

Cilene Rodrigues
Andrés Saab *Editors*

Formal Approaches to Languages of South America

 Springer

Formal Approaches to Languages of South America

Cilene Rodrigues • Andrés Saab
Editors

Formal Approaches to Languages of South America

 Springer

Editors

Cilene Rodrigues
Department of Literature and Letters
Pontifical Catholic University of Rio de
Janeiro (PUC-RIO)
Rio de Janeiro, Brazil

Andrés Saab
Argentine National Scientific and Technical
Research Council (CONICET)
University of Buenos Aires
Buenos Aires, Argentina

ISBN 978-3-031-22343-3 ISBN 978-3-031-22344-0 (eBook)
<https://doi.org/10.1007/978-3-031-22344-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our parents,

Andrés Saab:

*Professor Ada Morandi (Education –
Universidad Nacional de Luján, Argentina)
Professor Jorge Saab (History – Universidad
Nacional de la Pampa, Argentina)*

Cilene Rodrigues Paviani:

*Professor Therezinha Isaia Paviani
(Biology – Universidade de Brasília, Brazil)
Professor Aldo Paviani (Geography –
Universidade de Brasília, Brazil)*

*For their contributions to the scientific
development of South America and for their
thrust on us to continue their path.*

Preface

Editing a book is a challenging task. The very first step, choosing the collaborators, is already hard. Very often, editors must select from a set of excellent researchers a small subset. The subsequent steps, organizing, editing the chapters, and writing some introduction to the material, are not easier. Editors ask themselves if they will be able to put forward some leading ideas and organize the papers well, doing justice to the reputation of their authors, emphasizing the excellence of their research. To guide us through this process, we set down some criteria prior to the preparation of the present book. First, given our interests and know-how, we established a *formal-approach* criterion, assembling the book only with investigations focused on formal properties of South American languages. By applying this criterion, we narrow down the set of possible collaborators and the languages represented in the book. Although some local languages raised fascinating research questions, they haven't been fully investigated yet. Emergent sign languages are one example. By analyzing these languages, we might be able to make important contributions to investigation on language evolution, answering relevant questions about the architecture of grammar. Unfortunately, however, formal research on South American emergent sign languages is very limited for now. Second, as we had decided to offer to the public a book reflecting the linguistic diversity of South America centering on three main groups of native languages (indigenous languages, sign languages and languages of the colonizers), we applied a *linguistic-diversity* criterion. Therefore, we gave preference to research on different languages, representing, whenever possible, languages from different regions of South America. To obey this criterion, we had no choice but selecting some researchers and not others. Thus, we acknowledge that some excellent theoretical linguists were left out of our book project. Our third criterion was *local-research*, privileging, whenever possible, research conducted within South America. This criterion reflects our hope that this book will also bring visibility to local researchers, because, despite the barriers to the development of science in Latin America, they spare no efforts to do linguistics.

We are very pleased with the result of our work as a team. Each chapter of this book offers intriguing data and interesting theoretical considerations. Thus, we are grateful to our colleagues for having embraced the project, fabricating the book with

wonderful material. Among the qualities of the Springer editors that supported us, we emphasize their professionalism and their patience. As said, editing a book is a heavy task, and having to deal with deadlines and bureaucracies can add a lot to that. In our case, however, it did not because Springer has on its backstage a wonderful group of internal editors. Nevertheless, we did not want to overload Springer editors with work. Thus, we took to ourselves the responsibility of revising each chapter several times trying to format them in accordance with Springer Guidelines. We did not do this alone. Pablo Zdrojewski, one of the authors of Chap. 5, helped us a great deal. He did it based on friendship and love for good science. *Gracias por todo, Pablo!*

We do hope you, the reader, enjoy reading the book and may it contribute to your own research as yours contribute to ours.

Rio de Janeiro, Brazil
Buenos Aires, Argentina

Cilene Rodrigues
Andrés Saab

Contents

South-American Languages in a Formal Perspective	1
Cilene Rodrigues and Andrés Saab	
Part I South-American Sign Languages	
The Morpho-phonology of Nominal Plurality in Argentinian Sign Language (LSA)	25
Yanina Boria and Carlos Muñoz Pérez	
The Grammar of Agreement in Libras	57
Guilherme Lourenço	
Argument Structure in Peruvian Sign Language	79
Miguel Rodríguez-Mondoñedo	
Blending Libras and Portuguese: Acceptability Variables	107
Ronice Müller de Quadros, Diane Lillo-Martin, and Marilyn Mafra Klamt	
Part II South-American Romance Languages	
The Grammaticalization of <i>Igual</i> in Argentinean Spanish	133
Ángela Lucía Di Tullio, Mercedes Pujalte, and Pablo Zdrojewski	
Approaching the So-Called Topic-Subjects in Brazilian Portuguese from Below	171
Jairo Nunes and Mary A. Kato	
Is Chilean Spanish a Canonical Pro-drop Variety? On Subjecthood in Chilean Spanish	203
Iván Ortega-Santos	
Ways of Number Marking: English and Brazilian Portuguese	235
Roberta Pires de Oliveira	

Part III South-American Indigenous Languages	
Compounding Processes in Three Macro-Jê Linguistic Branches	257
Andrew Nevins and Mário André Coelho da Silva	
<i>Poro-/mba'e-</i> Antipassive Prefixation in Paraguayan Guarani	285
Bruno Estigarribia and Ernesto Luiz López Almada	
Argument Structure and Morphology in Cochabamba Quechua (with Occasional Comparison with Other Quechua Varieties)	311
Neil Myler	
Definiteness in A'ingae and Its Implications for Pragmatic Competition ..	347
Holly Zheng and Scott AnderBois	
Index	373

About the Editors and Authors

Editor's Bios

Cilene Rodrigues is an adjunct professor at Pontifical Catholic University of Rio de Janeiro (PUC-Rio) and research fellow at *Conselho Nacional de Pesquisa* (CNPq/Brazil). Her PhD thesis (University of Maryland, 2004) focused on licensing of null subjects in partial pro-drop languages. She is currently directing research on (a) the grammar of Brazilian Languages, especially colloquial Brazilian Portuguese, Tupi-Guarani languages, and Brazilian sign language (Libras); (b) evolution of language, with interdisciplinary investigations on language and psychosis, particularly schizophrenia, grammar and interjections, and verbal versus vocal behavior, with fieldwork on the vocal repertoire of Capuchin monkeys.

Andrés Saab studied literature and linguistics at the University of Buenos Aires and at the National University of Comahue (Argentina). In 2009, he defended his PhD dissertation on the theory of ellipsis. Currently, he is an independent researcher at the Argentine National Scientific and Technical Research Council (CONICET), having the Argentinian Society for Philosophical Analysis (IIF-SADAF-CONICET) as official work place, and an associate professor at the University of Buenos Aires. His main research topics are ellipsis, copy theory of movement, null subjects, and, more broadly, the syntax–interface connection. His research has been published by international journals and books (*Linguistic Inquiry*, *The Oxford Handbook of Ellipsis*, *Natural Language and Linguistic Theory*, *Lingua*, *Probus*, and *Studia Linguistica*, among others). He is also co-author of *Dimensiones del Significado. Una introducción a la semántica formal* (with Fernando Carranza, SADAF, Buenos Aires, 2021), co-editor of *Slurs and Expressivity. Semantics and Beyond* (with Eleonora Orlando, Lexington, USA, 2021), and co-editor of the volume *Romance Language and Linguistic Theory 2010* (with Irene Franco and Sara Lusini, John Benjamins, Amsterdam. 2012).

Authors' Bios

Scott AnderBois is Associate Professor of Cognitive, Linguistic, and Psychological Sciences at Brown University, having received his PhD in linguistics from UC Santa Cruz in 2011. His research uses primary fieldwork to explore questions in semantics and pragmatics, with a particular focus on the discourse effects of different sentence types (e.g., declarative, interrogative, imperative) and the ways different evidentials, miratives, and other discourse particles interact with them. He has explored such questions primarily in Yucatec Maya, Tagalog, English, and A'ingae (Cofán). He also is co-director of a community-engaged language documentation project working with speakers of A'ingae and others to create a multimedia, multi-purpose record of the language and culture of the A'i people.

Yanina Boria is a doctoral student of linguistics at Universidad de Buenos Aires. She holds a degree in language and literature from the same institution, and she is also a technical interpreter of Argentinian Sign Language. She works as an accessibility director of the National Agency of Disabilities (ANDIS), within the Argentinian government. She teaches sign language grammar as part of a degree in sign language interpretation and is also engaged at Asociación Civil de Artes y Señas (ADAS).

Mário André Coelho da Silva is an assistant professor at Universidade Federal do Acre, in Cruzeiro do Sul, Brazil, since 2022, and a postdoctoral researcher at Universidade Estadual de Campinas, Brazil. His research focuses on describing Macro-Je languages, especially Maxakalí and Krenák languages. His investigation concerns mostly phonological and morphological description, but recently it has engaged on historical comparison of the aforementioned languages as well. He also has experience in the training of indigenous educators.

Ángela Lucía Di Tullio is Full Professor of Spanish Grammar and Hispanic Philology at Universidad del Comahue (1973–2012). Since 2012, she has served as a researcher at Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso,” Universidad de Buenos Aires. She is the author of *Manual de gramática del español* (1997, 2005, 2010), *Políticas lingüísticas e inmigración: el caso argentino* (2003, 2010), and *Gramática del español para maestros y profesores del Uruguay* (with Marisa Malcuori). She is the editor of *El español de la Argentina: estudios gramaticales* (Eudeba, 2013). Also, she coedited *El español rioplatense: lengua, literatura, expresiones culturales* and *Universales vernáculos en la gramática del español* (both in Iberoamericana, 2011 and 2021), and some journals related with synchronic and diachronic grammar. She has participated in the elaboration of *New Grammar of Spanish Language* of the Spanish Royal Academy of Language (2003–2010). Invited by the universities of Lovaina, Naples, Girona, Santiago de Compostela, Valparaíso, Montevideo, México, Córdoba, she dictated graduate courses.

Bruno Estigarribia completed a *Maîtrise* in language sciences at Université Paris V-René Descartes-Sorbonne in 2001 and an MA in linguistics at Stanford University in 2005. He obtained his PhD in linguistics from Stanford University in 2007, with a dissertation investigating children's acquisition of the syntax of English questions. From 2007 to 2009, he was a National Institutes of Health T32 Postdoctoral Fellow, working on the language and cognitive development of children with neurodevelopment disabilities. He continued this work as research faculty in the cognitive science program in UNC's Psychology Department. Now, he is a professor in the Department of Romance Studies, in the Hispanic linguistics program. His current research focuses on syntax and semantics in Spanish (especially Argentinian Spanish), on the description of Paraguayan Guarani, and on the contact between these two languages. He has recently published the Open Access reference grammar *A Grammar of Paraguayan Guarani* with UCL Press. His work has been published in journals such as *the International Journal of American Linguistics*, *Glossa*, *Probus*, *Journal of Language Contact*, *Journal of Linguistics*, *Journal of Child Language*, *Cognitive Science*, *Applied Psycholinguistics*, *Journal of Speech, Language, and Hearing Research*, and the *International Journal of Language and Communication Disorders*, among others.

Mary A. Kato is retired full professor from the University of Campinas, Brazil, where she is currently a researcher collaborator. She holds an MA in English linguistics, a PhD in applied linguistics, and has devoted her career to the analysis of formal aspect of Portuguese grammars, with important contributions to comparative syntax, focusing on understanding the parameters of Brazilian Portuguese. She is a cofounder of *DELTA*, a Brazilian Scientific Journal dedicated to linguistics, and the editor relevant books including *The morphosyntax of Portuguese and Spanish in Latin America*, *Brazilian Portuguese and the Null Subject Parameter* and *Português Brasileiro: uma viagem diacrônica*.

Diane Lillo-Martin is board of trustees distinguished professor and head of the Department of Linguistics at the University of Connecticut, and a senior research scientist at Haskins Laboratories. Her research is on the structure and acquisition of sign languages, the development of bimodal bilingualism, and the structure of code blending.

Ernesto Luiz López Almada holds a BA in bilingualism, completed in 2011, and English as foreign language, completed in 2014, and he also completed his MA in higher education and research from Universidad Nacional de Itapúa, Paraguay. He is a categorized researcher of the National Research Incentive Program PRONII of the National Council of Science and Technology CONACYT from 2018 to date. He participated as a research associate, visiting professor, and collaborator with Western Washington University, Washington; Arizona State University, Arizona; and the University of North Carolina Chapel Hill, North Carolina, United States, in 2019, 2020, and 2022, respectively. He was assistant researcher in the Project 14 INV 222 "Multilingualism in the Department of Itapúa: The Official, Indigenous and Foreign Languages in Contact" in 2019 and main researcher in the research

project PINV 18 1284 called “Active Guarani Learning Resource: Development of Multimedia Resources for the Learning of the Guarani Language Online” funded by the National Council of Sciences and Technology CONACYT during 2020–2021. Currently, he is executive editor for *the Journal on Studies and Research of Academic Knowledge*, a multidisciplinary peer-reviewed academic journal. His research interest includes bilingualism, multilingualism, linguistics, translations, and language in contacts.

Guilherme Lourenço is currently Professor of Sign Language Linguistics at the Federal University of Minas Gerais (UFMG, Universidade Federal de Minas Gerais), Brazil. Dr. Lourenço earned his PhD from UFMG, Brazil, and was a visiting scholar in the Sign Language Linguistics Research Lab, within the Department of Speech, Language and Hearing Sciences at Purdue University, USA. His research interest includes sign language linguistics, with special focus on morphology, syntax, and semantics, and applications in language translation and interpreting.

Marilyn Mafra Klamt is an associate professor at the Federal University of Santa Catarina, Libras Department. Her research is on deaf literature, literature in Libras, and sign language linguistics.

Neil Myler is associate professor of linguistics at Boston University. His research interests include morphology, comparative morphosyntax, argument structure, the morphosyntax and semantics of possession, Quechua morphology and syntax, and English dialect syntax. He is the author of the monograph *Building and Interpreting Possession Sentences*, published by the MIT Press in 2016.

Ronice Müller de Quadros is full professor at the Federal University of Santa Catarina, Libras Department, since 2002, and CNPQ researcher. Her research is on Libras grammar, bimodal bilingualism, and language policies.

Carlos Muñoz Pérez is adjunct associate professor at Universidad Austral de Chile. He received his PhD in linguistics from Universidad de Buenos Aires in 2017 with the financial support of CONICET. His research focuses on syntactic theory, comparative grammar, and Spanish microvariation. He is particularly interested in the diversity of externalization mechanisms displayed by languages across the globe.

Andrew Nevins researches the contributions of minoritized languages to linguistic theory, including indigenous, diaspora, and signed languages of Brazil, and among his contributions involve the results of fieldwork and experimental linguistics conducted with the Maxakalí in Brazil. He is a professor at the Federal University of Rio de Janeiro and at University College London, and strongly committed to greater representativity and inclusion within the language sciences.

Jairo Nunes is a full professor at Universidade de São Paulo. He is author of *Linearization of Chains and Sideward Movement* (2004); co-author, with Norbert Hornstein and Kleantes K. Grohman, of *Understanding Minimalism* (2005); co-editor, with Norbert Corver, of *The Copy Theory of Movement* (2007); editor of

Minimalist Essays on Brazilian Portuguese Syntax (2009); co-author, with Cedric Boeckx and Norbert Hornstein, of *Control as Movement* (2010); and co-editor of the journal *Probus*.

Iván Ortega-Santos is Associate Professor of Spanish Linguistics at the University of Memphis. His work focuses on dialectal variation in Spanish, data collection standards in linguistic theory, and word order variation and the limits natural language imposes on it. His book *Focus-related Operations at the Right Edge in Spanish: Subjects and Ellipsis* has been published by John Benjamins.

Roberta Pires de Oliveira is a full professor at Federal University of Santa Catarina and a CNPq researcher. Her research interests center on syntax-semantics of nominal phrases across natural languages, with focus on Brazilian Portuguese nominal system. The scope of her research includes native languages as well as Rikbatksa (Macro-Jê), considering data from other methodologies, such as experimental semantics. Roberta has important contributions to the development of formal semantics in Brazil, and her curriculum features relevant publications, including the book *Brazilian Portuguese, Syntax and Semantics: 20 years of Núcleo de Estudos Gramaticais* and the textbook *Semântica formal: uma introdução*.

Mercedes Pujalte is Professor of Linguistics and Spanish Grammar at Universidad Nacional de General Sarmiento and at Universidad Nacional de Hurlingham. She also teaches semiotics at Universidad de Buenos Aires. Her major theoretical research interests are syntax and syntax-morphology interface, especially the relation between argument structure and functional structure. She has published several book chapters and articles about those aspects of Spanish Grammar.

Miguel Rodríguez-Mondoñedo is Professor of Linguistics and associate dean (Director de Estudios) in the Faculty of Letters and Human Sciences at Pontificia Universidad Católica del Perú. He is a syntactician, and the principal investigator in several projects on Peruvian Sign Language (LSP), working on the first LSP grammar and the first comprehensive LSP dictionary. He received his PhD from the University of Connecticut, Storrs.

Holly Zheng graduated from Brown University in 2022 with a Bachelor of Arts with Honors in Linguistics and a Bachelor of Arts in Computer Science. For her linguistics honors thesis advised by Scott AnderBois, she conducted field work in A'ingae and investigated the distribution and uses of anaphoric expressions in the language.

Pablo Zdrojewski is Assistant Professor of Spanish Grammar, Phonology, Morphology, Syntax, and Generative Grammar in Facultad de Filosofía y Letras at Universidad de Buenos Aires. He is also a researcher at Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso” in the same university. His main theoretical research interest are syntax, syntax-morphology interface, and information structure. His research has focused on clitic phenomena, such as clitic doubling, on differential object marking, and their variation across Spanish dialects. In addition, he has investigated some phenomena related to information structure

and to argument structure. He has published several journal articles and book chapters on these topics of Spanish.

Contributors

Scott AnderBois Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI, USA

Yanina Boria Universidad de Buenos Aires, Buenos Aires, Argentina

Mário André Coelho da Silva Universidade Federal do Acre, Rio Branco, Acre, Brazil
Universidade Estadual de Campinas, Campinas, Brazil

Ángela Lucía Di Tullio Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso” – Universidad de Buenos Aires, Buenos Aires, Argentina

Bruno Estigarríbia University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Mary A. Kato Universidade Estadual de Campinas, Campinas, Brazil

Marilyn Mafra Klamt Universidade Federal de Santa Catarina, Florianópolis, Brazil

Diane Lillo-Martin University of Connecticut, Storrs, CT, USA

Ernesto Luiz López Almada Departamento de Lenguas, Universidad Nacional de Itapúa, Encarnación, Paraguay

Guilherme Lourenço Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Ronice Müller de Quadros Universidade Federal de Santa Catarina, Florianópolis, Brazil

Carlos Muñoz Pérez Universidad Austral de Chile, Valdivia, Chile

Neil Myler Boston University, Boston, MA, USA

Andrew Nevins Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
University College London, London, UK

Jairo Nunes Universidade de São Paulo, São Paulo, Brazil

Iván Ortega-Santos University of Memphis, Memphis, TN, USA

Roberta Pires de Oliveira Universidade Federal de Santa Catarina/CNPq, Florianópolis, Brazil

Mercedes Pujalte Universidad Nacional de Hurlingham-CONICET – Universidad Nacional de General Sarmiento, Argentina

Cilene Rodrigues CNPq – Conselho Nacional de Pesquisa and Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

Miguel Rodríguez-Mondoñedo Pontificia Universidad Católica del Perú, San Miguel, Peru

Andrés Saab IIF-SADAF-CONICET and University of Buenos Aires, Buenos Aires, Argentina

Pablo Zdrojewski Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso” – Universidad de Buenos Aires, Buenos Aires, Argentina

Holly Zheng Brown University, Providence, RI, USA

South-American Languages in a Formal Perspective



Cilene Rodrigues and Andrés Saab

1 Linguistic Diversity in South America

South America extends, north to south, from Punta Gallinas in Colombia to the Drake Passage in Cape Horn and, east to west, from Cabo Branco in Brazil (Ponta do Seixas) to Punta Pariñas in Peru, being the fourth largest world's continent in territory and 50th in population. It is one of our last storehouses of natural resources, including fauna, flora, mineral, and hydric reserves. However, for centuries, South America's natural reserves have been increasingly depleted without much concern and planning. The present book is devoted to another, even less minded, often ignored, treasure of South America: linguistic diversity. Although some of us, linguists, defend that language is part of our genetic endowment, while others take it to be a cultural asset, we all agree that languages are a humankind patrimony, and linguistic preservation, documentation, and analysis are priorities, especially when minority languages are considered. Thus, studies on languages of South America are first concern.

One fourth of the language families of the world is located in South America (Campbell 2012), with the majority of local languages being spoken by communities with much less than 1 million speakers. Also, there are about 625 living languages in South America, but, according to Lewis (*Ethnologue: languages of the world*, 2009), 179 of them are dying.

C. Rodrigues (✉)

CNPq – Conselho Nacional de Pesquisa and Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

e-mail: crodrigues@puc-rio.br

A. Saab

IIF-SADAF-CONICET and University of Buenos Aires, Buenos Aires, Argentina

e-mail: andres.saab@uba.ar

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

C. Rodrigues, A. Saab (eds.), *Formal Approaches to Languages of South America*,

https://doi.org/10.1007/978-3-031-22344-0_1

Recently, linguists from different areas have shown an increasing interest in these languages. Nevertheless, the conducted studies are mostly typological in nature, being specially devoted to the indigenous languages, 420 in total. The languages of the colonizers, which include Brazilian Portuguese, non-peninsular varieties of Spanish and French, and a number of local sign languages, are usually left out of purview in studies about native languages of South America.

The languages European colonizers brought to South America underwent processes of transformation giving rise to new, local grammars. For example, there are major parametric differences that clearly set apart Brazilian Portuguese and varieties of South-American Spanish from their European origins. Some of these parametric differences will be discussed in the second part of the book.

As for sign languages, although present classifications and counting are uncertain, there are around 30 sign languages in South America, which are composed of numerous dialects. Similarly, to Brazilian Portuguese and the local varieties of Spanish, some of South-American sign languages were subject to influences from non-autochthone, overseas sign languages. For example, Libras – Brazilian Sign Language – was influenced by French Sign Language and Argentinian Sign Language by Italian Sign Language. However, they should also be classified as native languages of South America, as they have their own linguistic systems, which emerged within local deaf communities.

2 Formal Approaches to Human Languages

The perspective under which each chapter of this book approaches the linguistic diversity in South America is deeply rooted in the tradition of formal grammar. Human languages are complex objects, the result of certain genetic disposition of humans and myriads of historic culture encoded in the atoms (i.e., roughly, words or morphemes) that form the vocabulary of each particular language. We can see languages from two points of view: (i) as mere corpora attested in oral or written production or (ii) as a cognitive capacity, partially responsible for the attested corpora. This characterization roughly corresponds to the well-known Chomskyan distinction between E-language (E for external) and I-language (I for internal) (Chomsky 1986). Both objects can be represented formally. On the one hand, the utterances that form particular E-languages can be segmented and decomposed in various ways, given certain degree of abstraction. A reasonable linguistic decomposition would attribute to a single utterance several levels of linguistic analysis, each of which corresponds to the phonetic, phonological, morphological, syntactic, and semantic dimensions of the said utterance. On the other hand, genetic attributes endow humans with a recursive system of atom combinations, whose ultimate outputs are language sentences, externally interpreted by systems of thought and language production, i.e., the so-called linguistic interfaces (Chomsky 1995). In a sense, what is common to all humans, at least under normal circumstances, is then some sort of Universal Grammar (UG). UG can also be formally represented

through a complex system of *principles* (or equivalently, of *natural laws*). If such system is correctly modeled, then linguistic theory can be conceived of as a theory of an essential aspect of human cognition.¹

Now, under normal conditions, any child is exposed to the imperfect input of an E-language (or more than one) on the basis of which she will be ending up generating her own I-language. Myriads of children under similar conditions will acquire similar I-languages in the same period of time. The key to understanding why the acquisition process is relatively fast is genetic endowment, more specifically, UG, which in combination with external data will help the child to fix her I-language. Now, external data depend on many factors that are in essence external to the language faculty itself. The historical and sociocultural complexity of such a vast territory like South America will obviously impact on the development and growing of many local grammars, which also interact among themselves in complex ways. For instance, a child who was born in the city of Buenos Aires in the 1940s was probably exposed to many fragments of Italian dialects, varieties of Peninsular Spanish, and even Galician. In addition, the same child was also exposed to sociolectal varieties of Rioplatense Spanish that at that time was already relatively fixed. As noted by Muñoz Pérez and Saab (in prep.), this entire mass of linguistic evidence does not count as a consistent set generated by any particular I-language. However, in a relatively short period of time, she will produce utterances entirely determined by her own I-language. This situation is replicated in myriads of Buenos Aires neighborhoods, where at least half of the immigrants came from different parts of Italy in that period of time. Any particular I-language acquired by those children is a pure individual language, but given that the external factors conditioning the process of language acquisition are quite similar for each of these children, their I-languages will have many common properties regarding their phonetics, phonology, morphology, and syntax. A close observation can be made about Brazilian Portuguese, which is not a language in itself, but a set of dialects defined by different regional and socioeconomic populations of speakers. Language contact in Paraguay, where local Spanish coexists with different varieties of Paraguayan Guarani, including a language mixing, provides us with yet another similar acquisition issue (Estigarribia 2020).

Given this perspective, the tasks for any plausible formal approach to human language would be (i) to describe the type of knowledge internalized in the mind of the individual that fixed a particular I-language (varieties of Spanish, Quechua, Chinese, Brazilian Sign Languages, and so on) and (ii) to abstract from the complex body of available evidence general principles (i.e., natural laws) common to all particular I-languages. Such tasks require a comparative approach to human languages. In addition, comparisons must take place at different macro- and microscopic levels. As lucidly argued by Kayne (2005), microscopic dialectal

¹ But note that even if one does not ascribe to the idea that linguistic theory is a branch of cognitive psychology, one can still characterize syntax as a complex system of axioms under which each sentence of a given language is formally derived as a theorem of such axiomatic system.

comparison allows the linguist to isolate the properties under study in a sort of natural laboratory. As is clear, comparing a certain property P in two close varieties, in which other factors remain identical, considerably reduces the risk of introducing unwanted confounding factors and allows to the linguist to focus on P alone. Methodologically speaking, this is an obvious advantage, specially, if it is true that macro and micro differences among language types are just a matter of degree, as claimed by Kayne himself. But given our current state of knowledge on linguistic diversity, the language = dialect thesis cannot be categorically asserted, and there is the possibility that micro-linguistic differences interact in various complex ways with macro-parameters of the well-known type (Baker 2008). Roberts (2019) seems to agree with the latter position and provides a hierarchy of parameters that would underlie both big qualitative linguistic types but also minimal differences sometimes, reduced to a mere lexical item (his nano-parameters). We don't want to take a particular stance regarding this fascinating debate, but we are convinced that any methodology, including macro and micro approaches to language diversity and modality comparisons among oral and sign languages, is welcome and all have different contributions to make to a better understanding of the language faculty. This is one of the essential reasons this book is divided into three different parts, each of one approaching linguistic diversity from different, but needed, points of view, from dialectological to typological approaches and from sign to oral languages.

3 Part I: South-American Sign Languages

Not long ago, linguists were mostly preoccupied in gathering comparative data verifying points of convergence and divergence between sign and oral languages in order to establish whether sign languages (SLs) were natural languages or not. Bloomfield's (1933) characterization of SLs as primitive gestural systems and Hockett's (1960) assumption that the vocal-auditory channel is one of the defining features of human language had previously contributed to an erroneous understanding of natural as synonymous of oral. But this understanding was soon challenged. Stokoe's (1960) groundbreaking work, further developed by Friedman (1977) and Klima and Bellugi (1979), among others, on the phonemics of ASL (American SL), indicated that SLs are identical to oral languages in duality of patterning (Hockett 1960), which pointed towards an identity between sign and oral languages in terms of combinatorial power. In the 1980s and 1990s, comparative research on grammatical structure (Liddell 1980, among others), on acquisition (Petitto 1997, among others), and on the brain structure and function of language (Bellugi et al. 1983; Poizner et al. 1987) cemented, as an uncontroversial empirical fact, the conclusion that deaf and hearing populations have the same linguistic ability.

The history of SL linguistics has had consequences far beyond the study of SLs in itself. First, it shows that applying analytic tools to the analysis of a linguistic

phenomenon, providing an explicit account, is a necessary step in deciding whether that phenomenon reflects some natural systematic body of knowledge or not. The aforementioned findings on ASL phonemics emerged because a structural analysis of the elementary units of the language was applied first by Stokoe and later on by Friedman on phonotactics and by Klima and Bellugi's explicit comparisons with oral language phonology. Second, it paved the way to a perspective of language according to which the linguistic code of a language L is to be differentiated from L's mode of externalization (Sandler and Lillo-Martin 2006). While oral and sign languages present patent differences in mode of externalization, their underlying grammatical cores are quite comparable, resulting from the same brain mechanisms and cognitive architecture. This resonates with the fundamentals of generative grammar theory, according to which all languages share deep-seated structural regularities (I-language – Chomsky 1965, 1986; see Sect. 2), with linguistic diversity reflecting differences at PF. Thus, modality (oral-aural and gestural-visual) raises interesting issues related to externalization (e.g., linearity vs. simultaneity), rather than pointing towards differences in grammar architecture. This change in perspective promoted a balanced and informed discussion on the centrality of gestures to human communication and its effects on the components of language, resulting in a more refined view of the role played by iconicity in language development, production, and processing (Perniss et al. 2010; Meteyard et al. 2015; Schlenker 2022, among others) and on the relationship between speech, sign, and gesture (Messing and Campbell 1999; Kendon 2008; Goldin-Meadow and Brentari 2017, among others).

As SL linguistics is now a well-established field of research, there is an increasing number of investigations on different SLs and a great number of comparative studies focusing on SLs. But still, it is to be acknowledged that, much like what is observed for oral languages, current formal investigations concentrate mostly on ASL and sign languages of large communities within Europe. The bulk of cross-linguistic comparative studies promote comparisons among these languages or with them, leaving non-American and non-European SLs out of purview. This is not necessarily an unnatural outcome – comparisons are based on those languages because we know more about them, but still, it indicates a pressing need for research on less well-known languages.

The *ethnologue* website lists 154 SLs in total (129 deaf community sign languages and 27 shared sign languages; <https://www.ethnologue.com/subgroups/sign-language>). Of those, 13 are in South America, 10 languages of deaf communities, and 3 shared languages. This is, however, an under-representation of South America SLs. Brazil alone is a linguistic nebula, where more than 12 emerging SLs were already identified, being located mostly within indigenous communities (Silva and de Quadros 2019; Almeida-Silva and Nevins 2020). Mapping, describing, documenting, and analyzing these languages should be of most important consideration, as they can provide us with information about I-language evolution (e.g., how linguistic complexity emerges) and how language development is shaped by biological and environmental forces (the nature/nurture debate). Particularly, emerging SLs are a fertile empirical ground to investigations on synchronic and

diachronic relation between gestures and signs (Sandler et al. 2014; Meir et al. 2003). Thus far, we don't have comprehensive formal analyses of South-American emerging SLs, but hopefully in a near future we will.

The first part of our book is dedicated to well-established deaf communities' languages SLs of South America. The four chapters that compose it focus on Brazilian Sign Language (Libras, *Língua Brasileira de Sinais*), Argentinian Sign Language (LSA, *Lengua de Señas Argentina*), and Peruvian Sign Language (LSP, *Lengua de Señas Peruana*). These languages have different degrees of official recognition. Libras was legally recognized as the national language of the deaf community of Brazil in 2002, and, in 2021, bilingual education to deaf students (Libras, Brazilian Portuguese) became mandatory by law. Libras is signed in all regions of the country, presenting some dialectal variation that hasn't been extensively mapped up to now. For a detailed presentation and analyses of Libras, we refer the reader to the book *Brazilian Sign Language Studies* (2020), organized by Ronice Müller de Quadros. LSA is the national sign language of Argentina, and, according to Massone and Martínez (2015), it presents regional lexical and phonetic variation. Despite the presence of local strong committees and organizations in its defense, LSA hasn't been legally recognized as a national language yet. LSP is composed of many generational and regional varieties (Parks and Parks 2010; Clark 2017) and it does not really figure as an official national language yet, although in 2010, the Peruvian government recognized it as a Peruvian language (Rodríguez-Mondoñedo [this volume](#)). Thus, all in all, although Libras, LSA, and LSP may well be secured in number of signers, they are still vulnerable in that they haven't been fully accepted as part of the linguistic identity of the nations where they are located. See Ramsey and Quinto-Pozos (2010) for an elaborate overview of political and social issues standing in the way of sign language transmission in Latin America. We, the editors, concur with local deaf communities and SL researchers in supporting these languages, as well as emerging and isolated SLs, as part of our linguistic heritage.

Chapter “[The Morpho-phonology of Nominal Plurality in Argentinian Sign Language \(LSA\)](#)” of the present book by Yanina Boria and Carlos Muñoz Pérez, presents a comprehensive study of nominal plurality in LSA, which is expressed either via sideward reduplication of the lexeme (e.g., *child*, *children*) or via reduplication of a meaningless epenthetic classifier (e.g., *pencil*, *pencils*). Thus, the chapter's driving question is: how does this plural allomorphy come about? Boria and Muñoz Pérez advance a unified analysis, suggesting that LSA has only one plural morpheme, which is expressed by repetitions of an arch-like movement carrying over the noun. As the plural morpheme lacks inherent phonological features, it needs to be combined with a nominal base at PF. When the phonological features of the nominal base impede it from hosting the plural affix, a sign-movement epenthesis à la Brentari (1998) occurs to rescue the derivation, resulting in an intruder classifier form.

In chapter “[Argument Structure in Peruvian Sign Language](#),” Miguel Rodríguez-Mondoñedo's quest is providing a formal analysis for the fact that LSP types of verbal classifiers vary in function of predicate types. Taking verbal classifiers to

be exponents of a functional head akin to little v (see Benedicto and Brentari 2004 for a proposal along these lines), Rodríguez-Mondoñedo suggests that the observed variation is predicted by the Unaccusative Hypothesis. Handling classifiers, which occur with transitive predicates, are agentive heads that agree with the internal argument in form/shape feature, while theta marking the external argument. Conversely, entity classifiers are non-agentive elements, heading, thus, an unaccusative structure. They agree with the internal argument in form feature, but they do not project an external argument. On its turn, body-part classifiers occur with unergative predicates. Assuming unergative predicates to be hidden transitive structures, formed by incorporation of an internal argument into the verb (Hale and Kayser 1993), Rodríguez-Mondoñedo suggests that LSP unergative verbs are formed via incorporation of a body-part inalienable possessed noun. Thus, body-part classifiers are similar to handling classifiers in that they agree with the internal argument and theta mark the external argument.

Classifiers are tough to work on because there is a lot of variability in the systems employed by languages (Aikhenvald 2000, 2005). Also, classifier constructions in SLs have a somewhat unexpected behavior, presenting some unique properties that might be related to the visual-gestural modality (Emmorey 2003; Sandler and Lillo-Martin 2006). In spite of these issues, Boria and Muñoz Pérez's and Rodríguez-Mondoñedo's chapters, focusing on different classifiers systems, provide us with a generalization: classifiers are exponents of functional categories (presumably little v in Rodríguez-Mondoñedo's analysis and Num⁰ in Boria and Muñoz Pérez's). This is by no means unique of SLs; see, for example, Watanabe's (2006) and Simpson's (2021) analyses for numeral classifiers in Japanese and Korean. Thus, the functional status of classifiers might be a universal, which Boria and Muñoz Pérez's and Rodríguez-Mondoñedo's analyses are carving out of LSA and LSP.

Chapters “The Grammar of Agreement in Libras” and “Blending Libras and Portuguese: Acceptability Variables” are dedicated to Libras. Chapter “The Grammar of Agreement in Libras” authored by Guilherme Lourenço, calls for a revision of verbal agreement in SL. It is traditionally assumed that in these languages, verb agreement is expressed via directionality: verb path movement between the referential locations specified for the verbal arguments (Padden 1988; Sandler and Lillo-Martin 2006; Lillo-Martin and Meier 2011, among others). Lourenço disagrees with this view, arguing that agreement is expressed via co-localization, not movement. Co-localization, as defined by the author, is a change in the location of the verb to match that of an argument. A corpus of Libras was analyzed for verb agreement, and almost half of the tokens displayed co-localization (agreement). Verbs that did not were either verbs with a first-person argument or body-anchored verbs. Lourenço takes first-person marking to be default agreement in Libras, and reasons that body-anchored verbs block co-localization because their point of articulation cannot be altered at PF.

A great body of work in many different and unrelated languages has pointed towards the universality of agreement, but how verbal agreement is expressed in SLs is a moot question, with literature offering many different accounts (see Quer

(2021) for an overview). One interesting feature of Lourenço's account is that it ties the proposed PF process of co-articulation to a minimalist syntax, arguing that co-articulation is a PF realization of Agree applied to phi-features. Hence, Lourenço's proposal seeks for a unified, universal syntax.

The last chapter of Part I, “**Blending Libras and Portuguese: Acceptability Variables**” by de Quadros, Lillo-Martin and Klamt, brings forward an experimental study on code-blending. An acceptability judgment task, followed by an explicit elicitation of the target items, was conducted with 22 codas, all natives of Libras and spoken Brazilian Portuguese, to verify constraints on blended sentences. The following linguistic factors were manipulated: match/mismatch in word order (OV with VO; post-verbal with pre-verbal negation), (b) match/mismatch voice (active with passive), and (c) match/mismatch in idiomatic reading (idiomatic with non-idiomatic). Results show a preference of matching structures. Mismatching in voice as well as in idiomatic reading blocks code-blending. Conflicting word orders received lower rating, with mismatches in the position of negation being more accepted. Overall, these results indicate that blending requires both derivations to be convergent, syntactically compatible and mapped into a single proposition.

We start our presentation of SLs by calling attention to the fact that I-language is stable throughout modality. de Quadros, Lillo-Martin and Klamt's experimental observations provide beautiful evidence for this. In fact, they argue, in accordance with the Synthesis Model (Lillo-Martin et al. 2016; de Quadros et al. 2020), that code-blendings are cases in which a single derivation has multiple externalizations. This is, spoken and signed utterances are produced by the very same computational system.

4 Part II: Linguistic Innovations at the South of the *Romania Nova*

The second part of this book is devoted to the analysis of certain remarkable properties of the grammars of Brazilian Portuguese and two varieties of South-American Spanish (Chilean and Argentinian), the two main Romance languages spoken in the *Romania Nova*.² The *Romania Nova* area makes pretty evident to linguists how dynamic language change is. In effect, the three varieties of Romance explored in the chapters below have some formal particularities that are substantive for a formal perspective to language evolution. To a great extent, this is intimately related to a set of complex interactions of historic, cultural and social facts, which

² Romania Nova is a term coined by Francisco Ordóñez and Mary A. Kato to refer to the Romance languages spoken in the Americas and to the research project #14 within the *Asociación de Lingüística y Filología de América Latina (ALFAL)*. The reader can visit the website of the project at <https://sites.google.com/view/romania-nova/p%C3%A1gina-principal>, where a list of activities including recent publications is available.

are constitutive of the enormous human diversity in the South-American territory. Brazilian Portuguese, a broad term covering many local grammars within Brazil, illustrates this point paradigmatically. As is well-known, the European Portuguese that the colonizers brought to Brazil (“a língua das caravelas”; see Galves (2020)) was progressively affected by the aforementioned factors, giving rise to a more stable situation in the twentieth and twenty-first centuries. For instance, the person and number verbal paradigms and the pronominal systems were radically eroded at least since the nineteenth century (see Duarte (1993, 2000)). The first-person plural pronoun *nós* was replaced by the third-person singular *a gente*; the distinction between *tu* and *você* vanished in favor of the unique second-person pronoun *você*, which is morphologically third-person singular, and the same happened in the second-person plural slot, with *vocês* as the unique second-person pronoun. This obviously impacted on the verbal paradigm, which in most Brazilian Portuguese dialects contain only three morphological distinctions: first-person singular, third-person singular, and third-person plural – *eu amo*, *você ama*, *ele/ela ama*, *a gente ama*, *vocês amam*, *eles/elas amam* (*I love, you love, he/she loves, we love, you love, they love*). Compared with Latin-American varieties of Spanish, which lost only the second-person plural *vosotros*, these changes in the pronominal and verbal paradigms in Brazilian Portuguese produced a radical typological change as far as the interpretation, syntactic distribution, and phonetic realization of subjects are concerned. In effect, the evidence recollected in the last 30 years leads to the conclusion that most varieties of Brazilian Portuguese pertain to the partial *pro*-drop type. This type is known for a series of perplexing properties, in particular, the almost absolute absence of referential null subjects, the presence of null generics, and hyper-raising effects in finite clauses (see Rodrigues (2004) for Brazilian Portuguese and Finish, Holmberg (2005, 2010) for Finish, and Kato (1999, 2000), Barbosa et al. (2005), Modesto (2000), Ferreira (2000), and Nunes (2020) for Brazilian Portuguese). According to a prominent approach to the topic, the erosion of agreement morphology in the verbal paradigm is associated to an impoverishment of the T^0 node. This sort of functional impoverishment seems to generalize to other formal heads in the Brazilian Portuguese spine. For instance, in chapter “Approaching the So-Called Topic-Subjects in Brazilian Portuguese from Below,” Kato and Nunes revise the properties of the so-called “topic-subject” construction, a well-known characteristic of Brazilian Portuguese, non-attested in the European varieties of Portuguese. Two relevant examples are provided below:

- (1) a. [Os relógios] **quebraram** o ponteiro.
 the watches broke-3PL the arm
 “The arms of the watches broke.”
- b. [Essas gavetas] **cabem** muita coisa.
 these drawers fit-3PL many thing
 “Many things can fit in these drawers.”

Mainstream approaches in the Lusophone tradition correlate the emergence of this property precisely to the fact that Brazilian Portuguese is becoming a non-*pro-drop* language. Yet, Kato and Nunes convincingly show that this is indeed a fake correlation and claim that the crucial trigger that gave rise to this “construction” must be found within other formal properties of the clause. More specifically, they propose that with the reduction of structural Case assignment possibilities at the νP level, Brazilian Portuguese started to make pervasive use of inherent Case within the verbal domain. This is a substantive result on two counts: (i) first because it clarifies the empirical domain by establishing where the locus of change is regarding topic-subjects, and (ii) because it shows that the change that Brazilian Portuguese suffered goes well beyond the T^0 node. If Nunes and Kato are correct, Brazilian Portuguese suffered a massive reduction or impoverishment of the formal content of the set of functional heads that compose the sentence structure, a fact that obviously impacts the syntactic distribution, interpretation, and realization of predicate arguments.

Other formal changes in most Portuguese varieties spoken in Brazil also affected crucial aspects of phrase interpretation. A prominent example in this domain is the meaning of bare singular nominals. In chapter “[Ways of Number Marking: English and Brazilian Portuguese](#),” Roberta Pires de Oliveira critically revises different versions of Chierchia’s nominal parameter on the basis of comparison between English and Brazilian Portuguese, two languages of Type I, according to Chierchia’s parameter.³ A crucial difference between the two languages is that Brazilian Portuguese, but not English, allows for bare singulars in argument position, a fact that seems to contradict one important prediction of the parameter:

- (2) a. Gato tem bigode.
 cat has whisker
 “Cats have whiskers.”
 b. * Cat has whisker.

After showing why previous attempts to solve this problem fail, Pires de Oliveira proposes a microparameter which puts the burden of the explanation in the locus of atomicity. According to the author, English is a noun-centered language, that is, a language in which the first layer of the nominal projection projects atomicity. In turn, Brazilian Portuguese is a determiner-centered language, that is, a language in which atomicity is computed later in the derivation, at the DP level. This implies that a singular bare noun in English is always atomic and cannot be interpreted as an individual in argument position, although it can be re-interpreted as a mass noun through well-known coercion processes. In Brazilian Portuguese, a bare singular is

³ “Type I are languages where numerals combine directly with some nouns, but not with others, which require insertion of a measure phrase. English is an example of Type I, because of the contrast between *three chairs* and **three blood(s)*, *three ounces/drops of blood*.” (Pires de Oliveira [this volume](#)).

underspecified for atomicity and, consequently, can be converted into an individual by particular conditions.

Although the changes that Peninsular Spanish suffered in different areas of America didn't seem to affect the typological nature of the new local grammars, or not at least in the same evident way Brazilian Portuguese varieties show, the innovations imposed by these local grammars are extremely substantive from a phonological, morphological, and syntactic point of view. The two studies on Chilean and Argentinian Spanish contained in Part II of this book analyze different dimensions of the relevant grammars (syntactic and lexical dimensions, respectively). In chapter “[Is Chilean Spanish a Canonical Pro-Drop Variety? On Subjecthood in Chilean Spanish](#),” Iván Ortega-Santos explores the extent to which Chilean Spanish meets the classic properties of the null-subject parameter. Concretely, he claims (i) that Chilean Spanish is not a canonical *pro*-drop language, but (ii) that it is not a partial *pro*-drop language of the Brazilian Portuguese or Finish type, either. On the one hand, using data from corpora as well as from acceptability judgments, Ortega-Santos argues that Chilean Spanish displays some nontrivial overlapping with Caribbean Spanish with respect to *pro*-drop properties, a Latin America variety well-known for not fitting the *pro*-drop pattern consistently. On the other hand, a brief comparison with Brazilian Portuguese, a typical partial *pro*-drop language, shows that Chilean Spanish, even being a non-canonical representative of this linguistic type, still pertains to the family of *pro*-drop languages.

In turn, chapter “[The Grammaticalization of Igual in Argentinean Spanish](#)” by Ángela Lucía Di Tullio, Mercedes Pujalte and Pablo Zdrojewski, focuses on two characteristic uses of the word *igual* “equal” in two Spanish varieties: Argentinean and Peninsular Spanish. As is well-known, among the many functions *igual* can have in the sentence, most of them shared by any Spanish dialect, Argentinean Spanish has also a particular use which can be characterized as concessive:

- (3) Está lloviendo a cántaros. *Igual* voy a tu casa.
 is raining to pitchers equal go.1SG to your house
 “It is raining cats and dogs. But I’ll go to your house.”

[Argentinian Spanish]

In Peninsular Spanish varieties, instead, like in many other Latin-American ones, *igual* can be used as an epistemic marker, an option banned in Argentinian Spanish. Thus, in a sentence like (4), *igual* can be straightforwardly paraphrased (or even replaced) by *quizás* “maybe,” clearly pointing out to the conclusion that the original adjective was reanalyzed as an epistemic adverb expressing uncertainty.

- (4) *Igual* mañana nieva.
 equal tomorrow snows
 “Maybe it will snow tomorrow.”

[Peninsular Spanish]

The authors’ observation is that the epistemic and the concessive uses of *igual* are the result of two different grammaticalization paths, which bifurcated at a stage where *igual* got the value of a focalizing adverb in both varieties. Concretely, both dialects (and related ones) began to bifurcate when they took different grammaticalization paths for exactly the same original adjective. Thus, whereas Peninsular Spanish grammaticalized *igual* from comparative constructions of identical possibility, Argentinean Spanish followed a different path taking as a crucial step the grammaticalization of *igual* as a concessive head.

In sum, the four chapters we offer to the reader in this second part cover lexical, syntactic, and semantic particularities of some of the aforementioned Romance varieties at the south of the *Romania Nova* with the aim (i) of providing a better description of the relevant varieties (in some cases by correcting certain empirical observations in the previous literature and, in others, by adding novel data), and (ii) of offering new theoretical insights which are of great concern in formal approaches to language evolution/change, such as the nature of grammaticalization and the hierarchies of parameters (Di Tullio, Pujalte, and Zdrojewski), the impact that the Case/Agreement system has in the emergence of innovative constructions (Kato and Nunes), the status of the *pro*-drop parameter (Iván Ortega Santos), and the semantic parameter of number marking (Pires de Oliveira).

5 Part III: Indigenous South-American Languages

South America was the last habitable continent to be populated by humans, and its prehistory is wrapped up in mystery. However, although most of the past is just silence, being still lost in the mists of time, recent comparative interdisciplinary investigations have brought light to some of the gaps, incrementing bit by bit our knowledge about the process of peopling the territory. According to the conservative hypothesis, this process started around 11,000 years ago. Following a coastal pacific route, a wave of decedents of North-American Paleoindians – originally from Eurasia – crossed Panama, reaching and settling down in the uppermost west part of South America, expanding later towards east and south (Lynch 1999). This hypothesis is challenged by recent findings that pull back to thousands of years the primal human arrival in the continent (Gruhn 2020; Prates et al. 2020). Particularly, archeological data from the Monte Verde complex site in south-central Chile suggests that humans were in South America before 15 kya (Dillehay et al. 2015, 2019). More dramatically, a recent stone artifact found in the site of Pedra Furada, state of Piauí, Brazil, helped dating the presence of modern *Homo sapiens*

in that place in 24 kya (Boëda et al. 2021). Be this dating as conflicting as it is, it still indicates that indigenous South Americans are ancient local populations, being rightfully our oldest natives.

Anchoring the time-depth of South Amerindians to our current time and interests, consider Quechua and Tupi languages. Historical and linguistic evidence points to Central Peru as the homeland of Proto-Quechua, which coexisted in that territory with Proto-Aymara, maintaining with it close linguistic exchanges (Adelaar 2012a, b; Emlen 2017). Afterwards, Quechua expanded geographically, branching out in two different varieties: Quechua I and II, followed by a subdivision within Quechua II (Heggarty 2005). Thus, although Quechua is often called the language of the Incas, its origins are much older than the rise of Inca Empire, in the Cuzco region, at the beginning of the thirteenth century (Adelaar and Muysken 2004). As for Tupi, convergent information from history, linguistics, genetics, and archeology indicates that Proto-Tupi emerged around 5000 years ago in central-western Amazon, in an inter-fluvial zone bounded by the rivers Amazon, Tocantins, Madeira, and Guaporé, a region within the actual Brazilian state of Rondônia (Rodrigues 1964; Brochado 1984; Noelli 2008; Rodrigues and Cabral 2012, among others). Around 3000 years ago, Tupi diversified in different linguistic varieties, which correspond to the ten families that currently form the Tupi stock. Some of these varieties expanded in a fast radial fashion towards south, north, and east. About 2000 years ago, Guarani, a linguistic subgroup of the Tupi-Guarani family, was already in the lowlands of South America, in the Paraná-Paraguay basin (Urban 1996). Thus, the prehistory of South-American Indians is to be told from a linguistic perspective as well, considering grammatical consequences of ancient migrations and encounters with other languages. This is an understudied topic, with almost no formal investigation being conducted so far.

Some of the remarkable prehistorical migrations within South America left linguistic footprints that can still be traced. Consider, for example, the so-called founder effect. Much like what is observed in genetics, when a subgroup of speakers of a language A spread across a vast territory, reaching unconnected distances from A and founding a new language B, a founder effect is observed: the founders' language (B) exhibits less phonemic variability than the original language (A) (Mayr 1963; Cavalli-Sforza 2001; Atkinson 2011). This is observable within the Tupi stock. Compared to Tupi languages still located in central-western Amazon, South Guarani languages have less variability in their vowel inventories. Almost all Guarani languages have a 6x2 harmonic vocalic system (6 oral, 6 nasal), while Amazonian Tupi languages present among themselves greater vowel variability (Rodrigues 2020). Interestingly reconstruction studies ascribe to Proto-Tupi-Guarani this very same harmonic vowel inventory (Lemle 1971; Rodrigues and Dietrich 1997; Meira and Drude 2015), suggesting that Guarani grammars are very conservative with respect to vowels. Following the CV hypothesis of Nespor and colleagues (Nespor et al. 2003; Hochmann et al. 2011, among others), Rodrigues (2020) associates this finding with the observation that Guarani languages are quite similar in structure (Rodrigues 1985), raising, thus, the hypothesis that there is less

parametric variation among Guaraní languages than among the Tupi languages that stayed within Amazon.

Although $\frac{1}{4}$ of the family languages of the world are in South America, the number of languages in the continent is rather small: there are currently 53 indigenous language families and 55 isolates, but only 420 indigenous languages in total (Campbell 2012). This unbalanced family-language ratio (approximately 6.9 languages per family) is not observed in older continents, such as Africa, Asia, and Europe (Nicols 1990; Nettle 1999; Seifart and Hammarström 2017), raising questions related to the linguistic profile of the South American prehistory. Although these questions haven't been fully addressed yet, they have large implications for theories of language and language evolution. This unbalance might reflect language loss as discussed below, but it might have resulted from territorial expansions as well. Long-distance, disconnecting migrations of speaking populations could have led to severe phonemic founder effects, which, in its turn, brought about linguistic differentiations first at the phonemic level then at the structural level, creating, thus, multiple speciation within one single linguistic lineage. Obviously, verification of this hypothesis requires a better understanding of how phonemic changes affect grammar.

Events of peopling and migrations might have shaped language diversity in prehistory South America in unexpected ways; and in addition to this, the recent past and the present indicate reduction of diversity. The European colonization took its tolls, reducing dramatically the number of local languages in the whole territory (Kaufman 1994; Crevels 2012). For example, the Charrúa language that flourished in the south of Uruguay in the pre-Columbian period died during the first half of the nineteenth century due to dispossession of their speakers by the Spanish founding fathers of modern Uruguay (Kaufman 1994). The Brazilian territory lost 75% of its indigenous languages during and after Portuguese colonization (Rodrigues 2014), and, according to information on the 2021 edition of the *ethnologue website* (<https://www.ethnologue.com/guides/how-many-languages-endangered>), 62% of Brazilian living languages are endangered. This is not the highest percentage, as it reaches 86% in Chile. From our point of view, this is an alarming situation. Most of these languages are still poorly documented, underdescribed and underanalyzed. Also, although in the last decades there has been a significant increase in linguistic awareness and recognition of indigenous languages as full-fledged complex grammars, accompanied by an increment on local research, it is to be emphasized that the bulk of the investigations conducted so far are typological and descriptive, with formal analyses being quite limited for now.

This scarcity of systematic and deep analyses is in tension with the fact that important considerations about language and cognition are quite often based on empirical evidence coming from understudied languages. To exemplify this, consider Gordon's (2004) work on Pirahã (Mura language, Brazilian Amazon) and Pica et al.'s (2004) work on Mundurucu (Tupi language, Brazilian Amazon). Their experimental observations led to the conclusion that language underlies numerical cognition, supporting cognitive tasks involved in exact arithmetic. This showcases the importance of promoting detailed investigations on understudied languages,

while calling attention to the potential of South-American languages for research on human neurocognition.

The present publication undertakes efforts to promote formal studies on South-American indigenous languages, aiming at adding to a better understanding of local grammars and to the development of a general theory of human language. The four chapters that compose the third part of the book are devoted to analytical studies of Quechua, Paraguayan Guarani, A'ingae, and Macro-Jê languages, addressing different levels of analysis.

Chapter “[Compounding Processes in Three Macro-Jê Linguistic Branches](#)” elaborated by Andrew Nevins and Mário André Coelho da Silva, centers on compoundhood in Macro-Jê, a stock located in the south of Amazon River, mostly within the Brazilian territory, but also, in smaller numbers, in Argentina and Paraguay (Nikulin 2020). Although classifications of Macro-Jê are uncertain, the stock comprises approximately 10 families and 40 languages, including those already extinct (Rodrigues 1999; Nikulin 2020). Nevins and Coelho da Silva study aims at identifying grammar-specific conditions on compound formation in four Macro-Jê extant languages: Maxakalí (Maxakalí family), Krenak (Krenak family), Xerente, and Xavante (Jê family). Although in these grammars, compounds are more productive than derivational mechanisms in word formation, Nevins and Coelho da Silva show that, due to language-specific conditions, different criteria are needed to identify compoundhood. In Maxakalí, phonological criteria, such as lexical stress and number of syllables, might not be effective in identifying compounds. Thus, semantic properties related to reference must be considered. Conversely, compounds in Krenak are easily identified via phonological criteria, such as place of assimilation and nasal-voice interaction, where a nasal feature in the last segment of the first root triggers devoicing in the first segment of the following root. In contrast, noun form might be a good criterion for detecting compoundhood in Xerente and Xavante, as these languages have two forms of nouns, the first one restricted to the rightmost position (first root) within a compound and the second occurring everywhere, but not in the last position.

Paraguayan Guarani, a Tupi-Guarani language, is the focus of chapter “[Poro-/mba'e- Antipassive Prefixation in Paraguayan Guarani](#).” Bruno Estigarribia and Ernesto Luiz López Almada argue, against typological previous analyses (e.g., Heaton 2017, 2020), that Modern Paraguayan Guarani uses the prefixes *poro* and *mba'e* as antipassive voice markers rather than incorporated nouns. The proposal is that insertion of these prefixes yields syntactically intransitive argument structures that are semantically dyadic with a nonspecific, generic internal argument. The following pieces of evidence are given in favor of antipassivization and against noun-incorporation: (a) noun-incorporation does not explain the affix *poro*, which has no corresponding lexicalized noun; (b) the noun *mbá'e* (*thing*) might incorporate into the predicate, but, if so, it triggers a reading of the predicate as an institutionalized and socially important activity, which is not observed in the data at hand; and (c) while structures with noun-incorporation allow for a manifested object, structures with *poro* and *mba'é* do not.

In our assessment, chapters “[Compounding Processes in Three Macro-Jê Linguistic Branches](#)” and “[Poró-/mba’e- Antipassive Prefixation in Paraguayan Guarani](#)” are very good theoretical exercises. Nevins and Coelho da Silva show the need for establishing general, non-language-specific criteria for identification and characterization of word-formation processes being at the same time sensitive to language-specific properties in deciding which criterion is to be applied to which language. Estigarribia and López Almada show that separating “apples from oranges” is a crucial step at descriptive level. They do not provide us with the structure of Paraguayan Guarani antipassives, but give us clean data, contributing, thus, from a cross-linguistic perspective, for formulation of a general morphosyntactic theory of antipassivization.

Neil Myler, the author of chapter “[Argument Structure and Morphology in Cochabamba Quechua \(with Occasional Comparison with Other Quechua Varieties\)](#)”, is also interested in the morphosyntactic properties of argument structure, but his research language is Quechua, in particular the Cochabamba variety (variety of Quechua IIC, Bolivia), and his focus is on causative, reflexive, and applicative morphemes. Working on the interface between syntax and morphology, Myler concludes that (a) the reflexive marker *ku*, a portmanteau morpheme, is better analyzed as a pronominal clitic merged in an argument position, reducing the predicate valency before adjoining to Voice⁰ via head movement; (b) the morpheme *pu* is the head of high applicative phrasal that does not project arguments, but serves as a host for displaced oblique arguments, and (c) *chi* is a vP-selecting causative functional element.

The last chapter “[Definiteness in A’ingae and Its Implications for Pragmatic Competition](#)” is dedicated to A’ingae, also known as Cofán, an isolate language spoken in Ecuador Colombia. In both countries, A’ingae is endangered, having a very low number of speakers, but in Colombian it is in greater danger due to a lack of protective language policies (Dabkowski 2021). The authors of chapter “[Definiteness in A’ingae and Its Implications for Pragmatic Competition](#)”, Holly Zheng and Scott AnderBois, analyze the semantics of definiteness based on fieldwork data from A’ingae spoken in Ecuador. In this language, bare nouns are freely available to encode indefiniteness and unique and anaphoric definiteness, despite it having dedicated indefinite anaphoric definite forms, *fae* and *tsa*, respectively, with *tsa* being a stronger exclusive definite anaphoric form. This challenges leading accounts for the distribution of definite forms based on hard pragmatic competitions, such as *Maximize Presupposition!*, developed by Schwarz (2013) for German, and *Index!*, developed by Jenks (2018) for Mandarin. Assuming these pragmatic principles to be universal, it is predicted that in any language with dedicated forms, complementary distribution will be observed, with the anaphoric definite form being used whenever possible. After demonstrating that A’ingae’s non-complementary distribution between bare nouns and *tsa* in the expression of anaphoric definiteness counterexamples the prediction of competition principles put forward in the literature, Zheng and AnderBois propose a semantic-based analysis, suggesting that A’ingae bare singulars have an existential component, but lack an anti-uniqueness condition. As such, they have no presupposition, being felicitous as indefinite as

well as in unique and anaphoric definite contexts. Anaphoric *tsa*, on the other hand, is semantically specified for uniqueness and familiarity, being, thus, restricted to contexts where a unique and familiar referent is provided.

Chapters “Argument Structure and Morphology in Cochabamba Quechua (with Occasional Comparison with Other Quechua Varieties)” and “Definiteness in A’ingae and Its Implications for Pragmatic Competition” highlight the importance of comparative research involving less-studied languages. Myler’s chapter provides us with an analysis of morphosyntax microparametric variations at the argument structural level. Centering on Cochabamba Quechua, it promotes comparisons with other Quechua varieties such as Tarma (variety of Quechua I, center of Peru), Cajamarca (variety of Quechua IIA, northern of Peru), and Santiago del Estero (variety of Quechua IIC, Argentina). By doing so, Myler’s analysis adds to the general syntax of argument structure, especially his observations on how the c-selection properties of functional heads interact and determine the final structure of a verbal thematic domain. Zheng and AnderBois’ chapter focuses on macroparametric variations on the expression of definiteness, showing how comparative analyses are important for revising and tuning theoretical hypothesis. Very often grammatical principles are put forward on the basis of evidence from well-known languages, because we know more about these languages, but, as Zheng and AnderBois fully demonstrate, these principles are to be carefully tested in other languages. This is the real gist of comparative work.

6 Conclusions

The following chapters offer an overview of a subset of the different grammatical systems found in South America, emphasizing research questions important to the present stage of linguistic theory, while raising issues to be addressed and better understood in a near future. Also, the content of this book, particularly Parts I and II, invites us to revise the expressions *Native languages of South America*. All the languages currently spoken in the continent are native languages. Their synchronic grammars resulted from changes that occurred here, in the south hemisphere, some in response to external forces such as language contact and language expansions. As Jorge Drexler, a Uruguayan singer and composer, beautifully puts it in his song *Movimiento*:

Yo no soy de aquí

Pero tú tampoco

De ningún lado del todo

De todos lados un poco

Products of science are restricted by time, space, current interests, and what is available at a given moment. Thus, as emphasized above, this book does not offer a complete picture of South America linguistic diversity. For instance, the so-called immigrant dialects (e.g., Italian, German, Finnish, and Welsh), French varieties spoken in French Guiana, and languages of quilombola communities were

not represented in this volume. While acknowledging these limitations, we take them as a motivation for continuing our work.

References

- Adelaar, Willem F.H. 2012a. Historical overview: Descriptive and comparative research on south American Indian languages. In *The indigenous languages of South America: A comprehensive guide*, ed. Lyle Campbell and Verónica Grondona, 1–58. Berlin/Boston: Walter de Gruyter.
- . 2012b. Modeling convergence: Towards a reconstruction of the history of Quechuan-Aymaran interaction. *Lingua* 122: 461–469.
- Adelaar, Willem F.H., and Pieter C. Muysken. 2004. *The languages of the Andes*. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization*. Oxford: Oxford University Press.
- . 2005. Classifiers and noun classes: Semantics. In *Encyclopedia of languages and linguistics*, ed. Keith Brown, 463–471. Oxford: Elsevier.
- Almeida-Silva, Anderson, and Andrew Ira Nevins. 2020. Observações sobre a estrutura linguística da Cena: a língua de sinais emergente de Várzea Queimada (Piauí, Brasil). *Linguagem & Ensino - Revista do Programa de Pós-graduação em Letras da Universidade Federal de Pelotas* 23 (4): 1029–1053.
- Atkinson, Quentin D. 2011. Phonemic diversity supports a founder effect model of language expansion from Africa. *Science* 332: 346–349.
- Baker, Mark. 2008. The macroparameter in a microparametric world. In *The limits of syntactic variation*, ed. Theresa Biberauer, 351–373. Amsterdam: John Benjamins.
- Barbosa, Pilar, Maria Eugênia L. Duarte, and Mary A. Kato. 2005. Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 4: 11–52.
- Bellugi, Ursula, Howard Poizner, and Edward Klima. 1983. Brain organization for language: Clues from sign aphasia. *Human Neurobiology* 2 (3): 155–170.
- Benedicto, Elena, and Diana Brentari. 2004. Where did all the arguments go? Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory* 22 (4): 743–810.
- Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
- Boëda, Eric, Marcos Ramos, Antonio Pérez, Christine Hatté, Christelle Lahaye, Mario Pino, David Hérisson, Ignacio Clemente-Conte, Michel Fontugne, Guillaume Guérin, Ximen Villagran, Janaina C. Santos, Lucas Costa, Lucie Germond, Gisele Ahmed, Nelson E. Delacroix, Amelie Da Costa, Carolina Borges, Sirley Hoeltz, María Felice Gluchy, Grégoire van Havre, Christophe Griggo, Livia Lucas, Iderlan de Souza, Sibeli Viana, André Strauss, Jennifer Kerner, and Niéde Guidon. 2021. 24.0 kyr cal BP stone artefact from Vale da Pedra Furada, Piauí, Brazil: Techno-functional analysis. *PLoS ONE* 16 (3): e0247965.
- Brentari, Diana. 1998. *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.
- Brochado, José P. 1984. *An ecological model of the spread of pottery and agriculture into eastern South America*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Campbell, Lyle. 2012. Classification of the indigenous languages of South America. In *The indigenous languages of South America: a comprehensive guide*, ed. Lyle Campbell and Verónica Grondona, 59–166. Berlin/New York: Mouton/de Gruyter.
- Cavalli-Sforza, Luigi L. 2001. *Genes, people and language*. Los Angeles: University of California Press.
- Chomsky, Noam. 1965. *Aspect of the theory of syntax*. Cambridge, MA: MIT Press.
- . 1986. *Knowledge of language: Its nature, origins, and use*. New York: Praeger.
- . 1995. *The minimalist program*. Cambridge, MA: MIT Press.

- Clark, Brenda. 2017. Sign language varieties in Lima, Peru. *Sign Language Studies* 17 (2): 222–264.
- Crevels, Mily. 2012. Language endangerment in South America: The clock is ticking. In *The indigenous languages of South America: A comprehensive guide*, ed. Lyle Campbell, Verónica Grondona, and Verónica, 167–233. Berlin/Boston: Walter de Gruyter.
- Dabkowski, Maksymilian. 2021. A'ingae (Ecuador and Colombia) – Language snapshot. *Language Documentation and Description* 20: 1–12.
- de Quadros, Ronice Müller, ed. 2020. *Brazilian sign language studies*. Boston/Berlin/Lancaster: De Gruyter Mouton & Ishara Press.
- de Quadros, Ronice Müller, Diane Lillo-Martin, and Deborah Chen Pichler. 2020. Blending languages: Bimodal bilinguals and languages synthesis. In *Sign multilingualism*, ed. Ulrike Zeshan, Jennifer Webster, and Anastasia Bradford. Boston/Berlin/Lancaster: De Gruyter Mouton & Ishara Press.
- Dillehay, Tom D., Carlos Ocampo, Jose Saavedra, Andre O. Sawakuchi, Rodrigo Vega, Mario M. Pino, Michael P. Collins, Linda S. Cummings, Iván Arregui, Ximena S. Villagran, Gelvam A. Hartmann, Mauricio Mella, Andrea Gonzáles, and Goerge Dix. 2015. New archeological evidence for an early human presence at Monte Verde, Chile. *PLoS ONE* 10 (11): e0141923. <https://doi.org/10.1371/journal.pone.0141923>.
- Dillehay, Tom D., Carlos Ocampo, Jose Saavedra, Mario Pino, Linda S. Cummings, Peter Kováčik, Claudia Silvag, and Rodrigo Alvarh. 2019. New excavations at the late Pleistocene site of Chinchihuapi I, Chile. *Quaternary Research* 92: 70. <https://doi.org/10.1017/qua.2018.145>.
- Duarte, Maria Eugênia L. 1993. Do pronomes nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In *Português Brasileiro: Uma viagem diacrônica (Homenagem a Fernando Tarallo)*, ed. Ian Roberts and Mary A. Kato, 107–128. Campinas: Editora da UNICAMP.
- . 2000. The loss of the ‘avoid pronoun’ principle in Brazilian Portuguese. In *The null subject parameter in Brazilian Portuguese*, ed. Mary A. Kato and Esmeralda V. Negrão, 17–36. Frankfurt-Madrid: Vervuert-Iberoamericana.
- Emlen, Nicholas. 2017. Perspectives on the Quechua-Aymara contact relationship and the lexicon and phonology of pre-Proto-Aymara. *International Journal of American Linguistics* 83 (2): 307–340.
- Emmorey, Karen. 2003. *Perspective on classifiers constructions in sign language*. Mahwah: Lawrence Erlbaum Associates.
- Estigarribia, Bruno. 2020. *A grammar of Paraguayan Guarani*. London: UCL Press.
- Ferreira, Marcelo. 2000. *Argumentos nulos em Português Brasileiro*. Master Thesis, UNICAMP.
- Friedman, Lynn. 1977. Formational properties of American sign language. In *On the other hand: New perspectives in ASL*, ed. Lynn Friedman, 13–56. New York: Academic Press.
- Galves, Charlotte. 2020. Mudança sintática no português brasileiro. *Cuadernos de la ALFAL* 12 (2): 17–43.
- Goldin-Meadow, Susan, and Diana Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Science* 40: e46. <https://doi.org/10.1017/S0140525X15001247>.
- Gordon, Peter. 2004. Numerical cognition without words: Evidence from Amazon. *Science* 306 (5695): 496–499.
- Gruhn, Ruth. 2020. Evidence grows that peopling of the Americas began more than 20,000 years ago. *Nature* 584: 47–48.
- Hale, Kenneth, and Samuel J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The view from building 20*, ed. Kenneth Hale and Samuel J. Keyser, 53–110. Cambridge, MA: The MIT Press.
- Heaton, Raina. 2017. *A typology of antipassives, with special reference to Mayan*. Ph.D. Dissertation. University of Hawai'i at Mānoa.
- . 2020. Antipassives in crosslinguistic perspective. *Annual Review of Linguistics* 6: 131–153.
- Heggarty, Paul. 2005. Enigmas em el origem de las lenguas andimas: aplicando nuevas técnicas a las incógnitas por resolver. *Revista Andina* 40: 9–80.

- Hochmann, Jena-Rémi, Silvia Benavides-Varela, Marina Nespor, and Jacques Mehler. 2011. Consonants and vowels: Different roles in early language acquisition. *Developmental Science* 14: 1467–1687.
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203: 88–111.
- Holmberg, Anders. 2005. Is there little Pro? Evidence from Finnish. *Linguistic Inquiry* 36: 533–564.
- . 2010. The null generic subject pronoun in Finnish: A case of incorporation in T. In *Parametric variation: Null subjects in minimalist theory*, ed. T. Biberauer, A. Holmberg, I. Roberts, and M. Sheehan, 200–230. Cambridge, UK: Cambridge University Press.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49: 501–536.
- Kato, Mary A. 1999. Strong and weak pronominals in the null subject parameter. *Probus* 11: 137.
- . 2000. The partial *pro*-drop nature and the restricted VS order in Brazilian Portuguese. In *Brazilian Portuguese and the null subject parameter*, ed. Mary A. Kato and Esmeralda Negão, 223–258. Madrid: Vervuert-Iberoamericana.
- Kaufman, Terrence. 1994. The native languages of South America. In *Atlas of the world's languages*, ed. Christopher Moseley and Ronald E. Asher, 46–76. London: Routledge.
- Kayne, Richard. 2005. Some notes on comparative syntax, with special reference to English and French. In *The Oxford handbook of comparative syntax*, ed. Guglielmo Cinque and Richard Kayne, 3–69. New York: Oxford University Press.
- Kendon, Adam. 2008. Some reflections on the relationship between “gesture” and sign. *Gesture* 8 (3): 348–366.
- Klima, Edward, and Ursula Bellugi. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.
- Lemle, Miriam. 1971. Internal classification of the Tupi-Guarani linguistic family. In *Tupi studies*, ed. David Bendor-Samuel, 107–129. Norman: Summer Institute of Linguistics Publications.
- Liddell, Scott. 1980. *American sign language syntax*. New York: Mouton.
- Lillo-Martin, Diane, and Richard P. Meier. 2011. On the linguistic status of ‘agreement’ in sign language. *Theoretical Linguistics* 37 (3/4): 95–141.
- Lillo-Martin, Diane, Ronice Müller de Quadros, and Deborah Chen Pichler. 2016. The development of bimodal bilingualism: Implications for linguistic theory. *Linguistic Approach to Bilingualism* 6 (6): 719–755.
- Lynch, Thomas F. 1999. The earliest South American lifeways. In *The Cambridge history of the native peoples of the Americas. Volume III: South America. Part I*, ed. Frank Salomon and Stuart B. Schwartz, 188–263. Cambridge: Cambridge University Press.
- Massone, María Ignacia, and Rocío Anabel Martínez. 2015. Argentine sign language. In *Sign languages of the world. A comparative handbook*, ed. Julie Bakken Jepsen, Goedele De Clerck Sam Lutalo-Kiingi, and William B. McGregor, 71–104. Boston/Berlin/Lancaster: De Gruyter Mouton & Ishara Press.
- Mayr, Ernst. 1963. *Animal species and evolution*. Cambridge, MA: Harvard University Press.
- Meir, Irit, Wendy Sandler, Carol Padden, and Mark Aronoff. 2003. Emerging sign languages. In *Oxford handbook of deaf studies, language, and education*, ed. Marc Marschark and Patricia Elizabeth Spencer, vol. 2, 267–280. Oxford: Oxford University Press.
- Meira, Sérgio, and Sebastian Drude. 2015. A summary reconstruction of Proto Mawetí-Guarani segmental phonology. *Boletim do Museu Paraense Emílio Goeldi* 10: 275–269.
- Messing, Lynn S., and Ruth Campbell, eds. 1999. *Gesture, speech and sign*. Oxford: Oxford University Press.
- Meteyard, Lotte, Emily Stoppard, Stefano Cappa, and Gabriella Vigliocco. 2015. When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia* 76: 264–274.
- Modesto, Marcello. 2000. *Null subjects without ‘rich’ agreement*. PhD. Dissertation, University of Southern California.
- Muñoz Pérez, Carlos, and Andrés Saab. in prep. American Spanish dialects: Chile, River Plate, Paraguay. In *A guide to Spanish Dialects: Descriptive and theoretical aspects of linguistic*

- variation in the Hispanic world*, ed. Ángel J. Gallego and Cristina Sánchez López. Oxford: Oxford University Press.
- Nespor, Marina, Marcela Peña, and Jacques Mehler. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2: 2003–2229.
- Nettle, Daniel. 1999. Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proceedings. National Academy of Sciences. United States of America* 96 (6): 3325–3329.
- Nicols, Johanna. 1990. Linguistic diversity and the first settlement of the new world. *Language* 66 (3): 475–521.
- Nikulin, Andrey. 2020. *Proto-Macro-Jê: Um estudo reconstrutivo*. Doctoral thesis. Brasília: Universidade de Brasília.
- Noelli, Francisco S. 2008. The Tupi expansion. In *Handbook of South American archaeology*, ed. Helaine Silverman and Isbell H. William, 659–670. New York: Springer.
- Nunes, Jairo. 2020. Hyperalçamento em português brasileiro. *Cuadernos de la ALFAL* 12 (2): 199–227.
- Padden, Carol A. 1988. *Interaction of morphology and syntax in American sign language*. New York: Garland Publishing.
- Parks, Elizabeth, and Jason Parks. 2010. A sociolinguist profile of the Peruvian deaf community. *Sign Language Studies* 10 (4): 409–441.
- Perniss, Pamela, Robin L. Thompson, and Gabriella Vigliocco. 2010. Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology* 1. <https://doi.org/10.3389/fpsyg.2010.00227>.
- Petitto, Laura Ann. 1997. In the beginning: On the genetic and environmental factors that make early language acquisition possible. In *The inheritance and innateness of grammar*, ed. Myrna Gopnik, 45–69. Oxford: Oxford University Press.
- Pica, Pierre, Cathy Lemer, Véronique Izard, and Stanislas Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306 (5695): 499–503.
- Pires de Oliveira, Roberta. this volume. Chapter 9: Ways of number marking: English and Brazilian Portuguese. In *Formal approaches to languages of South America*, ed. Cilene Rodrigues and Andrés Saab. Cham: Springer.
- Poizner, Howard, Edward Klima, and Ursula Bellugi. 1987. *What the hands reveal about the brain*. Cambridge: The MIT Press.
- Prates, Luciano, Gustavo G. Politis, and Ivan S. Perez. 2020. Rapid radiation of humans in South America after the last glacial maximum: A radiocarbon-based study. *PLoS One* 15 (7): 1–22.
- Quer, Josep. 2021. Verb agreement: Theoretical perspective. In *The Routledge handbook of theoretical and experimental sign language research*, ed. Josep Quer, Roland Pfau, and Anika Herrmann, 95–121. London/New York: Routledge Taylor & Francis Group.
- Ramsey, Claire, and David Quinto-Pozos. 2010. Transmission of sign languages in Latin America. In *Sign languages*, ed. Diane Brentari, 46–73. Cambridge: Cambridge University Press.
- Roberts, Ian. 2019. *Parameters hierarchies and universal grammar*. Oxford: Oxford University Press.
- Rodrigues, Aryon D. 1964. Classificação do tronco linguístico Tupi. *Revista de Antropologia* 12: 99–104.
- . 1985. Relações internas na família linguística Tupí-Guaraní. *Revista de Antropologia* 27: 33–53.
- . 1999. Macro-Jê. In *Amazonian languages*, ed. Robert M.W. Dixon and Alexandra Y. Aikhenvald, 165–206. Cambridge: Cambridge University Press.
- Rodrigues, Cilene. 2004. *Impoverished morphology and A-movement out of Case domains*. PhD. Dissertation, Universidad de Maryland.
- Rodrigues, Aryon D. 2014. Endangered languages in Brazil. *DELTA* 30: 447–463.
- Rodrigues, Cilene. 2020. Founder effect in Tupian languages. *Revista Diadorim* 22: 65–97.
- Rodrigues, Aryon D., and Ana S. Cabral. 2012. Tupian. In *The indigenous languages of South America: A comprehensive guide*, ed. Lyle Campbell and Veronica Grondona, 495–574. Berlin/New York: Mouton/de Gruyter.

- Rodrigues, Aryon D., and Wolf Dietrich. 1997. On the linguistic relationship between Mawé and Tupi-Guarani. *Diachronica* 14: 265–304.
- Rodríguez-Mondoñedo, Miguel. this volume. Chapter 4: Argument structure in Peruvian sign language. In *Formal approaches to languages of South America*, ed. Cilene Rodrigues and Andrés Saab. Cham: Springer.
- Sandler, Wendy, and Diane Lillo-Martin. 2006. *Sign languages and linguistic universals*. Cambridge: Cambridge University Press.
- Sandler, Wendy, Mark Aronoff, Carol Padden, and Meir Irit. 2014. Language emergence: Al-Sayyid Bedouin sign language. In *The Cambridge handbook of linguistic anthropology*, ed. N.J. Enfield, Paul Kockelman, and Jack Sidnell, 250–284. Cambridge: Cambridge University Press.
- Schlenker, Phillippe. 2022. *What it all means: Semantics for (almost) everything*. Cambridge, MA: The MIT Press.
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistic Compass* 7 (10): 534–559.
- Seifart, Frank, and Harard Harmmarström. 2017. Languages isolates in South America. In *Languages isolates*, ed. Lyle Campbell, Alex Smith, and Thomas Dougherty, 260–286. London: Routledge.
- Silva, Diná Souza, and Ronice Müller De Quadros. 2019. Línguas de sinais de comunidades isoladas encontradas no Brasil/Sign languages of isolated communities found in Brazil. *Brazilian Journal of Development* 5 (10): 22111–22127.
- Simpson, Andrew. 2021. Revisiting the structure of nominals in Japanese and Korean: Mixed headedness vs. pure head-finality. *Natural Language & Linguistic Theory* 40: 573–597.
- Stokoe, William C. 1960. *Sign language structure. An outline of the visual communication system of the American Deaf*, Studies in linguistics: Occasional papers 8. Buffalo: University of Buffalo Press.
- Urban, Greg. 1996. On the geographical origins and dispersions of Tupian languages. *Revista de Antropologia* 39: 61–104.
- Watanabe, Akira. 2006. Functional projections for nominals in Japanese: Syntax of classifiers. *Natural Language & Linguistic Theory* 24: 241–306.

Part I
South-American Sign Languages

The Morpho-phonology of Nominal Plurality in Argentinian Sign Language (LSA)



Yanina Boria and Carlos Muñoz Pérez

1 Introduction

Reduplication is a well-attested pluralization strategy in sign languages. Our focus in this chapter is the functioning of this strategy in Argentinian Sign Language (LSA), the natural language of the deaf community of Argentina. As originally pointed out by Massone (1996), nominal plurals are expressed in the language through repetitions of an arched movement. This behavior is attested in the pair of examples in (1) and (2) corresponding, respectively, to the singular and plural forms of the noun CHILD. As can be seen in (2), the plural CHILDREN is realized as a series of short arches that altogether compose a movement towards the ipsilateral side. Each of these arches seemingly “repeats” the noun CHILD.

(1)



CHILD.SG

Y. Boria
Universidad de Buenos Aires, Buenos Aires, Argentina

C. Muñoz Pérez (✉)
Universidad Austral de Chile, Valdivia, Chile
e-mail: carlos.munozperez@uc.cl

(2)



CHILD.PL

The phenomena associated with the exteriorization of nominal plurality in LSA are more complex than this general description suggests. As the example in (2) shows, the plural in this case is expressed over the lexeme, i.e., the noun CHILD is “repeated” to convey a plural interpretation. This form of externalization sharply contrasts with cases in which plurality is expressed over a *classifier* (CL), i.e., a type of sign denoting prototypical characteristics of a nominal referent (Supalla 1986; Zwitserlood 2012). Consider the pair in (3) and (4). The example in (3) corresponds to the singular form of the noun PENCIL. As can be observed, this noun is articulated in the chin area.

(3)



PENCIL.SG

In contrast to CHILDREN in (2), the nominal base of the noun PENCILS does not “carry” the plural, but a classifier does. That is, the reduplication pattern is expressed on a classifier rather than on the noun, e.g., (4). The introduction of the classifier form in (4) does not seem to have any additional semantic effect, i.e., (3) and (4) form a minimal pair contrasting singular and plural features.

(4)



PENCIL CL.PL

In principle, these data seem to show that the plural morpheme in LSA has two (or more) allomorphic variants. This line of analysis has been advanced for similar alternations in nominal plurality in German Sign Language (DGS) by Pfau and Steinbach (2005, 2006). In short, these authors argue that plural allomorphy in DGS is constrained by phonological properties of the underlying nominal base.

In this chapter, we advance a rather different characterization for the LSA data. While building on Pfau and Steinbach's observations, we claim that the LSA patterns in (1) to (4) are not the product of an allomorphic alternation. Instead, we argue that LSA has a single plural morph. This element is an affix that needs to be combined with a nominal base complying with certain phonological requirements. In scenarios in which these requirements are not met, a phonological mechanism that we call *handshape insertion* provides a host for the affix in the form of a classifier. The proposal is framed within a general analysis of the exponence of nominal plurality in LSA.

The chapter is structured as follows. In Sect. 2, we describe what we take to be the underlying phonological representation of the plural morpheme in LSA. As part of our proposal, we introduce the hypotheses that (i) certain properties of the plural exponent are epenthetic in nature and (ii) the classifier attested in examples such as (4) is introduced in the phonological representation as a proper host for the plural affix. Section 3 presents a typology of nouns in LSA; the resulting classes are based on the phonological properties determining whether the plural affix can be attached to a noun. In Sect. 4, we explore some predictions of the hypothesis in (i). Section 5 contains the conclusions.

This work is framed in the perspective that studying the grammar of sign languages contributes to the visibility and valorization of its users. The research and data collection were conducted with and within the Deaf community.

2 The Phonological Representation of Plurality in LSA

As pointed out by Pfau and Steinbach (2005, 2006) and Steinbach (2012), sign languages may convey (nominal) plurality in different ways. Reduplication is one

of these strategies, but even reduplication comes in distinct variants. On the one hand, *simple reduplication* involves the repetition of a movement once and in the same place. On the other, *sideways reduplication* (or “triplcation”) involves (i) the realization of a certain sign and (ii) its repetition two extra times through arc-like movements. These repetitions must be one next to the other in slightly different places; thus, sideways reduplication always requires displacement in space. As the examples in (2) and (4) attempt to show, LSA displays sideways reduplication; this strategy is the only morpho-syntactic means to convey nominal plurality in the language. See Massone (1996: 274), Valassina (1997: xxvi), and Massone and Martínez (2012: 1–2) for preliminary descriptions.

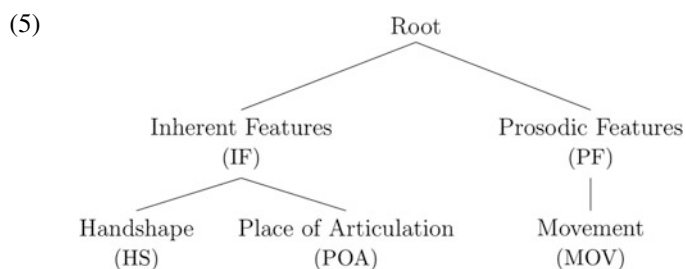
In this section, we aim at providing an explicit characterization of the exponence of nominal plurality in LSA. To do so, we first introduce some basic notions of the *Prosodic Model* (Brentari 1998), a specific theory of sign language phonology in which we will frame our discussion.

Sign languages differ from oral languages in the perceptual channel through which they are expressed: viso-gestual and auditory-oral, respectively. A characteristic feature of the viso-gestual modality are the elements from which the phonetic material is composed, e.g., hands, arms, torso, head, eyes, eyebrows, mouth, and cheeks. These are the articulators responsible for the phonological primitives of sign languages. Traditional literature organizes these phonological features into five major groups: (i) *configuration* (i.e., the shape of the hand), (ii) *location* (i.e., the body area in which the sign is articulated), (iii) *orientation* (of the palm with respect to the location), (iv) *movement* (i.e., how the hand moves with respect to the location), and (v) *non-manual features*; see Stokoe (1960), Brentari (2012), and Sandler (2012), among many others, for relevant discussion.

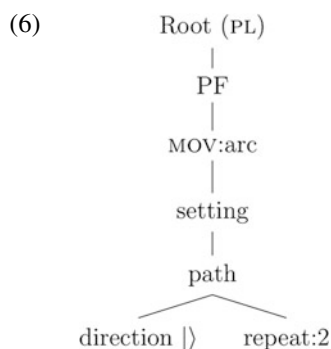
The *Prosodic Model* of sign language phonology (Brentari 1998; Fenlon et al. 2018) takes these primitives and organizes them into structured representations of features that follow Dependency Theory (Anderson and Ewen 1987; van der Hulst 1993). These are binary branching hierarchical structures that organize phonological features into segmental units. Within the *Prosodic Model*, these representations allow us to describe lexical units in terms of combinations of phonological properties. This is possible in sign languages because (i) the phonological unit resulting from this analysis is equivalent to a syllable in spoken languages, and (ii) most signs are monosyllabic.

According to the *Prosodic Model*, phonological features combine to form a *root*, which is roughly the phonological representation of a monosyllabic word. A root branches into two main types of features: *Inherent Features* (IF) and *Prosodic Features* (PF). This distinction allows us to capture the parallel between roots and syllables. The PF branch specifies sequential articulations of features that result in movement; within the root, movement plays a function similar to that of syllabic nucleus in spoken languages. The IF branch, on the other hand, specifies static properties of the active and passive articulators, i.e., the *Handshape* (HS) and the *Place of articulation* (POA), respectively. In this model, *Orientation* is derived from the relation between HS and POA. We will attempt to avoid more specific and complex parts of the theory; further aspects of the system will be introduced

as they become relevant. The general organization of the major classes of features according to the Prosodic Model is schematized in (5).



Given this framework, we propose an analysis of the plural morpheme PL in LSA in which the phonological exponent of this element consists of a set of Prosodic Features. This means that there are no Inherent Features (i.e., no Handshape and no Place of Articulation) that are part of the lexical entry of PL. Its underlying phonological representation only specifies (i) that the relevant movement must be in the shape of an arc, (ii) that this movement must be repeated twice, and (iii) that the repetitions proceed in some direction (i.e., that they do not take place in exactly the same location). Note that this specification of features does not specify why this type of reduplication is *sideways*, i.e., towards the ipsilateral side. We will return to this issue below. In sum, the phonological representation we assume for PL is sketched in (6).



Since the plural morpheme lacks any Inherent Features, it follows that it cannot be exteriorized by itself. This observation is rather natural: a well-formed sign cannot consist of movement alone; there must be a certain specification of the properties of the active articulator undergoing the movement, i.e., the hand must have some configuration. Under this observation, the representation in (6) entails that the exponent of the plural morpheme must be phonologically attached to a root providing a Handshape in order to be spelled out. In other words, PL is an affix in need of an appropriate (nominal) host.

We contend that there are phonological conditions constraining the set of nouns to which the plural morpheme can attach. That is, there are cases in which the phonological properties of a nominal base N “clash” with the phonological properties attributed to PL in (6). We consider two types of scenarios in which this happens. First, a noun N having a fixed Place of Articulation (POA) cannot undergo sideways reduplication, e.g., if N is articulated at the top of the head, the signer cannot execute arc-like movements with her head towards the ipsilateral side. Second, a noun N cannot undergo sideways reduplication if N has Prosodic Features of its own specifying a movement with a different shape or direction, e.g., if N already involves a thrilled movement, it cannot also undergo an arc-like movement. In short, the phonological properties of the noun need to be compatible with those of the affix for affixation to take place.

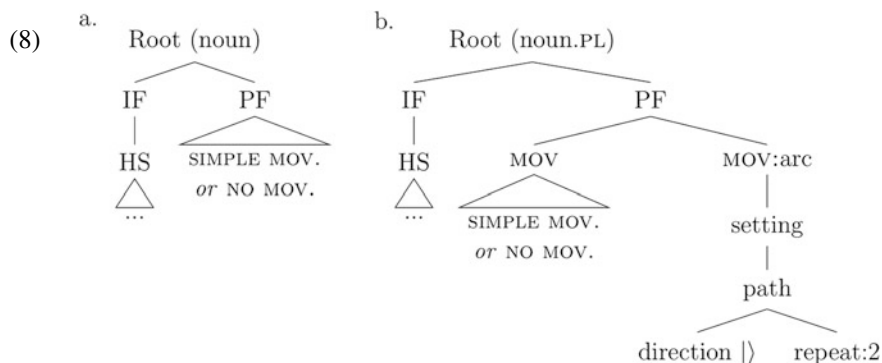
As we will see below, nouns with very partial specifications of Prosodic Features can host the plural affix. Adapting Pfau and Steinbach’s (2005, 2006) terminology, we will refer to these nouns as having *simple movement*. As Brentari (1998) puts it, simple movements involve PF tiers that branch only once, i.e., there is a single feature specifying the properties of the movement; *complex movements*, on the other hand, have two or more branching nodes specifying further characteristics of the movement. At the phonetic level, simple movements involve a single path movement or local change, while complex movements involve two or more co-occurring path or local (sub)movements (Brentari 1998: 130). This distinction is one of metrical structure: roots with simple movements are equivalent to “light” syllables, while roots with complex movements are equivalent to “heavy” syllables. Thus, the claim here is that phonologically “light” (monosyllabic) nouns can undergo sideways reduplication.

In Sect. 3, we discuss some specific examples of these restrictions based on LSA data. For now, we establish the conditions on plural affixation as in (7).

(7) PHONOLOGICAL CONSTRAINTS ON PLURAL AFFIXATION IN LSA

- a. POA BLOCKS PLURAL AFFIXATION: if the POA features of a relevant Root are specified, PL cannot be attached to that Root.
- b. COMPLEX MOVEMENT BLOCKS PLURAL AFFIXATION: if the PF of a relevant Root determines a complex type of movement, PL cannot be attached to that Root.

