. (2009). Uncommon patterns of core term marking and case terminology. *Lingua 119*(3), 445–459.

Creissels, D. (2010). Benefactive applicative periphrases: A typological approach. In F. Zúñiga and S. Kittilä (Eds.), *Benefactives and Malefactives: Typological perspectives and case studies*, pp. 29–69. Amsterdam: John Benjamins Publishing Company.

Crevels, M. (2011). Who did what to whom in Magdalena (Itonama). *International Journal of American Linguistics 77*(4), 577–594.

Crevels, M. (2012a). Itonama. In M. Crevels and P. Muysken (Eds.), *Lenguas de Bolivia, Vol. II: Amazonía*, pp. 233–294. La Paz: Plural editores.

Crevels, M. (2012b). Language endangerment in South America: The clock is ticking. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 167–233. Berlin: De Gruyter Mouton.

Crevels, M. and P. Muysken (2005). Inclusive-exclusive distinctions in the languages of central-western South America. In E. Filimonova (Ed.), *Clusivity: Typology and case stu*, pp. 313–340. Amsterdam: John Benjamins Publishing Company.

Crevels, M. and H. van der Voort (2008). The Guaporé-Mamoré region as a linguistic area. In P. Muysken (Ed.), *From Linguistic Areas to Areal Linguistics*, Volume 90 of *Studies in Language Companion Series*, pp. 151–179. Amsterdam: John Benjamins Publishing Company.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Crowell, T. H. (1979). *A grammar of Bororo*. Doctoral dissertation, Cornell University.

Cruz, A. d. (2011). *Fonologia e Ggamática do Nheengatú: A língua geral falada pelos povos Baré, Warekena e Baniwa*. Doctoral dissertation, Vrije Universiteit Amsterdam.

Curnow, T. (1997). *A grammar of Awa Pit (Cuaiquier): An indigenous language of south-western Colombia*. Doctoral dissertation, Australian National University.

Cysouw, M. (2003). *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

Cysouw, M. (2005). Quantitative methods in typology. In R. Kölher, G. Altmann, and R. Piotrowski (Eds.), *Quantitative Linguistik: Ein internationales Handbuch*, pp. 554–578. Berlin: Walter de Gruyter.

Cysouw, M. (2009). Reconstruction of morphosyntactic function: Nonspatial usage of spatial case marking in Tsezic. *Language 85*(3), 588–617.

Cysouw, M. (2011). Inclusive/Exclusive distinction in verbal inflection. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Danielsen, S. (2007). *Baure: An Arawak Language of Bolivia*, Volume 6 of *Indigenous Languages of Latin America*. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), Universiteit Leiden.

de Hoop, H. and P. de Swart (2008). Cross-linguistic variation in differential subject marking. In H. de Hoop and P. de Swart (Eds.), *Differential Subject Marking*, pp. 1–16. Dordrecht: Springer.

de Swart, P. (2007). *Cross-linguistic variation in object marking.* Doctoral dissertation, Radboud Universiteit Nijmegen.

DeLancey, S. (1997). Mirativity: The grammatical marking of unexpected information. *Lingusitic Typology 1*, 33–52.

Denevan, W. (1992). Native American populations in 1492: Recent research and a revised hemispheric estimate. In W. Denevan (Ed.), *The Native Populations of the Americas in 1492* (2 ed.)., pp. xvii–xxix. Madison: University of Wisconsin Press.

Denevan, W. (2001). *Cultivated Landscapes of Native Amazonia and the Andes.* Oxford: Oxford University Press.

Derbyshire, D. C. (1979). *Hixkaryana.* Amsterdam: North-Holland.

Derbyshire, D. C. (1985). *Hixkaryana and Linguistic Typology*, Volume 76 of *Summer Institute of Linguistics: Publications in Linguistics.* Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Derbyshire, D. C. (1987). Morphosyntactic areal characteristics of Amazonian languages. *International Journal of American Linguistics 53*(3), 311–326.

Dickinson, C. (2002). *Complex predicates in Tsafiki.* Doctoral dissertation, University of Oregon.

Dik, S. (1997). *The Theory of Functional Grammar.* Berlin: Mouton de Gruyter.

Dillehay, T. D. (2008). Profiles of Pleistocene history. In H. Silverman and W. Isbell (Eds.), *The Handbook of South American Archaeology*, pp. 29–43. New York: Springer.

Dixon, R. M. W. (1979). Ergativity. *Language 55*, 57–133.

Dixon, R. M. W. (1986). Noun classes and noun classification in typological perspective. In C. Craig (Ed.), *Noun Classes and Categorization*, pp. 105–112. Amsterdam: John Benjamins Publishing Company.

Dixon, R. M. W. (1994). *Ergativity.* Cambridge: Cambridge University Press.

Dixon, R. M. W. (1997). *The Rise and Fall of Languages.* Cambridge: Cambridge University Press.

Dixon, R. M. W. (2000). A typology of causatives: form, syntax and meaning. In R. M. W. Dixon and A. Y. Aikhenvald (Eds.), *Changing Valency: Case studies in transitivity*, pp. 30–83. Cambridge: Cambridge University Press.

Dixon, R. M. W. (2004). *The Jarawara Language of Southern Amazonia.* Oxford: Oxford University Press.

Dixon, R. M. W. (2010). *Basic Linguistic Theory. Volume 2: Grammatical Topics.* Oxford: Oxford University Press.

Dixon, R. M. W. and A. Aikhenvald (1999). Introduction. In R. M. W. Dixon and A. Aikhenvald (Eds.), *The Amazonian Languages*, Cambridge Language Surveys, pp. 1–22. Cambridge: Cambridge University Press.

Dixon, R. M. W. and A. Y. Aikhenvald (1997). A typology of argument-determined constructions. In J. Bybee, J. Haiman, and S. A. Thompson (Eds.), *Essays on Language Function and Language Type. Dedicated to T. Givón*, pp. 71–113. Amsterdam: John Benjamins Publishing Company.

Dixon, R. M. W. and A. Y. Aikhenvald (Eds.) (2000a). *Changing Valency: Case studies in transitivity.* Cambridge: Cambridge University Press.

Dixon, R. M. W. and A. Y. Aikhenvald (2000b). Introduction. In R. M. W. Dixon and A. Y. Aikhenvald (Eds.), *Changing Valency: Case Studies in Transitivity*, pp. 1–29. Cambridge: Cambridge University Press.

Donohue, M. and S. Wichmann (Eds.) (2008). *The Typology of Semantic Alignment.* Oxford: Oxford University Press.

d'Orbigny, A. D. (1839). *L'homme américain (de l'Amérique Méridionale): considéré sous ses rapports physiologiques et moraux*, Volume 4. Paris: Pitois-Levrault et C.

dos Anjos, Z. (2011). *Fonologia e gramática Katukina-Kanamari.* Doctoral dissertation, Vrije Universiteit Amsterdam.

dos Santos, J. P. (2008). *Marcas pessoais, concordância de número e alinhamento em Xavánte.* Ma thesis, Universidade de Brasília.

Dowty, D. R. (1991). Thematic proto-roles and argument selection. *Language 67*(3), 547–619.

Drude, S. (2011). 'Derivational verbs' and other multi-verb constructions in Awetí and Tupí-Guaraní. In A. Y. Aikhenvald and P. C. Muysken (Eds.), *Multi-verb Constructions: A view from the Americas*, pp. 213–254. Leiden: Brill.

Drude, S. and S. Meira (2015). A reconstruction of Proto-Maweti-Guaraní. *Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas*, to appear.

Dryer, M. (1986). Primary objects, secondary objects, and antidative. *Language 64*(4), 808–845.

Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language 13*(2), 257–292.

Dryer, M. S. (1992). The Greenbergian Word Order Correlations. *Language 68*, 81–138.

Dryer, M. S. and M. Haspelmath (2011). The World Atlas of Language Structures Online.

Duff-Tripp, M. (1997). *Gramática del idioma Yanesha' (Amuesha)*, Volume 43 of *Serie Lingüística Peruana*. Lima: Ministerio de Educación and Instituto Lingüístico de Verano.

Dunn, M., S. C. Levinson, E. Lindström, G. Reesink, and A. Terrill (2008). Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language 84*(4), 710–759.

Dunn, M., A. Terrill, G. Reesink, R. A. Foley, and S. C. Levinson (2005). Structural Phylogenetics and the Reconstruction of Ancient Language History. *Science 309*(5743), 2072–2075.

Emperaire, J. (1955). *Les Nomades de la Mer*. Paris: Gallimard.

Enfield, N. J. (2005). Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology 34*, 181–206.

Epps, P. (2006). The Vaupés melting pot: Tucanoan influence on Hup. In A. Y. Aikhenvald and R. M. W. Dixon (Eds.), *Grammars in Contact: A Cross-Linguistic Typology*, pp. 267–289. Oxford: Oxford University Press.

Epps, P. (2008). *A Grammar of Hup*. Berlin: Mouton de Gruyter.

Eriksen, L. (2011). *Nature and culture in crehistoric Amazonia: Using G.I.S. to reconstruct ancient ethnogenetic processes from archaeology, linguistics, geography, and ethnohistory*. Doctoral dissertation, Lund Universitet.

Eriksen, L. and A. V. Galucio (2014). The Tupian expansion. In L. O'Connor and P. C. Muysken (Eds.), *The Native Languages of South America: Origins, Development, Typology*, pp. 177–199. Cambridge: Cambridge University Press.

Estevam, A. M. (2011). *Morphosyntaxe du xavante, langue jê du Mato Grosso (Brésil)*. Doctoral dissertation, Université Denis Diderot (Paris 7).

Evans, N. and S. Levinson (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences 32*(5), 429–492.

Everett, C. (2006). *Gestural, perceptual, and conceptual patterns in Karitiana*. Doctoral dissertation, Rice University.

Everett, D. L. and B. Kern (1997). *Wari': the Pacaás Novos language of Western Brazil*. London and New York: Routledge.

Facundes, S. (2000). *The language of the Apurinã people of Brazil (Maipure/Arawak)*. Doctoral dissertation, State University of New York at Buffalo.

Facundes, S. d. S. (2002). Historical linguistics and its contribution to improving knowledge of Arawak. In J. D. Hill and F. Santos-Granero (Eds.), *Comparative Arawakan histories: rethinking language family and culture area in Amazonia*, pp. 74–96. Urbana, IL: University of Illinois Press.

Fargetti, C. M. (2001). *Estudo fonológico e morfossintático da língua Juruna*. Doctoral dissertation, Universidade Estadual de Campinas.

Faust, N. and E. E. Loos (2002). *Gramática del Idioma Yaminahua*. Lima: Instituto Lingüístico de Verano.

Fausto, C., B. Franchetto, and M. J. Heckenberger (2008). Language, ritual and historical reconstruction: towards a linguistic, ethnographical and archaeological account of Upper Xingu Society. In K. D. Harrison, D. S. Rood, and A. Dwyer (Eds.), *Lessons from Documented Endangered Languages*, pp. 129–158. Amsterdam: John Benjamins Publishing Company.

Fernández Garay, A. (1998). *El Tehuelche: Una lengua en vías de extinción*, Volume 15 of *Estudios Filológicos*. Valdivia: Facultad de Filosofía y Humanidades, Universidad Austral de Chile.

Ferreira, H. P. (2012). Argument marking patterns across Yanomami languages. *Paper presented at the Workshop on Linguistic Typology and Language Contact in South America, Nijmegen*.

Filimonova, E. (Ed.) (2005). *Clusivity: Typology and case studies of the inclusive-exclusive distinction*. Amsterdam: John Benjamins Publishing Company.

Fillmore, C. (1968). The case for case. In E. Bach and R. Harms (Eds.), *Universals in Linguistic Theory*, pp. 1–90. New York: Holt, Rinehart and Winston.

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specified tree topology. *Systematic Zoology 20*(4), 406–416.

Fleck, D. W. (2003). *A grammar of Matses*. Doctoral dissertation, Rice University.

Fleck, D. W. (2010). Ergativity in the Mayoruna branch of the Panoan family. In S. Gildea and F. Queixalós (Eds.), *Ergativity in Amazonia*. Amsterdam: John Benjamins Publishing Company.

Floyd, S. I. (2011). Re-discovering the Quechua adjective. *Linguistic Typology 15*, 25–63.

Foley, W. A. (2005). Semantic parameters and the unaccusative split in the Austronesian language family. *Studies in Language 29*(2), 385–430.

Foley, W. A. and R. D. Van Valin Jr. (1984). *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Franchetto, B. (2011). Evidências linguísticas para o entendimento de uma sociedade multilíngue: O Alto Xingu. In B. Franchetto (Ed.), *Alto Xingu: uma sociedade multilíngue*, pp. 3–38. Rio de Janeiro: Museu do Indio - FUNAI.

Frank, P. S. (1985). *A grammar of Ika (Chibchan, Colombia)*. Doctoral dissertation, University of Pennsylvania.

Gabas Jr., N. (1999). *A grammar of Karo, Tupí (Brazil)*. Doctoral dissertation, University of California at Santa Barbara.

Galucio, A. V. (2001). *The morphosyntax of Mekens (Tupi)*. Doctoral dissertation, The University of Chicago.

Galucio, A. V. (2005). Puruborá: Notas etnográficas e lingüísticas recentes. *Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas 1*(2), 159–192.

Galucio, A. V. and A. F. Nogueira (2012). Comparative study of the Tupari branch of the Tupi family: contributions to understanding its historical development and internal classification. In *Memorias del V Congreso de Idiomas Indígenas de Latinoamérica, 6-8 de octubre de 2011, Universidad de Texas en Austin*, Austin, pp. 1–28.

Gibson, M. L. (1996). *El Munichi: Un idioma que se extingue*. Pucallpa: Instituto Lingüístico de Verano.

Gildea, S. (1994). Semantic and pragmatic inverse: 'Inverse alignment' and 'inverse voice' in Carib of Surinam. In T. Givon (Ed.), *Voice and Inversion*, pp. 187–230. Amsterdam: John Benjamins Publishing Company.

Gildea, S. (1998). *On Reconstructing Grammar: Comparative Cariban morphosyntax*. Oxford: Oxford University Press.

Gildea, S. (2002). Reconstructing Pre-Proto-Tupi-Guarani main clause grammar. In *Atas do I Encontro Internacional do grupo de trabalho sobre línguas indígenas da ANPOLL*, Volume I, pp. 315–326. Belém: Universidade Federal do Pará.

Gildea, S. (2012). Lingusitic studies in the Cariban family. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 441–494. Berlin: De Gruyter Mouton.

Gildea, S. and D. Payne (2007). Is Greenberg's 'Macro-Carib' viable? *Boletim do Museu Paraense Emílio Goeldi: Ciências Humanas 2*(2), 19–72.

Girard, V. (1971). *Proto-Takanan Phonology*. Berkeley and Los Angeles: University of California Press.

Girón, J. M. (2008). *Una gramática del Wãnsöjöt (Puinave)*. Doctoral dissertation, Vrije Universiteit Amsterdam.

Givón, T. (2000). Internal reconstruction: As method, as theory. In S. Gildea (Ed.), *Reconstructing grammar: comparative linguistics and grammaticalization*, pp. 107–159. Amsterdam: John Benjamins Publishing Company.

Givón, T. (2001). *Syntax: An introduction*, Volume 2. Amsterdam: John Benjamins Publishing Company.

Goldberg, A. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Golluscio, L. (2010). Ditransitive constructions in Mapudungun. In A. Malchukov, M. Haspelmath, and B. Comrie (Eds.), *Studies in Ditransitive Constructions: A Comparative Handbook*, pp. 710–756. Berlin: De Gruyter Mouton.

Gomes, D. M. (2006). *Estudo morfológico e sintático da língua Mundurukú*. Doctoral dissertation, Universidade de Brasília.

González, H. A. (2005). *A grammar of Tapiete (Tupi-Guarani)*. Doctoral dissertation, University of Pittsburgh.

Goodwin Gómez, G. (1990). *The Shiriana dialect of Yanam (Northern Brazil)*. Doctoral dissertation, Columbia University.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of grammar*, pp. 73–113. Cambridge, MA: MIT press.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford: Stanford University Press.