Plural affixation to a noun N can occur whenever the constraints in (7a) and (7b) do not apply, i.e., if the noun N is a Root with no POA and simple or no movement. This scenario is schematized in (8), in which (8a) represents the phonological properties of the nominal base, and (8b) the result of attaching the plural morpheme to it. As the representations show, plural affixation involves a phonological process that applies within a single syllable, e.g., if a noun is monosyllabic, the result of applying plural affixation to it is also a monosyllabic word. There is a change, however, on the metrical structure of the resulting form: while (8a) is a “light” syllable, (8b) is a “heavy” syllable.



This pattern of externalization can be attested in the plural noun CHILDREN in (2). As can be seen in this example, sideways reduplication is expressed together with the nominal base, i.e., the exponent of the plural morpheme is realized over the Handshape corresponding to the noun CHILD.

The following subsections develop further important aspects of our proposal.

2.1 Ipsilateral Movement as an Epenthetic Property

As discussed, the underlying representation of the plural morpheme in (6) does not contain features stating that the two arc-like movements must proceed towards the ipsilateral side; the same observation applies for the resulting representation in (8). In the Prosodic Model, a change in the POA obtained through movement is encoded in the *setting* node. In both (4) and (6), *setting* has no specification determining the realization of its dependent node *path*. In other words, the instance of *sideways* reduplication that should result from these underlying representations is not truly “sideways” yet.

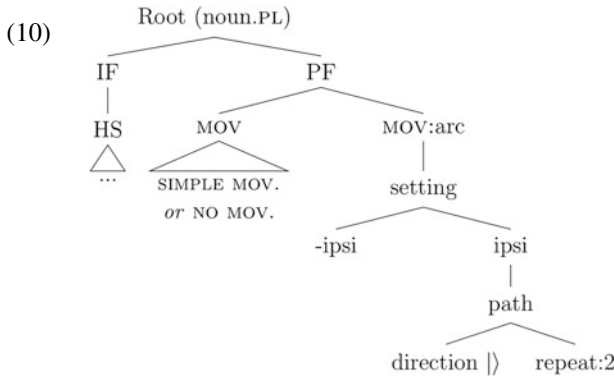
According to Brentari (1998), the *setting* node can receive features by default. In these circumstances, *setting* must introduce movements articulated by the shoulder. There are two types of movement that can be realized as default values for *setting*: *contralateral-ipsilateral* or *top-down*. This is formally captured in (9).

(9) ORDERING OF SETTINGS IN THE DEFAULT CASE (Brentari 1998: 153)

When no *setting* order is indicated in the input, the following default settings are used: [contra] will occur before [ipsi], and [top] will occur before [bottom].

The basic intuition is the following. The plural exponent is expressed through two arc-like movements that cannot be realized in the same place. This means that the articulator must move somewhere. Instead of executing random movements, there

are two default options: contra-ipsi and top-down. Clearly, the plural morpheme makes use of the alternative in which there is movement in the horizontal plane, towards the ipsilateral side. This means that the *setting* features in the phonological representation of the plural noun are merely epenthetic. The result of introducing an epenthetic value to the representation in (8b) is schematized in (10).



We discuss data supporting this mechanism in Sect. 4. From now on, for didactic reasons, we will include the specification of the *setting* features in our phonological representations.

2.2 Handshape Insertion

What about nominal bases that cannot host the plural affix? As discussed, these are nouns in LSA that fit at least one of the descriptions in (7), i.e., (i) nouns with a specified POA or (ii) nouns with complex movements. Do these elements lack a plural form?

We maintain that the phonological system of LSA applies a “last resort” type of mechanism in these cases. Since the plural affix cannot attach to these nouns, a new exponent is inserted in the surface representation to host it. Phonologically, the element in question is a hand configuration, i.e., a configuration consisting of a set of Handshape features. We refer to the operation introducing this element as *handshape insertion*.

Our claim is that handshape insertion applies whenever plural affixation is blocked in the sense of (7), i.e., it is an operation triggered by phonological constraints. Under the assumption that the root in (5) is equivalent to a syllable in spoken languages, with PF depicting the properties of the syllabic nucleus, the mechanism of handshape insertion can be taken to be parallel to consonant epenthesis. Certain instances of consonant epenthesis have been proposed to apply when there is a consonant slot that needs to be filled but there is no lexically

underlying consonant available (Broselow 1995, i.a.). This is roughly, we argue, what happens in LSA: the plural morpheme has no specification for IF, and its PF cannot form a root structure together with the exponent of the lexical noun due to the phonological restrictions in (7); therefore, a set of IF providing a Handshape is introduced to support the PF expressing the plural.

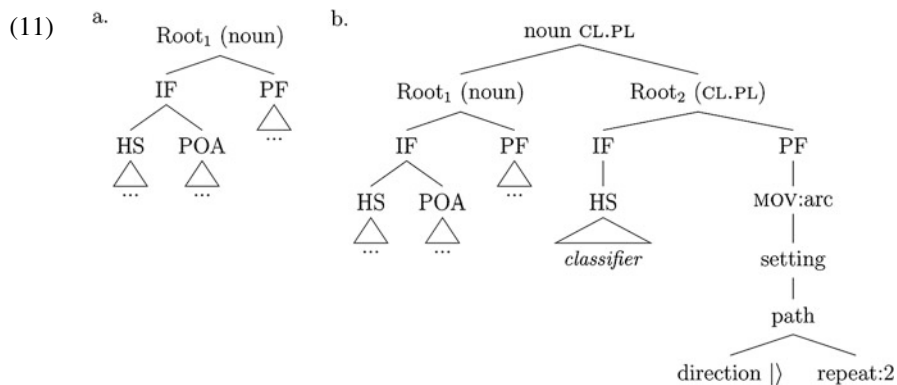
We contend that handshape insertion can only handle sets of Handshape features that already function as units in the language. This means that there is a closed class of signs in LSA providing exponents for the mechanism of handshape insertion. Our assumption is that such a class corresponds to the type of element known as *classifier* (Supalla 1986; Zwitserlood 2012). Thus, the observable result of applying handshape insertion in the contexts of (7) is the realization of the plural affix over a classifier.

Classifiers, also known as *depicting handshapes*, constitute a poorly understood grammatical category that is found in most sign languages. Basically, they are morphemic elements that have no specific lexical meaning, but establish an iconic relation with a nominal referent by depicting its salient characteristics, e.g., shape, size, function, weight, and edges; for instance, the classifier in (4) seemingly represents the size for a prototypical pen. Similar quasi-anaphorical relations hold between classifiers and nouns referring to people, “legged” entities, squarish objects, and so on: the classifier establishes an iconic accordance with perceived properties of a nominal antecedent. Given this behavior, classifiers have been analyzed both as agreement markers (e.g., Glück and Pfau 1998; Zwitserlood 2003; Benedicto and Brentari 2004) or as pronominal forms (e.g., Chang et al. 2005). At the syntactic level, classifiers never occur alone; they are always part of so-called classifier constructions. They typically appear in combination with verbs indicating a referent’s movement through space, although they have also been consistently attested as part of nominal inflections (e.g., Pizzuto and Corazza 1996; Herbert 2018). At the phonological level, classifiers always consist of hand configurations.

While the trigger to apply handshape insertion is purely phonological, we contend that the choice of a specific classifier in a given grammatical context is a matter of morpho-syntactic structure. In other words, phonology dictates when to introduce a classifier, but morpho-syntax determines its shape. The data we discuss in this chapter does not allow us to advance any concrete proposal regarding the latter mechanism, but a preliminary conjecture can be offered. As a working hypothesis, we conceive classifiers as dissociated morphemes in the terminology of Embick (1997) and Embick and Noyer (2001), i.e., as morphemic material that is inserted post-syntactically and reflects structural properties indirectly. Thus, once handshape insertion applies, the form of the corresponding classifier obtains as a function of configurational or featural values in the underlying syntactic representation. We take that an analysis on these lines offers a rationale for the iconic accordance observed between a nominal and its classifier, although much work is still required to advance an explicit proposal.

The basic functioning of handshape insertion is illustrated in (11). Take a noun N that cannot host the plural affix, e.g., an N that has POA or complex movement; the phonological structure corresponding to N is labeled as *Root₁* in (11a). In

order to produce the plural form of an N of that sort, a proper base for the plural affix must be introduced. Thus, a set of Handshape features corresponding to a classifier is inserted in the phonological representation. Being a hand configuration, the classifier itself consists only of an IF branch. Therefore, when combined with the PF of the plural affix, they form a new root structure, i.e., *Root*₂ in (11b). The unit formed by *Root*₁ and *Root*₂ is basically a dissyllabic plural noun, in which the plural morpheme is realized on the second syllable, i.e., *Root*₂. Thus, handshape insertion has the effect of generating a dissyllabic form from a monosyllabic one.



If this analysis is on the right track, the mechanism of handshape insertion provides an account for the distribution and functioning of a number of classifier forms in LSA. As mentioned, classifiers are ubiquitous elements in sign languages, but there is no consensus on how they should be analyzed; see Zwitserlood (2012) for relevant discussion. According to our proposal, at least a subset of the classifier forms attested in LSA appear in the surface representation to provide Handshape features for functional morphemes that are expressed through movement alone. Thus, the corollary of our proposal is that (at least some) classifiers serve as support units for morpho-phonological processes.

The idea that classifiers may have a support role in the expression of nominal plurality is not totally new. In their discussion of nominal inflection in Italian Sign Language (LIS), Pizzuto and Corazza (1996:185) make very similar observations regarding sideways reduplication.

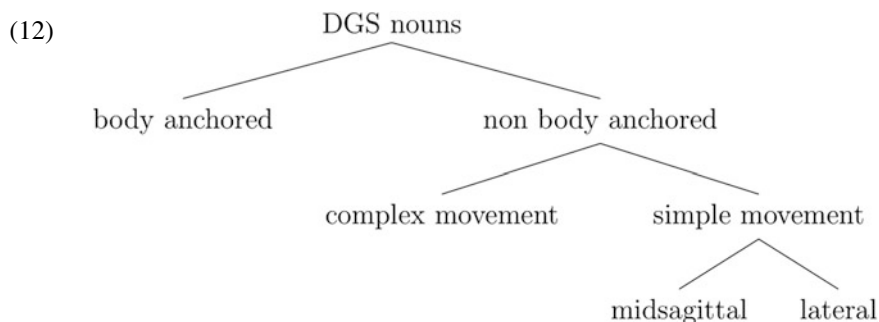
Nouns articulated on, or close to the signer's body cannot be modified to mark the numerosity inflection in the same manner that is observable in nouns articulated in neutral space: their handshapes cannot be relocated and reduplicated in the signing space, but must retain the same point of articulation they have in citation form. The use of classifier signs easily overcomes this morphophonological constraint. For example, suppose the signer wishes to express the meaning of 'many binoculars'. χ BINOCULARS is a noun that can neither be inflected in space, nor undergo the nonmanual numerosity inflection. But a classifier sign can be produced immediately after the noun, and this classifier can be inflected in space for numerosity in the same manner as neutral space nouns can.

In the following sections, we analyze a number of pluralization patterns attested in LSA. The discussion of the data strongly confirms the theoretical approach advanced throughout this section.

3 A Phonological Typology of Nouns in LSA for the Expression of Plurals

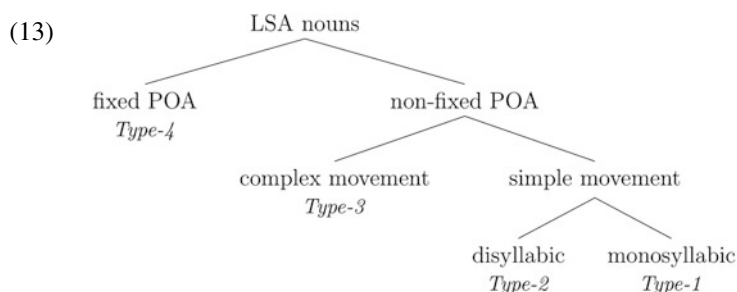
The data from LSA to be presented throughout this section was gathered by Yanina Boria in collaboration with Carolina Galvez, our consultant from the Argentinian Deaf Community. A comprehensive protocol (Van Herreweghe and Vermeerbergen 2012) consisting of sign language observation, data production, data elicitation, and data review was employed. As a first step in the collection of the data, public access recordings in LSA, dictionaries, and virtual applications were consulted; among the dictionaries used are *Diccionario de Lengua de Señas Argentina* (Valassina 1997), *Lengua de Señas Argentina: Análisis y Vocabulario Bilingüe* (Massone and Machado 1992), *LSA en Familia* (Proyecto DANE), and *Señario* (Confederación Argentina de Sordos 2019). A total of 360 nouns (together with their plural forms) was analyzed. From this corpus, 9 representative data points were selected for discussion. Our consultant signed and filmed the singular and plural forms for this reduced sample. These examples were further corroborated by informants from two non-governmental organizations (Círculo de Sordos and Asociación Civil de Artes y Señas).

We take as a starting point the work by Pfau and Steinbach (2005, 2006). These authors capture the patterns of externalization of nominal plurality in German Sign Language (DGS) by classifying the nouns in the language according to their phonological properties. They argue that there are two major groups of nouns: nouns that are body-anchored, and nouns that are not body-anchored. The second group ramifies into two subgroups depending on the type of movement they involve: nouns with complex movement, and nouns with simple movement that, in its turn, can be further divided into two subtypes depending on their place of articulation: they can be midsagittal or lateral. The scheme in (12) summarizes the typology.



This classification relates with distinct pluralization strategies in the language. That is, the plural of body-anchored nouns and nouns with complex movement is spelled out as a null morph, i.e., it receives zero marking. Plural nouns that are articulated in the midsagittal plane are realized as simple reduplication. Finally, plural nouns with a lateral place of articulation involve sideways reduplication. The availability of distinct exponents for the plural morpheme in DGS together with these strong correspondences suggests an analysis of the pattern in terms of phonologically conditioned allomorphy; this is precisely the approach taken by Pfau and Steinbach. Basically, they propose that the selection of the exponent for the plural morpheme depends on the phonological context in which this element appears.

Our analysis of nominal plurals in LSA closely follows the classification by Pfau and Steinbach (2005, 2006). However, in contrast to these authors, we claim that the plural morpheme in LSA does not exhibit an allomorphic alternation, i.e., the phonological representation in (6) constitutes the only exponent for PL in the language. We contend that all discernible differences in the phonological manifestation of the plural morpheme are mainly due to (i) the mechanism of handshape insertion (see Sect. 2.2) and to a lesser degree to (ii) contexts in which there is no ipsilateral movement epenthesis (in the sense of Sect. 2.1). Our discussion for now will focus on the first factor. As argued, handshape insertion applies depending on the phonological properties of the nominal base to which the plural morpheme needs to be attached. We advance the typology of nouns in (13) to capture the distribution of handshape insertion in LSA.



In principle, the classification in (13) looks similar to its DGS counterpart in (12). However, it includes a number of distinctions that can be made on the basis of the Prosodic Model. To begin with, consider nouns classified as Type-1. These are monosyllabic nouns (i.e., they have only one root structure) that involve a Handshape undergoing a simple movement, e.g., a straight path. Nouns pertaining to this class are CHILD, HOUSE, BUILDING, TREE, COIN, BALL, CITY, PIZZA, TABLE, and PLATE, among many others; in our sample, 104 of the 360 collected nouns belong to this class. The functioning of Type-1 nouns is firstly illustrated with the noun CHILD in (1). As can be seen in (14), the underlying phonological

representation of CHILD is just a Handshape; it has no POA or PF. The top-bottom movement attested in (1) is a “by default” realization of the PF tier in the quotation form of the sign; this follows the rule for epenthetic movement in (9). We understand epenthetic movements like this one to be always simple straight movements in LSA, just as Brentari (1998: 131) proposes for American Sign Language (ASL). Since straight movements are assumed to be the “default” type of movement in LSA, they do not need to be specified in the underlying representation of the root.

(14) Root (CHILD)



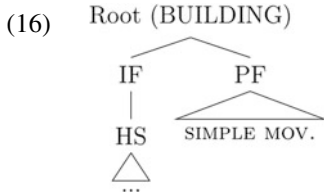
Many nouns pertaining to Type-1 do not involve an epenthetic movement. In contrast to CHILD, the underlying phonological form of a noun like BUILDING does partially specify Prosodic Features. As (15) shows, BUILDING involves a simple bottom-top movement. In this case, such movement is an invariable part of the phonological realization of the noun in all contexts, and therefore it does not display a “by default” realization in its quotation form.

(15)



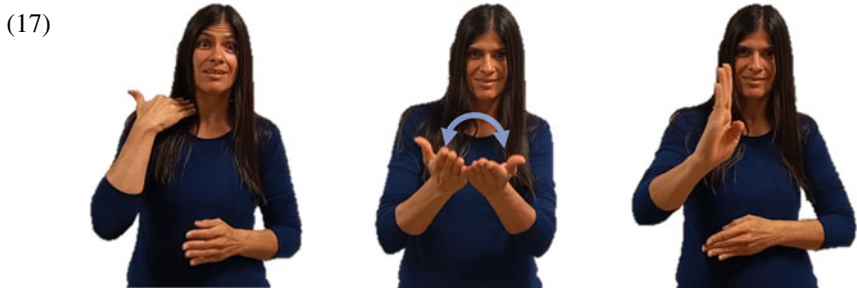
BUILDING.SG

For concreteness, we assume that this movement depends on the specification of a single node in the PF tier, i.e., setting, and that the rest of the PF branch remains “free”; as discussed earlier, the result of this sort of specification is a simple movement. For ease of presentation, we label the underlying PF branches of nominal roots as having simple or complex movements instead of offering a complete phonological description for each sign.



No matter whether they have simple movement due to epenthesis or lexical properties, all Type-1 forms are monosyllabic nouns with no POA, i.e., they have no fixed place of articulation.

Type-2 nouns also lack POA features and exhibit simple movements. What distinguishes them from Type-1 nouns is that they are disyllabic, i.e., they involve the conjunction of two root structures. Nouns pertaining to this class are BOOK, CAR, PEA, FRUIT, SHIP, BICYCLE, and GARLIC, among many others; in total, 95 of the 360 nouns collected belong to Type-2. As examples of this class, consider BOOK and BICYCLE in (17) and (18), respectively.



BOOK.SG

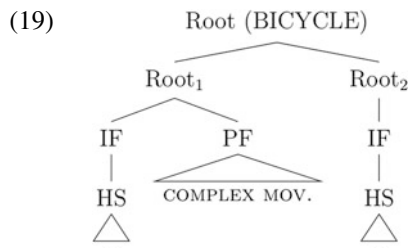


BICYCLE.SG

As the schematic description in (19) shows, these nouns have two roots, i.e., *Root₁* and *Root₂*, each of them with their own phonological properties. In both cases, *Root₁* is specified for Handshape, POA, and complex movement.¹ *Root₂*, on the

¹ We take BOOK to involve two roots, one involving complex movement, i.e., the first two captures in (17), and one with simple movement, i.e., the last capture in (17). However, a potential analysis

other hand, has only Handshape features; therefore, the movement in *Root*₂ is an epenthetic simple movement. As can be seen, these nouns are classified as Type-2 due to the properties of *Root*₂, which exhibits simple movement and no POA, just like monosyllabic nouns of Type-1.²



Just like elements of Type-1 and Type-2, nouns pertaining to Type-3 do not have a fixed place of articulation. What distinguishes Type-3 nouns is that their final (and in most cases only) root exhibits complex movements, i.e., movements involving parallel submovements. This is the class in the typology of (13) with the least number of samples within our corpus: only 5 nouns from a total of 360. These are BOX, DONUT, MOTORBIKE, TRUCK, and FURNITURE. The functioning of Type-3 nouns is exemplified with BOX in (20).

(20)

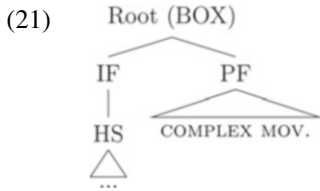


BOX.SG

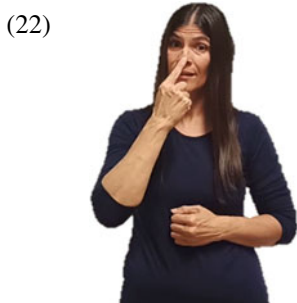
in terms of three roots with simple movement could also be possible. Further research in the phonology of LSA is needed to better establish a method for syllable separation and a language-specific metric for simple/complex movement. The issue does not alter any significant aspect of the analysis proposed here.

² Type-2 includes many nouns with complex morphological structure, i.e., nouns that are obtained through derivation or composition. For instance, BICYCLE seems to be a deverbal noun; its second syllable is a classifier that seemingly has a nominalizing function. Further research on the morphological structure of nouns in LSA is needed to provide a better characterization of the internal structure of these signs.

As can be seen, this noun has Handshape features in both articulators and undergoes a change of configuration and arm movement in both of them. This is summarized in (21) as involving complex movement.

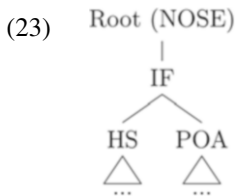


Nouns pertaining to Type-4 are the only ones in our typology that are specified for POA features in their underlying phonological representation; this means that they are fixed to a certain place of articulation. In our corpus, a total of 161 nouns was classified as Type-4 forms; NOSE, DEER, PENCIL, STOMACH, APPLE, EYEBROW, PARROT, (COCHLEAR) IMPLANT, and RABBIT are some examples of this class. Consider the case of NOSE in (22), which is invariably articulated on the nose.

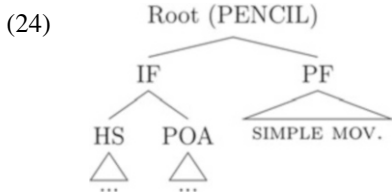


NOSE.SG

As illustrated in (23), this noun is phonologically composed of Handshape and POA, but lacks a specification for PF, i.e., it involves no movement at all.

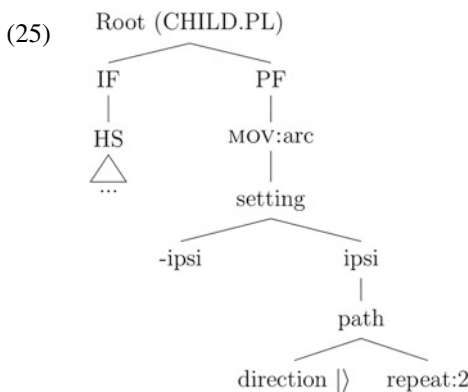


Type-4 nouns can display simple movement within their own place of articulation. This is the case of PENCIL, illustrated in (3). In this example, the noun incorporates a local movement towards the chin. This is codified as simple movement in the root structure of (24).



We contend that the typology advanced in (13) allows to predict the distribution of handshape insertion in LSA, just like Pfau and Steinbach’s classification in (12) captures the patterns of allomorphy in DGS. Basically, we observe that LSA nouns pertaining to Type-1 and Type-2 can host a plural affix without the need of introducing a classifier. Nouns of Type-3 and Type-4, on the other hand, trigger handshape insertion. As discussed earlier, this follows from the constraints on plural affixation in (7). That is, a hand configuration is needed if the nominal base cannot host the plural affix. The exponent of the plural affix is a complex type of movement that needs to be applied over a Handshape; if the phonological properties of the noun do not allow this exponent to be expressed, e.g., in virtue of having a fixed place of articulation (7a), or because the noun already exhibits a complex sort of movement (7b), plural affixation cannot take place. Thus, nouns of Type-3 and Type-4 are predicted to require a classifier to surface as plural nouns.

Consider first the Type-1 noun CHILD, already discussed in (14). As said, the underlying phonological representation of CHILD consists only of Handshape features; this element has no POA nor PF. Thus, the constraints in (7) do not apply to it, and, therefore, plural affixation can take place. This prediction is borne out in (2), where the plural affix is realized over a single root together with the nominal base. The following scheme describes the resulting form.



The same behavior is observed with nouns of Type-1 that are specified for simple movement in the PF tier. This is the case of BUILDING. As can be seen in (26), the plural form of this noun involves sideways reduplication of its nominal base.

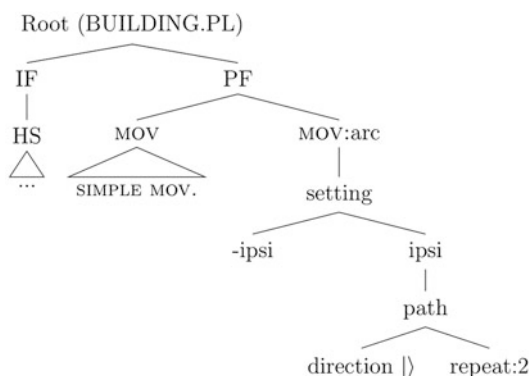
(26)



BUILDING.PL

As this example shows, the plural exponent can be realized together with other movements as long as these are simple movements. Thus, we confirm that the restriction in (7b) is on the right track, and that metrical weight is a key factor governing plural affixation in LSA. The representation in (27) summarizes the analysis.

(27)



Type-2 nouns also allow for direct plural affixation. By comparing the plural forms of BOOK and BICYCLE in (28) and (29) with their singular counterparts in (17) and (18), it is possible to observe that the material undergoing sideways reduplication is the final part of the nominal base.

(28)



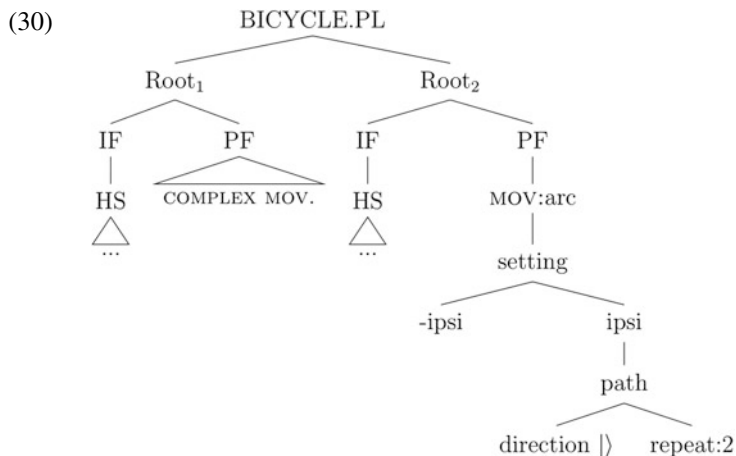
BOOK.PL

(29)



BICYCLE.PL

This scenario is predicted by our analysis. As discussed, Type-2 nouns are disyllabic in the sense that they are composed of two root structures, $Root_1$ and $Root_2$. As pointed out in (19), $Root_2$ consists of Handshape features only and, therefore, can host the plural affix. The examples in (28) and (29) confirm this prediction. The result is the phonological representation sketched in (30).



In a nutshell, plural affixation with both Type-1 and Type-2 nouns is a phonological process that applies to “light” syllables. The result is a “heavy” syllable exhibiting complex movement. In both cases, the plural form of the noun has the same number of syllables as its singular counterpart.

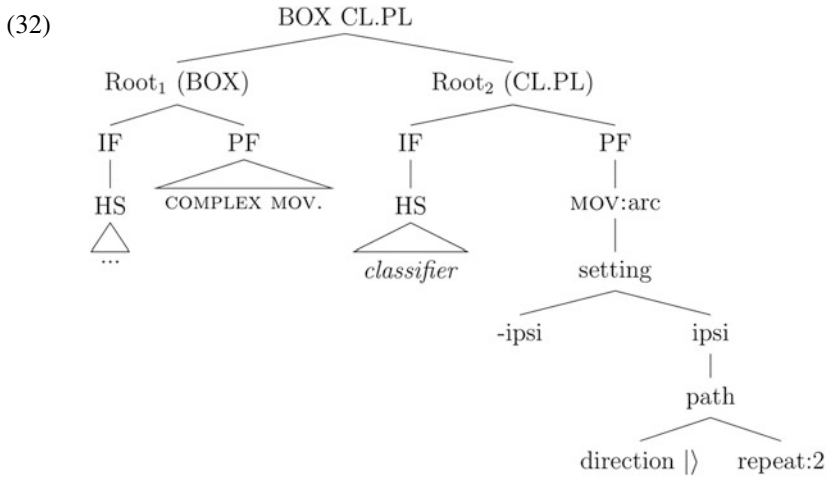
Nouns pertaining to Type-3 cannot host the plural affix. As discussed with respect to BOX in (21), these nouns have a PF specification that determines a complex movement. Consequently, they fit the constraint in (7b) and reject plural affixation. The ungrammaticality of plural affixation to Type-3 nouns is exemplified in (31) with what would be the plural form of BOX. If plural affixation was possible in this case, the plural BOXES should be realized as a sequence of complex movements that move towards the ipsilateral side. This is considered unacceptable by LSA speakers.

(31)



*BOX.PL

The grammatical way to construct the plural form of BOX is through handshape insertion. That is, since the plural morpheme cannot attach to BOX due to the restriction in (7b), a set of Handshape features needs to be inserted in the phonological representation to host the plural. By introducing a Handshape, the classifier makes available a new root structure that has neither POA nor complex movements, i.e., it is a “light” syllable. The plural affix attaches to this new root and converts it into a “heavy” syllable with complex movement.



The resulting plural form for BOX is exemplified in (33). As can be seen, the plural exponent is spelled out on a (two-handed) classifier consisting only of Handshape features.



Nouns pertaining to Type-4 also require handshape insertion. As the discussion regarding the nouns NOSE in (3) and PENCIL in (22) indicates, Type-4 nouns have POA features, i.e., they have a fixed place of articulation. Thus, according to the constraint in (7a), they cannot host the plural affix. This restriction is exemplified in (34) and (35) with the ungrammatical forms that are obtained from directly attaching the plural affix to NOSE and PENCIL. As can be seen, these forms would require moving the whole place of articulation (nose or chin, respectively) towards the ipsilateral side. LSA speakers find these realizations strongly unacceptable.

(34)



*NOSE.PL

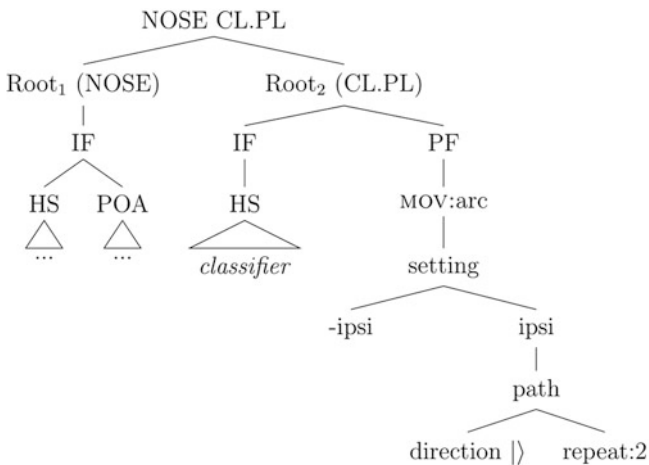
(35)



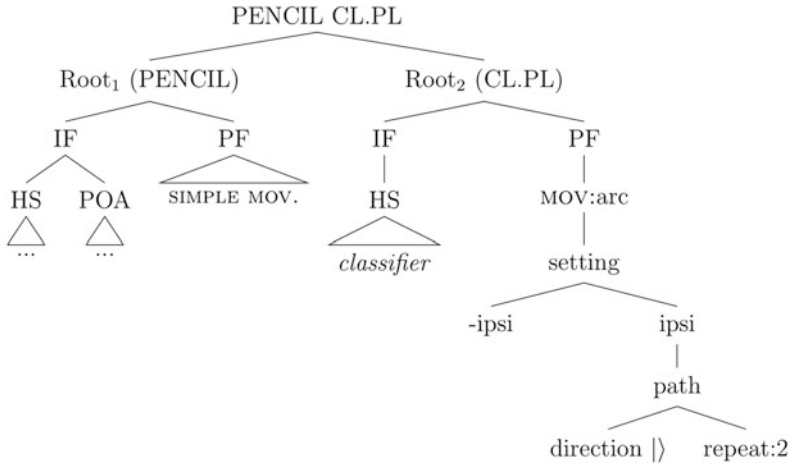
*PENCIL.PL

The adequate plural forms of these nouns require inserting a hand configuration to host the plural affix. This process applies just as described for Type-3 nouns. That is, the classifier provides a Handshape over which the movement corresponding to the plural affix can be expressed. The corresponding phonological representation for these plural forms is sketched in (36) and (37).

(36)



(37)



Thus, the acceptable plural forms for NOSES and PENCILS must involve classifiers, just as shown in (38) and (4), respectively.

(38)



There are two main conclusions that can be drawn from this analysis. First, the alternation between distinct forms of pluralization in LSA depends on the phonological properties of the nominal base; this is in line with Pfau and Steinbach’s (2005, 2006) proposals for DGS. Second, the variants in the realization of nominal plurality in LSA do not involve allomorphy. As seen, the exponent for the plural morpheme always remains the same, i.e., it has the phonological properties described in (6).

This contrasts sharply with the behavior of nominal plurality in DGS, which, in principle, does seem to involve allomorphic variants.³

A final qualification is in order. All the examples discussed in this paper belong to the *native lexicon* of LSA; no systematic analysis of items pertaining to the *non-native lexicon* was conducted. As already known, the relevant difference between these two is that the items from the non-native lexicon are constructed with phonological features that are not part of the language (e.g., the manual alphabet); see Brentari (1998) and Padden (1998) for relevant discussion. Our corpus had only five nouns that belong to the non-native lexicon: BREAD, ZOO, BAR, CLUB, and DAY. In all these cases, the plural forms are formed through handshape insertion.

4 More on the Epenthetic Nature of Ipsilateral Movement

As discussed in Sect. 2, the underlying phonological representation of the plural morpheme does not specify features for sideways movement. That is, the schematic structure in (6) requires (i) that the exponent of plural must move in the shape of an arc, (ii) that this movement must be repeated twice, and (iii) that the repetitions do not take place in the same location; the “sideways” aspect of this movement, i.e., the fact that it proceeds from the contralateral towards the ipsilateral side, was argued to be epenthetic. In this section, we present some patterns supporting this characterization.

Authors like Brentari (1998) and Geraci (2009) have extensively argued that movement epenthesis is a standard trait of sign languages. In ASL, for instance, Brentari (1998) observes that certain verbs display path movement in their quotation form but lose this movement when forming a compound sign. She interprets this pattern as involving a “by default” movement in the quotation form. Her rationale is schematized in (39).

- (39) a. A sign requires movement to be well-formed (just like a syllable requires a nucleus).
 b. Some signs have no underlying specification for movement; when they are uttered in isolation, they are spelled out with epenthetic movement.
 c. These signs might not require movement epenthesis in certain contexts; this is the case of compounding in ASL, as the resulting compound is already specified for movement.

³ Under this analogy, our proposal resembles the hypothesis that the vowel *-e* found in Spanish pairs such as *flor* “flower” vs. *flores* “flowers” is epenthetic (e.g., Harris 1980). According to this analysis, *-s* and *-es* are not allomorphs. Instead, *-s* is the only plural exponent in both cases, while *-e* is inserted due to requirements of syllabic structure. Obvious differences aside, our analysis follows a parallel line of reasoning.

An analogous reasoning led us to propose that the ipsilateral movement displayed by the plural affix in LSA is epenthetic. That is, there are grammatical contexts in which the plural is realized as arc movements that do not develop towards the ipsilateral side. Following the logic depicted in (39), this means that the “ipsilaterality” of sideways reduplication is a “by default” value.

Evidence for this claim comes from cases in which the plural noun is accompanied by a locative modifier. In these scenarios, the locative is instantiated by a prenominal sign that is spelled out in a certain place of articulation. The setting features corresponding to the plural noun are restricted to that place of articulation, in such a way that the arc movements of the plural morpheme must be realized within it. In other words, the locative specifies a signing space for the noun, either in its singular or plural forms.

Consider the following example based on the behavior of the noun HOUSE.

(40)



HOUSE.SG

HOUSE is a noun of Type-1, i.e., it is a monosyllabic word that can host the plural affix without the need of introducing a classifier. As can be seen in (41), the plural form of this noun exhibits sideways reduplication with ipsilateral (epenthetic) movement.

(41)



HOUSE.PL

However, this realization is very different if the plural NP is modified by a locative. Consider the example in (42), which translates as “the houses in the mountain.” As can be seen, the place of articulation introduced by the locative MOUNTAIN becomes the signing space for the noun; in particular, the reduplication pattern of the plural morpheme does not involve movement towards the ipsilateral side in this case, but a series of arches within the designated place of articulation. Pfau and Steinbach (2005, 2006) refer to this type of realization as *random reduplication*.

(42)



MOUNTAIN IX.ADV HOUSE.PL

This sort of pattern is attested with nouns pertaining to any of the four classes depicted in (13). Plural forms that require handshape insertion behave much in the same way as HOUSE in (40). Consider as a further example the case of COOKIE in (43). This is a noun classified as Type-4, i.e., it has a fixed place of articulation in the chin area (and also displays a short simple movement).

(43)



COOKIE.SG

As a Type-4 noun, its plural form COOKIES requires a classifier; the example in (44) shows how the classifier undergoes sideways reduplication.

(44)



COOKIE CL.PL

However, if the locative PLATE modifies the noun COOKIES, e.g., as in “cookies on a plate,” the movement of the classifier carrying the plural morpheme is constrained to the signing space depicted by PLATE, and it does not proceed towards the ipsilateral side.

(45)



PLATE IX.ADV COOKIE CL.PL

This type of phenomenon is standard for sign languages. As observed by Quer et al. (2017), the visual-gestural modality allows to mark location features on nouns; if the location is not part of the quotation form of the noun, the noun is realized with a “by default” location. Our proposal is that the exteriorization of nominal plurality in LSA follows the same logic: the movement of the plural morpheme is ipsilateral unless the noun displays some location.

5 Concluding Remarks

In this chapter, we have advanced an analysis of the functioning of nominal plurality in LSA. We have argued that the plural morpheme in this language behaves as an affix, in the sense that it needs a nominal base to be spelled out. The plural has only one exponent, which is expressed as repetitions of an arched movement carrying over the noun. All remaining variation in how plural nouns are exteriorized lies on two phonological mechanisms: handshape insertion and movement epenthesis.

We conceive handshape insertion as a rescue operation that applies in the phonological component. It applies whenever a plural affix is phonologically incompatible with the nominal base. In these representations, a classifier, i.e., a sign consisting of Handshape features only, appears to provide a configuration for the articulator undergoing plural movement. We further advanced a typology of nouns in LSA based on the phonological properties that trigger handshape insertion.

We have also argued that the “sideways” directionality of the plural morpheme is a “by default” option in LSA that is not part of the underlying phonological representation of the plural morpheme. This follows from the observation that plural nouns do not display ipsilateral movement in contexts in which they have locative modifiers. While this chapter provides a rather explicit approach to the morpho-phonology of nominal plurality in LSA, significant work is still needed in this area. We hope this paper contributes to encouraging further research on the grammar of sign languages, particularly of those spoken in South America.

Acknowledgments We would like to thank Carolina Galvez, Círculo de Sordos, and Asociación Civil de Artes y Señas for their collaboration and support during this research. Usual disclaimers apply. Carlos Muñoz Pérez acknowledges the financial support from ANID/FONDECYT project 3190813.

References

- Anderson, John M.A., and Colin J. Ewen. 1987. *Principles of dependency phonology*. Cambridge: Cambridge University Press.
- Benedicto, Elena, and Diane Brentari. 2004. Where did all the arguments go? Argument-changing properties of classifiers in ASL. *Natural Language and Linguistic Theory* 22: 743–810.
- Brentari, Diane. 1998. *A prosodic model of sign language phonology*. MIT Press.

- . 2012. Phonology. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 21–54. Berlin: De Gruyter.
- Broselow, Ellen. 1995. Skeletal positions and Moras. In *The handbook of phonological theory*, ed. John A. Goldsmith, 175–205. Oxford: Blackwell.
- Chang, Jung-hsing, Su Shiou-fen, and James H.-Y. Tai. 2005. Classifier predicates reanalyzed, with special reference to Taiwan sign language. *Language and Linguistics* 6 (2): 247–278.
- Confederación Argentina de Sordos. 2019. *Señario*. <https://cas.org.ar/2019/05/24/senario-de-terminos-y-expresiones-en-lengua-de-senas-argentina-descarga/>. Accessed 16 Dec 2020.
- Embick, David. 1997. *Voice and the interfaces of syntax*. Doctoral dissertation, University of Pennsylvania.
- Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32 (4): 555–595.
- Fenlon, Jordan, Kearsy Cormier, and Diane Brentari. 2018. The phonology of sign languages. In *The Routledge handbook of phonological theory*, ed. S.J. Hannahs and Anna R.K. Bosch, 453–475. Abingdon/New York: Routledge.
- Geraci, Carlo. 2009. Epenthesis in Italian sign language. *Sign Language & Linguistics* 12 (1): 3–51.
- Glück, Susanne, and Roland Pfau. 1998. On classifying classification as a class of inflection in German sign language. In *Proceedings of ConSOLE VI*, ed. Tina Cambier-Langeveld, Aniko Lipták, and Michael Redford, 59–74. Leiden: SOLE.
- Harris, John. 1980. Nonconcatenative morphology and Spanish plurals. *Journal of Linguistic Research* 1: 15–31.
- Herbert, Marjorie. 2018. A new classifier-based plural morpheme in German sign language (DGS). *Sign Language & Linguistics* 21 (1): 115–136.
- Massone, María Ignacia. 1996. *Gramática y léxico de la lengua de señas Argentina*. Doctoral dissertation, University of Buenos Aires.
- Massone, María Ignacia, and Emilia Machado. 1992. *Lengua de Señas Argentina: Análisis y Vocabulario Bilingüe*. Buenos Aires: Edicial Universidad.
- Massone, María Ignacia, and Rocío Martínez. 2012. *Curso de Lengua de Señas Argentina (parte IIIId)*. https://cultura-sorda.org/wp-content/uploads/2015/03/Massone_Martinez_Curso_LSA_2012-Parte_IIIId.pdf. Accessed 12 Dec 2020.
- Padden, Carol A. 1998. The ASL lexicon. *Sign Language & Linguistics* 1 (1): 39–60.
- Pfau, Roland, and Markus Steinbach. 2005. Plural formation in German sign language: Constraints and strategies. In *Gebärdensprachen: Struktur, Erwerb, Verwendung*, ed. Helen Leuninger and Daniela Happ, 111–144. Hamburg: Helmut Buske Verlag.
- . 2006. Pluralization in sign and in speech: A cross-modal typological study. *Linguistic Typology* 10 (2): 135–182.
- Pizzuto, Elena, and Serena Corazza. 1996. Noun morphology in Italian sign language (LIS). *Lingua* 98: 169–196.
- Quer, Josep, Carlo Cecchetto, Caterina Donati, Carlo Geraci, Meltem Kelepir, Roland Pfau, and Markus Steinbach. 2017. *SignGram blueprint: A guide to sign language grammar writing*. Berlin: De Gruyter.
- Sandler, Wendy. 2012. The phonological organization of sign languages. *Lang Linguist Compass* 6 (3): 162–182.
- Steinbach, Markus. 2012. Plurality. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 112–136. Berlin: De Gruyter.
- Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication system of the American deaf*, Studies in linguistics: Occasional papers 8. Buffalo: University of Buffalo.
- Supalla, Ted. 1986. The classifier system in American sign language. In *Noun classes and categorization*, ed. Colette Craig, 181–214. Amsterdam: John Benjamins.
- Valassina, Sara. 1997. *Diccionario Lengua de Señas Argentina-Español*. Buenos Aires: Ministerio de Cultura y Educación de la Nación.
- Van der Hulst, Harry. 1993. Units in the analysis of signs. *Phonology* 10: 209–241.

- Van Herreweghe, Mieke, and Myriam Vermeerbergen. 2012. Data collection. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 1023–1045. Berlin: De Gruyter.
- Zwitserslood, Inge. (2003). *Classifying hand configurations in Nederlandse Gebarentaal (sign language of the Netherlands)*. PhD Dissertation, Utrecht University.
- . 2012. Classifiers. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 158–186. Berlin: De Gruyter.

The Grammar of Agreement in Libras



Guilherme Lourenço

1 Introduction

Signed languages are natural human languages that emerged from and within Deaf communities all over the world. Just as spoken languages, there is a wide range of variation and differences among them, so that each signed language has its own grammar system. Libras – *Língua Brasileira de Sinais*, Brazilian Sign Language – is the signed language used by the urban Deaf communities in Brazil. Its history can be traced back to 1857, when the first school for Deaf people was established in Rio de Janeiro. It may very well be that other sign systems/languages existed before that in Brazil, but it was the foundation of this first Deaf school that gave rise to a national signed language that is used across the country nowadays.¹

When it comes to linguistic analysis, many formal approaches have been adopted to describe different aspects of the grammar of Libras. In this chapter, I focus on the system of verb agreement and how it interacts with different syntactic structures in the language. Building on previous works (Lourenço 2018b; Lourenço and Wilbur 2018), I will argue that verb agreement is pervasive in the grammar under discussion and that classical lexical divisions, such as the existence of “agreement verb classes,” should be eliminated. I also claim that agreement in Libras, and possibly in other signed languages, can be derived from syntactic mechanisms that are modality-independent.

¹ Libras is not the only signed language in Brazil. There are at least ten other already documented signed languages that can be found in villages (rural areas) or in indigenous areas (de Quadros and da Silva 2017).

G. Lourenço (✉)

Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

To support my claim, I bring language data extracted from the Libras Corpus Project (de Quadros et al. n.d.). The Libras Corpus is an “open-access online comprehensive corpus of Libras, that is wide-ranging, empirically derived and theoretically and methodologically driven” (de Quadros et al. n.d.). The corpus comprises different initiatives to collect and register different types of signing – from vernacular to literary signing and from Deaf people of different ages and from different parts of Brazil. All the Libras examples provided in this chapter are extracted from the Corpus.

This chapter is structured as follows. Section 2 presents some general aspects of Libras grammar, focusing on describing the basic structure of the clause, the grammatical roles of nonmanual markers, and the use of sign space to encode (grammatical) person information. Section 3 describes how agreement is marked in Libras and how it interacts with the signing space. In Sect. 4, a syntactic derivation of agreement is pursued, assuming a generative grammar framework. Section 5 discusses how pervasive agreement is in the Libras Corpus data, arguing that agreement is always marked in Libras, unless it is not available syntactically or there is a phonological restriction in the verb (body-anchoring). Some final remarks are provided in Sect. 6.

2 Some Relevant Aspects of Libras Grammar

Current descriptions of Libras syntax have been focused mainly on the structure of the clause. Word order (Lourenço and de Quadros 2020; de Quadros 1999; Royer 2019), negative constructions (Arrotéia 2005; Lourenço 2015), aspectual and/or tense marking (Bertucci and Finau 2018; Figueiredo and Lourenço 2020; Finau 2004), modality (Ferreira-Brito 1990; Xavier and Wilcox 2014), and verb morphology (Felipe 1998; Lourenço 2018b; Lourenço and Wilbur 2018; de Quadros and Karnopp 2004; de Souza 2016) are some of the topics already investigated. In this section, a brief overview of Libras clause structure will be provided.

Libras has a basic SVO word order that is always grammatical in matrix sentences, regardless of verb type or verb morphology. Other derived orders, such as SOV, OSV, and VOS, are possible, but only under specific syntactic configurations – such as focalization, topicalization, and object shift (Lourenço and de Quadros 2020; de Quadros 1999). Some examples of plain SVO sentences are provided as follows (Figs. 1 and 2):

- (1) NOW I HAVE AGE SIX-ONE
“Now I am 61 years old.”
- (2) TEACHER ₃GIVE₁ TEST
“The teacher gave me the test.”



Fig. 1 NOW I HAVE AGE SIX-ONE (Libras Corpus)



Fig. 2 TEACHER ₃GIVE₁ TEST (Libras Corpus)

As many other signed languages documented around the globe, Libras makes use not only of manual information, but also of grammatical nonmanual markers (NMMs). Similarly, to what has been argued for American Sign Language (ASL) by Wilbur and Patschke (1999) (see also Neidle et al. (2000)), we have claimed that upper face NMMs and head movement function as syntactic markers in Libras (Figueiredo and Lourenço 2019; Lourenço 2018a). Just to give some examples, as earlier noticed by de Quadros (1999), topicalization in Libras is marked with raised eyebrows (_re) (3) and focus constructions are accompanied by a head nod (_hn) (4).

- (3) Topicalization: (de Quadros and Karnopp 2004: 147)

_re
 SOCCER JOHN LIKE
 “Soccer, John likes.”

- (4) Focus construction: (de Quadros and Karnopp 2004: 153)

I (LOSE) BOOK _hn
 “I LOST my book.”

Nonmanual markers have also been described for other types of constructions. For instance, head and eyebrow movements have been found in questions – both polar and content ones (de Quadros 2011). Figueiredo and Lourenço (2019) have reported raised eyebrows also in relative clauses, conditionals, and (adverbial) subordinate clauses.

Another aspect of Libras grammar that needs to be mentioned is that it makes extensive use of spatial information, as also commonly found in other signed languages. In this sense, space is recruited as a grammatical mechanism and its three-dimensional structure is exploited to convey different types of linguistic information, playing a role not only in syntax, but also in phonology, morphology, semantics, and discourse (Perniss 2012).

The most relevant aspect of the grammar of space for the topic under scrutiny in this chapter is the association of specific points in space with discourse referents. By means of establishing points in space, signed languages mark (grammatical) person information and create a very modality-specific system of making and tracking reference. This is so because “unlike oral [spoken] languages where space is referred to, in sign languages, space is physically available for representation” (Padden 1990: 118). A classic tripartite classification, firstly proposed by Friedman (1975), is provided in Fig. 3. Notice that pointing inward, toward the signer’s body, marks first-person reference, whereas pointing toward the addressee is associated with second-person reference. Pointing away from both the signer and the addressee marks third-person reference.

Additionally, when it comes to third-person reference, an arbitrary point is picked. This creates a semantic mapping between an entity (x) and an abstract geometric point² (p) in space. I have called this association between (p) \rightarrow (x) a *location*, to make clear that I am not referring only to the point in space, but to this specific semantic mapping.

Given this discussion on how points in space can be associated with individuals resulting in a location, let us now describe verb agreement in Libras.

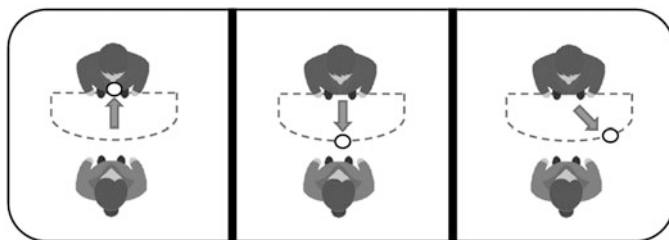


Fig. 3 Pointing locus for first person, second person, and third person, respectively (Lourenço 2018b:62)

² For a discussion on the status of this geometric point in space, see Wilbur (2013), Quer (2011), and Lourenço (2018a, Chap. 3).

3 Agreement in Space

Once there are locations (geometrical points in space (p) linked to specific referential entities (x)), a group of verbs can be modified in such a way that the beginning point and the endpoint of their movement will coincide with the location associated with their arguments. This systematic modification of the verb based on the locations of its arguments has been called *verb agreement*. This is the case, for example, of the verb ANSWER in Libras (Fig. 4).

Traditionally, only verbs that have a path movement between two points in space have been classified as agreement verbs in signed languages (Padden 1990, *inter alia*). However, we (Lourenço 2018b; Lourenço and Wilbur 2018) have challenged this notion by claiming that verbs that do not move through space, but still have their location changed in order to match the location of an argument, are also marked for agreement. The verb WORK is an example of a verb that is signed in respect of a single point in space, but has its location changed to match the locus of an argument (Fig. 5).

We, thus, provide a clear definition of agreement based solely on the mechanism of location matching between controllers and targets:

Verb Agreement in Sign Languages

A verb shows agreement with its argument(s), when the verb's location is changed in order to match the location of the argument(s), a process called co-localization.

Fig. 4 The verb ANSWER in Libras. In this example, the verb departs from the second-person locus toward the body of the signer, meaning “you answer to me” (Libras Corpus)



Fig. 5 The verb WORK in Libras. In this example, the location of the verb is changed, and the sign is produced in a lateral area of the signing space. This location has been previously assigned to an argument of the sentence (Libras Corpus)



Fig. 6 The verbs *LIKE* (left) and *SUFFER* (right) in Libras are body-anchored verbs and, therefore, cannot have their location changed (Libras Corpus)



When we consider that all instances of location matching are agreement, the majority of verbs in Libras can be agreement marked. However, some verbs are not capable of having their location changed because their point of articulation is lexically specified and, therefore, cannot be altered. These verbs are those that are articulated on (or close to) a specific location of the signer's body and are called body-anchored verbs (Fig. 6).

Considering that body-anchoring is a phonological specification of the sign, we have stated that all Libras verbs can show agreement unless they are phonologically restricted not to do so. This claim is important because it has been observed that in many signed languages only a few verbs are marked for agreement (Mathur and Rathmann 2012; Padden 1988); thus, some authors have argued against calling this mechanism agreement. Instead, some analyses have proposed that the movement between loci is the result of a fusion of morphemic and deictic gestural elements (Liddell 2000, 2011, *inter alia*). Some corpus data from British Sign Language (BSL) has been presented in order to support the gestural analysis (Fenlon et al. 2018), building on the claim that this movement/location modification is not obligatory in the language.

Strong arguments against the gestural analysis have already been presented in the literature (Lillo-Martin and Meier 2011; Lourenço and Wilbur 2018; Quer 2011; Wilbur 2013). Additionally, I have shown that, if the right marker is considered – change of location and not path movement – agreement is indeed the rule, not the exception (Lourenço 2018b, 2020b).

Still, one may wonder why agreement sometimes “fails” to show up, even when the verb could be marked for agreement. Taking some data from the Libras Corpus, I will show that this apparent “optionality” of agreement can be explained if we take a closer look at the different syntactic environments. Before presenting the corpus data, the next section will outline how this spatial agreement system can be handled within the scope of a formal theory of syntax.

4 Deriving Agreement Syntactically

Since the earliest formal analysis of agreement in signed languages, two main approaches have been pursued in the literature in order to explain this phenomenon: thematic accounts, mainly in terms of lexical-conceptual structures (Bos 2017; Meir 1998, 2002), and pure syntactic ones, which assume that agreement is the result

of underlying syntactic structures and operations (Costello 2015; Lourenço 2018b; Padden 1988; Pfau et al. 2018; de Quadros 1999).³

One piece of evidence in favor of a syntactic account for agreement in Libras is that agreeing and non-agreeing (body-anchored) verbs behave syntactically differently in the language (Lourenço 2017; Lourenço and de Quadros 2020; de Quadros 1999; de Quadros and Lillo-Martin 2010). The presence or absence of agreement impacts on word order flexibility, null argument licensing, and negation distribution.

As shown before, Libras is an SVO language, but the presence of agreement allows for more flexibility in the ordering of the signs. On the other hand, non-agreeing verbs show a more restricted word order.

Order	Type of verb	Grammaticality	Example
SVO	Agreement verbs	✓	JOHN _a aHELP _b MARY _b
	Non-agreeing verbs	✓	JOHN _a aLIKE MARY _b
SOV (object-shift constructions)	Agreement verbs	✓	JOHN _a MARY _b aHELP _b
	Non-agreeing verbs	*	*JOHN _a MARY _b LIKE
OSV (object topicalization)	Agreement verbs	✓	<MARY _b > JOHN _a aHELP _b
	Non-agreeing verbs	*	*<MARY _b > _{topic} JOHN _a LIKE
		✓, iff OSVO (Resumptive strategy)	<MARY _b > _{topic} JOHN _a LIKE IX _b Lit. ‘Mary, John likes her.’

Additionally, agreement verbs only license argument dropping, whereas non-agreeing verbs do not (de Quadros 1995, 1999).

- (5) (I) ₁HELP₂ (YOU)
 “I help you.”
- (6) *(I) LIKE *(YOU)
 “I like you.”

Finally, agreement also influences the distribution of negation. In Libras, negation is marked by two different elements: the sign NO (lexical negation) and a negative NMM (glossed as ______{neg}). The following examples show the structure of negation in agreeing constructions and in non-agreeing ones. Note that preverbal

³ For good reviews of the different approaches, see Mathur and Rathmann (2012) and Quer (2021).

$(x) \rightarrow (p)$ is a discourse option.⁴ Therefore, discourse [location] will be merged during the syntactic computation inside the DP. Given the very close relationship between location and person (Fig. 3), I assume that [location] is inserted on D° .

Since agreement verbs do not have a lexical location, they enter the derivation with an unvalued [location:___] feature. Double agreement verbs, the ones that have directional path movement, have two phonological slots for location specifications and, therefore, two unvalued [location:___] features. Single agreement verbs, which are articulated in respect to a single point in space, have only one phonological slot for location specification and, therefore, a single unvalued [location:___]. Finally, body-anchored verbs are inserted into the derivation already fully specified for location [location:val].

- (9) a. Double agreement verbs:
Two unvalued [location] features \rightarrow [location:___]VERB[location:___]
b. Single agreement verbs:
One unvalued [location] feature \rightarrow VERB[location:___]
c. Non-agreeing verbs:
Lexically valued [location] feature \rightarrow VERB[location:val]

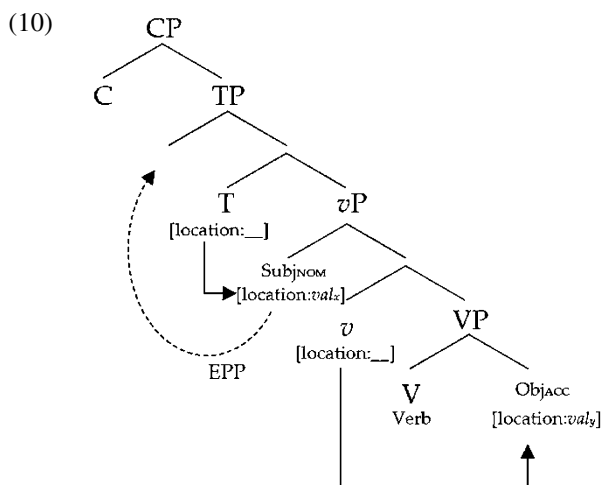
Once the relevant feature is identified, one must describe the underlying structure and operations that trigger verbal agreement in Libras. In some recent frameworks of minimalist syntax, agreement has been treated as the spell-out of a specific *Agree* operation – a probe-goal relation – in which the goal values a given feature (usually φ -features) of the probe. Subject agreement has been taken as the result of the *Agree* operation that involves the φ -probe that is merged on C° and percolates down to T° (Chomsky 2008; Miyagawa 2010) and the φ -features of the subject DP. Object agreement is also the result of an *Agree* relation, but between the φ -probe in v° and the φ -features of an object DP.

As stated in (9), double agreement verbs bear two agreement markers, subject and object agreement. Considering that [location] is the relevant feature under discussion here and that it is part of the φ -feature bundle, I will, for the sake of simplicity, only annotate the location probe on T° and on v° as [location:___]. So, the other types of φ -probes and φ -features are intentionally omitted.⁵ The [location:___] feature in T° probes down the tree and agrees with the external argument, the subject

⁴ The intended meaning of “discourse” here is the idea that a given value for [location] is not part of the lexical specification of a noun or a verb. Another possibility would be to call this type of [location] feature “referential [location],” reinforcing the idea that a location is a mapping between an entity (x) and a geometric point (p). Regardless of terminology, the main idea here is that [location] is a formal feature relevant for syntactic computation. Recently, we have claimed that the choice of the geometrical point (p) comes from the interaction of different formal features, such as person, definiteness, and specificity (Lourenço 2020a).

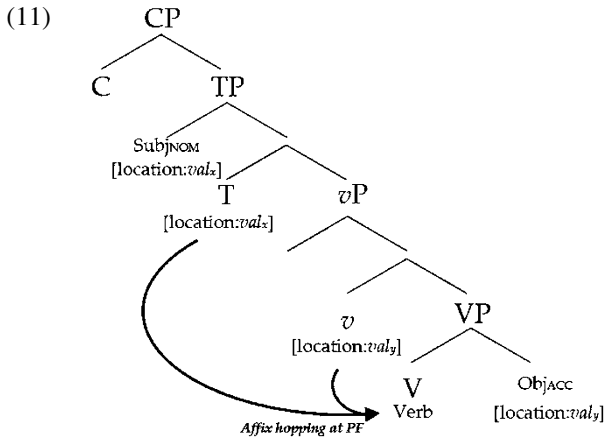
⁵ It has been argued that each unvalued feature acts like an independent probe (Pesetsky and Torrego 2007). Therefore, the probing operation of the [location:___] feature is independent of the other φ -probes present in the functional head.

DP. In the same fashion, the [location:___] feature in v° probes down and agrees with the object DP. The *Agree* relations are illustrated in (10):



It has been argued that the verb in Libras does not move up to T° (Lourenço 2014; de Quadros 1999), not even to v° (Lourenço 2014; Lourenço and Duarte 2014). However, the now valued [location:val] features in T° and in v° need to be pronounced on the verb. This could be analyzed as a case of affix hopping, also present in other spoken languages, such as English. In simple terms, affix hopping is the result of a morphological operation by which an affix that sits on T° , for instance, is lowered onto a verb. This attachment occurs in a post-syntactic component, at PF, and it gives rise to the spelling out of the agreement markers on the verb ($[_{location:val.x}]VERB[_{location:val.y}]$).

Building on this derivation, Lourenço (2018a) analyzes other types of verbs in Libras and their respective agreement patterns, by adopting a minimalist syntactic spine (C-T-v-V), the minimal operations *Merge* and *Agree*, the feature [location], and the distinction between valued and unvalued features. In a nutshell, the proposal is that agreement verbs enter the derivation with an unvalued [location:___] feature that receives its value by agreeing with different arguments, during the syntactic computation. The difference between double and single agreement verbs rests in how many unvalued [location:___] features the verb has in its lexical specification. On the other hand, non-agreeing verbs, which are body-anchored verbs, have a valued [location:val] feature as part of their lexical entry. The presence of a lexically specified value blocks verb agreement and also impacts on the subsequent syntactic derivations.



5 The Pervasiveness of Agreement in Corpus Data

In spite of the fact that many formal analyses of verb agreement in different sign languages have treated this issue as a truly linguistic phenomenon, proposing different formal implementations (Costello 2015; Janis 1995; Lillo-Martin and Meier 2011; Lourenço 2018b; Meier 1990; Meir 2002; Pfau et al. 2018; de Quadros 1999; Wilbur 2013), some researchers have argued that what we have called agreement is not agreement at all. As mentioned before, the opposing view is that this type of verb modification is actually a fusion of morphemic and gestural elements (Fenlon et al. 2018; Liddell 2003, 2011).

One piece of argument against the agreement analysis is that the verb modification is claimed to be optional and not obligatory, as one would (maybe wrongly) expect from an agreement system. For instance, Fenlon et al. (2018) observed that this type of verb modification is not really systematic and straightforward in some data taken from the British Sign Language Corpus. The authors claimed that the modification of the verb could be conditioned by different factors – such as coreference, verb position, presence of constructed action, person, and animacy – and that modification seems to be optional. Fenlon and collaborators recognize that the fact that the marking is not consistently present in the data does not constitute evidence against the agreement analysis, as agreement optionality is also observed in spoken languages. It is important to note that the traditional definition of agreement/modification as directional path movement was used by Fenlon et al. (2018). Thus, it would be interesting to see how pervasive agreement really is in the data, if one considers agreement to be co-localization.

With that in mind, we decided to look at the Libras Corpus examining if the so-called optionality of agreement is equally true for Libras. I will now show some preliminary results of an ongoing investigation.

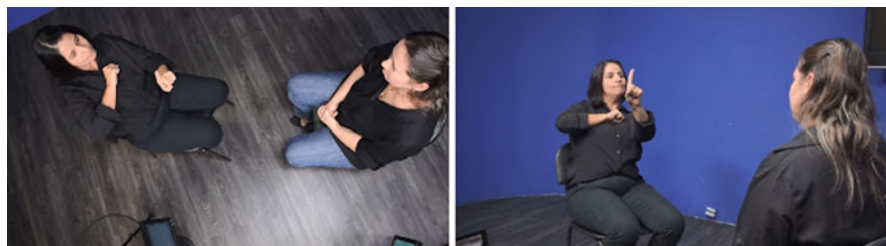


Fig. 7 The views from the top and the frontal camera

For this study, the signing of three Deaf people (two women), all of them from the state of Minas Gerais, was analyzed. The recordings are part of the Libras Corpus Project (de Quadros et al. [n.d.](#)) that comprises different initiatives to collect and register different types of signing – from vernacular to literary signing – and from Deaf people of different ages and from different parts of Brazil. The three videos we analyzed were interviews about the subjects’ life stories and they sum a total of 01 h, 38 m, 28 s of video data.

All data collection within the Libras Corpus Project makes use of four different cameras, so signing was captured from four different angles. For our particular study, the frontal camera and the top camera were relevant and helped us identify the position of the hands in space, as shown in Fig. 7. All transcription and annotations were done on ELAN (Crasborn and Sloetjes 2008).

The sample was analyzed for the presence versus the absence of verb agreement, assuming co-localization as the agreement marker. Verbs that had agreement were further classified in single or double agreement. Also, single agreement verbs were further analyzed with respect to the location on which the agreement marker took place (see Fig. 3): first-person location, second-person location, or third-person location (ipsilateral or contralateral side of the body). Non-agreeing verbs had their phonological shapes classified as body-anchored or non-body-anchored, since we have posited that body-anchoring is a phonological feature that blocks agreement marking.

Only data from the Deaf participants were analyzed, the interviewer signing was excluded. Additionally, lexical items that were not clearly a verb were not included in the study, neither were the signs that could even have a verb meaning but were produced with a mouthing (mouth movement that mimics a spoken word) that referred to a noun word (e.g., the verb TEACH signed while the signer produces the spoken word “education”). Finally, due to their complex polymorphemic structure, classifier constructions were also excluded from our data. At the end, a total of 1570 verb tokens were identified and analyzed.

A primary observation is that agreement in our data was much more pervasive and present than what has usually been assumed in the literature as you can see in Table 1. If we go back to claims explicitly stating that “the agreement process in sign languages is restricted to a smaller set of verbs” (Mathur and Rathmann

Verbs with agreement:	Verbs without agreement:
Double agreement	Body-anchored verbs
Subject location:	Non-body-anchored verbs
First person	
Second person	
Third person	
Object location:	
First person	
Second person	
Third person	
Single agreement	
First-person location	
Second-person location	
Third-person location	

Table 1 Tabulation of agreement vs. non-agreement verb tokens

	Count	% of total
<i>Verbs with agreement</i>	735	46.82%
Double agreement	412	26.24%
Single agreement	323	20.57%
First-person location	50	3.18%
Second-person location	9	0.57%
Third-person location	263	16.75%
Ipsilateral side	155	9.87%
Contralateral side	108	6.88%
<i>Verbs without agreement</i>	835	53.18%
Body-anchored verbs	483	30.76%
Non-body-anchored verbs	352	22.42%

2012: 152), we can see that this is not the case. Almost half of the verb tokens in our data displays agreement (46.82%). Therefore, agreement cannot be considered a marginal phenomenon in Libras grammar.

Additionally, it is worth mentioning that if one considers only double agreement (path movement between two locations) as agreement marker – as it has been traditionally assumed in previous works on the matter – only 26.24% of verb tokens would be considered to have agreement. Thus, we would miss half of the occurrences of agreement in our data, leading us to the wrong conclusion that agreement is the exception and not the rule. This is why we have claimed that agreement should be analyzed only in terms of location and not movement. Additionally, we have argued that movement modifications are associated to the event structure of the predicate and aspectual modifications, while location modifications (co-localization) on the verb are related to agreement (Lourenço and Wilbur 2018:76).

A closer look at the single agreement verbal tokens, those with co-localization but not path movement, reveals that most of the occurrences are instances of third-

person agreement. As discussed earlier, third-person locations are mapped on the contralateral or the ipsilateral side of the signing space. Second-person agreement was scarce in the data, but this could be an effect of the type of discourse we analyzed. Considering that the signers were talking about their own life stories and that they were being interviewed, it is reasonable to assume that most of the discourse would not make reference to the addressee. However, it is interesting to notice that first-person agreement marking was also not that frequent.

First-person locus has usually been described as the space closer to the signer's body or chest (Friedman (1975), among others). However, one of the challenges of analyzing first-person agreement by co-localization is identifying how close is close enough to be considered actual first-person agreement. When it comes to double agreement marking, it is easier to identify first-person markers because the hand(s) moves from or toward the signer's body, and direction of movement is an extremely salient property. In this case, there are two timing units, each one with its own location specification, creating, thus, person-feature distinctions. Therefore, it is clear that the verb CATCH shown in Fig. 8 has its location specifications changed in order to match first-person and third-person locations. However, it is not that easy to discriminate first-person marking when it comes to single agreement verbs. The criterion we adopted to categorize a verb as carrying first-person agreement was that the hand should be brought closer to the signer's body, compared to a non-agreeing item. In Fig. 9, you can compare the position of the hand of the verb OPPRESS that we have considered to show first-person agreement and the sign UNTIL that is not marked for agreement at all. Notice, nevertheless, that this difference is not crystal clear, which will lead us to question the notion of first-person location in what follows.

Let us now turn our attention to the non-agreeing verbs. Remember that in our descriptions of agreement as co-localization, we explicitly state that verbs will show agreement unless they are constrained not to do so; and this constraint is

Fig. 8 Verb CATCH marking first-person agreement at the first slot/timing unit and third-person agreement at the second slot/timing unit (Libras Corpus)





Fig. 9 On the left, the verb *OPPRESS* showing first-person agreement and, on the right, the sign *UNTIL*, which is not agreement marked (Libras Corpus)

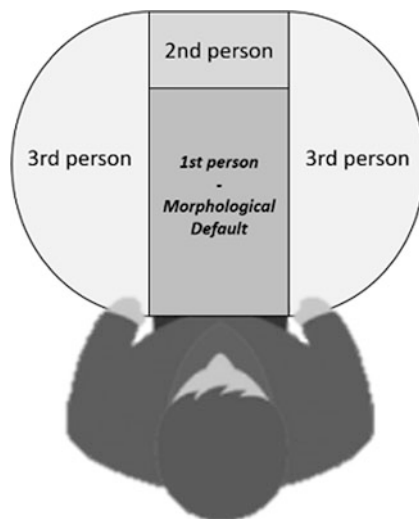
phonological in nature, to wit: body-anchoring. As we can see in Table 1, 30.76% of the verbs in our sample do not show agreement and are body-anchored verbs. Moreover, if we conflate the body-anchored non-agreeing verbs with the verbs that do show agreement (both single and double), we end up with the observation that 77.58% of all the verb tokens behave exactly as predicted. This observation alone seems to contribute to the analysis of Libras as an agreement language, challenging claims that the systematic modification of verbs in signed languages is not grammatical agreement or that it is just a peripheral phenomenon in the grammar.

Still, there are 22.42% of verbs that are not body-anchored and yet were not marked for agreement. Why don't these verbs agree? Does this suggest an optionality in the system? Or are all those verbs exceptions to the agreement rule in the language? Good answers to these questions are yet to be formulated, but we already have some initial observations that are quite interesting in itself, while strengthening the agreement analysis we propose.

Firstly, it is to be noticed that, in our sample, 154 of the 352 non-body-anchored non-agreeing verbs have a first-person subject. To put it differently, in 44.25% of the times that agreement “fails” to occur, the verb had a first-person agreement. There are two possible ways to approach this. First, we may argue that first-person agreement is indeed optional, being less frequently marked in Libras. Second, we may take first-person agreement as default agreement; hence all unmarked occurrences we observed are actually instances of morphological default agreement. Pfau et al. (2018), for instance, have analyzed the omission of subject agreement marker as an instance of default agreement. In this sense, the location right in front of the body of the signer would be a morphological agreement default.

That said, we can now come back to the previous discussion on how close to the body the verb must be signed in order to be considered as marked for first-person agreement. Actually, we would argue that the distribution of the space morphology presented in Fig. 3 should be revised to a more accurate mapping of the space, as represented in Fig. 10. In this representation, the whole space in front of the signer's body is considered as first-person marker, which functions as a morphological default in the language. This would also explain why it has been so hard to set

Fig. 10 Distribution of person features in the signing space



a criterion for assuming a first-person agreement in our sample and will create a simple opposition in the space system: close to the body (first person as default) vs. far from the body (second person).⁶

By assuming this morphological default, we are now left with the other 198 verb tokens that are not body-anchored, do not show agreement, and do not have a first-person subject. Considering that this represents only 12.61% of the verb tokens in our data, it would be already safe to assume that our model of grammatical agreement has been demonstrated to be adequate to deal with this systematic modification of verb locations in Libras. The fact that agreement is not always there is definitely not an argument against calling it agreement. It is well known that morphosyntactic features can be optionally marked, not only in signed languages but in spoken languages as well. Case markers, verb agreement markers, and number concord, for instance, are just examples of formal features that can be deleted or realized by a morphological default in many languages, especially in colloquial speech. Therefore, it should be fine to assume that some “non-occurrence” of agreement is natural and even expected.

However, when we look at this remaining data individually, some properties of the sentences in which these verbs appear catch our attention, especially regarding their syntactic configuration. Some selected examples will be given below.

⁶ One may argue that this analysis still faces the problem of how to draw the distinction between “close” and “far” in the signing space. However, we should recall that the actual physical point is not relevant for the organization of the grammatical space. Instead, the geometric point is (Quer 2011; Wilbur 2013). Geometric points are categorically perceived and bear distinctiveness properties. Therefore, the relevant notion is not *where* is “close” and *where* is “far,” but the categorically perceived distinction between “close” and “far.”

First, many examples⁷ seem to be of non-finite predicates. As it is generally assumed in generative theory, there is a close relation between finiteness and agreement. Therefore, it is not a surprise that some instances of verbs that are not marked for agreement in our data, examples (12) and (13) below, are actually non-finite verbs.⁸ There are also examples of [V + V] structures, in which the second predicate does not show agreement. These could be also considered to be non-finite (14).

- (12) PALM-UP SEW_∅ PALM-UP ARCHIVE_∅ PALM-UP SIGN-LANGUAGE ADVISE₁.
“(I) was advised/taught how to sew and how to archive in sign language.”
- (13) PUT_∅ INTERPRETER START_∅ SECOND YEAR
“Putting interpreters started (when I was) in my second year.”
- (14) DAD FAMILY [HAVE USE_∅] GESTURE OWN LANGUAGE POSSESSIVE_A OWN.
“My dad’s family ‘used’ gestures, their own language.”

Another relevant observation regards the features present in the subject of the sentence. More specifically, agreement will only be marked if the subject does have a location feature. For instance, the verb *START* in (16) has a clausal subject that is not mapped into a point in space. Since there is no $(x) \rightarrow (p)$ mapping, there is no location and, therefore, no verb agreement.

This also seems to be the case when the subject itself is body-anchored. Considering that body-anchoring does not constitute a location feature on D° , it is not available for agreement in Libras (Lourenço 2018a), and therefore there is no verb agreement. This is illustrated in (15), where the subject *DEAF* is a body-anchored sign.

- (15) DEAF START_∅ TALK-ORAL_∅
“Deaf started to speak orally.”

One last interesting piece of data to be discussed is the sentence in (16) and in Fig. 11.

In this example, the verb *HAVE-NOT* occurs twice and curiously its first occurrence has agreement, whereas the second does not. Verb doubling in Libras

⁷ Remember that the lack of agreement here could indeed be analyzed as default agreement morphology, meaning that the verb is signed in the signing space right in front of the signer, as the above discussion of first-person agreement as the morphological default shows.

⁸ We want to be careful here, though, because there are no descriptions of (non-)finiteness in Libras, so it is hard to assure that these examples indisputably show non-finite predicates.

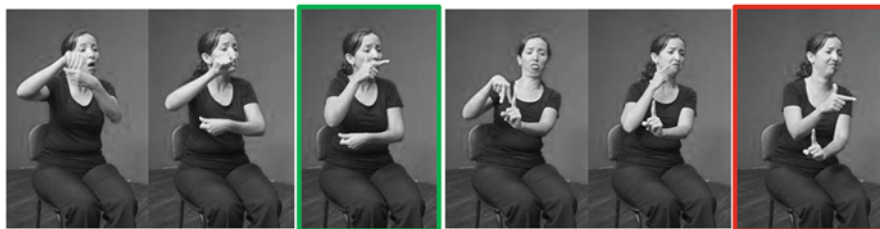


Fig. 11 INSIDE_A ASSOCIATION_A HAVE-NOT_A SEE-BAD DEAF $\text{HAVE-NOT}_\emptyset$ (Libras Corpus)

- (16) INSIDE_A ASSOCIATION_A HAVE-NOT_A SEE-BAD DEAF $\text{HAVE-NOT}_\emptyset$
 “In the [deaf] association, there is no bad perspective on deafness.”

has been analyzed as focus constructions (de Quadros 1999; de Quadros and Lillo-Martin 2010), in which the focalized verb sits on a position higher in the clause, being spelled out at the end of the sentence. What is interesting in these cases is that the lower occurrence of the verb which occupies the canonical verb position in the sentence, being within the syntactic domain in which agreement takes place, is marked with agreement, while the higher occurrence, which is outside the agreement syntactic domain, is not.

In sum, what we have shown above indicates that the so-called optionality for agreement does not seem to hold when a more fine-grained description of what agreement looks like in signed languages is assumed and when we identify the specific restrictions (phonological or syntactic) that can block agreement marking. This leads us to a stronger hypothesis about agreement in Libras:

Agreement in Libras

All verbs are always marked for agreement (default vs. specific morphology), unless they are phonologically constrained not to do so (body anchored) or agreement is not available syntactically.

6 Final Remarks

In this chapter, I presented an outline of the grammar of agreement in Libras. Not only I demonstrated that agreement in space, that is, typical of signed languages, can be analyzed by implementing formal theoretical tools, but I have also shown that, although agreement in signed languages might seem quite different from agreement

in spoken languages, the grammatical operations that make agreement available in natural languages are universal and not specific to a given modality.

This approach also supports the idea that the systematic modification of the location of verbs in signed languages is indeed syntactic agreement and that it is not an exceptional or a peripheral phenomenon in Libras grammar. By looking at corpus data, I have shown that agreement is not restricted to a subset of verbs, being actually pervasive and more productive than it has been argued, thus challenging claims that it is not agreement. Also, by positing phonological (body-anchoring) and syntactic (structural) constraints on this derivational mechanism, we provide a better understanding of the contexts in which it “fails” to show up.

References

- Arrotéia, Jéssica. 2005. *O papel da marcação não-manual nas sentenças negativas em Língua de Sinais Brasileira (LSB)*. MA Thesis, Universidade Estadual de Campinas.
- Bertucci, Roberlei Alves, and Rossana Aparecida Finau. 2018. Uma descrição inicial do presente perfeito na Libras. *Cadernos de Estudos Lingüísticos* 60: 71–89. <https://doi.org/10.20396/cel.v60i1.8649763>.
- Bos, Heleen F. 2017. An analysis of main verb agreement and auxiliary agreement in NGT within the theory of conceptual semantics (Jackendoff 1990). *Sign Language and Linguistics* 20: 228–252.
- Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory*, ed. Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta, 133–166. Cambridge, MA: The MIT Press.
- Costello, Brendan. 2015. *Language and modality: Effects of the use of space in the agreement system of Lengua de Signos Española (Spanish sign language)*. University of Amsterdam and the University of the Basque Country.
- Crasborn, Onno, and H. Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. In *Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation*.
- de Quadros, Ronice Müller. 1995. *As categorias vazias pronominais: uma análise alternativa com base na língua de sinais brasileira e reflexos no processo de aquisição*. PUCRS.
- . 1999. *Phrase structure of Brazilian sign language*. Pontifícia Universidade Católica do Rio Grande do Sul.
- . 2011. *Sintaxe das línguas gestuais*. Lisboa: Universidade Católica Editora.
- de Quadros, Ronice Müller, and Diná Souza da Silva. 2017. As comunidades surdas brasileiras. In *omunidades Surdas na América Latina: Língua – Cultura – Educação – Identidade*, ed. R.C. Zambrano and C.E.F. Pedrosa, 132–152. Florianópolis: Editora Bookess.
- de Quadros, Ronice Müller, and Lodenir Becker Karnopp. 2004. *Língua de sinais brasileira: estudos lingüísticos*. Artmed Editora.
- de Quadros, Ronice Müller, and Diane Lillo-Martin. 2010. Clause structure. In *Sign languages*, ed. Diane Brentari, 225–251. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511712203.012>.
- de Quadros, Ronice Müller, Deonísio Schmitt, Juliana Lohn, and Tarcísio de Arantes Leite. n.d. *Corpus de Libras*.
- de Souza, Isaac Gomes Moraes. 2016. *Concordância verbal e a hipótese do período crítico em Libras: Um estudo teórico-experimental*. Pontifícia Universidade Católica do Rio de Janeiro.
- Felipe, Tanya. 1998. *A relação sintático-semântica dos verbos e seus argumentos na língua brasileira de sinais (libras)*. Universidade Federal do Rio de Janeiro.

- Fenlon, Jordan, Adam Schembri, and Kearsy Cormier. 2018. Modification of indicating verbs in British sign language: A corpus-based study. *Language*: 0–59.
- Ferreira-Brito, Lucinda. 1990. Epistemic, alethic, and deontic modalities in a Brazilian sign language. In *Theoretical issues in sign language research*, ed. Susan D. Fischer and Patricia Siple, vol. 1, 224–260. Chicago: University of Chicago Press.
- Figueiredo, Lorena Mariano Borges, and Guilherme Lourenço. 2019. O movimento de sobranceiras como marcador de domínios sintáticos na Língua Brasileira de Sinais. *Revista da Anpoll* 1: 78. <https://doi.org/10.18309/anp.v1i148.1235>.
- . 2020. Analisando a Língua Brasileira de Sinais como uma língua sem-tense. *Scripta* 24: 361. <https://doi.org/10.5752/p.2358-3428.2020v24n51p361-396>.
- Finau, Rossana Aparecida. 2004. *Os sinais de tempo e aspecto na LIBRAS*. Universidade Federal do Paraná.
- Friedman, Lynn A. 1975. Space, time, and person reference in American sign language. *Language* 51: 940–961.
- Janis, Wynne D. 1995. A crosslinguistic perspective on ASL verb agreement. In *Language, gesture, and space*, ed. Karen Emmorey and Judy Reilly, 195–224. Psychology Press.
- Liddell, Scott K. 2000. Indicating verbs and pronouns: Pointing away from agreement. In *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, ed. Karen Emmorey and Harlan L. Lane, 303–320. Mahwah: Lawrence Erlbaum.
- . 2003. *Grammar, gesture, and meaning in American sign language. Grammar, gesture, and meaning in American sign language*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511615054>.
- . 2011. Agreement disagreements. *Theoretical Linguistics* 37: 161–172.
- Lillo-Martin, Diane, and Richard P. Meier. 2011. On the linguistic status of “agreement” in sign languages. *Theoretical Linguistics* 37: 95–141.
- Lourenço, Guilherme. 2014. *Concordância, caso e ergatividade em Língua de Sinais Brasileira: Uma proposta minimalista*. MA Thesis, Universidade Federal de Minas Gerais.
- . 2015. Sentential negation in Brazilian Sign Language. In III CIEL- Congresso Internacional de Estudos Linguísticos, Brasília, Brazil.
- . 2017. Assimetria entre verbos de concordância e verbos simples em Língua Brasileira de Sinais. *Entrepalavras* 7: 15–35.
- . 2018a. *Verb agreement in Brazilian sign language: Morphophonology, syntax and semantics*. Universidade Federal de Minas Gerais.
- . 2018b. Layering de informações visuais e a estrutura morfofonológica dos verbos em Libras. In *II Congresso Nacional de Pesquisas em Linguística e Libras*. Florianópolis: UFSC.
- . 2020a. Redefinindo o conceito de concordância verbal em Língua Brasileira de Sinais. In *Estudos da Língua Brasileira de Sinais*, ed. Carlos Henrique Rodrigues and Ronice Müller de Quadros, vol. V, 115–136. Florianópolis: Editora Insular.
- . 2020b. A morfologia do espaço na Língua Brasileira de Sinais: uma análise a partir de coordenadas cartesianas tridimensionais. In *XXXV Encontro Nacional da ANPOLL - ENANPOLL*.
- Lourenço, Guilherme, and Ronice Müller de Quadros. 2020. The syntactic structure of the clause in Brazilian sign language. In *Brazilian sign language studies*, ed. Ronice Müller de Quadros, 131–154. De Gruyter Mouton.
- Lourenço, Guilherme, and Fábio Bonfim Duarte. 2014. Caso e concordância em Língua de Sinais Brasileira: Investigando verbos de concordância regular e verbos de concordância reversa. *Veredas* 18.
- Lourenço, Guilherme, and Ronnie B. Wilbur. 2018. Are plain verbs really plain? Co-localization as the agreement marker in sign languages. *FEAST. Formal and Experimental Advances in Sign language Theory* 2. <https://doi.org/10.31009/feast.i2.06>.
- Mathur, Gaurav, and Christian Rathmann. 2012. Verb agreement. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 136–157. Berlin/Boston: De Gruyter Mouton.

- Meier, Richard P. 1990. Person deixes in American sign language. In *Theoretical issues in sign language research, volume 1: Linguistics*, ed. Susan D. Fischer and Patricia Siple, 175–119. Chicago: The University Chicago Press.
- Meir, Irit. 1998. *Thematic structure and verb agreement in Israeli sign language*. Hebrew University of Jerusalem.
- . 2002. A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory* 20: 413–450.
- Miyagawa, Shigeru. 2010. *Why agree? Why move: Unifying agreement-based and discourse-configurational languages*, Linguistic inquiry monograph 54. Cambridge: MIT Press.
- Neidle, Carol, Judy Kegl, Dawn Maclaughlin, Benjamin Bahan, and Robert G. Lee. 2000. *The syntax of American sign language: Functional categories and hierarchical structure*. Cambridge, MA: The MIT Press.
- Padden, Carol. 1988. *Interaction of morphology and syntax in American sign language*. New York/London: Garland Publishing.
- . 1990. The relation between space and grammar in ASL verb morphology. In *Sign language research: Theoretical issues*, ed. Ceil Lucas, 118–132. Washington, DC: Gallaudet University Press.
- Perniss, Pamela. 2012. Use of sign space. In *Sign language: An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 412–431. Berlin/Boston: De Gruyter Mouton.
- Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: Syntactic derivation and interpretation (in honor of Joseph E. Emonds)*, ed. Simin Karimi, Vida Samiian, and Wendy K. Wilkins, 262–294. Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/la.101.14pes>.
- Pfau, Roland, Martin Salzmann, and Markus Steinbach. 2018. The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: A Journal of General Linguistics* 3: 1–46.
- Quer, Josep. 2011. When agreeing to disagree is not enough: Further arguments for the linguistic status of sign language agreement. *Theoretical Linguistics* 37: 189–196.
- . 2021. Verb agreement: Theoretical perspectives. In *The Routledge handbook of theoretical and experimental sign language research*, ed. Josep Quer, Roland Pfau, and Annika Herrmann, 95–121. New York: Routledge.
- Royer, Miriam. 2019. *Análise da ordem das palavras nas sentenças em Libras do Corpus da Grande Florianópolis*. Universidade Federal de Santa Catarina.
- Wilbur, Ronnie B. 2013. The point of agreement: Changing how we think about sign language, gesture, and agreement. *Sign Language and Linguistics* 16: 221–258.
- Wilbur, Ronnie B., and Cynthia Patschke. 1999. Syntactic correlates of brow raise in ASL. *Sign Language and Linguistics* 2: 3. <https://doi.org/10.1075/sll.2.1.03wil>.
- Xavier, André Nogueira, and Sherman Wilcox. 2014. Necessity and possibility modals in Brazilian sign language (Libras). *Linguistic Typology* 18. <https://doi.org/10.1515/lingty-2014-0019>.

Argument Structure in Peruvian Sign Language



Miguel Rodríguez-Mondoñedo 

1 Introduction

Peruvian Sign Language (LSP, its Spanish acronym) is the language created by the Deaf community in Peru. It comprises several varieties, both geographical and generational—some of which are argued not to be mutually understandable (Parks and Parks 2009, 2010, Clarks 2017a, b). Here I examine the argument structure of the LSP variety spoken in Lima, the capital, with data from users between 14 and 40 years old—mainly taken from the PUCP corpus (Rodríguez-Mondoñedo et al. 2015). To my knowledge, no study has addressed the properties of argument structure of LSP; in fact, there are almost no studies describing any grammatical aspect of LSP—exceptions are Clark (2017b), Madrid (2018), Cuti (2018), Rodríguez-Mondoñedo and Arnaiz (2020), and Rodríguez-Mondoñedo (2021), but several investigations are in the making, as we will see.

Given that very little is known about LSP or the Peruvian Deaf community, I will spend some time explaining, in Sect. 2, its basic sociolinguistic traits, and in Sect. 3, the fundamentals of LSP grammar, including the issue of externalization. In Sect. 4, I will describe LSP argument structure, in particular, its connections with the classifier system. Section 5 will be dedicated to the questions this analysis raises

I'd like to express my deep thanks the several Deaf consultants that made research on LSP possible, and also to Alexandra Arnaiz, LSP interpreter, for her invaluable help. Many thanks, too, to Cilene Rodrigues and Andrés Saab for several comments that greatly improved the shape of this chapter. Any remaining flaw is my own.

M. Rodríguez-Mondoñedo (✉)
Pontificia Universidad Católica del Perú, San Miguel, Peru
e-mail: mmondon@pucp.pe

for studies on some independent issues. The last section presents the conclusions to be drawn from the present study.

2 An Invisible Language

LSP has been completely ignored by linguists and other academics for a long time. Notwithstanding a lexical repertoire with drawn pictures published by the Peruvian Department of Education before (MINEDU 1987), we can say that the first published account of LSP and Deaf culture was Paliza (1994), who, based on his own experiences as a leader in the Deaf community, featured a brief presentation of some aspects of the Peruvian Deaf experience, including a few remarks on LSP. According to Paliza, LSP has received, at least in the twentieth century, the influence of Spain Sign Language, given that Deaf leaders from Spain visited Peru, and the first School for the Deaf (today, CEBE 07 La Inmaculada de Barranco) was created in 1939 by religious people from Spain. It must be said that this school is oralist, i.e., its goal is to make Deaf people acquire oral skills, although it does not prohibit signing, which has promoted cultural interchanges between local Deaf people. As Paliza mentions, however, the main problem of the Deaf people in Peru is that they are isolated, without a full system for their education and complete development of their identity. Yet, the Inmaculada group was the seed for the foundation of the first National Association of the Deaf in 1958, whose members started the first discussions on LSP signs and their use.

Another landmark in the history of LSP is the arrival of Vernon Miller, a Deaf Baptist minister, who founded EFATA Church in 1971, and the first proper school for the Deaf in Villa El Salvador, one of the poorest districts in Lima at that time. The innovation that Miller brought to Deaf education was the use of sign language for school classes; this enormously impacted the Deaf community and produced two long-lasting results. First, a full generation of young Deaf students was educated and, in time, became empowered leaders of their communities. Second, given that Miller and his collaborators were American Sign Language (ASL) signers, ASL influenced LSP in various ways, creating another layer of fragmentation in the language, and steamed linguistic conflicts regarding the “purity” of local signs.¹ Other religious denominations, such as Adventists and Jehovah’s Witnesses, also brought ASL to their work with Deaf people in Peru.

In fact, the few researchers that have investigated LSP have found that there is a great deal of variation inside. Parks and Parks (2009) were the first to attempt a sociolinguistic survey on the language. Interviewing Deaf individuals in several major cities in the country, they concluded that the vast majority of Deaf signers

¹ It is interesting to notice how effortlessly a prescriptive discourse arises in a linguistic community. Of course, this is not specific of Deaf communities, as can be easily confirmed by the myriad of prescriptivist discussions oral language speakers have all over the world.

considered Lima LSP variety the most prestigious and the one they could understand well, even if most of them had the perception that LSP was not uniform across all Peru. Clark (2017b) claims that in Sivia, Ayacucho, there is a Deaf community signing a very distinct version of LSP, and Clark (2017a) even observes that, in Lima, younger generations of Deaf people sign a significantly different variety. We limit the data for the present work to younger Lima signers (20–40 years old).