Greenberg, J. H. (1988). The first person inclusive dual as an ambiguous category. *Studies in Language 12*, 1–18.

Grinevald, C. (2000). A morphosyntactic typology of classifiers. In G. Senft (Ed.), *Systems of Nominal Classification*, pp. 50–91. Cambridge: Cambridge University Press.

Grinevald, C. (2004). Classifiers. In G. Booij (Ed.), *Morphology: An International Handbook on Inflection and Word Formation*, pp. 1016–1031. Berlin: Mouton de Gruyter.

Grondona, V. M. (1998). *A grammar of Mocovi*. Doctoral dissertation, University of Pittsburgh.

Gruber, J. S. (1965). *Studies in lexical relations.* Doctoral dissertation, MIT.

Grundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. Moravcsik, and J. Wirth (Eds.), *Studies in Syntactic Typology*, pp. 209–239. Amsterdam: John Benjamins Publishing Company.

Gualdieri, C. B. (1998). *Mocoví (Guaycurú): Fonologia e morfossintaxe.* Doctoral dissertation, Universidade Estadual de Campinas.

Guidon, N. and G. Delibrias (1986). Carbon-14 dates point to man in the Americas 32,000 years ago. *Nature 321*, 769–771.

Guillaume, A. (2006). Revisiting "split ergativity" in Cavineña. *International Journal of American Linguistics 72*(2), 159–192.

Guillaume, A. (2008). *A Grammar of Cavineña.* Berlin: Mouton de Gruyter.

Guillaume, A. and F. Rose (2010). Sociative causative markers in South American languages: a possible areal feature. In F. Foricic (Ed.), *Essais de Typologie et de Linguistique Générale. Mélanges offerts à Denis Creissels,*, pp. 383–402. Lyon: ENS Éditions.

Guirardello, R. (1999). *A reference grammar of Trumai.* Doctoral dissertation, Rice University.

Hale, A. (1980). Person markers: Finite conjunct and disjunct verb forms in Newari. In R. Trail (Ed.), *Papers in South-East Asian linguistics No.7*, pp. 95–106. Canberra: Pacific Linguistics.

Hardman, M. J. (2000). *Jaqaru.* München: Lincom Europe.

Hardman, M. J. (2001). *Aymara.* München: Lincom Europe.

Harris, A. C. and L. Campbell (1995). *Historical Syntax in a Cross-linguistic Perspective.* Cambridge: Cambridge University Press.

Haspelmath, M. (1987). *Transitivity alternations of the anticausative type.* Köln: Institut für Sprachwissenschaft der Universität zu Köln.

Haspelmath, M. (2009). Terminology of Case. In A. L. Malchukov and A. Spencer (Eds.), *The Oxford Handbook of Case*, pp. 505–517. Oxford: Oxford University Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in cross-linguistic studies. *Language 86*(3), 663–687.

Haspelmath, M. (2011a). On S, A, P, T, and R as comparative concepts in alignment typology. *Linguistic Typology 15*, 535–567.

Haspelmath, M. (2011b). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica 45*(1), 31–80.

Haspelmath, M. (2012). How to compare major word-classes across the world's languages. In T. Graf, D. Paperno, A. Szabolcsi, and J. Tellings (Eds.), *Theories of everything: in honor of Edward Keenan*, Volume UCLA Worki, pp. 109–130. Los Angeles: UCLA.

Haspelmath, M. (2013). Argument indexing: A conceptual framework for the syntactic status of bound person forms. In D. Bakker and M. Haspelmath (Eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, pp. 209–238. Berlin: De Gruyter Mouton.

Haude, K. (2006). *A grammar of Movima*. Doctoral dissertation, Radboud Universiteit Nijmegen.

Haude, K. (2009). Hierarchical Alignment in Movima. *International Journal of American Linguistics 75*(4), 513–532.

Haude, K. (2011). Argument encoding in Movima: The local domain. *International Journal of American Linguistics 77*(4), 559–575.

Hawkins, J. (1983). *Word Order Universals*. New York: Academic Press.

Heath, J. (1984). Language contact and language change. *Annual Review of Anthropology 13*, 367–384.

Heckenberger, M. J. and E. G. Neves (2009). Amazonian archaeology. *Annual Review of Anthropology 38*, 251–266.

Heckenberger, M. J., J. C. Russell, C. Fausto, J. R. Toney, M. J. Schmidt, E. Pereira, B. Franchetto, and A. Kuikuro (2008). Pre-Columbian urbanism, anthropogenic landscapes, and the future of the Amazon. *Science 321*(5893), 1214–1217.

Heckenberger, M. J., M. J. Schmidt, J. R. Toney, and C. Russell (2011). Cidades jardins do Alto Xingu no início do Antropoceno. *Paper presented at the 7th international meeting of the Society for the Anthropology of Lowland South America*.

Heggarty, P. and D. Beresford-Jones (Eds.) (2012). *Archaeology and Language in the Andes*. Oxford: Oxford University Press.

Hockett, C. (1958). *A Course in Modern Linguistics*. New York: Macmillan.

Holman, E. W., S. r. Wichmann, C. H. Brown, V. Velupillai, and A. Müller (2008). Explorations in automated language classification. *Folia Linguistica 42*(2), 331–354.

Hopper, P. and S. A. Thompson (1980). Transitivity in grammar and discourse. *Language 56*(2), 251–299.

Hornborg, A. (2005). Ethnogenesis, regional integration, and ecology in prehistoric Amazonia: Towards a systems approach. *Current Anthropology 46*(4), 589–620.

Jackendoff, R. (1976). Towards an explanatory semantic representation. *Linguistic Inquiry 7*(1), 89–150.

Janssen, D. P., B. Bickel, and F. Zúñiga (2006). Randomization tests in language typology. *Linguistic Typology 10*, 419–440.

Jelinek, E. (1984). Empty categories and non-configurational languages. *Natural Language and Linguistic Theory 2*, 39–76.

Jensen, C. (1990). Cross-referencing changes in some Tupí-Guaraní languages. In D. L. Payne (Ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*, pp. 117–158. Austin: University of Texas Press.

Jensen, C. (1998). Comparative Tupí-Guaraní morphosyntax. In D. C. Derbyshire and G. K. Pullum (Eds.), *Handbook of Amazonian Languages*, Volume IV, pp. 487–618. Berlin: Mouton de Gruyter.

Jensen, C. (1999). Tupi-Guarani. In R. M. W. Dixon and A. Aikhenvald (Eds.), *The Amazonian Languages*, Cambridge Language Surveys, pp. 125–164. Cambridge: Cambridge University Press.

Jolkesky, M. P. D. V. (2010). Reconstrução fonológica e lexical do Proto-Jê Meridional. Master's thesis, Universidade Estadual de Campinas.

Jolkesky, M. P. V. (2009). Macro-Daha: reconstrução de um tronco lingüístico do noroeste amazônico. Paper presented at the I Congresso Internacional de Lingü\'istica Histórica (ROSAE), 26-29 de Julho de 2009.

Jung, I. (2008). *Gramática del Páez o Nasa Yuwe: Descripción de una lengua indígena de Colombia*. München: Lincom Europe.

Kaufman, T. (1990). Language history in South America: What we know and how to know more. In D. L. Payne (Ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*, Texas Linguistics Series, pp. 13–73. Austin: University of Texas Press.

Keenan, E. (1976). Towards a universal denition of 'subject'. In C. Li (Ed.), *Subject and Topic*, pp. 305–333. New York: Academic Press.

Keenan, E. L. and B. Comrie (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry 8*, 63–99.

Keenan, E. L. and M. S. Dryer (2007). Passive in the worlds languages. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Vol. I: Clause Structure*, pp. 325–361. Cambridge: Cambridge University Press.

Kemmer, S. (1993). *The Middle Voice*. Amsterdam: John Benjamins Publishing Company.

Key, M. R. (1968). *Comparative Tacanan phonology: With Cavineña phonology and notes on Pano-Tacanan relationship*, Volume 50 of *Janua Linguarum. Series Práctica*. Den Haag: Mouton.

Kibort, A. and G. C. Corbett (2008). Number.

Klimov, G. A. (1974). On the character of languages of active typology. *Linguistics 131*, 11–25.

Koehn, E. and S. S. Koehn (1986). Apalai. In D. C. Derbyshire and G. K. Pullum (Eds.), *Handbook of Amazonian Languages*, Volume I, pp. 33–127. Berlin: Mouton de Gruyter.

Krasnoukhova, O. (2012). *The noun phrase in the languages of South America*. Doctoral dissertation, Radboud Universiteit Nijmegen.

Lafone Quevedo, S. A. (1896). Lenguas Argentinas: Grupo Mataco-Mataguayo: Dialecto Vejoz. *Boletín del Instituto Geográfico Argentino XVII*, 121–176.

Lakoff, G. (1977). Linguistic gestalts. In W. Beach, S. Fox, and S. Philosoph (Eds.), *Papers from the 13th Regional Meeting of the Chicago Linguistic Society*, pp. 236–287. Chicago: Chicago Linguistic Society.

Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Langacker, R. W. and P. Munro (1975). Passives and their meaning. *Language 51*(4), pp. 789–830.

Lathrap, D. W. (1970). *The Upper Amazon*. London: Thames and Hudson.

Lazard, G. (1998). *Actancy*. Berlin: Mouton de Gruyter.

Lazard, G. (2002). Transitivity revisted as an example of a more strict approach in typological research. *Folia Linguistica 36*, 141–190.

Leite, Y. (1989). As cisões no sistema referencial de pessoa em Tapirapé.

Lévi-Strauss, C. (1948). Tribes of the right bank of the Guaporé River. In J. H. Steward (Ed.), *The Tropical Forest Tribes*, Volume 3 of *Handbook of South American Indians*, pp. 371–379. Washington: Smithsonian Institution, Washington: Bureau of American Ethnology.

Levin, B. (1993). *English Verb Classes and Alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Levin, B. and M. Rappaport Hovav (1995). *Unaccusativity: At the syntax-lexical semantics interface.* Cambridge, MA: MIT press.

Lewis, M. P. (Ed.) (2009). *Ethnologue: Languages of the world* (16 ed.). Dallas: SIL International.

Lima, S. O. d. (2008). A estrutura argumental dos verbos na língua Juruna (Yudja). Master's thesis, Universidade de São Paulo.

Loukotka, v. (1968). *Classification of the South American Indian Languages.* Los Angeles: Latin American Center, University of California.

Mace, R. and M. Pagel (1994). The comparative method in anthropology. *Current Anthropology 35*(5), 549–564.

Maddison, W. P. and D. R. Maddison (2011). Mesquite: A Modular System for Evolutionary Analysis. Version 2.75.

Malchukov, A., M. Haspelmath, and B. Comrie (2010). Ditransitive constructions: A typological overview. In A. Malchukov, M. Haspelmath, and B. Comrie (Eds.), *Studies in Ditransitive Constructions: A comparative handbook*, pp. 1–64. Berlin: De Gruyter Mouton.

Martins, S. A. (2004). *Fonologia e gramática Dâw.* Doctoral dissertation, Vrije Universiteit Amsterdam.

Maslova, E. and V. P. Nedjalkov (2011). Reciprocal constructions. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* München: Max Planck Digital Library.

McGregor, W. (1989). Greenberg on the first person inclusive dual: Evidence from some Australian languages. *Studies in Language 13*, 437–458.

McMichael, C. H., D. R. Piperno, M. B. Bush, M. R. Silman, A. R. Zimmerman, M. F. Raczka, and L. C. Lobato (2012). Sparse Pre-Columbian human habitation in Western Amazonia. *Science 336*(6087), 1429–1431.

Meira, S. (1999). *A grammar of Tiriyó.* Doctoral dissertation, Rice University.

Meira, S. (2000a). *A reconstruction of Proto-Taranoan: phonology and morphology.* München: Lincom Europe.

Meira, S. (2000b). The accidental intransitive split in the Cariban family. In S. Gildea (Ed.), *Reconstructing Grammar: Comparative linguistics and grammaticalization*, pp. 201–230. Amsterdam: John Benjamins Publishing Company.

Meira, S. and S. Drude (2013). Sobre a origem histórica dos "prefixos relacionais" das línguas Tupí-Guaraní. *Cadernos de Etnolingüística 5*(1), 1–30.

Meireles, D. M. (1989). *Guardiães da Fronteira: Rio Guaporé, século XVIII*. Petrópolis: Vozes.

Meireles, D. M. (1991). O complexo cultural do marico: Sociedades indígenas dos ríos Branco, Colorado e Mequens, afluentes do médio Guaporé. *Boletim do Museu Paraense Emílio Goeldi, Série Antropologia 7*(2), 209–269.

Mel'cuk, I. (1994). The inflectional category of voice: Towards a more rigorous denition. In B. Comrie and M. Polinsky (Eds.), *Causatives and Transitivity*, pp. 1–46. Amsterdam: John Benjamins Publishing Company.

Merlan, F. (1985). Split intransitivity: functional oppositions in intransitive inflection. In J. Nichols and A. Woodbury (Eds.), *Grammar Inside and Outside the Clause*, pp. 324–362. Cambridge: Cambridge University Press.

Metraux, A. (1928). *La Civilisation Matérielle des Tribus Tupi-Guarani*. Paris: Paul Geuthner.

Métraux, A. (1948). The Paressi. In J. H. Steward (Ed.), *The Tropical Forest Tribes*, Volume 3 of *Handbook of South American Indians*, pp. 349–360. Smithsonian Institution, Washington: Bureau of American Ethnology.

Migliazza, E. C. (1985). Languages of the Orinoco-Amazon region: Current status. In H. E. Manelis Klein and L. Stark (Eds.), *South American Indian Languages: Retrospect and Prospect*, pp. 17–139. Austin: Texas University Press.

Miller, E. (1992). Arqueologia nos Empreendimentos Hidrelétricos da Elertonorte. Resultados Preliminares. Technical report, Eletronorte, Brasilia.

Miller, M. (1999). *Desano Grammar*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Mithun, M. (1991). Active/agentive case marking and its motivations. *Language 67*(3), 510–546.

Monserrat, R. M. F. (1976). Prefixos pessoais em Aweti. *Publicações do Museu Nacional, Série Lingüística III*.

Monserrat, R. M. F. and M. F. Soares (1983). Hierarquia referencial em línguas Tupi. *Ensaios de Lingüística 9*, 164–187.

Moore, D. (1984). *Syntax of the language of the Gavião Indians of Rondônia, Brazil*. Doctoral dissertation, City University of New York.

Morse, N. L. and M. B. Maxwell (1999). *Cubeo Grammar*. Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Mortensen, C. A. (1999). *A Reference Grammar of Northern Embera Languages*. Dallas: SIL International and University of Texas at Arlington.

Mueller, N. (2013). *Tense, aspect, modality and evidentiality marking in South American indigenous languages.* Doctoral thesis, Radboud Universiteit Nijmegen.

Murdock, G. P. (1951). South American culture areas. *Southwestern Journal of Anthropology 7*(4), 415–436.

Muysken, P. (1977). *Syntactic developments in the verb phrase of Ecuadorian Quechua.* Lisse: Peter de Ridder Press.

Muysken, P. (1986). A note on passive-like statives in Quechua. *Revue Québécoise de Linguistique 15*(2), 109–118.

Muysken, P. (Ed.) (2008). *From Lingusitic Areas to Areal Linguistics.* Amsterdam: John Benjamins Publishing Company.

Neves, E. G. (1998). *Paths in dark waters: Archaeology as indigenous history in the Upper Rio Negro Basin, Northwest Amazon.* Doctoral dissertation, Indiana University.

Neves, E. G. (2006). *A Arqueologia da Amazônia.* Rio de Janeiro: Jorge Zahar.

Neves, E. G. (2008). Ecology, ceramic chronology and distribution, long-term history and political chance in the Amazonian floodplain. In H. Silverman and W. Isbell (Eds.), *The Handbook of South American Archaeology*, pp. 359–379. New York: Springer.

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language 62*(1), 56–119.

Nichols, J. (1990). Linguistic diversity and the first settlement of the New World. *Language 66*(3), 475–521.

Nichols, J. (1992). *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.