As the Deaf community became more organized and empowered, with an increasing access to education and information, Deaf associations and cultural and sports clubs started to flourish, and Deaf culture slowly became more visible. Thus, Deaf people became more conscious about their rights and more capable of moving their organization toward political actions. After years of lobbying by Deaf activists in the public sphere, the National Congress declared LSP an official Peruvian language in 2010 (Law 29535). This made possible the creation of the first public school for the Deaf (CEBE Ludwig van Beethoven), where classes are taught in LSP, for primary education. Later other schools were implemented for secondary education as well, with classes being offered in Spanish, with the use of interpreters. In addition, a number of small private schools have been created, also teaching in LSP, mainly under the leadership of parents of Deaf children and collaborators.

So far, this resembles the path of many other Deaf communities, which have grown and flourished around the educational system available to them. Many challenges rest ahead, however. The official recognition of LSP is not complete. In 2011, a Law of Languages (Law 29735) was approved, and not only did it not mention LSP at all, it established a rule for a language to be considered inside the scope of publicly funded linguistic policy, namely, that it should be “originaria” (lit. original), by which it means a language present in the territory before the arrival of Spanish (article 3, Law 29735). Since LSP has not left any written testimony before the twentieth century, it has been excluded from the linguistic services the State provides to minority oral communities. As such, there are no dictionaries, plans for standardization, training or certification for interpreters (all LSP interpreters are self-educated), and the like. LSP is not even listed in the roll of Peruvian languages kept by the Department of Culture, and it is frequently excluded when government institutions talk about linguistic diversity in the country, even when the officials in charge are linguists, a condition that has not prevented them, unfortunately, from claiming that LSP does not merit their attention (Olivero 2016). As a result, LSP has been almost invisible in the cultural media, and in general in the political discussion about cultural policies.

This situation blocks the expansion of the few advances obtained by the Deaf community, and it certainly slows down the fight against the linguistic deprivation suffered by Deaf children, which is widespread in the country (see Rodríguez-Mondoñedo (2020) for a discussion). Also, it makes parents and educators less likely to attempt an education fully in LSP and more likely to prefer the “inclusive” option (mainstreaming), which in Peru basically consists of seating the child in a room full of hearing students, with classes in Spanish that she/he must somehow understand—the results are, of course, disastrous, as carefully discussed in Goico (2019).

The situation is so precarious that we do not even have a reliable number of LSP signers in the country. The last census (INEI 2017) informs that 10,447 individuals reported they “learnt to speak” with LSP, so we can assume that there are at least 10,447 native LSP signers. On the other hand, the very same census reports 25,763 individuals who “neither hear, nor speak” and 192,285 individuals with “speaking disability.” Also, the first census for people with disabilities (INEI 2012) counted half a million people with hearing impairments (but it most likely included senior citizens with severe hearing loss, who are not really members of the Deaf community culturally speaking). Since no further clarifications are provided by the census, we can estimate 200,000 as potentially the highest number of Deaf individuals in the country.

Given this state of affairs, it is a sign of hope that in recent years, an interest in LSP has sprouted among young linguists, who have started researching on the grammatical aspects of the language. Madrid (2018) is the first study on LSP classifiers, Cuti (2018) investigated the anthroponomic system, Rodríguez-Mondoñedo and Arnaiz (2022) analyze some LSP copula constructions, Arnaiz (2021) and Catalán (2021) researched the properties of the Spanish written system by LSP signers, Ramos (2022) analyzes LSP agreement, Cerna-Herrera y Ramos (2022) studied relative clauses in LSP, and Malca and Domínguez (2022) studied metonymy in sign creation. In addition, more work is coming: on LSP distinctive phonological features (Raico, in preparation), LSP questions (Carlín, in preparation), mouthing (Mateo, in preparation), pronominal system (Cerna, in preparation), and hopefully much more.

The present investigation aims to contribute to the development of LSP linguistics by examining argument structure in the language. In the next section, we will present some basic and relevant aspects of LSP grammar.

3 LSP Grammar and Externalization of Arguments

As mentioned, we know little about LSP grammar; most issues are yet to be investigated. Nevertheless, a growing body of research is forming, and we aim at giving some basic detail about the language grammar in this section, offering a brief overview, limited to syntactic properties, not only because they are the most relevant for this chapter, but also because most of the research done so far on LSP has focused on syntax.

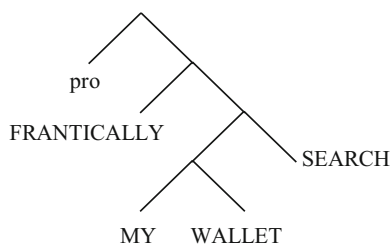
First, LSP, as many other sign languages, allows great mobility of constituents, but it seems to be a canonical SOV language, at least judging from the relatively high number of sentences that show this order, and also because native signers judge

SOV sentences as natural, except for psychological verbs, which tend to be SVO.² An example is given in (1):³

- (1) frantically
MY WALLET SEARCH
“I frantically search for my wallet.”

There are several details worth noticing in (1), besides OV order. As most (if not all) sign languages do, LSP makes use of simultaneous signs. In this case, the adverbial “frantically” is a non-manual sign, since it is produced with a face-sign, which remains in place during the whole sentence production. Of course, this is the natural result of having a language that can make use of a tridimensional space, but it means that the externalization device is able to “flat” a syntactic structure allowing for simultaneous expression of two or more structural layers. In that sense, we assume that (1) has a structure in line with the representation sketched in (2):

(2)



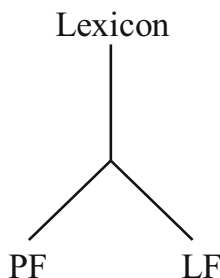
In order to focus on externalization matters, I am abstracting away from labels and a finer-grained structure, presenting only the bare-bone hierarchical structure.

² For the purpose of this chapter, a “native signer” is a LSP signer who has acquired her language from the beginning of her life, i.e., a Deaf individual who has been raised in a Deaf family. This is crucial because, given that Spanish is a SVO language, LSP interpreters (all of them Spanish speakers) tend to sign in SVO fashion, and their LSP (which appears on some TV news programs) is the one with the wider reach among the Deaf community (with occasional complaints about how some specific interpreters sign, however). In fact, interpreters’ representatives are very often asked by government officials to discuss policies regarding the Deaf community’s language. Although Deaf representatives are also invited to those meetings, their opinions have, in the best-case scenario, the same weight as that of interpreters, and quite often they are previously vetted by interpreters. This happens because the Deaf community is not sufficiently organized, and their leaders do not necessarily have the social capital needed to put forward a strong representation. How this affects LSP structure is yet to be studied. With respect to word order, our data points toward SOV, without necessarily precluding alternative analyses. Judy Shepard-Kegl (p.c.) suggested, for instance, that LSP could have a Ground-Figure order.

³ Sentence (1) was offered by the Deaf consultant after being asked (in LSP) to produce an example to illustrate the usage of the verb SEARCH.

Adopting the classical model (3) (Chomsky 1995), (2) is a pre-spell-out structure, ready to be shipped to the interfaces:

(3)



A tree like (2) in a model like (3) is thought to be linearly unordered, that is, (2) only expresses the hierarchical relations between the signs. Thus, in principle, it could be exteriorized in any order, even simultaneously. However, there are two constraints an externalization process must respect. First, following Kayne's (1994) insight, linearization must piggyback on hierarchical (asymmetric) relationships, being, therefore, extremely local, unable to target discontinuous terminal nodes—banning, for instance, a sequence like <MY, SEARCH> because there is not a single node that contains MY and SEARCH and only MY and SEARCH. Second, it must comply with the requirements imposed by the materialization channel. For oral languages, this means that terminal nodes must be externalized one by one, in linear order, as Saussure famously stated (*Cours* §3, 145). However, as witnessed by (1) sign languages are not restricted in this manner. A question arises, thus, with respect to externalization in these cases: how does it happen?

It must be noticed that sign languages are also subject to sequential linearization. In (2), MY and WALLET are signed in that order, complying with Saussure's dictum—in fact, sign languages, similarly to oral languages, can adopt different word orders (see Sandler and Lillo-Martin (2006) for discussion). An important question is how FRANTICALLY can be exteriorized at the same time as the other terminal nodes.⁴

Several attempts have been made to deal with the details of externalization in sign languages (Wilbur 2003; Sandler and Lillo-Martin 2006; Vermeerbergen and Crasborn 2007; Napoli and Sutton-Spence 2010; Kremers 2012; Lourenço and Wilbur 2018, among others). It has been claimed that up to four propositions can be simultaneously expressed in a sign language (Napoli and Sutton-Spence 2010: 650–653), under the assumption that a proposition is simply the assemblage of a predicate with its arguments. I will assume a layering system (Wilbur 2003; Lourenço and Wilbur 2018), according to which various pieces of information can be externalized

⁴ The issue of simultaneity is not restricted to the externalization of syntactic constituents. It pervades sign language phonology as well. For instance, in (2) the sign MY has different phonological features simultaneously appearing: the handshape (all fingers selected) and the location and movement (directed toward the signer's chest)—if the same handshape were directed toward the addressee, it would mean YOUR. For a recent discussion about the phonological complexities involved in simultaneous signs, see Sandler (2017).

at the same time as long as the articulation of one item does not affect the articulation of other items.

In the case of sentence (1), we are concerned with **FRANTICALLY**, which is externalized using the face articulator. In addition, it can be said that arms and even hands are conveying the same meaning; after all, when articulating **SEARCH**, the arms and hands are more tense and move faster. Furthermore, although in this sentence the verb is articulated in the neutral space (pointing slightly down), if the location of the search were to be specified (up or right, for instance), the verb articulation will point to the search-location. In other words, the different layers of the articulator *hands* convey different information: place of articulation indicates the location where the event took place, whereas tension and pace together with the upper part of the face express how the event developed (i.e., manner) and so on. The face, being an articulator, has its own layers, typically its upper part (eyes and forehead) and lower part (below nose). In (1), both face layers are used. The frown (the upper layer) conveys the property of being worried, desperate. In addition, it must be noticed that the tongue is a bit out in (1), with slightly puffed cheeks (the lower layer), which specifically conveys the desperate manner. This means that the transcription in (1) has been simplified. Properly, we should tease apart the upper and lower layer of the face; the latter is a bound morpheme attached to the verb (cheeks are puffed only during the articulation of **SEARCH**), whereas the frown (the upper layer) is active during the whole sentence.⁵

In other words, in (1) we have the following associations between layers of articulation and meanings (putting aside the object and verb):

- (4) Upper face → desperate
 Lower face → (frantically) → MANNER
 Arms and hands tension and pace of movement → MANNER
 Hands place of articulation → LOCATION

According to this, a more accurate translation of (1) would be:

- (5) Worrying about it, I frantically search for my wallet.

As it is easy to see, the possibility of having more than one argument-predicate relation simultaneously expressed complicates sign language argument structure in considerable ways. Given that the study of LSP is still in its infancy, in this chapter, I will limit myself to the most basic structures, namely, the relationships between verbs and arguments when both are expressed with manual articulators. Nevertheless, the issue of simultaneity will come back, since it is crucial to the understanding of classifiers, which are externalized by the hands.

⁵ This means we can understand the upper layer as a secondary predicate, something like (5). Or maybe it is a way to introduce a conventional implicature (in the sense of Potts (2005)), as suggested by Andrés Saab (p.c).

4 LSP Basic Argument Structure

As it has been proposed by many different authors (see Hale and Keyser (1993) and Pyllkänen (2002) for some key works and Williams (2015) for a comprehensive overview), a predicate can have between zero and three arguments, which gives us four types. At least since the Unaccusative Hypothesis (Perlmutter 1978), it is assumed that 1-argument predicates come in two different kinds: unaccusative (which realizes the internal theme-type argument) and unergative (which realizes the external agent-type argument). This raises the number of possible predicates to five. It is, thus, not surprising that sign languages, which are full languages, exhibit the full range of predicate-types as they do. What is interesting, though, is that at least a subset of predicate-types matches different material forms (i.e., handshapes). This raises questions about the conceptualization of the predicate when there is an externalization mechanism that can use space and form.

There is a correlation between the predicate-type and the type of classifier a verb can combine with (see Zwitserlood (2003), Sandler and Lillo-Martin (2006), Benedicto and Brentari (2004), Benedicto et al. (2007), Geraci and Quer (2014), and Kimmelman (2022), a.o.):

- (6) a. Transitive predicates combine with handling classifiers (HC).
- b. Unaccusative predicates combine with entity classifiers (EC).
- c. Unergative predicates combine with body part classifiers (BC).

Before unpacking these generalizations, it is necessary to explain how we treat classifiers. This is not as straightforward as we would like it to be, since it remains a heavily discussed issue in the literature on sign languages (see Schembri (2003) and Kimmelman (2022) for critical overviews). I will adopt some aspects of the view developed by Zwitserlood (2003) and Sandler and Lillo-Martin (2006), among others, according to which at least some classifiers participate in agreement relations and are incorporated into the predicate. Thus, I will adapt Madrid (2018) analysis of LSP classifiers, not only because he adopts a compatible view, but also because it is the only study on LSP classifiers to date. These works, however, conceive agreement in a way closely related to subject agreement in oral languages. Here I will adopt a wider view of agreement, but limiting it to just a sub-product of the operation *Agree* (Chomsky (2000) and much work after it), and I will say nothing about the traditional functional dependencies *Agree* seems to cause (being a subject, for instance).⁶

⁶ It is well known that traditional functional notions do not properly capture the complexities involved in the construction of the clause, and many researchers consider them an epiphenomenon (see McCloskey (1997) for a thorough discussion). This has not prevented them from using these expressions (*subject, object*, etc.) in an informal way, a practice that I follow here.

4.1 *Classifiers*

The concept of classifier is used to describe the function of certain morphological systems in oral languages which separate nouns in different classes. According to Aikhenvald (2019), there are seven classification systems in oral languages: gender, numerical, noun properties, possessive, verbal, locative, and deictic. A few examples from oral languages are enough to understand why the term *classifier* is used to name similar systems in sign languages:

(7) Nominal intrinsic property classifiers

Dâw tog *Dâw*
 CL:human girl
 “a girl” (lit. human girl)
 (Aikhenvald 2019, example (3))

(8) Location classifiers

Wis-uh tarak-e-gu a-**hakwa**-t un *Palikur*
 1PL-EXCL push-COMPL-3F 3neu-CL:into.liquid-DIR waterway
 “We push it (the canoe) into the water.”
 (Aikhenvald 2019, example (8))

(9) Deictic classifiers

re-**rak** *Mandan*
 this-CL:sitting
 “this (sitting)”
 (Aikhenvald 2019, example (9))

(10) Verbal classifiers

Nah watak-**buk**-e ini mawru *Palikur*
 1SG untie-CL:linear-completive this:NEUTER cotton
 “I untied the string completely.”
 (Aikhenvald 2019, example (7))

In (7), the classifier *dâw*, a homonym of the language name, is a free morpheme and distinguishes nouns based on their intrinsic property (being human, in this case). In (8), the classifier *-hakwa* is a bound morpheme and categorizes the location


toward which the object (not mentioned in the sentence, but discursively available) is moved—notice that in this case, what is classified is not the unmentioned object (the canoe) but the noun that expresses its location (i.e., the waterway). In (9), the classifier *-rak* is also a suffix and refers to the position of the referred object (discursively available). In (10), the suffix classifier *buk* refers to an object with a linear form (a string of cotton). Classifiers, thus, can be free or bound morphemes and relate to various semantic dimensions both intrinsic (like being human or having a given form) and extrinsic (like position or location). Also, they can refer both to nouns expressed in the sentence or to entities only discursively available.

It is not a surprise, then, that the term “classifier” was chosen to define signs that seem to display this pattern (see Supalla (1982, 1986) and Meir (2001), among many others). It must be said that there is no consensus on this issue among researchers, with positions falling all over the place, from categorically denying that there is any similarity with oral languages to simply ignoring comparisons and discussing classifiers on their own terms, including its conception as agreement markers, clitics, and incorporated nouns (see Emmorey (2003), Sandler and Lillo-Martin (2006: 76–93, 344–351), Zwitserlood (2012), and Madrid (2018: 14–45), among others, for overviews and discussions). Obviously, this is not the place to set these debates, but since my goal is to describe the basic argument structure of LSP, and classifiers (as we will see) are intimately connected to it, I must choose a theoretical way to address them. As already mentioned, I will consider the type of classifiers shown here as resulting from the operation *Agree* (Chomsky (2000) and subsequent work), partially siding with Zwitserlood (2003, 2008) and Madrid (2018), with certain differences to which we will come back.

Furthermore, this chapter is limited to just three types of LSP classifiers: handling, whole entity, and body parts. The reason for this is that these types have been linked to three different kinds of predicates in a fairly uniform way (Benedicto and Brentari (2004), Benedicto et al. (2007), and Geraci and Quer (2014), among others), and my goal is to check these correlations in LSP.

To close this section, I will provide an example for each type.

Handling classifiers (HC) represent entities held or moved by some agent (generally human). They roughly correspond to instrumental classifiers from Supalla’s (1986) original insight on the matter.⁷ (11) is an LSP example (Fig. 1):

- (11) Handling classifier (HC)
 HOLD-TREMBLING-CL: 
 ““He holds it (a bottle) trembling.””

(Adapted from Madrid 2018: 104)

There are a few points to stress with respect to (11), besides the handshake, which matches the form of the sentence object (a bottle), which is the primary function of

⁷ Madrid (2018), following Zwitserlood (2003, 2008, 2012), divides body part classifiers into two groups, assigning them either to handling classifiers or to whole entity classifiers. We are keeping body part classifiers together as a group.

Fig. 1 He holds it trembling as his heart pumps. (Taken from Madrid 2018: 104)



HC. First, note that the handshape not only expresses the form of the object, but also the action of holding. Second, there is a third meaningful simultaneous layer in the same hand, namely, the movement of the hand, which conveys the manner (trembling). Third, since this sentence is part of a longer narrative, the object (the bottle) has already been expressed, as a part of a previous sentence (in which the bottle is opened). Fourth, the non-dominant hand expresses a pumping heart, adding yet another meaningful layer, in fact, a different proposition: “The heart is pumping” (see the previous sections on layers of meaning).⁸ Here, I will focus on the hand expressing the HC.

Whole entity classifiers (EC) represent referents (individuals and objects) by expressing some salient property of an object or an individual (e.g., schematic form, intrinsic properties, and the like). They roughly correspond to Supalla’s (1986) semantic classifiers. In (12), we have two LSP examples simultaneously expressed by each hand (Fig. 2):



Again, there are some observations to make regarding sentences (12a) and (12b), besides the fact that they hold different ECs, one for an individual, the other for a plane object (a door, in this case). First, each one expresses a different proposition (i.e., different predicate-argument relations), simultaneously. Second, in both cases, the hand not only expresses the entities (person or door), but also what happened to them (being there or being opened). Third, these sentences are also part of a larger narrative, in which the signer tells the story of a girl who arrives at a door, tries to

⁸ An interesting issue is the handshape of the non-dominant hand. It has a pointing finger (but not a classifier). In this case, it is not pointing to the bottle; actually, it is not pointing to anything. I hypothesize that it expresses the stage-level nature of the predicate (*pumping*); if this is correct, the non-dominant hand is exteriorizing a different proposition, and the pointing finger would be a copula-type expression—see Rodríguez-Mondoñedo and Arnaiz (2022) for the suggestion that LSP may have, or it is developing, a copula from the pronominal form, in line with similar typological evolutions in several languages.

Fig. 2 (As she was about to leave) The door opened for her. (Taken from Madrid 2018: 91)



(12) Whole entity classifier (EC)

- a. *Dominant hand:* location_{ipsolateral}-CL: 
 “‘The person is in front of it (the door)’.”
- b. *Non-dominant hand:* location_{ipsolateral}-OPEN-CL: 
 “‘It (the door) opened’.”
- (Adapted from Madrid 2018: 91)

open it, and fails to do so, and after that she starts to leave and the door just opens by itself. Fourth, the girl is represented by the slight turn of the head, expressing that she is looking away from the door when it opens for her.⁹ I will focus on the ECs.

Body part classifiers (BPC) were proposed originally by Supalla (1982, 1986), and there have been some questions regarding its inclusion into a singular class—see Zwitserlood (2012) for discussion. Some body parts represent themselves: mouth or eyes, for instance. Others are represented by handshapes; for example, the legs are represented by selecting the index and the middle finger. Here, I will focus on the last type. (13) is an LSP example.

In this case, as we see in Fig. 3, there is also an EC that represents a sidewalk in the non-dominant hand. Additionally, in the dominant hand, the movement forward represents the direction of the walking, and the movement up and down of the palm-hand expresses the manner in which the walking happens. Thus, here we have three externalization layers expressing three different predicate-argument relations. Also,

⁹ Notice that I am not counting the head as a classifier in this case; rather, it is part of another complex system of role shifting, constructed action, and meaningful use of the signing space. We will not discuss this issue here (see Perniss (2012), Lillo-Martin (2012), and references therein).

Fig. 3 She walks forward, going up and down the sidewalk. (Taken from Madrid 2018: 89)



(13) Body part classifier (BPC)

WALK-FORWARD-CL: 

““She walks forward.””

(Adapted from Madrid 2018: 104)

we must notice that the BPC expresses not only the body part (the legs) but also the act of walking. As in the previous cases, (13) is part of a larger narrative, where a girl, introduced at the beginning of the story, is going up and down a sidewalk as she walks forward. Thus, in this instance, we have two different types of classifiers, EC and BPC, but we will focus on the BPC.

Summarizing, thus, in this section, we have seen that LSP classifiers form complex structures, indeed full classifier constructions with simultaneous layers of externalization. They can refer to discursively available entities and can express various semantic dimensions, intrinsic (being a person, or a flat object) or extrinsic (relative position, movement).¹⁰ There is an issue, though, that I have not yet properly addressed: are these classifiers bound or free morphemes? We will not entertain an answer to this question here but let me describe the problem we face. Let’s take the ECs in (12) as an illustration. First, the EC refers to the door, but, as mentioned, it does not only express the door, but also its closing—although the closing is expressed by the hand movement, not the handshape. This allows for an incorporation analysis (following Sandler and Lillo-Martin (2006: 83–84),




¹⁰ Thus, this is not so different from oral language classifiers, although see below.

among others), where the classifier and the movement become bound morphemes—obviously, given that a movement cannot happen without an entity. However, if we turn to the second EC, the one that refers to the person who arrived at the door, the situation is less clear. Putting aside the arrival (which is a different event not pictured in Fig. 2), in (12) the EC expresses the person, but also her position with respect to the door. It is not clear if that relational location, created not just by the presence of the EC, but also by the presence and position of the other EC (the door), is a lexical head to which the EC incorporates. An alternative could be, for instance, that the position is pragmatically inferred, and therefore the EC would be a free morpheme (something like a deictic classifier). I will not set this issue here, but I would like to point out that classifiers seem to simultaneously express both arguments and predicates. Here I side with the analyses that contend that this is a modality effect, and that, syntactically, we must separate arguments and predicates in different heads.

4.2 *Classifiers and Argument Structure*

We have seen in the previous section that a classifier is an iconization representing, among other things,¹¹ a participant in a predicate: it can mimic the way objects are handled, the form of the object, or a body part involved in the action. For illustration, I add here a few more LSP classifiers identified by Madrid (2018), some of which I have already examined (Table 1).

Table 1 Sampling of LSP classifiers

Type of classifier	Handshapes
a. Handling	
b. Entity	
c. Body parts	

Adapted from Madrid (2018)

¹¹ We will not discuss other types of classifiers—see Supalla (1986), Emmorey (2003), Zwitserlood (2003, 2008, 2012), Sandler and Lillo-Martin (2006), and Madrid (2018), i.a., for further types and careful discussion.

We have also presented in (6), repeated as (14) for convenience, correlations between these classifiers and different argument structures in sign languages:

- (14) Correlations between argument structure and certain classifiers
- a. Transitive predicates combine with handling classifiers (HC).
 - b. Unaccusative predicates combine with entity classifiers (EC).
 - c. Unergative predicates combine with body part classifiers (BC).

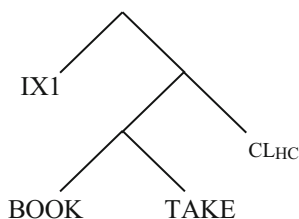
In what follows, I provide examples of these correlations and a proposal for a syntactic analysis for each of them. The reader should keep in mind that I will present only the bare bones of the syntactic structure. My goal is focused on the differences between each argument structure, showing how each of them gets externalized; hence, I will ignore null heads and labels.

HCs combine with transitive predicates:¹²

- (15) IX1 BOOK TAKE+CL_{HC}
 “I took a book.”

In this sentence, both subject (IX1) and object (BOOK) are expressed. Additionally, the verb TAKE takes the form of the HC classifier. The syntactic structure for (15) is straightforward, if we assume that the HC is an independent head that sits in a functional projection—see Gallego (2020) for an overview of the multiple middle periphery phenomena and the various functional projections in different oral languages. As mentioned, I also assume that the verb and functional heads are final, whereas specifiers are initial. Under these assumptions, it is enough to assume that the HC classifier probes the object BOOK, agrees with it, and then incorporates into it, resulting in what seems a regular object agreement relation. To see how this happens, let’s start with the base-generated structure in (16). Here, the verb TAKE has merged first with the internal argument BOOK, and then [BOOK TAKE] merges with the HC classifier; and the new constituent merges with the external argument (IX1):¹³

(16)



¹² Notice that in (15) we are using IX to represent a pointing sign; in this case, IX1 will be the signer pointing to herself to say “I.”

¹³ Remember we are ignoring null heads, for instance, the T head. There is no evidence that LSP has a morphological expression of Tense, which of course does not preclude a null T. To avoid any commitment, I have opted for labelless trees.

Remember that the externalization of (16) will end up with TAKE and the CL_{HC} forming a single sign, as (15) shows. The question, then, is how we achieve this. There are various ways to produce this result. For instance, TAKE can undergo incorporation by head movement, moving to the next head CL_{HC}, forming a complex morpheme. But recent developments in the interface between syntax and morphology have brought upon other possibilities. We could say that we have a span (in the sense of Svenonius (2016, 2020)) with a treelet composed of those two heads, meaning the lexicon can store parts of the tree (the spans) and insert the corresponding form at once. No matter which analysis we adopt here, there are a couple of crucial aspects that need to be refined.

First is the feature matching between BOOK and the CL_{HC}. BOOK must have the relevant features to value the unvalued features of the CL_{HC}. Adapting Madrid (2018) featural system for LSP (in turn based on Zwitserlood (2003) for sign language of the Netherlands), I propose that BOOK has the following valued features:

- (17) BOOK
 [form: straight]
 [form: flat]

On the other hand, the HC has its features [form] unvalued:

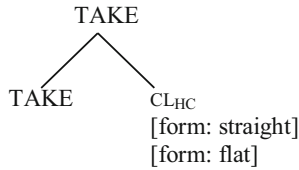
- (18) CL_{HC}
 [form: ____]
 [form: ____]

Notice that (18) is a bundling of formal features; assuming a post-syntactic insertion, the specific exponents have to be filled after the corresponding domain had been shipped to the interfaces, as predicted by (3). But before that, the operation *Agree* must value the features in (18), under match with BOOK:

- (19) CL_{HC} BOOK
 [form: ____] → [form: straight]
 [form: ____] → [form: flat]



After being valued, the CL_{HC} incorporates into TAKE, forming a complex head:

(20)

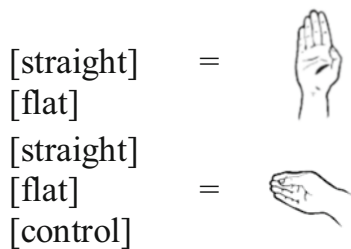


In order to provide the proper phonological form for (20), we must explain something regarding the features of TAKE. We have seen in the previous section that the phonological expression of this verb is just the movement, which in this instance goes from the object toward the subject. As already mentioned, every movement implies something that moves; therefore, the classifier (the handshape) must incorporate into the predicate (the movement) to support it morphologically.

The second aspect in (20) is the assignment of thematic roles. Under standard assumptions, BOOK (the internal argument) receives the Theme role from the verb TAKE, and IX1 (the external argument) receives the Agent role from the classifier. This means that the CL_{HC} may be understood as an agentive small *v*, which introduces the external argument and assigns it a theta role. Benedicto and Brentari (2004) make a comparable proposal, identifying HCs as agentive, although they employ a finer grained structure. Now, there is nothing in (20) that expresses the fact that this is an agentive classifier, even after valuation in (19). To capture that, we will need an additional feature. We can use Zwitserlood’s (2003) and Madrid’s (2018) feature [control].

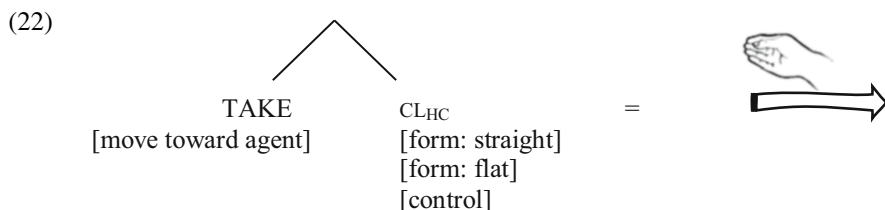
Notice that this is not really a semantic feature, and it must not be confused with the Agent role. [control] makes a crucial contribution to the phonological expression of the classifier. If we do not include it, the form of the classifier would be  and not . The reason for this is that the form features of the classifier have taken the values [straight] and [flat] from BOOK, resulting in a handshape with the fingers extended and closed. Adding the [control] feature will trigger the grabbing form, bending the fingers:

(21) a.



b.

Thus, after incorporation to small v , the phonological form of the complex head produced by TAKE and CL_{HC} will be externalized as the hand pulling a straight and flat object:¹⁴



In other words, the feature composition of (22), a result of incorporating the CL into TAKE, is externalized as the movement of the hand from the locus assigned to BOOK to the locus assigned to the agent (hence, the curvy arrow in (22) moving toward the taker).

We have kept a minimal structure, representing only the operation Merge, but nothing precludes us from implementing a more complex configuration. For instance, we could express (22) in a Ramchand's (2008) style, with multiple heads expressing different aspects of the event, or we could deploy a full Lexical Conceptual Structure (Jackendoff 1990). Since my purpose here is just to represent LSP basic argument structure, I will not explore this possibility.

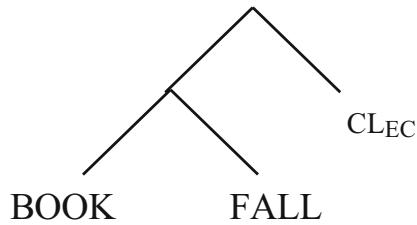
Let's turn now to ECs. As mentioned, they combine with unaccusative predicates, that is, with predicates that only have internal argument:


- (23) BOOK FALL+CL_{EC}
 "The book fell."

In (23), there is no external argument and, consequently, no Agent. The internal argument is expressed in an independent phrase (BOOK), with the classifier heading a functional projection:

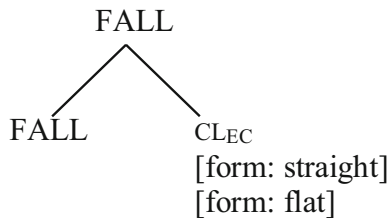
¹⁴ Notice that the picture in (22) is a 2D rendition of a 3D handshape and movement. Furthermore, the handshape can be expressed in terms of its phonological features (selected fingers, closure, etc.; see Brentari (2019) for a recent overview of sign language phonology). For LSP, see Raico (in preparation). With respect to the classifier, here we are only dealing with the valuation of its features, not its interpretation status; see Pesetsky and Torrego (2007) for the claim that interpretability and valuation are different aspects of the features.

(24)



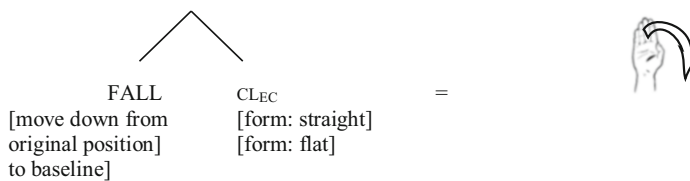
Notice that this has the same internal argument as (15), a transitive sentence; hence, it has exactly the same features. Yet, the classifier is different: now we have . Given our analysis in (22), it should be apparent why this is the case: the classifier lacks the feature [control]. Other than that, the *Agree* operation takes place exactly like in (19), and incorporation of the EC into FALL renders a complex head:

(25)



Like in (16), we need to separate the movement (which is the phonological content of the verb) from the handshape (the phonological content of the classifier). In (23), FALL corresponds to a movement down from the original position to the baseline. Then, after incorporation, the externalization will be as follows:¹⁵

(26)



Note that this predicts a wide range of variation with respect to both the handshape and the movement, as it is indeed the case. The former depends on the shape of the object, and the latter will change according to the baseline. A book on a bookshelf moves from a vertical to a horizontal position, if the baseline is one of the

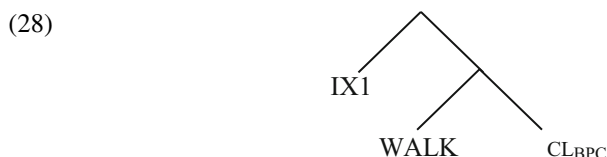
¹⁵ The caveat from fn. 14 also applies here.

shelves, but further down if the baseline is the ground. If the object is, for instance, an apple that falls from a tree, the handshape will be round, and the baseline will be the ground (which can be either included in the sentence or discursively available).

Finally, we arrive to BPCs, which combine with unergative predicates, i.e., with predicates that require only external arguments:

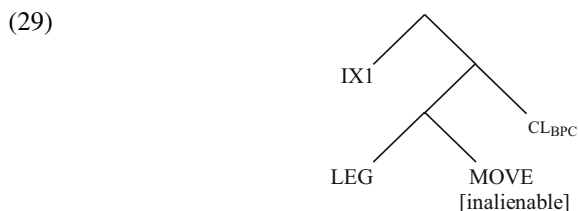
- (27) IX1 WALK+CL_{BPC}
 “I walk.”


In (27), there is no internal argument and, therefore, no Theme. The external argument is a pronominal form (IX1), and the classifier is in a functional projection. We could entertain a base-generated structure like (28):



Following an already familiar process, now the BPC should be incorporated into the verb. Notice that in this case *Agree* cannot happen, since the probe (CL_{BPC}) does not c-command any goal. A question arises, then, with respect to how the classifier will acquire its shape. We cannot resort to a default shape, since it does vary with respect to the type of body part involved. For instance, if the walker were a chicken, the classifier will turn into a chicken finger shape (three selected fingers in both hands), which will move as walking. The only conclusion is that (28) cannot be the structure for the unergative classifier construction.

In fact, it has been suggested that unergative verbs project more than meets the eye. Hale and Kayser (1993) proposed that unergative verbs have transitive structures, with an object incorporated into the verb. It is with this object that the classifier will agree, under a process akin to the one devised for (15). For (27), this would mean a structure like (29), where the verb MOVE selects an inalienable possessed noun as its complement:



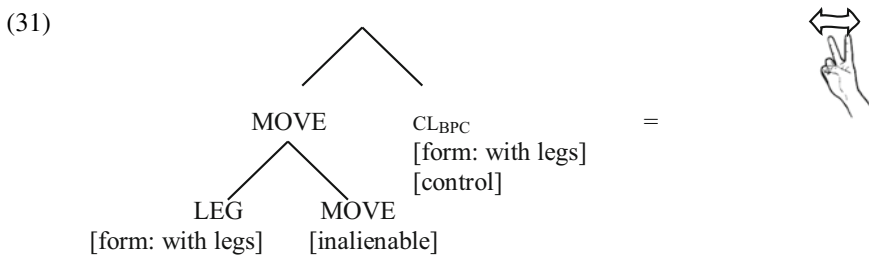
Given that Agent is assigned in (29), the classifier now must have the unary feature [control], but it has its [form] features unvalued. Now the classifier can value its [form: ___] feature under *Agree*, in a procedure similar to (19), obtaining the shape . For BPC, I codify the [form] feature in a way similar to Zwitserlood (2003) and Madrid (2018), with the value [with legs]:

(30)

$$\begin{array}{ccc}
 \text{CL}_{\text{BPC}} & & \text{LEG} \\
 [\text{form: } ___] & \longrightarrow & [\text{form: with legs}] \\
 [\text{control}] & &
 \end{array}$$

We should remember that the features [form: with legs] and [control] are interpretable formal features, which will need to be interpreted phonologically after arriving at the corresponding interface (PF in (3)), and semantically at the other interface (LF in (3)). Furthermore, we must notice that LEG has just the necessary features to be merged with the verb and be interpreted by the interfaces.¹⁶

After valuation, the classifier is enabled to be interpreted in the phonological component, but before that, it must incorporate into the complex head already formed by the incorporation of LEG into the verb. The result is (31):



This means that a literal translation of (27) is not “I walk,” but “I leg”; in other words, the denominal English verb “to walk” corresponds to the denominal LSP verb “to leg.” This is a very common occurrence in denominal verbs across languages. For instance, in Jaqaru (Aymaran family, spoken in Peru), the word for “to drink” and for “water” is the same, namely, *uma*, which means that in Jaqaru the literal equivalent of “to drink” is “to water” (Rodríguez-Mondoñedo 1999: 25). In fact, exactly the same happens in English, where “walk” can be a noun or a verb. It is tempting to interpret “to walk” as “to make a walk,” but paraphrases like these

¹⁶ If we were to implement (30) in terms of Distributed Morphology, we would say that LEG just has an index, and that Agree makes possible the sharing of its index with the classifier. Later, that index will be interpreted as “with legs.”

have different meanings (as early observed by Fodor (1970)); so, English “to walk” also results from the incorporation of a noun “walk” into its predicate. Likewise, in LSP the verb for “to walk” is literally “to leg.”

So far, we have seen that there is a correspondence between the shape of the classifier and the argument structure of some verbs. This is in line with previously observed patterns in other sign languages (Benedicto and Brentari 2004; Benedicto et al. 2007; Geraci and Quer 2014, among others).

As mentioned, I have analyzed the very bare bones of argument structure, without going further into the structure of the event or the whole clause, which are different endeavors. In fact, the main insight from Talmy’s (1985) groundbreaking work is that the verb cannot bear by itself the responsibility for event structure. For instance, Manner and Path are also categories that need to be included (see Supalla (1982) and Talmy (2007) for some applications to sign languages), but I assume, following Hale and Kayser (2005), that argument structure is a separate component of grammar, and that we must start from argument structure and then try to deduce event structure, and not the other way around. I hope I have contributed to this effort from the point of view of LSP.¹⁷

5 Some Loose Ends

In this final section, I would like to address some issues not considered in the previous sections that might be sources of misunderstandings.

First, I would like to highlight the relation between form and meaning in LSP classifiers and its comparison with similar oral systems. Clearly form and meaning are simultaneous, and the suggested task is disentangling them analytically. We have found in the Unaccusative Hypothesis and some of its structural interpretations a way to do so. We must keep in mind, though, that these tools were proposed having oral language data in mind; it is indeed remarkable that this theoretical apparatus can say something about LSP, a sign language. Clearly, these are welcome news by someone (like me) who aims to find the ultimate properties of the Faculty of Language, which can, of course, be manifested in sign languages. Nevertheless, we must be wary of any Procrustes’ inclination. In fact, I believe that, after examining sign languages, we can look back at oral languages and acknowledge some features we didn’t notice were there—as it is already happening in the domain of gestures (see Abner et al. (2015) for an overview) or the anaphora system (see Schlenker (2017)). In that sense, departing from contrarian suggestions, it is not only valid but indeed necessary to continue the comparison, trying hard to find connections between sign and oral languages. So, yes, simultaneity does provide new challenges,

¹⁷ Kimmelman (2022) shows cases where the relation between event structure and classifier type in at least some sign languages is not so straightforward. More research is needed, of course, but the point here is that argument structure does not necessarily predict event structure.

but it should not be a reason for discontinuing this enterprise—see Sandler and Lillo-Martin (2006) for a thoroughly detailed exposition of the comparative project, and for LSP. See Pérez Silva (2021) and Rodríguez-Mondoñedo (2021).

In addition, we must be aware of the real scope of the analytic tools. For instance, the operation *Agree*, which is a key feature of the current analysis, should not be confused with the normal use of agreement (as in for instance “subject-verb agreement”). Although the latter could be understood as a sub-case of the first, in no way *Agree* is limited to subject agreement, not even to A-dependencies, but it extends to A-bar, including operator-variable relations, and more. In the present work, I have used it to carve the shape of LSP argument structure, identifying its analytic primitives. No claim can be made from this analysis about the possibility (or not) of sign languages to have agreement. No tree in the present work includes any functional head related to traditional agreement (no T, for instance). In that sense, we do not need to explain why in sign languages there seems to be a preference for object agreement, contrary to the typological trend that gives prominence to subject agreement. We have not shown object (or subject) agreement, we have only shown how *Agree* can explain the regularities in the shape/externalization of argument structure.

Of course, the proposal is fully compatible with the claim that there is agreement in sign languages, including LSP. Lourenço and Wilbur (2018, and this volume) for LIBRAS and Ramos (2022) for LSP provide, I believe, important arguments in favor of the idea that this is indeed the case (see also Pfau et al. (2018) and references therein). If agreement must code dependencies using a physical channel (to be heard or to be seen), then it should be no surprise that, if the channel is sound, agreement is expressed as co-sounding; but if the channel is visual space, then it is expressed as co-locating.

Also, in LSP, as well as in other sign languages, not all verbs can incorporate classifiers, regardless of the nature of the predicate. In these cases, arguments are represented only by pronominal forms (when not lexically expressed):

- (32) MARY_a LIKE IX_b
 “Mary likes him.”

Observing that these verbs are often anchored on the body (in LSP, LIKE is articulated on the front part of the neck), and following Meir et al. (2007), in these cases, the body is used to introduce the external argument, given that the body has the highest prominence in a visual system. Furthermore, LIKE shows no iconicity (in the handshape, index and thumb fingers close on each other, pointing to the neck), that is, the components of its predicate structure have not been codified in different morphemes, but just as one arbitrary handshape. It must, thus, resort to the default introducer of external arguments, the body.

6 Conclusion

I have examined how LSP builds different predicate structures and concluded that, at least for a subset of them, we can identify sub-components of the predicate, using the Unaccusative Hypothesis together with the operation *Agree*. This explains the correlations between type of classifiers and type of argument structures, given that the handshapes of the classifiers are built from the features of the corresponding arguments. I believe these results strengthen the force of the comparative Oral-Sign Language project, which aims to better understand the nature of the Language Faculty, a task that cannot be completed without taking in consideration languages from different modalities.

Let me end this chapter by reminding the reader that the fight for full recognition of sign languages as natural languages is not over, not even after legal recognition, as witnessed by the situation of LSP described in the second section. A no-small contribution to this fight is unraveling the inner workings of sign language grammars.

References

- Abner, Natasha, Kensy Cooperrider, and Susan Goldin-Meadow. 2015. Gesture for linguists: A handy primer. *Language and Linguistics Compass* 9 (11): 437–451.
- Aikhenvald, Alexandra Y. 2019. Noun categorization devices a cross-linguistic perspective. In *Genders and classifiers: A cross-linguistic typology*, ed. Alexandra Y. Aikhenvald and Elena I. Mihás, 1–29. Oxford: OUP.
- Arnaiz, Alexandra. 2021. *La lengua de señas peruana como vehículo de aprendizaje del castellano escrito como segunda lengua*. MA thesis. Pontificia Universidad Católica del Perú. Lima: Perú.
- Benedicto, Elena, and D. Brentari. 2004. Where did all the arguments go? Argument-changing properties of classifiers in ASL. *Natural Language and Linguistic Theory* 22: 743–810.
- Benedicto, Elena, S. Cvejanov, and J. Quer. 2007. Valency in classifier predicates: A syntactic Analysis. *Lingua* 117: 1202–1215.
- Brentari, Diane. 2019. *Sign language phonology*. Cambridge University Press.
- Catalán, Yedirel. 2021. *Características morfológicas y morfosintácticas en el español escrito por escolares signantes de la lengua de señas peruana (LSP)*. Tesis de Licenciatura. Lima: UNMSA.
- Cerna-Herrera, Francisco, y César Ramos. 2022. Cláusulas relativas en lengua de señas peruana (LSP). *Quintú Quimín. Revista De lingüística* 6 (Q063): 1–16.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step*, ed. R. Martin et al., 89–155. Cambridge, MA: MIT Press.
- Clark, Brenda. 2017a. Sign language varieties in Lima, Peru. *Sign Language Studies* 17 (2): 222–264.
- . 2017b. *A grammatical sketch of Sivia sign language*. Ph.D. Dissertation, University of Hawai'i at Mānoa.
- Cuti, Zanabria Elizabeth. 2018. *Sistema Antroponímico en la Lengua de Señas Peruana*. Tesis de Licenciatura. Lima: UNMSM.
- Emmorey, Karen, ed. 2003. *Perspectives on classifier constructions in sign languages*. Psychology Press.

- Fodor, Jerry A. 1970. Three reasons for not deriving “kill” from “cause to die”. *Linguistic Inquiry* 1 (4): 429–438.
- Gallego, Ángel. 2020. Morpho-syntactic Variation in Romance v: A Micro-parametric Approach. *Probus* 32 (2): 401–437.
- Geraci, Carlo, and Josep Quer. 2014. Determining argument structure in sign languages. In *Structuring the Argument: Multidisciplinary research on verb argument structure*, ed. Asaf Bachrach, Isabelle Roy, and Linnaea Stockall, 45–60. John Benjamins.
- Goico, Sara. 2019. The impact of ‘inclusive’ education on the language of deaf youth in Iquitos, Peru. *Sign Language Studies* 19 (3): 348–372.
- Hale, Kenneth, and Samuel J. Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In *The view from building 20: Essays in linguistics in Honor of Sylvain Bromberger*, ed. K. Hale and J. Keyser, 53–109. Cambridge: MIT Press.
- Hale, Kenneth, and Samuel Jay Keyser. 2005. Aspect and the syntax of argument structure. The syntax of aspect: Deriving thematic and aspectual interpretation. In *The syntax of aspect: Deriving thematic and aspectual interpretation*, ed. Nomi Erteschik-Shir, Tova R. Rapoport, and Tova Rapoport, 11–41. Oxford: Oxford University Press.
- INEI. 2012. *Primera Encuesta Nacional Especializada sobre Discapacidad ENEDIS 2012*. <https://conadisperu.gob.pe/observatorio/estadisticas/encuesta-nacional-especializada-sobre-discapacidad/>. Accessed 7 Apr 2022.
- . 2017. *CPV 2017 Censo de Población y Vivienda 2017*. Lima: Perú. <http://censo2017.inei.gob.pe/>. Accessed 7 Apr 2022.
- Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.
- Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kimmelman, Vadim. 2022. Argument structure in sign languages. *Annual Review of Linguistics* 8: 19–38.
- Kremers, Joost. 2012. The syntax of simultaneity. *Lingua* 122 (9): 979–1003.
- Lillo-Martin, Diane. 2012. Utterance reports and constructed action. In *Sign language. An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 365–387. De Gruyter Mouton.
- Lourenço, G., and R.B. Wilbur. 2018. Are plain verbs really plain? Co-localization as the agreement marker in sign languages. *FEAST* 2: 68–81.
- Madrid, Vega Rodrigo. 2018. *Clasificadores en la Lengua de Señas Peruana (LSP)*. Tesis de Licenciatura. Lima: PUCP.
- Malca, Marco, and Frank Domínguez Chenguayen. 2022. Conceptual metonymy in the creation of concrete nominal signs in Peruvian sign language: Towards a metonymic typology. In *Concepts, discourses, and translations*. Second language learning and teaching, ed. B. Lewandowska-Tomaszczyk and M. Trojszczak, 81–112. Cham: Springer.
- McCloskey, James. 1997. Subjecthood and subject positions. In *Elements of grammar: Handbook in generative syntax*, ed. Liliane Haegeman, 197–235. Dordrecht: Kluwer.
- Meir, Irit. 2001. Verb classifiers as noun incorporation in Israeli sign language. In *Yearbook of morphology 1999*, ed. G. Booij and J. Van Marle, 299–319. Dordrecht: Kluwer.
- Meir, Irit, Carol Padden, Mark Aronoff, and Wendy Sandler. 2007. Body as subject. *Journal of Linguistics*. 43: 531–563.
- MINEDU. 1987. *Lenguaje de Señas Peruanas*. Lima: Ministerio de Educación.
- Napoli, D.J., and R. Sutton-Spence. 2010. Limitations on simultaneity in sign language. *Language* 86: 647–662.
- Olivero, Hugo. 2016. Los sordos peruanos y su lengua ignorada. *Lucidez*. <https://lucidez.pe/los-sordos-peruanos-y-su-lengua-ignorada-por-hugo-olivero/>. Accessed 7 Apr 2022.
- Paliza, Farfan Alberto. 1994. The problem of the Peruvian deaf person. In *The Deaf Way: Perspectives from the International Conference on Deaf Culture*, ed. C.J. Erting, R.C. Johnson, D.L. Smith, and B.D. Snider, 804–810. Washington, DC: Gallaudet University Press.
- Parks, Elizabeth, and Jason Parks. 2009. *Sociolinguistic survey report of the deaf community of Peru*. Summer Institute of Linguistics Electronic Survey Report 019.

- . 2010. A sociolinguistic profile of the Peruvian deaf community. *Sign Language Studies* 10 (4): 409–441.
- Pérez Silva, Jorge. 2021. Forma y sustancia en las señas del lenguaje: un análisis del plano de la expresión de la lengua de señas peruana. In *Actualidad y futuro del pensamiento de Eugenio Coseriu. Estudios de teoría del lenguaje, descripción lingüística, dimensión textual y lingüística peruana*, ed. Carlos Garatea, Jorge Wiese, and Marta Fernández Alcaide, 288–304. Madrid: Universidad de Sevilla.
- Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 4th annual meeting of the Berkeley linguistics society*, 157–189. UC Berkeley.
- Perniss, Pamela. 2012. Use of sign space. In *Sign language. An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 412–431. De Gruyter Mouton.
- Pesetsky, D., and E. Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture*, ed. S. Karimi, V. Samiiian, and W. Wilkins, 262–294. Amsterdam: John Benjamins.
- Pfau, Roland, Martin Salzmann, and Markus Steinbach. 2018. The syntax of sign language agreement: Common ingredients, but unusual recipe. *Glossa: A Journal of General Linguistics* 3 (1): 1–46.
- Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.
- Pyllkkänen, Liina. 2002. *Introducing arguments*. Doctoral dissertation. Massachusetts Institute of Technology.
- Ramchand, G.C. 2008. *Verb meaning and the lexicon. A first phase syntax*. Cambridge: CUP.
- Ramos, César. 2022. *Los verbos de concordancia doble de la lengua de señas peruana (LSP): un análisis del uso del espacio*. Tesis de Licenciatura. Pontificia Universidad Católica del Perú. Lima: Perú.
- Rodríguez-Mondoñedo, Miguel. 1999. La formación de palabras en aimara tupino. *Boletín de la Academia Peruana de la Lengua* 31: 131–160.
- . 2020. Privación lingüística: el caso de la lengua de señas. *AIEDI* <https://www.aiedi.org/2020/01/17/privacion-linguistica-el-caso-de-la-lengua-de-senas/>. Accessed 7 Apr 2022.
- . 2021. Modificación y desarrollo en lenguas orales y de señas. In *Actualidad y futuro del pensamiento de Eugenio Coseriu. Estudios de teoría del lenguaje, descripción lingüística, dimensión textual y lingüística peruana*, ed. Carlos Garatea, Jorge Wiese, and Marta Fernández Alcaide, 305–323. Madrid: Universidad de Sevilla.
- Rodríguez-Mondoñedo, Miguel, and Alexandra Arnaiz. 2022. A copula in Peruvian Sign Language. *En semantics of under-represented languages of the Americas (SULA 11)*, 161–170. GLSA, Amherst: University of Massachusetts.
- Rodríguez-Mondoñedo, Miguel, Sonia Maruenda, and Alexandra Arnaiz (Compiladores). 2015. *Archivo Digital de la Lengua de Señas Peruana*. <https://repositorio.pucp.edu.pe/index/handle/123456789/124183>. Accessed 7 Apr 2022.
- Sandler, Wendy. 2017. The challenge of sign language phonology. *Annual Review of Linguistics* 3: 15.1–15.21.
- Sandler, Wendy, and Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge: CUP.
- Schembri, Adam. 2003. Rethinking “classifiers” in signed languages. In *Perspectives on classifier constructions in sign languages*, ed. Emmorey Karen, 3–34. Mahwah: Lawrence Erlbaum Associates.
- Schlenker, Philippe. 2017. Sign language and the foundations of anaphora. *Annual Review of Linguistics* 3: 149–177.
- Supalla, Ted. 1982. *Structure and acquisition of verbs of motion and location in American sign language*, PhD thesis, UCSD.
- . 1986. The classifier system in American sign language. In *Noun classes and categorization*, ed. C. Craig, 181–214. Amsterdam: John Benjamins.
- Svenonius, Peter. 2016. Spans and words. In *Morphological metatheory*, ed. Daniel Siddiqi and Heidi Harley, 201–224. John Benjamins.
- . 2020. A span is a thing: A span-based theory of words. *NELS* 50.

- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In *Language typology and syntactic description III: Grammatical categories and the lexicon*, ed. Timothy Shopen, 57–149. Cambridge: CUP.
- . 2007. Lexical typologies. In *Language typology and syntactic description*, ed. Timothy Shopen, vol. III, 66–168. Cambridge: CUP.
- Vermeerbergen, L. Leeson, and O. Crasborn, eds. 2007. *Simultaneity in sign language: Form and function*. Amsterdam: John Benjamins.
- Williams, Alexander. 2015. *Arguments in syntax and semantics*. Cambridge: CUP.
- Wilbur, Ronnie. 2003. Modality and the structure of language: Sign languages versus signed systems. In *Oxford handbook of deaf studies, language, and education*, ed. Marc Marschark and Patricia E. Spencer, 332–346. New York: Oxford University Press.
- Zwitserslood, Ingeborg. 2003. *Classifying hand configurations in sign language of the Netherlands*. Doctoral dissertation, Utrecht University.
- . 2008. Morphology below the level of the sign. Frozen forms and classifier predicates. In *Signs of the time: Selected papers from TISLR 2004*, ed. Josep Quer, 251–272. Hamburg: Signum.
- . 2012. Classifiers. In *Sign language. An international handbook*, ed. Roland Pfau, Markus Steinbach, and Bencie Woll, 158–185. De Gruyter Mouton.

Blending Libras and Portuguese: Acceptability Variables



Ronice Müller de Quadros, Diane Lillo-Martin, and Marilyn Mafra Klamt

1 Bimodal Bilingualism

The bimodal bilinguals who are the focus of this study are adults who grew up in deaf families, with either one or both parents deaf signers. Our broader study includes Codas from both the USA and Brazil, but in this chapter, we will present the analysis of one specific study based on the Brazilian data alone. The broader study examined the linguistic characteristics of participants by analyzing languages in three modes: speech, sign, and the combination of sign and speech in which both languages are blended simultaneously. The results presented here are based on a study focused on the acceptability of code-blends among Codas, supplemented by elicited production data.

This is a special case of bilingualism, known as bimodal bilingualism, and it displays many of the same properties of unimodal bilingualism, but also some unique forms (Emmorey et al. 2008). We focus on different possibilities of combining the two languages, which are not possible when the languages are of the same modality. Thus, uniquely, bimodal bilinguals can produce both languages at the same time, because the languages used primarily employ different modalities for production and perception. Following Emmorey et al. (2008), we call such productions code-blends.

Code-blending frequently involves spelling out a single meaning in the two languages, although often one modality will contain more grammatical or content

R. Müller de Quadros (✉) · M. M. Klamt
Universidade Federal de Santa Catarina, Florianópolis, Brazil
e-mail: ronice.quadros@ufsc.br; marilyn.mafra@ufsc.br

D. Lillo-Martin
University of Connecticut, Storrs, CT, USA
e-mail: diane.lillo-martin@uconn.edu

information than the other. Examples of “full” and “partial” code-blending are given in (1)–(2).¹

(1) Full blending

LIBRAS	IX(eu) I	COMPRAR bought	VESTIDO dress	DV(estampa) print	FLOR flower
BP	Eu	comprei	vestido	estampa	flor
'I bought a flower print dress.'					

<https://youtu.be/L1GpoO5qkCY>

(2) Partial blending

LIBRAS	IX(eu) I	VISITAR visited	FAMÍLIA family	DEM there	TER... have
LIBRAS	QUATRO four	TV+ TV			
BP	quatro	tevé			
'I visited a family who had four TVs.'					

<https://youtu.be/SXvxEWPfU2Q>

Our main goal in this chapter is to understand better what the derivational mechanisms underlying code-blends are. To do so, we conducted an experimental study consisting of a judgment task and an elicitation task. As we will discuss, a factor manipulated in the experiments was the type of syntactic structure, in which we included transitive, negative, and passive sentences as well as idiomatic expressions. By considering these specific structures, we are able to analyze how syntactic differences between Brazilian Portuguese and Libras are accommodated

¹ The examples use the following notation: On the LIBRAS tier, capitalized words are glosses for signs in Libras and lower-case is used to provide English translations for the glosses. On the BP tier, lower-case is used for words spoken in Portuguese. The Portuguese glosses for the signs employ the identification conventions for each sign established in Libras Signbank (<https://signbank.libras.ufsc.br/>) when they were available; for the signs that were not already identified in the Signbank, we coded the sign by a new specific gloss to be added to the Signbank later. Pointing is annotated with IX (for “index”). In the formatting of examples, the signs and spoken words are aligned following the timing used in the blending as produced by the participant.

under blending. These differences can result in either congruent or incongruent code-blends. In incongruent code-blends, each language is produced following its own derivational possibilities in such a way that consequently, the sentences produced through signs and speech are produced with contrasting word orders. Congruent blends, on the other hand, display a variety of types of synthesis of the two languages without differing in word order. For example, there may be a one-to-one correspondence between content words in sign and speech; or there may be more information produced in one or the other language; or one language might use a language-specific form, while the other produces a near translation (an example is found when classifiers are produced in Libras alongside a phrase in Portuguese; see Quadros et al. 2020a, b).

In order to explain code-blendings, we considered two theoretical proposals: Branchini and Donati's (2016) analysis, based on blendings in Italian Sign Language (LIS)/spoken Italian, and the so-called Synthesis Model (Lillo-Martin et al. 2010, 2016; Koulidobrova 2012, 2016). These proposals are considered and discussed in view of the results obtained in our study.

The present study grows out of a long-term investigation conducted by us on the Development of Bimodal Bilingualism (funded by NIH; see <https://slla.lab.uconn.edu/bibibi/>). In the main project, we collected and analyzed longitudinal spontaneous production data from bimodal bilingual children in the USA and Brazil, with ages ranging from 18 months to 6 years (Chen Pichler et al. 2016; Lillo-Martin et al. 2010, 2014, 2016; Quadros et al. 2012, 2014; Quadros 2017). We also conducted language studies with older children, ages 4–7, to assess their language use at the phonological, morphological, lexical, syntactic, and discourse levels (Quadros et al. 2015). Our sample included *Kodas* (Coda kids) and Deaf native signers who received one or two cochlear implants and were developing bilingually in sign and speech (Davidson et al. 2014; Goodwin and Lillo-Martin 2019).

The theoretical proposal of the Synthesis Model was adopted as an account for the unique forms observed when the bimodal bilinguals show evidence of cross-linguistic influence and code-blending (Lillo-Martin et al. 2010, 2016; Koulidobrova 2012, 2016). Following MacSwan's proposal (2000, 2005), we considered that the bilingual language architecture is essentially the same as that for monolinguals, except for the existence of two lexicons. However, we added the basic concepts of *Distributed Morphology* (Halle and Marantz 1993), in particular, the idea that the input to a derivation is an abstract element not specified for phonological form, with Vocabulary Insertion happening late in the derivation. The DM approach provides not only a useful broad conception compatible with code-blending (late insertion), but also an explicit way for forming predictions, some of which will be discussed in the current chapter. We dubbed this approach *Language Synthesis*, intending to convey the idea that the computational system of grammar can synthesize pieces from multiple languages while running a single derivation. We suggested this proposal predicts the existence of code-blending whenever it is not prevented by the multiple articulatory interfaces, as is the case for bimodal bilinguals. Figure 1 provides a representation of the architecture of the Synthesis Model.

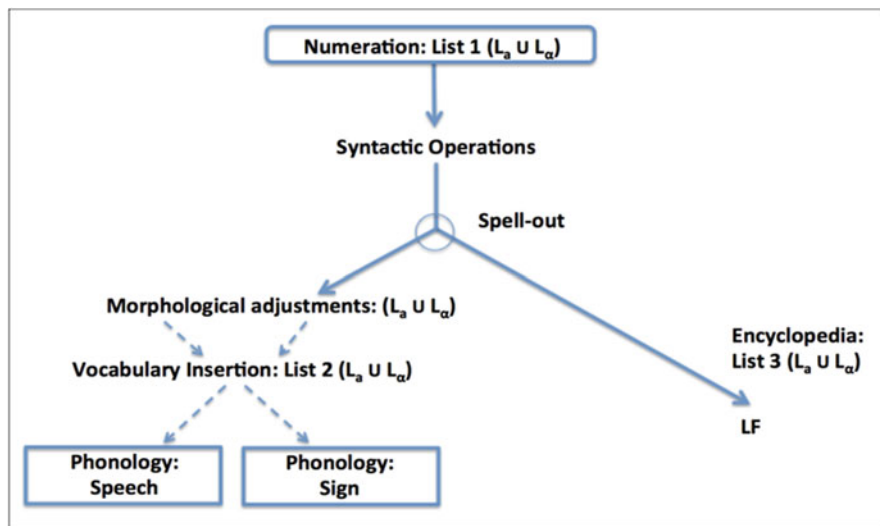


Fig. 1 Language synthesis model. (Lillo-Martin et al. 2016: 730; reproduced with permission)

Note that, under this approach, only one derivation feeds spell-out, starting from a single numeration that may contain abstract elements (features) associated with both languages. It is only after spell-out that elements from two languages are simultaneously introduced. Of course, this last step is not part of standard DM assumptions, since the theory was proposed in the absence of knowledge about code-blending.

The study we present here is designed to assess constraints on code-blending, aiming at understanding the derivational procedure that gives rise to this phenomenon. To do so, we started by investigating which types of blends are allowed and which ones are ruled out. As already pointed out (Emmorey et al. 2005, 2008; van den Bogaerde and Baker, 2005, 2008; Lillo-Martin et al. 2010, 2016), the output of a synthesis has to be a single proposition. If so, code-blends should not be able to express different concepts coming from the two different modalities. This prediction is tested most prominently in our acceptability judgment study in connection with idiomatic expressions.

In general, co-insertion of near translation equivalents (Emmorey et al. 2008; Lillo-Martin et al. 2010, 2014, 2016) should pose no problem. As illustrated in (1)–(2) above, whenever productions in the two languages follow the same surface word order, they are labeled congruent, although, generally, one of the blended languages provides the functional categories that determine the course of the derivation. In the most common cases, the two languages use parallel structures, so that the blending allows both languages to use structures generally produced in monolingual settings. Such examples are common. In other cases, word order comes from one of the languages (the primary language of the sentence) to the possible exclusion of word-order preferences observed in the secondary language. When this happens, a

derivation is generated, but we predict it is not fully acceptable due a preference for congruence at the word-order level.

As discussed earlier, incongruent code-blending occurs when the two languages use overtly different word orders (e.g., OV and VO). According to the Synthesis Model, incongruent word orders can be generated if linearization is a late operation, so that producing one order in speech and a different order in sign is possible. However, in these cases difficulties in processing might be observed, since the opposing word orders can lead to an increased memory load. Our prediction is that different linearization of strings under a single node may be acceptable, at least for short strings which impose lower processing costs. Across signers, however, such cases might not be rated as highly as congruent ones. (3) exemplifies a case of different linearizations, as in this example Libras and PB contrast with respect to the linear order between the negation and the verb.

(3) Blending with different word orders

LIBRAS	HOMEM man	ANDAR walk	NÃO not	IR went	ÔNIBUS bus
BP	O homem	não	andou	foi	de ônibus
‘The man did not walk, he took the bus’.					

<https://youtu.be/WMUqXQEKQlk>

Other types of syntactic restrictions might apply as well in code-blending, stemming from a single derivation process. For example, even in favorable pragmatic contexts we do not expect blends such as (4), combining transitive and intransitive argument structures, to be acceptable.

(4) Unexpected blending: different argument structures

LIBRAS		CASA INCENDIAR house burn
BP	O homem	incendiou a casa
‘The man burned the house/ the house burned.’		

<https://youtu.be/YbFBQMwclCo>

A different conclusion about incongruent blends is provided by Branchini and Donati (2016), based on examples of code-blending in Italian Sign Language (LIS) and spoken Italian. According to the authors’ findings, it is possible for bimodal bilinguals to produce completely distinct structures in the two languages

simultaneously, while the morphology and prosody of each language are preserved. The authors give the following example, which involves a word-order change along the lines of the example in (3).

(5) Italian/LIS

It:	Cosa	ha	mangiato	la	rana?
	what	have.3SG	eat-PTCP	the	frog
LIS:	FROG		EAT	[WHAT] _{WH}	

“What did the frog eat?”

(Branchini and Donati 2016, p.11)

This observation led Branchini and Donati to propose that the computational system can run two different derivations at the same time. However, they do not propose any specific constraints on these combinations. Thus, we understand that they may predict (4) to be possible, although they provide no examples of this sort.

In order to verify the licensing of syntactically incongruent structures and, afterwards, compare Branchini and Donati’s proposal with the proposal put forward in the Synthesis Model, we designed an acceptability judgment experimental study. To the best of our knowledge, this is the first study on code-blending using this specific experimental methodology, although it has been successfully used in studies of code-switching by unimodal bilinguals (cf. Schütze and Sprouse 2014). We expected participants to be able to judge the acceptability of code-blends, based on our previous discussions about code-blending with bimodal bilingual linguists.

In order to supplement our acceptability judgment task, we also ran a follow-up study, an elicitation task, with Codas.

2 The Present Study²

The acceptability judgment task and the elicitation task conducted by us aimed at identifying possible restrictions that may apply to blended sentences of Libras and BP. The tasks include stimuli items with word-order differences among the two languages to address the general question of “what word-order differences are observable in blended structures?”. By investigating both possible and impossible instances of blending, we hope we are able to verify the adequacy of the Synthesis Model as a formal explanation for code-blending in Libras and BP.

² This research was approved by an ethical committee and the participants were volunteers who provided an informed consent, in accordance with the Brazilian CNS-CONEP resolution No. 196/96 version 2012, CAAE: 84511918.0.0000.0121.

2.1 *Participants*

22 Brazilian bimodal bilingual adults completed the acceptability judgment task, and a subset of 6 of these participants completed the elicitation task. All of them have typical hearing and were raised in households that use Libras as the primary language with one or two Deaf parents.

The selected final sample of participants contained individuals with high proficiency in both Libras and BP and individuals with more varied proficiency in Libras. We asked participants to provide a self-assessment of their own skills in Libras and in BP, on a scale from 1 “not fluent” to 7 “very fluent, native.” We also assessed their vocabulary in both Libras and Portuguese using a picture-based assessment (Swadesh List; Swadesh 1971). Tables 1 and 2 present detailed information regarding participants, including their self-assessment of their signing and speech skills, and their vocabulary task results for both languages. Table 1 presents a general summary of each participant including their gender (F for female and M for male); their educational level; their self-assessment in each language where they had assigned from 1 (low fluency) to 7 (high fluency); and their score in the vocabulary test conducted in our experimental set of tests.

2.2 *Materials and Procedure*

The acceptability judgment task was designed to verify participants’ acceptance of a variety of blended sentences, including congruent cases (i.e., simultaneously produced sentences following the same word order, which also were grammatical in both languages), and incongruent cases, in which the word order used in the two languages contrasts. We also considered 12 cases of co-insertion of the type commonly observed in code-blending, expected to be considered fully acceptable, and code-blended utterances designed to be clearly unacceptable by violating the syntax of both languages; in both cases these were used as fillers.

The target stimuli were produced by a fluent bimodal bilingual model and presented to participants, on video, in a quasi-randomized order, so that participants did not view more than two items of the same type in a row.

Participants viewed the video items and rated each one as fully unacceptable (1), intermediate (2), fully acceptable (3), or cannot judge, as illustrated in Fig. 2. Practice items were placed at the beginning of the experimental session to familiarize participants with the task. To set the scale endpoints clearly, during the practice session the experimenter discussed with participants the relative acceptability of different kinds of code-blending, considering extreme cases that are completely acceptable or completely unacceptable. Participants were encouraged to follow their first instinctive reaction in providing responses. Responses equal to or above 2.7 (average score) were classified as HIGH (fully acceptable), from 1.6 to 2.6 MIX (intermediate), and up to 1.5 LOW (fully unacceptable).

Table 1 Judgment task: participants' background information and proficiency in Libras and BP

Participant	Gender	Education level	Fluency self-assessment in BP ^a	Fluency self-assessment in Libras ^a	Libras vocabulary	BP vocabulary
0301	F	Complete high education level	7	7	77,1	97,0
0302	M	Complete high education level	7	6	73,0	95,0
0401	M	Complete high education level	7	7	99,2	99,0
0402	F	Master degree	7	7	95,5	98,0
0403	F	Incomplete high education level	7	7	95,1	95,0
0404	M	Complete high education level	7	7	96,3	96,2
0405	F	Complete high education level	6	7	94,3	96,0
0406	F	Complete high education level	5	7	95,1	100,0
0407	F	Incomplete high education level	6	6	85,3	91,6
0408	F	Master degree	7	7	98,4	94,8
0501	F	Incomplete high education level	6	7	96,3	97,0
0502	F	Complete high education level	6	7	96,3	99,0
0503	F	Complete high education level	7	7	97,1	98,0
0504	F	Incomplete high education level	7	5	77,9	92,0
0505	M	Doctor degree	7	7	93,0	98,0
0506	F	Master degree	6	6	99,6	98,0
0507	F	Complete high education level	6	5	93,1	98,0
0508	F	Complete high education level	7	6	95,9	94,0
0509	F	Incomplete high education level	7	7	95,9	98,0
0510	F	Complete high education level	7	7	93,9	94,0
0601	M	Complete high education level	7	4	66,1	89,2
0602	F	Complete high education level	6	6	97,9	94,2

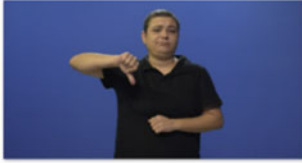
^aThe Codas were asked to self-select one of the numbers in the fluency gradation in both Portuguese and Libras, with 1 meaning “not fluent” and 7 “very fluent, native”

Table 2 Elicitation task: participants' background information and proficiency in Libras and BP

Participant	Gender	Education level	Fluency self-assessment in BP ^a	Fluency self-assessment in Libras ^a	Libras vocabulary	BP vocabulary
20201	F	Master degree	7	7	92,1	97,2
20202	F	Doctor degree	7	7	97,2	95,3
20203	F	Complete high education level	7	7	98,4	99,1
20204	F	Master degree	7	7	95,5	98,0
20205	F	Complete high education level	7	6	95,92	94,0
20206 ^b	F	Incomplete high education level	7	6	–	–

^aThe Codas were asked to self-select one of the numbers in the fluency gradation in both Portuguese and Libras, with 1 meaning “not fluent” and 7 “very fluent, native”

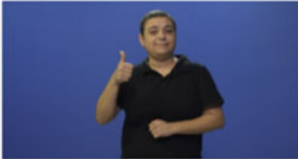
^bThis participant did not complete the vocabulary tasks

1

Fully unacceptable (1)

 2

Intermediate (2)

 3

Fully acceptable (3)

 Não posso julgar

Cannot judge

Fig. 2 Judgment scale

The type of sentential structure was manipulated as an independent variable. The types were: transitive structures, declarative negative sentences, passives, and sentences containing idiomatic chunks. Sentences representing these types were placed under the *congruent* condition if there were no word-order mismatches between the two languages. Mismatch cases were placed under the *incongruent* condition.

Transitive Sentences

Both Libras and BP are SVO languages. However, Libras optionally or obligatorily uses SOV in certain constructions which includes sentences containing verbs marked with aspect, agreement, spatial location, or handling classifiers (Quadros 1999; Quadros and Karnopp 2004). The sentences from the judgment task listed below illustrate congruent and incongruent blendings of transitive sentences:

(6) Congruence at word-order level (VO)

LIBRAS	IX (ele) He	DORMIR CL <i>sleep</i> CL(person-lying-in)	CAMA bed
BP	Ele	tá dormindo	na cama
<i>'He is sleeping in the bed.'</i>			

https://youtu.be/Piwny_TCKPY

(7) Incongruence at word-order level (Libras, OV; BP, VO)

LIBRAS	MULHER woman	TEMPO time	PARAR _[aspect] stop
BP	A mulher	parou	o tempo
LIBRAS	CONGELAR frozen	TODO all	PESSOAS people
BP	e congelou	todas	as pessoas
<i>'The woman stopped time and froze all the people.'</i>			

<https://youtu.be/TiLDWL5ApCc>

(8) Congruence at word-order level (Libras, OV; BP, follows Libras OV)

LIBRAS	IX(eu) I	CHAMAR ask	MULHER woman
BP	Eu	chamo	mulher
LIBRAS	AJUDAR help	CASA house	LIMPAR clean
BP	ajuda	casa	limpar
<i>'I ask the woman to help clean the house.'</i>			

<https://www.youtube.com/watch?v=bnuSknur5iY>

(9) Incongruence at word-order level (Libras, OV; BP, VO)

LIBRAS	HOMEMa man	GATOb cat	aATIRARb shoot
BP	O homem	atirou	no gato
‘The man shot the cat.’			

<https://youtu.be/QpmlffUUdcY>

Negative Sentences

In this sentential type, we included items in which BP followed the Libras negation-final word order, as in (10), and structures in which Libras is negation final while BP preserves its canonical pre-verbal negation as in (11).

(10) Congruence at word-order level (negation in final position)

LIBRAS	MENINO ELE-MESMO IX (ele) boy by-himself	ESCOLHER choose	SORVETE ice-cream
BP	O menino	escolheu	sorvete
LIBRAS	CHOCOLATE chocolat	ABACAXI pineapple	NÃO no
BP	de chocolate	abacaxi	não
‘The boy chose chocolate ice-cream, not pineapple.’			

<https://youtu.be/kSd6o7IxZgQ>

(11) Incongruence at word-order level (Libras, post-verbal neg.; BP, pre-verbal neg.)

LIBRAS	COMIDA food	RS PICANTE RS hot	NÃO not
BP	A comida	gaúcha	não é picante
LIBRAS	MAS but	BAHIA Bahia	FS (é) is
BP	mas	a baiana	é
‘Gaucho food is not spicy, but Bahian food is.’			

<https://youtu.be/PH4AinX9ZX0>

Passive Sentences

Portuguese has verbal passive construction, but there is no fully analogous construction in Libras. In order to explore this, we first considered structures along the lines discussed in Villanueva (2010) for American Sign Language (ASL). Villanueva considered the possibility that in ASL, analogues to passives are found in which the agent is “unfocused,” where the Agent theta-role is not mapped onto the sentence structure or morphologically mapped onto the verb. We did not find examples exactly analogous to this in our data.

Another type of agent de-focusing construction observed in Libras is similar to the high-locus construction as described by Barberà and Hofherr (2017) for Catalan Sign Language (LSC). In Libras, loci in the signing space can be used for reference both in the pronominal system and in verb agreement. In the usual case, loci are used at roughly chest height. In contrast, when employing the high-locus construction the verb is signed with reference to a subject locus in a relatively high area of the signing space, roughly at head height. An overt determiner phrase (DP) may be signed at this locus, but it would consist of a non-specific indefinite pronoun. Whether overt or not, the subject with a verb using a high locus is thus interpreted as non-specific, indefinite. Although high-locus structures background the agent, Barberà and Hofherr (2017) argue that sentences of this type do not behave exactly as a passive, as they do not involve a reduction in transitivity and because there is no evidence that the object is promoted to subject. According to them, in these constructions, there is a deficient referentiality of the subject, without any change in transitivity, comparable to a construction without a specified subject. This seems to hold in Libras as well, since structures of this type maintain the grammatical agentive subject, in contrast to the true passivation process of BP.

There are, therefore, multiple ways of de-focusing the subject of a transitive sentence in Libras, but none of them results in a passive structure analogous to that of BP. As far as we know, there is no true passive construction in Libras. Given these observations, what options are there for blendings involving passives in BP? A short passive (with no *by*-phrase) in BP can be blended with the high-locus construction in Libras, as in (12). The sentence in Libras follows the same order as the passive sentence in BP; however, the verb ROUBAR (“to steal”) is produced with movement from the neutral space towards the head (from lower to higher, since ROUBAR is a backwards verb). This movement is made to a high point in space, which corresponds to an indeterminate subject in Libras, which is not pronounced due to its indeterminacy, enabling Libras to overlap with BP with respect to passivization.

(12) Congruence at the word-order level

LIBRAS	CARTEIRA DV(carreira-no-bolso) wallet(wallet-into-pocket)	HOMEM IX(ele) man he	ROUBAR(down-up) stolen
BP	Carteira	homem	foi roubada
'The wallet of the man was stolen.'			

<https://youtu.be/6zaUWWRUtz0>

A structure in which the *by*-phrase is maintained in BP was combined with a structure in Libras using the emphatic pronoun ELE-MESMO (“by himself”) to mark the agent. In (13), the emphatic pronoun indicates that Machado is the person who wrote the book. These blendings can be derivable, but unpreferred given that the use of ELE-MESMO to convey agency is not generally employed in Libras.

(13) Congruence at word-order level

LIBRAS	DEM that	LIVRO book	ESCREVER write
BP	Aquele	livro	escrito
LIBRAS	ELE-MESMO by-himself	FS (Machado) m-a-c-h-a-d-o	
BP		Machado	
'That book was written by Machado.'			

<https://youtu.be/K81TDxIL484>

Incongruent combinations of BP passives with active sentences in Libras, as in (14), should not be possible, as they cannot be the output of a single derivational procedure.

(14) Incongruence at word-order level

LIBRAS	SENHOR IX(ele) Sir he	ELE-MESMO by-himself	IX(ele) he
BP	O escravo		
LIBRAS	COMPRAR IX(ele) buy he	ESCRAVO slave	
BP	foi comprado	pelo senhor	
'The slave was bought by the master.'			

<https://youtu.be/NAXXHIdVWF0>

Idioms

There seems to be some semantic requirements on blending, particularly that the combined sentences must make only one coherent proposition. Blends violating this requirement are readily rejected, even if they are syntactically congruent. Such cases were tested in our acceptability judgment task through blends involving idiomatic expressions. We included among our target items four combinations of idiomatic expressions with non-idiomatic sentences that were congruent at the meaning level (one sentential idiom and three VP idioms), as in (15), where the BP idiom *bateu as botas* (“die”) is blended with the verb MORRER “to die” in Libras. Two blends of an idiomatic expression with lexically equivalent sentences without semantic correspondence (one sentential and one VP idiom) were also tested. We expected these cases to be rejected, as a single proposition is not delivered when the BP idiom is interpreted idiomatically, while the Libras content is interpreted literally. In (16), for instance, the sentences from Libras and BP are lexically equivalent; nevertheless, they are semantically mismatched as the sentence in Libras is not interpreted as an idiom.

(15) Correspondence at the semantic level

LIBRAS	PADRE priest	MORRER died
BP	O padre	bateu as botas
‘The priest died.’		

<https://youtu.be/Owq4Oitrcg0>

(16) Non-correspondence at the semantic level

LIBRAS	PADRE <i>priest</i>	DV (cair) <i>beat</i>	DV (botas) <i>boots</i>
BP	O padre	bateu	as botas
‘The priest beat the boots.’			

<https://youtu.be/Dux20No9qIc>

In total, the acceptability judgment task consisted of 31 target items. The sentences used as stimuli were chosen because they contrast congruent and incongruent structures in Libras and Portuguese with analogous meaning.

The judgment task was complemented with additional data collected directly with Codas explicitly discussing the target structures used in the acceptability judgment task. This elicitation was conducted in a conversational setting, in which

participants were asked to say how they would produce sentences with the target structures. We asked them to produce the sentences in various possible ways in Libras, trying to produce them using blends of Libras and Portuguese. In this follow-up task, which included 23 target items, we aimed at collecting sentences in which the usual word order for each language would be incongruent in some way, focusing particularly on instances in which the word order in Libras is verb final, while BP would hold its canonical SVO order. This set included sentences with handling verbs or depicting signs (Laszakovits et al. 2022). We also considered sentences with BP passive constructions since passives are not found in Libras, as discussed previously. When participants exhibited a high level of difficulty in blending, they were asked to produce the possible analogous sentences in Libras only, so that we could verify the structures allowed in this language.

3 Results

In general, participants assigned high scores to items in the congruent conditions. In contrast, items in the incongruent conditions received intermediate or low rates.

Results for the transitive sentence type indicate a high level of acceptance if both languages use the same word order, as in example (6). Incongruent combinations, with OV in Libras and VO in BP (examples (7) and (9) above), received low scores. However, blends in which BP follows Libras in using OV word order, as in (8), received high scores. Thus, we may say that there is a preference for maintaining the same word order even if this results in violation of BP word order.

Blended structures with both languages following the position of the negation in Libras (negation final, as in (10)) received high ratings. In addition, blended structures in which Libras used negation final while BP used pre-verbal negation (example (11)) received high ratings. This latter case is an example showing that incongruent blends can be generated due to late linearization.

Combinations like (13), where a bonafide passive with a *by*-phrase in BP is combined with a structure in Libras in which the agent is marked by the emphatic pronoun ELE-MESMO (“he himself”), were judged as intermediate. Blends like (14) were widely rejected by participants. These are cases of word-order incongruence stemming from structural mismatches, not simply late linearization.

We further explored restrictions on passive blends in the elicitation task, and the results indicate that congruent structures are always preferred. In the productions obtained, functional categories come from one of the blended languages (considered the primary language), with the secondary language following the imposed structure. Participants were asked to blend sentences with the same meaning but with different syntactic structures, such as a passive in BP and an active statement in Libras, but, as expected given the results obtained in the acceptability judgment task, they halted, responding that it was not possible. In some cases, they did not blend, producing an active sentence in Libras only. This is illustrated in (17)–(18).

When the experimenter insisted on blending, participants tried examples like (19), but rejected them.

(17)

Experimenter: A menina foi machucada pelo cachorro
 the girl was hurt by-the dog

Participant:

LIBRAS (only)	CACHORRO dog	CL(pessoa) CL(person)	MORDER bite
LIBRAS (only)	PESSOA person	MULHER woman	
'The dog bit the woman.'			

<https://youtu.be/4p121zGGJNs>

(18)

Experimenter: A Maria foi presa pela polícia
 the Mary was arrested by-the police

Participant:

LIBRAS (only)	IX the	POLÍCIA police	PRENDER arrest	FS(maria) m-a-r-i-a
'The police arrested Mary.'				

<https://youtu.be/Eo3dOZIKfvY>

(19)

Experimenter: A Maria foi presa pela polícia.
 the Maria was arrested by-the police

Participant:

LIBRAS	POLÍCIA police	PRENDER arrest	FS (maria) m-a-r-i-a
BP	Polícia	predeu	Maria
'The police arrested Maria.'			

<https://youtu.be/xSvuQX-z85o>

In sum, based on the results from the elicitation task, it is clear that Libras does not structure passives as BP does. Thus, blends involving passives are not possible as such blends would impose non-equivalence in meaning and form.

As expected, combinations involving idiomatic expressions were acceptable if meaning consistency was observed. Thus, while (17) was fully accepted, (18) was fully rejected.

In the elicitation task, participants resisted blending idioms, including cases with semantic correspondence as in (15) and (20) and (21) below. Blends with literal translation equivalents ((21) and (23)) were not spontaneously produced either. These were produced only when the experimenter strongly encouraged the participants to do code-blending, but, in this situation, participants sought possible signs that could provide a good match with the meaning of the BP idiom trying to maintain structural congruency. Sometimes partial blends were produced with some parts of the BP idiom being conveyed in Libras. Interestingly, category matching was observed (e.g., noun with noun, verb with verb), even when the words had different meanings (as in (20)).

(20)

LIBRAS	MÃE mother	TER have	NAMORADO boyfriend	IDADE age
BP	Minha mãe	tem	um namorado	de
LIBRAS	40 forty	&=how	IX(ela) she	DAR+ give
BP	quarenta	anos	e ela	dá tudo
LIBRAS	TUDO all	DAR give	PESSOA person	IX (ele) he
BP	de mão	beijada	pra	ele

'My mom has a boyfriend that is 40 years old and she all "hand kissed" to him.'
(this means that she gives everything to him)

<https://youtu.be/2CQkoh9G848>

(21)

LIBRAS	MÃE IX(ela) mother	PEGAR get	NAMORADO PESSOA boyfriend person	IDADE age
BP	Minha mãe	arrumou	um namorado	de
LIBRAS	40 forty	IX(ela) she	DAR+ give	DEM-MÃO her-hand
BP	quarenta anos	e ela	dá tudo	de mão
LIBRAS	BEIJAR-MÃO kiss-hand	DAR-MÃO give-hand	&=hand-up	
BP		beijada	pra ele	

'My mom got a boyfriend that is 40 years old and she gives all "hand kissed" to him.'

https://youtu.be/XX7grqL_CAs

(22)

LIBRAS		IX(ele) he	TRABALHAR work	POR-ISSO because
BP	Então	ele	trabalha	
LIBRAS	AMOR love	PROFISSÃO job		
BP	por amor		camisa	

'He works "by love of the shirt".' (this means that he works because he loves the job)

<https://youtu.be/eOcyR0IQwdw>

(23)

LIBRAS		IX(ele) he	TRABALHAR work	
BP	Então	ele	trabalha	
LIBRAS	PORQUE because	AMOR love	ROUPA shirt	
BP		por amor	à camisa	

'He works "by love of the shirt".'

<https://youtu.be/KaJwotDjcjQ>

In addition, code-switching was used as a strategy to produce structures with parallel meaning. In (21), for example, the participant used a gesture sign to express BEIJAR-MÃO, showing the act of kissing the hand as parallel meaning together with DAR-MÃO, which is only one sign in Libras. However, (21) was not judged as completely acceptable by its producer. In (22), there is no oral speech during the sign PROFISSÃO, and no sign during the oral production of *camisa*. The same is observed in (23), where PORQUE from Libras has no match in BP. Thus, participants code-switched whenever they did not find an appropriate way to produce blending while keeping congruency of form and consistency of meaning.

In the elicitation task, idioms from Libras were included and participants were asked to blend them with BP, as in (24) and (25). A translation expressing the same meaning in BP was often used preserving structural congruence. Interestingly, participants were less reluctant to produce blends with literal translations of Libras in BP, as in (24). Notice that in (25) code-switching is observed. Also, the signed idioms are very short (two signs).

(24)

LIBRAS	SINAL(André) sign(andr�)	OLHO eye	CARO expensive
BP	O Andr� tem um	olho	caro
'Andr� has a rich eye.'			

<https://youtu.be/XM2yhN6Rw14>

(25)

LIBRAS	SINAL(Andr�) sign(andr�)	OLHO CARO eye expensive	
BP	O Andr� tem um	olhar	diferenciado
'Andr� has a genuine way of seeing (things).'			

<https://youtu.be/1uxIS0Wdv1I>

Some participants declared that blends of idiomatic expressions are strange or funny.

4 Discussion

Overall observations from our acceptability judgment task and from our elicitation follow-up are in accordance with the predictions made by the Synthesis Model. Two important aspects of the model are relevant here: (a) blends must convey a single proposition, and (b) blends are outputs of a single derivation, which is derived by functional features coming from one of the blended languages or from both of them.

In the acceptability judgment task, congruent blends received high scores. This result is in accordance with the observations from the elicitation task, where congruent blends were readily produced.

As shown above, congruent blends may have the structure coming from one language, with the other just following along. This was accepted in general, but there was preference for preserving the structure from Libras, with BP tagging along.

Word-order incongruences received lower ratings. However, these incongruences were accepted in some cases when the two modalities employ different linearization orders. This was shown in the target sentences with the negation morpheme placed in different syntactic positions across the two languages. These sentences were overall accepted by participants.

While in the acceptability judgment task, participants judged several types of incongruent blends to be acceptable, in the elicitation task, when asked to actually produce these blends, the very same participants showed a tendency to respond in only one language, thus avoiding the blending process. Also, when explicitly requested to produce a blend, participants generally choose to structure the given sentence in one of the languages, using fewer words in the other language, or switching back and forth between the two languages and blending only in congruent parts. Thus, we may conclude that participants showed a general preference for congruent blends.

Importantly, semantic compatibility seems to be a stronger requirement than syntactic compatibility. That is, the semantic requirement of a single proposition must be met before syntactic congruency can even be considered. Participants also showed a strong preference for conveying meaning using Libras. This preference may be related to a pragmatic preference for Libras. This preference is illustrated by the observations obtained from idiomatic expressions. Even if syntactically congruent, blends that combined a BP idiom with a literal word-for-word translation equivalent in Libras were rejected. Since the productions in Libras were not interpreted idiomatically, this causes a semantic mismatch. Furthermore, in the elicitation task, participants strongly resisted blending idiomatic expressions from BP with Libras, although when explicitly requested, blends were produced with Libras matching the meaning of the BP idioms (e.g., example (20)). However, blends in the opposite direction – idiomatic expressions from Libras with literal word-for-word translation equivalents in BP – had higher acceptability responses, against our predictions. We interpret these findings as supporting the idea that the derivation itself allows blends of idiomatic expressions with literal meanings; however, restrictions on how meaning is conveyed in sign take preference over restrictions coming from BP. Altogether, we may say that blends of idioms are possible only in restricted cases. Furthermore, such blends may require a stronger mental effort, perhaps for processing reasons. Further research is needed for a more complete account of this.

Some comments on our methodology are in order. Kimmelman (2021) argues that, given modality-related and sociolinguistic specificities of sign languages, there are important methodological considerations for using the acceptability judgment methodology in sign linguistics. In the face of this observation, we consider several

factors. Importantly, sign languages exist in contact with spoken languages, which can lead to signers modifying their sign to adjust to the spoken language (Lucas and Valli 1989). In addition, signers are frequently in contact with less-fluent signers, who might be using contact signing or even an artificial form of signing that purposely follows the structure of the spoken language. Such sociolinguistic factors may interfere in results from judgment tasks, as observed by Kimmelman (2021). Furthermore, we noticed in our elicited data from bimodal bilinguals that some of them produced blended sentences naturally. However, others produced blended utterances as a kind of signed Portuguese, not using the natural type of co-production. There are differences between these two kinds of blended productions. These factors may influence the rated acceptability of the blended sentences presented in our task.

In general, the results reported above support the main theoretical assumption of the Synthesis Model, according to which a single derivational procedure underlies code-blends. In Branchini and Donati's (2016) proposal, code-blending involves two independent derivations, one in each language. The authors do not discuss any constraints on the simultaneous production of two languages, but we can assume that some kind of "one proposition" constraint is imposed on their analysis, that in blends, "the utterance is complete and meaningful only if the two fragments are integrated" (Branchini and Donati 2016, p. 22). Our interpretation is that this statement would not rule out various types of mismatching structures, such as actives blended with passives, which we found to be strongly rejected.

LIS and Italian, the languages studied by Branchini and Donati, are not as similar to each other as Libras and BP are. For example, while Italian uses basic SVO order, LIS uses SOV. Thus, structures that are both congruent and independently attested in LIS/Italian are much less common when compared to Libras/BP. It may well be that the need to produce incongruent blends in order to satisfy the demands of each language leads to a different outcome with respect to acceptability. We will leave this as open possibility, observing, nevertheless, that at least, blends from languages with similar structural constraints, particularly similar word orders, are strong evidence in favor of a single shared deviational procedure.

Overall, our results can be summarized as follows: bimodal bilinguals, when producing Libras and BP, may use a series of strategies to produce blends that are convergent in both languages, prioritizing formal and semantic congruence. Functional features from one of the languages might be selected to drive the derivation, with the other language just tagging along. Strategies used to avoid incongruent blends include code-switching, partial blending, and lexical fillings in the secondary language.

Acknowledgments We gratefully acknowledge the contribution of bimodal bilingual participants. None of this work would be possible without their generous involvement in our studies. We are also thankful to our collaborators and research assistants for their effort, especially Deanna Gagne, who has been a primary collaborator for the US version of the judgment task. This research is supported in part by Award Number #1734120, from the National Science Foundation (NSF), and by Award Number # 440337/2017-8 CNPQ. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF and CNPQ.

References

- Barberà, Gemma, and Patricia Cabredo Hofherr. 2017. Backgrounded agents in Catalan Sign Language (LSC): Passives, middles, or impersonals? *Language, Baltimore* 93 (4): 767–798.
- Branchini, C., and C. Donati. 2016. Assessing lexicalism through bimodal eyes. *Glossa: A Journal of General Linguistics* 1 (1): 48. <https://doi.org/10.5334/gjgl.29>.
- Chen Pichler, Deborah, Julie A. Hochgesang, Diane Lillo-Martin, Ronice Müller de Quadros, and Wanette Reynolds. 2016. Best practices for building a bi-modal bi-lingual bi-national child corpus. *Sign Language Studies* 16 (3): 361–388.
- Davidson, Kathryn, Diane Lillo-Martin, and Deborah Chen Pichler. 2014. Spoken English language development among native signing children with cochlear implants. *Journal of Deaf Studies and Deaf Education* 19 (2): 238–250.
- Emmorey, Karen, Helsa B. Borinstein, and Robin Thompson. 2005. Bimodal bilingualism: Code-blending between spoken English and American sign language. In *ISB4: Proceedings of the 4th international symposium on bilingualism*, ed. James Cohen, Kara T. McAlister, Kellie Rolstad, and Jeff MacSwan, 663–673. Somerville: Cascadia Press.
- Emmorey, K., H.B. Borinstein, R. Thompson, and T.H. Gollan. 2008. Bimodal bilingualism. *Bilingualism: Language and Cognition* 11 (1): 43–61. <https://doi.org/10.1017/S1366728907003203>.
- Goodwin, Corina, and Diane Lillo-Martin. 2019. Grammatical morphemes in the elicited speech of bilingual signing children with cochlear implants. *Journal of Deaf Studies and Deaf Education* 24 (4): 435–447.
- Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20: Essays in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel J. Keyser, 111–176. Cambridge, MA: MIT Press.
- Kimmelman, Vadim. 2021. Acceptability judgments in sign linguistics. In *Cambridge handbook of experimental syntax*, ed. Grant Goodall. San Diego: University of California.
- Koulidobrova, Elena. 2012. *When the quiet surfaces: 'Transfer' of argument omission in the English of ASL-English bilinguals*. Ph.D. dissertation, University of Connecticut.
- . 2016. Language interaction effects in bimodal bilingualism: Argument omission in the languages of hearing ASL-English bilinguals. *Linguistic Approaches to Bilingualism*, online first.
- Laszakovits, Sabina, Quadros, Ronice Müller de, Emily Jo Noschese, Diane Lillo-Martin, and to appear. 2022. Object shift in ASL and Libras. In *The size of things II: Movement, features, and interpretation*, ed. Sabine Laszakovits and Zheng Shen. Berlin: Language Science Press.
- Lillo-Martin, Diane, Quadros, Ronice Müller de, Elena Koulidobrova, and Deborah Chen Pichler. 2010. Bimodal bilingual cross-language influence in unexpected domains. In *Language acquisition and development: Proceedings of GALA 2009*, ed. João Costa, Ana Castro, Maria Lobo, and Fernanda Pratas, 264–275. Newcastle upon Tyne: Cambridge Scholars Press.
- Lillo-Martin, Diane, Quadros Ronice Müller de, Deborah Chen Pichler, and Zoe Fieldsteel. 2014. Language choice in bimodal bilingual development. *Frontiers in Psychology* 5: 1163. (Language Sciences Research Topic: Language by mouth and by hand). <http://journal.frontiersin.org/Journal/10.3389/fpsyg.2014.01163/abstract>. <https://doi.org/10.3389/fpsyg.2014.01163>.
- Lillo-Martin, Diane, Quadros, Ronice Müller de, and Deborah Chen Pichler. 2016. The development of bimodal bilingualism: Implications for linguistic theory. Invited keynote paper for epistemological issue. *Linguistic Approaches to Bilingualism* 6: 719–755.
- Lucas, Ceil, and Clayton Valli. 1989. *Language contact in the American deaf community*. Academic, Published by Elsevier Inc.
- MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: Evidence from code-switching. *Bilingualism: Language & Cognition* 3: 37–54.
- . 2005. Codeswitching and generative grammar. *Bilingualism: Language & Cognition* 8: 1–22.
- Quadros, Ronice Müller de. 1999. *Phrase structure in Brazilian sign language*. Língua: Tese de Doutorado. Pontifícia Universidade Católica do Rio Grande do Sul. Programa de Pós-Graduação em Letras.

- . 2017. *Língua de herança: língua brasileira de sinais*. Porto Alegre: Editora Penso.
- Quadros, Ronice Müller de, and Lodenir Karnopp. 2004. *Estudos linguísticos: língua de sinais brasileira*. Porto Alegre: Editora ArtMed.
- Quadros, Ronice Müller de, Diane Lillo-Martin, and Deborah Chen Pichler. 2012. Bi-national bi-modal bi-lingual corpora of child language. In *Proceedings of the VIIth GSCP (Gruppo di Studi sulla Comunicazione Parlata) international conference: Speech and corpora*, ed. Heliana Mello, Massimo Pettorino, and Tommaso Raso, 87–92. Firenze: Firenze University Press.
- . 2014. Methodological considerations for the development and use of sign language acquisition data. In *Spoken corpora and linguistic studies*, ed. Tommaso Raso, Heliana Mello, and Massimo Pettorino, 84–102. John Benjamins Publishers.
- Quadros, Ronice Müller de, Deborah Chen Pichler, Diane Lillo-Martin, Carina Rebello Cruz, L. Viola Kozak, Jeffrey Levi Palmer, Aline Lemos Pizzio, and Wanette Reynolds. 2015. Methods in bimodal bilingualism research: Experimental studies. In *The Blackwell guide to research methods in sign language studies*, ed. Eleni Orfanidou, Bencie Woll, and Gary Morgan, 250–280. Blackwell.
- Quadros, Ronice Müller de, K. Davidson, D. Lillo-Martin, and K. Emmorey. 2020a. Code blending speech with depicting signs. *Linguistic Approaches to Bilingualism* 10 (2): 290–308. <https://doi.org/10.1075/lab.17043.qua>.
- Quadros, Ronice Müller de, Diane Lillo-Martin, and Deborah Chen Pichler. 2020b. Blending languages: Bimodal bilinguals and language synthesis. In *Sign Multilingualism*, ed. Ulrike Zeshan, Jennifer Webster, and Anastasia Bradford, 201–226. DeGruyter Mouton/Ishara Press.
- Schütze, Carson, and Jon Sprouse. 2014. Judgment Data. In *Research methods in linguistics*, ed. Devyani Sharma and Robert Podesva, 22–50. Cambridge University Press.
- Swadesh, Morris. 1971. In *The origin and diversification of language*, ed. Post-Mortem by Joel Sherzer. Chicago: Aldine.
- van den Bogaerde, Beppie, and Anne Baker. 2005. Code-mixing in mother-child interaction in deaf families. *Sign Language & Linguistics* 8 (1/2): 151–174.
- . 2008. Bimodal language acquisition in Kotas. In *Hearing, mother father deaf: Hearing people in deaf families*, ed. Michele Bishop and Sherry L. Hicks, 99–131. Washington, DC: Gallaudet University Press.
- Villanueva, Miako Noel. 2010. *Non-agent focused constructions in American Sign Language: A Cognitive Linguistic Analysis*. Dissertation (Doutorado em Filosofia) – Department of Linguistics and the Graduate School of Gallaudet, Washington.

Part II
South-American Romance Languages

The Grammaticalization of *Igual* in Argentinean Spanish



Ángela Lucía Di Tullio, Mercedes Pujalte, and Pablo Zdrojewski

1 Introduction

Words expressing identity, similarity, equivalence, and equality seem to be especially subject to grammaticalization processes. A very well-known case is the comparative adjective *mismo* [“same”]. This term has been grammaticalized into the reflexive form *mí mismo* [“myself”], *ti mismo* [“yourself”], and *sí mismo* [“herself/himself”].

The grammaticalization of this kind of expressions involves not just the pronominal domain. The adjective *semejante* [“similar”], for instance, includes comparative forms (1); it can behave as a demonstrative adjective (2) and can even function, in some contexts, as a degree modifier (3).

- (1) Leí un examen *semejante* al de Juan.
read.PST.1SG an exam similar to=the of Juan
“I read an exam similar to Juan’s.”
- (2) No me imaginé que le otorgaran
not CL.1SG.REF imagine.1SG that CL.3SG.DAT grant.FUT.3PL
semejante distinción.
similar distinction
“I did not imagine that she would be awarded with such a distinction.”

Á. L. Di Tullio · P. Zdrojewski (✉)
Instituto de Filología y Literaturas Hispánicas “Dr. Amado Alonso” – Universidad de Buenos Aires, Buenos Aires, Argentina
e-mail: aditulio@filo.uba.ar; pablozd@filo.uba.ar

M. Pujalte
Universidad Nacional de Hurlingham-CONICET – Universidad Nacional de General Sarmiento, Argentina
e-mail: mercedespujalte@gmail.com

- (3) ¡Qué *semejante* tontería!
 what *similar* foolishness
 “That is tremendous nonsense!” (Peru, CdE database)

As in the case of *mismo* [“same”] and *semejante* [“similar/alike”], the Spanish adjective *igual* [“equal”] has undergone several grammaticalization processes. This chapter focuses on some of them and their consequences in terms of micro-parametric variation. Particularly, in Peninsular Spanish, the grammaticalization of *igual* resulted in an epistemic modality marker expressing uncertainty—like “maybe/perhaps,” as in (4). In contrast, in Argentinean Spanish, the grammaticalization mainly gave rise to an operator of concession, similar to “anyway” (5):¹

- (4) Igual mañana voy a tu casa.
 equal tomorrow go.1SG to your home
 “Maybe I will go to your home tomorrow.”
- (5) Está lloviendo. Igual voy a ir a tu casa.
 is raining equal go.1SG to go.INF to your home
 “It is raining. But I will go to your home anyway.”

Within the generative framework, it is commonly agreed that grammaticalization processes imply, in general terms, the insertion of lexical categories in functional positions. Accordingly, Roberts and Roussou (2003) and Roberts (2010), among others, consider that grammaticalization paths involve *successive upward reanalysis along the functional hierarchy, and this is thus how we define grammaticalization paths* (Roberts and Roussou 2003, p. 202).

Following Cavarani-Pots (2020), in this chapter we explore the idea that the transition from lexical categories to purely functional categories entails intermediate stages of semi-lexicalization.² In other words, it could be asserted that semi-lexicalization is not a category itself, but a property of a lexical item in the initial stages of a process, which eventually results in the complete grammaticalization of a lexical item (see Klockmann (2017) and Cavarani-Pots (2020), among others). Consistently, we could assume that roots get a semi-lexical status whenever they are inserted in the functional domain of a different root. This is what happens, precisely, with the adjective of *igual*, which adopts a semi-lexical status when used

¹ We use the terms Peninsular Spanish, American Spanish, and Argentinean Spanish for simplicity. Each term should be understood as referring to a relevant set of varieties spoken in those territories. Although the phenomena analyzed here cannot be circumscribed to a single variety or country, it is worth noting that the epistemic *igual* is characteristic of most of the varieties spoken in Spain and the concessive *igual* is typical of the varieties spoken in Argentina. In this paper we study mainly the Argentinean and the Peninsular values of *igual*, but much work remains to be done with respect to the diatopic variation of these uses.

² We consider the parameters provided by Sanromán Vilas and Carrasco Gutiérrez (2019) and by Cavarani-Pots (2020) to define whether a vocabulary item is a lexical, semi-lexical, or functional category.

as a comparative adjective or adverb. At this stage, the grammaticalization process bifurcates into several paths, which expand the aforementioned process, from its use as a degree adverb to an operator of modality and of concession. The highest degree of grammaticalization of *igual* corresponds to its use as an adversative particle:

- (6) mi gran triunfo fue que detractores terminaron viniendo
 my great triumph was that detractors ended=up coming
 al pie . . . Igual, hubo muchos que siguieron en la suya . . .
 to=the foot equal have many that continued in the theirs
 “My greatest success was that the opponents fell at our feet... However,
 many still did not follow suit.”

(Fuentes, 2011: 80)

This approach to the grammaticalization process is consistent with Hopper and Traugott’s (1993) proposal, which states that the shift from category A to category B does not follow the pattern in (7a), but the one in (7b):

- (7) a. A uniformly > B
 b. A > B/A (> B)

The pattern in (7b) indicates that the grammaticalization of A into B does not mean that A ceases to exist. On the contrary, both forms could be still available and can also go through subsequent grammaticalization processes (see Hopper and Traugott (1993) and Cavarani-Pots (2020)). For instance, *semejante* originated as a comparative adjective (1) and developed in the sixteenth century as a demonstrative adjective (2) and later as a degree modifier (3). These three alternatives, as we already have seen, coexist in several dialects of Modern Spanish.

In this chapter, we explore the hypothesis that the process described by Hopper and Traugott occurs with *igual* in Spanish. We suggest that the epistemic and the concessive uses of *igual* belong to two different grammaticalization paths, which bifurcate at a stage where *igual* gets the value of a focalizing adverb (see Sect. 5).

The chapter is organized as follows. In Sect. 2, we describe the comparative uses of *igual*, whose behavior does not seem to be subject to the variation across Spanish dialects. Section 3 presents an analysis of the properties of *igual* as an epistemic marker expressing uncertainty, typical of Peninsular Spanish. Section 4 examines the use of *igual* as a concessive operator in Argentinean Spanish. In Sect. 5, we spell out our analysis of the grammaticalization process that *igual* undergoes and put forward a proposal to explain the variation between the epistemic use and the concessive use. Section 6 outlines the main conclusions of the chapter.

2 On the Properties of *Igual*

The adjective *igual* and the homonymous adjectival adverb exhibit a distinctive combination of lexical and functional properties. As it is expected of lexical words,

it is a predicate that selects arguments: thus, in *igual a su hermana (en simpatía)* [“like her sister (in terms of friendliness)”], the arguments are the PP *a su hermana* and the standard of comparison, which is commonly implicit. The comparative adjective encodes the functional information of the comparative construction of equality *tan simpática como su hermana* [“as friendly as her sister”], that is, the functional elements that make up the discontinuous constituent *tan . . . como* [“as . . . as”]. This equivalence exposes the comparative functional attribute denoted by *igual*. Therefore, it will be considered a semi-lexical word (Sanromán Vilas and Carrasco Gutiérrez 2019).³

Furthermore, *igual* has given rise to two functional non-comparative uses, as a sentential particle with different values in the Spanish-speaking world. In Argentinean Spanish and other varieties from America,⁴ it is interpreted as a concessive operator, whereas in most varieties of Peninsular Spanish it is interpreted as an epistemic modality marker. The challenge posed by the development of these specific uses is to find a common thread that links the lexical starting point of the process to the subsequent grammaticalization stages. Consequently, we will begin by characterizing the comparative values of *igual*, since we consider they represent the first stage in the grammaticalization process of this lexical item.

2.1 *Igual: A Comparative Predicate*

The comparative feature defines the predicate *igual* both in its adjectival (8) and adverbial (9) uses:

- (8) a. Martín es igual a / que su papá.
Martín is equal to that his father
“Martín is just like his father.”
- b. El precio de este auto es igual
the price of this car is equal
al del mío / que el del mío.
to=the of=the mine that the of=the mine
“The price of this car is equal to mine.”

³ Besides the comparative use, the adjective *igual* has a purely lexical meaning equivalent to *plano* “plain”: *una tierra muy igual* “very plain lowlands.”

⁴ The *Diccionario de americanismos* (DAMER) records this value in Paraguay, Argentina, Chile, Uruguay, Colombia, Bolivia, Ecuador, República Dominicana, Cuba, and Panama. Nevertheless, given the complexity of the data, research on the uses and properties of *igual* in these varieties would be required.

- (9) a. Pedro habla igual que Carlos.
 Pedro speaks equal that Carlos
 “Pedro speaks just like Carlos.”
- b. No pienso igual que vos.
 not think equal that you
 “I do not think like you.”

As expected from other comparative predicates, the second term selected by the adjective in (8) is introduced by the preposition *a* [“to”], which alternates with *que* [“than”]. In contrast, the comparative coda of the adverb in (9) is only construed with the particle *que*. On the same note, when two infinitival clauses denoting events or facts are compared (10a,b), there is an alternation between *a* and *que*, as in (8), but in the case of finite clauses (10c), only the preposition is used as a complement:

- (10) a. Mentir a un muerto es *igual* que mentir=se a
 lying to a dead=person is equal that lying=CL.3.REF to
 uno mismo.
 one=self
 “Lying to a dead person is the same as lying to oneself.”
- b. Elegir una carrera es *igual* a elegir un estilo de vida.
 choosing a career is equal to choosing a style of life
 “Choosing a professional career is the same as choosing a lifestyle.”
- c. Que lo haya reconocido no es *igual*
 that CL.3SG.ACC has acknowledged not is equal
 a que se haya disculpado.
 to that CL.3.REF has apologized
 “The fact that he acknowledged it is not the same as apologizing.”

Igual, like other comparative predicates, is also a symmetrical predicate that admits two patterns: in the first one, illustrated in (8) and (9), the compared terms can be found in different positions in the sentence; in the second one, they are coordinated as in (11a–c). In both cases, *igual* can be modified by the diminutive suffix *-ito* or the adverb *exactamente* [“exactly”], both implying accuracy; or it can be modified by approximative adverbs (11b), but never by degree adverbs, since it is a non-gradable predicate, as opposed to the term *parecido* (*muy parecido*) [“similar (very similar)”]:

- (11) a. Martín y su papá son iguales / igualitos.
Martín and his father are equal equal.DIM
“Martín and his father are alike/exactly alike.”
- b. El precio de este auto y el del mío son casi iguales.
the price of this car and the of=the mine are almost equal
“The price of this car and mine are almost the same.”
- c. Pedro y Carlos hablan igualito.
Pedro and Carlos speak equal.DIM
“Pedro and Carlos speak exactly alike.”
- d. No pensamos exactamente igual.
not think exactly equal
“We do not really think alike.”

In all the examples above, the first term is the subject or one of the conjoined subjects, but, as shown in (12), other syntactic functions or other explicit or implicit elements can be compared, as the implicit *hoy* [“today”] in (12d).

- (12) a. A Martín lo veo igualito a su padre.
ACC Martín CL.3M.SG.ACC see.1SG equal.DIM to his father
“I think Martín is exactly like his father.”
- b. A Martín y a su padre los veo igualitos.
ACC Martín and ACC his father CL.3M.PL.ACC see.1SG equal.DIM
“I think Martín and his father are exactly alike.”
- c. No pienso igual que {vos/ustedes} {sobre eso /acerca de eso}
not think.1SG equal that you.SG/you.PL on that/about of that
“I do not think the same as you {on that matter/about that matter}.”
- d. No pienso [høy] igual que hace veinte años.
not think.1SG today equal that make twenty years
“I do not think the same today as I did twenty years ago.”

Usually, the second term of the comparison can be retrieved from context, be it from the previous discourse or the communicative situation. Both the anaphoric and deictic uses are illustrated in (13):

- (13) a. Me invitaron a la fiesta. Me gustó
 CL.1SG.ACC invited.3PL to the party CL.1SG.DAT liked.3SG
 igual que la del año pasado.
 equal that the of=the year passed
 “I was invited to a party . . . It was as good as last’s year.”
- b. [Context: Someone is pointing to a chair, which is somehow different to other chairs in the same room.]
 Esta silla no es igual, ¿no?
 this chair not is equal not
 “This chair is not the same, is it?”

In contrast, when it expresses identity in mathematical terms, the adjective *igual* only allows the prepositional complement and rejects the diminutive and elative suffixes, as well as the coordinated version:

- (14) a. Dos más dos es igual /?? igualito /?? igualísimo a cuatro.
 two plus two is equal equal.DIM equal.SUP to four
 “Two plus two equals four/??really equals/??most equal.”
- b.* Dos más dos es igual que cuatro.
 two plus two is equal that four
- c.* Dos más dos y cuatro son iguales.
 two plus two and four are equal

Moreover, the *igual* that appears in (14) is purely attributive, while the comparative *igual* of examples (8)–(13) has attributive uses (see (13b)), but it can also act as a modifier inside a DP (15)–(16) (Fernández Alcaide 2011):

- (15) a. un bebé igual de cara a su papá
 a baby equal of face to his father
 “a baby with the face of his father”
- b. un coche igual al mío en precio
 a car equal to=the mine in price
 “a car with the price of mine”

- (16) a. dos personas iguales de cara
 two people equal of face
 “two people with the same face”
 b. los autos iguales en precio
 the cars equal in price
 “Cars equal in price”

As shown in the examples in (8), *igual* establishes a global comparison between the compared terms, whether they are entities (8a) or magnitudes (8b); the standard of comparison is implicit and must be retrieved from the context or situation. Conversely, in (15) and (16), the compared notion is overtly expressed, and it can have a quantity or quality nature. Therefore, there is a difference between this structure and the canonical comparative ones, where the standard of comparison, always explicit, only has a quantity nature.

Igual belongs to the same class of symmetrical comparatives as *idéntico* [“identical”], *similar* [“similar”], *semejante* [“alike”], *opuesto* [“contrary”], *diferente* [“different”], and *distinto* [“unlike”]. Nonetheless, these comparative predicates only select prepositional complements and do not give rise to syntactic constructions, such as comparative structures. Neither are they turned into functional elements, as is the case with *igual* as an adverb of degree: *igual de caros* [“as expensive as”]. Both characteristics, however, are shared by *igual* and *mismo* (Di Tullio 2012a).

In addition to the descriptive uses for *igual* presented so far, there are other contexts where the comparative value of *igual* coexists with other values. These uses introduce subtle differences in meaning that are relevant to consider in the discussion about the grammaticalization path of this item (see (17)–(19)). As can be observed from symmetrical predicates in general, the comparative adverb *igual* acquires meanings associated with addition, which is the reason why (17a) can be paraphrased by a coordinated structure as (17b):

- (17) a. Juan compró tres libros, igual que mi padre.
 Juan bought three books equal that my father
 “Juan bought three books just like my father.”
 b. Juan compró tres libros, y mi padre también.
 Juan bought three books and my father too
 “Juan bought three books and so did my father.”

In a different context, as noted by Fuentes (2011), *igual* gets a correlative distribution pattern: *Igual A que B*, when coordinated two propositions. The initial position is associated with a focus movement of *igual* to the left periphery of the sentence (Llopis 2018).

- (18) Igual voy a caballo que a pie.
 equal go.1SG by horse than by foot
 “I could ride a horse or go walking all the same.”

This configuration, in which two predicates are conjoined, is particularly relevant since it conveys values associated to possibility:

- (19) Igual toca su concierto para piano y orquesta
 equal performs her concert for piano and orchestra
 con una sinfónica, que se presenta en solitario . . .
 with a symphony than CL.3.REF performs in alone
 “She may very well perform her Concert for Piano and Orchestra with a
 symphony or alone...” (Spain, CREA database)

Thus far, we have considered the comparative uses of *igual*. The following sections will deal with the non-comparative uses of *igual*, especially its use as an epistemic and concessive discourse marker.

3 *Igual* as an Epistemic Marker Expressing Uncertainty

As mentioned above, the adverb *igual* presents non-comparative values that are subject to dialectal variation. One of those emerges from its use as an epistemic marker expressing uncertainty, analogous to other markers such as *quizá*, *capaz* (*que*), *tal vez*, *posiblemente* [“*perhaps, it could be that, maybe, possibly*”]. This value is typical of Peninsular Spanish and it is particularly uncommon in Argentinean Spanish (*Diccionario de la lengua española* [DLE], *Nueva gramática de la lengua español* [NGLE] 2009; Martín Zorraquino 2011; Di Tullio 2012a; García Negroni and Marcovecchio 2013, 2014). The first uses in Spain can be traced back to the mid-twentieth century, around the same time when, as will be discussed, its use with concessive value emerges in several varieties of the Spanish spoken in America. Indeed, *igual* in (20) is interpreted as a marker indicating the possibility that the fact expressed in the proposition actually occurs, i.e., the possibility that it snows.

- (20) Igual mañana nieva.
 equal tomorrow snows
 “Maybe it will snow tomorrow.”

The aforementioned literature assumes that the grammaticalization process involved in the meaning of uncertainty in *igual* shifts from the original comparative use in structures such as (21) and (22) to the epistemic value of (23) (see Sect. 2).

(21) [Context: A: If you had won the lottery, what would you have preferred, buying a car or a motorbike?]

B: Hombre, igual me habría apetecido una cosa
 man equal me would.have.1SG preferred one thing
 que la otra.
 than the other

“I would have very well preferred any of them.”

(Martín Zorraquino 2011, p. 403)

(22) Igual pudo ser la guerrilla que el ejército.
 equal could be.inf the guerrilla than the army

“It could have very well been the guerrilla or the army.”

(Di Tullio 2012a, p. 100)

More specifically, NGLE (2009) and Martín Zorraquino (2011) argue that comparative structures provide the semantic basis for that shift. Since the epistemic *igual* only expresses one of the compared facts, whose interpretation denotes that the implicit one could also occur, it is interpreted as a possibility.

Regarding its syntactic properties, this epistemic marker is defined by being a sentence particle that takes an initial position in the sentence.

(23) Igual estás bien un año y al siguiente te
 equal are.2SG okay one year and to=the next CL.2SG.REF
 encuentras mal. No se puede asegurar qué sucederá.
 find.2SG bad not CL.3.REF can assure what happen.FUT

“You may very well be okay one year and feel bad the next one. You never know.”

(Martín Zorraquino 2011, p. 404)

Moreover, as it is the case with other particles from the same class, this variant of *igual* can only precede the structure it modifies (cf. 24a–b), and it cannot be followed by a pause (24c).

- (24) [Context: A: Pedro will be working all day].
- a. B: Igual viene esta tarde.
 equal comes this afternoon
 “He may come this afternoon.”
- b. B: * Viene esta tarde igual.
- c. B: * Igual, viene esta tarde.

Regarding its syntactic distribution, the epistemic *igual* can be preceded by conjunctions, as in (25) and (26):

- (25) [Context:
- A: Sometimes I get the impression that, when I am with my friends, you look at me as if saying: why are you here? Do you know?
- B: What are you saying?!]

A: O igual son imaginaciones mías.
 or equal are imaginations mine
 “Or maybe it is just my imagination.” (DPDE)

- (26) [Context: People talking about a Christmas Eve party. Someone says
- A: I don’t know what they’ll buy. How many people are coming?]
- B: ¿mm? pues igual somos veinticinco o treinta
 hmm well equal are.1PL twenty-five or thirty
 “Hmm we may very well be twenty-five or thirty.” (DPDE)

Igual may well be preceded by a topic:

- (27) Porque a José igual su padre le dice oye,
 because to Jose equal his father CL.3SG.DAT tells hey
 voy yo en=vez de dejarle cogex el coche.
 go.1SG I instead of let.INF=CL.3SG.DAT take.INF the car
 “Because José’s father may very well tell him ‘hey, I’m going’, instead of letting him use the car.” (DPDE)

The following example illustrates that *igual* can not only precede an overt subject, but it can also follow the subject provided that it is topicalized:

- (28) a. Tu hermana igual necesita ayuda.
 your sister equal needs help
 “Your sister may very well need help.” (Hölderlin)
- b. Ese igual no vuelve por aquí en su vida,
 that=one equal not come=back for here in his life,
 vete a saber.
 go.CL.2SG to know
 “He may not come back ever again, who could know?”
 (NGLE, 25.14ñ)

Furthermore, the *Diccionario de partículas discursivas del español* [DPDE] and Santos Río (2003) point out to the fact that this epistemic marker can appear in isolation in an answer (29). Yet, Santos Río states that structures like (30), where there is some kind polarity item, are more frequent than (29):

- (29) [Context:
 A: Do you think they will come?
 B: (Pues) igual.
 (well) igual
 “(Well) maybe.” (Santos Río 2003, p. 414)
- (30) a. [Context:
 A: “Had you been born 50 years ago, would you have been a missionary Sister?”]
 B: Igual sí, seguramente.
 equal yes, surely
 “Well, yes, most probably.” (Spain, CdE database)
- b. Mira no sé cómo eres. Igual sí que tienes
 look not know.1SG how are.2SG equal yes that have.2SG
 sobrepeso. O igual no.
 overweight or equal not
 “Look, I don’t know how you look. You may very well be
 overweight. Or maybe not.” (Spain, CdE database)

Compared with other epistemic markers expressing uncertainty, *igual* cannot occur with the subjunctive mood, and it is exclusively combined with the indicative.

- (31) a. Tal vez viene /venga esta
 maybe comes.PRS.IND.3SG comes.PRS.SUBJ.3SG this
 tarde.
 afternoon
 “She may come this afternoon.”
- b. Igual viene / *venga esta tarde.
 equal comes.PRS.IND.3SG comes.PRS.SUBJ.3SG this afternoon

Moreover, Santos Río (2003) argues that it is different from other kindred markers because it cannot be used in interrogatives:

- (32) [Context:
 A: Juan didn’t come to testify.]
 B: ¿Tal=vez tenga miedo?
 maybe have.SUBJ.2SG fear
 “Maybe, is he afraid.”
 B’:* ¿Igual tiene miedo?
 equal have afraid

Following the restrictions observed for the position of *igual* in relation to conjunctions, topicalized structures, and preverbal subjects, it seems safe to assume that this particle takes a relatively low position in the left periphery of the sentence, probably in ModEpistemicP (see Sect. 5).

That being said, the grammaticalization of comparative adverbs as epistemic markers expressing uncertainty is not uncommon for Spanish at all, since *igual* can be grouped with other adverbial expressions, such as *a lo mejor* [lit. “to the better”] and *lo mismo* [lit. “the same”], *en una de esas* [lit. “in one of them”], and *capaz que* [lit. “it is able that”].

- (33) a. A lo mejor no sea tan malo envejecer
 to the best not be.SUBJ.3SG that bad grow=old.INF
 y morir.
 and die.INF
 “Maybe growing old and dying is not that bad.” (NGLE, 25.4n)
- b. Lo mismo un tío fogoso te la corta.
 the same a man passionate CL.2SG.DAT CL.3FSG.ACC cut
 “A passionate man may rip [your clothes].” (NGLE, 30.11i)

Out of the five markers shown above, Argentinean Spanish only makes use of *a lo mejor, en una de esas y capaz*. *Igual* with this relevant value is not part of this dialect. In fact, this value is particularly observed in colloquial speech of Spain and seems to get different values in other American dialects.

4 *Igual* as a Concessive Marker and Emphatic Marker

In Sect. 2 we mention that *igual*, as other symmetrical predicates, gets values associated with addition. This value of addition would allow its co-occurrence with the adversative conjunction *pero* [“but”] as in:

- (34) a. niños más sucios, desnutridos y harapientos,
 children more dirty malnourished and ragged
 pero igual de solos, igual de asustados
 but equal of lonely equal of scared
 “dirtier, deeply malnourished and ragged children, but just as lonely, just as scared”
 (Spain, CREA database)
- b. Del mismo curso sí que es, pero igual la ponen
 of=the same class yes that is, but equal CL.3F.SG.ACC put
 en otro grupo
 in another group
 “She is from the same class, but she may as well be assigned to another group.”
 (Spain, CREA database)

The examples in (34), which belong to most of the dialects of Spain, show that *pero igual* can get both a comparative (34a) and an epistemic (34b) value, akin to the one discussed in the previous section. In contrast, in many dialects spoken in America (see fn. 4) this construction encodes a concessive value (35b), even though the original comparative value is preserved (35a):

- (35) a. Este autor considera un estudio económico de los problemas,
 this author considers a study economic of the problems
 pero igual metodología podría utilizarse para un estudio
 but equal methodology could=be used for a study
 jurídico.
 legal
 “This author takes into account an economic assessment of the problems, but the same methodology could be applied to a legal study.”
 (Colombia, CREA database)

- b. Le compro ropa de las mejores marcas,
 CL.3SG.DAT buy.1SG clothes from the best brands
 pero igual se las ingenia para parecer
 but equal CL.3.REF CL.3F.PL.ACC manages to look=like
 un adefesio
 a freak
 “I buy her the best clothing brands, but she still manages to look
 ridiculous.” (Chile, CREA database)

This concessive value is the central or most recurring meaning *igual* gets in Argentinean Spanish. In this variety, *igual* functions as an emphatic particle that triggers an inference which counteracts an implicit presupposition in the communicative context. Hence, in (36), the presence of *igual* overrides an implicit affirmation according to which, in case it rains, it would be “normal” not to go to your house.

- (36) Está lloviendo. Igual voy a tu casa.
 is raining equal go.1SG to your home
 “It is raining. But I’ll go to your home.”
 Implicit affirmation: if it rains, I normally do not go to your house.

This semantic contribution is sometimes complemented with an additional emphatic value, which reinforces the affirmation “*I go to your house*,” especially when *igual* occupies the final position in the sentence:

- (37) Está lloviendo. Voy a tu casa igual.
 is raining go.1SG to your home equal
 “It’s raining. I’ll go to your home anyway.”

The concessive value of *igual* can be clearly observed in the following dialogue, adapted from a similar example in Hernanz (2010):

- (38) [Context:
 A: Julia is a very good basketball player.]
 B: Igual es bajita.
 equal is short.DIM
 “She is short anyway.”
 B’: # Igual es alta.
 equal is tall

In this conversation, it is evident that while (38B) is a possible response—since the logical conclusion conveyed in the affirmation *Julia is a very good basketball player* is ruled out (basketball players are not usually short)—(38B') is not a possible response, as there is no conflict between the conclusion that follows from speaker A's affirmation and B's statement: being tall does not contradict the possible conclusions that can be derived from being a basketball player.

That said, without the adverb *igual*, sentence (36) does not trigger any inference, and the juxtaposition of sentences can only be interpreted as the summation of two facts. In this sense, *igual* seems to behave as the conjunction *pero* ["but"] in its concessive use (e.g., *Julia is short, but she plays basketball well*). In other words, the presence of *igual* leads the way towards such inference (Portolés 2001; Martín Zorraquino and Portolés 1999).

However, in contrast with *pero*, this discourse marker can co-occur with the conjunctions *y* ["and"] (39b) and *pero* ["but"] (cf. 35b):

- (39) a.* Estuvo lloviendo a cántaros ayer, y pero
 was raining to pitchers yesterday and but
 la pasamos genial.
 CL.3F.SG.ACC passed.1PL great
- b. Estuvo lloviendo a cántaros ayer, e igual
 was raining to pitchers yesterday and equal
 la pasamos genial.
 CL.3F.SG.ACC passed.1PL great
- “Yesterday, it was raining cats and dogs, and we still had a great time.”

According to Portolés (2001), (39a) is ungrammatical due to a redundancy issue, since both conjunctions display the same function: connecting two clauses. Then, it follows that *igual* is not a conjunction for it does not have the same function as *pero* or *y*.

Moreover, in contrast with other discourse markers of reformulation expressing distance (Martín Zorraquino and Portolés 1999)—e.g., *en cualquier caso* ["in any case"]—counter-argumentative connectives, e.g., *sin embargo* ["nevertheless"], or concessive connectives, e.g., *aun así* ["even so"], *igual* is part of the tonal distribution of the sentence.

This use of *igual* as an operator of concession is very common in Argentinean Spanish, but it is not present in other varieties such as the ones spoken in Spain. Interestingly, written records of this use, similar to what happens with the epistemic use discussed in the previous section, are relatively recent and can be traced back to the mid-twentieth century.

Below, we will present its main syntactic properties and attempt to explain the different semantic and pragmatic contributions this particle makes to the sentence where it appears.

4.1 *The Concessive Value of Igual*

As previously mentioned, *igual* as an operator of concession is a discourse marker that triggers an inference, which overrides an implicit conclusion. Therefore, the statement where it is included cannot initiate discourse. In addition, like in the case with epistemic *igual*, the concessive value rejects modifiers that can be combined with the comparative, such as diminutives, elatives, or precision adverbs: *igualito*, *igualísimo*, and *casi igual*, respectively.

- (40) a. Está lloviendo. * Igualito/*Igualísimo voy a tu casa.
 is raining equal.DIM/equal.SUP go.1SG to your home
- b. Está lloviendo. * Casi igual voy a tu casa.
 is raining almost equal go.1SG to your home

Igual can appear preceding independent clauses (36) or accompanied by the concessive conjunction *pero* (35b), in which case it is mandatory that *igual* be preceded by *pero*:

- (41) * Está lloviendo a cántaros, igual pero voy a tu casa.
 is raining to pitchers equal but go.1SG to your home

The discourse marker *igual* can also introduce the main clause in a concessive construction:

- (42) Aunque está lloviendo a cántaros, igual voy a tu casa.
 even=though is raining to pitchers equal go.1SG to your home
 “Even though it is pouring down, I will go to your home all the same.”

Both in the adversative and the concessive constructions, *igual* is optional: if it is not present, the sentence is still grammatical and the truth conditions of the statement are not affected. Its contribution seems to be circumscribed to reinforcing the concessive character of the construction.

- (43) a. Está lloviendo a cántaros, pero ya salgo para allá.
 is raining to pitchers but yet go=out.1SG to there
 “It’s pouring down, but I’m on my way.”

- b. Aunque está lloviendo a cántaros, voy a tu casa.
 even=though is raining to pitchers go.1SG to your home
 “Even though it’s pouring down, I’m going to your home.”

One major difference with the epistemic *igual* is that the concessive use cannot be used as the answer to a total or partial question:

(44) [Context:

A: I’m going to the stadium.

B: Even though it’s raining?]

A: *Igual voy a la cancha/*Igual sí.
 equal go.1SG to the stadium/equal yes

(45) [Context:

A: You had very little money, didn’t you? What did you buy?]

B: *Igual compré un libro carísimo
 equal bought.1SG a book expensive.SUP

Besides, it can occur in interrogative (46) and imperative clauses (47):

- (46) Está lloviendo a cántaros, ¿igual vas a ir a la cancha?
 is raining to pitchers equal go.2SG to go.INF to the stadium
 “It’s pouring down. Are you still going to the stadium?”

- (47) a. Aunque te aprieten, igual ponete los
 even=though CL.2SG.ACC tighten equal put.2SG=CL.2SG.REF the
 zapatos.
 shoes
 “Even though they are tight, put your shoes on all the same.”
 b. Ya no está lloviendo, pero igual no salgas.
 already no is raining, but equal no go=out.2SG
 “It’s not raining anymore, but still don’t go out.”

In contexts of modality, *igual* can precede the modal operators of possibility and necessity given in (48) and (49), in which case it is out of their scope. For instance, in (48) B’s proposition is true if there exists a possible world where a fellow betrays me. In that statement, *igual* overrides the implicit presupposition that arises from A’s statement (if the person could not be identified, nobody betrayed nobody).

- (48) [Context:
 A: The person who caused the disaster was not identified.]
 B: Igual es posible que algún compañero me haya
 equal is possible that some fellow CL.1SG.ACC have.2SG
 delatado.
 snitched
 “It is likely that some fellow has snitched on me all the same.”
- (49) Aunque esté lloviendo a cántaros, igual tenés que
 even=though is raining to pitchers equal have.2SG that
 ir al colegio.
 go.INF to=the school
 “Even though it’s raining cats and dogs, you still have to go to school.”

In this sense, it can be observed that whenever *igual* is heading the sentence, it is not interpreted within the elliptical site of verbs of cognition, which could indicate that *igual* is out of the scope of the operator of belief. In (50), the counter-expectation induces *igual* only affects the main sentence (*I will go to the party*), but not the complement of *creer*: Julia could be a person for whom the fact that it is raining is not an issue. So, *porque a ella la lluvia nunca la afecta* (50b) is a possible continuation for *Julia también va a ir a la fiesta*, but not for *Igual yo voy a ir a la fiesta* (50c):

- (50) a. Está lloviendo a cántaros. Igual yo voy a ir a la
 is raining to pitchers equal I go.1SG to go.INF to the
 fiesta y creo que Julia también [~~va a ir~~
 party and believe.1SG that Julia too go.3SG to go.INF
~~a la fiesta~~
 to the party
 “It’s raining cats and dogs. I will go to the party anyway and I believe so will Julia.”
- b. porque a ella la lluvia nunca la afecta.
 because ACC her the rain never CL.3F.SG.ACC affects
 “Because the rain never affects her.”
- c. # porque a mí la lluvia nunca me afecta
 because ACC me the rain never CL.1SG.ACC affects

Instead, whenever it occupies the final position, it is interpreted within the scope of the operator of belief and, therefore, within the elliptical site:

- (51) Está lloviendo mucho. Yo voy a ir a la fiesta igual y creo
 is raining a=lot I go to go to the party equal and believe
 que Juan también [~~va~~ ~~a~~ ~~ir~~ ~~a~~ ~~la~~ ~~fiesta~~ ~~igual~~]
 that Juan too go.3SG to go.INF to the party equal
 “It’s raining cats and dogs. I will go to the party anyway and I believe
 so will Juan.”

In light of these facts, it would be safe to conclude that the discourse marker *igual* introducing a proposition is located on the left periphery of the sentence, above modal operators. The position of this operator is high enough to precede (clitic-left)-dislocated topics (52), although it can follow them as well (53):

- (52) a. Aunque llueva, igual a Juan lo voy
 even=though rains.SUBJ equal ACC Juan CL.3M.SG.ACC go.1SG
 a mandar al colegio.
 to send.INF to=the school
 “Even though it’s raining, I will take Juan to school anyway.”
- b. No tengo plata, pero igual yo voy a comprar tu libro.
 not have money but equal I go to buy.INF your book
 “I don’t have money, but I will buy your book anyway.”
- c. No tengo plata. Igual, tu libro, lo voy
 not have money equal your book CL.3M.SG.ACC go.1SG
 a comprar.
 to buy.INF
 “I don’t have money. Your book, I will buy it anyway.”
- (53) a. Aunque llueva, a Juan, igual lo voy a
 even=though rains ACC Juan equal CL.3M.SG.ACC go.1SG to
 mandar al colegio.
 send.INF to=the school
- b. No tengo plata, pero yo igual voy a comprar tu libro.
 not have money but I equal go.1SG to buy.INF your book
- c. No tengo plata. Tu libro, igual lo voy a
 not have money your book equal CL.3MSG.ACC go.1SG to
 comprar.
 buy.INF

In Sect. 5, we will state that the concessive value of *igual* stems from its merge in the head of ConcessiveP.

4.2 *The Emphatic Value of Igual*

In the beginning of this section, we stated that *igual* gets an emphatic value in a wide range of contexts. This value is evident in sentences as (54).⁵

- (54) Está lloviendo, pero voy a la fiesta igual.
 is raining but go.1SG to the party equal
 “It’s raining, but I’m going to the party anyway.”

Igual is not restricted to the final position within the clause, as can be seen in the following examples:

- (55) a. Aunque llueve, voy a la fiesta igual.
 even=though rains go.1SG to the party equal
 b. Aunque llueve, voy igual a la fiesta.
 even=though rains go.1SG equal to the party
 “Even though it’s raining, I’m still going to the party.”

In contrast to *igual* in initial position, this adverb is stressed in examples (54) and (55), which accounts for its emphatic value. This focalized element could be indicating that the subsequent constituents, if any, are dislocated.

Note that this emphatic position is similar to the position of other typical constructions of Argentinean Spanish, which include emphatic negation (56) (Di Tullio 2012b) or generalized *nomás* [“lit: no more”] (57) (Gutiérrez-González and Zdrojewski 2010), among other typical emphatic constructions found in this variety:⁶

⁵ In this context, there could also be a coordinating *aunque* [“even though”], like in (i):

- (i) Cayó ayer ante Rusia, aunque se clasificó igual.
 “It was defeated by Russia even though it could classify anyway.”

Interestingly, when *igual* is combined with the coordinating *aunque*, it also conveys the epistemic value, which is characteristic of the Peninsular varieties:

- (ii) Hacer esto es un poco triste, aunque igual es porque lo hago mal.
 “Doing this is a bit sad, yet it may be because I do it poorly.”

⁶ It is worth noting, however, that the emphatic *nomás* is more restricted than *igual* with respect to its positions in the clause. In fact, *nomás* appears almost exclusively in the final position.

- (56) a. ¡No voy nada!
not go.1SG nothing
“I’m not going at all!”
- b. ¡No te doy nada, el libro!
not CL.2SG.DAT give.1SG nothing the book
“I’m not giving you the book at all.”
- (57) Andá nomás.
go.2SG no=more
“Go away, it doesn’t matter.”

Placed within the clause, *igual* is within the scope of epistemic and modal operators.

- (58) a. Aunque esté lloviendo, es necesario que vayas igual.
even=though is raining is necessary that go.2SG equal
“Even though it’s raining, you should go all the same.”
- b. Está lloviendo, pero es posible que vaya igual.
is raining but is possible that go.1SG equal
“It’s raining, but I may go all the same.”

It should be noted that, in contrast with the initial position which typically conveys concession, this position is not available whenever the discourse context is incompatible with the presence of an emphatic value:

- (59) # Julia es muy buena basquetbolista. Es baja igual.
Julia is very good basketball=player. Is short equal

The example (59) illustrates that *igual* in the final position does not always add a concessive value or, at least, the counter-expectation is more difficult to infer; yet, it is not impossible, as shown in example (60).

- (60) a. ¿No querés? Hacerlo igual.
not want.2SG do=it equal
“Don’t you want to do it? Do it anyway.”
- b. Aunque no me dejes, voy al cine igual.
even=though not CL.1SG.ACC let.2SG go.1SG to=the cinema equal
“Even though you won’t let me, I’m going to the cinema anyway.”

This final position, however, is preferred by the speakers when they intend to highlight the emphatic value of the proposition, its meaning being closer to obligation or necessity.

The existence of these two different values for *igual*, associated with two different positions within the sentence, accounts for the grammaticality of (61):

- (61) Ya sé que no me invitaste a la fiesta.
 already know that not CL.1SG.ACC invite.2SG to the party
 Igual yo voy a ir igual.
 equal I go.1SG to go.INF equal
 “I know that you didn’t invite me to the party. Anyway, I will still go.”

The first instance of *igual* corresponds to its concessive use, while the second one to the emphatic use.

The concessive and emphatic values of *igual* usually overlap, which is the reason why in many cases *igual* can be described as a concessive emphatic operator. In fact, with the right intonation, initial *igual* could instantiate both values. Thus, for instance, a possible answer for (44) would be (62):

- (62) [Context:
 A: I’m going to the stadium.
 B: Even though it’s raining?]
 A: Sí, IGUAL voy a la cancha.
 yes equal go.1SG to the stadium
 “Yes, ANYWAY I’m going to the stadium.”

The meaning can be paraphrased as *I do not care what you think: I’m going to the stadium*. In this case, with the emphatic *igual*, the implicit proposition that is canceled is not the one that emerges from a concessive construction, but from an inference that A assumes is underlying B’s proposition: *do not go to the stadium to watch a game on this weather*.

All in all, in this section we have described the value of *igual* as a concessive emphatic operator in Argentinean Spanish and we have pointed out the essential differences with the epistemic operator characteristic of other varieties such as the ones spoken in Spain. In what follows, we will attempt to provide an account of the properties described so far in light of grammaticalization theory.

5 Semi-lexicality, Grammaticalization, and Variation

The previous sections outlined the comparative uses of *igual*, whose properties are uniform among all Spanish dialects, and the non-comparative uses, which exhibit different values in the Spanish-speaking world. This section will delve into these grammaticalization processes and their variation.

Roberts and Roussou (2003) and Roberts (2010), among others, assume that the grammaticalization of a morpheme is the result of successive upward reanalysis along the functional hierarchy of the clause. In (63) and (64), we provide a summary of the hierarchy of functional categories presented in Cinque (1999, p. 106) and Munaro (2010, p. 153).

- (63) [Mood_{speech-act}P [Moode_{evaluative}P [Mood_{evidential}P
 [Mod_{epistemic}P [T_(Past)P [T_(Future)P
 [Mood_{irrealis}P [Mod_{necessity}P [Mod_{possibility}P
 [Asp_{habitual}P [Asp_{repetitive(I)}P [Asp_{frequentative(I)}P
 [Mod_{volitional}P [Asp_{celerative(I)}P [T_(Anterior)P
 [Asp_{terminative}P [Asp_{continuative}P [Asp_{perfect}P
 [Asp_{retrospective}P [Asp_{proximative}P [Asp_{durative}P
 [Asp_{generic/progressive}P [Asp_{prospective}P [Asp_{Sgcompletive(I)}P
 [Asp_{PIcompletive}P [VoiceP [Asp_{celerative(II)}P
 [Asp_{repetitive(II)}P [Asp_{frequentative(II)}P [Asp_{Sgcompletive(II)}P . . .
- (64) [DiscourseP [ConcessiveP [CounterfactualP
 [EvaluativeP [TopicP [Focus/ InterP [FinP]]]]]]]]

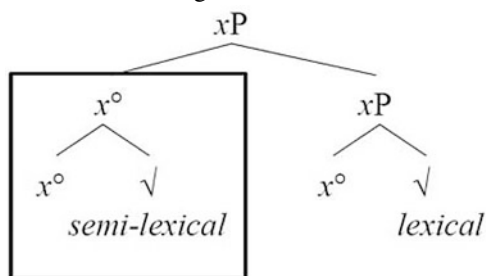
In addition, Roberts and Roussou (2003, p. 36) highlight “that much of the allegedly continuous or cline-like nature of grammaticalization is due to multiple “lexical splits”; [. . .], the different readings attributed to a single lexical item correspond to different positions in which it may be merged in the clause structure.” As we will see below, precisely, not only do these multiple “lexical splits” provide the different values obtained in the interpretation of *igual*, but they also determine the variation attested between, for example, Peninsular and Argentinean Spanish.

Moreover, Hopper and Traugott (1993) insist on the idea that changes operating in grammaticalization processes are not abrupt, but gradual. Even though there is no consensus as to how to delimit the different stages of the grammaticalization process, it is usually agreed that it could follow the descriptive pattern in (65).

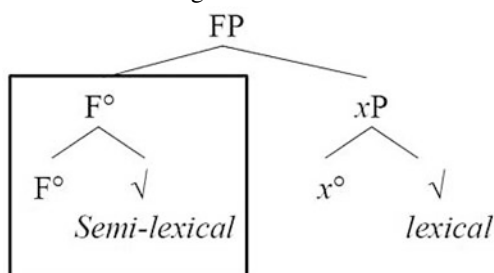
- (65) Content item > Grammatical word > Clitic > Inflectional affix

According to (65), grammaticalization processes are associated to three phenomena: phonological simplification, semantic bleaching, and loss of (head) movement. Evidently, in the case of *igual*, the process involved is not as radical as proposed by (65); instead, its multiple degrees of grammaticalization seem to build semi-lexical categories, as defined by Emonds (1985), Van Riemsdijk (1998), and Corver and Van Riemsdijk (2001). Many recent proposals (Klockmann (2017) and Cavirani-Pots (2020), among others) consider that semi-lexicality is the result of the first stages of a grammaticalization process, which involves the insertion of a root in the functional domain of another one. It is in these early stages of grammaticalization that a lexical item begins to acquire a functional meaning. Cavirani-Pots (2020), particularly, proposes two semi-lexical stages. The first one involves the insertion of a categorized root in the closest functional domain of another categorized root (66a). In the second stage, the root forms a complex head with a functional head (66b). The grammaticalization process could continue with the complete functionalization of the item, which becomes a functional head itself (66c).

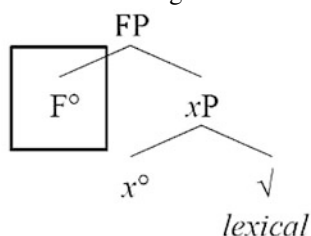
(66) a. Semi-lexical stage I



b. Semi-lexical stage II



c. Functional stage

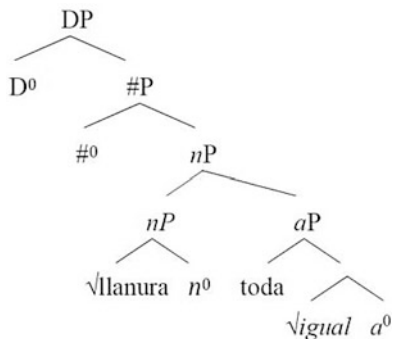


Except for the cumulative meaning of *igual* (see example (6), Sect. 1), all the aforementioned grammaticalized variations of *igual* can be regarded as semi-lexical categories, with different degrees of grammaticalization. Definitely, the patterns that *igual* exhibits in Spanish suggest that there are more than two semi-lexical stages in its grammaticalization path. Below, we will describe them together with their multiple splits.

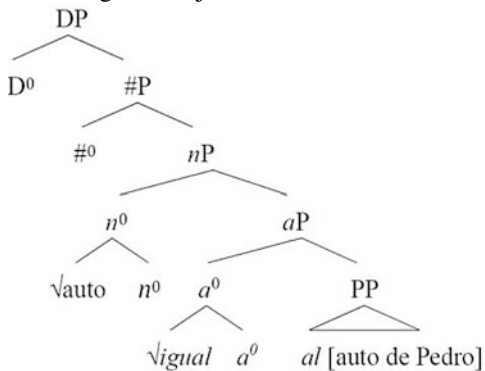
In Spanish, the adjective *igual* has a lexical variant that corresponds to an adjective of quality, meaning “plain” or “uniform” (see fn. 3). This variation preserves the etymological value of *igual*, which comes from the relational adjective *aequalis* in Latin. The first step in the grammaticalization process involves its semi-lexicalization into a comparative adjective (67b). In this step, the descriptive meaning of *igual* shifts to a comparative meaning, and the monadic argument structure is replaced by a dyadic one.

(67) *Semi-lexical stage I*

- a. Caminamos por una llanura toda igual. [adjective of quality]
 walked.1PL through a lowland all equal
 “We walked on the plain lowlands.”



- b. Compró un auto igual al auto de Pedro. [comp. adjective]
 bought.3SG a car equal to=the car of Pedro
 “She bought a car just like Pedro’s car.”



The adjective is re-categorized as a comparative adverb, which later on, in a second grammaticalization stage, becomes an adverb of degree:

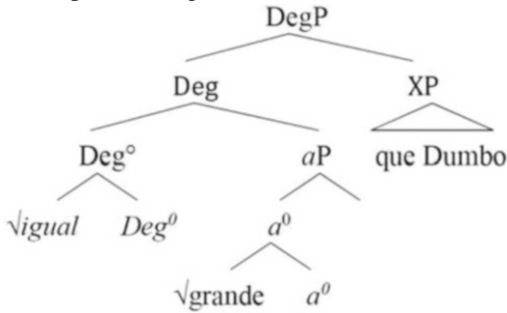
(68) Comparative adverb > Degree adverb



On this stage, it constitutes a closed class with *más* (more) and *menos* (less):

(69) *Semi-lexical stage II*

un elefante igual de grande que Dumbo [adverb of degree]
 an elephant equal of big that Dumbo
 “an elephant as big as Dumbo.”



Furthermore, the re-categorization of *igual* as a comparative adverb constitutes a major step towards its grammaticalization as an epistemic marker expressing uncertainty, and as a concessive marker.

Llopis (2018) states that this process goes through three main stages. The first one is determined by the frequent use of a comparative structure throughout the nineteenth century, which involved focus movement of the adverb *igual* to the left periphery, as shown in (70):

- (70) a. [FocP ... [FinP El decano trabaja igual que el rector]]
 the dean works equal than the chancellor
 “The dean works just as much as the chancellor.”
- b. [FocP Igual ... [FinP trabaja el decano ~~igual~~ que el rector.]]
-

Llopis highlights that, in this construction, the second term of the comparison is usually elided. As a result of this process, the equivalence between the overt constituent and the implied one is emphasized or more salient:

- (71) a. El rector trabaja muchas horas. El decano trabaja igual
 the chancellor works many hours the dean works equal
 que el rector.
 than the chancellor
 “The chancellor works many hours. The dean works just as much as the chancellor.”
- b. El rector trabaja muchas horas. Igual trabaja el decano
 the chancellor works many hours equal works the dean
 [que el rector].
 than the chancellor
 “The chancellor works many hours. So does the dean.”

This configuration gives rise to the “comparative constructions of identical possibility” (*CCIP*), as the one in (72) (see Sect. 2), which emerged, according to Llopis, in the mid-nineteenth century.

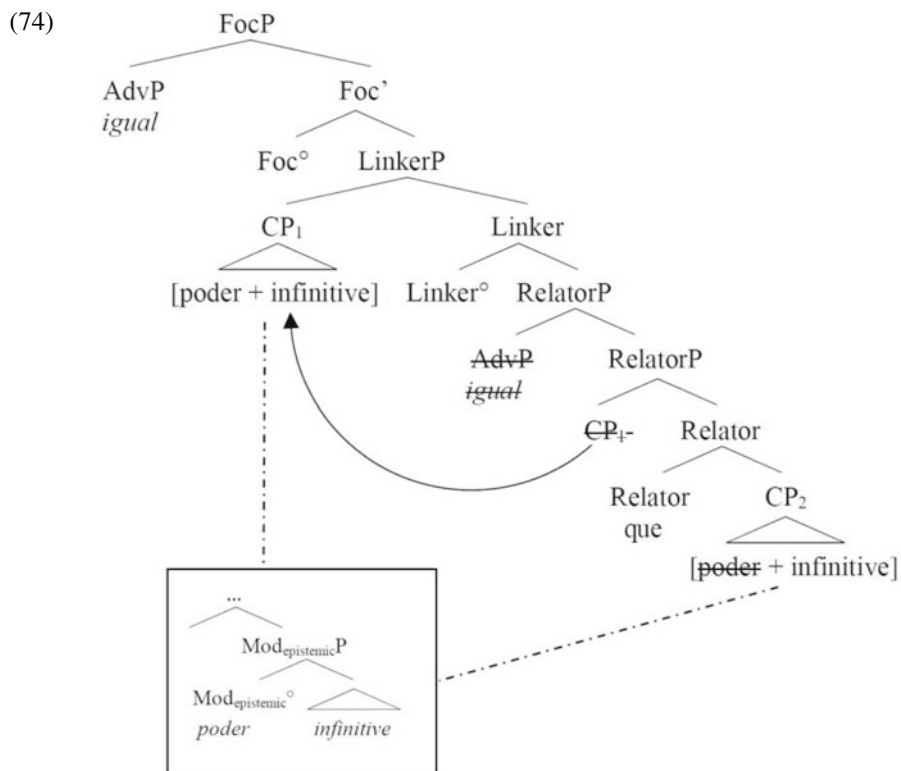
- (72) a. Igual puede Juan decir una mentira que decir la verdad.
 equal may Juan tell.INF a lie that tell.INF the truth
 “Juan may very well tell a lie or tell the truth.”
- b. Igual puede ser cuestión de semanas que de meses.
 equal may be.INF issue of weeks that of months
 “It may as well take weeks or months.” (Llopis 2018, p. 166)

In terms of description, CCIPs with *igual* are made up of a focalized *igual*, the auxiliary *poder* as part of the 1st term of comparison, the conjunction *que* and the 2nd term of comparison:

- (73)
- | | | | | | | | |
|-------|---|------------------------------------|---|-----------------|------|------------------------------------|------|
| Igual | [| poder | + | infinitivo ...] | que | [... | ...] |
| equal | | Aux _{epistemic} | | infinitive | that | | |
| | | | | | | | |
| | | 1 ST TERM OF COMPARISON | | | | 2 ND TERM OF COMPARISON | |

Nonetheless, the details of this analysis are far from evident. It is worth pointing out that this kind of comparative structures expresses a comparison of equality. So, for the time being, we follow Sánchez López’s (2013), who analyzes comparatives of equality as predicative structures in the sense of Den Dikken (2006).

In these configurations, the 1st term of comparison (CP₁) is the subject of the predication, and the 2nd term of comparison (CP₂) is the predicate. The relation of these constituents is mediated by a Relator^o head *que*, and *igual* is an adjunct to RelatorP. We will assume that the 1st term of comparison moves to Spec,LinkerP and *igual* moves to FocP. The analysis of this kind of structures is schematically represented in (74). Notice that both CPs present a ModEpistemic^o head in their inflectional layer. However, such a head can remain silent in CP₂.



Regarding interpretation, most comparisons are semantically associated with addition. In fact, the sentences in (70), repeated in (75a,b) below, could be paraphrased as (75c).

- (75) a. El decano trabaja igual que el rector.
 the dean works equal that the principal
 “The dean works just as much as the principal.”
- b. Igual trabaja el decano que el rector.
 equal works the dean that the chancellor
 “The dean works as much as the chancellor.”
- c. El decano y el rector trabajan
 the dean and the chancellor work
 de la misma manera / en igual cantidad.
 of the same manner in equal quantity
 “The dean and the chancellor work alike/equally.”

In this sense, Llopis suggests that, as opposed to comparative constructions (75c), CCIPs (76) trigger a generalized conversational implicature which excludes one of the compared terms. In other words, the presence of the epistemic auxiliary *poder* in these comparative structures favors a disjunctive reading.

- (76) Igual puede Juan decir una mentira que decir la verdad.
 equal may Juan tell.INF a lie than tell.INF the truth
 “Juan may very well tell a lie or tell the truth.”
 ⇨ Juan can tell a lie \vee he can tell the truth.

As occurs with focalizing comparative structures with *igual* (70), in CCIPs the 2nd term of the comparison is usually elided, which results in both options being possible. Therefore, if the 2nd term of comparison is elided as in (77a), it is implied that the opposite option is also possible (77b):

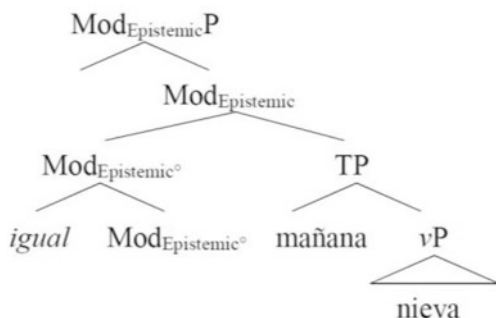
- (77) a. Igual puede decir una mentira.
 equal may tell.INF a lie
 “He may tell a lie.”
 b. Puede decir la verdad.
 can tell.INF the truth
 “He can tell the truth.”

In effect, this seems to be the value expressed by the epistemic *igual* typical of Peninsular Spanish, in examples such as (20), repeated in (78). In these contexts, it is possible to propose that *igual* is reanalyzed as an epistemic adverb expressing uncertainty:

- (78) a. *Igual* mañana nieva.
 equal tomorrow snows
 “Maybe it will snow tomorrow.”
 b. Igual[EPIST]: is possible $p \vee \neg p$

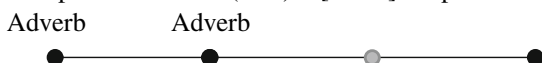
Note that in (79), *igual* is not part of a reduced or simplified comparative structure. We consider that the informal tree diagram in (79) is a plausible description of (78a), where *igual* forms a complex head with $\text{Mod}_{\text{Epistemic}}^{\circ}$.

(79)



The grammaticalization path of the epistemic variation of *igual* could be then schematized as in (80):

(80) Comparative > Focal(ized) > [CCIP] > Epistemic marker



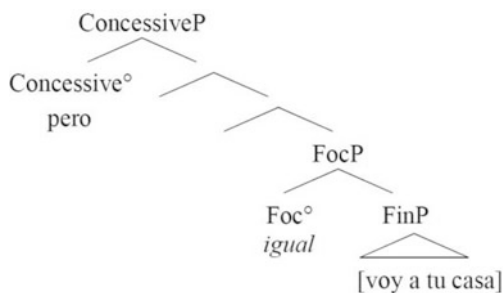
We are now ready to turn to the analysis of the grammaticalization process of concessive *igual* in Argentinean Spanish. As has been previously mentioned, both the epistemic and the concessive uses of *igual* emerged in the mid-twentieth century. We suggest that the lexical split that gives rise to both forms takes place when *igual* turns into a focal adverb.

In this variety, the focalized value of *igual* first co-occurs with the adversative conjunction expressing concession (81):

(81) Está lloviendo a cántaros, pero igual voy a tu casa.
 is raining to pitchers but equal go.1SG to your home
 “It’s raining cats and dogs, but I’ll go to your home anyway.”

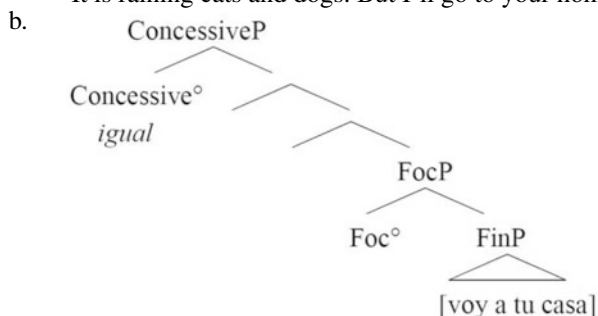
Following the structure sketched by Munaro (2010) for concessive structures, we propose that the left periphery of the main clause in (81) is along the lines of (82):

(82)



The following step is clear. The *igual* in Foc° is attracted to Concessive° ; consequently, *igual* gets a concessive value and is reanalyzed as a marker of concession. This process generates the variant in (83a), analyzed in (83b):

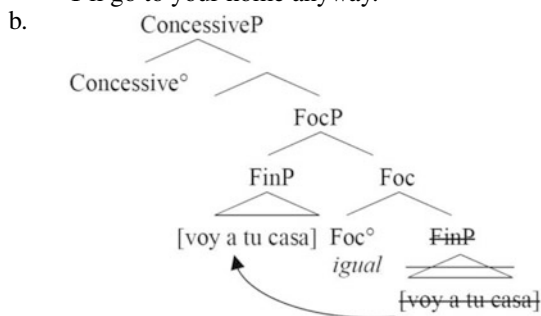
- (83) a. Está lloviendo a cántaros. *Igual* voy a tu casa.
 is raining to pitchers equal go.1SG to your home
 “It is raining cats and dogs. But I’ll go to your home.”



Subsequently, this operator of concession expands into the adversative construction and the concessive construction with *aunque*, losing its focal character.

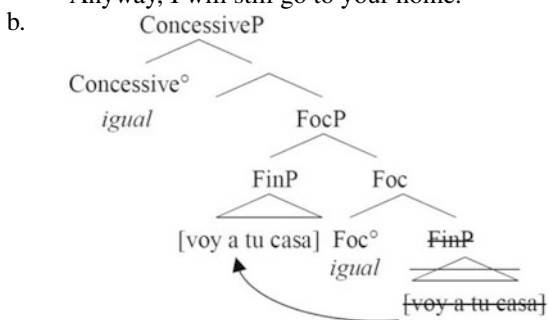
However, it is important to note that both the focalized and the concessive values coexist in Argentinean Spanish. *Igual* with focalized value is used in emphatic contexts (see Sect. 4.2), where it is typically in the final position. In this case, it is possible to assume that there is remnant movement to FocP :

- (84) a. Voy a tu casa *igual*.
 go.1SG to your home equal go.1SG
 “I’ll go to your home anyway.”



The existence of these two values of *igual* can be verified through their co-occurrence as in (85) (see also (61) in Sect. 4). Notice that only the higher *igual* expresses a concessive value, while the one at the end is emphatic.

- (85) a. *Igual* voy a tu casa *igual*
 equal go.1SG to your home equal
 “Anyway, I will still go to your home.”



The grammaticalization path of concessive *igual* would then be the one described in (86):

- (86) Comparative > Focal(ized) > *pero igual* > Concessive marker
 Adverb Adverb



Finally, we will briefly refer to the most grammaticalized variant, which corresponds to the use of *igual* as cumulative particle (6), repeated in (87).

- (87) mi gran triunfo fue que detractores terminaron viniendo
 my great triumph was that detractors ended=up coming
 al pie... Igual, hubo muchos que siguieron en la suya...
 to=the foot equal have many that continued in the theirs
 “My greatest success was that the opponents fell at our feet... However,
 many still did not follow suit.”

(Fuentes 2011: 80)

This value is related to the ones already discussed; as it has already been observed, the comparative structures are somehow associated with addition. This grammaticalization path seems to be independent from the previous two, since it is attested in all Spanish varieties. In this particular case, it is possible that the focalizing *igual* has been reanalyzed as an additive particle. For this reason, the grammaticalization process we propose for this use can be described as follows:

- (88) Comparative > Focal(ized) > Additive Particle
 Adverb Adverb



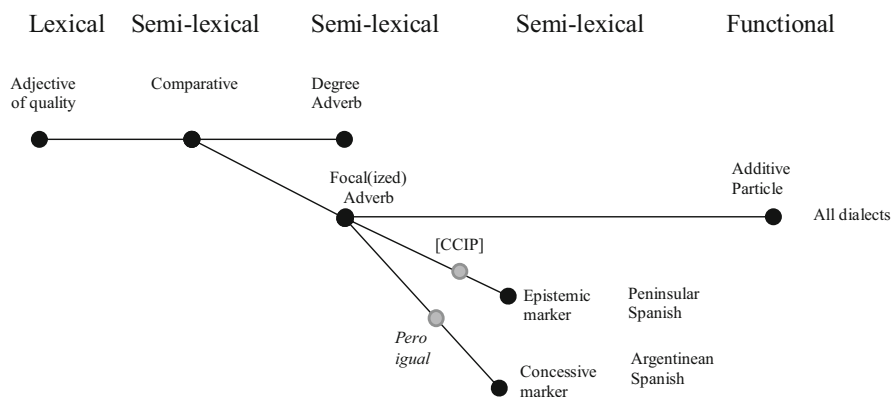


Fig. 1 Cline of grammaticalization of *igual*

In sum, this section aimed at capturing the grammaticalization processes that gives rise to the concessive *igual* characteristic of Argentinean Spanish, while exploring its connection to the grammaticalization processes of the epistemic *igual* and the *igual* expressing addition, which is found in other varieties of Spanish. Figure 1 presents the complete grammaticalization path with its multiple lexical splits.

6 Final Remarks

One of the aims of the micro-parametric agenda is to determine the minimal units of syntactic variation (Kayne 2005). Therefore, the main contribution of this work has been to establish the locus of variation in Spanish dialects within a specific domain: the grammaticalization of the adjective *igual*. Particularly, we observed that this adjective adopts a semi-lexical status when it is used first as a comparative and, afterwards, as a degree element. The insertion of *igual* in the sentential inflectional layers implies an expansion of the same grammaticalization path. It is in this specific part of the path that we observe variation across Spanish dialects. On the one hand, in most varieties of Peninsular Spanish, *igual* is grammaticalized as an epistemic modal operator; on the other hand, in Argentinean Spanish and other dialects spoken in America, it shifts into an emphatic operator of concession. Thus, *igual* would be specified for different formal features in these Spanish dialects.

Regarding the parametric agenda (Kayne 2005; Baker 2008; Biberauer and Roberts 2015, 2017; Roberts 2012, 2019), it is noteworthy that the previously analyzed values occur precisely on the left periphery of the sentence, the most salient area for diatopic variation in Spanish. As we mentioned before (see Sect. 4.2), Argentinean Spanish have a set of constructions linked to emphasis, which includes the emphatic *igual* discussed before. Nevertheless, it is not easy to

establish where these differences between the Spanish varieties are located and, more specifically, if they can be analyzed as a kind of variation linked to a micro-parameter—a small subclass of functional heads share a formal feature—or to a nano-parameter: one or more idiosyncratic lexical items are specified for a formal feature. According to Roberts (2019) “what distinguishes the parameter types is the classes of lexical entries in the functional lexicon which the presence/absence of a given formal feature (or set of formal features) ranges over” (Roberts 2019, p. 76). Thus, the micro-parameters affect a subset of certain functional heads, such as some of the heads located in the left periphery of the clause; but the nano-parameters involve individual lexical entries, as it seems the case of *igual*. In this sense, the parametric variation which is observed in this chapter could be an example of a nano-parameter, since it involves a single lexical item specified for different formal features in different Spanish dialects. However, it is possible that a micro-parameter linked to the left periphery of the sentences converges in the same phenomenon. It remains for future works to evaluate these possibilities.

In terms of grammaticalization theory, we have been able to, on the one hand, provide evidence in favor of Hopper and Traugott’s (1993) hypothesis, which considers the coexistence of lexical items that reflect different stages and forking paths of the grammaticalization process of a word, while, on the other hand, we have emphasized the fact that, in the case under discussion, there happens to be more than one bifurcation at the same stage of the grammaticalization process. Oddly enough, those paths virtually originated around the same period, the mid-twentieth century. Therefore, it is possible that those two open paths will give rise to different functional words, which in turns deepen the differences between dialects.

Acknowledgments We want to thank the editors for inviting us to participate in this project. We are very grateful for the detailed and insightful comments we received from them. All errors and omissions are our own. Also, we would like to thank Juan José Arias and Rocío Maure for helping us with the proofreading and English writing of the paper and Romina Trebisacce for the discussion on some examples.

References

- Baker, Mark. 2008. The macroparameter in a microparametric world. In *The limits of syntactic variation*, ed. Theresa Biberauer, 351–373. Amsterdam: Benjamins.
- Biberauer, Theresa, and Ian Roberts. 2015. The clausal hierarchy, features and parameters. In *Beyond functional sequence*, ed. Ur Shlonsky, 295–313. Oxford: Oxford University Press.
- . 2017. Parameter setting. In *The Cambridge handbook of historical syntax*, ed. Adam Ledgeway and Ian Roberts, 134–162. Cambridge: Cambridge University Press.
- Cavirani-Pots, Cora. 2020. *Roots in progress. Semi-lexicality in the Dutch and Afrikaans verbal domain*. Amsterdam: LOT Dissertation series.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.
- Corver, Norbert, and Henk van Riemsdijk. 2001. Semi-lexical categories. In *Semi-lexical categories: The content of function words and the function of content words*, ed. Norbert Corver, and Henk van Riemsdijk, 1–22. Berlin, Mouton de Gruyter.

- DAMER=Real Academia Española and Asociación de Academias de la Lengua Española. 2010. *Diccionario de americanismos*. <https://lema.rae.es/damer/>.
- Den Dikken, Marcel. 2006. *Linkers and relators*. Cambridge, MA: MIT Press.
- Di Tullio, Ángela. 2012a. *Igual: un comparativo sin igual*. In *Léxico y sintaxis*, ed. Mabel Giammatteo and Laura Ferrari, 93–105. Mendoza: Editorial FFyL-UNCuyo y SAL.
- . 2012b. Enunciados ecoicos focalizados en el español rioplatense. In *Español al sur*, ed. Gerardo Canepa Álvarez, María José González, and Cristina Pippolo, 41–48. Montevideo: Administración Nacional de Educación Pública.
- DLE= Real Academia Española. Diccionario de la lengua española, 23.^a ed., [online version 23.4]. <https://dle.rae.es>.
- DPDE= Briz, Antonio, Salvador Pons, and José Portolés. 2008. *Diccionario de partículas discursivas del español*. <http://www.dpde.es>.
- Emonds, Joseph. 1985. *A unified theory of syntactic categories*. Dordrecht: Foris.
- Fernández Alcaide, María. 2011. *Igual: historia y valores*. Adjetivo y construcciones. In *Sintaxis y análisis del discurso hablado en español. Homenaje a Antonio Narbona*, ed. José Jesús de Bustos Tovar, Rafael Cano-Aguilar, Elena Méndez García de Paredes, and Araceli López Serena, 509–524. Sevilla: Editorial Universidad de Sevilla.
- Fuentes, Catalina. 2011. Conexión y debilitamiento asertivo: *igual*, igualmente, lo mismo. In *Gramática y discurso. Nuevas aportaciones sobre partículas discursivas del español*, ed. Ramón González Ruiz and Carmen Llamas Sáiz, 75–103. Pamplona: Eunsu. Ediciones Universidad de Navarra.
- García Negroni, María Marta, and Ana María Marcovecchio. 2013. No todo da lo mismo: de la comparación al distanciamiento. *El caso de igual*. *Oralia* 16: 167–186.
- . 2014. *Igual a un lado y otro del Atlántico*. Un origen común para dos valores argumentativos. In *Marcadores del discurso: Perspectivas y contrastes*, ed. María Marta García Negroni, 141–157. Buenos Aires: Santiago Arcos.
- Gutiérrez-Gonzalez, Yurena, and Pablo Zdrojewski. 2010. *Focus and doubled DPs: Interactions and Intersections*. Poster presented at XX Colloquium on Generative Grammar, Universitat Pompeu Fabra, Barcelona, España. 18–20 March 2010.
- Hernanz, María Lliüsa. 2010. Assertive Bien in Spanish and the Left Periphery. In *Mapping the Left Periphery*, ed. Paola Benincà and Nicola Munaro, 19–62. New York: Oxford University Press.
- Hopper, Paul, and Elizabeth Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kayne, Richard. 2005. *Movement and silence*. Oxford: Oxford University Press.
- Klockmann, Heidi. 2017. *The design of semi-lexicality. Evidence from case and agreement in the nominal domain*. Utrecht: LOT Dissertation series.
- Llopis, Ana. 2018. Sobre la pragmaticalización de lo mismo e *igual* como marcadores epistémicos. In *Enunciado y discurso: estructura y relaciones*, ed. Ester Brenes, Marina González, and Francisco Grande Alija, 153–179. Sevilla: Editorial Universidad de Sevilla.
- Martín Zorraquino, María Antonia. 2011. Juan vendrá *igual* mañana (que vino ayer)/*Igual* vendrá Juan mañana. . . . In *60 problemas de gramática dedicados a Ignacio Bosque*, ed. María Victoria Escandell Vidal, Manuel Leonetti Jungl, and Cristina Sánchez López. Madrid: Ediciones AKAL.
- Martín Zorraquino, María Antonia, and José Portolés. 1999. Los marcadores del discurso. In *Gramática descriptiva de la lengua española*, ed. Ignacio Bosque and Violeta Demonte, vol. 3, 4051–4213. Madrid: Espasa Calpe.
- Munaro, Nicola. 2010. Toward a hierarchy of clause types. In *Mapping the left periphery*, ed. Paola Benincà and Nicola Munaro, 125–162. New York: Oxford University Press.
- NGLE= Real Academia Española and Asociación de Academias de la Lengua Española. 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa-Calpe.
- Portolés, José. 2001. *Marcadores del discurso*. Madrid: Ariel.

- Roberts, Ian. 2010. Grammaticalization, the clausal hierarchy and semantic bleaching. In *Gradience, gradualness and grammaticalization*, ed. Elizabeth Traugott and Graeme Trousdale, 45–73. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- . 2012. Macroparameters and Minimalism: a programme for comparative research. In *Parameter theory and linguistic change*, ed. Charlotte Galves, Sonia Cyrino, Ruth Lopes, Filomena Sândalo, and Juanito Avelar, 320–335. Oxford: Oxford University Press.
- Roberts, I. 2019. *Parameters hierarchies and universal grammar*. Oxford: Oxford University Press.
- Roberts, Ian, and Anna Roussou. 2003. *Syntactic change: A Minimalist approach to grammaticalisation*. Cambridge: Cambridge University Press.
- Sánchez López, Cristina. 2013. Cuanto antes, mejor, y otras comparativas proporcionales. In *Las estructuras comparativas*, ed. Luis Sáez and Cristina Sánchez López, 309–336. Madrid: Visor Libros.
- Sanromán Vilas, Begoña and Carrasco Gutiérrez, Ángeles. 2019. Introducción. Categorías semiléxicas. Haciendo balance entre el contenido léxico y la función gramatical. *Revista Española de Lingüística* 49(1):7–20. <https://doi.org/10.31810/RSEL.49.1>.
- Santos Río, Luis. 2003. *Diccionario de partículas*. Salamanca: Luso-Española de Ediciones.
- van Riemsdijk, Henk. 1998. Categorical feature magnetism: The endocentricity and distribution of projections. *The Journal of Comparative Germanic Linguistics* 2: 1–48.

Corpora

CREA= Real Academia Española, Corpus de Referencia del Español Actual.

CdE= Mark Davies, Corpus del Español.

Hölderlin= Vallejo A (1984) Hölderlin. In; Primer acto: cuadernos de investigación teatral, pp. 80–102.

Approaching the So-Called Topic-Subjects in Brazilian Portuguese from Below



Jairo Nunes and Mary A. Kato

1 Introduction¹

Since the seminal work by Pontes (1987), the literature on Brazilian Portuguese syntax (henceforth BP) has paid special attention to the so-called “*topic-subject*” constructions, that is, constructions where an apparent topic functions as a subject, controlling verbal agreement (see, e.g., Galves (1987, 1998), Kato (1989), Lunguinho (2006), Avelar and Galves (2011), Munhoz and Naves (2012), de Andrade and Galves (2014), Nunes (2016, 2017), and Kato and Ordóñez (2019)). These include, among others, constructions involving “possessor raising,” as in (1a), and “locative agreement,” as in (1b).

¹ This research has received partial support from CNPq (grants 303195/2019-3 – first author – and 304954/2021-7, second author). We would like to thank Janayna Carvalho, Cláudia Coelho, Renato Lacerda, Ezekiel Panitz, Cilene Rodrigues, and Andrés Saab for valuable comments on a previous version of this chapter.

J. Nunes (✉)
Universidade de São Paulo, São Paulo, Brazil
e-mail: jmnunes@usp.br

M. A. Kato
Universidade Estadual de Campinas, Campinas, Brazil
e-mail: mary.kato@gmail.com

- (1) a. [Os relógios] **quebraram** o ponteiro.
 the watches broke-3PL the arm
 “The arms of the watches broke.”
- b. [Essas gavetas] **cabem** muita coisa.
 these drawers fit-3PL many thing
 “Many things can fit in these drawers.”

In (1a), the plural possessor *os relógios* “the watches” associated with the noun *ponteiro* “arm” surfaces in a preverbal position and triggers third-person plural agreement on the verb.² In (1b), in turn, the plural locative argument associated with the unaccusative verb *cabem* “fit” is not headed by the expected preposition *em* “in” and is realized in a preverbal position, also triggering verbal agreement.³

² In English, a “subject-oriented” language, sentences corresponding to (1a) involve the verb *have* and a possessive pronoun, as illustrated in (ia) below. European Portuguese also has structures analogous to (ia), as shown in (ib). We will see below that the emergence of constructions such as (1a) in BP is related to the loss of third-person possessive pronouns in the language.

- (i) a. The watches had their hands broken.
 b. *European Portuguese:*
 Os relógios tiveram **seus** ponteiros quebrados.
 the watches had their hands broken
 “The watches had their hands broken.”

In Japanese, a “topic-oriented” language, the possessor may surface with a topic marker and the possessee, with nominative Case, as illustrated in (iia) below. BP also allows constructions analogous to (iia), as illustrated in (iib). Due to space limitations, in this chapter we will not be able to discuss constructions such as (iib) in BP. For relevant discussion, see, e.g., Pontes (1987), Kato (1989), Galves (1998), Bastos-Gee (2011), and Nunes (2016).

- (ii) a. *Japanese:*
 Kono tokei-wa hari-ga oreta.
 clock-TOP hand-NOM broke
 “The watch had its hand broken.”
- b. *Brazilian Portuguese:*
 O relógio, o ponteiro quebrou.
 the watch the arm broke
 “The watch had its arm broken.”

³ We take constructions with weather verbs like (i) below, where an apparent prepositionless locative adjunct controls verbal agreement, to be analyzed along the lines of unaccusative verbs like (1b), with the locative being generated as an internal argument (see Sect. 4 below).

- (i) [Essas cidades] **chovem** muito no verão.
 these cities rain-3PL much in-the Summer
 “It rains a lot in these cities during Summer.”

Details aside, the prevailing view in the literature is that these constructions arose in BP as it ceased to be a canonical pro-drop language due to the weakening of its verbal agreement inflection and became a topic-prominent or discourse-oriented language.⁴ From this perspective, as BP is in the course of becoming a non-pro-drop language, its [Spec,TP] tends to be overtly filled. In the case of unergative and transitive verbs, the external argument moves to this position, generally excluding the possibility of VS order commonly found in canonical Romance pro-drop languages. In the case of verbs lacking external arguments, VS is still allowed, as the subject may be licensed VP-internally. This in turn paves the way for elements other than the standard subject to occupy [Spec,TP] in constructions without an external argument. Given that BP also displays characteristics of a topic-prominent language, it is commonly held that topics came to be allowed to occupy [Spec,TP], yielding “topic-subject constructions” like the ones in (1).

Although the factors mentioned above certainly play a role in the emergence of constructions like (1) in BP, in this chapter we argue that they are not the primary causes of this new development in BP as they relate to portions of structure that are too high in the clausal domain. Assuming the general framework of the Agree-based model (Chomsky 2000, 2001), we argue instead that “topic-subjects” are regular subjects in the sense that they are derived by A-movement of a DP from a position within *v*P to [Spec,TP]. In other words, the relevant diachronic changes that gave rise to “topic-subject” constructions in BP actually involve changes in its *v*P and DP layers.

The chapter is organized as follows. In Sect. 2, we make some qualifications regarding the putative association of constructions like the ones in (1) with topics, showing that focus constructions display the same pattern of agreement seen in (1). In Sect. 3, we discuss diachronic changes affecting Case-licensing within *v*P and DP in BP, which yielded “topic-subject” constructions as a byproduct. In Sect. 4, we show why A-movement of the “topic-subject” to [Spec,TP] does not violate minimality. Section 5 discusses some fine-grained distinctions noted in Sect. 3 regarding person asymmetries and resumption in “topic-subject” constructions. Finally, Sect. 6 concludes the chapter.

2 “Topic-Subjects” Are Not Topics, But Subjects

Before we tackle the derivation of “topic-subject” constructions proper, a couple of comments are in order. “Topic-subject” constructions are not, strictly speaking, restricted to topics. The underlined constituents in (2) below, for example, involve

⁴ For relevant discussion, see, e.g., Pontes (1987), Kato (1989), Galves (1998), Negrão (1999), Modesto (2008), and Duarte and Kato (2008).

different types of foci and the resulting constructions are as well formed as the ones in (1). The sentence in (3), in turn, shows that “topic-subject” constructions can be appropriate answers for out-of-the-blue questions, showing that they do not necessarily convey a categorical judgment (in the sense of Kuroda (1979)), as we would expect if they were true topic constructions.⁵

(2) a. **Focus with *só* “only”**

Só esses relógios **quebraram** o ponteiro.
 only these watches broke-3PL the arm
 “Only these watches had their arms broken.”

b. **Focus with *nem* “not even”**

Nem essas gavetas **cabem** muita coisa.
 nor these drawers fit-3PL many thing
 “Not even these drawers are large enough.”

c. **Contrastive focus**

OS RELÓGIOS **acabaram** a bateria (não os celulares)
 the watches finished-3PL the battery not the cell.phones
 “The batteries of the watches (not the cell phones) are dead.”

d. **Aggressively non-D-linked *wh*-constituents**

Que diabo de carro vai fundir o motor depois de passar
 what devil of car goes melt the engine after of pass
 pela revisão?
 through-the revision
 “What kind of car has its engine stopped soon after it leaves the garage?”

e. **D-linked *wh*-constituents**

A: – Que cidades **chovem** muito no verão?
 which cities rain-3PL much in-the Summer
 “In which cities does it rain a lot during Summer?”

Information focus

B: – Rio e São Paulo **chovem** muito no verão.
 Rio and São Paulo rain-3PL much in-the Summer
 “It rains a lot in Rio and São Paulo during Summer.”

⁵ We thank Renato Lacerda (p.c.) for the observation regarding (3).

- (3) A: – O que aconteceu?
 what happened
 “What happened?”
 B: – O celular acabou a bateria.
 the cell.phone finished the battery
 “The battery of the cell phone is dead.”

These observations are relevant for two reasons. First, they attenuate the putative connection between constructions such as the ones in (1) and topics. To put it in different words, (2) and (3) show that whatever makes constructions like (1) grammatical in BP, it should be structural in nature, rather than informational (see Lacerda (2020) for relevant discussion). In what follows, we will keep using the term *topic-subject* only for presentation purposes.

The second reason why the grammaticality of the sentences in (2) is relevant is that the types of focalization seen in (2a–d) cannot be base-generated (see, e.g., Cinque (1990)), as independently shown by the fact that the focalized constituent is sensitive to islands and is incompatible with a resumptive pronoun:⁶

- (4) a. * Só esse relógio a Maria conversou com o cara
 only this watch the Maria talked with the guy
 que queria vender (ele).
 that wanted sell it
 “Maria talked with the guy that wanted to sell only this watch.”
 b. * Nem essa gaveta o João ficou contente depois de
 nor this drawer the João stayed content after of
 consertar (ela).
 fix it
 “John did not get happy even after fixing this drawer.”

⁶ D-linked *wh*-constituents like the one in (2e), on the other hand, may be base-generated in BP. Accordingly, they are compatible with resumptive pronouns and do not display island effects, as illustrated in (i).

- (i) Que livro todo mundo que leu (ele) resolveu mudar de vida?
 which book every world that read it resolved change of life
 “Which book is such that everyone who read it decided to change his life?”

- c. * O RELÓGIO a polícia prendeu o ladrão
 the watch the police arrested the thief
 que roubou (ele) não o celular.
 that stole it not the cell.phone
 “The police arrested the thief that stole THE WATCH, not the cell phone.”
- d. * Que diabo de carro o João se arrependeu depois de
 what devil of car the João himself repented after of
 comprar (ele)?
 buy-SUBJ it
 “What kind of car did John regret having bought?”

This indicates that the focalized constituents in (2a–d) are not base-generated and have reached their surface position via movement. All things being equal, the same conclusion should also apply to the classic “topic-subject” constructions in (1). In other words, the grammaticality of constructions such as (2a–d) refutes the recurring idea that the subject of “topic-subject” constructions is directly merged in [Spec,TP]. Thus, the null hypothesis is that the DPs that trigger verbal agreement in constructions such as the ones in (1) reach the standard subject position ([Spec,TP]) via movement, pretty much like the ones in (2a–d). From [Spec,TP], they may eventually move to higher A'-positions in the left periphery – an issue that is orthogonal to the derivation of “topic-subject” constructions itself.

Another fact that corroborates this conclusion was observed by Galves (1998), who noted that “topic-subject” constructions are incompatible with resumptive pronouns, thus contrasting with regular topic constructions, as illustrated in (5).

- (5) a. [Os relógios], **quebrou** o ponteiro **deles**.
 the watches broke-3SG the arm of-them
- a'. * [Os relógios] **quebraram** o ponteiro **deles**.
 the watches broke-3PL the arm of-them
 “The arms of the watches broke.”
- b. [Essas gavetas], **cabe** muita coisa **nelas**.
 these drawers fit-3SG many thing in-them
- b'. * [Essas gavetas] **cabem** muita coisa **nelas**.
 these drawers fit-3PL many thing in-them
 “Many things can fit in these drawers.”

The lack of verbal agreement with the topic in (5a) and (5b) signals that they instantiate a standard topic construction and the compatibility with the associated pronoun shows that the topic can be base-generated. In turn, the ungrammaticality of the “topic-subject” constructions in (5a') and (5b'), which display verbal agreement with the alleged topic, is to be expected if the subject cannot be generated where it surfaces and must therefore compete with the associated pronoun for the relevant position within *vP*.

So, the interim conclusion we reach is that, unless proven otherwise, the DP that triggers verbal agreement in a “topic-subject” construction in BP is not necessarily a topic (although it may later move to a topic position) and comes to occupy the subject position by moving from a ν P-internal position. From the perspective we are exploring here, this indicates that the availability of “topic-subject” constructions is not related to some special property to be ascribed to the upper part of the clausal domain.

3 Changes Within ν P and DP

Our starting point is Kato and Ordóñez’s (2019) proposal regarding the diachronic source for the emergence of “topic-subject” constructions in BP, based on a comparison between BP and Dominican Spanish, a language that has also been analyzed as becoming non-pro-drop (see, e.g., Toribio (1996), Ordóñez and Olarrea (2008), and Kato (2012a)). The authors show that although the two languages exhibit common properties with respect to phenomena related to the loss of null subjects, Dominican Spanish does not allow “topic-subject” constructions, as illustrated in (6) below. Kato and Ordóñez also show that the grammatical versions of (6) in Dominican Spanish involve CLLD with a third-person dative clitic, instead of “topic-subjects,” as can be seen in (7).