Nimuendajú, C. (1982). *Textos Indigenistas.* São Paulo: Edições Loyola.

Noelli, F. S. (1996). As hipóteses sobre o centro de origem e rotas de expansão dos Tupi. *Revista de Antropologia 39*, 7–53.

Nogueira, A. F. (2011). Wayoro emeto: Fonología segmental e morfossintaxe verbal. Master's thesis, Universidade de São Paulo.

Nonato, R. and F. Sandalo (2007). Uma comparação gramatical, fonológica e lexical entre as famílias Guaikurú, Mataco e Bororo: Um caso de difusão areal? *Boletim do Museu Paraense Emílio Goeldi: Ciências Humanas 2*(2), 91–107.

Nonato, R. B. (2008). Ainore Boe egore: Um estudo descritivo da língua Bororo. Master's thesis, Universidade Estadual de Campinas.

Nordenskiöld, E. (1920). *The Changes in the Material Culture of Two Indian tribes under the Influence of New Surroundings*, Volume 2 of *Comparative Ethnographical Studies*. Göteborg: Elanders.

Nordenskiöld, E. (1924). *The ethnography of South-America seen from Mojos in Bolivia*. Göteborg: Elanders.

O'Connor, L. (2014). Structural features and language contact in the Isthmo-Colombian Area. In L. O'Connor and P. C. Muysken (Eds.), *The Native Languages of South America: Origins, development, typology*, pp. 73–101. Cambridge: Cambridge University Press.

O'Connor, L. and P. C. Muysken (Eds.) (2014). *The Native Languages of South America: Origins, development, typology*. Cambridge: Cambridge University Press.

Olawsky, K. (2006). *A Grammar of Urarina*. Berlin: Mouton de Gruyter.

Orr, C. and R. E. Longacre (1968). Proto-Quechumaran. *Language 44*(3), pp. 528–555.

Overall, S. (2007). *A grammar of Aguaruna*. Doctoral dissertation, LaTrobe University.

Pachêco, F. B. (2001). *Morfossintaxe do verbo Ikpeng (Karíb)*. Doctoral dissertation, Universidade Estadual de Campinas.

Palácio, A. P. (1984). *Guató: A língua dos índios Canoeiros do Rio Paraguai*. Doctoral dissertation, Universidade Estadual de Campinas.

Palácio, A. P. (2004). Alguns aspectos da língua Guató. *LIAMES 4*, 163–170.

Payne, D. (1990). Morphological characteristics of lowland South American languages. In D. Payne (Ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*, pp. 213–242. Austin: University of Texas Press.

Payne, D. (1994). The Tupí-Guaraní inverse. In B. Fox and P. J. Hopper (Eds.), *Voice: Form and function*, pp. 313–340. Amsterdam: John Benjamins Publishing Company.

Payne, D. L. (2001). Review of Amazonian Languages, ed. by Alexandra Aikhenvald and R. M. W. Dixon. *Language 77*, 594–598.

Pearsall, D. (2008). Plant domestication and the shift to agriculture in the Andes. In H. Silverman and W. Isbell (Eds.), *The Handbook of South American Archaeology*, pp. 105–121. New York: Springer.

Perlmutter, D. (1978). Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 6th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, pp. 157–189. University of California.

Pet, W. (2011). *A Grammar Sketch and Lexicon of Arawak (Lokono Dian)* (SIL e-Book ed.). SIL International.

Peterson, D. A. (2007). *Applicative constructions*. Oxford: Oxford University Press.

Piperno, D. and D. Pearsall (1998). *The Origins of Agriculture in the Lowland Neotropics*. San Diego: Academic Press.

Polinsky, M. (2011a). Antipassive constructions. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Polinsky, M. (2011b). Applicative constructions. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Primus, B. (1999). *Cases and Thematic Roles: Ergative, accusative and active*. Tübingen: Niemeyer.

Queixalós, F. (2010). Grammatical relations in Katukina-Kanamari. In S. Gildea and F. Queixalós (Eds.), *Ergativity in Amazonia*, pp. 235–283. Amsterdam: John Benjamins Publishing Company.

Ramirez, H. (1994). *Le parler Yanomami des Xamatauteri*. Ph. D. thesis, Aix-en-Provence: Université de Provence.

Regúnaga, M. A. (2012). *Tipología del género en lenguas indígenas de América del Sur*. Doctoral dissertation, Universidad Nacional del Sur.

Ribeiro, E. R. (2010, October). Nimuendajú was right: The inclusion of the jabutí language family in the Macro-Jê stock1. *International Journal of American Linguistics 76*(4), 517–570.

Rodrigues, A. D. (1964). A classificação do tronco lingüístico Tupí. *Revista de Antropologia 12*, 99–104.

Rodrigues, A. D. (1985a). Evidence for Tupi-Carib relationships. In H. E. Manelis Klein and L. Stark (Eds.), *South American Indian Languages: Retrospect and prospect*, pp. 371–404. Austin: Texas University Press.

Rodrigues, A. D. (1985b). Relações internas na família lingüística Tupi-Guarani. *Revista de Antropologia 27/28*, 33–53.

Rodrigues, A. D. (1999). Tupi. In R. M. W. Dixon and A. Aikhenvald (Eds.), *The Amazonian Languages*, Cambridge Language Surveys, pp. 107–124. Cambridge: Cambridge University Press.

Rodrigues, A. D. (2007). As consoantes do Proto-Tupí. In A. S. A. C. Cabral and A. D. Rodrigues (Eds.), *Línguas e Culturas Tupí*, pp. 167–203. Campinas: Curt Nimuendaju.

Rodrigues, A. D. and A. S. A. C. Cabral (2012). Tupían. In L. Campbell and V. Grondona (Eds.), *The Indigenous Languages of South America: A comprehensive guide*, pp. 495–574. Berlin: De Gruyter Mouton.

Rodrigues, A. D. and W. Dietrich (1997). On the linguistic relationship between Mawé and Tupí-Guaraní. *Diachronica 14*(2), 265–304.

Rodrigues, C. L. R. (1995). *Étude morphosyntaxique de la langue Xipaya (Brésil)*. Doctoral dissertation, Université Paris VII.

Rojas Berscia, L. M. (2013). *La sintaxis y semántica de las construcciones causativas en el chayahuita de Balsapuerto*. Master's thesis, Pontificia Universidad Católica del Perú.

Romero-Figeroa, A. (1997). *A Reference Grammar of Warao*. München: Lincom Europe.

Roosevelt, A. C. (1991). *Moundbuilders of the Amazon: Geophysical archaeology on Marajó Island, Brazil*. San Diego: Academic Press.

Roosevelt, A. C. (1994). Amazonian anthropology: Strategy for a new synthesis. In A. C. Roosevelt (Ed.), *Amazonian Indians from Prehistory to the Present*, pp. 1–29. Tucson, AZ: University of Arizona Press.

Rose, F. (2003). *Morphosyntaxe de l'Emerillon: Langue Tupi-Guarani de Guyane Française*. Doctoral dissertation, Université Lumière Lyon 2.

Rowan, O. and E. Burgess (2008). *Gramática Parecis*. Anápolis, GO: Summer Institute of Linguistics.

Saitou, N. and M. Nei (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution 4*(4), 406–425.

Sakel, J. (2004). *A Grammar of Mosetén*. Berlin: Mouton de Gruyter.

Sakel, J. (2007). Types of loans: Matter and pattern. In Y. Matras and J. Sakel (Eds.), *Grammatical Borrowing in Cross-linguistic Perspective*, pp. 15–30. Berlin: Walter de Gruyter.

Sakel, J. (2011). Moseten and Chimane argument coding : A layered system. *International Journal of American Linguistics 77*(4), 537–557.

Sans, P. (2012). Does Bésiro have a "quadripartite" alignment? *Paper presented at the Society for the Study of the Indigenous Languages of the Americas annual meeting.*

Santo Tomás, D. d. (1560). *Grammatica o Arte de la Lengua General de los Indios de los Reynos del Peru.* Valladolid: Francisco Fernández de Córdova.

Sapir, E. (1917). Review of Het Passieve Karakter van het Verbum Transitivum of van het Verbum Actionis in Talen van Noord-Amerika (The Passive Character of the Transitive Verb or of the Active Verb in Languages of North America) by C. C. Uhlenbeck. *International Journal of American Linguistics 1*(1), 82–86.

Schachter, P. and T. Shopen (2007). Parts-of-speech systems. In T. Shopen (Ed.), *Language Typology and Syntactic Description, Volume 1: Clause Structure*, pp. 1–60. Cambridge: Cambridge University Press.

Schleicher, C. O. (1998). *Comparative and internal reconstruction of the Tupi-Guarani language family.* Doctoral dissertation, University of Wisconsin.

Seifart, F. (2005). *The structure and use of shape-based noun classes in Miraña (North West Amazon).* Doctoral dissertation, Radboud Universiteit Nijmegen.

Seifart, F. (2014). Valency classes in Bora. In A. Malchukov and B. Comrie (Eds.), *Valency Classes: A Comparative Handbook*, pp. to appear. Berlin: De Gruyter Mouton.

Seki, L. (1990). Kamaiurá (Tupí-Guaraní) as an active-stative language. In D. L. Payne (Ed.), *Amazonian Linguistics: Studies in Lowland South American Languages*, pp. 367–391. Austin: University of Texas Press.

Seki, L. (1999). The Upper Xingu as an incipient linguistic area. In R. M. W. Dixon and A. Y. Aikhenvald (Eds.), *The Amazonian Languages*, pp. 417–430. Cambridge: Cambridge University Press.

Seki, L. (2000). *Gramática Kamaiurá: Língua Tupi-Guarani do Alto Xingu.* Campinas: Editora Unicamp.

Sherzer, J. (1973). Areal linguistics in North America. In T. A. Sebeok (Ed.), *Linguistics in North America*, Volume 10 of *Current Trends in Linguistics*, pp. 749–795. Den Haag: Mouton.

Sherzer, J. (1976). *An Areal-typological Study of American Indian languages North of Mexico.* Amsterdam: North Holland Publishing Co.

Shibatani, M. (1976). The grammar of causative constructions: a conspectus. In M. Shibatani (Ed.), *Syntax and Semantics: Vol. VI, The grammar of causative constructions*, pp. 1–40. New York: Academic Press.

Shibatani, M. (2004). Voice. In G. Booij, C. Lehmann, and J. Mugdan (Eds.), *Morphology: A handbook on inflection and word formation*, Volume 2, pp. 1145–1165. Berlin: Mouton de Gruyter.

Shibatani, M. and P. Pardeshi (2001). The causative continuum. In M. Shibatani (Ed.), *The Grammar of Causation and Interpersonal Manipulation*, pp. 85–126. Amsterdam: John Benjamins Publishing Company.

Siewierska, A. (1999). From anaphoric pronoun to grammatical agreement marker: Why objects don't make it. *Folia linguistica 33*(2), 225–251.

Siewierska, A. (2003). Person agreement and the determination of alignment. *Transactions of the Philological Society 101*(2), 339–370.

Siewierska, A. (2004). *Person*. Cambridge: Cambridge University Press.

Siewierska, A. (2011a). Alignment of verbal person marking. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Siewierska, A. (2011b). Passive constructions. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Siewierska, A. (2011c). Third person zero of verbal person marking. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structure Online*. München: Max Planck Digital Library.

Silva, L. d. J. (2011). *Morphosyntaxe du Rikbaktsa*. Doctoral dissertation, Université Denis Diderot – Paris 7.

Silva, R. G. P. d. (2010). *Estudo morfossintático da língua Sateré-Mawé*. Doctoral dissertation, Universidade Estadual de Campinas.

Silva, W. d. L. (2012). *A descriptive grammar of Desano*. Doctoral dissertation, University of Utah.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages*, pp. 112–171. Atlantic Highlands, NJ: Humanities Press.

Smeets, I. (2008). *A Grammar of Mapuche*. Berlin: Mouton de Gruyter.

Song, J. J. (2011). Nonperiphrastic causative constructions. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.

Stenzel, K. (2008). Kotiria 'differential object marking' in cross-linguistic perspective. *Amerindia 32*, 153–182.

Steward, J. H. (1948). Tribes of the Montaña: An introduction. In J. H. Steward (Ed.), *The Tropical Forest Tribes*, Volume 3 of *Handbook of South American Indians*, pp. 507–533. Washington: Smithsonian Institution, Washington: Bureau of American Ethnology.

Talmy, L. A. (1976). Semantic causative types. In M. Shibatani (Ed.), *Syntax and Semantics: Vol. VI, The grammar of causative constructions*, pp. 43–116. New York: Academic Press.

Telles, S. (2002). *Fonologia e gramática Latundê/Lakondê*. Doctoral dissertation, Vrije Universiteit Amsterdam.

Terraza, J. (2009). *Grammaire du Wichí: Phonologie et morphosyntaxe*. Doctoral dissertation, Université du Québec à Montréal.

Terrill, A. (1997). The development of antipassive constructions in Australian languages. *Australian Journal of Linguistics 17*(1), 71–88.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.

Thomason, S. G. (2001). *Language Contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, S. G. and T. S. Kaufman (1988). *Language Contact, Creolization and Genetic Linguistics*. Berkeley and Los Angeles: University of California Press.

Torero, A. (1970). Lingüística e historia de la Sociedad Andina. *Anales Científicos de la Universidad Agraria 8*(3-4), 231–264.

Torero, A. (2002). *Idiomas de los Andes: Lingüística e Historia*. Lima: Editorial Horizonte.

Tovar, A. and C. L. Tovar (1984). *Catálogo de las Lenguas de América del Sur: Con clasificaciones, indicaciones tipológicas, bibliografía y mapas*. Madrid: Gredos.

Trillos Amaya, M. (1996). *Categorias Grammatical del Ette Taara: Lengua de los chimilas*, Volume 10 of *Lenguas Aborígenes de Colombia: Descripciones*. Bogotá: CESO-CCELA, Universidad de los Andes.

Valenzuela, P. (2000). Ergatividad escindida en Waripano, Yaminawa y Shipibo-Konibo. In H. van der Voort and S. van de Kerke (Eds.), *Indigenous Languages of South America*, pp. 111–128. Leiden: Research School of Asian, African, and Amerindian Studies (CNWS), Leiden University.

Valenzuela, P. (2004). El sistema de marcación de caso ergativo-escindido en wariapano: análisis sincrónico y comparativo. In Z. E. Fernández, A. Fernández Garay, and A. Álvarez González (Eds.), *Estudios en Lenguas Amerindias: Homenaje a Ken L. Hale*, pp. 309–369. Hermosillo, Mexico: Universidad de Sonora.

Valenzuela, P. M. (1997). Basic verb types and argument structures in Shipibo-Conibo. Master's thesis, University of Oregon.

Valenzuela, P. M. (2003). *Transitivity in Shipibo-Konibo grammar*. Doctoral dissertation, University of Oregon.

Vallejos Yopán, R. (2010). *A grammar of Kokama-Kokamilla*. Doctoral dissertation, University of Oregon.

van de Kerke, S. (2009). El Leko. In M. Crevels and P. Muysken (Eds.), *Lenguas de Bolivia. Tomo I: Ámbito Andino*, Volume 1 of *Lenguas de Bolivia*, pp. 287–332. La Paz: Plural editores.

van de Kerke, S. and P. C. Muysken (2014). The Andean matrix. In L. O'Connor and P. C. Muysken (Eds.), *The Native Languages of South America: Origins, Development, Typologynguages of South*, pp. 126–151. Cambridge: Cambridge University Press.

van der Meer, T. (1985). Case marking in Suruí. In D. L. Fortune (Ed.), *Porto Velho Workpapers*, pp. 208–230. Brasília: Summer Institute of Linguistics.

van der Voort, H. (2004). *A Grammar of Kwaza*. Berlin: Mouton de Gruyter.

van Gijn, E. (2006). *A grammar of Yurakaré*. Doctoral dissertation, Radboud Universiteit Nijmegen.

van Gijn, R. (2010). Middle voice and ideophones, a diachronic connection: The case of Yurakaré. *Studies in Language 34* (2), 273–297.

Van Valin Jr., R. D. (1980). On the distribution of passive and antipassive constructions in universal grammar. *Lingua 50*, 303–332.