(6) Dominican Spanish (Kato and Ordóñez 2019: (27) and (25))

a. * El reloj rompió las agujas.
the clock broke.3SG the needles

Intended: “The hands of the clock broke.”

b. * Estos bosques llueven mucho.
these forests rain.3PL a.lot

Intended: “It rains a lot in the forests.”

(7) Dominican Spanish (Kato and Ordóñez 2019: (30) and (28))

a. A este reloj se le rompió la aguja.
to this clock REFL DAT.3SG broke.3SG the needle
“The clock’s hand broke.”

b. A estos bosques les llueve mucho.
to these forests DAT.PL rain.3SG a.lot

“In these forests, it rains a lot.”

Table 1 Third-person accusative and dative clitics and possessive pronouns in BP in the eighteenth and twentieth centuries

	Third-person singular			Third-person plural		
	Acc clitic	Dat clitic	Poss pronoun	Acc clitic	Dat clitic	Poss pronoun
Eighteenth-century BP	o (MASC) a (FEM)	lhe	seu (MASC.SG) sua (FEM.SG) seus (MASC.PL) suas (FEM.PL)	os (MASC) as (FEM)	lhes	seu (MASC.SG) sua (FEM.SG) seus (MASC.PL) suas (FEM.PL)
Twentieth-century BP	–	–	–	–	–	–

Interestingly, as Kato and Ordóñez observe, BP has lost its series of third-person pronominal clitics, as well as its third-person possessive pronouns, as illustrated in the chart above, adapted from Kato (1993b) (Table 1).⁷

This simplification in the pronominal paradigm of BP led the authors to conclude that the real trigger for the emergence of “topic-subject” constructions in BP was the impoverishment in its clitic and possessive systems, and not in its verbal agreement paradigm. Based on the work by Barros (2006) and Torres Morais (2016), Kato and Ordóñez propose that until the nineteenth century, the functional skeleton of the extended projection of ν P in BP included a dative phrase (in the sense of Landau (1999)), whose specifier was licensed with dative Case and its head was realized as a dative clitic, yielding CLLD constructions analogous to (7). Once third-person dative clitics became null, their associates could no longer be licensed with dative Case and had to move to [Spec,TP] to have their Case licensed, yielding “topic-subject” constructions.

In this chapter we assume the gist of Kato and Ordóñez’s analysis, modifying some details of its technical implementation. In particular, we show below that the standard assumption that dative clitics are θ -marked in their argument position before undergoing cliticization suffices to account for the facts under discussion and we provide a novel analysis of “topic-subject” constructions with locatives, as they do not seem to have a clitic structure as their diachronic source.⁸ We follow Barros (2006), Torres Morais and Salles (2016), Gonçalves and Miguel (2019), and

⁷ For relevant discussion, see, e.g., Omena (1978), Tarallo (1983), Oliveira e Silva (1984), Duarte (1986), Galves (1989), Cerqueira (1993), Kato (1993b), Nunes (1993), Cyrino (1997), Torres Morais (2007), Torres Morais and Berlinck (2006), and Torres Morais and Salles (2010, 2016).

⁸ Old Portuguese had a locative clitic *hi* “there,” which could be doubled by a locative expression, but this clitic was lost a couple of centuries before the emergence of “topic-subject” constructions and the loss of third-person dative clitics and possessive pronouns in BP. For relevant discussion on the disappearance of locative clitics, see Castilho (2012).

Kato and Ordóñez (2019) in assuming that until the nineteenth century, the extended projection of vP in BP included a projection that was able to assign dative Case, yielding external possessor constructions in the sense of Vergnaud and Zubizarreta (1992), as illustrated by the sentences in (8) below, by the nineteenth-century author Machado de Assis. In (8a), the dative element in bold is an R-expression and in (8b), a clitic pronoun.

- (8) (Nineteenth-century BP; Torres Morais and Salles 2016: (33b–c))
- a. Clara não tinha sequer tempo de remendar a roupa
 Clara not had hardly time to mend the clothes
ao marido.
 the husband.DAT
 “Clara hardly had time to mend her husband’s clothes.”
- b. Clara estendeu a mão ao marido como a amparar-**lhe**
 Clara extended her hand the husband.DAT as to give-3SG.DAT
 o ânimo.
 the support
 “Clara extended her hand to her husband, as a way to give him support.”

Once vP became unable to assign dative Case in BP, external possessor constructions with an R-expression such as (8a) simply died out in the grammar. Let us then consider how DP_1 in the abstract transitive vP structure depicted in (9) below can be licensed, after BP lost the dative licensing projection at the vP level, as well as its third-person dative clitics and third-person possessive pronouns.

- (9) [TP T [vP DP₃ [v [vP V [DP_2 D . . . [NP N DP₁]]]]]]

In (9), DP_2 is arguably Case-licensed by v , and DP_3 by T. Given that the extended projection of v is no longer able to license dative Case in BP, the only possibility for DP_1 to have its Case licensed is to receive inherent Case by the noun that θ -marks it (see Chomsky 1986). If the inherently Case-marked DP_1 in (9) is a first- or second-person pronoun, it may be realized as a possessive pronoun, as illustrated in (10a) below, or as a dative clitic, as illustrated in (10b).

- (10) a. A Maria segurou [a [{**minha/sua**]_i [mão t_i]]
 the Maria held the my/your hand
- b. A Maria {**me/te**]_i segurou [a [mão t_i]]
 the Maria me.DAT/you.DAT held the hand
 “Maria held my hand.”

On the other hand, if the inherently Case-marked DP₁ in (9) is a third-person pronoun, neither possibility is available, as shown in (11a) and (11b) below, for BP has lost the genitive and dative forms for third-person pronouns. A third-person pronoun occupying the position of DP₁ in (9) may, however, be realized preceded by the dummy preposition *de*, as shown in (11c), a possibility that is also available for R-expressions, as shown in (11d), but not for first- and second-person pronouns, as shown in (11e).

- (11) a. * [A Maria]_i segurou a **sua**_k mão.
 the Maria held the his hand
- b. * A Maria **lhe** segurou a mão.
 the Maria him.DAT held the hand
- c. A Maria segurou a mão **dele**.
 the Maria held the hand of-him
 “Maria held his hand.”
- d. A Maria segurou a mão **do Pedro**.
 the Maria held the hand of-the Pedro
 “Maria held Pedro’s hand.”
- e. * A Maria segurou a mão {**de mim/de você**}.
 the Maria held the hand of me/of you
 “Maria held {my/your} hand.”

The contrast between (11c,d) and (11e) may be described as showing that *de* is only allowed if the expression it licenses does not have an independent form for the realization of inherent Case (genitive or dative, in this particular scenario). In other words, *de*-insertion is a last resort strategy for the realization of the inherent Case assigned by the noun in (9).

Notice that the asymmetries seen in (10) and (11) between first and second persons, on the one hand, and third person, on the other, were handled based solely on the Case properties and Case realizations within DP₂ in (9), in a way quite independent from the properties of *v*. This leads us to expect that these asymmetries should not be restricted to transitive verbs, as in (9), but could also be found with unaccusative verbs. Kato and Ordóñez (2019) show that this is indeed the case. Take the monoargumental unaccusative structure in (12) below, for instance. Like what we saw above, DP₁ in (12) is assigned inherent Case and is realized as a possessive pronoun or a dative clitic if it is a first- or second-person pronoun (see (13a–b)) or by an oblique form preceded by the preposition *de* if it is a third-person pronoun (see (13c)).

- (12) [TP T [_{VP} V [_{DP2} D ... [_{NP} N DP₁]]]]

- (13) a. Ainda não nasceu a {minha/sua} barba.
 still not was.born the my/your beard
- b. Ainda não {me/te} nasceu a barba.
 still not me.DAT/you.DAT was.born the beard
 “I/you have not grown a beard yet.”
- c. Ainda não nasceu a barba **dele**.
 still not was.born the beard of-him
 “He has not grown a beard yet.”

Given that unaccusative structures do not involve an external argument, DP₁ in (12) may also be Case-licensed in BP by moving to [Spec,TP], yielding “topic-subject” constructions. Interestingly, “topic-subject” constructions also display sensitivity with respect to person, with third-person pronouns (and R-expressions) yielding the best outputs, as shown in (14).⁹

- (14) a. %* Eu ainda não nasci a barba.
 I still not was.born.1SG the beard
 “I have not grown a beard yet.”
- b. %?? Você ainda não nasceu a barba.
 you still not was.born the beard
 “You have not grown a beard yet.”
- c. {Ele/o João} ainda não nasceu a barba.
 he/the João still not was.born the beard
 “{He/João} has not grown a beard yet.”

A similar state of affairs is found with unaccusative structures with two internal arguments, as sketched in (15) below. Given that unaccusative verbs are not associated with structural Case (Burzio’s Generalization), at least one of the arguments of V in (15) must be assigned inherent Case. Suppose that V assigns inherent dative Case to DP₁. DP₁ then surfaces as a dative clitic if it is a first or second person, but as an oblique form preceded by the preposition *a* if it is a third-person pronoun, as shown in (16). (15) may also give rise to a “topic-subject” construction, again displaying person sensitivity, with third-person pronouns and R-expressions being the best results, as shown in (17).

⁹ The amelioration effect with the pronoun *você* in (14b) (and in (17b) and (39b) below) is undoubtedly related to the fact that this pronoun, as well as its plural counterpart *vocês*, triggers third-person agreement, despite being a second-person pronoun from a semantic point of view. It is thus unsurprising that speakers display more variation in their judgments when “topic-subject” constructions involve the second-person pronouns *você* and *vocês*. Due to space considerations, we will put further discussion of this variation aside.

- (15) [TP T [_{VP} [_{VP} DP₂ V DP₁]]]
- (16) a. Me/te faltou sorte.
 me.DAT/you.DAT lacked luck
 “I was/you were unlucky.”
 b. Faltou sorte a ele.
 lacked luck to he
 “He was unlucky.”
- (17) a. * Eu faltei sorte.
 I lacked.1SG luck
 “I was unlucky.”
 b. %?? Você faltou sorte.
 you lacked luck
 “You were unlucky.”
 c. Ele faltou sorte.
 he lacked luck
 “He was unlucky.”
 d. Meus times faltaram sorte.
 my teams lacked.3PL luck
 “My teams were unlucky.”

The data in (14) and (17) seem to show that inherent Case assignment by N and V is obligatory when the relevant DP is a first- or second-person pronoun, thus blocking its movement to [Spec,TP] (see (14a,b)/(17a,b)), but optional when the relevant DP is a third-person expression, optionally allowing it to move to [Spec,TP], yielding a “topic-subject” construction (see (14c)/(17c,d)). This conceptually odd result may however be disentangled if we observe that the relevant difference is likely to be related to the loss of third-person possessive pronouns and dative clitics in BP. In other words, inherent Case assignment may be taken to apply obligatorily in the cases discussed above and the additional possibility available to third-person expressions may have to do with the realization of inherent Case. Under the standard assumption that an inherent Case is linked to a θ -role, we tentatively propose that if the target of inherent Case assignment cannot morphologically realize the Case it has received, two different repair strategies can be employed (with equal derivational cost): (i) the expression is realized with default Case morphology and a linker is added to encode the dependency relation with respect to the θ -marking head, or (ii) the expression “repels” the Case associated with the θ -role in the sense that it does not incorporate the Case morphology/specification associated with the θ -role, thus remaining active for the purposes of Case and agreement.

The possibility in (i) is illustrated by sentences such as (11c), (13c), and (16b). Crucially, with the loss of third-person dative and accusative clitics and third-person

possessive pronouns in BP, the nominative form came to be used and licensed in all syntactic positions, for nominative is the default Case in BP. In other words, what we described in (11c), (13c), and (16b) as a third-person pronoun in its oblique form seems to be more adequately described as the default form preceded by a linker (the same applies to the R-expression in (11d), for instance). As for the possibility (ii), it is exemplified by “topic-subject” constructions such as (14c) and (17c,d), where the argument of N in (14c) and V in (17c,d) repels the “unrealizable” Case assigned and moves to [Spec,TP], where it gets licensed with nominative Case in the standard way. Importantly, a similar derivation is not available to first- and second-person pronouns (see (14a,b) and (17a,b)) because the last resort saving strategy is not applicable, for the inherent Case assigned can be morphologically realized.

Assuming that something along these lines may be on the right track, let us now examine some details of the derivation of “topic-subject” constructions.

4 The Role of Inherent Case in “Topic-Subject” Constructions

We saw in Sect. 3 that the loss of the projection licensing dative Case at the vP level, coupled with the loss of third-person dative clitics and third-person possessive pronouns, considerably reduced the possibilities for Case licensing within vP in BP. However, given that we are talking about the vP level, structural Case is not the only possibility for Case licensing. Crucially, vP is also a thematic domain and, therefore, a domain where inherent Case may be available. Our proposal is that with the reduction of structural Case assignment possibilities at the vP level, BP came to make pervasive use of inherent Case within the verbal domain, extending it to the adjectival and nominal domains, as well.

This innovation developed is interesting in that it provides a single solution for the two general puzzles posed by “topic-subject” constructions in BP: (i) how the relevant DPs are Case-licensed and (ii) why the relevant A-movement involved in these constructions does not violate minimality (see Nunes (2017)). Let us consider each of them in turn.

Take the alternation in (18) and (19), for example.

- (18) a. Quebrou o ponteiro dos relógios.
 broke-3SG the arm of-the watches
- a'. [TP_{pro_{expl}} [_{vP} [_{VP} quebrar [_{DP} o [_{NP} ponteiro [_{DP} os relógios]]]]]]
- |_____↑inherent Case |_____↑inherent Case
- b. [os relógios] quebraram o ponteiro.
 the watches broke-3PL the arm
 “The arms of the watches broke.”
- b'. [TP [os relógios]_i T [_{vP} [_{VP} quebrar [_{DP} o [_{NP} ponteiro *t*_i]]]]]]
- |_____↑inherent Case

- (19) a. Cabe muita coisa nessas gavetas.
fit-3SG many thing in-these drawers
- a'. [TP_{pro_{expl}} T [_vPV [_vP [muita coisa] [cabe [_{DP} essas gavetas]]]]]
inherent Case ↑ ____ || ____ ↑ inherent Case
- b. [essas gavetas] cabem muita coisa.
these drawers fit-3PL many thing
“Many things can fit in these drawers.”
- b'. [TP [_{DP} essas gavetas] T [_vPV [_vP [muita coisa] cabem *t*_i]]]
inherent Case ↑ ____ |

In (18a'), the verb *quebrar* and the noun *ponteiro* both assign inherent Case to their arguments, which are then licensed in situ. The inherent Case assigned by the noun is morphologically realized with the help of the preposition *de* (see Sect. 3). Recall that a given expression may also have the option of repelling a given inherent Case if it cannot morphologically realize it (see Sect. 3). This option is exercised in (18b'), where the DP *os relógios* “the watches” repels the inherent Case assigned by *ponteiro* “arm” and then moves to [Spec,TP], where it triggers verbal agreement and is assigned nominative Case. As for (19a), the verb *cabem* “fit” assigns inherent Case to both of its arguments, which surface in situ, and the inherent Case assigned to the locative is realized with the help of the preposition *em* “in.” If the complement exercises the option of repelling the inherent Case assigned by *cabem*, as sketched in (19b'), it may get licensed by moving to [Spec,TP], triggering verbal agreement and receiving nominative Case in the standard way (cf. (19b)).

One could say that the absence of *em* in (19b) is not a strong argument for taking *em* in (19a) as a marker of inherent Case, for the verb *cabem* “fit” selects a location for its complement and *em* can be used as a true preposition with a locative meaning. There is however independent evidence that the preposition *em* in BP may be used as a realization of inherent Case in contexts unrelated to a locative meaning. Discussing the role of *em* in marking the complement of verbal nominalizations with the suffix *-ada* in BP, as shown in (20) below, Scher (2004:198) points out that *em* cannot be inserted to Case-mark the subject of a small clause, as illustrated by the contrast in (21). Given that *os papeis* “the papers” is an argument of the verb in (21a) but not in (21b), Scher interprets the contrast in (21), showing that *em* in these contexts is a realization of inherent Case.¹⁰

¹⁰ Andrade and Galves (2014) propose that “topic-subject” constructions such as (ia) and (iia) below are launched from the structures in (ib) and (iib), respectively, where R is a relator in the sense of den Dikken (2006) and P is a null preposition that gets incorporated into R.

(20) A Maria deu uma organizada nos dados.
 the Maria gave a organize-NOMZ in-the data
 “Maria has organized the data.”

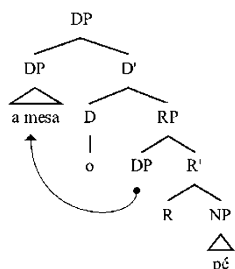
(21) a. O João deu uma classificada nos papeis.
 the João gave a classify-NOMZ in-the papers
 “João has classified the papers.”

b. * O João deu uma classificada nos papeis como
 the João gave a classify-NOMZ in-the papers as
 interessantes.
 interesting
 “João has classified the papers as interesting.”

(i) Andrade and Galves (2014: 118, 137):

a. A mesa quebrou o pé.
 the table break-PAST-3SG the foot
 “The table leg broke.”

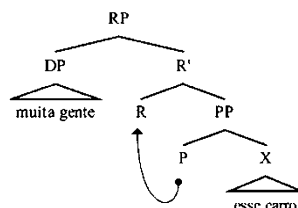
b.



(ii) Andrade and Galves (2014: 120, 138):

a. Esse carro cabe muita gente.
 this car fit-3SG many people
 “Many people fit in this car.”

b.



Addressing the issue of the Case of the postverbal DP in “topic-subject” constructions, Andrade and Galves (2014: fn. 5) suggest in passing that “the V/Root valued this element with inherent Case, much in spirit of Belletti (1988).” It should be noted, though, that in Andrade and Galves’s analysis, the postverbal DP in “topic-subject” constructions such as (19b) are located in the specifier of RP, which is the actual complement of the verb (cf. (iib)). However, as argued by Chomsky (1986) and Belletti (1988), one does not find the analogue of ECM when θ -relations are involved, that is, a given head may assign Case to the specifier of its complement, but not a θ -role (see the contrast in (21), for instance). Given the necessary association between inherent Case and θ -role (see Chomsky 1986), it appears that Andrade and Galves’s analysis must resort to another sort of Case licensing for “topic-subject” constructions like (iia) (see Nunes (2022) for additional problems and further discussion).

The discussion to be presented below in the text is framed under the standard assumption that a given head H may only assign θ -roles/inherent Case to elements in its minimal domain (i.e., elements with which H or a projection of H has merged). For the purpose of exposition, we assume Chomsky’s (1986) original coarse-grained distinction between structural (non- θ -related) and inherent (θ -related) Case, putting aside Woolford’s (2006) refinements, according to which nonstructural Case should be divided in two subclasses: inherent Case, which is predictably associated with specific θ -roles, and lexical Case, which is idiosyncratically linked to specific lexical items.

Also relevant to the present discussion is the fact that directional verbs in BP underwent a diachronic change replacing the preposition *a* “to” by the preposition *em* “in” as the head of their complements (see, e.g., Wiedemer 2013), as illustrated in (22) below. This change seems to have connected the double role played by *em* as an independent true preposition with locative meaning and as a marker of inherent Case assignment (see (21)), according well with our proposal that BP underwent a diachronic change that greatly expanded the use of inherent Case in its grammar.

- (22) a. O João foi **no** mercado.
 the João went in-the market
 “João went to the market.”
- b. A Maria já chegou **em** casa.
 the Maria already arrived in house
 “Maria has already arrived home.”
- c. O Pedro veio **na** festa.
 the Pedro came in-the party
 “Pedro came to the party.”
- d. A Maria levou o filho **no** cinema hoje.
 the Maria took the son in-the movies today
 “Maria took her son to the movies today.”

Let us now consider the connection between inherent Case and minimality, by examining the English data in (23).

- (23) a.* [Mary seems to **him_k** [*t* to like **John_k**]]
 b. [**Mary_i** seems to him [*t_i* to be nice]]
 c.* [**To him**] seems *t* [Mary to be nice]

The sentence in (23a) displays a Principle C effect, suggesting that the preposition does not prevent the pronoun from c-commanding into the embedded clause. That being so, the fact that it does not block movement of the embedded subject on (23b) becomes rather puzzling. Chomsky (1995:306) observes that the experiencer should get inherent Case within the lexical VP shell and suggests (fn. 77) that the preposition *to* is adjoined to the experiencer DP as a kind of Case-marker, explaining why it does not affect c-command. Building on these premises, Nunes (2008a) proposes that inherent Case renders a given element inert for the purposes of A-movement, as shown in (23c), removing it from computations of Relativized Minimality.¹¹ To put it in different terms, inherent Case makes an element transparent for A-movement across it. This is exactly what happens in

¹¹ To be precise, inherent Case may be assigned alone or in association with structural Case (for instances of *quirky* Case assignment, see, e.g., Zaenen et al. (1985)). Only when it is assigned in

(18b') and (19b'). The DPs *o ponteiro* “the arm” in (18b') and *muita coisa* “many things” in (19b') do not block A-movement of the “topic-subject,” for they have become transparent after receiving inherent Case.¹²

As shown in Nunes (2017), this proposal also provides a straightforward account of “mixed” and “extralong” “topic-subject” constructions such as (24) and (25).

- (24) a. [Esses porta-malas] **cabem** muita coisa na lateral.
 these car-trunks fit-3PL many thing in-the lateral
 “Many things can fit on the side of the trunk of these cars.”
 b. [TP [DP esses porta-malas]_i T [_vPV [_{VP} [DP muita coisa] [cabem [DP a lateral t_i]]]]]
- (25) a. [Esses barcos] **diminuíram** o tamanho da hélice do motor.
 these boats diminished-3PL the size of-the fan of-the engine
 “These boats had the size of the fans of their engine reduced.”
 b. [TP [DP esses barcos]_i T [_vPV [_{VP} diminuíram [DP o tamanho [DP a hélice [DP o motor t_i]]]]]]]
 |—↑inherent Case |—↑inherent Case |—↑inherent Case

From the perspective taken here, apparently complex structures like the ones in (24) and (25) receive the same analysis as (18b) and (19b): given that the potential interveners receive inherent Case, they become transparent for A-movement across them, as sketched in (24b) and (25b), in the same way the pronoun in (23b) does not block A-movement of the embedded subject. In (24b), the verb *cabem* assigns inherent Case to both the theme in its specifier and the locative in its complement (which is realized as *em* in the morphological component) and they become inert for the purposes of intervention. The argument of *lateral*, which has exercised the option of repelling inherent Case, can then move to [Spec,TP] to get Case-licensed without incurring in a minimality violation. As for (25b), the verb *diminuir*

isolation does it render an element inert for A-movement purposes. See Sect. 5 below for further discussion of quirky Case.

¹² Janayna Carvalho (p.c.) brought to our attention the relevance of constructions such as (ia) below (see, e.g., Meireles and Caçado (2020)), which are also found in European Portuguese. From the perspective of our proposal, sentences such as (ia) are derived along the lines of (ib), where the main verb assigns inherent Case to its DP complement, whose head is then realized with the preposition *em* “in,” and *v* agrees with *o João* in the specifier of DP₂, valuing its Case as accusative. Crucially, DP₁ does not induce a minimality violation for the agreement relation between *v* and DP₂ because it has received inherent Case.

- (i) a. Ela beijou o João no rosto.
 she kissed the João on-the face
 “She kissed João on the cheek.”
- b. [_vPV [_{VP} beijou [DP₁ [DP₂ o João]_i [D' o rosto t_i]]]]]
 |—↑structural Case

“diminish” and the nouns *tamanho* “size” and *hélice* “fan” assign inherent Case to their arguments and the DP *esses barcos* “these boats” repels the inherent Case assigned by the noun *motor* “engine.” *Esses barcos* then receives nominative Case after moving to [Spec,TP] and the inherent Case assigned by the nouns is realized as the preposition *de*, yielding the “topic-subject” construction in (25a), which involves an “extralong” instance of A-movement, but all the potential interveners have become inert after receiving inherent Case.

The amplification of the use of inherent Case in BP had widespread consequences in the grammar. Take the contrast in (26) below, for example.

- (26) a. [Esse livro]_i, todo mundo [que gosta Ø]_i vira pacifista.
 this book every world that likes becomes pacifist
 “Everybody who read this book becomes a pacifist.”
- b. Todo mundo gosta *(de) filme de detetive.
 every world likes of movie of detective
 “Everyone likes detective movies.”

The sentence in (26a) has a null object within a relative clause island, suggesting that it is a *pro* linked to the base-generated topic, rather than a trace. In turn, (26b) shows that the verb *gostar* “like” is not a Case assigner, requiring the preposition *de* in order to license its complement. That being so, one wonders how *pro* in (26a) is Case-licensed. To account for this sort of problem, Ferreira (2000) has proposed that *pro* in BP may be defective in not having a Case feature. However, Kato (2010, 2012b) shows that this cannot be what is behind the grammaticality of sentences such as (26a), for the apparently exceptional licensing of *pro* is lexically conditioned. The complement of the verb *rir* “laugh,” for instance, must also be preceded by the preposition *de*, as shown in (27). However, whereas *gostar* licenses a null object linked to a topic, *rir* doesn’t, as shown in (28).

- (27) a. Ela não gostou *(d)o palhaço.
 she not liked of-the clown
 “She didn’t like the clown.”
- b. Ela não riu *(d)o palhaço.
 she not laughed of-the clown
 “She didn’t laugh at the clown.”
- (28) a. [Aquele palhaço]_i, ela não gostou *pro*_i
 that clown she not liked
 “That clown, she didn’t like him.”
- b. * [Aquele palhaço]_i, ela não riu *pro*_i
 that clown she not laughed
 “That clown, she didn’t laugh at him.”

Kato and Nunes (2009) argue that the difference between these two verbs is that *rir* selects for a PP headed by *de*, whereas *gostar* assigns inherent Case to its complement, which is realized as *de* if the complement is phonetically realized. Interesting empirical evidence for this proposal is the fact that *gostar* licenses an inherently Case-marked reflexive/reciprocal clitic, but *rir* doesn't, as shown in (29) (see Nunes (2008b)).

- (29) a. Eles **se** gostam muito.
 they REFL.CL like much
 "They like each other a lot."
 b. * Eles **se** riram bastante.
 they REFL.CL laughed much
 "They laughed a lot at each other."

Kato and Nunes (2009) also show that the salient resort to inherent Case in the grammar of BP also accounts for the existence of what Tarallo (1983) called PP-chopping relatives in the language. In his seminal work on relative clauses in BP, Tarallo postulated three types of relativization strategies, each of which with a different sociolinguistic status: the standard strategy with movement of a PP, as illustrated in (30a) below; the resumptive strategy with an overt resumptive pronoun, as illustrated in (30b); and the chopping strategy, with no visible presence of the relevant PP, as illustrated in (30c).

- (30) a. a pessoa com quem eu conversei
 the person with who I talked
 b. a pessoa que eu conversei com ela
 the person that I talked with her
 c. a pessoa que eu conversei
 the person that I talked
 "the person I talked to"

For Tarallo, the chopping version in (30c) involves a null resumptive pronoun and the preposition is deleted in the phonological component, as BP does not allow preposition stranding. However, Kato (1993a) observes that if the chopping strategy involved preposition deletion because prepositions cannot be left stranded in BP, the counterpart of (30a) given in (31a) below should be grammatical under the derivation sketched in (31b), where the relative pronoun *quem* undergoes A'-movement and the stranded preposition is deleted in the phonological component.

- (31) a. * a pessoa quem eu conversei
 the person who I talked
 “the person I talked to”
 b. a pessoa [quem_i eu conversei e~~om~~ t_i]

Kato and Nunes (2009) reanalyze Tarallo’s proposal in terms of phonetic realization. According to them, the verb *conversar* “talk” assigns inherent Case to its complement, which is realized as the preposition *com* “with” if the argument is phonetically realized. This is transparent in the case of (30b), as the overt resumptive pronoun is realized in its thematic position. In (30a), the verb assigns inherent Case to the relative pronoun before it moves to [Spec,CP] and it then surfaces accompanied by the preposition. In (30c), the preposition is not phonetically realized because the complement of the verb (*pro*) has no phonetic realization. Finally, under the assumption that the relative pronoun *quem* cannot be base-generated in its surface position (see Kato and Nunes (2009) for arguments and relevant discussion), it must have merged with the verb before moving to the left periphery, and once it receives inherent Case from *conversar*, it must surface with the preposition, explaining why (31a) is not acceptable.

It is thus not surprising that verbs of movement that came to take the preposition *em* preceding their locative complement (see (22)) freely allow chopping relatives, as illustrated in (32) below. This is exactly what we should expect if *em* may be a realization of inherent Case, as proposed above.

- (32) a. o mercado que o João foi
 the market that the João went
 “the market João went to”
 b. o lugar que a Maria chegou
 the place that the Maria arrived
 “the place Maria arrived at”
 c. a festa que o Pedro veio
 the party that the Pedro came
 “the party Pedro came to”
 d. o cinema que a Maria levou o filho
 the movies that the Maria took the son
 “the movie theater Maria took her son to”

Kato and Nunes (2009) also observe that their reanalysis of Tarallo’s (1983) account of chopping relatives also extends to free relatives in BP. Lessa de Oliveira (2008) notes that free relatives in BP can be of the “chopping” variety, as illustrated in (33) below, whose embedded verbs require a PP complement, as shown in (34).

- (33) a. Eu vou visitar quem_i você simpatiza muito.
 I go visit who you sympathize much
 “I’m going to visit who you like a lot.”
- b. Eu encontrei o que você gosta.
 I found what you like
 “I found what you like.”
- (34) a. O João simpatiza *(com) a Maria.
 the João sympathizes with the Maria
 “João likes Maria.”
- b. O João gosta *(de) romances.
 the João likes of novels
 “João likes novels.”

If “chopping” free relatives like (33) also involve inherent Case assignment to an object *pro*, one would expect contrasts such as the one in (28) to be replicated in free relative clauses. Kato and Nunes (2009) show that this prediction is borne out, as shown in (35).

- (35) a. * O João sempre critica quem ele ri.
 the João always criticizes who he laughs
 “João always criticizes whoever he laughs at.”
- b. O João sempre critica quem ele gosta.
 the João always criticizes who he likes
 “João always criticizes whoever he likes.”

As seen above, *gostar* assigns inherent Case, but *rir* doesn’t. Hence, the embedded object position in (35b) can be licensed (if it is *pro*), but not the embedded object position in (35a) (regardless of whether it is *pro* or a trace).

Let us finally examine another domain where we can directly see the connection between inherent Case and A-minimality. Galves (1987) has observed that *tough*-predicates in BP display a very distinctive behavior. In addition to the standard interpretation as the object of the embedded predicate, the subject of a *tough*-construction such as (36), for instance, may also be interpreted as the embedded subject.

- (36) O João é difícil de elogiar.
 the João is difficult of praise-INF
Tough-interpretation: “It is hard to praise João.”
 Raising interpretation: “João rarely praises someone.”

Nunes (2008a) notes that the adjectival predicates that allow the crosslinguistic uncommon subject reading optionally take a preposition before its infinitival complement, as illustrated in (37a) and (38a) below. Interestingly, the optionality does not hold constant across different constructions. If the preposition is present, the infinitival cannot move to the matrix subject position, as shown in (37b) and (38b). Conversely, the embedded subject can only move to the matrix subject position if the preposition is present, as illustrated in (37c) and (38c).

- (37) a. É difícil [(d)esses jornalistas elogiarem alguém].
 is difficult of-these journalists praise-INF-3PL somebody
 “It is rare for these journalists to praise someone.”
- b. (*D)esses jornalistas elogiarem alguém é difícil.
 of-these journalists praise-INF-3PL somebody is difficult
 “For these journalists to praise someone is very rare.”
- c. [esses jornalistas]_i são difíceis *(de)_i elogiarem alguém.
 these journalists are difficult of praise-INF-3PL somebody
 “These journalists rarely praise someone.”
- (38) a. Não estava previsto (para) as aulas começarem amanhã.
 not was predicted for the classes start-INF-3PL tomorrow
- b. (*Para) as aulas começarem amanhã não estava previsto.
 for the classes start-INF-3PL tomorrow not was predicted
 “It was not expected that the classes should start tomorrow.”
- c. As aulas estavam previstas *(para) começarem amanhã.
 the classes were predicted for start-INF-3PL tomorrow
 “The classes were not expected to start tomorrow.”

Nunes (2008a) argues that these prepositions are actually realization of an inherent Case optionally assigned by the impersonal predicates to their infinitival complement, much like what Chomsky (1995) has proposed for the preposition *to* preceding the experiencer of raising constructions like (23). Like what we saw with the prepositioned experiencer in (23b), the prepositioned infinitivals in (37b) and (38b) cannot undergo A-movement because they have already been Case-marked. The inherent Case also renders the infinitival transparent for A-movement from within it, in the same way we saw in “topic-subject” constructions like (25); hence, A-movement from within the infinitival (37c) and (38c) can only take place if the preposition is present.

To summarize, with the weakening of structural Case licensing at the vP level in BP seen in Sect. 3, the grammar came to explore to its limits the other possibility for Case licensing, namely, inherent Case assignment. This expansion of the use of inherent Case reverberated across different domains in the grammar, yielding as byproducts apparent Caseless null objects, chopping relative clauses, hyper-raising constructions out of infinitivals, and topic-subject constructions.

5 Further Issues on Person Asymmetries and “Resumption” in “Topic-Subject” Constructions

As we saw in Sect. 3, Kato and Ordóñez (2019) observe that “topic-subject” constructions display a sensitivity to the value of the feature [person]. Whereas third-person pronouns (and R-expressions) function as good candidates as “topic-subjects,” first- and second-person pronouns in general do not yield well-formed results or exhibit considerable variation across speakers (see fn. 9), as seen in (14), repeated here in (39).

- (39) a. %* Eu ainda não nasci a barba.
 I still not was.born.1SG the beard
 “I have not grown a beard yet.”
- b. %?? Você ainda não nasceu a barba.
 you still not was.born the beard
 “You have not grown a beard yet.”
- c. {Ele/o João} ainda não nasceu a barba.
 he/the João still not was.born the beard
 “{He/João} has not grown a beard yet.”

This asymmetry seems to be lexically conditioned, though. In (40) below, for instance, there is no difference of acceptability among the different persons in the “topic-subject” versions.

- (40) a. Inchou o meu pé.
 swelled the my foot
- a'. Eu inchei o pé.
 I swelled.1SG the foot
 “My foot got swollen.”
- b. Inchou o seu pé.
 swelled the your foot
- b'. Você inchou o pé.
 you swelled the foot
 “Your foot got swollen.”
- c. Inchou o pé dele.
 swelled the foot of-he
- c'. Ele inchou o pé.
 he swelled the foot
 “His foot got swollen.”

Rodrigues (2020) in fact argues that possessor raising constructions in BP have a subclass in which the moving element moves directly to [Spec,TP] and another one in which the moving element lands in an intermediate position, where it receives the θ -role of affected entity. According to her, the contrast between the two sentences in (41) below can be accounted for if *esturricar* “burn” is a member of the second subclass, and accordingly, it requires that the moving element be affected, hence the pragmatic oddity of (41b), for the pig is dead and cannot be affected in the relevant sense. In turn, *cair* “fall” belongs to the first class and is not subject to this pragmatic restriction, for the possessor moves directly to [Spec,TP]; hence, (42) is acceptable even if the falling event occurs after Lincoln’s death.

(41) (Rodrigues 2020, glosses and translation added)

- a. Eu esturriquei o dedo na frigideira.
 I burnt the finger in-the frying.pan
 “I burned my finger on the frying pan.”
- b.# O porco esturricou a costela na frigideira.
 the pork burnt the ribs in-the frying.pan
 “The pork’s ribs got burned.”

(42) (Rodrigues 2020, glosses and translation added)

- O Lincoln caiu os dentes (depois de morto).
 the Lincoln fell the teeth after of dead
 “Lincoln’s teeth fell off (after he was dead).”

What is relevant for our purposes is that the subclasses identified by Rodrigues seem to correlate with the person asymmetry noted by Kato and Ordóñez (2019). Specifically, verbs that are not associated with an affectedness θ -role in Rodrigues’s sense are the ones that display person sensitivity. The verb *cair* “fall,” for example, admits possessor raising with an R-expression, as seen in (42), or a third-person pronoun, as shown in (43a), but not with a first-person pronoun, as shown in (43b). By contrast, verbs that arguably involve the affectedness θ -role such as *esturricar* “burn” and *inchar* “swell” do not display person restrictions, as seen in (41a) and (40).

- (43) a. Elas caíram o cabelo.
 they.F fell-3PL the hair
 “Their hair fell out.”
- b.%* Eu caí o cabelo.
 I fell-1SG the hair
 “My hair fell out.”

From an abstract point of view, the behavior of these two types of possessor raising in BP resembles the patterns of control and raising of DPs marked with quirky Case in Icelandic. The embedded main verbs of (44) below, for example, assign quirky dative to their complements. Dative morphology is preserved in standard raising constructions, as shown in (44a), but not in control constructions, as shown in (44b).

(44) *Icelandic*:

- a. Mönnunum/*Mennirnir virðist báðum hafa verið hjálpað.
 men.the.DAT/*NOM seems both.DAT have been helped.DFLT
 “The men seem to have both been helped.”
 (Sigurðsson 2008)
- b. Hann/*Honum vonast til að verða bjargað af fjallinu.
 he.NOM/*DAT hopes for to be rescued.DFLT of the.mountain
 “He hopes to be rescued from the mountain.”
- (Andrews 1990)

As is well known, quirky Case appears to involve a mixture of inherent and structural Case (see footnote 11). On the one hand, it behaves like inherent Case in establishing a connection between a specific θ -role and a specific piece of morphological information; on the other hand, it behaves like structural Case in its need to be licensed by a ϕ -complete probe. Assuming the Movement Theory of Control, Boeckx et al. (2010a, b) accounted for contrasts like the one in (44), by assuming that the additional θ -role assignment present in control breaks the connection between morphology and θ -role in inherent Case. The derivation of (44b), for example, proceeds along the lines sketched in (45) (with English words for convenience).

- (45) a. *Assignment of quirky Case*:
 [rescued **he**_[θ 1-DAT] from the mountain]
- b. *Movement to [Spec,TP]*:
 [_{TP}**he**_[θ 1-DAT] to be rescued *t* from the mountain]
- c. *Applications of Merge*:
 [_{vP}**v**_[θ 2] [_{VP} hopes [_{CP} C [_{TP}**he**_[θ 1-DAT] to be rescued *t* from the mountain]]]]]
- d. *Movement and θ -assignment*:
 [_{vP}**he**_[θ 2+ θ 1] [_{v'} [_{VP} hopes [_{CP} C [_{TP}*t* to be rescued *t* from the mountain]]]]]]]
- e. *Movement to [Spec,TP] and nominative Case assignment*:
 [_{TP}**he**_{[θ 2+ θ 1]-Case:NOM} [_{T'} T [_{vP} [_{v'} [_{VP} hopes [_{CP} C [_{TP}*t* to be rescued *t* from the mountain]]]]]]]]]

In (45a) the verb assigns quirky dative Case to the pronoun, which then moves to the embedded TP, yielding (45b). After the matrix *v* enters the derivation in (45c), the embedded subject moves to [Spec,*vP*] and receives an additional θ -role, as shown in (45d). Boeckx, Hornstein, and Nunes argue that the assignment of an additional θ -role in (45d) ends up breaking the previously established connection between θ_1 and dative Case, for there are two θ -roles associated with the pronoun.

- (48) a. Eu inchei o (meu) pé.
I swelled.1SG the my foot
“My foot got swollen.”
- b. Eu arranhei o (meu) braço.
I scratched.1SG the my arm
“My arm got scratched.”
- c. Ele quebrou o braço (dele) no jogo.
he broke.3SG the arm of-him in-the game
“He broke his arms during the game.”
- (49) a. Eles ainda não nasceram a barba (*deles).
they still not were.born the beard of-they
“They haven’t grown a beard yet.”
- b. Os bebês cresceram o cabelo (*deles).
the babies grew.3PL the hair of-they
“The babies’ hair has grown.”

The class of verbs that do not display person asymmetries optionally allow the realization of a possessor within the internal argument, as seen in (48), whereas the class of verbs that require that “topic-subjects” be third person do not, as seen in (49). Notice that the verbs in (48) belong to the subclass that has an additional affectedness θ -role, but not the ones in (49). This entails that if the possessor gets independently licensed within the internal argument in (48), another element could in principle be independently merged in the specifier of ν P and be assigned the affected θ -role, yielding the versions of (48) with the possessive pronoun overtly realized. This possibility is not available in (49) even if the possessor is realized with inherent Case, for there is no additional θ -role to be assigned to the “topic-subject.”

This correlation between person sensitivity and “resumption” also seems to account for variation among speakers with respect to specific lexical items. Take the verb *furar* “puncture,” for example. Sentences such as (50a) below, with the “topic-subject” being an R-expression, are uniformly judged well formed by BP speakers, whereas sentences such as (50b) (under the relevant non-agentive “topic-subject” reading) find variation among speakers (the first author, for instance, admits them, but the second author doesn’t). Interestingly, speakers who allow (50b) also allow the corresponding sentences with a possessive pronoun in (51). This can be captured if the difference is reduced to whether or not the grammar of individual speakers encodes the unaccusative verb *furar* with an additional affectedness θ -role. If it does, all the sentences in (50) and (51) will be allowed; if it doesn’t, only (50a) is permitted.

- (50) a. O carro furou o pneu ontem.
 the car punctured the tire yesterday
 “The car had a flat tire yesterday.”
- b. % Eu furei o pneu ontem.
 I punctured-1SG the tire yesterday
 “I had a flat tire yesterday.”
- (51) a. % Eu furei o **meu** pneu ontem.
 I punctured-1SG the my tire yesterday
 “I had a flat tire yesterday.”
- b. % Eu furei o pneu do **meu** carro ontem.
 I punctured-1SG the tire of-the my car yesterday
 “My car had a flat tire yesterday.”

The data in (48)–(51) thus provide independent support to both Rodrigues’s (2020) distinction between two subclasses of possessor raising constructions in BP and our account of the person asymmetries observed by Kato and Ordóñez (2019).

6 Conclusion

“Topic-subject” constructions have received a lot of attention in the syntactic literature on BP. Details aside, the prevailing view is that the emergence of this type of construction in the grammar of BP is somehow related to the fact that it is no longer a canonical pro-drop language. Following Kato and Ordóñez (2019), we argued in this chapter that although BP’s becoming a non-pro-drop language has surely contributed to this innovation in the grammar, it is certainly not its primary cause. Developing Kato and Ordóñez’s insight, we proposed that with the weakening of structural Case licensing at the ν P level and the loss of third-person dative clitics and third-person possessive pronouns, BP came to make extensive use of inherent Case. This ended up amplifying the possibilities for a Caseless DP to move to [Spec,TP], trigger verbal agreement, and receive nominative Case, as potential interveners have been rendered inert for minimality computations after receiving inherent Case (see Chomsky (1995) and Nunes (2017)). The general conclusion is that “topic-subject” constructions in BP are in fact a byproduct of changes at the ν P level and its pronominal system.

Although there remain technical details to be further worked out, we believe that a major achievement of our proposal is that it provides a unified analysis for a series of distinct phenomena in BP that are apparently unrelated to “topic-subject” constructions such as the change of the preposition associated with directional verbs, the pervasive use of chopping relatives, and the emergence of hyper-raising out of infinitivals licensed by prepositions.

References

- Andrews, Avery. 1990. Case structures and control in Modern Icelandic. In *Modern Icelandic syntax*, ed. J. Maling and A. Zaenen, 187–234. San Diego, CA: Academic Press.
- Avelar, Juanito, and Charlotte Galves. 2011. Tópico e concordância em português brasileiro e português europeu. In *Textos Selecionados, XXVI Encontro da Associação Portuguesa de Linguística*, ed. A. Costa, I. Falé, and P. Barbosa, 49–65. Lisbon: APL.
- Barros, Evângela. 2006. *Construções de posse com clítico no PB: Percurso diacrônico*. Doctoral dissertation, Universidade Federal de Minas Gerais.
- Bastos-Gee, Ana. 2011. *Information structure within the traditional nominal phrase: The case of Brazilian Portuguese*. Doctoral dissertation, University of Connecticut.
- Belletti, Adriana. 1988. The case of unaccusatives. *Linguistic Inquiry* 19: 1–34.
- Boeckx, Cedric, Norbert Hornstein, and Jairo Nunes. 2010a. *Control as movement*. Cambridge: Cambridge University Press.
- . 2010b. Icelandic control really is A-movement: Reply to Bobaljik and Landau. *Linguistic Inquiry* 4: 111–130.
- Castilho, Célia. 2012. *Fundamentos sintáticos do português brasileiro*. São Paulo: Contexto.
- Carqueira, Vicente. 1993. A forma genitiva *dele* e a categoria de concordância (AGR) no português brasileiro. In *Português brasileiro: Uma viagem diacrônica*, ed. I. Roberts and M.A. Kato, 129–161. Editora da UNICAMP: Campinas.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin and use*. New York: Praeger.
- . 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- . 2000. Minimalist inquiries: The framework. In *Step by step: Essays on Minimalist Syntax in honor of Howard Lasnik*, ed. R. Martin, D. Michaels, and J. Uriagereka, 81–155. Cambridge, MA: MIT Press.
- . 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Cinque, Guglielmo. 1990. *Types of A-bar dependencies*. Cambridge, MA: MIT Press.
- Cyrino, Sonia. 1997. *O objeto nulo no português do Brasil: Um estudo sintático-diacrônico*. Londrina: Editora da UEL.
- de Andrade, Aroldo, and Charlotte Galves. 2014. A unified analysis for subject topics in Brazilian Portuguese. *Journal of Portuguese Linguistics* 13: 117–147.
- den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, MA: MIT Press.
- Duarte, Maria Eugênia. 1986. *Variação e sintaxe: Clítico acusativo, pronomes lexical e categoria vazia no português do Brasil*. Master's thesis, Pontifícia Universidade Católica de São Paulo.
- Duarte, Maria Eugênia and Mary A. Kato. 2008. A mudança paramétrica e a orientação para o discurso. Paper presented at the XXIV Encontro Nacional da Associação Portuguesa de Linguística, Braga.
- Ferreira, Marcelo. 2000. *Argumentos nulos em português brasileiro*. M.A. thesis, Universidade Estadual de Campinas.
- Galves, Charlotte. 1987. A sintaxe do português brasileiro. *Ensaios de Linguística* 13: 31–50.
- . 1989. O objeto nulo no português brasileiro: Percurso de uma pesquisa. *Cadernos de Estudos Linguísticos* 17: 65–90.
- . 1998. Tópicos, sujeitos, pronomes e concordância no português brasileiro. *Cadernos de Estudos Linguísticos* 34: 7–21.
- Gonçalves, Anabela, and Matilde Miguel. 2019. Porque é que os relógios não quebram os ponteiros em português europeu. In *Estudos linguísticos e filológicos oferecidos a Ivo Castro*, ed. E. Carrilho, A.M. Martins, S. Pereira, and J. Silvestre, 713–738. Lisboa: Centro de Linguística da Universidade de Lisboa.

- Kato, Mary A. 1989. Tópico e sujeito: Duas categorias em sintaxe? *Cadernos de Estudos Linguísticos* 17: 109–132.
- . 1993a. Recontando a história das relativas. In *Português brasileiro: Uma viagem diacrônica*, ed. I. Roberts and M.A. Kato, 223–261. Editora da Unicamp: Campinas.
- . 1993b. The distribution of pronouns and null elements in object position in Brazilian Portuguese. In *Linguistic perspectives on Romance languages: Selected papers from the XXI Linguistic Symposium of Romance Languages*, ed. W.J. Ashby et al., 225–235. Amsterdam/Philadelphia: John Benjamins.
- . 2010. Optional prepositions in Brazilian Portuguese. In *Romance Linguistics 2008: Interactions in Romance*, ed. K. Arregi, Z. Fagyal, S. Montrul, and A. Tremblay, 171–184. John Benjamins: Amsterdam/Philadelphia.
- . 2012a. Brazilian Portuguese and Caribbean Spanish: Similar changes in *Romania Nova*. In *Parameter theory and linguistic change*, ed. C. Galves, S. Cyrino, R. Lopes, M.F. Sândalo, and J. Avelar. Oxford: Oxford University Press.
- . 2012b. Caso inerente, Caso “default” e ausência de preposições. In *Por amor à linguística: Miscelânea de estudos linguísticos dedicados à Maria Denilda Moura*, ed. A. Sedrins, A. de Castilho, M. Sibaldo, and R. de Lima, 86–99. Edufal: Maceió.
- Kato, Mary A., and Jairo Nunes. 2009. A uniform raising analysis for standard and nonstandard relative clauses. In *Minimalist essays on Brazilian Portuguese Syntax*, ed. J. Nunes, 93–120. John Benjamins: Amsterdam/Philadelphia.
- Kato, Mary A., and Francisco Ordóñez. 2019. Topic-subjects in Brazilian Portuguese and clitic left dislocation in Dominican Spanish: Clitics and null subjects. *Syntax* 22: 229–247.
- Kuroda, Shige-Yuki. 1979. *Generative grammatical studies in the Japanese language*. New York: Garland.
- Lacerda, Renato. 2020. *Middle-field syntax and information structure in Brazilian Portuguese*. Doctoral dissertation, University of Connecticut.
- Landau, Idan. 1999. Possessor raising and the structure of VP. *Lingua* 107: 1–37.
- Lessa de Oliveira, Adriana. 2008. *As sentenças relativas em português brasileiro: Aspectos sintáticos e fatos de aquisição*. Doctoral dissertation, Universidade Estadual de Campinas.
- Lunguinho, Marcus. 2006. Partição de constituintes no português brasileiro: características sintáticas. In *Língua, gramática e discurso*, ed. D.E. Silva, 133–147. Cânone: Goiânia.
- Meireles, Letícia, and Márcia Cançado. 2020. A alternância parte-todo com verbos transitivos no PB: Um caso de fatoração de argumento. *D.E.L.T.A.* 36 (1): 2020360103.
- Modesto, Marcello. 2008. Topic prominence and null subjects. In *The limits of syntactic variation*, ed. Theresa Biberauer, 375–406. John Benjamins: Amsterdam/Philadelphia.
- Munhoz, Ana, and Rozana Naves. 2012. Construções de Tópico-Sujeito: uma proposta em termos de estrutura argumental e de transferência de traços de C. *Signum* 15: 245–265.
- Negrão, Esmeralda. 1999. *O português brasileiro: Uma língua voltada para o discurso*. “Tese de livre-docência”, Universidade de São Paulo.
- Nunes, Jairo. 1993. Direção de cliticização, objeto nulo e pronome tônico na posição de objeto em português brasileiro. In *Português brasileiro: Uma viagem diacrônica, 207–222*, ed. I. Roberts and M.A. Kato. Editora da UNICAMP: Campinas.
- . 2008a. Inherent Case as a licensing condition for A-movement: The case of hyper-raising constructions in Brazilian Portuguese. *Journal of Portuguese Linguistics* 7: 83–108.
- . 2008b. Preposition insertion in the mapping from Spell-Out to PF. *Linguistics in Potsdam 28: Optimality Theory and Minimalism: Interface Theories*, 133–156.
- . 2016. Subject and topic hyper-raising in Brazilian Portuguese: A case study on reference sets for economy computations. In *Morphosyntax of Spanish and Portuguese in Latin America*, ed. M.A. Kato and F. Ordóñez, 107–134. Oxford: Oxford University Press.
- . 2017. Circumventing ϕ -minimality: On some unorthodox cases of A-movement in Brazilian Portuguese. In *Romance languages and linguistic theory 12: Selected papers from the 45th Linguistic Symposium on Romance Languages (LSRL)*, ed. R. Lopes, J. Avelar, and S. Cyrino, 159–183. John Benjamins: Amsterdam/Philadelphia.
- . 2022. “Topic-subject” constructions in BP and minimality within the Agree model. *Estudos Linguísticos e Literários* 72: 299–324.

- Oliveira e Silva, Giselle. 1984. Variação no sistema possessivo de terceira pessoa. *Tempo Brasileiro* 78 (79): 54–72.
- Omena, Nelise. 1978. *Pronome pessoal de terceira pessoa: Suas formas variantes em função acusativa*. Master's thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- Ordóñez, Francisco, and Antxon Olarrea. 2008. Microvariation in Caribbean/non Caribbean Spanish interrogatives. *Probus* 18: 59–96.
- Pontes, Eunice. 1987. *O tópico no português do Brasil*. Campinas: Pontes.
- Rodrigues, Cilene. 2020. Estruturas com alçamento de possuidor e a restrição de entidade afetada: Evidência para controle como movimento. *Cuadernos de la ALFAL* 14 (2): 228–250.
- Scher, Ana Paula. 2004. *As construções com o verbo leve dar e as nominalizações em -ada no português do Brasil*. Doctoral dissertation, Universidade Estadual de Campinas.
- Sigurðsson, Halldór. 2008. The case of PRO. *Natural Language and Linguistic Theory* 26: 403–450.
- Tarallo, Fernando. 1983. *Relativization strategies in Brazilian Portuguese*. Doctoral dissertation, University of Pennsylvania.
- Toribio, Almeida J. 1996. Dialectal variation in the licensing of null and referential subjects. In *Aspects of Romance Linguistics*, ed. C. Parodi, C. Quicoli, M. Saltarelli, and M.L. Zubizarreta, 409–433. Washington, DC: Georgetown University Press.
- Torres Morais, Maria Aparecida. 2007. *Dativos*. “Livre docência” thesis, Universidade de São Paulo.
- Torres Morais, Maria Aparecida, and Rosane Berlinck. 2006. A caracterização do objeto indireto no português: Aspectos sincrônicos e diacrônicos. In *Para a história do português brasileiro: Novos dados, novas análises*, ed. R.V. Mattos e Silva and T. Lobo, vol. VI.1, 73–106. Salvador: Editora da UFBA.
- Torres Morais, Maria Aparecida, and Heloisa Salles. 2010. Parametric change in the grammatical encoding of indirect objects in Brazilian Portuguese. *Probus* 22: 181–209.
- . 2016. The external possessor construction in European Portuguese and Brazilian Portuguese. In *The Morphosyntax of Portuguese and Spanish in Latin America*, ed. M.A. Kato and F. Ordóñez, 204–235. Oxford: Oxford University Press.
- Vergnaud, Jean-Roger, and Maria Luisa Zubizarreta. 1992. The definite determiner and the inalienable constructions in French and in English. *Linguistic Inquiry* 23: 595–652.
- Wiedemer, Marcos. 2013. *Variação e gramaticalização no uso de preposições em contextos de verbo de movimento no português brasileiro*. Doctoral dissertation, Universidade Estadual Paulista, São José do Rio Preto.
- Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37: 111–130.
- Zaenen, Annie, Joan Maling, and Höskuldur Thráinsson. 1985. Case and grammatical functions. *Natural Language and Linguistic Theory* 3: 441–483.

Is Chilean Spanish a Canonical Pro-drop Variety? On Subjecthood in Chilean Spanish



Iván Ortega-Santos

1 Introduction

The goal of this chapter is to present the main properties of syntactic subjects in Chilean Spanish in comparison with better studied varieties of Spanish, such as Caribbean Spanish. Specifically, we will address the following research questions:

- (i) To what degree is Chilean Spanish similar/dissimilar to Caribbean Spanish?
- (ii) Is Chilean Spanish a canonical pro-drop variety?

To the best of my knowledge, this is the first time these questions are formally asked. Evidence will be provided for the view that the Null Subject properties of Chilean Spanish are not uniformly consistent with a positive specification of the Null Subject Parameter (Rizzi 1982).

The chapter is structured as follows: Sect. 2 introduces the general background on the Chilean dialect. Section 3 focuses on the Null Subject properties of the dialect under consideration in comparison with other varieties of Spanish, with a particular interest in Caribbean Spanish. Whenever pertinent, my claims about Chilean Spanish are supported by both corpus data and acceptability judgments. Section 4 interprets the results from Sect. 3, addressing the research questions raised above. Section 5 is dedicated to the conclusion.

A disclaimer is in order. The present research aims to set the stage for a broader dialectometric study of subjecthood across varieties of Spanish by pinpointing data points that can be gathered efficiently in existing corpora. The statistical

I. Ortega-Santos (✉)
University of Memphis, Memphis, TN, USA
e-mail: rtsantos@memphis.edu

analysis, which would add another piece to the picture I will present, is not included here. Also, wherever the discussion focuses on properties of Caribbean Spanish, microvariation within local varieties of Spanish (e.g., Dominican and Cuban Spanish) is left aside, unless relevant.

2 Background

Within generative grammar, studies of Spanish have devoted significant efforts to capture the properties of subjecthood (e.g., see Ordóñez 1997, Alexiadou and Anagnostopoulou 1998, Goodall 2001, Sheehan 2006, Camacho 2016, or Ortega-Santos 2016, among others), with the literature suggesting either an active or an inactive EPP requirement with consequences for the position posited for preverbal subjects (Spec,TP or Spec,TopP); see (1a) and (1b), respectively:

- (1) a. [TP Pedro lee [vP novelas]
 Pedro reads novels
 b. [TopP Pedro [TP lee novelas]]

Analysis (1a) is standardly assumed for English or French, that is to say, non-pro-drop languages, and therefore, the tendency is to see (1b) championed to capture the distinct behavior of Spanish and Null Subject Languages in general. Specifically, Null Subject Languages (of the Romance kind, cf. section 4.1) have been argued to have the following properties (Rizzi 1982; see Saab 2021 for recent discussion), which for current purposes I will refer to as canonical pro-drop features: use of null subjects unless surface semantic effects are present (e.g., focus; see section 3.2), availability of postverbal subjects (in the case of Spanish, VSO and VOS), null expletives, and lack of *that*-trace effects. These properties could be interpreted as evidence for the absence of an EPP requirement, (1b), (e.g., Ordóñez 1997; Alexiadou and Anagnostopoulou 1998), though the literature includes a number of proposals compatible with (1a), (e.g., Goodall 2001, Sheehan 2006, and Ortega-Santos 2016; see the latter work for a recent overview of the debate).

Caribbean Spanish dialects have received particular attention as they show linguistic properties believed to justify a distinct analysis of their subjecthood properties, with evidence provided by generativists and variationists. In general, the following has been reported as properties of these dialects: (i) use of overt subject pronouns without any requirement of surface semantic effects (contrast, emphasis, etc.), as exemplified in (2), in contrast to canonical Null Subject varieties (see Fernández Lagunilla and Rebollo 1995: 235 and Zagana 2002: 25); (ii) availability of overt subject expletives (specific to Dominican Spanish), as shown in (3); (iii) availability of Subject-Verb order in interrogative matrix sentences with non-D-linked argumental *wh*-phrases, as in (4); and (iv) widespread availability of the

SV order in adjunct infinitival clauses (see Pöll 2007 for recent discussion), as exemplified in (5).

- (2) Yo no lo vi, él estaba en
 I not CL.3SG.ACC saw he was in
 Massachussets acababa de llegar, pero muy
 Massachussets had-just of arrive.INF but most
 probablemente para el domingo pasado, que
 likely for the Sunday last that
 fue Día de las Madres allá, él estaba
 was Day of the Mothers there, he was
 en Nueva York ... Él estaba donde Eugenia, y
 in New York ... He was at Eugenia's and
 yo creo que él se va a
 I believe that he REFL.3SG will.3SG to
 quedar allá ...
 stay.INF there ...

‘I didn’t see him, he was in Massachusetts, he had just arrived, but quite probably by last Sunday, which was Mother’s Day there, he was in New York ... He was at Eugenia’s, and I think that he is going to stay there ...’

(Toribio 2000: 319)

- (3) a. Ello llegan guaguas hasta allá.
 it arrive.3PL buses till there
 ‘There arrive buses there.’
 b. Ello había mucha gente en lay-a-way
 it was many people on stand-by
 ‘There were a lot of people on stand-by.’

(Toribio 2000: 321)

- (4) ¿Qué tú quieres?
 what you want.2SG
 ‘What do you want?’

- (5) Para tú venir, hace falta un milagro.
 for you come.inf is need a miracle
 ‘A miracle is necessary for you to come.’

These characteristics led Toribio (2000) to put forward an active EPP analysis for Caribbean Spanish, though the dialect shows a high degree of complexity. In particular, it shows both pro-drop and non-pro-drop properties, as null subjects are still attested. Thus, in Toribio’s analysis, speakers of Caribbean Spanish are considered bidialectal: they speak both an active EPP variety, (1a), as suggested by (2)–(5), and a non-active EPP variety, (1b), as suggested by the availability of null subjects (seen in (2) as well).

Subsequent work on Caribbean Spanish (Martínez Sanz 2011) and on Afro-Hispanic varieties (Sessarego 2021) has refined Toribio's analysis arguing in favor of a view in which there is only one grammar, but two coexistent specifications of lexical items, giving rise to two different outputs, (1a) and (1b). This posited coexistence of the two variants within one grammar is in fact suggested by the language acquisition process and the resulting grammatical system, which includes cases of variable specification of linguistic features. It remains an open question, however, to what extent those properties are found in non-Caribbean Spanish and, most importantly, in other linguistic varieties within the so-called Bajeño dialectal area, which comprises both Caribbean and non-Caribbean Spanish, such as Chilean Spanish.

This chapter is devoted to analyzing subjecthood in Chilean Spanish, including but not limited to the properties exemplified in (2)–(5). It is argued that a mixed model including features of both Caribbean dialects and non-Caribbean null subject dialects best captures the properties of Chilean Spanish, in keeping with recent views that there are degrees of partiality in pro-drop properties (see Ticio 2018; Frascarelli and Jiménez-Fernández 2019). A critical discussion of the relevance of a number of criteria when determining pro-drop properties is provided.

2.1 *Background on Chilean Spanish*

Chilean Spanish is included in the Bajeño (Lowlands) dialectal area. As a dialect of Spanish, it is expected to be an SVO Null-Subject variety with flexible word order and rich verbal agreement morphology. This view, however, will be reassessed throughout the chapter. This section presents very briefly an overview of introductory works to the study of Chilean Spanish. (The list is far from comprehensive, as can be expected.)

The following large-scale atlases have been developed: *Atlas lingüístico – etnográfico del Sur de Chile (ALESUCh)* (Araya et al. 1973) and *Atlas Lingüístico y Etnográfico de Chile (ALECh) por regiones* (Wagner 1997). General introductions to the dialect including a summary of dialectal areas within the country can be found in Lipski (1996), Wagner (1996), and Palacios (2016). In turn, for a bibliography on the Chilean dialect, see Valencia (1995). Regarding the issue of how the syntax of Chilean Spanish compares to other varieties of Spanish, that is to say, the dialect distance (or dialect convergence) between this dialect and other varieties, see Ortega-Santos (2021), who studied the representativity and/or reliability of syntax journal data for Chilean Spanish, as well as the convergence rate between three dialects of the Bajeño dialectal area, namely, Chilean, Venezuelan, and Puerto Rican Spanish. In particular, a random sample of Spanish syntax data extracted from *Probus* (2006–2017) was reassessed through an acceptability judgment task conducted in the three dialects under consideration. The results for Chilean Spanish diverged the most from the two other varieties and from the journal data underscoring, thus, the need for further research on this dialect. This outcome is

also consistent with Moreno Fernández and Ueda's (2018) study of dialect cohesion in the Hispanic world, which focused mostly on non-syntactic features.

3 The Pro-drop Properties of Chilean Spanish

As noted in Sect. 2, according to the Null Subject Parameter (Rizzi 1982), Null Subject Languages prototypically license both null subjects in finite clauses (represented by *pro*) as well as postverbal subjects (VS order) while lacking *that*-trace effects. The descriptive generalization inherent in the Null Subject Parameter, namely, that those properties pattern together, has been challenged with crosslinguistic data. Gilligan (1987) shows a dissociation between lack of *that*-trace effects and the availability of null subjects, although an in-depth analysis of the potential counterexamples – controlling for potential interfering factors – is still missing. Moreover, a refinement of Parameter Theory in terms of hierarchically organized microparameters helps address empirical concerns (see Roberts and Holmberg 2010 for discussion of these two issues). While the debate on the Null Subject Parameter and parameters in general is still ongoing, establishing the exact properties of subjects in Chilean Spanish and determining its pro-drop properties is a goal in itself. Does it pattern with what is sometimes referred to as general Spanish, that is to say, Iberian and continental Latin American Spanish, which are well accepted as canonical Null Subject Varieties (though see Michnowicz 2015, Frascarelli and Jiménez-Fernández 2019, and Sessarego 2021 for relevant discussion on continental varieties)? Or does it pattern with Caribbean Spanish? Furthermore, this research helps fill a gap in our knowledge of microvariation and dialect distance within Spanish.

Since the pro-drop properties of Caribbean Spanish have figured prominently in the discussion of dialectal variation in subject properties Spanish, a good understanding of these varieties and how Chilean Spanish compares to them is essential to the current discussion. This section focuses on the following issues: morphological ambiguity and its alleged effect on subjecthood (Sect. 3.1), overt pronouns and their surface semantic effects (Sect. 3.2), inversion in *wh*-questions (Sect. 3.3), overt subjects of infinitives (Sect. 3.4), and the use of personal pronouns for inanimates and the use of generic *uno* “one” (Sect. 3.5). It is shown that Chilean Spanish displays a subset of the features associated with Caribbean varieties.¹

¹ *That*-trace effects arguably draw the line between pro-drop and non-pro-drop languages via the Empty Category Principle, ECP (Chomsky 1981). A long tradition assumes that subjects in pro-drop languages are less subject to locality constraints than in non-pro-drop languages. Support for this view and the ECP is also found in the extraction of subjects out of *wh*-islands, which are arguably possible in pro-drop languages, but not in non-pro-drop languages. Unfortunately, experimental studies on the locality of *wh*-islands have revealed that this observation does not hold and that *wh*-islands ban extraction of *wh*-subjects in both kinds of languages (see Sprouse et al.

3.1 *Morphological Ambiguity in Verbal Agreement and Subject Properties*²

As aforementioned, the Null Subject Parameter associates null subjects to rich agreement morphology (see Roberts and Holmberg 2010; D'Alessandro 2015 for detailed discussion). In keeping with this line of thought, researchers have linked the relatively high rates of overt pronouns in Caribbean Spanish to impoverishment of verbal morphology, both variationist sociolinguists (see Hochberg's 1986 functional compensation hypothesis and Flores-Ferrán 2007, Martínez Sanz 2011, and Martínez-Lara et al. 2021 for a critical review) as well as theoreticians (e.g., Suárez 1986). This being said, it has been noted that the issue is complex as languages without agreement also license pro-drop (so-called radical pro-drop or discourse pro-drop languages, e.g., Chinese or Japanese), and the uniformity in the agreement paradigm (rich agreement vs. zero agreement) has been taken as the decisive factor (Jaeggli and Safir 1989). For current purposes, it is arguably enough to notice that, at least in certain languages, overt agreement morphology is a requisite for null subject licensing (e.g., see D'Alessandro 2015). Chilean Spanish shows a prototypical Bajefiño feature relevant in this context, namely, the loss ([Ø]) or aspiration ([h]) of word final /s/ (e.g., Wagner 1996 and Lipski 1996). This may cause the ambiguity of 2nd person and 3rd person singular present tense, (6), and other tenses (preterit, imperfect, etc.), as *comes* in (6a) might be pronounced [komes], [komeh], or [kome]:³

- (6) a. Tú comes mucho.
 you eat.2SG much
 'You eat a lot.'
- b. Ella come mucho.
 she eats much
 'He eats a lot.'

This elision of the /s/ is precisely the main feature responsible for high verbal morphology ambiguity in Caribbean Spanish. Still, in Chilean Spanish, the present

2016 for Italian and Ortega-Santos et al. 2019 for Spanish, among others). Thus, I abstract away from the discussion of locality properties.

² For a state-of-the-art overview of the syntax of non-canonical pro-drop languages with an emphasis on Latin America, see Camacho (2016). As discussed by Camacho, for both Caribbean Spanish and Brazilian Portuguese, the partial loss of rich agreement morphology played a role in the change from pro-drop to partial pro-drop; nonetheless, the overlap in the non-canonical pro-drop properties of these two varieties is not complete.

³ A recurrent topic in the literature on the lenition of the -s across dialects is whether the loss of this sound correlates with a change in the preceding vowel, thus reducing the potential for ambiguity in a context like (6). For Chilean Spanish, see Bolyanatz (2020).

tense includes a distinct 2nd person morphology corresponding to *voseo*, an informal treatment that is being increasingly adopted across the social spectrum. This morphology is sometimes combined with the *tú* [“you”] pronoun (e.g., Lipski 1996, a.o.); *usted* is also used as a polite form of address, but may not be combined with the more colloquial *voseo* verbal morphology; *usted* is glossed as *youUSTED* and *vos* as *youVOS* in the chapter:

- (7) a. Vos comí mucho.
 youVOS eat much
 ‘You eat a lot.’
 b. Tú comí mucho.
 you eat much
 ‘You eat a lot.’

With regard to the more formal variety of Chilean Spanish, as expected from a Latin American variety, it does not make use of the *vosotros* [“you”] 2nd person plural pronoun typical of Iberian Spanish; *ustedes* is used instead, resulting, thus, in a reduction of the pronominal/verbal paradigm when compared to Iberian Spanish.

As a consequence, the Chilean dialect shows more ambiguity in its verbal paradigm than, say, standard Iberian Spanish, whose pronunciation leaves no room for ambiguity in (6), but less than Dominican Spanish, where there is no *voseo* verbal morphology, and the loss of the *-s* may cause (6) to be ambiguous (see Toribio 2000).

3.2 Overt Subject Pronouns and Their Surface Semantic Effects

High/low frequencies of usage of a certain syntactic feature cannot be equated with acceptable/unacceptable, as frequency and acceptability are different in nature (see Cornips and Gregersen 2016; Newmeyer 2013 for relevant discussion). Still, Caribbean Spanish has been associated with a relatively high use of overt pronouns, sometimes referred to as “overuse” of overt pronouns, as opposed to Iberian Spanish. Thus, frequency might provide a hint as to the status of a dialect, whether it is consistent, canonical pro-drop (Iberian Spanish), or non-canonical (Caribbean Spanish). With respect to this, Chilean Spanish seems closer to Iberian and mainland Spanish than to Caribbean Spanish. In Orozco and Hurtado’s (2021) meta-analysis of the data reported on the literature, the average rate of overt subject pronouns in the Caribbean across dialects is 38% as compared to 24% for mainland Latin American varieties (e.g., Lima, Mexico City, etc.) and 21% for Spain (see also Mayol’s 2012 overview of overt subject rates across dialects). For Chilean Spanish, Martínez-Lara et al. (2021) found a 25.3% of explicit subject pronouns in the PRESSEA corpus of Santiago de Chile in the context of finite verbs. This, in principle, suggests that

Chilean Spanish is a canonical Null Subject language, though, of course, whether the data taken into account by theoretical linguistics – which go beyond frequency of use – support this conclusion is an open question to be addressed next. In turn, Cifuentes' (1981) corpus analysis of interviews with educated speakers from Santiago de Chile reveals 35.95% of pronominal use; this being said, Cifuentes analyzed the use of overt pronouns co-referring with a previous pronoun; this restriction is not necessarily present in other studies, which may focus on co-referentiality irrespective of the overt pronoun or full DP status of the previous subject. This interfering factor makes it difficult to interpret the results in the context of the cited studies, as priming might have played a role in Cifuentes' results. Specifically, variationist studies of overt/null pronoun alternation have revealed that, when controlling for interfering factors, overt subject pronouns prime the use of another overt pronoun (e.g., Orozco and Hurtado 2021; see Mayol 2012 for discussion on priming across varieties of Spanish and its effect on canonical and non-canonical pro-dropness).⁴

Generative studies, based on acceptability judgments, have pointed out that overt subject pronouns in canonical null subject languages are acceptable, but arguably, they correlate with surface semantic effects such as contrast, focus, or emphasis or other semantic properties (see Chomsky's 1981 Avoid Pronoun Principle), as illustrated in (8)–(10).⁵

- (8) a. Ella_i piensa que (??ella_i) es buena
 she thinks that she is good
 persona.
 person
 ‘She believes she is a good person.’
- b. Ella_i piensa que ELLA_i es buena
 she thinks that she is good
 persona, no él.
 person not him
 ‘She believes she is a good person, not him.’

⁴ Sedano and Bentivoglio's (2014) analysis of subjecthood in eleven cities (ranging from Santiago to Havana and Madrid) using data from the PRESEEA project, specifically, the speech of one female and one male per city, reveals fairly similar percentages across cities with regard to the percentage of null subjects (but for the male from Madrid and the female from Mexico) as well as a strong tendency toward the use of preverbal pronouns across varieties. Still, a sample of only two speakers per city appears to be too small to reach reliable conclusions. In turn, see Frascarelli and Jiménez-Fernández (2019) for comparison of experimental results across various dialects (though not Chilean Spanish).

⁵ For recent discussion on the Avoid Pronoun Principle in Caribbean Spanish, see Camacho (2016), Ticio (2018), and Frascarelli and Jiménez-Fernández (2019), among others.

(9) [Context: coming from school, a child calls from the door]

A: ¿Quién es?
 who is
 ‘Who is it?’

B: Soy *(YO).
 is me
 ‘It’s me.’

(10) Ella estaba en la fiesta, pero *(él) no
 she was in the party but he not
 vino.
 came

‘She was at the party, but he did not come.’

(Flores-Ferrán 2007: 628)

Therefore, to study the pro-drop properties of a specific variety, it is necessary to focus on structural contexts where none of those surface semantic effects or any other independent factors are present. If only null subjects are licensed in these very neutral contexts, then we may conclude that the dialect under consideration is a canonical null subject variety. On the other hand, if overt subjects are allowed, then the dialect is a non-canonical null subject variety. Finite predicates are good testing grounds in Spanish because they may license both overt and null subjects, (see Pöll 2015 and references therein). The question, then, is to what extent this overt/null distinction is determined by surface semantics.

A qualitative analysis of the materials available online in the *PRESEEA Santiago de Chile* corpus reveals the use of overt subject pronouns in semantically neutral contexts, indicating that this dialect behaves similarly to Caribbean Spanish (irrespective of the overt pronoun rate discussed above). Specifically, co-referent overt first-person singular pronouns without surface semantics are found in subject position of embedded clauses, (11a); in conjoined clauses, (11b); and in adjunct clauses (ranging from *if*-clauses, (12), to relative clauses). The relevant subject pronoun appears in italics in the corresponding examples. In turn, the examples in (13)–(15) illustrate the use of third-person pronouns in semantically neutral context: Across sentences ((13) and (14)), in embedded contexts and under coordination (15). For discussion of the asymmetries between first- and second-person pronouns vs. third-person pronouns in the licensing of null subjects, see, for instance, Frascarelli 2018; it is the latter that are particularly relevant for the discussion of pro-drop

properties.⁶ [NB: The transcriptions of the corpus have been simplified by erasing clarifications regarding lexical items or simultaneous speech to increase readability; the original spelling, hesitations, and repetitions are maintained; the interview number of each example is included throughout.]

- (11) a. Yo me sentí muy indefenso/ en
 I refl felt.1SG very defenseless in
 ese sentido en decir como yo
 that sense in say.infl how I
 demuestro que yo no tenía nada
 demonstrate that I not had.1sg anything
 que ver (SCHI_H23_085)
 to do.inf
 ‘I felt vulnerable in that sense. How could I prove that I had nothing to do with that?’
- b. Sí yo les conté parece y yo
 yes I CL.3PL.DAT told seems and I
 les dije / y no / y yo soñé
 CL.3PL.DAT told and not and I dreamt
 así como que yo veía la nota
 this-way as-if that I saw the grade
 y veía que me iba muy bien
 and saw that to-me went very well
 ‘Yes, it seems I told them, I told them and no, and I dreamt like I saw the grade and I saw that I did pretty good.’ (SCHI_M13_079)

⁶ The PRESEEA project includes interviews with speakers about their life experiences, dramatic episodes, etc. As a consequence, it is particularly useful to study the behavior of the first-person singular pronoun. Research has shown that not all pronouns are used as frequently, *yo* being used more often than the rest (Flores-Ferrán 2007 and references therein). For a generative analysis of the overt/null alternation of first singular pronouns with *creer* ‘to believe’ and *saber* ‘to know’ in Central Iberian Spanish, see Herbeck (2021). According to him, overt *yo* is licensed by perspectival contrast. It is worth noting that the current research revealed the use of overt pronoun beyond that context. In turn, for discussion on the interpretation of null pronouns, particularly third-person pronouns, see Frascarelli (2018) and other works by this researcher.

- (12) a. Ya yo en la mañana me levanté
 already I in the morning REFL.1SG got.up
 bien / me levanté bien // pero cuando
 well REFL.1SG got.up well but when
 yo ya salí a la calle / yo salí
 I already went.out to the street I went.out
 mareada total (SCHI_M32_067)
 dizzy completely
 ‘Already in the morning I woke up ok, I woke up ok, but when I left for the street, I was totally dizzy.’
- b. Si yo quisiera saber yo podría ir a
 if I wanted to-know I could go.INF to
 la posta porque yo tengo mi ficha
 the urgent.care.center because I have my record
 ahí (SCHI_M32_067)
 there
 ‘If I wanted to know, I could go to the urgent care center, since my file is there.’

- (13) Ellos las fondas de ellos
 they the pop-up restaurant of them
 están en sus casas / ellos hacen sus
 are in their houses they cook their
 asaditos ellos ellos pasan sus días
 barbecues they they spend.3PL their days
 ahí con sus familias (SCHI_M32_067)
 there with their families
 ‘Their taverns are in their houses. They cook barbecues. They spend their days there with their families.’

- (14) El automóvil bueno lo facilitó ehh lo
 the car well it lent.3SG hm it
 facilitó // mi padrino mi padrino
 lent.3SG my godfather my godfather
 él tenía auto entonces lo facilitó pero él
 he had.3SG car so it lent.3SG but he
 sufrió un descuido / él usaba mucho el
 suffered.3SG a distraction he used.3SG much the
 el automóvil (SCHI_H32_061)
 the car
 ‘The car, well, my godfather lent it to us. My godfather, my godfather, he had a car, so he lent it to us, but he had a problem. He used to use the the car a lot.’

- (15) Él me ha hecho entender de que
 he CL.1SG.DAT has made understand.INF of that
 él nada juzga él sí que es de
 he nothing judges he yes that is of
 verdad / no es de cartón / él es
 truth not is of cardboard he is
 verdad sí / pues me ha visto
 truth yes since CL.1SG.ACC has seen
 en hartas paradas distintas y él nunca
 in many situations different and he never
 me ha dicho oye/ cómo se te
 CL.1SG.DAT has said hey how CL CL.2SG.DAT
 ocurre.
 occurs
 ‘He has made me realize that he is not judgmental. He really is genuine, not a fake. He is honest, since he has seen me in different situations and he has never told me “hey, how did you even think about that?”’

Thus, with regard to this criterion, Chilean Spanish has non-canonical pro-drop grammar, as overt subjects appear in contexts where null subjects are expected to be found. Still, the overt pronominal subject rate of this variety is low when compared to Caribbean Spanish (see the discussion in this section).⁷

3.3 *Inversion in Wh-Questions*

As shown in (4) above, repeated here as (16), the SV order in interrogatives without the wh-element being D-linked or an adjunct is a feature of Caribbean Spanish. This property is relevant in that it has been interpreted as an SVO/EPP requirement (see Brown and Rivas 2011 for relevant discussion), that is to say, the kind of word order enforced in English (though *do*-support may obscure this tendency in wh-questions). The Wh S V order, however, is not particularly felicitous in Chilean Spanish, according to speakers I consulted:⁸

- (16) ¿Qué tú quieres? *Caribbean Sp./*Chilean Sp.*
 what you want.2SG
 ‘What do you want?’

⁷ Cases of binding of a subject pronoun, whether null or overt, by a quantifier functioning as the main subject are not attested in the corpus, thus hinting at the potential limits of corpora when studying highly specific or infrequent structures. Those cases are relevant for the study of the “Avoid Pronoun Principle” as well.

⁸ Acceptability judgments of Chilean Spanish data throughout the paper were tested with three naïve native speakers.

While there is some controversy in the literature on Caribbean Spanish as to whether full DPs may intervene between the *wh*-phrase and the verb (see Ordóñez and Olarra 2006; Comínguez 2018 for an overview), Toribio (2000: 322–323) provides examples like (17), all of which are degraded in Chilean Spanish.

- (17) a. ¿Qué ese letrado dice?
 what that sign says
 ‘What does that sign say?’ *Dominican Sp./*Chilean Sp.*
- b. ¿Qué yo les voy a mandar a
 what I CL.3PL.DAT will.1SG to send.INF to
 esos muchachos?
 those boys
 ‘What am I going to send to those boys?’ *Dominican Sp./*Chilean Sp.*
- c. ¿Cuánto un médico gana?
 how much a physician earns
 ‘How much does a doctor earn?’ *Dominican Sp./?*Chilean Sp.*
- d. ¿Y con quién Fredi está allá?
 and with who Fredi is there
 ‘And with who is Fredi there?’ *Dominican Sp./?*Chilean Sp.*

As (18) and even (11a) indicate, non-inverted structures can be found in the PRESSEA corpus, but none of them contains non-D-linked *wh*-arguments:

- (18) Para qué yo voy a decir
 for what I will.1SG to say.INF
 que no?
 that not
 ‘Why would I say no?’
 (SCHI_M32_067)

In a similar vein, Lipski (1977: 62) points out that the loss of verbal agreement morphology does not correlate with the *Wh S V* order in Spanish varieties. Andalusian Spanish and Chilean Spanish are explicitly mentioned by the author as examples of the lack of correlation.

This difference between Caribbean Spanish and Chilean Spanish was corroborated in a search for questions with the sequence [*qué tú*] in the corpora *CREA* and *Corpus del Español NOW*. This sequence was chosen because *tú* is arguably the most frequent and/or widely accepted pronoun used in this context (see Ordóñez and Olarra 2006; see also Pöll 2015 and Orozco 2020, among others). The order [*qué tú*] seems relatively absent in the Chilean data (only one case in the first corpus, none in the second one). Puerto Rico and Cuba provide a number of cases often combined with the verb *creer* [‘think’] or *decir* [‘say’] (e.g. ¿*Qué tú crees/dices* (. . .)? [‘What do you think/say?’]), with or without an overt clausal complement,

but cases without a complement could potentially be analyzed as fixed expressions and, thus, might not be as relevant.⁹ Clearly, this kind of search gives us only a limited insight into the data, but it helps highlight the fact that the distribution of subject pronouns in Caribbean and Chilean Spanish is different.

Summing up, the Wh S V order is attested to a limited extent in Chilean Spanish, but its absence with non-D-linked wh-arguments implies that this dialect does not pattern with Caribbean Spanish, as seen in the absence of the [*que tú*] sequence in two different corpora.

3.4 *Infinitival Subjects*

Overt preverbal pronominal subjects in adjunct infinitival clauses are more widespread in Caribbean varieties (see (5)) than in Iberian Spanish (Toribio 2000, Flores-Ferrán 2007, and Ortiz-López et al. 2018, and references therein). As in the case of the Wh S V order, this property has been argued to illustrate the SVO/EPP property of the dialect. Chilean Spanish also allows for overt preverbal subjects in adjunct clauses in cases where Iberian Spanish rejects them, as shown by a preliminary search in the PRESSEA corpus and a review of the literature. The following instances of overt preverbal pronouns in adjunct infinitival clauses were located in the corpus:¹⁰

- (19) No me gusta que me
 not to-me pleases that to-me
 tuteen tiene que ser mucho para yo
 address.as.tú has that be.INF much for I
 aceptar a una persona (SCHI_M32_067)
 accept.INF to a person
 ‘I don’t like others to address me informally, I need to really know a
 person to accept him/her.’

⁹ Interestingly, the Dominican Republic provided only one case in the CREA as opposed to the Corpus del Español NOW, thus hinting at a bias in that corpus.

¹⁰ Since these are non-finite contexts, a priori, the licensing of an overt subject calls for an explanation. One of the hypotheses explored in the literature is that these infinitivals have abstract agreement, just like personal infinitives in Portuguese. Still, the absence of overt agreement morphology renders this proposal ad hoc (see Pöll 2007 for discussion).

- (20) Me acuerdo que para yo no
 REFL.1SG remember.1SG that for I not
 ver a la novia me tuve que
 see.INF to the fiancée REFL had.1SG that
 arrendar una casa cercana a la
 rent.INF a house close to the
 casa
 house (SCHL_H23_085)
 ‘I remember that I had to rent a house nearby to avoid seeing my
 fiancée before the wedding.’

- (21) Una vez por yo no saber
 one time for I not know.INF
 ‘One time, since I did not know’ (SCHL_H22_049)

Because of the low frequency of this structure, a search was also conducted in the *Corpus del Español NOW*, including the sequence *de yo* [“of + I”]. The choice of the preposition *de* [“of”] is motivated by the fact that conservative dialects of the Iberian kind reject overt preverbal subjects in this context, as opposed to contexts with the preposition *sin* [“without”], which allows overt preverbal subjects across dialects (see Mensching 2000). In turn, the choice of the first-person pronoun is motivated by the fact that sociolinguistic studies on the overt/null alternation have revealed that this is the pronoun that tends to be used more often overtly across varieties (see Flores-Ferrán 2007 and references therein). Examples from the Caribbean and Chile are frequently present in the corpus, whereas those from Spain were rare in comparison. Chile is the middle scale with the Caribbean and Spain at the extremes. The cases in (22)–(24), from Chilean Spanish, were found in the *Corpus del Español NOW*.

- (22) Estoy aburrido de yo tener que
 am.1SG bored of I have.INF that
 venir sin que él sepa. (18-06-2018 La Tercera)
 come.INF without that he knows
 ‘I am tired of having to come here without him knowing.’

- (23) Me lo sugirió al momento
 CL.1SG.DAT CL.3SG.ACC suggested.3SG at.the time
 de yo asumir el cargo. (29-09-2016 El Dínamo)
 of I assume.INF the position
 ‘He suggested it to me at the time I assumed the position.’

- (24) Al momento de yo querer empezar
 at.the time of I want.INF start.INF
 a arreglarme para amamantar a mi
 to prepare.INF.REFL to breastfeed.INF to my
 hijo, el administrador me
 son the manager CL.1SG.DAT
 dijo que...
 told.3SG that
 ‘At the time I wanted to start preparing to breastfeed the baby, the manager
 told me that...’
 (13-03-2014 La Nación)

Furthermore, the examples below taken from the literature exemplify preverbal subjects in infinitival clauses ((25) from Vidal (1980–1981: 947) and (26) from Contreras (1978:177)):

- (25) Se hizo fuerte [...] con estos dineros,
 REFL became.3SG strong with these funds
 no para usarlos personalmente,
 not to use.INF.CL.3PL.ACC personally
 pero queriendo que fueran usados en
 but wanting that were used in
 este congreso por el hecho de él
 this symposium for the fact that he
 haber sido eliminado de la
 have.INF been eliminated of the
 presidencia.
 presidency
 ‘He/she became strong thanks to these funds, not to use them himself/herself, but in
 an attempt to have them spent in this symposium, just because he/she was not the
 president anymore.’

- (26) Me mostraba su memoria para yo
 CL.1SG.DAT showed.3SG his/her report for I
 comenzar a leerla.
 start.INF to read.INF.CL.3SG.ACC
 ‘He/She showed me his/her report for me to start reading it.’

In turn, Lipski’s (1996) classic work on Latin-American Spanish notes the availability of preverbal subjects in infinitival clauses in the Antillas (Caribbean islands), Venezuela, Ecuador, Colombia, and Argentina, whereas Real Academia

(28) *Personal and impersonal pronouns* (Toribio 2000: 321)

<i>Uno</i>	<i>se</i>	<i>da</i>	<i>cuenta</i>	<i>que</i>	<i>uno</i>	<i>es</i>	<i>adulto</i>	<i>ya:</i>
one	REFL.3SG	realizes		that	one	is	adult	already
<i>nadie</i>		<i>te</i>		<i>controla,</i>	<i>nadie</i>		<i>va</i>	<i>a</i>
nobody		CL.2SG.ACC		controls	nobody		goes	to
<i>ver</i>		<i>tus</i>		<i>notas,</i>	<i>nadie</i>		<i>te</i>	
see.INF		your		grades	nobody		CL.2SG.DAT	
<i>dice si</i>		<i>tú</i>		<i>vas</i>	<i>o</i>	<i>no</i>	<i>vas.</i>	<i>Tú</i>
tells if		you		go.2SG	or	not	go.2SG	you
<i>haces lo</i>		<i>que</i>		<i>tú te</i>			<i>propones</i>	<i>a</i>
do.2SG the		that		you refl.2SG			decide	to
<i>hacer.</i>								
<i>do.</i>								

‘You realize that you are an adult: nobody controls you, nobody’s going to see your grades, nobody tells you if you can go or not. You do what you set out to do.’

This section analyzes these linguistic features in Chilean Spanish to establish whether the purported paradigmatic pressure could potentially have applied in that variety as well. Overt personal pronouns with non-human reference are not used productively in the dialect, at least for the speakers I consulted, which were highly educated. This being said, some naturalistic examples could be found, e.g., one in the literacy app *Bartolo*, which was used by the Chilean government throughout the pandemic, and one in PRESSEA, where the latter reveals the use of *él* [‘him’] for a vehicle, though not in subject function and, thus, is less relevant for current purposes.¹¹

¹¹ Possessives might be used with non-human reference:

i	<i>Vamos</i>	<i>a</i>	<i>comer</i>	<i>un</i>	<i>asado</i>	<i>con</i>	<i>su</i>	<i>carnecita,</i>
	will.1PL	to	eat.INF	a	barbeque	with	his	meat.DIM,
	<i>sus</i>	<i>papitas . . .</i>						
	its	potatoes.DIM						

‘We will have a barbeque with tasty meat, potatoes...’

Still another case of use of personal pronouns for horses is attested in the PRESSEA corpus:

ii	<i>Al</i>	<i>caballo</i>	<i>primero</i>	<i>se</i>	<i>le</i>	<i>muestra</i>	<i>donde</i>	<i>va</i>	<i>a</i>
	to.the	horse	first	SE.ARB	CL.3SG.DAT	shows	where	will.3SG	to
	<i>caminar /</i>	<i>ehh</i>	<i>donde</i>	<i>va</i>	<i>a</i>	<i>vivir /</i>	<i>donde</i>	<i>él</i>	
	walk.INF	hm	where	will.3SG	to	live.INF	where	he	
	<i>va</i>	<i>a</i>	<i>comer</i>						(SCHL_H21_013)
	will.3SG	to	eat.INF						

‘First, we show the horse where he will walk, where he will live, where he will it.’

(ii) is relevant in that other varieties may not use a personal pronoun in this context. This being said, horses are animates and, thus, the paradigm shift analysis is only partially supported.

- (29) Tú, Nico, recorrerás el mundo de
 you Nico travel.FUT.2SG the world of
 las palabras y descubrirás la
 the words and discover.FUT.2SG the
 magia que *ellas* esconden. (Bartolo app)
 magic that they hide.3PL
 ‘You, Nico, will travel around the world of the words and you will
 discover the magic they hide.’
- (30) Cuando trabajé con mi furgón / era
 when worked.1SG with my truck was
 diferente / igual gané plata y todo /
 different anyway made.1SG money and all
 pero yo quería / que el furgón me
 but I wanted.1SG that the truck CL.1SG.DAT
 diera / que yo a *él* lo
 gave.3SG that I to him CL.3SG.ACC
 podía meter en una empresa buena /
 could.1SG put.INF in a business good
 pero nunca lo pude hacer
 but never CL.3SG.ACC could.1SG do.INF
 ‘When I worked with my truck, it was different. I made money and
 all that, but I wanted the truck to help me earn a lot. I could have worked for a good firm
 with that truck, but I did not succeed.’
- (SCHI_H12_037)

Again, absence of a certain structure in a corpus is not a strong evidence that it is unacceptable. It might just be infrequent, or there might be a bias in the data present in the corpus. The examples in (29)–(30) suggest that use of personal pronouns for inanimates is part of the dialect and, to a certain extent, the idea that a paradigm shift à la Toribio may have taken place in the dialect finds support.

Uno [“one”] with first-person singular reference and with impersonal reference is also found in the PRESEEA corpus. In particular, *uno* with first-person singular reference is attested fairly frequently, due to the format of the PRESEEA, namely, interviews about a range of issues, including personal experiences, (31)–(33).

- (31) Antes veía teleseries porque *uno*
 before watched.1SG TV.series because one
 era chico y no lo dejaban
 was little and not CL.3SG.DAT let.3PL
 salir cuando llovía / a *uno* lo
 go.out.INF when raining to one CL.3SG.ACC
 dejaban ahí / pegado al televisor
 left.3PL there in front of.the television
 ‘Before, I used to watch TV series, because I was little and my parents
 did not allow me to go out when it was raining. They just had me watch
 TV.’ (SCHI_H21_013)
- (32) Si de repente pasa otra mina por delante /
 if suddenly walks another woman in front
 mío igual *uno* <ininteligible/> una mirada
 of-me maybe one unintelligible a look
 pero / no me gusta que después
 but not cl.1s.dat pleases that afterwards
 te haga show ni nada (...) porque
 CL.2SG.DAT makes scene not anything because
uno está con ella pues (SCHI_H12_037)
 one is with her after all
 ‘If a woman walks in front of me, I look at her, but I don’t like my wife to make a
 scene, because I am with her after all.’
- (33) La verdad que que esta gente está
 the truth that that these people is
 para sacarse el sombrero / o sea
 to take-off.INF.REFL the hat or be.SUBJ
uno justifica plenamente lo que *uno*
 one justifies completely the that one
 ve como expresión de cariño
 sees like expression of affection
 hacia los carabineros porque realmente /
 towards the policemen because really
 están al servicio de la comunidad
 are at.the service of the community
 ‘The truth is that these people are really great. That is to say, one
 justifies completely what one sees as a sign of affection towards the
 police, because they really serve the community.’ (SCHI_H23_085)

In turn, the example in (34) illustrates the use of *usted* as a generic pronoun, referring to people in general, including both the speaker and the addressee, in keeping with Toribio's (2000:320) claim about paradigmatic pressure that together with *uno*, *usted* and *tú* may tend to be used as generics, instead of "a null non-specific plural pronoun or an impersonal *se* construction."

- (34) Las micros andaban por todos lados / o
 the buses were.3PL everywhere or
 sea *usted* sabía que si salía
 be.SUBJ you_{USTED} knew.3SG that if left.3SG
 de su casa / *usted* decía ya aquí a
 of your house you_{USTED} said.3SG ok here to
 la esquina / pasa mi micro (SCHL_M32_067)
 the corner stops my bus
 'The buses were everywhere. That is to say, you knew that if you went
 out, you could say, ok, here at the corner is the bus line.'

Variationist research on the use of pronouns with impersonal reference, a.k.a. as nonspecificity of reference, has also noted Chilean Spanish patterns with Caribbean Spanish as opposed to Iberian Spanish in a number of respects. According to Cameron's (1997: 55–56) survey of existing research, nonspecificity of reference in second-person *tú* favors pronominal insertion in San Juan (Puerto Rico), in Buenos Aires (Argentina), and in Santiago (Chile) while disfavoring pronominal expression, relative to specific *tú*, in Madrid and Seville. For nonspecific *uno*, the same effect of nonspecificity is found in San Juan and Santiago, though the tendency is stronger in San Juan in comparison to Madrid.¹² The idea that a paradigm shift may have taken place in this dialect is, thus, supported, if the parallelism between Chilean and Puerto Rican Spanish is taken as evidence in this regard. However, this is an indirect piece of evidence concerning the null subject properties of Chilean Spanish. Hence, it is important to ponder on the value of the data presented here for an analysis of a pro-drop variety. In turn, the use of personal pronouns for inanimates is English-like, where one may refer to, say, a ship as "she." Partial pro-drop behavior or non-canonical pro-drop behavior is, to a certain degree, English-like as noted in Sect. 2 (though see Sect. 4.1 for further discussion). This being said, it is unclear that this feature draws the line between a canonical and a non-canonical null subject variety. As noted by Posio (2013: 259; see references therein), this feature is present in European Portuguese, but significantly less so in Madrid Spanish, although both varieties are considered pro-drop varieties.

¹² For discussion on generic/impersonal pronouns in the Chilean variety of Valdivia, see Hurtado Cubillos (2009).

With regard to impersonals, the link to pro-drop is slightly indirect. Pro-drop has been associated with SE impersonal constructions (see MacDonald and Maddox 2020). Thus, partial pro-drop varieties shift away from this construction and allegedly explore alternative ways of expressing impersonal meanings to varying degrees (though a partial pro-drop variety may still use SE impersonals in certain contexts; again, see Sect. 4.1). The behavior of Caribbean and Chilean Spanish might be a case in point. Thus, the data in this section are relevant for the discussion of the pro-drop properties of the dialect, but their relevance should be taken with a grain of salt.

4 Chilean Spanish as Non-canonical Pro-drop Language

Table 1 summarizes the linguistic properties unveiled so far. Iberian Spanish is included for comparison, Chilean Spanish being in the middle of a hypothetical scale, sharing properties with both Caribbean and Iberian Spanish.

Given the results in Table 1, it is now possible to answer the research questions, repeated here for the sake of exposition: (i) To what degree is Chilean Spanish similar/dissimilar to Caribbean Spanish? (ii) Is Chilean Spanish a canonical pro-drop variety?

Regarding the issue of surface semantic effects and overt pronouns, overt subject pronouns are used without a topic change, contrast, or focus. In turn, the use of preverbal subjects in prepositional infinitival clauses is allowed. Expletive *ello* is not attested (as expected, since this is a privative feature of Dominican Spanish).¹³ There is some fairly limited use of personal pronouns with inanimate reference, and *uno* is used relatively frequently with first-person reference. All together, these properties show that Chilean Spanish patterns with Caribbean Spanish rather than Iberian Spanish. Still, out of these criteria, only the use of overt pronouns without

Table 1 Summary of subject-related properties included in the literature on pro-drop for Caribbean, Chilean, and Iberian Spanish. +/– means that there is conflicting evidence, e.g., the property is attested, but it is not particularly frequent or widespread

	Caribbean Sp.	Chilean Sp.	Iberian Sp.
Overt pron. without surface semantics	+	+	–
Wh _{non-D-linked} S V	+	–	–
Preposition S infinitive	+	+	–
Use of personal pronouns for inanimates	+	+/–	–
Use of <i>uno</i> “one” with first-person reference	+	+	–

¹³ This being said, the status of *ello* as an expletive sitting in Spec,TP as part of the loss of pro-drop properties in Caribbean Spanish has been contested (see Gupton and Lowman 2013 and Camacho 2016 and references therein).

surface semantics stands out as particularly strong evidence in favor of the non-canonical pro-drop analysis. By that we mean that this is unexpected for a canonical pro-drop analysis, suggesting that either the variety at hand is non-canonical pro-drop or else the relation between the Avoid Pronoun Principle and pro-drop needs to be reconsidered.

Moreover, verb-subject inversion in interrogatives with argumental non-D-linked wh-phrases is not attested, in contrast to Caribbean Spanish. In particular, the absence of the [*qué tú*] sequence, contrasts with Caribbean Spanish. Given the various features that both varieties share, this difference calls for an explanation. Ordóñez and Olarrea (2006) provide empirical evidence that subjects in the Wh S V order tend to be pronominal as opposed to full DPs. This led them to argue that pronouns in Caribbean Spanish are weak in the sense of Cardinaletti (1997) and Cardinaletti and Starke (1999). Details aside, the weak/strong pronoun distinction arguably captures the difference between Caribbean and non-Caribbean dialects, where the latter have strong pronouns in their grammar – as opposed to weak pronouns.¹⁴ If this is right, it is possible to conclude that pronominal subjects in Chilean Spanish are not weak pronouns in contrast to Caribbean Spanish. Yet, other non-canonical pro-drop features are attested in the dialect.¹⁵

To conclude, thus, the present research adds Chilean Spanish to the growing number of Spanish non-canonical pro-drop varieties (see Sessarego 2021 for Latin-American Afro Hispanic varieties, and Frascarelli and Jiménez-Fernández 2019).¹⁶

¹⁴ Ordóñez and Olarrea (2006) argue that TP remnant movement takes place in wh-questions across dialects. Since weak pronouns appear in a low position in the clause, this movement does not change the surface word order in Caribbean Spanish. Given that strong pronouns are higher in the clause, the surface word order changes, giving rise to inversion in non-Caribbean Spanish.

¹⁵ Ordóñez and Olarrea (2006) provide evidence that light and morphologically simple pronouns (e.g., *tú*) are more likely to appear in the Wh SV order than *nosotros* and other heavier pronouns. Such tendencies are found elsewhere in Caribbean Spanish as the following results on word order in Cuban Spanish by Ortiz-López et al. (2018: 113) point out: “Explicit pronouns (*yo/ll, tú/you* (+/- Spec, *uno/one* (-Spec)) appear almost categorically in preverbal position, independent of the type of clause [+/- fin],” as opposed to more complex or heavier subjects. This suggests that a performance factor, namely, the tendency for complex or heavier phrases to appear at the end of the clause, might affect the distribution of overt subjects (see Ortiz-López et al. 2018 and references therein for discussion), potentially leading to the grammaticalization of these tendencies causing the grammar to make a difference between *yo/tú/uno* and other subjects. Note that the analysis of the difference between Chilean and Caribbean Spanish only addresses the behavior of pronouns. Still, see Comínguez (2018) for evidence that preverbal DPs may also be allowed in this context in Caribbean Spanish as well as an explanation that does not rely on the weak/strong pronoun distinction across dialects.

¹⁶ Frascarelli and Jiménez-Fernández argue in favor of degrees of partiality in that the same dialect may show partial null subject properties in a certain syntactic context but not in others. The following dialects show a certain degree of partiality: Andalusian, Castilian, Catalanian, Extremaduran, Caribbean, Mexican, and River Plate Spanish.

4.1 *Chilean Spanish vs. Partial Null Subject Languages*

Null Subject Languages are divided into various categories (see Roberts and Holmberg 2010, among others, for details):

- Canonical or consistent null subject languages
- Expletive null subject or semi pro-drop languages, which allow for expletive null subjects, but not argumental ones, e.g., German and various creole languages
- Discourse or radical pro-drop languages, that is, pro-drop with zero verbal morphology, e.g., Chinese and Japanese
- Partial null-subject languages, where null subjects are restricted by syntactic and morphological conditions, e.g., Finnish and Brazilian Portuguese

This section will briefly discuss whether Chilean Spanish is a partial pro-drop variety or not. First, asymmetries in licensing of null subjects determined by combinations of number or person features, while attested, for instance, in a partial null subject languages like Finnish (Roberts and Holmberg 2010), are, to the best of my knowledge, absent (though there are differences in the overt use of pronouns depending on these factors; see fn. 6 for relevant discussion). Thus, I will focus on two other defining properties of partial null subject languages: the distribution of null generic subject pronouns equivalent to English *one* and locality effects in antecedent-pronoun coreference.

The availability of null generic third-person singular pronouns has been claimed to be a significant property of partial pro-drop languages (Roberts and Holmberg 2010 and Holmberg and Sheehan 2010; see these works for proposals able to capture the divide between partial and canonical pro-drop languages with respect to this feature) and can be illustrated with examples like (35) from Brazilian Portuguese (data from Rodrigues 2004, quoted in Barbosa 2019):

- (35) É assim que faz o doce.
 is.3SG so that make.3SG the desert
 ‘This is how one makes the desert.’

In particular, the embedded null subject is interpreted as “people in general, including speaker and addressee” (Barbosa 2019: 490). Still, this reading is absent in Caribbean and Chilean Spanish (as well as non-Caribbean Spanish).¹⁷ According to Barbosa (2019), null generic third-person singular pronouns correlate with bare NP licensing. Inasmuch as (i) non-dislocated bare NPs are independently banned from preverbal subject positions (and postverbal position too, but for the subject of

¹⁷ Generic *uno* might be null when it co-refers with an overt impersonal *uno* (see Maddox 2016). Note that the generic meaning is conveyed with a SE clitic in Spanish and European Portuguese. Overt generic *uno* is also attested across Spanish (as opposed to *uno* with first-person reference), though varieties may differ in its productivity (see Sect. 3.5).

unaccusatives), (ii) this property, allegedly, has not changed in the Caribbean and Chilean pro-drop varieties of Spanish (as opposed to certain varieties/idioms of US Spanish), this might explain the absence of the said generic pronouns in Spanish. This is illustrated in (36) for the various dialects under discussion:

- (36) a. *(Los) perros ladran.
 the dogs bark.3PL
 ‘Dogs bark.’ *General, Caribbean and Chilean Sp.*
- b. *(El) café es muy sano.
 the coffee is very healthy.
 ‘Coffee is very healthy.’ *General, Caribbean and Chilean Sp.*

In contrast, the distribution of bare nouns varies from European to Brazilian Portuguese, where Brazilian Portuguese allows them in subject and object position as opposed to European Portuguese, which has a more restricted distribution (see Barbosa 2019 and references therein). This being said, the relevance of the presence/absence of a D-feature in T and its consequences for the distribution of bare NP’s has been called into question for Brazilian Portuguese (see Carvalho 2019 for discussion).

It is worth noting that impersonal third-person plural null subjects differ in partial and consistent null subject languages in that only the former can have an inclusive reading – if the construction is indeed available – as opposed to the latter, whose reading excludes speaker and addressee (original European Portuguese example from Barbosa 2019, translated into Spanish):

- (37) Aquí trabajan mucho.
 here work.3PL a.lot
 ‘Here they (people in general excluding the speaker and addressee) work a lot.’

Partial pro-drop languages as Brazilian Portuguese, Finnish, or Hebrew are also known for showing locality effects in the licensing of third-person null pronouns, meaning that the relationship between a null subject pronoun and its antecedent is local, whereas coreference with overt subject pronouns can be non-local (Rodrigues 2004, among others).¹⁸ Most cases of overt subject pronouns in Chilean Spanish were found in non-local environment (coordination, (11) and (15), or adjunct clauses, (12)), as opposed to embedded clauses; the latter environment

¹⁸ For recent discussion of variation in the locality properties of null subject in partial null subject languages and an account for the lack thereof, see Barbosa (2019) and references therein; see Frascarelli (2018) for experimental evidence challenging the received view on locality constraints in partial pro-drop varieties.

was comparatively less present in the corpus, and in many cases, the main verb was a quotation verb, as illustrated in (38):

- (38) a. Yo le había dicho a mi
 I CL.3SG.DAT had.3SG told to my
 señora que yo iba a volver
 wife that I was to return.INF
 como a las dos. (SCHI_H22_049)
 around at the two
 ‘I had told my wife that I was going to come back around two.’
- b. Yo le digo a ellos que yo
 I CL.3SG.DAT tell.1SG to them that I
 estoy enamorada de / y son mi vida
 am in love with and are my life
 entera (SCHI_M11_007)
 whole
 ‘I tell them that I love them and that they are my sunshine.’

In other words, it seems as if null subjects tend to be used in local environments, a fact that merits further research. This being said, null subjects in Chilean Spanish do not need to be bound by the closest local antecedent according to my informants, in contrast to Brazilian Portuguese or Finnish (e.g., see Barbosa 2019: 513 and Holmberg and Sheehan 2010):

- (39) Juan dice que María piensa que ganó la
 Juan says that Mary thinks that won.3SG the
 lotería.
 lottery
 ‘Juan₁ says that Mary₂ thinks that he₁/she₂ won the lottery.’

A priori, the data in this section suggests that Chilean Spanish is not a partial pro-drop variety, though the tendency to include overt pronouns in non-local environments needs to be explored. Thus, questions arise as to whether the results from this section can be reconciled with the results summarized in Table 1. One important contribution found in recent literature is that partiality in pro-drop properties is a matter of degrees (see Ticio 2018 and Frascarelli and Jiménez-Fernández 2019), and the data unveiled in this chapter can be interpreted in this light: The “overuse” of personal subject pronouns suggests that the Avoid Pronoun Principle is not as active in Chilean Spanish as one would think. This is not just a fact about first-person singular pronoun; rather the third-person singular is affected

as well. However, the most widely accepted partial pro-drop features – reviewed in this section – are not part of the grammar of this dialect.

4.2 *Other Issues and Future Research*

The syntactic features of Caribbean Spanish have been argued to be the result of language contact with English or African languages or African-based creoles and/or the corresponding demographics, but these are controversial theories (see Toribio 2000, Mayol 2012 and Pöll 2015 for discussion) as the distribution of the linguistic features that characterize these dialects does not necessarily correlate with historical language contact. Chilean Spanish instantiates one such case of dissociation between syntactic features and the said factors, arguing, thus, in favor of a more general approach. See Martínez Sanz (2011) and Sessarego (2021) for approaches based on language acquisition; see also Brown and Rivas (2011), Mayol (2012), and Camacho (2016) for an explicit discussion on how the change from null subject language to partial pro-drop may have taken place.

While various features of Chilean Spanish have been unveiled, the exact association of those features with sociolinguistic variables, such as level of education or age, is left for future research (see Sect. 2.2 for some relevant references). Also, potential variation within Chile has been left aside; the discussion has been exemplified with the PRESEEA Santiago de Chile data. Moreover, questions arise regarding the behavior of other Latin American varieties, particularly from the Bajeño area.

It worth noting that the present research – which sets the stage for a wider dialect distance analysis of subject properties in Spanish – has relied mostly on corpus data. Still, other sources of data (e.g., grammaticality judgment tasks) are needed to control for independent factor and analyze highly specific infrequent structures (e.g., see Ticio 2018 for various data points relevant to the canonical/non-canonical pro-drop status; for the relation between corpora and generative grammar, see Cornips and Gregersen 2016 and Newmeyer 2013).

5 Conclusion

The content of the present chapter has furthered our knowledge of cross-dialectal differences in subjecthood within Spanish as part of an ongoing dialectometric program focused on the syntax of Spanish. Specifically, the present study constitutes the first detailed characterization of the pro-drop properties of Chilean Spanish in light of the Null Subject Parameter (Rizzi 1982 and related discussions by Roberts and Holmberg 2010 and Barbosa 2019). It has been shown that Chilean Spanish patterns with Caribbean Spanish as opposed to Iberian Spanish in various respects, providing evidence that Chilean Spanish behaves as a non-canonical pro-

drop language – though not as a partial pro-drop variety – with respect to syntactic distribution of over subject pronouns. Furthermore, the pertinence of various data points for the study of pro-dropness has been evaluated.

Acknowledgments The research here has been done as part of the project research project PGC2018-093774-B-I00 of Spain's Ministry of Science, Innovation, and Universities (MICINN). I would like to thank Cilene Rodrigues and Andrés Saab for inviting me to participate in this volume and for their comments and suggestions. Marcela Hurtado and the audience of the Spanish Dialect Meeting V (Spadysin V), organized by the Universidad de Extremadura, Spain, also provided valuable feedback. All errors remain my own.

References

- Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16 (3): 491–539.
- Araya, Guillermo, Claudio Wagner, Constantino Contreras, and M. Mario Bernal. 1973. *Atlas lingüístico – etnográfico del Sur de Chile (ALESUCh)*. Santiago de Chile: Editorial Universitaria.
- Barbosa, Pilar P. 2019. *pro* as a Minimal nP: Toward a unified approach to Pro-Drop. *Linguistic Inquiry* 50: 487–526.
- Bolyanatz, Mariška A. 2020. Evidence for incomplete neutralization in Chilean Spanish. *Phonetica* 77: 107–130.
- Brown, Esther L., and Javier Rivas. 2011. Subject-verb word order in Spanish interrogatives: A quantitative analysis of Puerto Rican Spanish. *Spanish in Context* 8: 23–49.
- Camacho, José. 2016. The null subject parameter revisited: The evolution from null subject Spanish and Portuguese to Dominican Spanish and Brazilian Portuguese. In *The Morphosyntax of Portuguese and Spanish in Latin America*, ed. Mary A. Kato and Francisco Ordóñez, 27–48. Oxford: Oxford University Press.
- Cameron, Richard. 1997. Accessibility theory in a variable syntax of Spanish. *Journal of Pragmatics* 28: 29–67.
- Cardinaletti, Anna. 1997. Subjects and clause structure. In *The new comparative syntax*, ed. Liliane Haegeman, 33–63. New York: Longman.
- Cardinaletti, Anna, and Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In *Clitics in the Languages of Europe*, ed. Henk van Riemsdijk, 145–233. Berlin: Mouton de Gruyter.
- Carvalho, Janayna. 2019. Teasing apart 3rd person null subjects in Brazilian Portuguese. In *Romance languages and Linguistic Theory 15: Selected papers from 'Going Romance' 30*, ed. Ingo Feldhausen, Martin Elsig, Imme Kuchenbrandt, and Mareike Neuhaus, 238–254. New York: John Benjamins.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Kluwer Academic Publishers.
- Cifuentes, Hugo. 1981. La presencia y ausencia del pronombre personal sujeto en el habla culta de Santiago de Chile. *BFUCh* 31: 743–752.
- Comínguez, Juan Pablo. 2018. The nature and position of subjects in Puerto Rican Spanish wh-questions: Empirical evidence and theoretical implications. In *Current research in Puerto Rican linguistics*, ed. Melvin González-Rivera, 67–89. New York: Routledge.
- Contreras, Lidia. 1978. El pronombre acusativo en el español de Chile. *BFUCh* 29: 167–199.
- Cornips, Leonie, and Frans Gregersen. 2016. The impact of Labov's contribution to general linguistic theory. *Journal of Sociolinguistics* 20: 498–524.

- D'Alessandro, Roberta. 2015. Null subject. In *Contemporary linguistic parameters*, ed. Antonio Fábregas, Jaume Mateu, and Michael Putnam, 201–226. London: Bloomsbury Press.
- Fernández Lagunilla, Marina, and Alberto Anula Rebollo. 1995. *Sintaxis y cognición*. Madrid: Síntesis.
- Flores-Ferrán, Nidia. 2007. A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass* 1 (6): 624–652.
- Frascarelli, Mara. 2018. The interpretation of pro in consistent and partial null-subject languages: A comparative interface analysis. In *Null Subjects in generative grammar: A synchronic and diachronic perspective*, ed. Federica Cognola and Jan Casalicchio, 211–239. Oxford: Oxford University Press.
- Frascarelli, Mara, and Ángel L. Jiménez-Fernández. 2019. Understanding Partiality in pro-Drop Languages: An information-structure approach. *Syntax* 22: 162–198.
- Gilligan, Gary M. 1987. *A cross linguistic approach to the pro-drop parameter*. Ph.D. Dissertation, University of Southern California.
- Goodall, Grant. 2001. The EPP in Spanish. In *Objects and other subjects: Grammatical functions, functional categories and configurationality*, ed. William D. Davies and Stanley Dubinsky, 193–223. Dordrecht: Kluwer.
- Gupton, Timothy, and Sarah Lowman. 2013. An F projection in Cibeño Dominican Spanish. In *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, eds. Jennifer Cabrelli Amaro, Gillian Lord, Ana de Prada Pérez, and Jessi Elana Aaron, 338–348. Somerville: Cascadilla Proceedings Project.
- Herbeck, Peter. 2021. Perspectival factors and *pro*-drop: A corpus study of speaker/addressee pronouns with *creer* 'think/believe' and *saber* 'know' in spoken Spanish. *Glossa* 6 (1): 1–34.
- Hochberg, Judith G. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62: 609–621.
- Holmberg, Anders, and Michelle Sheehan. 2010. Control into finite clauses in partial null-subject languages. In *Parametric variation: Null subjects in Minimalist Theory*, ed. Theresa Biberauer, Anders Holmberg, Ian Roberts, and Michelle Sheehan, 125–152. Cambridge: Cambridge University Press.
- Hurtado Cubillos, Luz Marcela. 2009. La expresión de impersonalidad en el español de Chile. *Cuadernos de Lingüística Hispánica* 13: 31–42.
- Jaeggli, Osvaldo, and Ken Safir. 1989. *The null subject parameter*. Dordrecht: Foris.
- Lipski, John M. 1977. Preposed subjects in questions: Some considerations. *Hispania* 60: 61–67.
- . 1996. *El español de América*. 3rd ed. Madrid: Cátedra.
- MacDonald, Jonathan E., and Matthew L. Maddox. 2020. On impersonal se and null subjects in Romance. Paper presented at the Linguistic Symposium on Romance Languages 50, University of Texas at Austin.
- Maddox, Matthew. 2016. Null generic subjects in Spanish and the typology of null subject languages. Paper presented at the Linguistics Symposium on Romance Languages 46. Stony Brook University: March 31- April 3.
- Martínez-Lara, José A., Silvana Guerrero González, and Carlos González Vergara. 2021. (Yo) opino: la variación del sujeto pronominal en el habla de Santiago de Chile. In *Lingüística del castellano chileno: estudios sobre variación, innovación, contacto e identidad*, ed. Brandon Rogers and Mauricio Figueroa Candia, 283–308. Wilmington: Vernon Press.
- Martínez-Sanz, Cristina. 2011. *Null and Overt subjects in a variable system: The case of Dominican Spanish*. Ph.D. Dissertation, University of Ottawa.
- Mayol, Laia. 2012. An account of the variation in the rates of overt subject pronouns in Romance. *Spanish in Context* 9: 420–442.
- Mensching, Guido. 2000. *Infinitive constructions with specified subjects*. Oxford: Oxford University Press.
- Michnowicz, Jim. 2015. Subject Pronoun expression in contact with Maya in Yucatan Spanish. In *Subject Pronoun expression in Spanish: A cross-dialectal perspective*, ed. Ana M. Carvalho, Rafael Orozco, and Naomi Lapidus Shin, 101–119. Washington, DC: Georgetown University Press.

- Moreno Fernández, Francisco, and Hiroto Ueda. 2018. Cohesion and particularity in the Spanish dialect continuum. *Open Linguistic* 4: 722–742.
- Newmeyer, Frederick J. 2013. Goals and methods of generative syntax. In *The Cambridge Handbook of Generative Syntax*, ed. Marcel den Dikken, 61–92. Cambridge, U.K.: Cambridge University Press.
- Ordóñez, Francisco. 1997. *Word order and clause structure in Spanish and other romance languages*. Ph.D. Dissertation, City University of New York.
- Ordóñez, Francisco, and Antxon Olarrea. 2006. Microvariation in Caribbean/non-Caribbean Spanish interrogatives. *Probus* 18: 59–96.
- Orozco, Leonor. 2020. El papel de la interacción en la expresión de sujetos pronominales de segunda persona. *Cuadernos de Lingüística de El Colegio de México* 7: 1–40.
- Orozco, Rafael, and Luz Marcela Hurtado. 2021. A Variationist Study of Subject Pronoun Expression in Medellín, Colombia. *Languages* 6: 1–29.
- Ortega-Santos, Iván. 2016. *Focus-related operations at the right edge in Spanish: Subjects and Ellipsis*. Amsterdam: John Benjamins Publishing Company.
- . 2021. Dialect distance and data assessment in Chilean, Venezuelan and Puerto Rican Spanish. In *Lingüística del castellano chileno: estudios sobre variación, innovación, contacto e identidad*, ed. Brandon Rogers and Mauricio Figueroa Candia, 451–477. Wilmington, DC: Vernon Press.
- Ortega-Santos, Iván, Lara Reglero, and Jon Franco. 2019. Wh-Islands in L2 Spanish and L2 English: Between Poverty of the Stimulus and data assessment. *Fontes Lingvæ Vasconvm stvdia et documenta* 126: 435–471.
- Ortiz-López, Luis A., Ashlee Dauphinais, and Héctor Aponte Alequín. 2018. Cuban Spanish: Is it a Null Subject parameter Dialect? In *Cuban Spanish Dialectology: Variation, contact, and change*, ed. Alejandro Cuza, 97–118. Washington, DC: Georgetown University Press.
- Palacios, Azucena. 2016. Dialectos del español de América: Chile, Río de la Plata y Paraguay. In *Enciclopedia de Lingüística Hispánica*, ed. Javier Gutiérrez-Rexach, 330–340. New York: Routledge.
- Pöll, Bernhard. 2007. On the licensing of overt subjects in Spanish infinitival clauses. *Probus* 19: 93–120.
- . 2015. Caribbean Spanish = Brazilian Portuguese? Some comparative thoughts on the loss of pro-drop. *Studies in Hispanic and Lusophone Linguistics* 8 (2): 317–354.
- Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47: 253–291.
- Real Academia Española. 2010. *Nueva gramática de la lengua española*, 2 vols., Madrid: Espasa Calpe.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Roberts, Ian, and Anders Holmberg. 2010. Introduction: Parameters in Minimalist Theory. In *Parametric variation: Null subjects in Minimalist Theory*, ed. Theresa Biberauer, Anders Holmberg, Ian Roberts, and Michelle Sheehan, 1–57. Cambridge: Cambridge University Press.
- Rodrigues, Cilene. 2004. Impoverished Morphology and A-movement out of Case-Domains. Ph.D. Dissertation, University of Maryland, College Park.
- . 2021. *La gramática de los silencios: Balance y propuesta*. Ms., Universidad de Buenos Aires.
- Sedano, Mercedes, and Paola Bentivoglio. 2014. Uso del sujeto y del objeto directo en el habla de once ciudades hispanohablantes. In *Perspectives in the study of Spanish Language Variation: Papers in honor of Carmen Silva-Corvalán*, ed. Andrés Enrique-Arias, Manuel J. Gutiérrez, Alazne Landa, and Francisco Ocampo, 137–176. Santiago de Compostela: Universidade de Santiago de Compostela.
- Sessarego, Sandro. 2021. *Interfaces and domains of contact-driven restructuring aspects of Afro-Hispanic Linguistics*. Cambridge, MA: Cambridge University Press.
- Sheehan, Michelle. 2006. *The EPP and Null subjects in Romance*. Ph.D. Dissertation, Newcastle University, Newcastle upon Tyne.

- Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 38 (307): 344.
- Suñer, Margarita. 1986. Lexical subjects of infinitives in Caribbean Spanish. In *Studies in Romance Linguistics*, ed. Osvaldo Jaeggli and Carmen Silva-Corvalán, 189–203. Dordrecht: Foris.
- Ticio, Emma M. 2018. On Puerto Rican Spanish subjects. In *Current Research in Puerto Rican Linguistics*, ed. Melvin González-Rivera, 90–107. New York: Routledge.
- Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 110: 315–341.
- Valencia, Alba. 1995. *Chile*. Madrid: Arco Libros.
- Vidal, Manuel. 1980–1981. Los nexos adversativos en el habla culta de Santiago de Chile. *Homenaje a Ambrosio Rabanales BFUCh* 31: 963–977.
- Wagner, Claudio. 1996. Chile. In *Manual de dialectología hispánica: El español de América*, ed. Manuel Álvarez, 222–232. Barcelona: Ariel.
- . (dir.). 1997. *Atlas Lingüístico y Etnográfico de Chile (ALECh) por regiones*. <https://www.atlaslinguistico.cl/>. Accessed 1 Aug 2021.
- Zagona, Karen. 2002. *The Syntax of Spanish*. Cambridge: Cambridge University Press.

CORPORA

Corpus del Español NOW.

CREA= Real Academia Española, Corpus de Rereferencia del Español Actual.

PRESEEA = Corpus del Proyecto para el estudio sociolingüístico del español de España y de América.

Ways of Number Marking: English and Brazilian Portuguese



Roberta Pires de Oliveira

1 Introduction

Brazilian Portuguese (BrP) nominal system challenges both the nominal (Chierchia 1998) and the number parameters (Chierchia 2010, 2014, 2021) for the same reason: according to these models, a language that has number morphology, distinguishes mass and count nouns, may have Bare Plurals (BPs), but cannot have Bare Singulars (BSs).¹ Two solutions for explaining the grammaticality of BSs in BrP were explored in the literature: BSs are plural count nouns, or they are mass nouns. Assuming alternative theoretical perspectives, Schmitt and Munn (1999, 2002) and Müller (2002) agree that BSs are plural nouns for inclusive plurality, whereas BPs are exclusive, that is, the atoms are stripped out. Pires de Oliveira and Rothstein (2011) raised the view that BSs are mass noun. Experimental data do not support either of these views since, in quantity judgments, BSs are interpreted as counted and is measured by volume (Bevilaqua 2019). In response to these results, Rothstein and Pires de Oliveira (2020) argue that nouns in BrP are ambiguous between mass and count. In this chapter, I explore a different approach according to which BSs are underdetermined for grammatical atomicity (Rothstein 2017), that is, atomicity is not projected in the noun phrase. In a nutshell, I propose that English and BrP are Type I languages, but in English, the first layer of the nominal phrase projects atomicity, a grammatical operation that sorts the domain into atomic and plural sets of individuals, whereas in BrP, atomicity is projected by the “determiner.” The aim of this chapter is to develop this proposal and to investigate some of its consequences.

¹ It is not our aim to discuss the literature on the Bare Singular in BrP, see Ferreira (2021).

R. Pires de Oliveira (✉)

Universidade Federal de Santa Catarina/CNPq, Florianópolis, Brazil

The next section revises the data showing that English and BrP are Type I languages that differ with respect to the grammaticality of BSs in argument position – BrP allows it, English does not – and to the obligatoriness of the plural inflection in the noun: it is obligatory in English, but not in BrP. Section 3 shows that Chierchia (2010, 2021) cannot account for BrP and develops a modified version of his model, arguing that atomicity is projected by the noun in English and by the determiner in BrP. It also introduces two ways of modelling the determiner phrase in BrP. In Sect. 4, I take into consideration some consequences of this proposal.

2 English and BrP: Ways of Number Marking

Chierchia (2021: 22) develops “a theory of semantic variation, with a universal logical basis.” The chapter opens presenting three types of language, according to the parameter of number, and ends with an analysis where Type III languages cut across Type I and Type II. In this chapter, I focus only on Type I languages, though I come back to the issue of language variation in Sect. 4:

- (i) Type I are languages where numerals combine directly with some nouns, but not with others, which require insertion of a measure phrase. English is an example of Type I, because of the contrast between *three chairs* and **three blood(s), three ounces/drops of blood*.
- (ii) Type II are languages where numerals cannot combine with any nominal phrase (NP) directly. A classifier is always needed, independently of the noun being mass or count. Mandarin exemplifies this type.
- (iii) Type III are languages in which numerals freely combine with any type of noun. In combination with cognitively count nouns (e.g., *three cats*) numerals have the meaning they do in English. In combination with mass nouns, they have a “container” or a “quantity of” reading (not necessarily “standard” quantity of). Nez Perce is given as an example of Type III language, but along the chapter, the author argues that it should be analyzed as a subtype of Type I languages (I come back to this issue in Sect. 4).

Under this typology, BrP must be a Type I language, as I argue in the next section. However, as we will see, BrP is not like English, and it seems that the three other subtypes of Type I languages discussed by Chierchia (2021) are closer to BrP.

2.1 Nouns That Need a Measure Phrase, and Nouns That Don't

In BrP, there are some nouns that combine directly with numerals, (1a); and if they are combined with a measure phrase, the result is a marked construction, as in (1b), i.e., it is a construction that is not the usual way of saying in this language:

- (1) a. Comprei dois livro(-s).²
 bought.1SG two book(-PL)³
 ‘I bought two books.’
 b. # Comprei duas unidade(-s) de livro.⁴
 bought.1SG two unity(-PL) of book
 ‘I bought two unities of book.’

On the other hand, the direct combination of numerals with some other nouns is grammatically odd, as in (2a), whereas they are perfectly natural with a measure phrase, as in (2b):

- (2) a. # Comprei duas lama(-s).
 bought.1SG two mud(-PL)
 ‘I bought two muds.’
 b. Comprei duas bacia(-s) de lama.
 bought.1SG two bowl(-PL) of mud
 ‘I bought two bowls of mud.’

From examples (1) and (2), we can conclude that the numeral *dois/duas* (“two”) distinguishes mass and count nouns.⁵ There is no consensus in the literature on the theoretical definition of mass nouns. Descriptively, it is the “class” of nouns that do not naturally combine with numerals directly. Chierchia (2010, 2021) distinguishes things that we know that have stable atoms, that is, we know that their unity is stable across different situations, from the class of nouns, mass nouns, about which we know that their atoms are too vague to be counted, leaving always open the possibility of having an alternative unity. We might see a certain portion of water as one unity, but we can also imagine other partitions: “The ‘minimal components’ of mass nouns are specified too vaguely to be used in counting [...]” (Chierchia 2021: 22). I come back to this issue in the next sections.

At any rate, the examples above show that BrP patterns with English. Both languages share the main property of Type I languages: numerals distinguish count and mass nouns.

² (-s) indicates that the plural inflection in the noun is optional.

³ Glosses follow the Leipzig Glossing Rules and are restricted to the noun phrases, the focus of this paper. PL is plural.

⁴ # is used to signal that the sentence needs contextual cues to be interpreted.

⁵ As it is already known in the literature, there are ways of shifting a count noun into a mass noun as in *There is apple in the salad* (Link 1983, Pelletier 1975, among others), and in the opposite direction as well, mass nouns can be shifted to count, as in *Comprei dois vinho(-s)* (“I bought two wines”).

2.2 *Plural Inflection in the Noun and Not in the Noun*

The typology offered by Chierchia (2021) represents an improvement over his 2010 characterization, where number marking languages are defined by the property of having number features obligatorily in the noun: “This contrasts with what happens in number marking languages. I refer by this label to languages that have overt number features which obligatorily appear on nouns (and may appear to other components of the noun phrase)” (Chierchia 2010: 108). Although this is a property of English noun phrases, since number morphology is obligatorily marked in the noun, even if it is also marked in the determiner (e.g., *many boys*, **many boy*),⁶ it is not a property of BrP, where the plural inflection is optional in the noun.

In BrP, plural inflection distinguishes mass nouns since it does not combine with them naturally, as we saw in the last section. However, it is optional with count nouns, though it is obligatory in the determiner. In (3a), the absence of overt inflection in the nominal phrase is obligatorily interpreted as denoting an atomic individual; in (3b), the plural definite phrase is interpreted as about a sum of individuals.⁷ The ungrammaticality of (3c) shows that the plural inflection is obligatory in the determiner:

- (3) a. Compri o livro.
 bought.1SG the.SG book
 ‘I bought the book.’
- b. Compri o-s livro(-s)
 bought.1SG the-PL book(-PL)
 ‘I bought the books.’
- c. * Compri o livro-s
 bought.1SG the-SG book(-PL)

In English plural inflection is obligatory in the noun, whereas in BrP, it is obligatory in the determiner. The absence of inflection in the determiner is semantically equivalent to singularity, but not so in the noun.

The lack of overt inflection in the noun in BrP does not mean singularity, as exemplified in (4), where the BS is interpreted as plural; the next sections show that the absence of plural inflection in the noun in English is always singular.

⁶ Expressions like *a few books* where a plural noun seems to combine with a singular *a* are treated as having a singular empty noun. See Kayne (2005).

⁷ It is not the aim of this chapter to discuss the semantics of plural inflection, but as in other languages, it must include the atoms which are sums of themselves. See Chierchia (2010) for arguments that English plural is inclusive (against Chierchia 1998) and Pires de Oliveira (2019) for arguments that in BrP, the plural is inclusive. See also De Swart and Farkas (2010).

- (4) Comprei livro
 bought.1SG book
 ‘I bought books.’

Moreover, only the BS allows for mass interpretation, which led Pires de Oliveira and Rothstein (2011) to the claim that they are mass. If the BS in BrP were a plural predicate, it should not give rise to mass readings. However, questions as (5a) in quantity judgement tasks give rise to mass answers; (5b) in English are always measured:

- (5) a. Quem tem mais livro?
 who has more book
 b. Who has more book?

In quantity judgements tasks, participants answer questions as (5) choosing either a scenario that has fewer number of objects but with greater volume or a scenario with more objects and less volume.⁸ Brazilians interpreting (5a) oscillate between counting and measuring, with a tendency for counting. English speakers always go for volume with (5b).

In conclusion, although in BrP and in English plural inflection and numerals distinguish count and mass, in BrP, in contrast with English, plural inflection is optionally marked in count nouns. The noun in BrP sometimes looks like a singular predicate (3a), sometimes like a cumulative predicate, as in (4), and sometimes it allows both cardinal and mass readings, as in (5). A straightforward solution is to take BrP nouns to be ambiguous, as suggested by Rothstein and Pires de Oliveira (2020). I discuss this solution in the next sections.

2.3 *Bare Singulars in Argument Position: Grammatical or Not*

Chierchia’s (1998) *Nominal Parameter* distinguishes languages according to the syntactic “status” of bare nouns, whether these nouns function as arguments, as predicates, or as both. Mandarin is an argument language, since the noun is always bare in argument position, because there are no overt determiners, [+arg, –pred]. French, on the other hand, is a predicate language, because the noun is never an argument, but always a predicate, [–arg, +pred]. English is a language where nouns can be an argument and a predicate, [+arg, +pred]. Given this typology, once again

⁸ Quantity judgments play an important role in the investigation of the semantics of nouns since Barner and Snedeker (2005).

BrP patterns with English, since bare nouns can be in both positions, as an argument in (6a) and as a predicate in (6b):

- (6) a. Cachorro-s latem
 dog-PL bark
 ‘Dogs bark.’
- b. Os cachorro(-s) latem
 the-PL dog(-PL) bark
 ‘The dogs bark.’

However, BrP differs from English, since besides BPs, exemplified in (6a), BSs are also grammatical in argument position, as witnessed by (7a), in contrast with (7b), which is ungrammatical in English:

- (7) a. Gato tem bigode.
 cat has whisker
 ‘Cats have whiskers.’
- b. * Cat has whisker.

Schmitt and Munn (1999) were the first to notice that the BrP BSs were a challenge to Chierchia’s (1998) model, since it cannot generate a language that has plural inflection, BPs, and BSs.

Moreover, in the contexts where BSs are acceptable in English, a difference in the interpretation of the noun phrase arises. In (8a), for instance, we have the famous universal grinder (Pelletier 1975). The literature claims that *cat* is coerced into mass and interpreted as “smashed” cats. Crucially, it cannot be about individual cats. By contrast, (8b), from BrP, is true in a situation where there are cats all over the place and in a situation where there is “smashed” cat:

- (8) a. There was cat all over the place.
- b. Tinha gato em tudo quanto é lado.
 had cat in every much/many is side
 ‘There was cat all over the place.’ or ‘There were cats all over the place.’

The context in (9a) was discussed by Link (1983) as a case where the count noun is transformed into a mass noun. In his description of (9a), *apple* in English is a singular predicate that is transformed into mass because it is in an argument position, a place it cannot occupy. The literature has treated cases like (9a) as coercion from count to mass (see Chierchia 2010 and Rothstein 2017, among others). By contrast,

(9b), besides a mass interpretation, has the pragmatically implausible reading of units of apples:

- (9) a. There is apple in the salad.
 b. Tem maçã na salada.
 has apple in.the salad
 ‘There is apple in the salad’ or ‘There are apples in the salad.’

English and BrP are Type I languages, but in English, plural inflection is obligatory in the noun, whereas in BrP, it is optional in nouns and obligatory in the determiner. In BrP, the lack of number inflection in the noun is interpreted sometimes as singular, as plural, or as mass. In contrast, in English, BSs are singular. BSs are readily grammatical in BrP, but not in English. Finally, while in English count BSs must be coerced into mass, in BrP, BSs allow for both mass and count interpretations.

3 A Model Theoretic Approach to Language Variation

My goal in this section is explaining the data discussed so far. In 3.1, I show that Chierchia (2021) generates English but does not generate BrP; in 3.2, a proposal to derive both English and BrP is presented.

3.1 Chierchia’s Semantic Parameters

It is beyond the scope of this chapter to revise the theoretic approaches developed by Chierchia (1998, 2010, 2014, 2021), not only because they are slightly different but also because they get into issues that are not the focus of the present chapter. However, the data presented above on BrP BSs are a challenge to any of those versions, given that in all of them, count nouns without inflection (e.g., *book*, *bear*, *table*) are atomic predicates, and as such cannot be in argument position. The author is not always crystal clear about the mapping between linguistic forms and the model theoretic approach, but he is very clear about singular count nouns. In his 2021 paper, he explicitly assumes that a count noun without inflection (e.g., *bear*) denotes the set of (absolute) atomic individuals: “For singular properties like *bear*, it holds that $AT(bear) = bear$ ” (Chierchia 2021: 27).

In his approach, the mass/count distinction is a consequence of pressure from the cognitive apparatus, which is external to the language faculty. This cognitive apparatus distinguishes between objects and substances. Languages may vary in the way they actualize the logical schema, for instance, some languages have plural

inflection, others do not, but the schema is universal: “My inclination is to regard Base-line theories as a ‘schematism’ that supervenes on our natural computational capacities (i.e., something like ‘merge’ and ‘copy-merge’ or the untyped λ -calculus) under the pressure of a pre-existing conceptual space of ‘substances’ and ‘objects’” (Chierchia 2021: 36). There are some mismatches, as it is the case for fake mass nouns, such as *furniture*, but they are explained away as exceptions. Thus, the ontology is sorted into nouns that denote structures built out of stable atoms and those that are built from non-stable atoms. The prediction is that the mass/count distinction shows up in any language.

Chierchia suggests that language variation can be explained by a bifurcation at a very low level of derivation. Relying on ideas from distributed morphology, he understands that the “root” of a noun like *woman*, represented by $\sqrt{\text{woman}}$, denotes woman-atoms, represented by woman_c , woman-sums, WOMAN_c , and woman-kind, $\cap \text{WOMAN}_c$, as shown in (10a). At a very low level, the nominal stem both in English, a property-oriented language, $\sqrt{\text{woman}}$, and in Mandarin, a kind-orientated language, $\sqrt{\text{nürén}}$, has the same denotation, (10a). (10b) is the derivation of the noun in Mandarin and (10c) in English.

(10) Noun interpretation

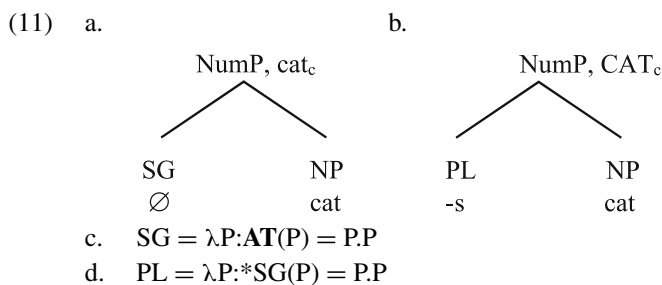
- a. $[[\sqrt{\text{woman}}]] = [[\sqrt{\text{nürén}}]] = \{\text{woman}_c, *$
 $\text{woman}_c (= \text{WOMAN}_c), \text{woman}_{k,C} (= \cap \text{WOMAN}_c)\}$
- b. Kind orientation
 $n, \lambda x_k. x_k ([[\sqrt{\text{nürén}}]]) = \cap \text{WOMAN}_c$
- $n, \lambda x_k. x_k$ $\sqrt{\text{nürén}}, [[\sqrt{\text{nürén}}]]$
- c. Property orientation
 $n, \lambda P_{\langle s, e \rangle}. P_{\langle s, e \rangle} ([[\sqrt{\text{woman}}]]) = \{\text{woman}_c, \text{WOMAN}_c\}$
- $n, \lambda P_{\langle s, e \rangle}. P_{\langle s, e \rangle}$ $\sqrt{\text{woman}}, [[\sqrt{\text{woman}}]]$
- d. i. Kind-oriented languages: $n \Rightarrow \lambda x_k. x_k$
 ii. Property-oriented languages: $n \Rightarrow \lambda P_{\langle s, e \rangle}. P_{\langle s, e \rangle}$
 (Chierchia 2021: 53)

In the nominal projection, little-*n* is the point of bifurcation between English and Mandarin: “The semantic role of little-*n* is similar to the role of functional heads: it introduces a (language specific) presupposition. This presupposition determines the type of the noun denotation that part takes into further semantic compositions higher up in the syntactic derivation: a property or a kind” (Chierchia 2021: 53). Type I languages are property-oriented languages, type $\langle e, t \rangle$ (see (10c)), whereas classifier languages such as Mandarin are kind-oriented languages, type *e* (see (10b)). Thus,

property-oriented languages select for properties, whereas Type II languages are kind-oriented, selecting for kinds. Here I restrict my attention to Type I languages.

Chierchia distinguishes three subtypes within Type I: English, Greek, and Nez Perce. These subtypes reflect the different behavior of plurality and numerals. In Greek, mass nouns combine systematically with plural inflection, but not with numerals. In Nez Perce, numerals combine with any “class” of nouns, and plural inflection combines with mass nouns.⁹ In this typology, BrP patterns with English.

His model theoretic proposal assumes, as it is common ground in contemporary semantics, that the domain U of individuals is ordered by a relation of “part of,” represented by “ \leq ” and closed under the sum operation, represented by “+.” In English, the absence of inflection, represented as SG, is an identity from an atomic predicate, generated by **AT**, the stable generator, to an atomic predicate, as in (11c). The result is the set of stable atoms:¹⁰ “only count properties have stable generator sets, for which I will use the boldface **AT(P)**” (Chierchia 2021: 30) Plural is also identity, but it applies to a set of sums of stable atoms, as represented in (11d):

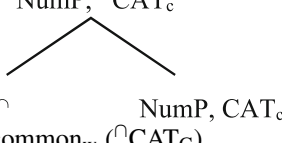


Notice that singular and plural are treated as presuppositions, that is, they restrict the domain of the function: “The presupposition of singular NPs is that they are true of P-atoms (where P is the basic denotation of the N-head); the presupposition of plural NPs is that they are sum-closed” (Chierchia 2021: 40). Number morphemes are then identity functions. Hence, *bear* is the atomic predicate and *bears* corresponds to the plural predicate: “the singular NP *cat* winds up denoting the atomic property cat_C and the plural NP *cats* the sum closed property CAT_C ” (Chierchia 2021: 40). The plural inflection corresponds to the star * operation in (11d).

⁹ They also differ with respect to the availability of “fake mass” nouns as *furniture*. This is a topic in itself and will not be discussed in this chapter.

¹⁰ The author makes a difference between $AT(P)$ and $\mathbf{AT}(P)$, where the first generates relative atoms. This difference is irrelevant for our purposes.

Since Chierchia (1998), the analysis of BPs goes along the same line of reasoning: it assumes, as Carlson (1977), that BPs always denote kinds but introduces the idea of type shifting and the constraint that it applies to cumulative denotations. In (12a), for instance, *cats* denotes the kind *cat*, as the derivation in (12b) shows:

- (12) a. Cats are common.
 b. NumP, \cap CAT_c

 c. common_w (\cap CAT_C)
- (Chierchia 2021: 33)

In (12b), the Number Phrase (NumP) is the outcome of applying plurality to the stem, as represented in (11b). It denotes the plural predicate, represented by capital letters, CAT_c. In argumental position, the predicate is shifted to a kind, via the down operator, represented by \cap , and the logical form in (12c) is, thus, derived. Down applies as a last resource only to cumulative predicates, that is, it is not defined for atomic predicates. This guarantees that BSs are ungrammatical in English, as exemplified in (7b) – **Cat has whisker* – given that atomic predicates are not cumulative. Note that it is implicitly assumed that there must be a NumP in the noun phrase in English.

This story accounts for English nicely, but it leads to the incorrect prediction that BSs are ungrammatical in BrP.

3.2 BrP Bare Singulars in a Model Theoretic Perspective

As presented above, Chierchia’s model leads to the wrong prediction that a count predicate without inflection in BrP is singular and that BSs are ungrammatical. One might keep the same description offered for English, arguing that in BrP, the noun without inflection is in fact plural, as Schmitt and Munn (1999, 2002) and Müller (2002) have done. In such an approach, *gato* (“cat”) is sometimes a plural predicate, a variant of *gatos* (“cats”), and sometimes a singular one. However, if BS were a plural predicate, the prediction, contrary to the facts, is that it should behave exactly as the bare plural. Experimental data shows that this is not the case: the BP is always counted, whereas the BS is sometimes counted, sometimes measured (Bevilaqua 2019). The examples in (13a) and (13b), counterparts of (8b) and (9a), respectively, do not allow for a mass interpretation. Thus, BPs give more information; they must be about sums of individuals. Sentences with BSs can be about individuals, but they can also be about smashed cats and apple stuff, as shown in Sect. 2.3.

The plural counterparts of (4) and (5), in (14a) and (14b), respectively, are not defined for situations where one is comparing non-cardinal dimensions:

- (13) a. Tem gato-s em tudo quanto é lado.
 has cat-PL in every much/many is side
 ‘There are cats everywhere.’
 b. Tem maçã-s na salada.
 has apple-PL in.the salad
 ‘There are apples in the salad.’
- (14) a. Comprei livro-s.
 bought.1SG book-PL
 ‘I bought books.’
 b. Quem tem mais livro-s?
 who has more book-PL?
 ‘Who has more books?’

The sentences in (14) are true in situations where the cardinality of books must be necessarily above an average quantity of individual books. In (4) and in (5), however, the BS allows for a situation that may involve comparing the weight or volume of books, independently of the cardinality. Thus, BSs in BrP are not a variant of BP, it says less, because it might be true in situations where cardinality is not involved.

Pires de Oliveira (2020) argues that the BS in BrP does not convey grammatical information about atomicity. It is underdetermined, following a suggestion made in Pelletier (2012) that nouns in the lexicon are both mass and count. In what follows, I develop this idea. Although my proposal relies on Chierchia, it has small differences to account for BS in BrP. As already mentioned, BrP patterns with English except for the grammaticality of BSs and for the optionality of inflection in the noun. These facts are connected, as I show below.

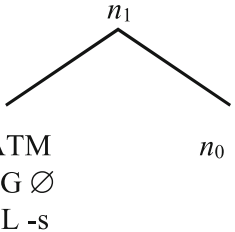
I assume that the root is uncategorized, denoting, thus, any realization both as noun and verb. The root $\sqrt{\text{mulher}}$ (“woman”) includes individuals, sums of individuals, portions, stuff, and events (*mulherizar*, to get more women, for instance¹¹). This is polysemy. The first layer of the nominal phrase turns the uncategorized root into a noun.¹² The nominalization kicks the event-related meanings out of the denotation and labels the noun as mass or count, if the distinction is active in the language. Thus, the nominal stem denotes all possible dimensions of a noun, atoms, sums, portions, and stuff. However, the projection of *n* organizes the domain, by labelling the noun mass or count. Thus, $[[n (\sqrt{\text{cat}})]]$ denotes singular cats, and sums

¹¹ From a Google search *Como mulherizar a organização?* (“How to have more women in the organization?”)

¹² This operation may correspond to the insertion to the thematic vowel in BrP, a topic I leave aside.

of cats, leaving cat portions, and cat stuff in the background, that is, they might be accessed, but there is an asymmetry introduced by the label.

In English, n projects atomicity. Thus, immediately after n is projected, another nominal layer is active in English, call it n_1 , where atomicity, a grammatical operation, sorts the domain into a set of atoms or a set of sums. If atomicity applies at the n layer, then the denotation of n_1 is a predicate that has grammatical atoms (Rothstein 2010, 2017). Atomicity is either singular or plural and corresponds to NumP in the syntax. We treat SG and PL as corresponding to the semantic operations of generating atoms and sums, respectively, but they might be seen as presuppositions, as Chierchia does:

- (15) a. 
- b. $\llbracket \text{SG} \rrbracket = \lambda P. P(x) \wedge \forall y(P(y) \wedge y \leq x \rightarrow y=x)$
- c. $\llbracket \text{PL} \rrbracket = \lambda P. P(x) \rightarrow \exists Y \subseteq P \wedge x = \cup Y$ ¹³

Singular, represented by SG, and corresponding to a null inflection, \emptyset , takes a predicate and returns the set of atoms, that is, an individual that if it has a part, the part must be itself. PL, corresponding to plural, takes a predicate and returns the sums, including the atoms.

If the predicate is an atomic, the shift to the kind cannot happen, because it is not cumulative and the derivation crashes. This explains the ungrammaticality of BSs in English, exemplified in (7b). In English, BSs surface in argument position only if some mechanism for rescuing the derivation is in place, coercion into a mass predicate, for instance. In (8a), the predicate *cat* is atomic; then, the only alternative is to coerce it into a mass denotation, cat-stuff, which is cumulative. (9a) is a food context; thus, *apple* must be reanalyzed as apple stuff (Frisson and Frazier 2005), again a cumulative predicate. Crucially, BSs cannot be in an argument position because they are atomic predicates, and they are so, in English, because atomicity is computed immediately after n . If they are atomic predicates, placed in cumulative contexts, they are massified. So, the interpretation of questions as (5b) is explained. BPs are grammatical because they are cumulative and down shifts the predicate into a kind. This is the same story we find in Chierchia, with the introduction of a level of atomicity immediately above n_0 .

¹³ There are different ways of defining plurality. This definition is close to Link's (1983).

Let us move to BrP. As in English, the uncategorized root turns into a noun by being merged with *n*, which labels the lexical item as count or mass. The denotation of nominal stem is as in English: atoms, sums, portions, and stuff; as in English, the label count backgrounds the portion and the stuff interpretations. However, the first layer in the nominal phrase, *n*, does not project atomicity. Thus, there is no grammatical information about atomicity. This contrasts with English. It explains why plural inflection is not obligatory in the noun in BrP, whereas it is in English. Moreover, since *n* in BrP does not compute atomicity, it may occur in argumental position, where the cumulative predicate is shifted into the kind:

- (16) a. Gato é comum no Brasil.
 cat is common in.the Brazil
 ‘Cats are common in Brazil.’
 b. $[DP \cap n_0 (\sqrt{\text{gat-}})]$

Note that there is no number projection in the noun phrase. The noun is underdetermined; thus, the predicate is cumulative, Down applies, and the result is the cat kind.

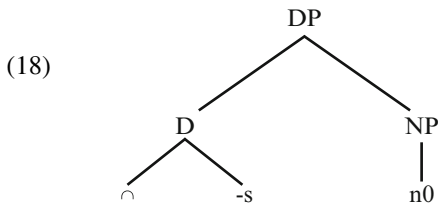
The meaning oscillation in the examples (5), (6), (7), (10b), and (11a) is explained as polysemy, which is allowed because of the absence of atomicity or NumP: BSs in BrP can be interpreted as cardinal, singular, or mass because there is no atomicity immediately after *n*. This explains why BrP BSs behave like mass nouns, as pointed out by Pires de Oliveira and Rothstein (2011); however, this is not because the BS is mass as the authors argue nor because the noun is ambiguous in BrP, as suggested by Rothstein and Pires de Oliveira (2020), but rather because there is no grammatical atomicity. The tendency to interpret the BS as cardinal interpretation is explained because the noun is labelled count; thus, objects are foregrounded. (8b) is by default interpreted as about cats; the smashed reading pops up only biased by context. However, the absence of atomicity let the interpretation free. Thus, there is no coercion to mass.

This analysis differs from those put forward in Schmitt and Munn (1999, 2002), Taveira da Cruz (2008), and Cyrino and Espinal (2011), though it shares with them the intuition that NumP is not projected in BrP BSs. These previous analyses offer different implementations, but they all share the idea that the noun phrase in the examples in (17) are syntactically different: the example (17a) involves NumP, whereas (17c) does not, as the representations sketched in (17b) and (17d) indicate:

- (17) a. Comprei o livro.
bought.1SG the book
'I bought the book.'
- b. [DP o [NumP SG [N livro]]] (= (21a))
- c. Livro serve pra ler.
book serves for read.INF
'Books are for reading.'
- d. [DP [N livro]]

Thus, *livro* is ambiguous between a singular predicate, in (17a), and a non-singular predicate in (17c). In that line of reasoning, the BP in BrP, exemplified below, in (19a) has a number projection. It seems, then, arbitrary when there is NumP and when not. In the end, the existence of BS in BrP is treated as an exception, a spurious DP. If, on the other hand, in BrP, atomicity is projected by the determiner, we explain the data without postulating ambiguity, and, more importantly, we have a principled way to explain number. I explore this avenue where the syntactic forms of the sentences in (17a) and (17b) are not as represented above in (17b) and (17d).

The core idea (See Pires de Oliveira 2021) is that in English, atomicity is projected by the noun (i.e., it is a noun-centered language), which explains why it does not allow BSs in argument position. In a determiner-centered language, atomicity is projected by the determiner. I am aware of the numerous issues that this proposal raises, but I also believe it is worthy of developing it. Consider, first, the BP in BrP. Plural inflection is attached to the determiner, as represented below:



Notice that there is no NumP, the plural inflection modifies the determiner that shifts to a kind when in argument position, as represented in (19b). Again, note that there is no NumP:

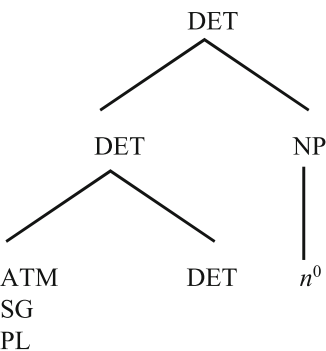
- (19) a. Gatos são comuns.
cats are common
- b. [DPⁿ n₁ -s [n₀ √N]]

The main difference with English is that in English, atomicity is projected at the first layer on of the noun phrase, whereas in BrP, the plural inflection modifies the determiner. The result is the same in each language, but the place where atomicity acts is not.

Moving to the definite article, assume the classical view, where it corresponds to the ι operator, $\langle\langle e,t\rangle,e\rangle$; it combines with a predicate that is true of just one individual in the context, i.e., it is a singleton, and the outcome is an individual.¹⁴ As we saw in (3), there is a semantic difference between the definite singular *o* and the definite plural *o-s*, repeated in (20) for convenience: (20a) is about a singular book. If the noun, *livro*, is a cumulative predicate in both realizations in (20), then singularity, which is obligatory for (20a), must come from the definite article:

- (20) a. Comprei o livro.
bought.1SG the.SG book
'I bought the book.'
- b. Comprei o-s livro(-s).
bought.1SG the-PL book(-PL)
'I bought the books.'

In our approach, represented below in (21a), atomicity is attracted by the determiner and is computed as a presupposition that restricts the noun. The definite article carries the presupposition of unicity, and a singular or a plural restriction is added. The description is given in (21b) for the singular and in (21c) for the plural:

- (21) a.
- 
- ```

graph TD
 DET1[DET] --- DET2[DET]
 DET1 --- NP[NP]
 DET2 --- ATM[ATM]
 DET2 --- DET3[DET]
 NP --- n0[n^0]
 ATM --- SG[SG]
 ATM --- PL[PL]

```
- b.  $\lambda P: \text{ATM}(P) \wedge P$  is a singleton.  $e_{\text{singular}}$
- c.  $\lambda P: * \text{ATM}(P) \wedge P$  is a singleton.  $e_{\text{plural}}$

<sup>14</sup> See, for instance, Heim and Kratzer (1998).

In (21b), the function applies to predicates that are atomic and a singleton, and the outcome is the only singular individual salient in the context, represented by  $e_{\text{singular}}$ , as in (20a). In (21c), the function is restricted to a plural predicate that is a singleton in the context, and the outcome is a plural individual,  $e_{\text{plural}}$ , as in (20b).

ATM does the same job. In BrP, the noun agrees with the determiner, whereas in English, the determiner agrees with the noun. At any rate, both English and BrP are generated as Type I languages.

## 4 Ways of Pushing It Even Further

I argued that although BrP and English are Type I languages, they differ with respect to the component that projects atomicity. In English, the first nominal layer projects atomicity, but in BrP, the determiner does so. This explains the contrast between these two languages with respect to the grammaticality of BSs and the obligatoriness or optionality of plural inflection in the noun. English is a noun-centered language, whereas BrP is a determiner-centered language.

It predicts different results with respect to sentence comprehension. Atomicity introduces the grammatical information that the noun is atomic. Thus, the BSs in English, if in argument position, should trigger a mass interpretation via coercion, which means that speakers take longer to process it because the semantic processor must revise the hypothesis about atomicity and to apply a mass function such as grinding. This is not true of BSs in mass contexts in BrP. If BrP BSs are underdetermined, and if there is no projection of atomicity, then they should not take longer to be processed. One might also empirically verify the claim about atomicity in the determiner versus atomicity in the noun.<sup>15</sup> If this piece of grammatical information is computed in different places, then it should be possible to “see” this difference in the behavior of speakers of these two languages. English speakers only get this information in the noun, *the boy*, whereas BrP speakers get this information with the article, *o menino*.

Moreover, if the proposal is on the right track, then there must be languages like BrP, that is, Type I languages that have both BSs and BPs, and obligatorily mark plural inflection in the determiner, but it is optional in the noun. Rikbaktsa, a Macro-Gê language, spoken in Brazil, seems to be one of these languages. Dellai et al. (2021) show that Rikbaktsa has no articles, definite or indefinite, but does have plural and mass morphology; plural morphology is not obligatorily marked in the noun, although it is in demonstratives. Thus, in Rikbaktsa, nouns without inflection, BSs, are not necessarily singular as it is the case in BrP. Again, there is a lot to be better understood, and we certainly need careful data in ideally all languages, but in particular in underrepresented ones.<sup>16</sup>

<sup>15</sup> Pires de Oliveira and Bezerra [in progress](#).

<sup>16</sup> Pires de Oliveira et al. (2021) ongoing research on (in)definiteness across languages.

The approach developed in this chapter explains the subtypes of Type I languages proposed in Chierchia (2021). We predict that both Greek and Nez Perce are languages that allow for BSs and BPs, that is, atomicity does not have to be immediately after *n*. Indeed, this seems to be the case for Greek (Lazaridou-Chatzigoga and Alexandropoulou 2013). In that respect, they pattern with BrP. We also predict that plural inflection is not obligatory in nouns in these languages. Greek systematically allows for the plural inflection with both mass and the count nouns, although the interpretation is not the same. This is an indication that atomicity is not in the first layer of the nominal derivation; if it were, this should not happen. Thus, instead of proposing, as Chierchia does, that Greek has two types of plurality, I propose that atomicity in Greek is not in the noun, and plural freely applies to mass and count. In Greek numerals distinguish the two nominal classes; thus, at this layer of the derivation, it must be the case that atomicity is playing a role, probably projected by the numeral. In Nez Perce, plural inflection and numerals are insensitive to the mass and count distinction. This might be evidence that atomicity is not projected by the noun, nor by plural inflection, nor projected by the numeral. The prediction once again is that this language has BSs and plural inflection is not obligatory in the noun, a topic waiting for future research.

## 5 Conclusion

The chapter investigated noun phrases in English and in BrP and argued that there is a minimal difference between them: the locus of atomicity. English is a noun-centered language and BrP a determiner-centered one. The first layer of the nominal projection in English projects atomicity, the semantic counterpart of NumP in syntax, whereas in BrP it does not. The absence of morphology in English is, then, always singular, that is, atomic. If the BS is atomic, it cannot be shifted to an individual by down, and the derivation either crashes or the noun is re-interpreted as mass, which is cumulative. In BrP, the BS carries no grammatical information about atomicity; thus, not only it may be shifted to the kind, but it also receives different interpretations in quantity judgment tasks. This is a micro-variation since both are Type 1 languages in Chierchia's typology. I have argued that the proposal put forward in this chapter gives a better explanation for the different subtypes of Type 1 languages. Moreover, I argued that Rikbatska might be a determiner-centered language, though it does not have articles. Finally, the proposal predicts processing differences.

More empirical data need to be taken into consideration for a better understanding of nominal phrases across languages. It is my hope that this investigation into the nominal domain helps us achieve a better understanding of language variation.

**Acknowledgments** Many thanks to Cilene Rodrigues and Andrés Saab for their careful reading. All the mistakes are my own responsibility.

## References

- Barner, David, and Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97 (1): 41–66.
- Bevilaqua, Kayron. 2019. *A semântica dos sintagmas nominais através das línguas: estudos experimentais sobre a distinção contável-massivo*. PhD dissertation in Letras, Ciências Humanas da Universidade Federal do Paraná. Curitiba. Available at: <https://acervodigital.ufpr.br/handle/1884/66270>.
- Carlson, Gregory N. 1977. *Reference to kinds in English*. PhD dissertation, University of Massachusetts, Amherst. Available at: <https://scholarworks.umass.edu/dissertations/AAI7726414/>.
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language and Semantics* 6: 339–405.
- . 2010. Mass nouns, vagueness and semantic variation. *Synthèse* 174: 99–149.
- . 2014. How universal is the mass/count distinction? Three grammars of counting. In *Chinese syntax: A cross-linguistic perspective*, ed. Audrey Li, Andrew Simpson, and Wei-Tien Dylan Tsai, 147–177. Oxford: Oxford University Press.
- . 2021. Mass vs. Count: Where do we stand? Outline of a theory of semantic variation. In *Things and stuff: The semantics of the count-mass distinction*, ed. Tibor Kiss, Jeff Pelletier, and Halima Husic, 21–54. Cambridge: Cambridge University Press.
- Cyrino, Sonia, and Maria Teresa Espinal. 2011. *Object BNs in Brazilian Portuguese. More on the NP/DP analysis*. Paris: Le neuvième Colloque de Syntaxe et Sémantique à Paris.
- de Swart, Henriette, and Donka Farkas. 2010. The semantics and pragmatics of plurals. *Semantics and Pragmatics* 3: 1–54.
- Dellai, Érica, Vitória Maria Jasperern, Léia de Jesus Silva Roberta Pires de Oliveira, Beatriz Martins Rachadel, and Bianca Maria de Souza. 2021. A distinção massa e contável na gramática Rikbaktsa (Macro-Jê). *LIAMES: Línguas Indígenas Americanas* 21: 1–18. Available at: <https://periodicos.sbu.unicamp.br/ojs/index.php/liames/article/view/8661408>.
- Ferreira, Marcelo. 2021. Bare Nominals in Brazilian Portuguese. In *Oxford handbook of grammatical number*, ed. Patricia Cabredo Hofherr and Jenny Doetjes. Oxford: Oxford University Press.
- Frisson, Steven, and Lyn Frazier. 2005. Carving up word meaning: Portioning and grinding. *Journal of Memory and Language* 53: 277–291.
- Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell.
- Kayne, Richard. 2005. Silent years, silent hours. In *Movement and silence*, 241–260. New York: Oxford University Press.
- Lazaridou-Chatzigoga, Dimitra, and Stavroula Alexandropoulou. 2013. A corpus study of Greek bare singulars: Implications for an analysis. *Revista da ABralin* 12 (1).
- Link, Godehard. 1983. The logical analysis of plurals and mass terms. In *Meaning, use, and the interpretation of language*, ed. Rainer Bäuerle, Arnim von Stechow, and Christoph Scherzke, 302–323. Berlin: Mouton de Gruyter.
- Müller, Ana Lúcia. 2002. Genericity and the denotation of common nouns in Brazilian Portuguese. *D.E.L.T.A.* 18: 287–308.
- Pelletier, F. Jeffry. 1975. Non-singular reference: Some preliminaries. *Philosophia* 5 (4): 451–465.
- . 2012. Lexical nouns are neither mass nor count, but they are both mass and count. In *Count and mass across languages*, ed. Diane Diane Massam, 9–26. Oxford: Oxford University Press.
- Pires de Oliveira, Roberta. 2019. O plural no português brasileiro e no inglês: comparando através das línguas. *Caderno de Squibs* 5 (1): 15–24.
- . 2020. Variação Semântica: o português brasileiro e o inglês. *Cuadernos de la ALFAL* 12 (2): 612–631.
- . 2021. Massa e contável: O Singular Nu e a variação entre as línguas. In *Estudios de Lengua y Gramática*, ed. Cecilia Bértola, Carolina Oggiani, and Ana Clara Polakof, 95–108. Montevideo: Prensa Universitaria.

- Pires de Oliveira, Roberta, and Gitanna Bezerra. in progress. *Online approaches to Bare Singulars in Brazilian Portuguese*.
- Pires de Oliveira, Roberta, Susan Rothstein, and Susan Rothstein. 2011. Bare Singular noun phrases are mass in Brazilian Portuguese. *Lingua* 121: 2153–2175.
- Pires de Oliveira, Roberta, Helena Guerra Vicente, Ana Clara Polakof and Romina Trebisacce. 2021. *(In)definiteness across languages*. International Cooperation Project. Grupo de Pesquisa CNPq.
- Rothstein, Susan. 2010. Counting and the mass/count distinction. *Journal of Semantics* 27 (3): 343–397.
- . 2017. *Semantics for counting and measuring*. Cambridge: Cambridge University Press.
- Rothstein, Susan, and Roberta Pires de Oliveira. 2020. Comparatives in Brazilian Portuguese: Counting and measuring. In *Mass and count in linguistics, philosophy, and cognitive science*, ed. Friederike Moltman, 141–158. Amsterdam: John Benjamins.
- Schmitt, Cristina, and Alan Munn. 1999. Against the nominal mapping parameter: Bare nouns in Brazilian Portuguese. In *Proceedings of NELS 29*, ed. Pius Tamanji, Masako Hirotsu, and Nancy Hall, 339–354. Amherst: GLSA.
- . 2002. The syntax and semantics of bare arguments in Brazilian Portuguese. *Linguistic Variation Yearbook* 2: 253–269.
- Taveira da Cruz, Ronald. 2008. *O SNU e a (pseudo) incorporação no PB*. PhD in Linguistics, Universidade Federal de Santa Catarina, Florianópolis. Available at: <https://repositorio.ufsc.br/handle/123456789/91630>.

**Part III**  
**South-American Indigenous Languages**

# Compounding Processes in Three Macro-Jê Linguistic Branches



Andrew Nevins and Mário André Coelho da Silva

## 1 Introduction

Languages around the world share morphological mechanisms that allow for the creation of new words, though these mechanisms are often characterized by distinct, language-specific properties. On the one hand, there is derivation, through which bound morphemes are affixed to a root so as to produce a new word related to that root. On the other hand, there is compounding, in which two or more lexical roots are combined to create another word altogether. Even though all languages employ compounding as a method of making new words (with its productivity greatly varying from language to language), what distinguishes a compound from a sequence of words can be difficult to define using general or universal criteria (i.e., the English stress difference found in *bláckboard* vs. *black bóard* does not have correlates in every language). For this reason, it is more likely for researchers to identify compounds based on language-specific conditions. This was indeed one of Bloomfield's concerns in his landmark (1939) paper "Menominee Morphophonemics": how to identify the morphophonemic processes occurring between subparts of a compound and to in turn use those to diagnose compoundhood. At the same time, it is now cross-linguistically agreed that there are some useful criteria which can be used as a starting point for elucidating

---

A. Nevins (✉)

Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

University College London, London, UK

e-mail: [andrewnevins@letras.ufrj.br](mailto:andrewnevins@letras.ufrj.br)

M. A. Coelho da Silva

Universidade Federal do Acre, Rio Branco, Acre, Brazil

Universidade Estadual de Campinas, Campinas, Brazil

e-mail: [mario.andre@ufac.br](mailto:mario.andre@ufac.br)

compounds in a given language, such as stress and phonological diagnostics for prosodic wordhood; syntactic impenetrability, inseparability and unalterability; and the behavior of compounds with respect to inflection (see Lieber and Štekauer (2009) for a comprehensive overview).

The current chapter aims to examine the methods employed for compound formation in the languages of three Macro-Jê branches of Brazil: the Maxakalí language (Maxakalí family), the Krenák language (Krenák family), and the Xerente and Xavante languages (Akuwẽ (or Central-Jê) branch of the Jê family). These four languages belong to the Eastern branch of Macro-Jê and are thus closer to one another than other languages within Macro-Jê (Nikulin 2020). As we will show, however, this proximity does not mean that the criteria used to identify a compound will remain constant in the four languages. Conversely, these four languages do have in common the higher tendency for compounding in word formation than for derivation, with only a few morphemes dedicated for the latter. Compared to Indo-European languages, all four fit well with the overall profile of an isolating language within Humboldt's (1836) original macro-typology.

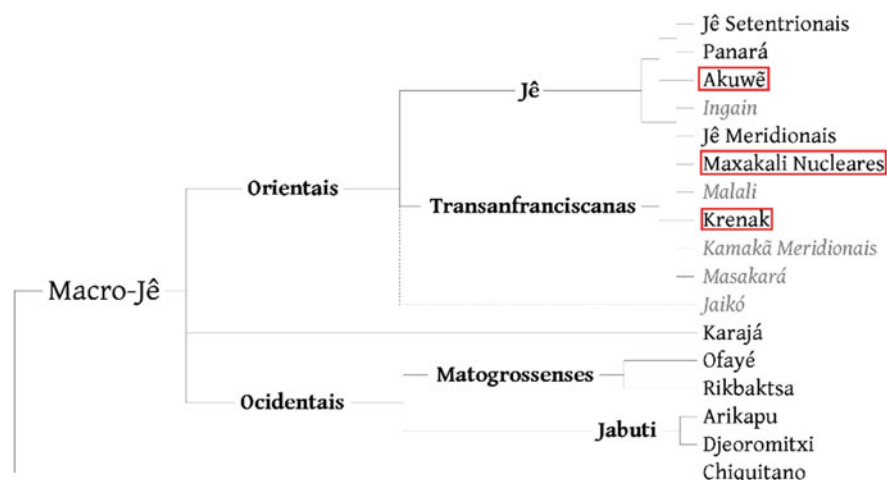
In the following section of this chapter, we briefly present general data about these four Brazilian languages. Section 3 will examine the derivation strategies used in those languages; even though derivation is not as productive as compoundhood in these languages, it is nevertheless beneficial to remark about this opposite side of the coin as well. Section 4 represents the core of the chapter, where we collate the characteristics of compoundhood in the four languages. First, we show the phonological criteria for identifying compounds in the studied languages. Then, we proceed to show the range of morphosyntactic possibilities found in the available data and detail our finding that all of the languages contain both endo- and exocentric compounds. We close the chapter by summarizing our findings and main conclusions.

## 2 The Languages Considered in This Study

The four languages considered in this study are members of the Macro-Jê language stock, pertaining to three different sub-branches of this family (indicated by the red boxes in Fig. 1). These three sub-branches are close to one another in that they form (alongside some other languages) the Eastern Macro-Jê branch. Maxakalí and Krenák are adjacent, each of them being the only extant languages of their respective families, and are grouped in what is called the Trans-Sanfranciscanian subbranch of Eastern Macro-Jê. Xerente and Xavante are the still used members of the Akuwẽ (also known in the literature as Central Jê) sub-branch of the Jê family.

The other languages in the Macro-Jê family are outside the scope of this chapter (for a map with the location of every Macro-Jê language, see Fig. 2), but it is clear that at least a few of the languages of other branches, such as Ofayé (Oliveira 2006: 121), also employ compounding as the main process of new word formation. Furthermore, Karajá (Ribeiro 2012: 55–6) and Rikbaktsa (Silva 2011: 177–82) have





**Fig. 1** Macro-Jê family. (Reproduced from Nikulin 2020)

complex phenomena related to compounding not found in all four of the languages analyzed in this chapter, for example, noun incorporation (see Sect. 4.2 for a brief discussion on noun incorporation in Xavante) – which is analyzed as a kind of compounding by some authors (Mithun 1984). Thus, compoundhood in Macro-Jê languages provides a rich field for both comparative and typological research.

In the map below, the Maxakalí family is indicated in blue, Krenák in yellow, and Akuwẽ in dark brown. The traditional territories of each family are indicated by the cross-hatched colors, and the current borders are demarcated by solid colors. The arrows indicate migrations registered by historical written sources. Although historically there were some Macro-Jê languages spoken outside modern Brazil, such as Ingain (a Southern Jê language) spoken in Argentina, the current extant languages are spoken exclusively within Brazilian territory.

Some phonologically shared characteristics among the four languages under analysis include a maximum syllable of CCVC<sup>1</sup>; the nearly complete absence of fricatives (the /s/ and /z/ phonemes in the Akuwẽ languages are recent innovations); and various phenomena related to nasality, such as the allophony of voiced oral and nasal stops, in which the former occurs in oral contexts and the latter adjacent to nasal vowels. Regarding their morphosyntax, Maxakalí, Krenák, and the Akuwẽ languages are ergative, with different kinds of splits: Maxakalí has an active-stative split in the irrealis mood (Silva 2020: 206–222), while Akuwẽ languages have a split related to the finiteness of the verb (Nikulin 2020: 286–292). Although Seki (2004)

<sup>1</sup> In the case of Maxakalí, the CCVC structure is present if one considers the analysis offered in Silva (2020). In Xerente, the phonetic structure allows the presence of more than two contiguous consonants, but this is due to phonological processes which elide underlying vowels. These processes are yet to be fully described.



**Fig. 2** Map of the historical distribution of the Macro-Jê languages. (Reproduced from Nikulin 2020)

and Campos (2021) claim that Krenák is an ergative language, there is at present not enough reliable data to determine whether it contains ergative splits or not.

These languages have relatively straightforward derivational morphology, sharing various morphemes between both nouns and verbs, and are all head-final. For a description of these languages, see Seki (2004) for Krenák, Estevam (2011) for Xavante, Sousa Filho (2007) and Cotrim (2016) for Xerente, and Campos (2009) and Silva (2020) for Maxakalí. For a genetic comparison of these four languages

(and the other Macro-Jê languages as well) with lexical, phonological, and partial morphosyntactic reconstructions, we refer the reader to Nikulin (2020).

The data for the analysis carried out in this chapter regarding compounding in these four languages were taken from several sources. The Maxakalí data used in this chapter comes from almost a decade of work by the second author together with the Maxakalí people. The Krenák data was taken from an unpublished wordlist collected by Seki (1982). Finally, data for the Xerente language was taken from the dictionary compiled by Krieger and Krieger (1994) and that for Xavante taken from the dictionary compiled by Lachnitt (1987). Sources other than these are cited individually above the respective examples (specifically, examples (1g), (8a–b), (12), and (29)).

## 2.1 Maxakalí Language

The Maxakalí language (*Tikmūūn yīy ax*, Glottocode [maxa1247]) is the only extant spoken language in the family of the same name.<sup>2</sup> Data from the first half of the 2010s put the count of the Maxakalí (Ethnonym: *Tikmūūn*)<sup>3</sup> at 2014 people, a number that has most likely increased in the years since (Instituto Socioambiental 2021). Since most of Maxakalí have their traditional language as their mother tongue, it is safe to assume there are more than 2000 speakers of the language.

These speakers presently reside in five territories (*Pananiy*, *Kōnāg Mai*, *Īmmoknāg*, *Apne Yīxux*, and *Apne Ixkot Hāmhipak*) in the northeastern state of Minas Gerais in Brazil. Their original lands were larger historically: Maxakalí and other speakers of closely related dialects had lands ranging from southern Bahia up through the Doce River as their traditional home circa the nineteenth century in the Minas Gerais state. If one includes the now-extinct Maxakalí languages, their traditional lands were even larger, reaching up to the border of the Minas Gerais and Rio de Janeiro states, otherwise known as the territory of the now-extinct Koropó people. The Maxakalí languages were grouped as *Goytacás* by von Martius (1867), but there are no written sources of the language spoken by the eponymous group, who were located near the Koropó in the northern part of the Rio de Janeiro state (cf. the city “Campo dos Goytacazes,” some 275 km northeast of the city of

<sup>2</sup> There are two other peoples that originally spoke Maxakalí languages, Pataxó and Pataxó-Hāhāhāe, which are now undertaking the revitalization of their traditional languages. Since there was an intergenerational rupture in the transmission of their original languages, the ones now being brought to light are heavily influenced by Portuguese. Such influence (i.e., of the foreign adopted language) is well-known and documented in revitalization efforts: that was the case for Hebrew and for some Australian Aboriginal languages, for example (Zuckermann and Walsh 2011).

<sup>3</sup> The Maxakalí ethnonym is, according to the Maxakalí themselves, a compound, formed by *tik* “man” *mūg* “1.PL” *ūn* “woman,” roughly meaning “Us, men, and women.” The elision of the *g* in *mūg* is regular.

**Table 1** Maxakalí consonants

|                    | Labials              | Dentals                | Palatals              | Velars               | Glottal        |
|--------------------|----------------------|------------------------|-----------------------|----------------------|----------------|
| Voiceless plosives | /p/ [p]<br><p>       | /t/ [t̞]<br><t>        | /ç/ [tʃ]<br><x>       | /k/ [k]<br><k>       |                |
| Voiced plosives    | /b/ [b] ~ [m]<br><m> | /d/ [d̞] ~ [n̞]<br><n> | /j/ [dʒ] ~ [ɲ]<br><y> | /g/ [g] ~ [ŋ]<br><g> |                |
| Fricative          |                      |                        |                       |                      | /h/ [h]<br><h> |

**Table 2** Maxakalí vowels

|        | Front          |                | Central        |                | Back                 |                      |
|--------|----------------|----------------|----------------|----------------|----------------------|----------------------|
| Closed | /i/ [i]<br><i> | /ĩ/ [ĩ]<br><ĩ> | /i/ [i]<br><u> | /ĩ/ [ĩ]<br><ũ> | /u/ [u] ~ [o]<br><o> | /ũ/ [ũ] ~ [õ]<br><õ> |
| Open   | /ɛ/ [ɛ]<br><e> | /ẽ/ [ẽ]<br><ẽ> | /a/ [a]<br><a> | /ã/ [ã]<br><ã> |                      |                      |

Rio de Janeiro). Relating the Maxakalí and Goytacás people is a speculation that still requires further data to be confirmed.

The Maxakalí language has nine consonants and ten vowels in its phonemic inventory, according to Silva (2020). Tables 1 and 2 present both classes.

In the tables above, the phonemes are shown between slashes with their main allophones in square brackets. Orthographic graphemes used by the Maxakalí communities are shown in angled brackets and will be the representation used in the remainder of the chapter, except when phonological/phonetic examples need to be used for clarity purposes.

## 2.2 Krenák Language

The Krenák language ([krɛˈnak iˈma], Glottocode [kren1239]) is the sole member of its family. Regarding historical sources, it seems that various groups spoke closely related dialects of the same language with no register of a different language of the same group. Today, the ethnic Krenáks (Ethnonym: [m<sup>b</sup>uˈru<sup>g</sup>ɪ]) total around 434 individuals, mainly from the [nakrɛˈhɛ] and [ɪ<sup>g</sup>utˈkrak] groups, located in 3 different territories (*T.I. Krenák*, in the Brazilian state of Minas Gerais; *T.I. Vanuíre*, in the state of São Paulo; and *T.I. Krenrehe*, in the state of Mato Grosso, where the last 2 territories are situated outside their traditional lands) (Instituto Sociambiental 2021).

The Krenák speakers were denominated by the pejorative exonym *Botocudos* in historical literature due to the ear and lip plugs (*botoque*, in Portuguese) traditionally worn in past centuries. In some historical written sources, they were also called *Aimorés*, which has a Tupí origin. The name Krenák, in contrast, is said to mean “head in the land” (Krenak et al. 1997), but a more accurate meaning is “head’s land,” a compound made by /krɛn/ [ˈkrɛn] “head” + /nāk/ [ˈnak] “land” and was the

name of a chief of a [m<sup>b</sup>u'ru<sup>g</sup>ŋ] group. [m<sup>b</sup>u'ru<sup>g</sup>ŋ] groups were usually known by toponyms and/or the name of a chief.

Unlike Maxakalí, Krenák is a critically endangered language with very few native speakers left: the number of speakers is likely less than a dozen people. The severity of the situation was already alarming in the 1980s and can be traced back to a series of attacks and public policies against the Krenák. This systematic crusade resulted in high language loss dating from the 1950s when the Krenák were expelled from their traditional lands, though some speakers did return to their traditional lands some 20 years later. During the military dictatorship in Brazil (1964–1985), the autocratic government established the *Reformatório Indígena Crenac*, a barbaric prison for indigenous people from all around the country, inside the Krenák traditional lands. The *Reformatório* witnessed a number of crimes committed by the government, including torture of the inmates and strict control of the entry and exit of the Krenák in their own land, with the threat of forced labor for any indigenous non-compliance. The *Reformatório* was shut down in 1972, and the Krenák were forcibly transferred to the *Fazenda Guarani*, a reservation outside their traditional lands, before returning to the Doce River in 1980 (Emmerich and Monserrat 1975, Cristófar-Silva 1986, Seki 1992). Seki (1992) counted 15 speakers in the beginning of the 1990s, while Frassetto (2019: 136) counts only 5 fluent speakers in one of the three lands in which they are located (*T.I. Vanuíre*). The fluent speakers, according to Cristófar-Silva (1986), were all born in or before the 1950s, meaning the younger ones would be nearing 70 years old presently.

Krenák is also a relatively under-described language. Even though there are a somewhat large number of works about the language, they are in most part composed of word lists collected by people untrained in linguistics in the nineteenth and first half of the twentieth centuries. Two notable exceptions are the vast work of the Russian ethnologue H. H. Manizer in 1915, which can be found in the *Peter the Great Museum of Anthropology and Ethnography* (Kunstkamera) in Saint Petersburg, Russia, and the lists collected by Curt Nimuendajú in 1939, whose originals were lost in the 2018 fire that destroyed almost the entire archive of *Museu Nacional* in Rio de Janeiro.<sup>4</sup> Phonological analyses undertaken by Cristófar-Silva (1986), Pessoa (2012), and Seki (2004) present brief descriptions of the phonemic inventory and some basic aspects of the morphosyntax of the language. The phonological basis used in this work is based on the one proposed by Nikulin and Silva (2020: 10–3) in a comparative work between the Krenák and Maxakalí languages (Tables 3 and 4), with the addition of the nasal vowel /ã/ due to the ongoing research of the second author of this chapter.

The analysis in Nikulin and Silva (2020) differs from preceding analyses in that, according to these two authors, the post-oralized allophones are derived from phonological nasal consonants, and therefore, a voiced plosive series is not posited. Ongoing research using historical data conducted by the second author of this chapter indicates that voiceless nasals can be predicted from morphological context

<sup>4</sup> Thankfully, we had obtained and digitized copies of those lists before the fire. Unfortunately, we still have not managed access to the Manizer archive in Russia.

**Table 3** Krenák consonants

|                    | Labials                     | Dentals                       | Palatals                     | Velars                | Glottal |
|--------------------|-----------------------------|-------------------------------|------------------------------|-----------------------|---------|
| Voiceless plosives | /p/ [p]                     | /t/ [t̪]                      | /c/ [tʃ]                     | /k/ [k]               |         |
| Voiced nasals      | /m/ [m] ~ [m <sup>b</sup> ] | /n/ [n̪] ~ [n̪ <sup>d</sup> ] | /ɲ/ [ɲ] ~ [ɲ <sup>dz</sup> ] | /ŋ/ [ŋ <sup>g</sup> ] |         |
| Voiceless nasals   | /m̥/ [m̥]                   | /n̥/ [n̥]                     | /ɲ̥/ [ɲ̥]                    | /ŋ̥/ [ŋ̥]             |         |
| Continuants        | /w/ [w]                     | /r/ [r]                       | /j/ [j] ~ [z̥]               |                       | /h/ [h] |

**Table 4** Krenák vowels

|          | Front   |         | Central |           | Back    |           |
|----------|---------|---------|---------|-----------|---------|-----------|
| Non-open | /i/ [i] | /ĩ/ [ĩ] | /ə/ [ə] | /ɜ̃/ [ɜ̃] | /u/ [u] | /ũ/ [ũ]   |
| Open     | /ɛ/ [ɛ] | /ê/ [ê] | /a/ [a] | /ã/ [ã]   | /ɔ/ [ɔ] | /ɔ̃/ [ɔ̃] |

in most cases and may not be phonemic, but as the available data does not allow for a conclusive analysis, we regard them as separate phonemes here.

Because the Krenák language has no established orthography, with each community using different strategies for writing their language, we present the data using only phonological and/or phonetic representations.