Van Valin Jr., R. D. and R. J. LaPolla (1997). *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.

Vidal, A. (2001). *Pilagá grammar (Guaykuruan Family, Argentina)*. Doctoral dissertation, University of Oregon.

Vidal, A. (2008). Affectedness and viewpoint in Pilagá (Guaykuruan): a semantically aligned case-marking system. In M. Donohue and S. Wichmann (Eds.), *The Typology of Semantic Alignment*, pp. 412–430. Oxford: Oxford University Press.

Viegas Barros, P. J. (2005a). Algunas semejanzas gramaticales macro-Guaicurú – macro-Jê. Ms.

Viegas Barros, P. J. (2005b). *Voces en el Viento: Raíces lingüísticas de la Patagonia*. Buenos Aires: Ediciones Mondragon.

Walker, R. S. and L. A. Ribeiro (2011). Bayesian phylogeography of the Arawak expansion in lowland South America. *Proceedings of the Royal Society of London, Series B January 19*, 1–7.

Walker, R. S., S. Wichmann, T. Mailund, and C. J. Atkisson (2012). Cultural phylogenetics of the Tupi language family in lowland South America. *PLoS ONE 7*(4).

Weber, D. J. (1983). *A grammar of Huallaga (Huánuco) Quechua*. Doctoral dissertation, University of California at Los Angeles.

Weber, D. J. (1989). *A Grammar of Huallaga (Huánuco) Quechua*. Berkeley and Los Angeles: University of California Press.

Weber, D. J. (1996). *Una Gramática del Quechua del Huallaga (Huánuco)*. Lima: Ministerio de Educación and Instituto Lingüístico de Verano.

Whitehead, N. L. (2003). *History and Historicities in Amazonia*. Lincoln: University of Nebraska Press.

Wilkins, D. P. (1996). Morphology. In H. Goebl, P. H. Nelde, Z. Starý, and W. Wölck (Eds.), *Contact Linguistics: An International Handbook of Contemporary Research*, pp. 109–117. Berlin: Walter de Gruyter.

Willey, G. R. (1971). *An Introduction to American Archaeology, Vol. 2: South America*. Englewood Cliffs, NJ: Prentice Hall.

Winford, D. (2005). Contact-induced changes: Classification and processes. *Diachronica 22*(2), 373–427.

Wise, M. R. (1986). Grammatical characteristics of PreAndine Arawakan languages of Peru. In D. C. Derbyshire and G. K. Pullum (Eds.), *Handbook of Amazonian Languages*, Volume I, pp. 567–642. Berlin: Mouton de Gruyter.

Wise, M. R. (1990). Valence-changing affixes in Maipuran Arawakan languages. In D. Payne (Ed.), *Amazonian Linguistics. Studies in Lowland South American Languages*, pp. 89–116. Austin: University of Texas Press.

Witzlack-Makarevich, A. (2010). *Typological variation in grammatical relations*. Doctoral dissertation, Universität Leipzig.

Witzlack-Makarevich, A. and B. Bickel (2013). Handbook of Grammatical Relations Questionnaire.

Zariquiey Biondi, R. (2011). *A grammar of Kashibo-Kakataibo*. Doctoral dissertation, LaTrobe University.

Zúñiga, F. (2006a). *Deixis and Alignment: Inverse systems in indigenous languages of the Americas*. Amsterdam: John Benjamins Publishing Company.

Zúñiga, F. (2006b). *Mapudungun*. München: Lincom Europe.

Zúñiga, F. (2007). The discourse-syntax interface in northwestern Amazonia: Differential object marking in Makú and some Tucanoan languages. In *Language Endangerment and Endangered Languages: Linguistic and Anthropological Studies with Special Emphasis on the Languages and Cultures of the Andean-Amazonian Border Area*, pp. 209–227. Leiden: Publications of the Research School of Asian, African and Amerindian Studies (CNWS), University of Leiden.

Zúñiga, F. (2012). Animacy, inversion, and redundancy in Amerindian languages: Elements for a crosslinguistic survey. *Paper presented at the 54th International Congress of Americanists, Vienna*.

Zwicky, A. (1985). Clitics and particles. *Language 61*(2), 283–305.

Samenvatting in het Nederlands

Dit proefschrift begint met een introductie van de ingewikkelde puzzel van de Zuid-Amerikaanse talen waarmee wetenschappers die zich bezighouden met de taaldiversiteit van de Nieuwe Wereld en de reikwijdte van mogelijke structurele variatie in taal zich geconfronteerd zien. Het stelt dat een gedetailleerde studie van de verschillende coderingsstrategieën voor participanten in taaluitingen de reikwijdte van structurele variatie in de talen van dit werelddeel naar voren kan brengen, vooral nu zoveel nieuwe informatie over de inheemse talen gedurende de laatste jaren beschikbaar is gekomen. De typologische en areale patronen die deze gegevens bevatten kunnen ons helpen de oorsprong van de taaldiversiteit beter te begrijpen en kunnen ons inzicht verschaffen in de complexe relaties tussen de populaties die deze talen spraken.

Het hier gepresenteerde werk vertegenwoordigt het eerste uitgebreide onderzoek van de strategieën waarmee de argumenten van enkelvoudige hoofdzinsconstructies in Zuid-Amerikaanse talen worden gemarkeerd. Het maakt gebruik van een geografisch en genealogisch representatief sample van 74 Zuid-Amerikaanse talen, dat in totaal 40 verschillende taalfamilies en isolaten vertegenwoordigt. Vanwege de taalspecifieke aard van vele taalkundige categorieën, maakt deze studie gebruik van semantisch gedefiniëerde vergelijkingsconcepten, die als een onafhankelijke maatstaf kunnen dienen voor de vergelijking van verschillende morfosyntactische structuren. De typologische benadering van de taalkundige gegevens berust op de identificatie van verschillende semantische participantenrollen en de manier waarop deze participanten in de morfosyntaxis verschillende argumentenrollen uitdrukken. In deze studie gaat de meeste aandacht uit naar de codering van de argumentenrollen door verschillende markeerders, zoals persoons- en naamvalsmarkering, en hoe deze argumentenrollen samenwerken om grammaticale relaties te vormen in deze verschillende grammaticale domeinen.

Hoofdstuk 1 schetst de doelstellingen van het proefschrift en plaatst deze studie in de context van de Zuid-Amerikaanse geschiedenis, geografie en etnolinguïstische diversiteit. De geselecteerde talen worden gepresenteerd en verdeeld

in geografische gebieden die reeds eerder voorgestelde culturele en taalkundige area's omvatten of ermee overeenkomen, waarbij elk gebied wordt bepaald door geografische kenmerken, zoals bergketens, grote rivieren en ecologische zones.

Hoofdstuk 2 verklaart de methodologie die in dit proefschrift wordt gebruikt, door de belangrijkste vergelijkingsconcepten te definiëren en de criteria vast te stellen waarmee het bereik van vergelijking binnen het zeer diverse taalsample wordt bepaald. Ook de structurele questionnaire en de coderingsprocedure worden hier gepresenteerd.

Hoofdstuk 3 richt zich op de reikwijdte van variatie in patronen van verbale argumentmarkering, in het bijzonder met betrekking tot persoonsmarkering. Verschillende kenmerken, zoals de aanwezigheid, positie, alignment en fusie van argumenten binnen verschillende groepen markeerders, en hoe deze verschillende groepen markeerders op het predikaat samengaan om deze argumenten uit te drukken, worden alle behandeld als belangrijke componenten van het markeringspatroon.

Hoofdstuk 4 over complexe verbale markeringspatronen onderzoekt twee gebieden van verbale argumentmarkering die in de typologische literatuur vaak worden behandeld als verschillende alignmenttypen: hiërarchische markeringspatronen en gespleten intransitiviteit. In dit hoofdstuk wordt gesteld dat deze grammaticale patronen door de beheersing van enkele semantische voorwaarden eenvoudig kunnen worden opgedeeld in een aantal verschillende structurele parameters en alignmentpatronen, waarmee tevens een meer omvattende vergelijking mogelijk wordt gemaakt tussen deze en de andere talen van het continent met verschillende indexatiepatronen.