### 2.3 *Akuwẽ Languages (Xerente and Xavante)*

There are two *Akuwẽ* (also known as Central Jê) languages still spoken: *Xerente* (*Akwẽ mrmẽze*) and *Xavante* (*A'uwẽ mreme*), both spoken by people of the same name (respective ethnonyms: *Akwẽ*, Glottocode [xere1240], and *A'uwẽ*, Glottocode [xava1240]). These *Akuwẽ* languages, together with the extinct *Xakriabá* and *Akroá* languages, form a sub-branch of the larger Jê family. *Xerente* is spoken by 3509 people in two territories (*Xerente* and *Funil*) in the municipality of Tocantínia, state of Tocantins (Instituto Socioambiental 2021), while *Xavante*, one of the indigenous languages with the most speakers in Brazil, is spoken by some 22,256 people in 10 territories (*Areões*, *Chão Preto*, *Marãiwatsédé*, *Marechal Rondon*, *Parabubure*, *Pimentel Barbosa*, *Sangradouro*, *São Marcos*, *Ubawawẽ*, and *Wededzê*) in the state of Mato Grosso (Instituto Socioambiental 2021). Even though most *Xerente* and *Xavante* speak their mother tongues as their first language, almost all *Xerente* are also highly proficient in Brazilian Portuguese. The second author of this chapter taught linguistic classes for *Xerente* and *Xavante* students in the Intercultural Education undergraduate course at the Federal University of Goiás, and according to the students, there is dialectal variation in both languages, although this still needs to be properly researched.

Although the phonemic inventories of the two languages are highly similar (Tables 5 to 8), both languages present some innovations. Among the most relevant are the changes of Proto-*Akuwẽ* \**k* to *Xavante* /ʔ/, while *Xerente* retains the /k/ (e.g., *Xerente* *pikõ*, *Xavante* *pi'õ*, tracing to Proto-*Akuwẽ* \**pikõ* “woman”). There is also a series of vowel elisions in *Xerente*, while *Xavante* (or at least some dialects) still preserves the structure C(C)V(C) of the protolanguage (compare, e.g., *Xerente*

**Table 5** Xerente consonants

|                      | Labials                    | Dentals                     | Fricatives     | Velars         | Glottal        |
|----------------------|----------------------------|-----------------------------|----------------|----------------|----------------|
| Voiceless obstruents | /p/ [p]<br><p>             | /t/ [t̪]<br><t>             | /s/ [s]<br><s> | /k/ [k]<br><k> |                |
| Nasals               | /m/ [m] ~ [b]<br><m> ~ <b> | /n/ [n̪] ~ [ɲ]<br><n> ~ <d> |                |                |                |
| Continuants          | /w/ [w]<br><w>             | /r/ [r]<br><r>              | /z/ [z]<br><z> |                | /h/ [h]<br><h> |

**Table 6** Xavante consonants

|                      | Labials                    | Dentals                     | Fricatives/Affricates               | Glottals       |
|----------------------|----------------------------|-----------------------------|-------------------------------------|----------------|
| Voiceless obstruents | /p/ [p]<br><p>             | /t/ [t̪]<br><t>             | /s/ [s] ~ [ts]<br><ts>              | /ʔ/ [ʔ]<br><ʔ> |
| Nasals               | /m/ [m] ~ [b]<br><m> ~ <b> | /n/ [n̪] ~ [ɲ]<br><n> ~ <d> | /ɲ/ [ɲ] ~ [z] ~ [dz]<br><nɲ> ~ <dz> |                |
| Continuants          | /w/ [w]<br><w>             | /r/ [r]<br><r>              |                                     | /h/ [h]<br><h> |

**Table 7** Xerente vowels

|        | Front          |                | Central         |                  | Back           |                |
|--------|----------------|----------------|-----------------|------------------|----------------|----------------|
| Closed | /i/ [i]<br><i> | /ĩ/ [ĩ]<br><ĩ> | /i/ [i̠]<br><ú> |                  | /u/ [u]<br><u> |                |
| Mid    | /e/ [e]<br><ê> | /ẽ/ [ẽ]<br><ẽ> | /ə/ [ə]<br><â>  | /ə̃/ [ə̃]<br><ã> | /o/ [o]<br><ô> | /õ/ [õ]<br><õ> |
| Open   | /ɛ/ [ɛ]<br><e> |                | /a/ [a]<br><a>  |                  | /ɔ/ [ɔ]<br><o> |                |

*krda*, Xavante *'rada*, reconstructable as Proto-Akuwẽ *\*krada* “old”). The phonemic inventories of the two languages are given below.<sup>5</sup>

Like the Maxakalí data, we present the examples from Xerente and Xavante using the established orthography for both languages unless it becomes necessary to present in phonetic transcriptions for the sake of clarity.

### 3 Derivation in the Four Languages

Before we proceed to the discussion about compounding in the four Macro-Jê languages under consideration, we briefly analyze the morphemes used in derivation

<sup>5</sup> It is important to note that the fricatives from Xerente and fricatives/affricates from Xavante come etymologically from palatals and are thus indicated in a separate column in Tables 5 and 6, since they still have palatal allophones in some cases: [j] for Xerente and Xavante /s/ in coda, [j̠] for Xavante /ɲ/ and some Xerente /n/ in coda, and [ɲ] for Xavante /ɲ/ before a nasal vowel.

**Table 8** Xavante vowels

|        | Front          |                | Central         |                | Back           |                |
|--------|----------------|----------------|-----------------|----------------|----------------|----------------|
| Closed | /i/ [i]<br><i> | /ĩ/ [ĩ]<br><ĩ> | /i/ [i̠]<br><y> |                | /u/ [u]<br><u> |                |
| Mid    | /e/ [e]<br><e> | /ẽ/ [ẽ]<br><ẽ> | /ə/ [ə]<br><ô>  | /õ/ [õ]<br><ã> | /o/ [o]<br><ô> | /õ/ [õ]<br><õ> |
| Open   | /ɛ/ [ɛ]<br><é> |                | /a/ [a]<br><a>  |                | /ɔ/ [ɔ]<br><o> |                |

in each of these languages so as to demonstrate that this is a restricted, though productive, method for word formation.

### 3.1 Derivation in Maxakalí

In Maxakalí, there are three nominalizer morphemes, one verbalizer, which also functions as a causative when suffixed to verbal roots, and another causative suffix with restricted occurrence. These are the only derivational morphemes in the language, which is in keeping with its overall classification as an isolating language akin to Mandarin Chinese. All of these morphemes, except for one of the nominalizers and the second causative, also have other functions besides being category-changing affixes (Campos 2009: 132–6, Silva 2020: Chapter 4).

The first nominalizer is the clitic *ax* used to derive inanimate nouns from intransitive verbs (1a–b) and animate nouns from transitive verbs (1c–d). When following a verb, it has the function of expressing future tense (1e–f). Campos (2009: 132) explains that it can also be used as nominalizer at the sentential level, as (1g) (adapted from Campos 2009: 132). In this case, however, since the nominalized verb has as its internal argument the postposition *nõ* 3.INS, instead of *hã* (which is used after nouns), we reanalyze it as a verb nominalization preceded by a noun (something like “seller of meat” instead of “meat-seller”):<sup>6</sup>

Notice that the examples in (1) can be ambiguous: the words in the right column of (1a–d) can be used to indicate the future tense of the verbs, and the words in (1e–f) can be analyzed as nouns derived from the bases (the suffix *-vaalaa* in Hindi/Urdu similarly shares the function of both nominalizer and marker of immediate future). Syntactic, semantic, and pragmatic contexts are used to disambiguate each meaning in Maxakalí.

The second nominalizer in Maxakalí is *hãm*, which has the historic form *ãm* that is no longer productive. This morpheme stems from the word for “thing” and is

<sup>6</sup> The glosses used in this work are as follows: ANIM, animated; DAT, dative; INAN, inanimate; INS, instrumental; INTR, intransitive; NEG, negative; NMZ, nominalizer; PL, plural; SG, singular; TRANS, transitive.



- (1) a. *-xit* ‘eat.INTR’ → *-xit ax* ‘food’  
 b. *-yĩy* ‘speak’ → *-yĩy ax* ‘language, speech’  
 c. *xat* ‘ask for’ → *xat ax* ‘client’  
 d. *kix* ‘kill.PL’ → *kix ax* ‘assassin’  
 e. *-kutex* ‘sing’ → *-kutex ax* ‘will sing’  
 f. *xupak* ‘listen’ → *xupak ax* ‘will listen’  
 g. *mũnũy -tut -yĩn nõ menex ax*  
 deer -mother-meat 3.INS sell NMZ  
 ‘butcher’ ← lit.: ‘seller of cow’s meat’

used to form passive and inanimate nouns from both transitive and intransitive verbs ((2)a–b). The unproductive *ãm* is placed before intransitive verbs (2c–d):

- (2) a. *-pakut* ‘sick’ → *hãm pakut* ‘sickness’  
 b. *ãgtux* ‘tell, narrate’ → *hãm ãgtux* ‘story’  
 c. *-muk* ‘cook.INTR’ → *ãmmuk* ‘food, meal’  
 d. *-hok* ‘NEG’ → *ãmhok* ‘no’

The last nominalizer is *oknãg*, which has a privative meaning and can be postposed to nominal bases (3). It can be postposed to verbal bases as well, but its function in this case remains unclear.

- (3) a. *payenet* ‘farmer’ → *payenet oknãg* ‘without farmer(s)’  
 b. *yãy kup* ‘one’s own leg’ → *yãy kup oknãg* ‘easy prey’ (lit.: ‘without one’s own leg’)

The last two morphemes presented in this subsection are prototypically causatives. The first, *-nãhã*, is productive and undergoes an assimilation of its initial consonant to the place of articulation of the last consonant of the base (4a–b). It can also be suffixed to nouns to form verbs (4c–d). The second suffix, *-a*, is not productive and can be found in verbal roots ending in velar consonants (4e–f).

- (4) a. *-xuxi* ‘cold’ → *-xuxinãhã* ‘make something get cold’  
 b. *yũmũg* ‘understand’ → *yũmũgãhã* ‘teach’  
 c. *-yĩkox* ‘mouth’ → *yĩkoyãhã* ‘imitate’  
 d. *kotpex* ‘beiju’<sup>7</sup> → *kotpegãhã* ‘make beiju’  
 e. *-kumuk* ‘bad’ → *-kumua* ‘make something bad and/or useless’  
 f. *-yok* ‘straight’ → *-yoa* ‘lift up, erect’

In summary, Maxakalí includes derivational morphology for agentives, nominalizations, verbalizations, privative predicates, and causatives.

<sup>7</sup> *Beiju* is a kind of tortilla made from manioc starch, and its cooking is widespread among South American lowland peoples.

### 3.2 Derivation in Krenák

In the available descriptive work on Krenák, there is almost no mention of derivational morphology. However, the word lists allow one to identify at least some affixes, though it is difficult to make generalizations about them.

Seki (2004: 137–8) glosses, in two examples, the prefix /rə-/ as a causative-comitative: /rə-nĩŋ/ “bring,” from /nĩŋ/ “come.” In her 1982 word list, one can also find the following examples:

- (5)
- |    |                               |   |                                        |
|----|-------------------------------|---|----------------------------------------|
| a. | /rə-mũŋ/ ‘take away, carry’   | ← | /mũŋ/ ‘go’                             |
| b. | /rə-mət/ ‘finish, stop.TRANS’ | ← | /mət/ [1 <sup>m</sup> bət] ‘stop.INTR’ |
| c. | /rə-kwãŋ/ ‘have’              | ← | ? /kwãŋ/ ‘belly’                       |
| d. | /rə-ŋẽŋẽŋ/ ‘bend’             | ← | /ŋẽŋẽŋ/ ‘???’                          |
| e. | /rə-kõmũ/ ‘begin’             | ← | /kõmũ/ ‘???’                           |

We were able to find the bases for the (possible) causative-comitatives in (5a–b), but not for those in (5c–e). For the label causative-comitative, it is unclear if that is the real meaning of /rə-, since example (5b) may indicate a simple causative.

Seki’s (1982) list also includes examples of another possible causative, namely, a suffix in the form /-Vŋ/ [-ʔV<sup>g</sup>ŋ], in which /V/ is a copy of the last vowel of the base. We present the data in (6) below:

- (6)
- |    |                            |   |                                    |
|----|----------------------------|---|------------------------------------|
| a. | /ɲãɛ-ɛŋ/ ‘get close’       | ← | /ɲãɛŋ/ ‘close’                     |
| b. | /jɛ-ɛŋ/ ‘leave’            | ← | /jɛk/ ‘put’                        |
| c. | /ĩmpɛ-ɛŋ/ ‘dry something’  | ← | /ĩmpɛ/ [ĩm <sup>h</sup> dʒɛ] ‘dry’ |
| d. | /jica-aŋ/ ‘heat something’ | ← | /jica/ ‘hot’                       |

There is also another verb which seems to have the same causative suffix for which we have not found the correspondent base: /pɛ-ɛŋ/ “light, burn” (cf. the discussion of the example /cɔn-pɛk/ “fire” on Sect. 4). All of the bases in the examples have a velar or no final coda, but this may reflect a skew in the data.

Finally, there is a derivational morpheme /ãm/, which seems to be used as a nominalizer like Maxakalí *hãm* ~ *ãm*, and of which it is probably cognate. Again, similar to the Maxakalí cognate, it has the meaning of “thing, something” and seems to form inanimate nouns. We provide some examples of the nominalizer /ãm/ below:

- (7) a. /ã̃m-ã̃ŋut/ ‘food’ ← /ã̃ŋut/ [ã̃'ŋ<sup>9</sup>ut] ‘eat.TRANS’  
 b. /ã̃m-him/ [ã̃m<sup>b</sup>i<sup>b</sup>m] ‘night’ ← /him/ [hi<sup>b</sup>m] ‘black’  
 c. /ã̃m-ca/ ‘field for farming’ ← /ca/ ‘mow, clear, scrape’

There are some other words beginning with /ã̃m/ that are most likely also nominalizations, but we could not find their bases in the available sources. Some examples are /ã̃mŋru/ [ã̃m'ŋ<sup>9</sup>ru] “noise”; /ã̃mmuru/ [ã̃m<sup>b</sup>u<sup>9</sup>ru] “wind, cold”; /ã̃mŋun/ [ã̃m'dzu<sup>d</sup>n] “day, today”; /ã̃mmijik/ [ã̃m<sup>b</sup>i'z̥ik] “manioc”; and /ã̃mŋək/ [ã̃m'dzək] “shadow,” among others.

### 3.3 Derivation in Xerente and Xavante

Descriptions of Xerente and Xavante offer more derivational affixes. For Xerente, Cotrim (2016) lists nine different suffixes:

- rê “attenuative”
- zawre “intensive (nouns)”
- kta(b) “intensive (state)”
- hu “abundance of”
- di (and allomorphs -ki and -ti) “existential”
- kō (and the allomorph -tō) “privative”
- ri (and allomorphs -rī and -r) “nominalizer (action)”
- kwa “nominalizer (agent)”
- ze “nominalizer (circumstance = instrument, place, action)”

We consider only the last three of them to be true derivational suffixes as they may be analyzed as inflectional suffixes (-rê, -di) due to their grammatical nature. The others are descriptive roots (-zawre, -kta(b), -kō) as they behave syntactically the same way as other descriptive roots like psê “good,” pre “red,” and so on. The inalienable noun (-hu) also may be considered a descriptive root, since it demonstrates the same syntactic behavior of other part-whole relationship words such as -nkra “branch” and -pa “root.”

Xavante is likewise described as having many derivational affixes, many of them cognate to the Xerente ones. Estevam (2011) presents the following: -re “diminutive,” -wawê “augmentative,” si(P)- ~ nhi(P)- “nominal applicative,” simi- ~ nhimi- “nominal antipassive” m̃ã- “singulative,” ró(P)- “antipassive,” -'wa “nominalizer (agent),” and -dzé “nominalizer (instrument, place, action).” For similar reasons, we only consider the final two as true derivational affixes.

Examples are given in (8) for the three Xerente derivational affixes and in (9) for the Xavante ones. Examples in (8a–b) were taken from Cotrim (2016: 113–4).

- (8) a. *kahu* ‘eat’ → *kahur* ‘eating’  
 b. *mō* ‘go’ → *mōrī* ‘going’  
 c. *rowahtu* ‘teach’ → *rowahtukwa* ‘teacher’  
 d. *pto* ‘sprout, germinate’ → *-ptokwa* ‘creator, generator, father’  
 e. *kupsbi* ‘cover’ → *kupsbize* ‘blanket, cape’  
 f. *nāmrā* ‘sit.SG’ → *nmrāze* ‘chair’
- (9) a. *mro* ‘gather, count’ → *mro’wa* ‘one who counts’  
 b. *petse* ‘heal, fix’ → *petse’wa* ‘healer’  
 c. *-dza’o* ‘hang’ → *-dza’odzé* ‘place for hanging’  
 d. *-hörö* ‘call, yell’ → *-hörödzé* ‘bugle’

In sum, Xavante and Xerente each have a number of derivational affixes, including nominalizations of both agent and instrument.

## 4 Compounding in the Four Languages

In this section, we will discuss the compounding in the four languages considered in this chapter: Maxakalí, Krenák, and Xerente and Xavante. As anticipated at the beginning of this chapter, compounding is a more productive and frequent way to expand the lexicon in Eastern Macro-Jê languages. The question arises as to how one can discern a compound from a sequence of separate words. As these languages have almost no inflectional morphology, defining whether a word is a compound or not can quickly become an arduous task. Thus, the distinction between compounds and two different words can be blurred, and morphology itself does not seem to help disentangle one from the other. To address this issue, we suggest phonological, morphosyntactic, and semantic criteria for the identification and classification of compounds in these languages.

### 4.1 Phonological Criteria for Identifying Compounds

Phonological criteria for the identification of compounds in the four Macro-Jê languages analyzed herein differ greatly. While phonology is not always a good criterium for the identification of compounds in Maxakalí, it helps in some cases in Krenák and frequently in the Akwẽ languages.

In Maxakalí, compounds are difficult to identify by phonological criteria alone, though there are some cases in which morphophonological alternations facilitate the process. For example, an intricate diagnostic involves the alternation between short and long forms in nouns. In this language, alienable monosyllabic roots of nouns have two allomorphs: one short and one long, shown in (10):

| (10) | SHORT FORM   | LONG FORM      |            |
|------|--------------|----------------|------------|
| a.   | <i>mãm</i>   | <i>mãhãm</i>   | ‘fish’     |
| b.   | <i>nãn</i>   | <i>nãhãn</i>   | ‘annatto’  |
| c.   | <i>ku</i>    | <i>kuhu</i>    | ‘firewood’ |
| d.   | <i>kutet</i> | <i>kutehet</i> | ‘bamboo’   |

The example in (10d) is phonologically monosyllabic, having an underlying representation /ktet/, with the first vowel being epenthetic.<sup>8</sup> According to Silva (2020: 195–205), short forms of the roots need to lengthen if they do not have a complement or if they do not fill the position of the complement themselves within a larger phrase. That is, when a monosyllabic nominal root is the sole morpheme in a phonological word, the noun must lengthen in order to form an iambic foot. Thus compounds, composed of at least two syllables (one from each root), never have long forms in their constituents. In (11), we present some examples:

|      |    |                                                         |   |                                                                    |
|------|----|---------------------------------------------------------|---|--------------------------------------------------------------------|
| (11) | a. | <i>mãmpata</i> ‘piaba’ <sup>9</sup>                     | ← | <i>mãm</i> ‘fish’ + <i>pata</i> ‘foot’ (* <i>mãhãmpata</i> )       |
|      | b. | <i>nãnhep</i> ‘red ink’                                 | ← | <i>nãn</i> ‘annatto’ + <i>hep</i> ‘liquid’ (* <i>nãhãnhep</i> )    |
|      | c. | <i>kugõy</i> ‘smoke’                                    | ← | <i>ku</i> ‘firewood’ + <i>gõy</i> ‘smoke’ (* <i>kuhugõy</i> )      |
|      | d. | <i>kutetkut</i> ‘bicho-da-taquara larvae’ <sup>10</sup> | ← | <i>kutet</i> ‘bamboo’ + <i>kut</i> ‘larvae’ (* <i>kutehetkut</i> ) |

There are also other syntactic relations in Maxakalí in which the long form rule applies, such as the position of the noun before or after the verb, inalienable possession marking, and the placement of a postposition in a phrase (Silva 2020: 191–205); therefore, the semantic definition related to a referent must also be considered.

Krenák, unlike Maxakalí, does not have root lengthening, but does have stress placed on the last syllable of lexical words, just like its “sister” language, Maxakalí. Even though the relevant data for Krenák is scarce, in most works of dubious quality for a complex linguistic analysis, there is one interesting phonological process we have observed that indicates compoundhood. This process is the voicing of a voiceless consonant following a nasal. Cristófar-Silva (1986: 86–7) was the first to describe this phenomenon in Krenák, explaining that every instance of a nasal consonant followed by a voiceless obstruent will lead to the voicing of the latter, such as in [‘mbøk] “fish,” which would be derived from the voiceless /npøk/. We

<sup>8</sup> Compare this with *mũnũy* “deer” which is underlyingly disyllabic, /bidĩç/, and does not have a “long” form \**mũnũhũy*.

<sup>9</sup> *Mãmpata* is used to name some species of fish from the Order Characiformes (Ferreira 2012: 105).

<sup>10</sup> *Bicho-do-taquara* is a moth whose edible larvae grow inside bamboo and which is also used as an hallucinogenic inducer.



certainty such as the one in (13f), as we don't know if its internal constituency is [X Y [Z]] or [X [Y Z]]. Therefore, the fact that the voicing of /kat/ does not occur may be related to this unknown fact.

Secondly, in all the compounding examples given in (13), the consonant undergoing voicing is an unvoiced velar /k/. There is another word which may show that this voicing is more pervasive in the language and that we may be dealing with a limitation in the data: the word meaning “fire” [tɕɔ̃m<sup>b</sup>ɛk]. If one analyzes this token as a compound of /cɔ̃n/ “wood” and /pɛk/ “burn,” the generalization holds because we have similar evidence in (12c) with the dative postposition undergoing a similar process. Even though /pɛk/ “burn” does not appear in the lists collected by trained linguists, it appears as *pek* “burn, armed” in the dictionary made by Rudolph (1909). It also may be the base of the verb /pɛ-ɛŋ/ “light, burn,” as presented in Sect. 3.2. The described assimilation occurs both at inflectional boundaries, as in example (12), and in compounds, and is thus indicative that one is dealing with just one phonological word in the latter case.

We offer one last observation regarding this assimilation. In some examples from the nineteenth-century word lists, as well as in the ones from the first half of the twentieth century (including those of Nimuendajú), there is some fluctuation regarding the representation of items like the ones seen in (13) in that there is no indication of voicing whatsoever. Therefore, it is rather common to see forms in which the nasal consonants do not trigger voicing of the unvoiced stop, like in <čoňkát> “canoe,” <čompék> “fire,” <kreňké> “hair,” etc. (Nimuendajú 1939), together with attestations similar to what we present in the example (13) above, where nasal consonants do indeed trigger the voicing of the unvoiced stops. Rather than counterexamples, we suppose that this kind of data was due to dialectal variation and/or language change, since data from the few more recent sources transcribed by trained linguists (published from the 1980s up to present day, such as Seki 1982 and Cristófaró-Silva 1986), are more “regular” in this sense.

Finally, phonological criteria for the identification of compounds in the Akuwẽ languages is possible because many nouns in these languages have two forms, though distinct from the alternation in Maxakalí. In both Xerente and Xavante, many words have a non-final and a final form. As the name suggests, the latter appears when it is the rightmost element of a phonological word and thus coincides with the elicitation or citation form, as it is the only root and therefore the one at the end of the word.

In contrast, the non-final forms are found only in contexts where they are not the last root of a complex word. Thus, these two forms yield a straightforward way for identifying a compound in that roots that have both of them will vary according to their position in the word. We present some examples of variable roots in Xerente (14a–c) and in Xavante (14d–f):

| (14) | NON-FINAL FORM  | FINAL FORM    |         |
|------|-----------------|---------------|---------|
| a.   | <i>btâ</i>      | <i>bdâ</i>    | ‘sun’   |
| b.   | <i>tpê</i>      | <i>tbê</i>    | ‘fish’  |
| c.   | <i>kunmã</i>    | <i>kuzâ</i>   | ‘fire’  |
| d.   | <i>’ētē</i>     | <i>’ēnē</i>   | ‘rock’  |
| e.   | <i>’utö</i>     | <i>’uhödö</i> | ‘tapir’ |
| f.   | <i>rowa’u’u</i> | <i>rowa’u</i> | ‘wind’  |

These are just a few pairs from a fairly large cache of roots that alternate between non-final and final forms. As seen above, both languages have this allomorphy, allowing it to be traced back to their proto language, which has its origins in the evolution of codas from the Proto-Jê language to Proto-Akuwẽ. Further examples of compounds in Xerente and Xavante are given in (15) and (16):

- (15) XERENTE
- a. *btâ* ~ *bdâ* ‘sun’ + *zas* ‘enter’ + *ze* NMZ  
→ *btâzasze* / \**bdâzasze* ‘west’
  - b. *tpê* ~ *tbê* ‘fish’ + *zdawa* ‘edge’ + *pre* ‘red’  
→ *tpêzdawapre* / \**tbêzdawapre* ‘pirarara fish (*Phractocephalus hemioliopus*)’
  - c. *kunmã* ~ *kuzâ* ‘fire’ + *nĩsdu* ‘tip’  
→ *kunmãĩsdu* / \**kuzãĩsdu* ‘flame’
- (16) XAVANTE
- a. *’ētē* ~ *’ēnē* ‘rock’ + *pa* ‘stream’  
→ *’ētēpa* / \**’ēnēpa* ‘rock stream (name of one of the traditional Xavante age groups)’
  - b. *’utö* ~ *’uhödö* ‘tapir’ + *nhitsi’re* ‘nose’ + *’u* ‘horn’  
→ *’utönhitsi’re’u* / \**’uhödönhitsi’re’u* ‘rhinoceros’
  - c. *rowa’u’u* ~ *rowa’u* ‘wind’ + *wahö* ‘cold’  
→ *rowa’u’uwahö* / \**rowa’uwahö* ‘cold wind’

In Xavante and Xerente, genitive phrases with inalienable nouns employ the same formula as compounds with the possessor to the left in a non-final form and the possessed to the right with a final form: *’utö para* ‘the foot of the tapir.’ (Note that the final form of the possessor (\**’uhödö para*) cannot occur in these genitive phrases.) This could lead one to analyze such phrases as compounds, but this situation seems to be similar to the difference between genitive phrases and



compounds in Semitic languages, such as Hebrew. For example, the free nominal *báyit* “house” has a construct (= possessed) form *beyt*. In a construct/genitive phrase *beyt morá* “house of a teacher,” the morphophonological structure is similar to a compound, as in *beyt midráš* “religious school” (< house + sermon) (Borer 2009).

Despite having the same form, there are syntactic differences between constructs and compounds in Hebrew and probably in Xavante and Xerente as well. Like in the Semitic languages, however, the use of a special allomorph according to the context in which a root appears is only one indicator that one is dealing with a compound, and not an exclusive diagnostic to identify compoundhood.

## 4.2 Morphosyntactic Criteria

In this section, we present the possible constituents of compounds in the languages under discussion. Though the most productive compounds in the languages are constructed with two nouns (N-N) or with a noun and a descriptive verb (N-V), there are other types of combination also used to produce new words.

We identified N-N compounds in all four languages with examples given in (17–20):

- (17) MAXAKALÍ
- a. *mĩmtut* ‘house’ ← *mĩhĩm* ‘tree, wood’ + *tut* ‘mother’
  - b. *tappetpet* ‘school’ ← *tappet* ‘paper’ + *pet* ‘home’
  - c. *kupukkox* ‘shotgun’ ← *kupuuk* ‘axe’ + *kox* ‘hole’
- (18) KRENÁK
- a. /kɛakãŋmak/ ‘pants’ ← /kɛakãŋ/ ‘clothes’ + /mãk/ ‘leg’
  - b. /cõŋjun/ ‘skewer’ ← /cõn/ ‘wood’ + /jun/ ‘tooth’
  - c. /tarukəkə/ ‘cloud’ ← /taru/ ‘sky’ + /kəkə/ ‘smoke’
- (19) XERENTE
- a. *wdêpa* ‘root’ ← *wdê* ‘tree’ + *pa* ‘arm’
  - b. *nrõto* ‘maripa palm’ ← *nrõ* ‘coconut’ + *to* ‘eye’
  - c. *kumdâmkwa* ‘traditional necklace worn by the bride’  
← *kumdâ* ~ *kumdâm* ‘capybara’ + *kwa* ‘tooth’

- (20) XAVANTE
- a. *hö're* 'pocket' ← *hö* 'bark, clothes' + *'re* 'hole, cavity'
  - b. *po'redzapu* 'earhole' ← *po're* 'ear' + *dzapu* 'hole'
  - c. *dzadaipro* 'saliva' ← *dzada* 'mouth' + *pro* 'liquid, powder'

There are also N-N-N compounds in all four languages, though they are much rarer than N-N examples. Some of the former are presented in (21–24) with their internal hierarchy indicated by the usage of brackets (except for Krenák, in which the available data do not allow us to make safe assumptions about structure):

- (21) MAXAKALÍ  
*Kotkuphi* 'Manioc fiber spirit'  
← [[*kot* 'manioc' + *kup* 'tree, horn'] + *hi* 'fiber']
- (22) KRENÁK  
*/pəŋkitəmkat/* 'wax'  
← */pəŋ/* 'honey, bee' + */kitəm/* 'eye' + */kat/* 'bark, skin'
- (23) XERENTE  
*nrōtowdê* 'maripa palm tree'  
← [[*nrō* 'coconut' + *to* 'eye'] + *wdê* 'tree']
- (24) XAVANTE  
*waprunhorōdupu* 'varicose veins'  
← [[*wapru* 'blood' + *nhorō* 'vein, string']] + *dupu* 'swelling']

Another productive compound patterning is when a noun is used as the head of the construction, followed by a verb, usually descriptive. Descriptive verbs in Macro-Jê languages are notionally similar to adjectives in European languages. Due to space restrictions, we will not present the reasoning behind the consideration of them as verbs. For an explanation of the absence of a proper adjectival class in Maxakalí, see Campos (2009:144–50) and Silva (2020: 211–5); for Krenák, we forward the reader to the paper by Seki (2004: 133–4); for Xerente, see Cotrim (2016); and for Xavante, see Estevam (2011: 69–76). For the latter language, Estevam (2011) uses the term static verb instead of descriptive verb, but the behavior of the class is overall the same across all four languages. Examples are given in (25–28) below:

- (25) MAXAKALÍ
- a. *xoktap* ‘black panther’ ← *xok* ‘animal’ + *tap* ‘dark’
  - b. *kokemax* ‘fox’ ← *kokex* ‘dog’ + *max* ‘false, pseudo’
  - c. *kupukhe* ‘scythe’ ← *kupuuk* ‘axe’ + *he* ‘bent’
- (26) KRENÁK
- a. /mĩjĩãjĩrum/ ‘(name of a [m<sup>b</sup>u<sup>9</sup>ɾ] group)’  
← /mĩjĩãjĩ/ ‘water’ + /jĩrum/ ‘white, clear’
  - b. /pɔjica/ ‘(name of another [m<sup>b</sup>u<sup>9</sup>ɾ] group)’  
← /pɔ/ ‘hand, foot’ + /jica/ ‘hot’
  - c. /mɔkma/ [m<sup>b</sup>ɔk<sup>m</sup>ba?] ‘piaba’  
← /mɔk/ ‘fish’ + /ma/ ‘flat’
- (27) XERENTE
- a. *târapre* ‘bronze’ ← *târa* ‘metal’ + *pre* ‘red’
  - b. *hâiwakro* ‘fever’ ← *hâ* ~ *hâi* ‘skin, bark’ + *wakro* ‘hot’
  - c. *kaktōka* ‘plumbeous pigeon (*Patagioenas plumbea*)’  
← *kaknō* ~ *kaktō* ‘picazuro pigeon  
(*Patagioenas picazuro*)’ + *ka* ‘white’
- (28) XAVANTE
- a. *Marãiwatsédé* ‘name of a Xavante Territory’  
← *marã* ~ *marãi* ‘forest’ + *watsédé* ‘dense, thick’
  - b. *bödödi’rã* ‘asphalt’ ← *bödödi* ‘path’ + *rã* ‘black’
  - c. *’ritu* ‘ruin’ ← *’ri* ‘house’ + *tu* ‘abandoned’

It is even possible to find compounds of the N-N-V type in the four languages, such as Xavante *pidzawatapa* “teapot,” formed by the roots *pidza* “bowl, pot” + *wada* ~ *wata* “beak” and *pa* “long,” though this kind of longer compound is incredibly sparse.

Another type of compound found, at least in Maxakalí, are the V-V constructions made with two verbal roots. These are few in number, and all the examples that are synchronically analyzable are formed with one of the three verbs meaning “get” followed by the verbs “go” or “come,” yielding a complex variety of verbs meaning various forms of “bring” or “take away.” The possible combinations are exemplified in Table 9.

**Table 9** Maxakalí verbs for “bring” and “take away”

| V1                       | V2                            |                                   |
|--------------------------|-------------------------------|-----------------------------------|
|                          | <i>nūn</i> ‘come’             | <i>mōg</i> ‘go’                   |
| <i>pa</i> ‘get.INAN.SG’  | <i>paxnūn</i> ‘bring.INAN.SG’ | <i>paxmōg</i> ‘take away.INAN.SG’ |
| <i>put</i> ‘get.ANIM.SG’ | <i>putnūn</i> ‘bring.ANIM.SG’ | <i>putmōg</i> ‘take away.ANIM.SG’ |
| <i>pop</i> ‘get.PL’      | <i>popnūn</i> ‘bring.PL’      | <i>popmōg</i> ‘take away.PL’      |

The compounding in these cases brings with it the semantic implications of movement of the verb’s internal argument (closer to the speaker with *nūn* and further away with *mōg*) as well as implications for (in)animacy and number (animacy is neutralized in the plural).<sup>11</sup> So, for example, if one wants to talk about bringing one’s puppy, they should use the verb *putnūn*, since the internal argument of the verb is singular and animate. Although it may be an unproductive kind of compounding, the Maxakalí case can be classified as an asymmetrical serial verb construction according to the definition and diagnostics proposed by Aikhenvald and Dixon (2006: 21–37) since it includes a semantically restricted class of motion verbs that serves as a modificational specification for the open, though small, class of event-specific “get” verbs.

We have not found compounding with two or more verbal roots in the other languages considered here, but this may be due to a skew in the available data. As we have said before, Krenák is an under-described language, and the data used for Xerente and Xavante comes from dictionaries which may have excluded this type of compounding.

Finally, Xavante is described as having noun incorporation in its repertoire (Estevam 2011: 376–92). Estevam describes nominal incorporation in Xavante as a “productive, regular, and semantically transparent process of constructive morphology” in which the incorporated argument of the verb may or not change the verbal valence. Though the classification of noun incorporation as a type of compounding is controversial since it may be considered a syntactic process rather than a lexical one (Murasugi 2014), we present in (29) some examples taken from Estevam (2011) for the sake of a complete overview of the possibilities:

<sup>11</sup> Notice also that the compound form of *pa* “get.INAN.SG” also acquires an excremental *-x*. We do not have an explanation as to why this coda appears, but it is probably reminiscent of the mood distinction still preserved in several verbs, such as *mōg* “go” (*realis* mood) and *mō* (*irrealis* mood), *nāhā* “fall” (*realis* mood) and *nāhāy* (*irrealis* mood), etc. The Akuweê languages also have a pervasive distinction between two forms of verbs, and even though the different forms are cognate to the Maxakalí ones, they encode finite (verbal) and non-finite (nominal) forms, the latter being used in subordinate clauses. A mood distinction is also morphologically marked in Krenák, at least in two verbs, /mũj/ ~ /mũ/ “go” and /nĩj/ ~ /nĩ/ “come,” with the second form of each pair serving as an imperative.

## (29) OBJECT INCORPORATION

- a. *pawaptob* ‘help’ ← *pa* ‘arm’ + *waptob* ‘offer’  
 b. *topo’o* ‘wake up’ ← *to* ~ *top* ‘eye’ + *po’o* ‘break’  
 c. *’ruiwapari* ‘hate’ ← *’ru* ~ *’rui* ‘rage’ + *wapari* ‘hear’

## INSTRUMENT INCORPORATION

- d. *pawaibu* ‘bring’ ← *pa* ‘arm’ + *waibu* ‘pick up’  
 e. *po’reha’ö* ‘forget’ ← *po’re* ‘ear’ + *ha’ö* ‘take’  
 f. *toptö’ö* ‘daze’ ← *to* ~ *top* ‘eye’ + *tö’ö* ‘die’

## CLASSIFICATORY INCORPORATION

- g. *hö’rêne* ‘drink’ ← *hö* ‘liquid’ + *’rêne* ‘eat’  
 h. *’öhuri* ‘ingest liquids’ ← *’ö* ‘(running) water’ + *huri* ‘ingest’  
 i. *tsaihuri* ‘have intercourse’ ← *tsi* ~ *tsai* ‘nourishment’ + *huri* ‘ingest’

In sum, we have classified compounds according to criteria such as headedness and category and provided N-N (and N-N-N), N-V, and V-V compounds, the latter of which bear similarity to serial verb constructions.

### 4.3 Semantic Description

In this final subsection, we present the semantic criteria for classification of compounds in the four languages. In all of them, both endocentric and exocentric compounds can be found. Endocentric compounds occur when one of the roots acts as the morphological head and the compound as a hyponym of this head. Examples are given in (30–33):

## (30) MAXAKALÍ

- a. *tappetpet* ‘school’ ← *tappet* ‘paper’ + *pet* ‘home, place’  
 b. *takox* ‘anus’ ← *ta* ‘buttock’ + *kox* ‘hole’  
 c. *mĩmkũĩn* ‘*Mĩmkũĩn* (a stick used to teach children the traditional songs)’  
 ← *mĩm* ‘tree, wood’ + *kũĩn* ‘striped (vertically)’

## (31) KRENÁK

- a. /ciŋkat/ ‘leather’ ← /ciŋ/ ‘meat’ + /kat/ ‘bark, skin’  
 b. /kənkε/ ‘eyebrows’ ← /kəŋ/ ‘forehead’ + /kε/ ‘hair, fur’  
 c. /m̃ĩp̃ñh̃im/ ‘coffee (drink)’ ← /m̃ĩp̃ñ/ ‘water’ + /h̃im/ ‘black’

## (32) XERENTE

- a. *dakrāihi* ‘skull’ ← *dakrā* ~ *dakrāi* ‘one’s head’ + *hi* ‘bone’  
 b. *pizawaku* ‘broth’ ← *piza* ‘pot’ + *waku* ‘soup, juice’  
 c. *kāpre* ‘flood’ ← *kā* ‘water’ + *pre* ‘red’

## (33) XAVANTE

- a. *wahiwa’u* ‘venom’ ← *wahi* ‘snake’ + *wa’u* ‘liquid (from inside)’  
 b. *’upadzu* ‘manioc flour’ ← *’upa* ‘manioc’ + *dzu* ‘powder’  
 c. *da’rādzépiré* ‘headache’ ← [[*da’rā* ‘one’s head’ + *dzé* ‘pain’]+ *piré* ‘heavy, difficult’]

Notice how the N-N compounds in (30a–b), (31a–b), (32a–b), and (33a–b) have a right-sided head. The compounds in these examples are always a hyponym of the head: “school” is a hyponym of “place” (30a), “eyebrows” are a hyponym of “hair, fur” (31b), and so on. Similarly, examples (30c, 31c, 32c), which are N-V compounds, and (32c), which is a N-N-V one, have the rightmost noun acting as the morphological head. In the first three cases, there is only one noun in each compound, and thus this noun necessarily acts as the head and therefore also the hypernym.

In contrast, exocentric compounds, the second semantic type described in this subsection, refer to compounds where none of the roots act as a semantic head related to the resulting word, and therefore none of the roots are hypernyms of the compound. Consequently, their meaning often cannot be grasped by the individual constituents. Examples are given in (34–37):

## (34) MAXAKALÍ

- a. *Kuptapyĩmmāg* ‘Orion’s belt’ ← *kuptap* ‘vulture’ + *yĩmmāg* ‘wing’  
 b. *āmnĩytut* ‘opossum’ ← *āmnĩy* ‘night’ + *tut* ‘mother’  
 c. *xax’āta* ‘(a kind of) armadillo’  
 ← *xax* ‘bark, skin’ + *āta* ‘red’

## (35) KRENÁK

- a. /ciŋmrɔŋ/ ‘stomach’ ← /ciŋ/ ‘meat’ + /mrɔŋ/ [m<sup>b</sup>rɔ<sup>g</sup>ŋ] ‘path’  
 b. /jukuanimiəŋ/ ‘rainbow’ ← /jukuan/ ‘boa constrictor’ + /imiəŋ/ [im<sup>b</sup>i’ə<sup>g</sup>ŋ] ‘urine’  
 c. /pɔŋɔɛk/ ‘horse’ ← /pɔ/ ‘hand, foot’ + /ŋɔɛk/ ‘round’

## (36) XERENTE

- a. *krizdawa* ‘door’ ← *kri* ‘house’ + *zdawa* ‘mouth’  
 b. *tāizâ* ‘hail’ ← *tā* ~ *tāi* ‘rain’ + *zâ* ‘seed’  
 c. *nānmākrānē* ‘caninana snake (*Spilotes pullatus*)’  
     ← *nōzâ* ~ *nānmā* ‘corn’ + *krā* ‘black’ + *nē* ‘similar’

## (37) XAVANTE

- a. *’ödawa* ‘beach’  
     ← *’ö* ‘(running) water’ + *dawa* ‘opening, entrance’  
 b. *hötöramreme* ‘tape-recorder’ ← *hötöra* ‘axe’ + *mreme* ‘speak’  
 c. *tsaihuri* ‘have intercourse’ ← *tsi* ~ *tsai* ‘nourishment’ + *huri* ‘ingest’

The relationship between the compound meaning and its parts is not always clear, but the roots sometimes provide a hint for meaning, like in example (36a), where the notion of “door” in Xerente can be literally translated as “mouth of the house.” Even though “door” is a hyponym of neither “mouth” nor “house,” the semantic relation between the compound and its constituent parts is implicit.

Another example can be found in (34c): *xax āta* “(a type of) armadillo” is distinct from a noun with a modifier as *xax āta* “red bark” (*xax* ‘bark’, *āta* ‘red’) just from its referent; when one knows that the *xax āta* armadillo has a bark with a reddish hue, the relation becomes more obvious. On the other hand, compounds such as the one in (35b) are completely ambiguous for non-speakers of Krenák, as in this example, the meaning of the word for “rainbow” is only remotely related to its parts, “boa constrictor” and “urine,” though it maybe has (or had) a cultural explanation, this is not apparent at a first glance. Likewise, the example in (34a) shows that the Maxakalí perceive the shape and form of the constellation known in English as the Orion’s Belt as being similar to a vulture’s wing, which may not be readily clear for someone outside their cultural background.

In sum, we have provided evidence for endocentric and exocentric compounds in terms of the semantic interpretation of the compound as a whole based on its constituent parts.

## 5 Conclusions

The data presented in this chapter aimed at shedding some light on compounding in Amerindian languages of Brazil with an isolating morphological profile, as at present, there is still a shortage of work that includes detailed patterns with breadth in the typological accounts about this subject. We presented phonological,

morphosyntactic, and semantic diagnostics for identifying this productive way of forming new words in four Macro-Jê languages.

Considering that the most well-studied case of a compounding processes and headedness in an isolating language is Mandarin Chinese (see, e.g., Packard 2000; Ceccagno and Basciano 2009; among others), we discussed data and patterns from other languages with a similar typological profile to broaden the understanding of compound formation and identification. In turn, we hope that a direct comparison of the issues of headedness and compounding criteria in the four Macro-Jê languages presented here can be of interest to future work in morphological typology and its theoretical modeling.

**Acknowledgments** This chapter is dedicated to the memory of Nanblá Gakran, a speaker-linguist of the Laklanō-Xokleng language (Macro-Jê), who passed away in June 2021. We would like to thank Carlo Sandro de Oliveira Campos, Sara Hockett, and the editors, for careful reading and suggestions.

## References

- Aikhenvald, Alexandra Y., and R.M.W. Dixon. 2006. *Serial verb constructions: A cross-linguistic typology*. Oxford: Oxford University Press.
- Bloomfield, Leonard. 1939. Menominee morphophonemics. *Études phonologiques dédiées à la mémoire de M. le Prince Trubetzkoy : Travaux de Cercle Linguistique de Prague* 8: 108–115.
- Borer, Hagit. 2009. Afro-Asiatic, Semitic: Hebrew. In *The Oxford handbook of compounding*, ed. Rochelle Lieber and Pavol Štekauer, 491–511. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0027>.
- Campos, Carlo Sandro de Oliveira. 2009. *Morfofonêmica e morfossintaxe do Maxakalí*. PhD dissertation. Belo Horizonte: Universidade Federal de Minas Gerais.
- . 2021. *Ergatividade cindida em Krenák?: Uma análise preliminar*. Talk presented at Amazônicas VIII. Available at: <https://www.youtube.com/watch?v=Ldvn6hGvTHE>.
- Ceccagno, Antonella and Bianca Basciano. 2009. Sino-Tibetan: Mandarin Chinese. In Lieber, Štekauer *The Oxford handbook of compounding*, Rochelle Lieber and Pavol Štekauer, 478–490. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0026>
- Cotrim, Rodrigo Guimarães Prudente Marquez. 2016. *Uma descrição da morfologia e de aspectos morfossintáticos da língua Akwẽ-Xerente (Jê Central)*. PhD dissertation. Brasília: Universidade de Brasília.
- Cristófaros-Silva, Thaís. 1986. *Descrição fonética e análise de alguns processos fonológicos da língua Krenák*. Master's thesis. Belo Horizonte: Universidade Federal de Minas Gerais.
- Emmerich, Charlotte, and Ruth Monserrat. 1975. Sobre os Aimorés, Krens e Botocudos: Notas lingüísticas. *Boletim do Museu do Índio – Antropologia* 3: 3–44.
- Estevam, Adriana Machado. 2011. *Morphosyntaxe du Xavante: Langue Jê du Mato Grosso (Brésil)*. PhD dissertation. Paris: Université Denis Diderot – Paris 7.
- Ferreira, Marco Túlio, and da Silva. 2012. *Ecologia histórica aplicada à gestão ambiental comunitária da Terra Indígena Maxakalí. Minas Gerais*. MA thesis. Belo Horizonte: Universidade Federal de Minas Gerais.
- Frassetto, Pedro Ternes. 2019. Krenak: Registros etnográficos e de viajantes no processo de revitalização linguística. In *Revitalização de línguas indígenas: O que é, como fazemos*, org. Wilmar D'Angelis, 133–153. Campinas: Curt Nimuendajú.



- Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Berlin: F. Dümmler Verlag.
- Instituto Socioambiental. 2021. *Quadro geral dos povos*. [https://pib.socioambiental.org/pt/Quadro\\_Geral\\_dos\\_Povos](https://pib.socioambiental.org/pt/Quadro_Geral_dos_Povos). Accessed 28 June 2021.
- Krenak, Maurício, José Carlos Krenak, Marcos Krenak, Osmar Krenak, and Itamar Krenak. 1997. *Conne pãnda rithioc Krenak: Coisa tudo na língua Krenak*. Belo Horizonte: MEC/UNESCO/SEE-MG.
- Krieger, Wanda Braidotti, and Guenther Carlos Krieger. 1994. *Dicionário escolar*. Rio de Janeiro: Junta de Missões Nacionais da Convenção Batista Brasileira.
- Lachnitt, Georg. 1987. *Romnhüsi'ubumro A'uwẽ mreme = Waradzu mreme / Dicionário Xavante = Português*. Campo Grande: Missão Salesiana de Mato Grosso.
- Lieber, Rochelle, and Pavol Štekauer. 2009. *The Oxford handbook of compounding*. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.001.0001>.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60 (4): 847–894. <https://doi.org/10.2307/413800>.
- Murasugi, Kumiko. 2014. Noun incorporation, nonconfigurality, and polysynthesis. In *The Routledge handbook of syntax*, ed. Andrew Carnie, Yosuke Sato, and Daniel Siddiqi, 283–303. London: Routledge. <https://doi.org/10.4324/9781315796604.ch14>.
- Nikulin, Andrey. 2020. *Proto-Macro-Jê: Um estudo reconstrutivo*. PhD dissertation. Brasília: Universidade de Brasília.
- Nikulin, Andrey, and Mário André Coelho da Silva. 2020. As línguas Maxakalí e Krenák dentro do tronco Macro-Jê. *Cadernos de Etnolinguística* 8 (1): 1–64.
- Nimuendajú, Curt. 1939. *Unveröffentlichte Sprachaufnahmen 1909–1940 – Band II*. Unpublished wordlists.
- Oliveira, Maria das Dores de. 2006. *Ofayé, a língua do povo do mel: Fonologia e gramática*. PhD dissertation. Maceió: Universidade Federal de Alagoas.
- Packard, Jerome L. 2000. Defining the word in Chinese. In *The morphology of Chinese: A linguistic and cognitive approach*, ed. Jerome L. Packard, 7–20. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486821.002>.
- Pessoa, Katia Nepomuceno. 2012. *Análise fonética e fonológica da língua Krenak e abordagem preliminar de contos Botocudo*. PhD dissertation. Campinas: Universidade Estadual de Campinas.
- Ribeiro, Eduardo Rivail. 2012. *A grammar of Karajá*. PhD dissertation. Chicago: The University of Chicago.
- Rudolph, Bruno. 1909. *Wörterbuch der Botokudensprache*. Hamburg: Fr. W. Thaden.
- Seki, Lucy. 1982. *Vocabulário Português-Botocudo (Krenak/Nakrehe)*. Unpublished wordlist.
- . 1992. Notas para a história dos Botocudo (Borum). *Boletim do Museu do Índio – Documentação* 4.
- . 2004. Aspectos da morfossintaxe Krenak: Orações independentes. *LIAMES* 4: 129–146. <https://doi.org/10.20396/liames.v4i1.1431>.
- Silva, Léia de Jesus. 2011. *Morphosyntaxe du Rikbaktsa (Amazonie brésilienne)*. PhD dissertation. Paris: Université Denis Diderot – Paris 7.
- Silva, Mário André Coelho da. 2020. *Tikmũn yĩ ax tinã xohi xi xahĩnãg – Sons e pedaços da língua Maxakalí: Descrição da fonologia e morfologia de uma língua Macro-Jê*. PhD dissertation. Belo Horizonte: Universidade Federal de Minas Gerais.
- Sousa Filho, Sinval Martins de. 2007. *Aspectos morfossintáticos da língua Akwẽ-Xerente (Jê)*. PhD dissertation. Goiânia: Universidade Federal de Goiás.
- Wetzels, Leo, and Andrew Nevins. 2018. Prenasalized and postoralized consonants: The diverse functions of enhancement. *Language* 94 (4): 834–866.
- Zuckermann, Ghil'ad, and Michael Walsh. 2011. Stop, revive, survive: Lessons from the Hebrew revival applicable to the reclamation, maintenance and empowerment of Aboriginal languages and cultures. *Australian Journal of Linguistics* 31 (1): 111–127.

# *Poro-/mba'e-* Antipassive Prefixation in Paraguayan Guarani



Bruno Estigarribia and Ernesto Luiz López Almada

## 1 Introduction

In this contribution, we argue in favor of the existence of an antipassive voice in modern Paraguayan Guarani. More precisely, we claim that *poro-/mba'e-* prefixation (1a, b) and (2a–c) is antipassive voice marking.<sup>1</sup>

---

<sup>1</sup> Examples are formatted and glossed following the Leipzig Glossing Rules (available at <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>; accessed 28 January 2021). Incorporated objects are indicated by the plus sign (+). Guarani third-person prefixes are unmarked for number and are always translated by “they/them/their” forms. The orthography of examples follows the guidelines given in Estigarribia (2020), which sometimes differs from other accepted orthographies, often with respect to the attachment of prefixes to their bases and postpositions to their complements.

Abbreviations: 1,2,3 First, second, third person; 1>2PL first-person agent, second-person plural patient; 1>2SG first-person agent, second-person singular patient; ACT active; ADJZ adjectivizer; AGD agent-demoting voice; ANTIP.H antipassive marker for human patients; ANTIP.NH antipassive marker for non-human patients; CAUS1 transitivity-causative voice; CAUS2 ditransitivizing causative voice; DEST destinative aspect/nominal future tense; DIM diminutive; DUB dubitative marker; EXCL exclusive of the addressee(s); FUT future tense/prospective aspect; IMP imperative; INACT inactive; INCL inclusive of the addressee(s); LOC locative (“in, on, at,” DOM); LOC2 oblique locative (“at, against, by”); MED medial demonstrative; NEG negation; NMLZ general nominalizer suffix; NMLZ.AG agentive nominalizer; NMLZ.PASS passive nominalizer; NMLZ.QUAL abstract nominalizer for qualities; NMLZ.REL relational nominalizer; NPOSS non-possessed form of relational roots; NPROX.PL general non-proximal plural demonstrative; PL plural; POSSM possessed form of relational roots (for non-third-person pronominal possessor); POSSM3 possessed form of relational roots (for third-person pronominal possessor); POST post-stative aspect/nominal past tense; Q question marker; SG singular; VERD veridical emphatic

---

B. Estigarribia (✉)

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA  
e-mail: [estigarr@email.unc.edu](mailto:estigarr@email.unc.edu)

E. L. López Almada

Departamento de Lenguas, Universidad Nacional de Itapúa, Encarnación, Paraguay  
e-mail: [ernesto.lopez@uc.edu.py](mailto:ernesto.lopez@uc.edu.py)

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023  
C. Rodrigues, A. Saab (eds.), *Formal Approaches to Languages of South America*,  
[https://doi.org/10.1007/978-3-031-22344-0\\_11](https://doi.org/10.1007/978-3-031-22344-0_11)

- (1) a. Aporojuka.  
 a-**poro**-juka  
 1SG.ACT-ANTIP.H-kill  
 ‘I kill (people).’ / ‘I am a killer.’
- b. Amba’ejogua.  
 a-**mba’e**-jogua  
 1SG.ACT-ANTIP.NH-buy  
 ‘I buy things.’ / ‘I am shopping.’
- (2) a. Oime oho oporohekávo.  
 o-ime                      o-ho                      o-**poro**-h-eka-vo  
 3.ACT-be                  3.ACT-go                  3.ACT-ANTIP.H-POSSM3-look-while  
 ‘They went to look for someone.’  
 (Guarania 2008:45)
- b. Oho ñemuhápe omba’ejoguávo.  
 o-ho                      ñemu-ha=pe                  o-**mba’e**-jogua-vo  
 3.ACT-go                  market-NMLZ=LOC          3.ACT-ANTIP.NH-buy-while  
 ‘They went to the market to buy (animals or things).’  
 (Guarania 2008:45)
- c. Amba’eñongatu.  
 a-**mba’e**-ñongatu.  
 1SG.ACT-ANTIP.NH-store  
 ‘I save.’ (literally, ‘I store things.’)  
 (Krivoshein de Canese and Acosta Alcaraz 1987:114)
- d. Mbo’ehára oporombo’e.  
 mbo’e-hára                  o-**poro**-mbo’e  
 teach-NMLZ.AG          3.ACT-ANTIP.H-teach  
 ‘The teacher teaches (people).’  
 (Krivoshein de Canese and Acosta Alcaraz 1987:114)

There is no published linguistic analysis of the prefixes *poro-* and *mba’e-* specifically. Traditional and scholastic grammars vary in their presentations. Gregores and Suárez (1967) list *poro-* (but not *mba’e-*) as a verbal prefix encoding an unspecified plural object, only used with a “reduced number of transitive stems” (p. 129). Zarratea (2012) calls *poro-* (but not *mba’e-*) *voz activa objetiva* (“active objective voice”). Guarania (2008) and Krivoshein de Canese and Acosta Alcaraz (1987) call it *voz subsuntiva* (“subsumptive voice”) for people (*poro-*) or for animals or things (*mba’e-*). Krivoshein de Canese and Acosta Alcaraz (1987) and Zarratea (2012) note that the resulting predicates are intransitive, a fact that we will take up in Sect. 4.4. There is a consensus in Paraguayan scholastic grammars that these prefixes are voice markers, although the category “antipassive” has never been used.

Published work in linguistic typology assumes that Paraguayan Guaraní lacks an antipassive (Dryer and Haspelmath 2013, based on Gregores and Suárez 1967;

Heaton 2017, 2020, based on Velázquez-Castillo 2008 and Jensen 1990). Generally, these prefixes are seen as a manifestation of a more general process of noun incorporation. For example, for Ayala (1996), *poro-* is an incorporated objective pronoun that creates an intransitive verb that functions exactly like a verb with noun incorporation.<sup>2</sup>

- (3) a. porojuka  
**poro**-juka  
 ANTIP.H-kill  
 'to kill people/someone' (Ayala 1996:134)
- b. aporopytyvõ  
 a-**poro**-pytyvõ  
 1SG.ACT-ANTIP.H-kill  
 'I help someone/people/my fellow beings/others' (Ayala 1996:134)

Velázquez-Castillo (1995a, b) includes *mba'e-* prefixation in her analysis of noun incorporation, based on the fact that the independent word *mba'e* means "thing(s)." Noun incorporation is exemplified in (4) (see further discussion in Sect. 5.2).

- (4) a. Amymbajuka.  
 a-**mymba**+juka  
 1SG.ACTIVE-**animal**+kill  
 'I hunt.'
- b. Ejepohéi.  
 e-je-**po**+(jo)héi  
 IMP.2SG-AGD-**hand**+wash  
 'Wash your hands.'

For Dietrich (2017), the prefixes *poro-* "generic human object" and *mba'e-* "(some)thing" fall under incorporation and word formation, yielding intransitive predicates with a habitual reading (but see Sect. 4.2). In her reconstruction of proto-Tupi-Guarani, Jensen (1998: 536) considers *\*ma'e* "thing" and *\*poro* "person" "generic morphemes" that can be incorporated into the verb to create intransitive verbs. Rodrigues (1953) also deemed this a manifestation of object incorporation to form intransitive verbs in Old Tupi (sixteenth and seventeenth centuries). Noun incorporation can indeed instantiate the antipassive for some languages (Polinsky 2017). However, although we agree that these prefixes are intransitivizers, we believe the noun incorporation analysis is incorrect for Paraguayan Guarani.

<sup>2</sup> In his analysis of the closely related Correntinean Guarani variety, Cerno (2011) puts forth an analysis of *poro-* and *mba'e* as pronominal incorporation.

## 2 The Antipassive Voice

Polinsky (2017: 308) defines antipassives as “constructions in which the logical object of a transitive (two-place) predicate is not realized as a direct object, but instead appears as a non-core argument or left unexpressed (but presupposed).” As Heaton (2017, 2020) notes, antipassives show a wide range of variation cross-linguistically, although it can be divided into two main groups: optional antipassives that remove the patient for particular predicate types and (often) obligatory antipassives that express a demoted patient and are usually rather productive. Both authors note that there are no necessary and sufficient conditions to identify antipassives unequivocally and that antipassives and noun incorporations are particularly difficult to discriminate in a principled way.

Here, we focus on the following properties of *poro-/mba'e-* prefixed predicates:

1. They alternate with transitive structures that are less marked.
2. They are used for generic, non-specific, or unknown patients.
3. Their subject is an agent.
4. Their patient argument is backgrounded in the discourse. As a result, they cannot be used to answer a question about the object or to refer anaphorically to a given referent.
5. They are compatible with both telic and atelic interpretations, and they yield both habitual and episodic readings, depending on the lexical aspect of the base predicate.
6. They do not admit an overt direct object or overt oblique complement with the patient role.
7. They are syntactically intransitive.
8. They are productive in the modern language.

Properties 1–4 and 7 are typical of antipassives cross-linguistically (Heaton 2017), and it is mainly on the basis of these that we propose an antipassive analysis. Properties 4–6 and 8 are shared by some antipassive types, but are not universal. Backgrounding of the patient is a function commonly noted (Cooreman 1994; Polinsky 2017), although patient foregrounding is also found (Heaton 2017).<sup>3</sup> Like Guarani, approximately 70–75% of the languages surveyed in Heaton (2017) do not admit an oblique patient (property 6), albeit the patient is always present in the semantic structure. Like Guarani, approximately 35% of her languages are highly productive (property 7). The antipassive markers are almost exclusively dedicated to this function (however, see Sect. 5.2).

---

<sup>3</sup> “Backgrounding” generally refers to processes that make event participants less syntactically prominent, less definite or individuated, less salient, sometimes indicating that they are less affected by a transitive event (in the case of patients). “Foregrounding,” on the other hand, highlights an argument, making it more salient and raising its syntactic prominence.

### 3 Relevant Aspects of the Grammar of Paraguayan Guarani

In this section, building mostly upon Estigarribia (2020), we provide a description of some aspects of Guarani grammar relevant to the grammatical topics we are addressing in the chapter.<sup>4</sup>

A phonological characteristic of Guarani that has morphosyntactic reflexes relevant here is the presence of nasal harmony. Simplifying somewhat, the stressed vowel of a roots is most often the final vowel since the language's roots are almost exclusively oxytone. If this stressed vowel is nasal, this nasality spreads several syllables to the left (regressive or anticipatory nasalization) and can spread one syllable to the right (progressive or preservatory nasalization). This creates oral/nasal allomorph alternations for a considerable number of affixes and enclitics. Oral allomorphs have oral or nasal-oral contour consonants and are used with oral bases. Nasal allomorphs used with nasal bases have fully nasal consonants. These alternations are indicated in the orthography by the following changes in spelling: <mb> → <m>; <nd> → <n>; <j> → <ñ>; sometimes <p> → <mb> or <m>, <k> → <ng> and <g> → <ḡ> as well. Stress is marked with an acute stress mark whenever it falls on a non-final vowel in a word (monomorphemic roots or multimorphemic words). If this vowel is nasal, the nasal tilde suffices to indicate stress as well. The use of the acute stress mark is borrowed from Spanish; however, the orthographic convention of marking oxytone words is different from that in Spanish, and it responds to the predominantly oxytone character of native Guarani roots.

Paraguayan Guarani is an agglutinating language. It has two clearly defined main lexical classes: nouns and verbs. (Evidence for the existence of adjectives and adverbs is scarce; see Dietrich 2017.) Intransitive verbs can be divided into an active and an inactive class (split intransitivity). The subject of active predicates is expressed with an active prefix from the *a-* set below, whereas the subject of inactive predicates is expressed with an inactive prefix from the *che-* set below (see Table 1). Note that Guarani differentiates a first-person plural inclusive that includes the addressees, from a first-person plural exclusive that excludes them (clusivity). All forms are unmarked for gender. The third person is unmarked for number as well: we will translate it as singular/plural “they/them/their” in all circumstances. The active set has prefix allomorphs ending in *i-* (see Table 1).

Transitive verbs mark with a prefix the higher participant in the 1 > 2 > 3 person hierarchy. If this participant is the agent, it receives active marking; if it is the patient, it receives inactive marking. Additionally, if the agent is first person and the patient second person, two special portmanteau prefixes are used: if the

<sup>4</sup> We describe here properties that are conserved for the modern language even on the face of the extensive contact with Spanish and that are in general use by speakers of Guarani, whether they use a more Guarani-oriented register or one closer to Jopara (Guarani-Spanish mixing).

**Table 1** Person-marking prefixes

|          | Active <i>a</i> - set | Inactive <i>che</i> - set |
|----------|-----------------------|---------------------------|
| 1SG      | <i>a(i)-</i>          | <i>che-</i>               |
| 2SG      | <i>re(i)-</i>         | <i>nde-/ne-</i>           |
| 1PL.INCL | <i>ja(i)-/ña(i)-</i>  | <i>ñande-/ñane-</i>       |
| 1PL.EXCL | <i>ro(i)-</i>         | <i>ore-</i>               |
| 2PL      | <i>pe(i)-</i>         | <i>pende-/pene-</i>       |
| 3        | <i>o(i)-</i>          | <i>i-/ij-/hi-/iñ-</i>     |

**Table 2** Indirect object pronouns

|          | Indirect object pronouns |
|----------|--------------------------|
| 1SG      | <i>chéve</i>             |
| 2SG      | <i>ndéve</i>             |
| 1PL.INCL | <i>ñandéve</i>           |
| 1PL.EXCL | <i>oréve</i>             |
| 2PL      | <i>peẽme</i>             |
| 3        | <i>(i)chupe</i>          |

patient is singular, the prefix is *ro-*; if it is plural, *po-*. (The number of the agent is irrelevant.)<sup>5</sup> Consequently, for transitive verbs, active prefix marking can only appear if the patient is third person.

A sizeable number of vowel-initial verbal roots require the use of relational prefixes. The prefix *h-* is the one used with active prefixes, portmanteau person prefixes, imperative prefixes, voice prefixes, and, importantly, *poro-/mba'e-* prefixation. The prefix *r-* is used with inactive person prefixes.<sup>6</sup>

Finally, bare verbs in Paraguayan Guarani are non-future: when unmarked for tense, in the absence of context, telic predicates are preferentially interpreted as past, atelic predicates as present. However, a given context can support either interpretation.

Guarani is also a postpositional language. Direct objects show differential object marking. If a direct object is third person non-human, it is unmarked; if it is third person human, it is marked with the locative enclitic *=pe* ‘in, on, at’ (4). This is the same marking used for indirect objects expressed by noun phrases (5). The special pronouns in Table 2 are used for pronominal indirect objects:

<sup>5</sup> Of note, many speakers of Paraguayan Guarani use the form *poro-* for the 1>2PL portmanteau prefix. This use should not be confused with the antipassive use discussed in the paper.

<sup>6</sup> The non-possessed prefix *t-* yields a nominal interpretation. The *h-* and *r-* prefixes are also used with nominal roots to mark third-person pronominal and non-third-person pronominal possessors, respectively.

- (5) a. Mario ohayhu ijaryisýpe.  
 Mario o-h-ayhu i-jarýi+sy=pe  
 Mario 3.ACT-POSSM3-love 3.INACT-grandmother+mother=LOC  
 ‘Mario loves his great-grandmother.’
- b. Che ha nde jajoguáta juky.  
 che ha nde ja-jogua-ta juky  
 I and you.SG 1PL.INCL.ACT-buy-FUT salt  
 ‘You and I will buy salt.’
- (Estigarribia 2020: 133)

- (6) Ome’ẽ ñanderúpe peteĩ aranduka.  
 o-me’ẽ ñande-ru=pe peteĩ aranduka  
 3.ACT-give 1PL.INCL.INACT-father=LOC one book  
 ‘They gave a book to our teacher.’

Some verbs subcategorize for oblique complements marked with the locative enclitic =*re(he)* ‘at, by’ (note the irregular form *hese* ‘at him/her/it/them’); others, with the ablative enclitic =*gui* ‘from’; yet a few others, with the comitative enclitic =*ndi(ve)* (see Sect. 4.3).

A few verbal roots beginning with *j-* or *ñ-* show apheresis of this first segment in the forms with personal inactive prefixes ending in *e*. This apheresis also happens in cases of noun incorporation. For example, *japo* ‘to make’ shows apheresis to +*apo*, and *johéi* ‘to wash’ shows apheresis to +*héi*.

## 4 Properties of *poro-/mba'e-* Prefixation

In this section, we exemplify the main semantic and syntactic properties of *poro-/mba'e-* prefixation. We believe that, taken together, these properties support an antipassive analysis for these prefixes.

### 4.1 *Poro-* and *mba'e-* Encode Generic Semantic Patients

As we mentioned at the beginning, verbal predicates can encode their patient with the prefixes *poro-* if the patient is human or *mba'e-* if the patient is non-human. As examples (1) and (2) show, prefixation with *poro-* yields a plural interpretation almost always translatable by ‘people,’ although sometimes it can be interpreted as ‘somebody.’ On the other hand, the prefix *mba'e-* can yield singular or plural



interpretations. *Poro-/mba'e-* prefixation entails reduced specificity of the patient and is used for unknown or generic referents. This low individuation is a common property of antipassives (Cooreman 1994; Polinsky 2017; Heaton 2020). Note that verbs prefixed with *poro-/mba'e-* take active person marking. If the base root is relational, they take the *h-* prefix (see (12)).<sup>7</sup>

*Poro-/mba'e-* prefixation has a second semantic property that is often found in antipassives and sometimes even thought to be defining (Heaton 2017, 2020). We note here that these prefixes force a non-coreferential reading (even though that reading is often somewhat odd for pragmatic reasons).

- (7) a. Aheka peteĩ mitã chenupãva'ekue kuehe ha katu ndajuhúi (ichupe) gueteri.  
 a-h-eka peteĩ mitã che-nupã-va'e-kue  
 1SG.ACT-POSSM3-see one child 1SG.INACT-beat.up-ADJZ-POST  
 kuehe ha katu nd-a-juhu-i (ichupe) gueteri  
 yesterday and just NEG-1SG.ACT-find-NEG (them) yet  
 'I am looking for a child that hit me yesterday but I haven't found them yet.'
- b. ?Aheka peteĩ mitã chenupãva'ekue kuehe ha katu ndaporójuhúi gueteri.  
 a-h-eka peteĩ mitã che-nupã-va'e-kue  
 1SG.ACT-POSSM3-see one child 1SG.INACT-beat.up-ADJZ-POST  
 kuehe ha katu nd-a-**poro**-juhu-i gueteri  
 yesterday and just NEG-1SG.ACT-**ANTIP.H**-find-NEG yet  
 'I am looking for a child that hit me yesterday but I haven't found anybody yet.'
- (8) a. Kalo ipy'aro Kolándive ha oinupãka ichupe.  
 Kalo i-py'aro Kola=ndive ha oi-nupã-ka ichupe  
 Carlos 3.INACT-angry Nicolas=with and 3.ACT-beat.up-CAUS2 them  
 'Carlos was mad at Nicolas and had him punished / asked somebody to beat him up.'
- b. #Kalo ipy'aro Kolándive ha oporoinupãka.  
 Kalo i-py'aro Kola=ndive ha o-**poro**-inupã-ka  
 Carlos 3.INACT-angry Nicolas=with and 3.ACT-**ANTIP.H**-beat.up-CAUS2  
 'Carlos was mad at Nicolas and had asked somebody to beat someone else up.'

<sup>7</sup> Incidentally, this *h-* prefix and the *i-* infix of areal prefixes were originally third-person object markers. The fact that the *h-* prefix survives under *poro-/mba'e-* prefixation shows that this function has been lost.

This is related to the common antipassive function of “backgrounding of the logical object” (see Polinsky 2017 for details). These prefixes cannot be used to answer a question about the object or to refer anaphorically to a given specific referent.

- (9) a. -Mba'epa ojuka Juan. -#Juan omba'ejuka.  
 mba'e=pa o-juka Juan Juan o-**mba'e**-juka  
 what=Q 3.ACT-kill Juan Juan 3.ACT-ANTIP.NH-kill  
 ‘-What did Juan kill? (Intended) -Juan killed things.’
- b. -Mávapepa ojuka Juan. -#Juan oporojuka.  
 máva=pe=pa o-juka Juan Juan o-**poro**-juka  
 who=LOC=Q 3.ACT-kill Juan Juan 3.ACT-ANTIP.H-kill  
 ‘-Who did Juan kill? (Intended) -Juan killed people.’
- c. -Mba'epa ojogua Juan. -#Juan omba'ejogua.  
 mba'e=pa o-jogua Juan Juan o-**mba'e**-jogua  
 what=Q 3.ACT-buy Juan Juan 3.ACT-ANTIP.NH-buy  
 ‘-What did Juan buy? (Intended) -Juan bought things.’

Importantly, *poro-/mba'e-* prefixation contrasts with another grammatical option Paraguayan Guaraní has: object drop. Whereas *poro-/mba'e-* prefixation encodes a generic, non-referential object, null objects are used for recoverable, salient objects (Tonhauser 2017).<sup>8</sup> In the example below, the null object of *ahéchávo* “that/while I see” is interpreted as co-referential with *tekoha* “world(s).”

- (10) Che ahayhu tekoha hekokangmíva [ . . . ] Avy'a ahéchávo Ø oñembosa'yju ha pytãũ.  
 che a-h-ayhu t-ekoha h-eko-kangy-mi-va  
 I 1SG.ACT-POSSM3-love NPOSSM-dwelling NPOSSM-NMLZ.QUAL-weak-DIM-ADJZ  
 a-vy'a a-h-echa-vo o-ñe-mbo-sa'yju  
 1SG.ACT-joy 1SG.ACT-POSSM3-see-while 3.ACT-AGD-CAUS1-yellow  
 ha pytã+ũ  
 and red+dark  
 ‘I love delicate worlds [ . . . ] I like to see them tinged of yellow and dark red.’ (Galeano Olivera, translation of “Cantares” by Antonio Machado/Joan Manuel Serrat)

<sup>8</sup> This property was already noted by Restivo ([1724] 2010). In his description of Jesuitic Guaraní, he noted that a form like *ahapy* does not mean “I burn,” but “I burn it,” and that *poro-* is necessary to express the “absolute” (i.e., intransitive) meaning “I burn.”

In the following example, the null object of *rejuhúma* “you already found” is interpreted as co-referential with *che rembikañy* “my lost thing.”

- (11) -[ . . . ] nderejuhukái chéve che rembikañy. Mbytetérupi ame’ëta ndéve rejuhúmaguive Ø.  
 nde-re-juhu-ka-i chéve che-r-embikañy  
 NEG-2SG.ACT-find-CAUS2-NEG to.me 1SG.INACT-POSSM-NMLZ.REL-get.lost  
 mbyte-te=rupi a-me’ë-ta ndéve re-juhu-ma=guive  
 half-very=around 1SG.INACT-give-FUT to.you.SG 2SG.ACT-find-already= since  
 ‘[ . . . ] (if) you cannot help me find what I lost. I will give you half when you have already found it.’ (Meza 2010: 62)

Of interest, the antipassive morphemes enter into syntagmatic relations with other voice morphemes to yield specific interpretive effects. For example, it can co-occur with agent demoting *je-* to demote both the logical subject and the logical object and yield an interpretation that foregrounds the event.

- (12) Mba’éichapa ojejapóne ojeporohayhu haḡua [ . . . ].<sup>9</sup>  
 mba’éicha=pa o-je-japó-ne o-je-**poro**-h-ayhu  
 how=Q 3.ACT-AGD-make-DUB 3.ACT-AGD-**ANTIP.H**-POSSM3-love  
 haḡua  
 for  
 ‘How is it possible to love [ . . . ]?’

## 4.2 *The Aspectual Interpretation of Poro- and mba’e- Is Unconstrained*

Antipassive markers often entail atelicity and habitual readings. Dietrich (2017) makes this claim specifically for *poro-/mba’e-* prefixation in Paraguayan and Correntinean Guaraní. However, it is clear that both *poro-* and *mba’e-* are compatible with both telic and atelic interpretations, and they can yield both default habitual and default episodic readings, depending on the lexical aspect of the base predicate.

<sup>9</sup> Example from <https://guaraniete.wixsite.com/guaraniete/aaa>; accessed January 31, 2021.

- (13) a. aporohayhu [state]  
a-**poro**-h-ayhu  
1SG.ACT-ANTIP.H-POSSM3-love  
‘I love people.’ [habitual] / ‘I am in love.’ [episodic]
- b. Reporomoñani. [activity]  
re-**poro**-mo-ñani  
1SG.ACT-ANTIP.H-CAUS1-run  
‘You make someone run.’ [episodic]
- c. Ha’e oporojuvy. [accomplishment]  
ha’e o-**poro**-juvy  
they 3.ACT-ANTIP.H-strangle  
‘They strangled someone.’ [episodic]
- d. Che aporojura. [achievement]  
che a-**poro**-jura  
I 1SG.ACT-ANTIP.H-lasso  
‘I lassoed someone.’ [episodic] (Ávalos Ocampos 2017: 259)
- (14) a. Aporojapi.  
a-**poro**-japi  
1SG.ACT-ANTIP.H-shoot  
‘I shoot people (generally).’ or ‘I am shooting someone.’ or ‘I shot people/someone.’
- b. Upe aja umi mbohapy mbokavusu yembe’yguive oporojapi [. . .]  
upe aja umi mbohapy mbo-kavusu  
MED.SG while NPROX.PL three CAUS1-boom  
y+r-embe’y=guive o-**poro**-japi  
water+POSSM-edge=since 3.ACT-ANTIP.H-shoot  
‘In the meantime, the three cannons opened fire from the coast [. . .]’  
(Asociación Cultural Mandu’arã, Wed Sept 4, 2019, 7:29am;  
Accessible at <https://www.facebook.com/asociacion.manduarã/posts/10157705660681458>)

Therefore, we conclude that *poro-/mba'e-* prefixation does not entail atelicity or habitual aspect.

### 4.3 *Poro- and mba'e-* Cannot Express Non-patient Arguments

*Poro-/mba'e-* prefixation is only used to express logical objects that are patients, mostly with verbs that take direct objects, although this restriction can be violated. Examples (15–17) show that verbs whose complements are ablatives (marked with =*gui*) or comitatives (marked with =*ndive*) cannot undergo *poro-/mba'e-* prefixation.

- (15) a. Epoi chehegui.  
 e-poi chehegui  
 IMP-let.go from.me  
 ‘Let go of me.’
- b. \*Aporopoi.  
 a-**poro**-poi  
 1.SG.ACT-ANTIP.H-let.go  
 Intended: ‘I free someone/people.’
- (16) a. Cheresarái ndehegui.  
 che-r-esarái ndehegui  
 1.SG.INACT-POSSM-forget from.you.SG  
 ‘I forgot you.’
- b. \*apororesarái  
 a-**poro**-r-esarái  
 1.SG.ACT-ANTIP.H-POSSM-forget  
 Intended: ‘I forgot someone/people.’
- c. \*amba’eresarái  
 a-**mba’e**-r-esarái  
 1.SG.ACT-ANTIP.NH-POSSM-forget  
 Intended: ‘I forgot something.’ / ‘I forgot (things).’
- (17) a. Añemyrõma nendive.  
 a-ñemyrõ-ma ne=ndive  
 1.SG.ACT-get.angry-already you.SG=with  
 ‘I already got angry at you.’
- (Ávalos Ocampos 2017: 228)
- b. \*aporoñemyrõ  
 a-**poro**-ñemyrõ  
 1.SG.ACT-ANTIP.H-get.angry  
 Intended: ‘I got angry at someone/people.’

Verbs with a patient expressed as an oblique complement marked with =*re*(*he*) (whether active as in (18) or inactive as in (19)) usually do not undergo *poro-/mba’e-*prefixation. This would be expected if direct objecthood was a necessary syntactic condition for use of these prefixes.

- (18) a. Aikotevẽ nderehe.  
 ai-kotevẽ                      nde=rehe  
 1.SG.ACT-need              you.SG=LOC2  
 'I need you.'
- b. \*a(i)porokotevẽ  
 a(i)-**poro**-kotevẽ  
 1.SG.ACT-**ANTIP.H**-need  
 Intended: 'I need someone.'
- c. \*a(i)mba'ekotevẽ  
 a(i)-**mba'e**-kotevẽ  
 1.SG.ACT-**ANTIP.NH**-need  
 Intended: 'I need something.'
- (19) a. Chemandu'a nderehe.  
 che-mandu'a                      nde=rehe  
 1.SG.INACT-remembrance      you.SG=LOC2  
 'I remember you.'
- b. \*Aporomandu'a.  
 a-**poro**-mandu'a  
 1.SG.ACT-**ANTIP.H**-remembrance  
 Intended: 'I remember people/someone.'
- c. \*Amba'emandu'a.  
 a-**mba'e**-mandu'a  
 1.SG.ACT-**ANTIP.NH**-remembrance  
 Intended: 'I remember something.'

However, for some *=re(he)*-taking verbs, *poro-/mba'e-* prefixation is possible, most commonly with *poro-*, that is, when the patient is human.

- (20) a. aporoñangareko  
 a-**poro**-ñangareko  
 1.SG.ACT-**ANTIP.H**-take.care  
 'I take care of someone.'
- b. aporondyvu  
 a-**poro**-ndyvu  
 1.SG.ACT-**ANTIP.H**-spit  
 'I spit at someone.'
- c. aporopoko  
 a-**poro**-poko  
 1.SG.ACT-**ANTIP.H**-touch  
 'I touch someone.'

Examples with non-human patients are grammatical, but they appear to be rare in the modern language (21).

- (21) a. amba'eñangareko  
 a-**mba'e**-ñangareko  
 1.SG.ACT-ANTIP.NH-take.care  
 'I take care of things.'
- b. amba'endyvu  
 a-**mba'e**-ndyvu  
 1.SG.ACT-ANTIP.NH-spit  
 'I spit at something.'
- c. amba'epoko  
 a-**mba'e**-poko  
 1.SG.ACT-ANTIP.NH-touch  
 'I touch something.'
- d. amba'ejavyky  
 a-**mba'e**-javyky  
 1.SG.ACT-ANTIP.NH-steal  
 'I steal.' / 'I am a petty thief.'

It appears, therefore, that a necessary licensing condition is that the predicate has a patient argument, and it is sufficient (but not strictly necessary, although this is usually the case) that this argument map to a direct object in the syntax. Human =*re*(he)-marked patients can undergo *poro-/mba'e-* prefixation under conditions that are not well understood at this time.

#### 4.4 *Poro-/mba'e-* Derived Predicates Are Syntactically Intransitive

In this section, we will show that *poro-/mba'e* prefixation creates derived syntactically intransitive but semantically dyadic predicates. By *syntactically transitive*, we mean predicates that overtly express a subject and a direct object, whereas *syntactically intransitive* predicates only allow one core argument to be expressed in the syntax. At the same time, we differentiate this from the valency of predicates, that is, the number of semantic participants in the event whether they are overtly expressed or not, as *predicate adicity*, that is, monadic, dyadic, or triadic predicates.<sup>10</sup>

<sup>10</sup> The distinction between syntactic (in)transitivity and semantic valency/predicate adicity that is assumed here is discussed with respect to antipassives in more detail in Heaton (2017: 7–9). It must be noted that the question of what the relevant notions are to explain the expression of predicate arguments in Guaraní still remains open. The reader can refer to the illuminating work of Velázquez-Castillo (2002, 2008) for some proposals.









- (28) a. Okañy chehegui cheju'i.  
 o-kañy            chehegui    che-ju'i  
 3.ACT-get.lost    from.me    1SG.INACT-frog  
 'I lost my frog'
- b. mba'e kañymby  
 mba'e            kañy-mby  
 thing            get.lost-NMLZ.PASS  
 'hidden things'

This suggests that intransitives can be used with *-py* and that this suffix requires the presence of a semantic patient in the event structure (perhaps an unaccusative predicate), not strict syntactic transitivity. The suffix *-py* encodes a referent that has entered or been caused to enter into a state. In the case of *kañy* "to get lost," the addition of *-py* gives *kañymby* "(something in the state of being) lost."<sup>14</sup>

Since we haven't been able to demonstrate that *poro-/mba'e*-prefixed predicates are transitive, we can look at possible tests for intransitivity. Intransitive predicates have few identifying morphosyntactic properties other than not admitting a direct object. As stated above, and repeated here for clarity, person prefix marking is fixed lexically for each intransitive verb: some take active marking, some inactive. Since all *poro-/mba'e*-prefixed verbs take active marking, this would be compatible with these predicates being members of the active intransitive class.

A possible candidate for identifying intransitive predicates is the causative prefix *m(b)o-*. This prefix is usually described a causative for intransitive predicates (Velázquez-Castillo 2002, among others).

- (29) a. Opuka.  
 o-puka  
 3.ACT-laugh  
 'They laugh.'
- b. Ambopuka chesy'pe.  
 a-mbo-puka                            che-sy=pe  
 1SG.ACT-CAUS1-laugh    1SG.INACT-mother=LOC  
 'I make my mother laugh.'
- (30) a. Okañy.  
 o-kañy  
 3.ACT-get.lost  
 'It got lost.'
- b. Amokañy péva.  
 a-mo-kañy                            péva  
 1SG.ACT-CAUS1-get.lost    that  
 'I hid that.'

<sup>14</sup> The suffix takes the form *-mby* because of progressive nasal harmony triggered by the palatal nasal *ñ* in the stressed syllable of the root.

However, *mbo-* cannot be prefixed to a *poro-/mba'e-* marked predicate.

- (31) \*amboporohayhu  
 a-mbo-**poro**-h-ayhu  
 1SG.ACT-CAUS1-**ANTIP.H**-POSSM3-love  
 Intended: 'I make someone love (people).'

Does this mean that *poro-/mba'e-* marked predicates are not intransitive? We do not think so. Prefixation by *mbo-* requires that the predicate be monadic, rather than being syntactically intransitive. *Mbo-* introduces a causee that performs an action with no patient (like *puka* "laugh" in (29)) or makes a patient enter a state (like *kañy* "get lost" in (30)). But *poro-/mba'e-* marked predicates are dyadic. Hence, an inability to be prefixed by *mbo-* is not evidence that a *poro-/mba'e-* marked predicate is not intransitive.

We conclude that *poro-/mba'e* prefixation does indeed create derived syntactically intransitive predicates that are nonetheless semantically dyadic due to the presence of a generic patient in argument structure. For that reason, a morphological operation such as causativization by *mbo-*, in spite of the intransitivity of the derived predicate, is blocked by the presence of the underlying semantic patient.

#### 4.5 *Poro-/mba'e- Prefixation Is Productive*

Contrary to the claim by Gregores and Suárez (1967) cited above, these prefixes are quite productive in modern Paraguayan Guaraní. The examples in (32) show *mba'e-* used with a wide range of predicates; in (33), we see similar data for *poro-*.<sup>15</sup>

<sup>15</sup> We provide the following example without glossing to illustrate how widespread the use of these markers is, even with borrowed roots like *trata* (from Spanish *tratar* "to treat").

- (i) [...] *oime umi omandasénteva, oñemombaretéva, omba'ejokóva, ijavusívova, oporomongy-hyje jekuáva térâ ñembotavyhápe, ombojoavýva, oñembohorýva térâ oporomofíndýva, upéicha avei oí teko oporomongy'áva, oporombo'apo'íva, oí avei pe acoso moral, oporotrata vaitereíva ha pe meña ñemyangekói.* "[...] there are those who want to command, who are bullies, who put a stop to things, who are abusive, who make people afraid consciously or unconsciously, who generate conflicts, who humiliate or make people fearful, also there are those who slander, who look down on others, there is also moral harassment, those who mistreat others excessively and harass sexually" (Sanchís and Espinosa 2014: 52).

- (32) a. Amba'ejoguáta ko ka'aru.  
 a-**mba'e**-jogua-ta ko ka'aru  
 1SG.ACTIVE-ANTIP.NONHUM-buy-FUT this afternoon  
 'I will go shopping this afternoon.'  
 (Velázquez-Castillo 1995b: 673)
- b. Amba'eporandusemi ndéve.  
 a-**mba'e**-porandu-se-mi ndéve  
 1SG.ACT-ANTIP.NH-question-want-DIM to.you  
 'I would like to ask you something.'  
 (Estigarribia 2020: 213)
- c. [...] ařma pendehegui amba'ejerure haguã.  
 a-tř-ma pendehegui a-**mba'e**-jerure haguã  
 1SG.ACT-feel.shame-already from.you.PL 1SG.ACT-ANTIP.NH-petition for  
 '[...] I already feel shame coming to you to ask for things.'  
 (Zarratea 2012[1981]: 40).
- d. Péicha omba'emombe'úmi ta'yrape Mbatovigua ypykue.  
 pe-icha o-**mba'e**-mombe'u-mi t-a'yra=pe  
 MED.SG-as 3.ACT-ANTIP.NH-tell-used.to NPOSSM-son.of.father=LOC  
 Mbatovi=gua ypy-kue  
 Mbatovi=from origin-POST  
 'Those were the stories the elders from Mbatovi used to tell their  
 sons.' (lit.: 'Thus used to recount (things) the elders from Mbatovi  
 to their sons.')
- (Zarratea 2012[1981]: 28)
- e. Guaranikuéra omba'ejuka ha oipirakutu.  
 guarani=kuéra o-**mba'e**-juka ha oi-pira+kutu  
 guarani=PL 3.ACT-ANTIP.NH-kill and 3.ACT-fish-pierce  
 'The Guaraní hunted (animals) and fished.'  
 (GuaraniMeme 2015)
- f. Peñepyrũ pamba'emombe'u.  
 pe-ñepyrũ pe-**mba'e**-mombe'u  
 2PL.ACT-begin 2PL.ACT-ANTIP.NH-tell  
 'Begin to tell stories.' (lit.: 'begin to tell (things)')
- (Zarratea 2012[1981]: 22).

- (33) a. Aporohayhuse.  
a-**poro**-h-ayhu-se  
1SG.ACTIVE-ANTIP.HUM-POSSM3-love-FUT  
'I want to love people/someone.' / 'I want to be in love.'
- b. Aju aporombo'évo.  
a-ju a-**poro**-mbo'e-vo  
1SG.ACT-come 1SG.ACT-ANTIP.H-teach-while  
'I came to teach (everyone / the people).'
- (Melià et al. 1995, 123).
- c. Mbokapī ha'e peteī mboka ojeipuruha oporoñongatu tēra oporojuka haḡua.  
mbokapīha'e peteī mboka o-jei-puru-ha o-**poro**-ñongatu  
rifle be one firearm 3.ACT-AGD-use-NMLZ.AG 3.ACT-ANTIP.H  
protect  
tēra o-**poro**-juka haḡua  
or 3.ACT-ANTIP.H-kill for  
'A rifle is a firearm that is used to protect (people) or to kill (people).'
- (Estigarribia 2020: 214)<sup>16</sup>
- d. Upe ñe'ẽmbohasa oñeme'ẽ poropytyvõ haḡuaíchante avañe'ẽ jekuaápe.  
upe ñe'ẽ+mbo-h-asa o-ñe-me'ẽ  
MED.SG language+MAKE1-POSSM3-pass 3.ACT-AGD-give  
**poro**-pytyvõ=haḡua-icha-nte ava+ñe'ẽ je-kuaa=pe  
ANTIP.H-help=for-as-only person+language AGD-know=LOC  
'The translations are given only to aid (people, readers, others) in the  
understanding of Guarani.'
- (Fernández 2002)
- e. Pe "género" niko oporodiferencia [ . . . ]  
pe género niko o-**poro**-diferencia  
MED.SG gender VERD 3.ACT-ANTIP.H-differentiate  
'But this "gender" differentiates people [ . . . ].' (Sanchís and Espinosa  
2014: 34)

## 5 Discussion

In this section we contrast the properties of the antipassive with those of noun-incorporation, and we will briefly touch on the grammaticalization of *mba'e-* as a nominalizing prefix.

<sup>16</sup> Example modified from <https://gn.wikipedia.org/wiki/Mbokapī>

### 5.1 *Contrasting the Antipassive with Noun Incorporation*

As mentioned at the beginning, since the word *mba'e* means “thing(s),” there is a tacit consensus among linguists working on Guarani that *mba'e*- prefixation (and by extension, *poro*- prefixation) is simply noun incorporation, itself well attested in the language. However, *poro*- is not an independent word in modern Paraguayan Guarani, and it has never been reconstructed as an independent root in proto-Tupi-Guarani, to the best of our knowledge. Current reconstructions suggest that *poro*- was a prefix with the value “generic human referent” in proto-Tupi-Guarani (Cabral 2001) and that it became extended in many languages to function as a 2PL object prefix.<sup>17</sup>

Moreover, noun incorporation has syntactic and semantic properties not shared by *poro*-/*mba'e*- prefixation (Velázquez-Castillo 1995a, b). Guarani has incorporation of body-part and non-body-part objects. We have seen that antipassive arguments must be patients (see Sect. 4.3), but non-body-part incorporated arguments can be, for example, postpositional manner complements (34).

- (34) a. tatu+pyvoi  
armadillo+kick  
'to kick like an armadillo'  
(Velázquez-Castillo 1995b: 675)
- b. mbói+juka  
snake+kill  
'to beat someone as if killing a snake'  
(Velázquez-Castillo 1995b: 675)

Moreover, subjects of intransitives can incorporate, although perhaps only for unaccusative verbs like *mano* “to die,” as in (35a,b), or even subjects of stative non-verbal predications (35c).

<sup>17</sup> This extension is the source of the 1>2PL portmanteau prefix in the active set and of the use mentioned above of *poro*- as 1>2PL portmanteau prefix by some modern-day speakers of Paraguayan Guarani.