Hoofdstuk 5 sluit het onderzoek naar verbale argumentmarkering af met een verkenning van de ontwikkeling van het persoonsmarkeringssysteem binnen de grote en structureel diverse Tupi taalfamilie. Er wordt gesteld dat de historische ontwikkeling van zulke verschillende patronen kan worden begrepen door de analyse van hun synchrone distributie in het licht van de huidige inzichten in de interne classificatie van taalfamilies. Dit hoofdstuk past een reconstructiemodel toe op de topologie (fylogenetische vertakkingspatronen) van twee verschillende classificaties en vergelijkt de resultaten met hetgeen bekend is uit traditioneel historisch taalkundig werk. Deze casestudy ondersteunt eerdere beweringen over de Tupi familie en brengt tevens een aantal nieuwe hypothesen voort over de tussenliggende ontwikkelingsfasen van deze patronen.

Hoofdstuk 6 onderzoekt de verschillende patronen van naamvalsmarkering van nominale en pronominale argumenten in de talen van het sample. Om maximale vergelijkbaarheid te garanderen wordt een aantal criteria vastgesteld om de pragmatische en referentiële effecten te kunnen beheersen die verschillende argumenten op de uitdrukking van naamval kunnen hebben. Bijzondere aandacht wordt ook besteed aan de naamvalsmarkering van niet-subject argumenten in ditransitieve constructies, en ook aan de effecten van animaatheid op de uitdrukking van naamval in transitieve constructies.

Hoofdstuk 7 onderzoekt de verschillende strategieën van talen om het argumentmarkeringspatroon van een bepaald predicaat te wijzigen door valen-

tieveranderende derivaties. De verschillende alternanties worden gedefinieerd op grond van hun effect op de valentie van het predicaat en het type semantische situatie dat ze uitdrukken. De verschillende valentieveranderende constructies worden zodanig gedefiniëerd dat zij optimaal vergelijkbaar zijn, en een aantal verschillende parameters van structurele variatie worden verkend.

De typologische kwesties die in dit proefschrift worden besproken helpen bij het vaststellen van de belangrijkste parameters van variatie binnen de verschillende grammaticale domeinen ten behoeve van een structurele questionnaire. De hierin gecodeerde typologische kenmerken vormen de basis voor een analyse van de geografische verspreiding van de argumentmarkeringskenmerken, die in hoofdstuk 8 aan bod komen. Zij vormen ook de empirische basis voor andere stellige beweringen in dit proefschrift.

In verschillende studies is gepoogd om taalkundige kenmerken te identificeren die zijn verbreid door taalcontact, hetgeen tot verscheidene voorstellen geeft geleid voor linguïstische area's binnen het Zuid-Amerikaanse continent. Op basis van eerdere kwantitatieve benaderingen van areale typologie schetst hoofdstuk 8 een procedure om de areale bepaaldheid van een een taalkundig kenmerk vast te stellen in een vooraf bepaalde regio en de geselecteerde talen die er worden gesproken. Meer dan 40 verschillende taalkundige kenmerken worden getest op hun areale aard op twee nieveaus van geografische analyse. Het eerste niveau vertegenwoordigt de hoofdregio's van het continent, zoals in hoofdstuk 1 genoemd. Het tweede analyseniveau verkent grotere macroregio's die door de samenvoeging van de eerdergenoemde hoofdregio's worden gevormd.

In de analyse wordt geëvalueerd hoe geschikt een bepaalde regio is als variabele om de typologische verspreiding van een bepaald kenmerk te voorspellen. De mogelijkheid dat deze verspreidingen geweten kunnen worden aan gedeelde overerving van een bekende gemeenschappelijke voorouder wordt beperkt door het gebruik van een genealogisch gedifferentieerd taalsample. De mogelijkheid dat deze verspreidingen het resultaat zijn van puur toeval of onafhankelijke taalinterne veranderingen wordt beperkt door het gebruik van statistische deductiemethoden waardoor alleen kenmerken worden geaccepteerd die met een waarschijnlijkheid van meer dan 95% niet aan toeval kunnen worden geweten. Indien een geografisch gebied een goede voorspellingsvariabele is voor de verspreiding van een bepaald kenmerk en het vóórkomen ervan niet aan toeval of gemeenschappelijke afstamming kan worden geweten, dan vertegenwoordigt de verspreiding van dit taalkundige kenmerk door taalcontact een mogelijke verklaring van dit distributiepatroon.

Deze analyse leidt tot interessante resultaten zowel op regionaal als macroregionaal niveau. De meest in het oog springende uitkomst van deze analyse is dat, hoewel een aanzienlijk aantal kenmerken een areale verspreiding bleken te hebben in de macroregio's van de Andes, westelijk Zuid-Amerika en oostelijk Zuid-Amerika, geen enkel kenmerk een areale verspreiding in het Amazonegebied vertoont. Dit lijkt erop te duiden dat het Amazonegebied geen goede kandidaat is voor een linguïstische area gebaseerd op de kenmerken die in deze studie zijn onderzocht. Een ander intrigerend resultaat is dat vele areale ken-

merken van de Andes gedeeld worden met de talen van westelijk Zuid-Amerika in het algemeen. Dit duidt erop dat vele sociohistorische processen waardoor deze taalkenmerken zijn verspreid ook invloed hebben gehad op de ontwikkeling van deze kenmerken in het westelijk Amazonegebied, in het bijzonder in de talen aan de bovenlopen van de zijrivieren van de Amazone, in de uitlopers van het Andesgebergte.

De in de gegevens waarneembare patronen met betrekking tot de oostelijke en westelijke macroregio's zijn niet geheel verrassend indien men hetgeen over de verspreiding van de inheemse volken van Zuid-Amerika uit andere disciplines bekend is in acht neemt. Archeologisch onderzoek heeft bijvoorbeeld aangetoond dat grote delen van her Amazonebekken relatief schaars bevolkt bleven na de expansieperiode van talen. De inheemse populaties van het Amazonegebied in de pre-expansieperiode waren zodoende geconcentreerd in specifieke geografische gebieden die de ideale context boden voor de verbreiding van taalkundige kenmerken en de convergentie van de talen tot een gemeenschappelijk structureel profiel. De expansieperiode zou deze kenmerken dan verder hebben verspreid, hetgeen tot een herformulering zou leiden van de sociale netwerken die multi-etnisch contact bevorderen en tot de distributie van de taalkundige kenmerken die we tegenwoordig waarnemen.

De resultaten die in dit proefschrift worden gepresenteerd suggereren dat het onderzoeksdomein moet worden uitgebreid met aanvullende grammaticale systemen en dat bestaande comparatieve gegevens verder moeten worden geanalyseerd met gebruik van soortgelijke kwantitatieve methoden. Het oost-west onderscheid, dat dusdanig opvalt in de argumentmarkeringsgegevens, moet verder worden getest met andere taalkundige kenmerken. Naarmate onze typologische kennis van de Zuid-Amerikaanse talen toeneemt, is het mogelijk om verder toe te werken naar een bredere synthese met de andere disciplines die zich bezighouden met de diversiteit en de prehistorie van de Nieuwe Wereld, zoals de archeologie, de genetica, geografie, etnografie en demografie. Zulk een meerzijdige benadering, die taal, cultuur, populatiedynamiek en ecologie integreert, is het meest geschikt om nieuwe inzichten te ontsluiten in de ontwikkeling van de inheemse gemeenschappen van het continent, en brengt ons daarmee geleidelijkaan dichterbij een beter begrip van de Zuid-Amerikaanse puzzel.

Curriculum Vitae

Joshua Birchall was born on February 20th, 1985 in Rockford, Illinois, USA. He earned a Bachelor's degree in Anthropology from the University of Montana in 2007. In 2008, he completed a Master's degree at the University of Montana in Linguistics, with a keen interest in the languages of the Americas. In 2009, he was awarded a Fulbright fellowship to conduct research on the Chapacuran language family at the Museu Paraense Emílio Goeli in Brazil, including extensive fieldwork with the endangered Oro Win language. The typological research that would eventually become his doctoral dissertation also began in 2009 as part of the ERC funded Traces of Contact research project at Radboud Universiteit Nijmegen.