- (35) a. Hoguemano che ka'avo.  
 h-ogue+mano che-ka'avo  
 POSSM3-leaf+die 1SG.INACT-vegetable  
 'The leaves of my vegetable are withered.'
- b. Ani reñemokuremanóti tupápe.  
 ani re-ñe-mo-kure+mano-ti tupa=pe  
 NEG.IMP 2SG.ACT-AGD-CAUS1-pig+die-NEG bed=LOC  
 'Don't be apathetic during sex in bed.' (lit.: 'don't act like a dead pig in bed.')
- c. Oñemokureramaguy.  
 o-ñe-mbo-kure+rama=guy  
 3.ACT-AGD-CAUS1-pig+manioc.branches=under  
 'They play dumb.' (lit.: 'they make as if a pig under the manioc branches')

(Ávalos Ocampos 2017: 145)

Furthermore, whereas the contribution of *poro-/mba'e-* prefixation to the meaning of the predicate is always compositional and transparent, a crucial property of noun incorporation in Paraguayan Guarani is that it can give rise to more or less opaque lexicalization effects. This can be seen in the colloquial expressions in (35b, c). Also, compare the non-incorporated (36a) with the incorporated (36b).

- (36) a. Tani omondo chéve ñe'ẽ pya'e ahahaña hendape.  
 Tani o-mondo chéve ñe'ẽ pya'e a-ha=haña h-enda=pe  
 Tani 3.ACT-send to.me speech fast 1SG.ACT-go=for POSSM3-place=LOC  
 'Tani sent me a message to meet with him soon.'
- b. Tani py'yi oñe'ẽmondo chéve Paraguáguive.  
 Tani py'yi o-ñe'ẽ+mondo chéve Paraguay=give  
 Tani often 3.ACT-speech+send to.me Asunción=since  
 'Tani communicated often with me from Asunción.'

(Velázquez-Castillo 1995b: 678)

Unlike for *poro-/mba'e-* antipassives, noun incorporated predicates can be prefixed by the causative *mbo-*. For example, there are several frequently used predicates formed from the root 'u "to ingest," such as *y'u* "to drink water," *ka'u* "to drink caña," and *kay'u* "to drink mate." These predicates can take the causative prefix to yield *mboy'u* "to make someone drink water," *monga'u* "to make someone drunk (on caña)," and *mongay'u* "to make someone drink mate" (Velázquez-Castillo 1995b: 680).

Noun incorporation does not necessarily yield syntactically intransitive predicates. Unlike *poro-/mba'e-* antipassives, predicates with an incorporated object can take an overt direct object, often cognate as in the case of (37).



- (37) [...] avakuéra opirakutu opaichagua pira [...]  
 ava=kuéra o-**pira**+kutu opa-icha=gua **pira**  
 person=PL 3.ACT-**fish**+pierce all-like=from **fish**  
 '[...] people fish all sorts of fish [...]' ([https://gn.wikipedia.org/wiki/San\\_Antonio](https://gn.wikipedia.org/wiki/San_Antonio);  
 Last modified 19 February 2020; Accessed 6 January 2021)

Finally, noun incorporation has lost much of its original productivity: it survives mostly in lexicalized cases or with specific verbs and verb-object combinations. We have already mentioned that the root -'u "to ingest" forms several lexicalized predicates. Other verbs that commonly incorporate are *japo* "to make" (with apheresis to -*apo*), *johéi* "to wash" (with apheresis to -*héi*), *juka* "to kill," *kutu* "to pierce," and *pete* "to slap." However, not every combination is possible. For example, from *óga* "house" and *jogua* "to buy" or *h-eka* "to seek," one cannot form *oga+jogua* with the intended meaning "to buy houses (habitually)" or *oga+heka* with the intended meaning "to look for houses (habitually)" or "house-hunting."<sup>18</sup>

## 5.2 *Mba'e- as a Result of Grammaticalization*

Notably, *mba'e-* also functions as a prefix with derivational nominalizing function. As Sansò (2017) notes, agent nominalizations are a common source of antipassives.<sup>19</sup> However, in this case, the source for the non-human antipassive seems to be the generic/indefinite element "thing" (Sansò 2017: sect. 2.2). *Mba'e-* as an antipassive prefix probably results from the further grammaticalization of the incorporated generic/indefinite noun *mba'e* "thing." Note that according to currently accepted reconstructions, Proto-Tupi-Guarani had a generic human object prefix \**poro-*, but no non-human counterpart. It is possible that an original object incorporation of generic *mba'e+* grammaticalized into a non-human antipassive prefix by analogy with *poro-*, in order to complete the voice paradigm. Therefore, *mba'e* would have given rise to both an abstract nominalizer and an antipassive via two distinct processes of grammaticalization. In addition, there may be some permeability between the human antipassive and abstract nominalizations, since *poro-* can apparently also yield nominalized predicates, albeit extremely infrequently (39). A more thorough evaluation of these relationships, synchronically as well as diachronically, is beyond the scope of this chapter.

<sup>18</sup> A reviewer suggests that given the meaning of -*poro* (generic human) and *mba'e-* (generic non-human), they could perhaps be treated as classifiers that incorporate into the verb. However, since Paraguayan Guarani has no classifier system, we do not think there is an independent basis for such an analysis.

<sup>19</sup> We thank Raina Heaton for this observation.

- (39) porokuaita  
**poro**-kuaita  
 ANTIP.H-order  
 'commandment' (Ávalos Ocampos 2017: 258)

## 6 Conclusion

In this chapter, we have proposed that the Paraguayan Guaraní prefixes *poro*- and *mba'e*- are antipassive voice markers. This antipassive yields syntactically intransitive predicates, which are nonetheless semantically dyadic, with an understood patient that is generic/non-specific. We provided evidence in the discussion that this analysis is more adequate empirically than one, in terms of noun incorporation. Finally, if this analysis is correct, the implication for a cross-linguistic theory of antipassives would be that Guaraní is another known language where the antipassive markers are not syncretic with other voice or aspect markers, in agreement with Heaton (2017, 2020) and contra Polinsky (2017).

## References

- Ávalos Ocampos, Celso. 2017. *Ñe'ẽryruguasú (Gran diccionario) Guaraní-Español Español Guaraní*. Asunción: El Lector.
- Ayala, José Valentín. 1996. *Gramática Guaraní*. Buenos Aires, Argentina: Ministerio de Educación de la Nación.
- Cabral, Ana Suelly. 2001. O desenvolvimento da marca de objeto de segunda pessoa plural em Tupí-Guaraní. In *Estudos sobre Línguas Indígenas*, ed. Ana Suelly Cabral and Aryon Rodrigues, vol. I. Belém: Universidade Federal do Pará.
- Cerno, Leonardo. 2011. *Descripción fonológica y morfosintáctica de una variedad de la lengua guaraní hablada en la provincia de Corrientes (Argentina)*. Ph.D. Dissertation. Universidad Nacional de Rosario, Argentina. [http://etnolingüística.wdfiles.com/local%2D%2Dfiles/tese%3Acerno-2011/cerno\\_2011\\_corrientes.pdf](http://etnolingüística.wdfiles.com/local%2D%2Dfiles/tese%3Acerno-2011/cerno_2011_corrientes.pdf).
- Cooreman, Ann. 1994. A functional typology of antipassives. In *Voice: Form and function*, ed. B. Fox and P. Hopper, 49–88. Amsterdam: Benjamins.
- Dietrich, Wolf. 2017. Word classes and word class switching in Guaraní syntax. In *Guaraní linguistics in the 21st century*, ed. Bruno Estigarribia and Justin Pinta, 158–193. Leiden: Brill.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info>. Accessed 21 Oct 2020.
- Estigarribia, Bruno. 2020. *A grammar of Paraguayan Guaraní*, Grammars of world and minority languages. London: UCL Press.
- Fernández, Manuel F. 2002. *Guaraní Renda*. Jahaikue Aty. Available at [https://guaranirenda.tripod.com/jahaikue\\_aty/index.htm](https://guaranirenda.tripod.com/jahaikue_aty/index.htm). Accessed 20 Jan 2021.

- Gregores, Emma, and Jorge A. Suárez. 1967. *A description of Colloquial Guaraní*. The Hague/Paris: Mouton & Co.
- GuaraniMeme. 2015. *Origen de la ciudad de Lambaré-Táva lambare rembiasa'imi*. Available at <http://guanimeme.blogspot.com/2015/02/origen-de-la-ciudad-de-lambare-tava.html>. Created 14 February 2015; Accessed 20 Jan 2021.
- Guarania, Félix de. 2008. *Tabla sinóptica para una nueva gramática Guaraní (2ª ed.)*. Asunción, Paraguay: Servilibro.
- Heaton, Raina. 2017. *A typology of antipassives, with special reference to Mayan*. Ph.D. Dissertation. University of Hawai'i at Mānoa.
- . 2020. Antipassives in crosslinguistic perspective. *Annual Review of Linguistics* 6: 131–153.
- Jensen, Cheryl. 1990. Cross-referencing changes in some Tupí-Guaraní languages. In *Amazonian linguistics: Studies in Lowland South American languages*, ed. Doris Payne, 117–160. Austin: University of Texas Press.
- Jensen, Cheryl Joice S. 1998. Comparative Tupi-Guarani morphosyntax. In *Handbook of Amazonian languages*, ed. Desmond C. Derbyshire and Geoffrey K. Pullum, 489–618. Berlin: Mouton de Gruyter.
- Krivoshein de Canese, Natalia, and Graziella Corvalán. 1987. *El español del Paraguay*. Asunción: Centro Paraguayo de Estudios Sociológicos.
- Meliá, Bartomeu, Luís Farré Maluquer, and A. Pérez Peñasco. 1995. *El guaraní a su alcance*. Asunción: CEPAG.
- Meza, Miguel Ángel. 2010. *Perurima pypore. Las huellas de Perurimá*. (Translation into Spanish: Mauro Lugo.) Asunción: Servilibro.
- Polinsky, Maria. 2017. Antipassive. In *The Oxford handbook of ergativity*, ed. Jessica Coon, Diane Massam, and Lisa Demena Travis, 308–331. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198739371.013.13>.
- Restivo, Paulo. (1724) 2010. *Arte de la Lengua Guaraní*. Transcription: Silvio M. Liuzzi. Online Edition: Jean-Michel Hoppan, Marc Thouvenot and Silvio M. Liuzzi. CELIA. [http://celia.cnrs.fr/FichExt/Paleographies/Guarani/Guarani\\_intro.htm](http://celia.cnrs.fr/FichExt/Paleographies/Guarani/Guarani_intro.htm).
- Rodrigues, Aryon Dall'Igna. 1953. Morfologia do verbo Tupi. *Letras* 1:121–152.
- Sanchís, Norma, and Ana Carmen Espinosa. 2014. *Herramientas para la promoción de los derechos laborales de las trabajadoras de casas particulares. Tembiporu oñemotenonde hagua umi derecho laboral oguerékóva, ogapýpegua kuña mba'apohára*. Córdoba: Fondo Mujeres del Sur. Accessible at [https://intercoonecta.aecid.es/Gestin%20del%20conocimiento/Manual\\_TRADO\\_Paraguay.pdf](https://intercoonecta.aecid.es/Gestin%20del%20conocimiento/Manual_TRADO_Paraguay.pdf).
- Sansò, Andrea. 2017. Where do antipassive constructions come from? A study in diachronic typology. *Diachronica* 34 (2): 175–218.
- Tonhauser, Judith. 2017. The distribution of implicit arguments in Paraguayan Guaraní. In *Guaraní Linguistics in the 21st century*, ed. Bruno Estigarribia and Justin Pinta, 194–230. Leiden: Brill.
- Velázquez-Castillo, Maura. 1995a. Noun incorporation and object placement in discourse. The case of Guaraní. In *Word order in discourse*, Typological studies in language 30, ed. Pamela A. Downing and Michael Noonan, 555–579. Amsterdam: John Benjamins Publishing Company.
- . 1995b. Noun incorporation in Guaraní: A functional analysis. *Linguistics* 33: 673–709.
- . 2002. Guaraní causative constructions. In *The grammar of causation and interpersonal manipulation*, Typological studies in language 48, ed. Masayoshi Shibatani, 507–534. Amsterdam: John Benjamins.
- . 2008. Voice and transitivity in Guaraní. In *The typology of semantic alignment*, ed. Mark Donohue and Søren Wichmann, 380–395. Oxford: Oxford University Press.
- Zarratea, Tadeo. 2012. *Kalaíto Pombéro Guaraníme*. Asunción: Servilibro. First edition 1981.

# Argument Structure and Morphology in Cochabamba Quechua (with Occasional Comparison with Other Quechua Varieties)



Neil Myler

## 1 Introduction<sup>1</sup>

In his groundbreaking discussion of Quechua word structure, Muysken (1981) argued, against a long tradition in the study of indigenous languages of the Americas (see, e.g., Pike 1945 on Mixtec; Weber 1976, 1983 on Huallaga Quechua), that the Quechua data pointed to the need for a strict division between morphology and syntax. Despite the pro-lexicalist stance of Muysken (1981), that paper went on to become one of the principal sources for Mark Baker in his establishment of the Mirror Principle (Baker 1985), which has since often been taken to strongly favor syntacticizing approaches to morphology (though see Grimshaw (1986) and Ackema and Neeleman (2004) for arguments that the Mirror Principle is equally compatible with lexicalist models).

In the four decades since Muysken's celebrated paper, numerous discussions of Quechua morphology have appeared, both from lexicalist and syntacticist standpoints. These discussions have involved many different types of morphology, but the so-called valency-changing morphology has been especially prominent since the very beginning: it was specifically Muysken's discussion of interactions between causatives, reflexives, and reciprocals (Muysken 1979, 1981: 295–297; 305–307) that went on to inform the Mirror Principle. A number of important theoretical and descriptive works have since appeared on these and other valency-changing suffixes, but to my knowledge, there are no published synopses of this literature, and not since Muysken's various related papers of the 1980s (1986,

---

<sup>1</sup> This paper is dedicated to the memory of Pieter Muysken.

---

N. Myler (✉)  
Boston University, Boston, MA, USA  
e-mail: [myler@bu.edu](mailto:myler@bu.edu)

1988, 1989) has there been a thoroughgoing discussion of the interactions of such morphemes and their implications for morphological theory. During the same period, a great many theoretical developments in the literature on argument structure and morphology have taken place (see Wood and Myler 2019 for a recent overview). A reassessment of valency-changing morphology in Quechua, taking into account subsequent empirical and theoretical advances, is therefore long overdue.

The purpose of this chapter is to begin such a reassessment by examining the interaction between argument structure and morphology in Cochabamba Quechua, a Quechua IIC variety of Bolivia (on the properties of this variety more generally, see Albó 1970; Bills et al. 1969; Lastra 1968; van de Kerke 1996). I will occasionally draw comparisons to Santiago del Estero Quechua (a Quechua IIC variety of Argentina), Cajamarca Quechua (a Quechua IIA variety of northern Peru), and certain Quechua I varieties of central Peru when the differences between these varieties and Cochabamba Quechua are especially instructive. The focus will be on the causative morpheme *-chi*, the so-called reflexive marker *-ku*, and the applicative morpheme *-pu*, as well as the interactions between them. We start with a brief overview of the *dramatis personae*.

The causative marker *-chi* is illustrated by the following examples:

(1) **Causative *-chi* in Cochabamba Quechua**

- |    |                                 |             |                  |
|----|---------------------------------|-------------|------------------|
| a. | Juan                            | tusu-n.     |                  |
|    | Juan                            | dance-3SUBJ |                  |
|    | 'Juan dances.'                  |             |                  |
| b. | Maria                           | Juan-ta     | tusu-chi-n.      |
|    | Maria                           | Juan-ACC    | dance-CAUS-3SUBJ |
|    | 'Maria makes Juan dance.'       |             |                  |
| c. | Juan Maria-ta                   | maylla-n.   |                  |
|    | Juan Maria-ACC                  | wash-3SUBJ  |                  |
|    | 'Juan washes Maria.'            |             |                  |
| d. | Gladys Juan-wan                 | Maria-ta    | maylla-chi-n.    |
|    | Gladys Juan-with                | Maria-ACC   | wash-CAUS-3SUBJ  |
|    | 'Gladys makes Juan wash Maria.' |             |                  |

Causativization in *-chi* has different consequences for argument marking depending on the arity of the main verb, in a familiar way. Specifically, the causee is marked with accusative case in the causative of an intransitive verb (compare (1a) and (1b)), but with the comitative/instrumental postposition *-wan* in the causative of a transitive verb (compare (1c) and (1d)).

The reflexive marker *-ku* is exemplified in (2).

(2) **Reflexive *-ku* in Cochabamba Quechua**

- a. Juan Marya-q chujcha-n-ta peyna-n.  
 Juan Maria-GEN hair-3POSS-ACC comb-3SUBJ  
 ‘Juan combs Maria’s hair.’
- b. Juan peyna-ku-n.  
 Juan comb-REFL-3SUBJ  
 ‘Juan combs himself.’

As has been observed for similar markers in many other languages, the Quechua “reflexive” marker *-ku* in fact has a number of other functions. Among these are marking a subset of verbs in the intransitive alternant of the anti-causative alternation as in (3), an autobenefactive/emotive use as in the first reading of (4), an anti-assistive use as in the second reading of (4), and a handful of idiosyncratic uses illustrated in (5).

- (3) a. Wayna qiru-ta p’aki-rqa. (Cochabamba Quechua)  
 boy glass-ACC break-PAST  
 ‘The boy broke the glass.’
- b. Qiru p’aki-ku-rqa.  
 glass break-REFL-PAST  
 ‘The glass broke.’
- (4) Tusuq tusu-ku-n. (Cochabamba Quechua)  
 dancer dance-REFL-3SUBJ  
 ‘The dancer dances (and gets really into it/enjoys it).’  
 ‘The dancer dances by himself (i.e., no one helps him).’
- (5) Munay ‘to want’ munakuy ‘to love’ (Cochabamba Quechua)  
 Yachay ‘to know’ yachakuy ‘to learn’

Finally, the applicative suffix *-pu* is illustrated in (6). While a benefactive use for this suffix is apparent in all dialects in which the suffix is productive, what uses *-pu* has beyond this is a matter of cross-linguistic variation within the family.

- (6) Juan Maria-paq tusu-pu-n. (Cochabamba Quechua)  
 Juan Marie-BEN dance-APPL-3SUBJ  
 ‘Juan dances for Mary.’

The question of how these morphemes interact is of central interest here. Of the twelve logically possible combinations, only two turn out to be permitted on the surface in Cochabamba Quechua:

- (7) a. <sup>OK</sup>chi-ku      g. \*-chi-ku-pu  
 b. \*ku-chi      h. \*ku-chi-pu  
 c. <sup>OK</sup>chi-pu      i. \*-ku-pu-chi  
 d. \*pu-chi      j. \*-chi-pu-ku  
 e. \*ku-pu      k. \*-pu-chi-ku  
 f. \*pu-ku      l. \*-pu-ku-chi

The generalizations listed in (7) can be distilled into two templatic statements, as follows:

- (8) *-chi* may co-occur with either of the other two suffixes, but must always come first.  
 (9) *-ku* and *-pu* can never co-occur.

One could encode (8) and (9) in the grammar as templatic ordering constraints (perhaps as constraints on bigrams, as in Ryan 2010) and leave the matter there. But I won't. Adopting a syntacticizing approach to morphology and argument structure, I will argue for the following conclusions: (i) the productive causative<sup>2</sup> *-chi* in Cochabamba Quechua is a verb-selecting causative in the sense of Pylkkänen (2008), which is to say that it c-selects a structure of category *v*P, not a VoiceP or a High ApplP; (ii) *-ku* is a pronominal clitic which always occupies some argument position in the structure, thereby effecting valency reduction, before moving and adjoining to the closest available Voice head; (iii) following Myler (2016, 2018), *-pu* is a High Applicative morpheme in Pylkkänen's (2008) sense, but it is not (always) an argument-introducer, instead (sometimes) serving as a landing site for movement of oblique arguments (hence, it is a "raising applicative" in the sense of Georgala et al. 2008; Paul and Whitman 2010; Georgala 2012; Nie 2019, 2020); and (iv) Cochabamba Quechua has a morphophonological rule deleting *-pu* when

<sup>2</sup> As well as appearing in productive causatives, there are lexicalized instances of *-chi* in some verbs, some of which alternate with *-ku* to mark the (anti-)causative alternation. I will not discuss lexicalized instances of *-chi* in this paper, but I assume that they spell out a little-*v* head, which is attached directly to the main verb root (and are hence root-selecting causatives in the taxonomy of Pylkkänen 2008).

it would precede *-ku*.<sup>3</sup> This account is superior to one involving surface affix-order templates, because, beyond the affix ordering, it explains a number of things that such surface templates have no bearing on. Individually, (i)–(iii) explain the syntactic and semantic properties of forms containing these morphemes in isolation. Collectively, they also explain the permissible ways in which these morphemes interact syntactically and semantically when combined. While (iv) is of course stipulated, we shall see that it explains a couple of otherwise mysterious semantic properties of *-ku*, and it correctly predicts that the ban on the order *\*-ku-pu* is of a much deeper sort than the ban on the order *\*-pu-ku*. The affix-ordering facts in (7) follow straightforwardly; there is no need for a template once the syntactic and semantic properties of the suffixes involved are properly analyzed.

The rest of the chapter is organized as follows. Section 2 introduces my assumptions about the nature of the thematic domain and goes on to present the arguments that productive *-chi* is a verb-selecting causative in Cochabamba Quechua. Section 3 turns to *-ku*, showing how a clitic analysis can account for its various uses and for the way *-ku*'s reflexive interpretation interacts with the causative *-chi*. In Sect. 4, I briefly rehearse arguments from Myler (2016, 2018) to the effect that *-pu* is a high applicative morpheme, and more specifically a “raising” applicative, whose variety of interpretations (and the variation across the family in which interpretations are available to it) are to be attributed to variation in the material that comes to occupy *-pu*'s specifier position. As already shown in Myler (2016), and recapitulated here, this account correctly predicts *-pu*'s semantic interactions with causative *-chi*, including ones that appear to violate the Mirror Principle. Section 5 turns finally to interactions of *-ku* and *-pu*. While *-ku* and *-pu* may not co-occur on the surface in Cochabamba Quechua, I will argue that syntactic structures combining the two are in fact attested (but they are obscured on the surface by the *-pu*-deletion rule alluded to in the previous paragraph). Section 6 summarizes the analysis and concludes with some general methodological remarks about how the interaction of affix order and interpretation should be investigated.

## 2 The Thematic Domain and *-chi*'s Place in It

The approach to the thematic domain adopted here is the one overviewed in Marantz (2013), which has been developed by Marantz and many others over a number of

---

<sup>3</sup> Suffix deletion rules of this sort have been proposed before in the Quechua literature, for other suffixes, but I am not aware of direct antecedents for a rule of *-pu*-deletion. Muysken (1988: 267, 1989: 44) proposes a rule deleting *-ku* when it would precede *-chi* in some varieties (this specific rule will not play a role in my analysis of Cochabamba Quechua). Weber (1976: 72–73) proposes rules deleting the first-person inclusive agreement suffix and the genitive case suffix under certain circumstances in Huallaga Quechua (a Quechua I language).

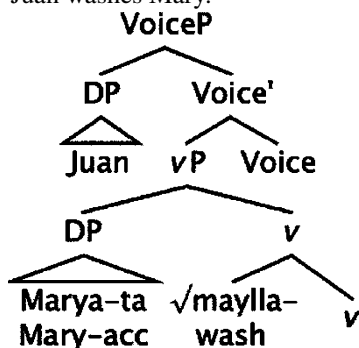


years.<sup>4</sup> A basic transitive clause in Cochabamba Quechua would, on this view, have the following partial structure:

- (10) Juan Marya-ta maylla-n. (Cochabamba Quechua)

Juan Mary-ACC washes-3SUBJ

‘Juan washes Mary.’



Here, Voice is the head responsible for introducing external arguments (Kratzer 1996 et seq.). “v” is responsible for verbalizing the root, which is adjoined to it. The resulting complex “v” head then merges with the direct object, which is thus the complement of the verb, as on standard analyses.

Let us now consider the case of the causative of a transitive clause like (10). In Cochabamba Quechua, the causatives of transitive verbs have the causee (i.e., the logical subject of the causativized predicate, *maylla-* “wash” in this example) surfacing in the instrumental case, as shown in example (11).<sup>5</sup>

- (11) Gladys Juan-wan Marya-ta maylla-chi-n. (Cochabamba Quechua)

Gladys Juan-with Mary-ACC wash-CAUS-3SUBJ

‘Gladys makes Juan wash Mary.’

As anticipated in the introduction, my claim regarding *-chi* is that it is a *verb-selecting causative* in the taxonomy of Pyllkänen (2008), rather than a *phase-*

<sup>4</sup> Analyses within this broad framework applied to various phenomena can be found in Alexiadou et al. 2015, Bruening 2013, Kastner 2020, Kratzer 1996, Myler 2016, Myler and Mali 2021, Nie 2020, Oseki 2017, Pyllkänen 2008, Schäfer 2008, Tyler 2020, Wood 2015, Wood and Myler 2019, and references cited in those works.

<sup>5</sup> Other Quechua varieties have somewhat different case-marking strategies for transitive causees, which include case alternations between instrumental, dative, and (in some rare cases) accusative. These alternations are often semantically consequential (frequently with regard to the level of volitionality of the causee; see especially Muysken 1979, Schoenfeld 2008). These are not at issue in the Cochabamba Quechua of the speakers I have worked with. For them, the causees of productive causatives of transitives are invariably marked with the instrumental postposition (although instrumental/accusative alternations do seem to have been permitted in the doculect of Bills et al. 1969: 98). For reasons of space and time, I cannot discuss this variation here. See also Treviño (1991, 1992, 1994) on similar variation in Spanish.



- (14) Gladys sach'a{-ta/\*-wan} urma-chi-rqa. (Cochabamba Quechua)  
 Gladys tree-ACC/with fall-CAUS-PAST  
 'Gladys {made the tree fall/felled the tree}.'

To account for this, I will assume that *-chi* is required to enter into a syntactic relationship of some kind with a DP in its c-command domain. This requirement could be encoded in a number of ways depending on one's other theoretical commitments (it could be implemented as an Agree requirement, or in terms of an abstract Case feature that *-chi* must discharge). The ungrammaticality of marking the causee with *-wan* in (13) and (14) will then follow: in the causative of an intransitive, embedding the causee in a PP will leave *-chi* bereft of a way to satisfy its needs.

In the rest of this section, I will present general arguments that the structure in (12) is the correct one for Cochabamba Quechua, before turning to a brief comparison with Tarma Quechua as analyzed by Muysken (1979, 1981). As we will see, the two varieties differ in a number of ways, which are explained if *-chi* can select a VoiceP or a  $\nu$ P in Tarma Quechua, but must select a  $\nu$ P in Cochabamba Quechua. My analysis is therefore one in which the locus of variation in the causative is the size of the verbal substructure embedded by the causative morpheme itself, a claim familiar from work on causatives in other language families (see Folli and Harley 2007; Harley 2017; Jung 2014; Key 2013; Legate 2014; and Piteroff and Campanini 2013, among many others).

Many diagnostics for the verb-selecting vs. Voice-selecting causative distinction revolve around the properties of the causee. The causee in a Voice-selecting causative construction is first-merged in spec-VoiceP, just as the external arguments of transitives usually are, and will hence pattern with subjects for the purposes of binding theory and agentive modification. In contrast, the causee in a verb-selecting causative will instead pattern against subjects with respect to the same diagnostics. A binding-theoretic diagnostic that is applicable here concerns reflexive-marking with *-ku*, which is subject-oriented in Quechua languages (see, among others, Muysken 1981: 456 on Tarma Quechua; Weber 1976: 56–57 on Huallaga Quechua; and van de Kerke 1996: 30–31 on Bolivian Quechua). In causative constructions, *-ku* can be bound by the causer argument, but not by the causee in Cochabamba Quechua (15). There is no way to convey "Juan makes Maria wash herself" using the causative and reflexive morphology in the language (Muysken 1988: 270 shows that the same is true in Chumbivilcas Quechua).<sup>6</sup>

<sup>6</sup> Unfortunately, I have not succeeded in identifying agentive adverbials or other modifiers in Cochabamba Quechua that would allow me to use this standard diagnostic for the presence of a VoiceP embedded under *-chi*. Another potential diagnostic that I haven't presented here pertains to Principle B of the binding theory: if a VoiceP is present, then the causee should act as a subject for the purposes of determining a binding domain, meaning that pronouns c-commanded by the causee should be able to be bound by an argument above the causee without inducing a Principle B violation. In a verb-selecting causative construction, however, the same configuration will trigger a Principle B violation. The relevant cases do pattern as expected in Cochabamba Quechua, as shown by these examples:

- (15) Juan Maria-wan maylla-chi-ku-n. (Cochabamba Quechua)  
 Juan Maria-with wash-CAUS-REFL-3SUBJ  
 ‘Juan has himself washed by Maria.’  
**NOT:** \*‘Juan makes Maria wash herself.’

A second fact about Cochabamba Quechua that is explained on the assumption that *-chi* in this variety is a verb-selecting causative has to do with affix order. Note that (15) would seem to violate the Mirror Principle: *-ku* relates to an argument associated with the embedded event, not the causing event, yet it surfaces outside *-chi*. Furthermore, forms in which *-ku* precedes *-chi* are invariably ungrammatical in the language.

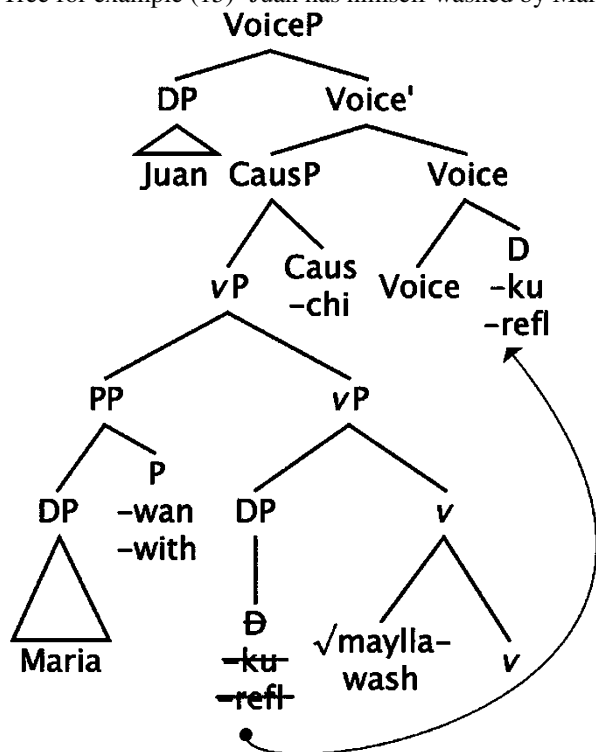
- (16) \*Maylla-ku-chi-n. (Cochabamba Quechua)  
 wash-REFL-CAUS-3SUBJ  
 ‘S/he has someone wash himself/herself.’

I will argue in the next section that *-ku* is an argument clitic and that it raises to adjoin to the nearest Voice head. Given this conclusion, the unavailability of the order *-ku-chi* is explained without further ado if *-chi* is a verb-selecting causative: *-chi*’s complement domain in Cochabamba Quechua is too small to accommodate VoiceP; therefore, it is also too small to contain *-ku*’s landing site. In any derivation in which *-ku* merges in an argument position lower than *-chi*, it will inevitably have to raise above *-chi* to find a landing site, invariably giving rise to the order *-chi-ku*. This also immediately dissolves the apparent problem for the Mirror Principle example (15) poses. This is illustrated in the following partial structure for that example.

- |      |                         |            |                      |                      |
|------|-------------------------|------------|----------------------|----------------------|
| (i)  | *Nuqa                   | Maria-wan  | maylla-chi-wa-ni.    | (Cochabamba Quechua) |
|      | I                       | Maria-with | wash-CAUS-1OBJ-1SUBJ |                      |
|      | ‘I make Maria wash me.’ |            |                      |                      |
|      |                         |            |                      |                      |
| (ii) | Nuqa                    | Maria-wan  | maylla-chi-ku-ni.    | (Cochabamba Quechua) |
|      | I                       | Maria-with | wash-CAUS-REFL-1SUBJ |                      |
|      | ‘I make Maria wash me.’ |            |                      |                      |

In (i), the relevant configuration with a first-person object marker (which, as Myler 2017 shows, is an object clitic pronoun) is ungrammatical; instead, reflexive *-ku* must be used to convey this meaning. This is the opposite of English *make* and *have* causatives, which are known to embed a substructure large enough to contain a VoiceP (and, indeed, more structure besides—see Bjorkman and Cowper 2013 for some discussion). However, because first-person (and, indeed, second-person) object clitics raise relatively high in Quechua languages, it could be that (i) would inevitably violate Principle B even in a variety with Voice-selecting causatives. Not having been able to work on such a variety directly, I have not been able to test this prediction.

- (17) Tree for example (15) 'Juan has himself washed by Maria'.



This account makes an interesting prediction about the micro-typology of affix ordering in the Quechua family. The prediction is this: if a Quechua language allows the affix order *-ku-chi*, it should also allow *-chi-ku*, but the converse should not hold. The order *-chi-ku* will always be available because it exploits the Voice head above *-chi*, present by hypothesis in all Quechua varieties; *-ku-chi* is possible in a variety only if, in addition, *-chi* can embed a structure large enough to contain a VoiceP in that variety. This prediction is apparently correct (Muysken 1979: 471, his (78)).

Let us now consider the phrase structural status assigned to the causee in this analysis (illustrated by the structures in (12) and (17)). The claim is that it is a PP adjunct in syntactic terms, not an argument DP, despite the fact that it appears to be interpreted as an argument of the embedded predicate semantically. This explains the fact that the causee is not a potential binder of *-ku* (it is not a subject), and it also gives rise to two further predictions: (i) the causee should not be eligible to be doubled by an object clitic, as a DP would be, and (ii) the causee, being an adjunct in syntactic terms, ought to be optional. Both predictions are correct.<sup>7</sup>

<sup>7</sup> Admittedly, given that Quechua allows null objects, the optionality of this PP is not as dispositive as it would be in some other languages. However, the fact that the implicit causee in such examples

- (18) Juan qam-wan nuqa-ta chiqni-chi-wa-n. (Cochabamba Quechua)  
 Juan you-with I-ACC hate-CAUS-1OBJ-3SUBJ  
 ‘Juan made you hate me.’
- (19) \*Juan qam-wan nuqa-ta chiqni-chi-su-nki.<sup>8</sup> (Cochabamba Quechua)  
 Juan you-with I-ACC hate-CAUS-2OBJ-2SUBJ  
 ‘Juan made you hate me.’
- (20) Juan nuqa-ta chiqni-chi-wa-n. (Cochabamba Quechua)  
 Juan I-ACC hate-CAUS-1OBJ-3SUBJ  
 ‘Juan made {him/her/someone} hate me.’

Example (18) shows that the direct object of the embedded predicate *chiqni* “hate” is eligible to be clitic doubled, as we would expect given its status as a DP. Indeed, it *must* be clitic doubled in Cochabamba Quechua, since clitic doubling is obligatory when possible in the language. Example (19) shows that the same privilege/obligation does not extend to the causee. Finally, example (20) demonstrates that the causee need not be included (with its surface absence being interpretable either as a null definite pronominal or through existential closure).

Before closing, let us address a pressing semantic question about the proposed structures: given that VoiceP is not present in the structure below CausP, how exactly is the external thematic role associated with the embedded predicate assigned to the causee? This issue of course arises in all languages that appear to have verb-selecting causatives, but the literature on such causatives since Pylkkänen (2008) has little to say on the matter. The expectation that the framework adopted here leads to, however, is that languages should vary in terms of how they resolve the semantic problem of integrating the causee of a verb-selecting causative into the composition. Further, this variation should be related in a principled way to the variation we see in the syntactic status of causees in verb-selecting causatives (on which see Harley 2017). The proposal I would like to make for Cochabamba Quechua is this: the causee is integrated into the event structure of the embedded predicate via a thematic role assigned to it by the instrumental postposition *-wan*.

---

need not be interpreted as definite suggests that a null pronoun is not the only available structure for such examples. Moreover, the clitic-doubling diagnostic is not vulnerable to the same potential objection.

<sup>8</sup> On the fact that the subject agreement morpheme in this example would reflect the features of the second-person object rather than those of the third-person subject, see Myler (2017) and references cited there.

The correctness of this approach is suggested by the surprising fact that *-wan*-marked causees in Cochabamba Quechua are in complementary distribution with “true” instruments, even if the two *-wan*-marked phrases are not adjacent to each other, as shown in (21). Either one of the PPs is fine on its own, as shown in (22) and (23).

- (21) \*Nuqa Juan-wan (pay-ta) k’aspi-wan maqa-chi-rqa-ni. (Coch. Quechua)  
 I Juan-with s/he-ACC stick-with beat-CAUS-PAST-1SUBJ  
 ‘I had Juan beat him/her with a stick.’
- (22) Nuqa Juan-wan maqa-chi-rqa-ni. (Cochabamba Quechua)  
 I Juan-with beat-CAUS-PAST-1SUBJ  
 ‘I had Juan beat him/her.’
- (23) Nuqa k’aspi-wan maqa-chi-rqa-ni. (Cochabamba Quechua)  
 I stick-with beat-CAUS-PAST-1SUBJ  
 ‘I had him/her beaten with a stick.’

More concretely, my proposal is this: *-wan* introduces a role which is like the traditional concept of *instrument* but distinct in that it is indifferent to the sentence/(in)animacy of its entity argument. It could be, as Jerro (2016, 2019) suggests for a Kinyarwanda morpheme that doubles as a causative and an instrumental applicative, that *-wan* encodes that an individual is an intermediate actor in a chain of causation. Example (21) is then excluded because it is internally contradictory: two entities are simultaneously asserted to be playing the same role in the same event.

Of course, this still leaves open the question of how the causees of intransitives receive their thematic roles, given that marking with *-wan* is unavailable for them as we saw in (13) and (14). For unaccusatives the answer is straightforward; the subjects of such predicates merge as the complement of their associated verb (or as the subject of a small clause which is itself the complement of the verb—see Irwin 2012). The solution for unergatives is less obvious. But if, following Tollan (2018), unergative external arguments are introduced in spec-*v*P rather than spec-VoiceP, then this problem evaporates also.

To summarize, we have seen several arguments that *-chi* in Cochabamba Quechua is a verb-selecting causative, in which the causee of a transitive verb is introduced in

a PP headed by the instrumental(-like) postposition *-wan*: the causee lacks subject properties (illustrated here via the diagnostic of reflexivization), cannot be clitic-doubled as it ought to be if it were a DP rather than a PP, and is in complementary distribution with canonical instruments introduced by the same postposition. The verb-selecting causative analysis also accounts for the fact that *-chi* must precede *-ku* in affix order in Cochabamba Quechua.

In all of these respects, Cochabamba Quechua differs from Tarma Quechua, a central Peruvian variety from the Quechua I branch of the family. Example (24) shows that the order *-ku-chi* is allowed in Tarma Quechua and that the causee can be the binder of reflexive *-ku* in this variety; example (25) shows that the order *-chi-ku* is also available, with the same meaning possibilities as in Cochabamba Quechua (these examples are from Muysken 1979: 457, his examples (34) and (35); the glosses are mine but differ trivially from Muysken's, and the translations given are his).

- (24) Maqa-ku-chi-n. (Tarma Quechua)  
 beat-REFL-CAUS-3SUBJ  
 'He causes someone<sub>i</sub> to beat himself<sub>i</sub>.'
- (25) Maqa-chi-ku-n. (Tarma Quechua)  
 beat-CAUS-REFL-3SUBJ  
 'He<sub>i</sub> causes someone to beat him<sub>i</sub>.'  
 'He lets himself be beaten.'

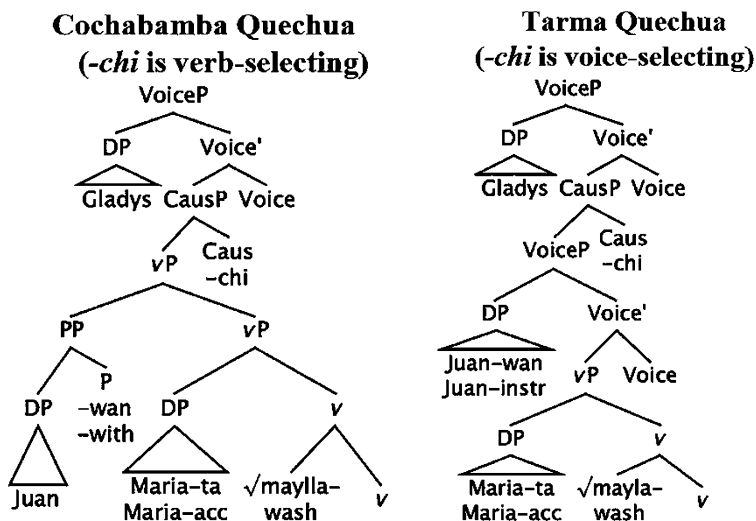
Similarly, in Tarma Quechua, either one of the causee or the direct object of the embedded verb can be clitic doubled, given that it has the appropriate features (these examples are from Muysken 1981: 307, his (70); again I have adjusted the glosses minimally to accord with the conventions of this chapter).

- (26) Ñuqa-wan Mañuku-ta maqa-chi-ma-n. (Tarma Quechua)  
 I-INSTR Manuel-ACC beat-CAUS-1OBJ-3SUBJ  
 'He causes me to beat Manuel.'
- (27) Mañuku-wan ñuqa-ta maqa-chi-ma-n. (Tarma Quechua)  
 Manuel-INSTR I-ACC beat-CAUS-1OBJ-3SUBJ  
 'He causes Manuel to beat me.'



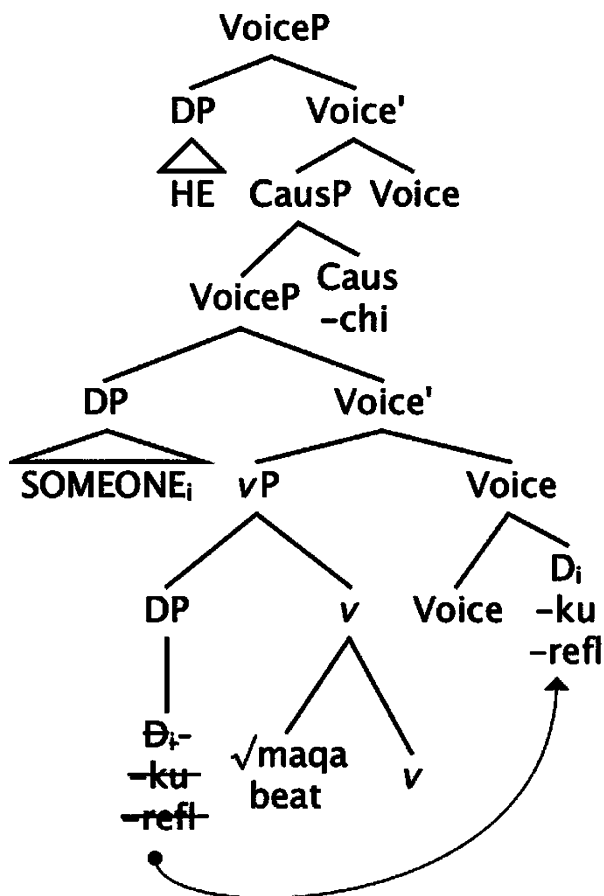
All this suggests that, despite being marked identically on the surface, the status of the causee in Tarma Quechua can be radically different from that of its Cochabamba counterpart: rather than being a PP adjoined to embedded  $vP$  in a structure like (26), it is a DP marked with instrumental case, occupying the specifier of a Voice phrase embedded underneath causative *-chi*. For convenience, I present partial structures for the same construction in each variety here (the verb “wash” is *mayla* in Tarma Quechua rather than *maylla* as in Cochabamba Quechua, but the two languages are otherwise string identical with respect to this example, for which reason I only present the example itself once):

- (28) Gladys Juan-wan Marya-ta maylla-chi-n.  
 Gladys Juan-with Mary-ACC wash-CAUS-3SUBJ  
 ‘Gladys makes Juan wash Mary.’



In Tarma Quechua, since the causative can embed a VoiceP, *-ku* can find a landing site below *-chi*. When this happens, *-ku* will be bound by the causee, because the latter is in the specifier of a Voice head, which is the notion of “subject” relevant to the binding of subject-oriented reflexives. Hence, we have the following structure for the example in (24), in which the causee is the most local available subject for *-ku* to take as an antecedent.

- (29) Maqa-ku-chi-n. (Tarma Quechua)  
 beat-REFL-CAUS-3SUBJ  
 'He causes someone<sub>i</sub> to beat himself<sub>i</sub>.'



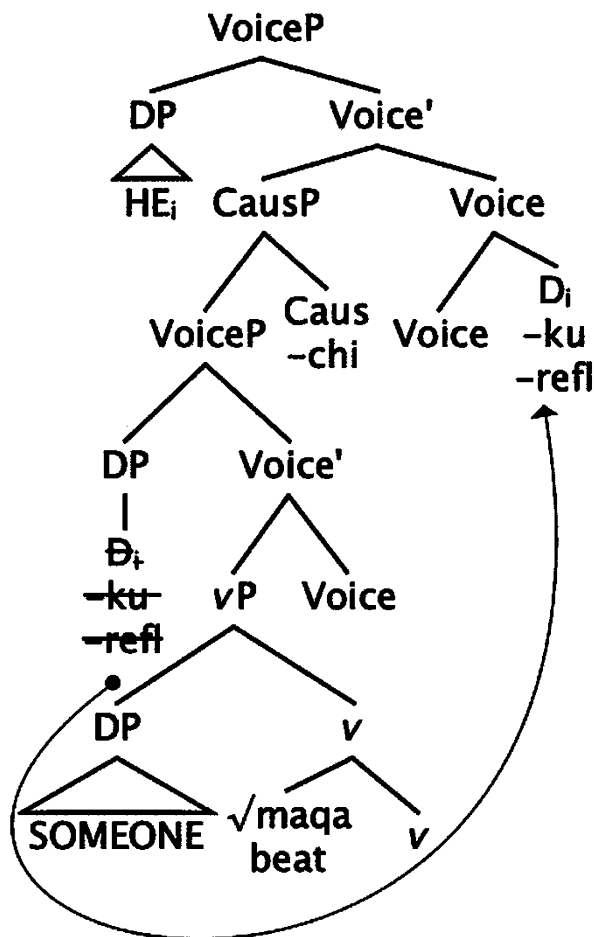
As we saw, Tarma Quechua also allows the order *-chi-ku*, in which case the same readings are available as in Cochabamba Quechua. I will assume that, as well as the structure in (24), *-chi* in Tarma Quechua may additionally embed a *vP*, giving rise to a structure identical to that presented for Cochabamba Quechua in (17).

One might expect it to be possible for *-ku* to be merged in the specifier of the embedded *VoiceP* in a structure like (29), before raising to adjoin to the matrix *Voice* head. The only possible binder for *-ku* in such a circumstance would be the causer (i.e., the subject in the specifier of the matrix *VoiceP*). If this were possible in Tarma Quechua, then the prediction is that example (25) should have the additional reading “He makes himself beat someone.” However, this reading is apparently unavailable for (25) in Tarma Quechua (Muysken 1979: 457–458; 462–464), although Muysken

cites Parker (1965) as indicating that such examples are grammatical on the relevant reading in Ayacucho Quechua, a Quechua IIC variety.

The problem we are faced with, then, is why the derivation in (30) is permitted in Ayacucho Quechua as described by Parker, but not in Tarma Quechua.

- (30) Maqa-chi-ku-n. (Hypothetical; apparently allowed in Ayacucho Q.)  
 beat-CAUS-REFL-3SUBJ  
 'He makes himself beat someone.'



Not having worked directly on the relevant Quechua languages, and seeing no straightforward way to rule out the derivation in (30), I will leave the question open here (though see Saab 2014, 2015, for a proposal regarding a prohibition in Spanish similar to the one that exists in Tarma Quechua).

There is an additional mystery regarding causative-applicative interactions, which I will have to leave open here. Given that applicative *-pu* is a high applicative (as argued in Myler 2016, 2018, and Sect. 4), and given that high applicatives merge above *vP* but below VoiceP (Pylkkänen 2008), all else held equal Cochabamba Quechua should permit the order *-chi-pu*, but not *\*-pu-chi*, whereas Tarma Quechua should permit both. This prediction is correct for Cochabamba Quechua, as we shall see in Sect. 4, but it is not correct for Tarma Quechua, which seems to have the same restriction as Cochabamba Quechua. In fact, no Quechua allows the order *\*-pu-chi*, as far as I know, although suffixal Appl-Caus orders are found in other language families (e.g., some Bantu languages allow it; see especially Satyo 1985 on isiXhosa). I have no explanation for this fact. Whatever the correct account is, however, it cannot involve analyzing all causatives in the Quechua family as verb-selecting, given the evidence I have presented in this section.

This concludes the analysis of the causative *-chi*. Parts of this account rely crucially on the assumption that reflexive *-ku* is an argument clitic. We turn to the evidence for that assumption in the next section.

### 3 *-ku* Is a Reflexive Argument Clitic

The possibility that *-ku* is a reflexive argument clitic immediately suggests itself given its variety of uses, which coincide with uses of such clitics in languages that uncontroversially have them (cf. Spanish *se* and its cognates in other Romance languages; see Saab 2020 for a recent analysis). For convenience, I repeat the examples showing this here as (31)–(34):

#### (31) Canonical Reflexive Uses

- a. Juan Marya-q                      chujcha-n-ta              peyna-n.  
     Juan Maria-GEN                  hair-3POSS-ACC      comb-3SUBJ  
     ‘Juan combs Maria’s hair.’
- b. Juan peyna-ku-n.  
     Juan comb-REFL-3SUBJ  
     ‘Juan combs himself.’

#### (32) Marking Certain Anticausatives

- a. Wayna qiru-ta                      p’aki-rqa.              (Cochabamba Quechua)  
     boy      glass-ACC                  break-PAST  
     ‘The boy broke the glass.’
- b. Qiru      p’aki-ku-rqa.  
     glass      break-REFL-PAST  
     ‘The glass broke.’

(33) **Autobenefactive/emotive and Anti-assistive Readings**

Tusuq tusu-ku-n.

dancer dance-REFL-3SUBJ

‘The dancer dances (and gets really into it/enjoys it).’

‘The dancer dances by himself (i.e., no one helps him).’

(34) **Idiosyncratic Uses**

Munay ‘to want’ munakuy ‘to love’

Yachay ‘to know’ yachakuy ‘to learn’

Of course, the range of uses in itself is merely *compatible* with an argument clitic analysis; it does not exclude the more traditional view that *-ku* is a verbal affix of some kind, spelling out some head in the extended projection of the verb (cf., e.g., the non-active Voice morphology of Modern Greek, as analyzed by Alexiadou et al. 2015, which overlaps heavily in function with *-ku*). What reason have we then for preferring one analysis over the other?

Myler (2017: 756–760) applies a range of diagnostics for argument clitichood to the Quechua object markers *-wa/-ma* (first person) and *-s(h)u* (second person), concluding that they are indeed object clitics rather than agreement affixes. Of these diagnostics, the most straightforward one to transpose to the case of *-ku* is that of clitic climbing in the context of a restructuring verb. Such behavior would be most unexpected of, for example, a Voice head, but is readily explained on an argument clitic account. It turns out that *-ku* can, in fact, climb from the verb it “belongs” to onto a matrix restructuring verb—compare (35) and (36). This is just as Myler (2017) showed for the object markers, as illustrated here for first-person *-wa* in examples (37) and (38).

(35) Maylla-ku-y-ta ati-n. (Cochabamba Quechua)

wash-REFL-INF-ACC can-3SUBJ

‘S/he can wash himself/herself.’

(36) Maylla-y-ta ati-ku-n. (Cochabamba Quechua)

wash-INF-ACC can-REFL-3SUBJ

‘S/he can wash himself/herself.’

(37) Pusa-wa-y-ta ati-n. (Cochabamba Quechua)

take-1OBJ-INF-ACC can-3SUBJ

‘S/he can take me.’

(38) Pusa-y-ta ati-wa-n. (Cochabamba Quechua)

take-INF-ACC can-1OBJ-3SUBJ

‘S/he can take me.’

Myler (2017: 760–765), examining the ordering of the object markers *-wa/-ma* and *-s(h)u* with respect to other verb morphology, concludes that these two markers come to occupy different cliticization sites in Quechua clause structure. While first-person *-wa/-ma* may raise to one of two positions (one below and one above Aspect), *-s(h)u* always raises to a position above TP. Given that *-ku* seems to be an argument clitic too, the natural question to ask is where in the clause structure *-ku* moves to. As anticipated in Sect. 2, it is the proposal of this chapter that *-ku* raises to the edge of the thematic domain in Cochabamba Quechua, and no higher. The evidence for this again comes from position relative to other verbal morphology. As the following example from van de Kerke (1996: 164, his (69)) illustrates, *-ku* precedes the other object markers if they co-occur.

- (39) Kaserita khuchi aycha-ta ranti-ku-wa-y. (Cochabamba Quechua)  
 client pig meat-ACC buy-REFL-1OBJ-INF  
 ‘Dear client! Buy yourself a piece of pork from me!’

Given that the lowest position *-wa* can occupy is below AspP, this forces the conclusion that *-ku* is still lower. Now recall from Sect. 2 that *-ku* must follow *-chi* on the surface. The facts follow if *-ku*'s landing site is at the borderline of the thematic domain with the IP domain (i.e., adjoined to the Voice head), exactly as proposed in the structure in (17).

Having dealt with where *-ku* ends up, let us ask where in the structure it starts off. The answer to this question varies depending on the construction, but the core that unites the cases is that *-ku* always merges in *some* argument position. Space constraints preclude my spelling out the details in full here, but the following is a brief overview of the possibilities. “True” reflexive readings presumably involve *-ku* starting off in the relevant internal argument position, as instantiated in (17). Where *-ku* marks anticausatives, I will follow Schäfer (2008) and Wood (2015) in proposing that these involve the reflexive merging in spec-VoiceP (with both it and Voice itself being interpreted as expletive for reasons discussed by Schäfer).<sup>9</sup> I will discuss the autobenefactive and anti-assistive readings illustrated by example (33) in Sect. 5. Even idiosyncratic uses such as (34) are compatible with the idea that *-ku* occupies some argument position; see Wood 2015: 84–88 and Chapter 6 on Icelandic *-st* for a range of possibilities.<sup>10</sup>

<sup>9</sup> In this circumstance, adjunction to the Voice head could take place under Morphological Merger (Marantz 1984; Matushansky 2006 et seqq.).

<sup>10</sup> For the specific idiosyncratic cases in (34), one could investigate the possibility that *munakuy* “to love” is merely the autobenefactive/emotive variant of *munay* “to want,” so that *munakuyki* “I love you” in Cochabamba Quechua is more literally “I want you, and I’m {emotionally invested in/enjoying} it to a high degree.” (See Sect. 5 for how this use of *-ku* is analyzed on the present approach). As for *yachakuy* “to learn” from *yachay* “to know,” the most natural account would be that *yachakuy* involves merging the root in question with an eventive rather than a stative little-*v* head; *yachakuy* would then be the anticausative of that eventive verb, derived as discussed in the

I will close this section with some brief cross-dialectal comparison. While I have not studied the matter in depth, my impression is that the uses to which *-ku* is put in Cochabamba Quechua are found throughout the family, so that the analyses sketched in the preceding paragraph ought to generalize to them straightforwardly. There are a couple of other uses attested in at least Cajamarca Quechua (a Quechua IIA variety spoken in rural communities surrounding the northern Peruvian city of Cajamarca), which deserve some comment. These are discussed by H. Coombs (n.d.) and Coombs-Lynch et al. (2003). One involves suppression of a direct object; this is illustrated in (40) and (41), which are from Coombs-Lynch et al. (2003: 87). I have added the morpheme segmentation and the gloss; the translations are my English rendering of the original Spanish.

(40) Akshu-ta-m            yanu-yka-ni.            (Cajamarca Quechua)  
 potato-ACC-EVID    cook-DUR-1SUBJ  
 ‘I am cooking potatoes.’

(41) Yanu-ku-yka-ni-mi.            (Cajamarca Quechua)  
 cook-REFL-DUR-1SUBJ-EVID  
 ‘I am cooking.’

The fact that this use involves argument suppression fits nicely into the general approach to the syntax of *-ku* taken here. However, the issue of how the interpretation arises from this syntax remains a mystery to me, as does the question of why other Quechua varieties do not allow *-ku* to suppress a direct object in this manner.

The second use for *-ku* in Cajamarca Quechua, which is not found in Cochabamba Quechua, is to form verbs from nouns. The resulting verbs mean “interact with **noun** in some canonical way.” These examples are from H. Coombs (n.d.: 12–13); the translations are mine from the original Spanish.

(42) **Denominal Verbs in *-ku* in Cajamarca Quechua**

|        |                        |           |                     |
|--------|------------------------|-----------|---------------------|
| kena   | ‘quena <sup>11</sup> ’ | kenakuy   | ‘play the quena’    |
| wawa   | ‘child’                | wawakuy   | ‘give birth’        |
| kanwa  | ‘canoe’                | kanwakuy  | ‘make a canoe’      |
| sigara | ‘cigarette’            | sigarakuy | ‘smoke a cigarette’ |

---

text for anticausatives in *-ku* more generally. In support of this analysis, note that *yachakuy* has a lexical causative doublet in the form of *yachachiy* ‘to teach.’

<sup>11</sup> The quena is a small wind instrument, traditional in the Andes.

Interestingly, the Icelandic morpheme *-st* has a similar use, illustrated in (43). Wood (2015: 252–259; the example is his (5b)) argues for an analysis of such *-st* verbs in which *-st* is an expletive direct object.

- (43) Ég er núna í Boston og er að **trompetast**. (Icelandic)  
 I am now in Boston and am to trumpet-ST  
 ‘Now I’m in Boston trumpeting.’

It seems clear to me that an extension of this analysis to the Cajamarca Quechua cases is desirable (it would effectively reduce this use of *-ku* to the object suppression subcase from the previous paragraph). What is unclear is, once again, why this use of *-ku* is not found everywhere in the family.

With that, two out of the three key players in this paper have been dealt with. I now turn to the third and final one: applicative *-pu*.

#### 4 *-pu* Is a High Applicative and a Raising Applicative

This section is a simplified précis of discussions published by Myler (2016: 202–224)<sup>12</sup> and Myler (2018); I refer the reader to those works for more empirical detail on the functions of *-pu* in Cochabamba Quechua and Santiago del Estero Quechua and for more explicit technical detail on the proposals, which I can only sketch here.

The suffix *-pu* is usually described as a benefactive suffix in the descriptive literature on Quechua, and this is indeed one of its most salient functions in the Cochabamba variety. This use of *-pu* is available with predicates where no transfer of possession is involved, and also with intransitive verbs, indicating that it is a High Applicative in the taxonomy of Pykkänen (2008):

- (44) Nuqa lapicero-ta jap’i-chka-ni. (Cochabamba Quechua)  
 I pen-ACC hold-DUR-1SUBJ  
 ‘I’m holding a pen.’
- (45) Nuqa (qam-paq) lapicero-ta jap’i-pu-chka-yki. (Cochabamba Q.)  
 I you-BEN pen-ACC hold-APPL-DUR-1SUBJ>2OBJ  
 ‘I’m holding a pen for you.’

<sup>12</sup> Beyond the simplifications I have made, there is one substantive difference between the present discussion and these earlier works: Myler (2016: 217) assumes without argument that Cochabamba Quechua causatives are Voice-selecting, a position I have presented evidence against in this paper, and also cites an example from van de Kerke (1996) which turns out, on investigation, not to be allowed by the speakers whose grammar I am characterizing here (since it involves cliticizing the causee in a causative construction).



This benefactive function of *-pu* is the most consistently available one across the dialects; that is, if a variety has *-pu* in its inventory at all, “benefactive” will, without fail, be among the functions that it has.<sup>13</sup> Alongside the benefactive, individual dialects often exhibit other functions for *-pu*, which vary considerably across the family. In Cochabamba Quechua, for example, *-pu* can be used with some verbs with a restitutive meaning, reminiscent of the English particle *back*:

- (46) Misk'i-n-ta            jaywa-y!    (Cochabamba Quechua)  
 sweet-3POSS-ACC    hand-INF  
 'Hand him his sweet!'
- (47) Misk'i-n-ta            jaywa-pu-y!    (Cochabamba Quechua)  
 sweet-3POSS-ACC    hand-APPL-INF  
 'Hand him his sweet back!'
- (48) Misk'i-n-ta            qu-y!    (Cochabamba Quechua)  
 sweet-3POSS-ACC    give-INF  
 'Give him his sweet!'
- (49) Misk'i-n-ta            qu-pu-y!    (Cochabamba Quechua)  
 sweet-3POSS-ACC    give-APPL-INF  
 'Give him his sweet back!'

With certain verbs of motion, *-pu* can contribute a sort of deictic goal argument corresponding rather closely to the English particle *away*:

- (50) Nuqa    ri-ni.            (Cochabamba Quechua)  
 I            go-1SUBJ  
 'I go.'
- (51) Nuqa    ri-pu-ni.            (Cochabamba Quechua)  
 I            go-APPL-1SUBJ  
 'I go away.'  
 'I go for him/her.'

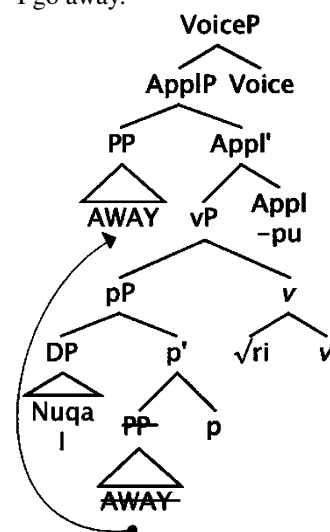
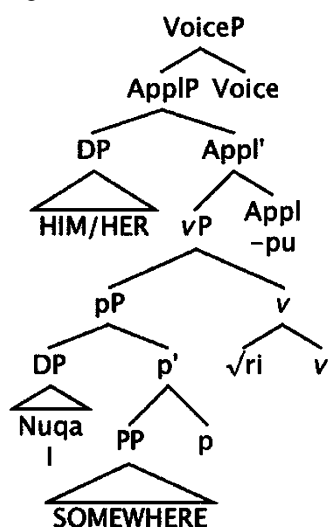
<sup>13</sup> I have no explanation for this fact, or for the fact that benefactive appears to be the most commonly attested interpretation associated with applicative morphemes cross-linguistically (Polinsky 2013).

As reflected in the translations for (51), *-pu* in such contexts is potentially ambiguous. Because there is no overt third-person object marker, (51) is also compatible with a benefactive reading. There is strong evidence that this is a structural ambiguity, rather than a lexical ambiguity involving a coincidentally homophonous *-pu* suffix meaning “away.” If an overt beneficiary in the form of a strong pronoun is added to the sentence, the “away” reading disappears.

- (52) Nuqa pay-paq ri-pu-ni. (Cochabamba Quechua)  
 I s/he-BEN go-APPL-1SUBJ  
 ‘I go for him/her’  
**NOT:** \*‘I go away for him/her.’

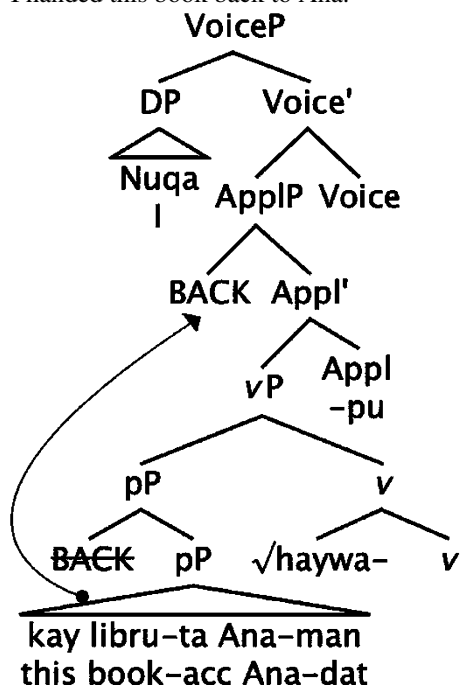
This is accounted for if there is just one *-pu*, which is capable of hosting exactly one element in its specifier. Example (52) is then unambiguous because the overt beneficiary argument is competing for the same specifier position with a silent element corresponding to English *away*, which contributes the “away” reading in (51). This, along with the lexical restrictions on the availability of the “away” reading and the restitutive reading exemplified in (46)–(49), can itself be explained if *-pu* is not responsible for introducing a thematic role, but is rather a licensing position for certain oblique arguments—a Raising Appl in the terminology of Paul and Whitman (2010) Georgala (2012), and Nie (2019, 2020). To illustrate, the two readings of (51) would have the following distinct structures:

- (53) Nuqa ri-pu-ni. (Cochabamba Quechua)  
 I go-APPL-1SUBJ  
 ‘I go for him/her.’



The idea is that the silent counterpart of *away* in the tree on the right in (53) is actually an argument of the verb semantically, and the role of the applicative is simply to license this silent oblique element.<sup>14</sup> Likewise, the lexically restricted restitutive reading can be accounted for in terms of a silent Quechua counterpart of the English particle *back*, which itself will only be capable of modifying certain kinds of PP complement, which in turn only certain verbs c-select. This analysis is illustrated in the structure below (the example is from van de Kerke 1996: 192, his (42)).

- (54) Nuqa kay libru-ta Ana-man haywa-pu-rqa-ni. (Cochabamba Q)  
 I this book-ACC Ana-to hand-APPL-PAST-1SUBJ  
 'I handed this book back to Ana.'



Now, as mentioned earlier, the non-benefactive readings of *-pu* vary in their availability across the family. Let us turn to Santiago del Estero Quechua, an Argentine variety from the same subfamily as Cochabamba Quechua (Quechua IIC), and whose common ancestor with the latter probably existed around 500 years

<sup>14</sup> Here I follow my earlier work in assuming that this licensing relationship involves movement into spec-AppIP, but *in situ* licensing (perhaps via Agree) would be a viable alternative.

ago.<sup>15</sup> In this variety, we find that the “away” reading and the restitutive reading are absent (see Albarracín and Alderetes 2013; Albarracín and Gómez Nazar 2021 for detailed discussions of the syntax and semantics of *-pu* in Santiago del Estero Quechua); this is illustrated for the “away” reading in (55).

- (55) Noqa ri-po-ra-ni. (Santiago del Estero)  
 I go-APPL-PAST-1SUBJ  
 ‘I went for him/her.’  
**NOT:** \*‘I went away.’

Such variation in the available readings can be captured very simply on the present approach, as a case of variation in what kinds of element may occupy spec-AppIP. Some parameter of variation regarding the occupants of spec-AppIP must clearly be countenanced in any case, because the two dialects differ visibly in the case-markings they allow on applied arguments. While both varieties allow the benefactive case suffix *-paq*, only Santiago del Estero Quechua permits accusative arguments in this position as well (the alternation has semantic consequences in some cases, as these examples illustrate).

- (56) Noqa pay-paq challwa-s-ta wachi-pu-ni. (Santiago del Estero Q.)  
 I s/he-BEN fish-PL-ACC scale-APPL-1SUBJ  
 ‘I scale fish so that s/he doesn’t have to.’  
 (adapted from Albarracín and Alderetes 2013: 7)
- (57) Noqa pay-ta challwa-s-ta wachi-pu-ni. (Santiago del Estero Q.)  
 I s/he-ACC fish-PL-ACC scale-APPL-1SUBJ  
 ‘I scale fishes {on him/for him to have}.’  
 (Lelia Albarracín (pers. comm.); gloss mine)

Cochabamba Quechua forbids accusative case on applied arguments in this and all other circumstances:

- (58) Nuqa Juan{-paq/\*-ta} challwa-s-ta escama-pu-ni. (Cochabamba Q.)  
 I Juan{-for/-ACC} fish-PL-ACC scale-APPL-1SUBJ  
 ‘I scale fishes for Juan.’

<sup>15</sup> On Santiago del Estero Quechua, generally, see Albarracín (2011); Alderetes (2001); Bravo (1956); Nardi (2002); and Prezioso and Torres (2006).

Conversely, Cochabamba Quechua allows genitive applied arguments in the context of a certain predicative possession construction (what Myler 2016 calls the BE-APPL construction); the corresponding<sup>16</sup> construction in Santiago del Estero Quechua exhibits accusative case-marking, and there are no contexts in the latter variety in which genitive-marked phrases appear as applied arguments.

- (59) Gladys-pata kallpa tiya-pu-n. (Cochabamba Quechua)  
 Gladys-GEN strength be-APPL-3SUBJ  
 ‘Gladys has strength.’
- (60) \*Gladys-ta kallpa tiya-pu-n. (Cochabamba Quechua)  
 Gladys-ACC strength be-APPL-3SUBJ  
 ‘Gladys has strength.’
- (61) \*Gladys-pa kallpa tiya-pu-n. (Santiago del Estero Quechua)  
 Gladys-GEN strength be-APPL-3SUBJ  
 ‘Gladys has strength.’
- (62) Gladys-ta kallpa tiya-pu-n. (Santiago del Estero Quechua)  
 Gladys-ACC strength be-APPL-3SUBJ  
 ‘Gladys has strength.’

Let us summarize what we have established so far in this section. Cochabamba Quechua (like many other Quechua varieties) has a suffix *-pu*. This suffix has the syntactic properties of a High Applicative in the sense of Pykkänen (2008), but does not appear to be responsible for the assignment of any one thematic role. Rather, it licenses certain arguments that gain their thematic role elsewhere: it is a raising applicative of the kind argued for by Paul and Whitman (2010) Georgala (2012), and Nie (2019, 2020). This conclusion has important implications for how we approach its interaction with other affixes, especially (and most relevantly to this chapter) the causative suffix *-chi*. I turn to this interaction now.

A key observation is that *-pu* must always follow *-chi*; the reverse order is ungrammatical.

<sup>16</sup> Although these constructions do correspond closely on the surface and in certain aspects of their syntax, they differ importantly in others and are radically different in terms of the possession relations they can express. See Myler (2018) for details.

- (63) Juan-wan mama-y-paq wasi-ta picha-chi-pu-rqa-ni. (Coch. Q.)  
 Juan-with mother-IPOSS-BEN house-ACC sweep-CAUS-APPL-PAST-1SUBJ  
 ‘I make Juan sweep the house for my mother.’
- (64) \*Juan-wan mama-y-paq wasi-ta picha-pu-chi-rqa-ni. (Coch. Q.)  
 Juan-with mother-IPOSS-BEN house-ACC sweep-APPL-CAUS-PAST-1SUBJ  
 ‘I make Juan sweep the house for my mother.’

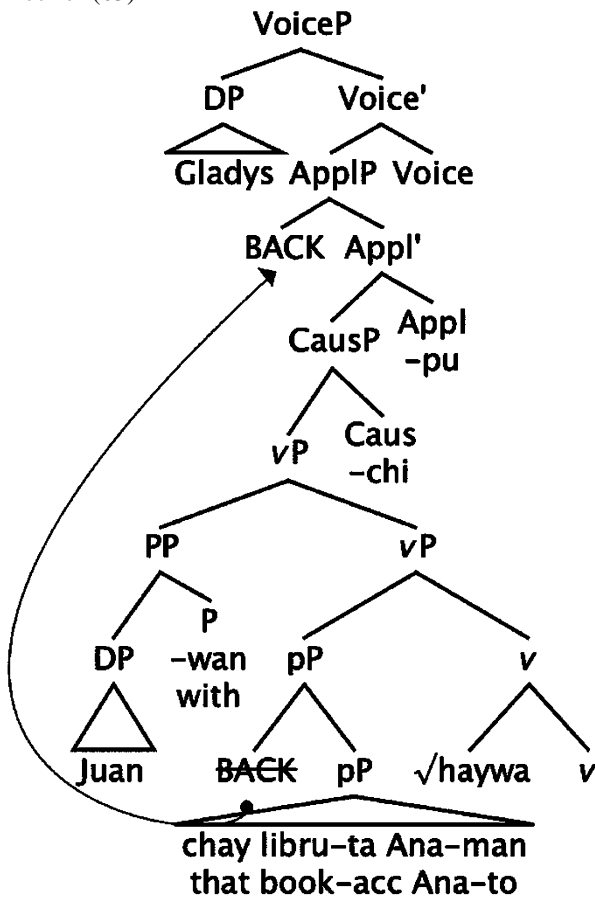
This is true even in cases, such as the restitutive reading, where the meaning associated with the applicative clearly “belongs to” the embedded event, rather than the causing event.

- (65) Gladys Juan-wan chay libru-ta Ana-man haywa-chi-pu-rqa. (Coch. Q.)  
 Gladys Juan-with that book-ACC Ana-to hand-CAUS-APPL-PAST  
 ‘Gladys made Juan hand that book back to Ana.’
- (66) \*Gladys Juan-wan chay libru-ta Ana-man haywa-pu-chi-rqa. (Coch. Q.)  
 Gladys Juan-with that book-ACC Ana-to hand-APPL-CAUS-PAST  
 ‘Gladys made Juan hand that book back to Ana.’

This is all explained by the analysis we assigned to *-chi* in Sect. 2, in conjunction with the analysis of *-pu* we have developed in this section. The order *-pu-chi-* is disallowed because *-chi* selects *vP* and *ApplP* does not satisfy this requirement (at least, not in Quechua).<sup>17</sup> The order *-chi-pu* is allowed and can convey meanings in which the applied argument belongs to the embedded event rather than the causing event because *-pu* is a Raising Applicative: the argument it licenses can raise from somewhere lower than *-chi*. This is illustrated in the following tree, which represents the structure of example (65).

<sup>17</sup> Recall from the end of Sect. 2 that the fact that the order *\*-pu-chi-* is disallowed in Tarma Quechua is still a mystery, given that that variety allows *-chi* to select VoiceP.

(67) Tree for (65)



All three of our players are now in the game, and we have covered the way in which *-chi* interacts with *-ku* and the way in which *-chi* interacts with *-pu*. All that remains is to discuss the final pairing: reflexive *-ku* and applicative *-pu*.

### 5 *\*-pu-ku/\*-ku-pu*: One Mystery’s Solution Leads to Another’s

The most striking fact about *-ku* and *-pu*’s interaction is that it never happens on the surface. That is, the two are in complementary distribution, regardless how one tries to order them. However, it seems that the two halves of this interdiction do not have the same status. The *-ku-pu-* ordering is ungrammatical, and there is no

morphological way to fix it; instead, the reading that would result from that ordering must resort to a postpositional phrase, as follows:

- (68) \*Maylla-ku-pu-ni. (Cochabamba Quechua)  
 wash-REFL-APPL-1SUBJ  
 ‘I wash myself for her.’
- (69) Pay-rayku maylla-ku-ni. (Cochabamba Quechua)  
 s/he-for.the.sake.of wash-REFL-1SUBJ  
 ‘I wash myself for her.’

It seems that Cochabamba Quechua used to permit *-ku* to precede *-pu* on the surface, with morphophonological lowering of the vowel of *-ku* to *-ka* (for all I know, this could be a point of synchronic variation to this day, but I will save space by continuing to discuss it as if this reflects two different diachronic stages of the language).<sup>18</sup> Bills et al. (1969: 195–196) provide a number of examples of the following sort (the gloss is mine, and I have altered the spelling to conform to modern orthographical conventions for Cochabamba Quechua, but the example is otherwise unchanged, and the translation is the one provided by Bills et al. 1969: 196):

- (70) Maylla-ka-pu-ni. (Doculect of Bills et al. 1963)  
 wash-REFL-APPL-1SUBJ  
 ‘I bathe myself for her.’

However, Gladys Camacho Rios (p.c.) tells me that these forms are no longer grammatical in contemporary Cochabamba Quechua with the relevant applicative-of-a-reflexive interpretation.<sup>19</sup>

This is exactly what we should expect given the preceding sections: *-pu*, as a high applicative, is a head in the thematic domain. *-ku*, being an argument clitic with the properties we have argued for, moves to adjoin to the Voice head in every structure it appears in. Therefore, based on the syntactic properties that these morphemes can be shown to have, it is expected that *-pu-ku* should be the only order available.

<sup>18</sup> On such morphophonological vowel lowering, which is found in many Quechua varieties, see Weber (1976: 79, 1989: 462–464) on Huallaga Quechua, Muysken (1981: 303–304) on Tarma Quechua, and Muysken (1988: 268, 274) on Chumbivilcas Quechua, among many others. This process affects and is triggered by a number of different morphemes in the Quechua verb in a way that resists generalizations based on natural classes and which sometimes takes place non-locally (such that trigger and target are separated by other, non-participating affixes).

<sup>19</sup> Synchronically, there is a single suffix *-kapu* in the language with an entirely different kind of semantics; see Camacho Rios (in progress).



The surprise is that not even this order is grammatical. Any attempt to reflexivize an applicative turns out to be ungrammatical on the expected surface manifestation:

- (71) \*Nuqa papa-ta wayk'u-pu-ku-ni. (Cochabamba Quechua)  
 I potato-ACC cook-APPL-REFL-1SUBJ  
 'I cook potato for myself.'

However, unlike example (68), in this case, there is a morphological means of expressing what (71) is trying to mean. It involves leaving out *-pu* altogether.

- (72) Nuqa papa-ta wayk'u-ku-ni. (Cochabamba Quechua)  
 I potato-ACC cook-REFL-1SUBJ  
 'I cook potato for myself.'

Quite generally, the meanings that one would expect to involve reflexivizing an applicative are ungrammatical if expressed using verbs in *-pu-ku* but are grammatical if expressed with *-ku* alone. This all suggests that the syntactic configurations that one would expect from a combination of *-pu* followed by *-ku* are in fact possible; their surface absence is to be explained by a morphophonological rule, of the following kind:

- (73)  $-pu \rightarrow \emptyset / \_ \_ -ku$

This suffices to explain a couple of uses of *-ku* that we have not addressed so far, specifically the autobenefactive/emotive and the anti-assistive uses. The autobenefactive/emotive use is fairly transparently a reflexivized benefactive applicative:

- (74) Tusuq tusu-~~pu~~-ku-n. (Cochabamba Quechua)  
 dancer dance-APPL-REFL-3SUBJ  
 'The dancer dances and gets really into it/enjoys it.'  
 'The dancer dances without help/by himself.'

As for the anti-assistive use, while I have no worked-out account of why an applicative morpheme combined with a reflexive morpheme should give rise to such a reading, isiXhosa (Nguni, Bantu; spoken in South Africa and Zimbabwe) makes it completely clear that such a syntax-semantics mapping is permitted by Universal Grammar:<sup>20</sup>

<sup>20</sup> Numerals in the glosses of isiXhosa examples indicate noun class morphology.

- (75) Usana lu-zi-hamb-el-e. (isiXhosa)  
 11.baby 11SUBJ-REFL-walk-APPL-PERF  
 ‘The baby walked by itself/without help.’  
 (Myler and Mali 2021: 15, their (31))

The variety described in Bills et al. (1969) does not appear to differ with contemporary varieties when it comes to *\*-pu-ku* sequences. No such sequences are mentioned as being possible in the relevant sections, and Bills et al. (1969: 73) indicate that, just as in the doculect of this paper, *-ku* could be used on its own with the meaning of a benefactive reflexive. They provide the following example (1969: 73; once again the gloss is mine but the translation is the original):

- (76) Ranti-ri-ku-y. (Cochabamba Quechua)  
 buy-polite-refl-inf  
 ‘Buy it for yourself!’

Presumably, then, Bills et al.’s Cochabamba Quechua differs from the contemporary variety described here in allowing *-ku* to left adjoin to Appl (rather than only having Voice as a potential cliticization site, as in the contemporary variety), but the two varieties coincide in having the *-pu* deletion rule I have proposed.

## 6 Conclusion: Investigating Affix Order and Interpretation in Quechua and Beyond

A key contribution of this chapter is a principled explanation of the core affix ordering facts involving the three morphemes that this paper has focused on, which are repeated as (77)–(79):

- (77) a. <sup>OK</sup>chi-ku      g. *\*-chi-ku-pu*  
 b. *\*ku-chi*      h. *\*ku-chi-pu*  
 c. <sup>OK</sup>chi-pu      i. *\*-ku-pu-chi*  
 d. *\*pu-chi*      j. *\*-chi-pu-ku*  
 e. *\*ku-pu*      k. *\*-pu-chi-ku*  
 f. *\*pu-ku*      l. *\*-pu-ku-chi*

The reason that *-chi* must precede the other two morphemes is that it is a verb-selecting causative, which means that it selects a structure too small to accommodate a High Applicative (which is what *-pu* is) and too small to contain a landing site for

- (78) *-chi* may co-occur with either of the other two suffixes, but must always come first.
- (79) *-ku* and *-pu* can never co-occur.

argument clitics like *-ku*, which we have established must move to adjoin to a Voice head. The order *-ku-pu* is always out in the contemporary variety because there is no cliticization site for *-ku* below the High Applicative *-pu*. The order *-pu-ku*, on the other hand, is derived, given the independently verifiable properties of the morphemes involved, and the meanings they can be used to construct. The fact that it does not surface is purely because of the *-pu*-deletion rule we have motivated.

At this point, we are able to see that all of these affix-ordering facts follow from the syntactic and semantic properties of the morphemes we have discussed, given the analyses we have motivated for them. The only exception is that meanings that we would expect to be conveyed by *-pu-ku* are instead conveyed by *-ku* on its own, a fact we have attributed to a morphophonological deletion rule targeting *-pu*. The general methodological lesson that suggests itself to me is this: one cannot make judgments about whether a given affix order is expected or unexpected without an extremely careful investigation of the syntactic and semantic properties of the affixes in question. Thus, many documented exceptions to the Mirror Principle may prove to be illusory on closer inspection.

**Acknowledgments** Most of the Quechua data reported here were gathered during fieldwork in Cochabamba, Bolivia, and Santiago del Estero, Argentina, which was supported by NSF Doctoral Dissertation Research Improvement Grant BCS-1324839. My profound thanks go to María Cardozo, Gladys Camacho Rios, and Cristina Puente Arista (in Cochabamba) and Dario Acosta, Rosinda Barreta, Casilda Chazarreta, Estela Chazarreta, Guillermo Chazarreta, Manuela Jimenez, Raquel Gómez, Teodosia González, Agustino Grano, Haydee Palavecino, and Aurella Quita (in Santiago) for sharing their languages with me. I'd also like to thank Gillian Gallagher for making my work in Bolivia possible; Lelia Albarracín for helping to set up my Argentina fieldwork; Héctor Andreani for his friendship and hospitality; and Gladys Camacho Rios (again) for her extraordinary forbearance and kindness in providing me with grammaticality judgments and other feedback over Facebook and email. Many thanks also to Cilene Rodrigues and Andrés Saab for their comments on this paper and for their work on this volume as a whole. The usual disclaimers apply.

## References

- Ackema, Peter, and Ad Neeleman. 2004. *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: Oxford University Press.
- Albarracín, Lelia Inés. 2011. *La Quichua: gramática, ejercicios, y diccionario quichua-castellano*. Buenos Aires: Editorial Dunken.
- Albarracín, Lelia Inés, and Jorge Alderetes. 2013. El sufijo *-pu* en el quichua de Santiago del Estero. Talk given at the *III Encuentro de Lenguas Indígenas Americanas*, San Carlos de Bariloche.

- Albarracín, Lelia Inés, and Regina Fernanda Gómez Nazar. 2021. Variación actancial en quichua santiagueño: los aplicativos (parte 1). *Lengua y Sociedad* 20 (1): 13–32.
- Albó, Xavier. 1970. *Social constraints on Cochabamba Quechua*. PhD thesis, Cornell University.
- Alderetes, Jorge R. 2001. *El quichua de Santiago del Estero. Gramática y vocabulario*. Tucumán: Universidad Nacional de Tucumán, Facultad de Filosofía y Letras.
- Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer. 2015. *External arguments in transitivity alternations: A layering approach*. Oxford: Oxford University Press.
- Baker, Mark. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16 (3): 373–415.
- Bills, Garland D., Vallejo C. Bernardo, and Rudolph C. Troike. 1969. *An introduction to spoken Bolivian Quechua*. Austin: University of Texas Press.
- Bjorkman, Bronwyn, and Elizabeth Cowper. 2013. Inflectional shells and the syntax of causative *have*. In *Proceedings of the 2013 annual conference of the Canadian Linguistic Association*, ed. Silvia Perpiñán, and Itziri Moreno-Villamar
- Bravo, Domingo A. 1956. *El quichua santiagueño. Reducto idiomático argentino*. Tucumán: Universidad Nacional de Tucumán.
- Bruening, Benjamin. 2013. *By*-phrases in passives and nominals. *Syntax* 16 (1): 1–41.
- Camacho Rios, Gladys. in progress. Untitled. PhD thesis, University of Texas at Austin.
- Coombs, Heidi. n.d. *El sufijo -ku en el quechua de Cajamarca*. ms, The Summer Institute of Linguistics.
- Coombs Lynch, David, Heidi Carlson de Coombs, and Blanca Ortiz Chamán. 2003. *Rimashun Kichwapi. Una introducción al quechua cajamarquino*. Lima: Atares artes y letras.
- Folli, Raffaella, and Heidi Harley. 2007. Causation, obligation, and argument structure: On the nature of little *v*. *Linguistics Inquiry* 38 (2): 197–238.
- Georgala, Effi. 2012. *Applicatives in their structural and thematic function: A minimalist account of multitransitivity*. PhD thesis, Cornell University.
- Georgala, Effi, Waltraud Paul, and John Whitman. 2008. Expletive and thematic applicatives. In *Proceedings of WCCFL 26*, ed. Charles B. Chang and Hannah J. Haynie, 181–189. Somerville: Cascadilla Proceedings Project.
- Grimshaw, Jane. 1986. A morphosyntactic explanation for the Mirror Principle. *Linguistic Inquiry* 17 (4): 745–749.
- Harley, Harley. 2017. The “bundling” hypothesis and the disparate functions of little *v*. In *The verbal domain*, ed. Roberta D’Alessandro, Irene Franco, and Angel Gallego, 3–28. Oxford: Oxford University Press.
- Irwin, Patricia. 2012. *Unaccusativity at the interfaces*. PhD thesis, New York University.
- Jerro, Kyle. 2016. *The syntax and semantics of applicative morphology in Bantu*. PhD thesis, University of Texas at Austin.
- . 2019. The causative-instrumental syncretism. *Journal of Linguistics* 53 (4): 751–788.
- Jung, Hyun Kyoung. 2014. *On the syntax of applicative and causative constructions*. PhD thesis, University of Arizona, Tucson.
- Kastner, Itamar. 2020. *Voice at the Interfaces: The syntax, semantics, and morphology of the Hebrew verb*. Berlin: Language Science Press.
- Key, Gregory. 2013. *The morphosyntax of the Turkish causative construction*. PhD thesis, University of Arizona, Tucson.
- Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. Johan Rooryck and Laurie Zaring, 109–137. Dordrecht: Kluwer.
- Lastra, Yolanda. 1968. *Cochabamba Quechua Syntax*. The Hague: Mouton.
- Legate, Julie Anne. 2014. *Voice and v: Lessons from Acehnese*. Cambridge, MA: MIT Press.
- Marantz, Alec. 1984. *On the nature of grammatical relations*. Cambridge, MA: MIT Press.
- . 2013. Verbal argument structure: Events and participants. *Lingua* 130: 152–168.
- Matushanky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37 (1): 69–109.
- Muysken, Pieter. 1979. Quechua causatives and logical form: A case study in markedness. In *Theory of markedness in generative grammar*, ed. Adriana Belletti, Luciana Brandi, and Luigi Rizzi, 445–473. Pisa: Scuola Normale Superiore.

- . 1981. Quechua word structure. In *Binding and filtering*, ed. Frank Heny, 279–329. London: Longmans.
- . 1986. Approaches to affix order. *Linguistics* 24: 629–643.
- . 1988. Affix order and interpretation: Quechua. In *Morphology and modularity*, ed. Martin Everaert and Mieke Trommelen, 259–279. Dordrecht: Foris.
- . 1989. The morphology/syntax interface in Quechua dialectology. In *Dialect variation and the theory of grammar*, ed. Paola Benincà, 41–51. Dordrecht: Foris.
- Myler, Neil. 2016. *Building and interpreting possession sentences*. Cambridge, MA: MIT Press.
- . 2017. Cliticization feeds agreement: A view from Quechua. *Natural Language and Linguistic Theory* 35: 751–800. <https://doi.org/10.1007/s11049-016-9351-y>.
- . 2018. Variation in the syntax and semantics of predicative possession in Quechua. *Proceedings of the workshop on the structure and constituency of the languages of the Americas*, 21.
- Myler, Neil, and Zoliswa Mali. 2021. Two places for causees in productive IsiXhosa morphological causatives. *Syntax* 24 (1): 1–43.
- Nardi, Ricardo L.J. 2002. *Introducción al quichua santiagueño*. San Miguel de Tucumán: Editorial Dunken.
- Nie, Yining. 2019. Raising applicatives and possessors in Tagalog. *Glossa* 4 (1): 139.
- . 2020. *Licensing arguments*. PhD thesis, New York University.
- Oseki, Yohei. 2017. *Voice morphology in Japanese argument structures*. Ms., New York University. <https://ling.auf.net/lingbuzz/003374>.
- Parker, Gary. 1965. *Gramática del quechua ayacuchano*. Lima: UNSM.
- Paul, Waltraud, and John Whitman. 2010. Applicative structure and Mandarin ditransitives. In *Argument structure and syntactic relations from a cross-linguistic perspective*, ed. Maia Duguine, Susana Huidobro, and Nerea Madriaga, 261–282. Amsterdam: John Benjamins.
- Pike, Kenneth. 1945. A problem in morphology syntax division. *Acta Linguistica* 5: 125–136.
- Pitteroff, Marcel, and Cinzia Campanini. 2013. Variation in analytic causative constructions: A view on German and Romance. *Journal of Comparative Germanic Linguistics* 16: 209–230.
- Polinsky, Maria. 2013. Applicative constructions. In *The World atlas of language structures online*, ed. Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/109>, Accessed 23 Mar 2016.
- Prezioso, Mabel, and Rubén Torres. 2006. *Quechua Tukuypaj*. Nuestra América.
- Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, MA: MIT Press.
- Ryan, Kevin. 2010. Variable affix order: Grammar and learning. *Language* 86: 758–791.
- Saab, Andrés. 2014. Syntax or nothing: Some theoretical and empirical remarks on implicit arguments. *Borealis: An International Journal of Hispanic Linguistics* 3 (2): 125–183.
- . 2015. On long-distance theta-role assignment. *Lingua* 160: 91–126.
- . 2020. Deconstructing voice: The syntax and semantics of *u*-syncretism in Spanish. *Glossa* 5 (1): 1–50. <https://doi.org/10.5334/gjgl.704>.
- Satyo, Sizwe C. 1985. *Topics in Xhosa verbal extension*. PhD thesis, University of South Africa, Pretoria.
- Schäfer, Florian. 2008. *The syntax of (anti-)causatives*. Amsterdam: John Benjamins Publishing Company.
- Schoenfeld, Robert Aaron. 2008. *The role and manifestation of the causative morpheme ‘-chi’ in Cuzco Quechua*. MA thesis, University of Pittsburgh.
- Tollan, Rebecca. 2018. Unergatives are different: Two types of transitivity in Samoan. *Glossa: A Journal of General Linguistics* 3 (1): 1–41.
- Treviño, Esthela. 1991. Derived objects, inner aspect, and the structure of VP. Talk given at *NELS* 22, University of Delaware.
- . 1992. Subjects in Spanish causative constructions. In *Romance languages and modern linguistic theory*, ed. Paul Hirschbühler and E.F.K. Joerner, 309–324. Amsterdam: John Benjamins.
- . 1994. *Las causativas del español con complemento de infinitivo*. Mexico: Colegio de México.

- Tyler, Matthew. 2020. *Argument structure and argument-marking in Choctaw*. PhD thesis, Yale University
- van de Kerke, Simon C. 1996. *Affix order and interpretation in Bolivian Quechua*. PhD thesis, University of Amsterdam.
- Weber, David. 1976. *Suffix-as-operator analysis and the grammar of successive encoding in Llacon Huánuco Quechua*. Yarinacocha Peru: Instituto Lingüístico de Verano.
- . 1983. The relationship of morphology and syntax: Evidence from Quechua. In *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session*, 161–181.
- . 1989. *A Grammar of Huallaga (Huanuco) Quechua*. Los Angeles: University of California Press.
- Wood, Jim. 2015. *Icelandic morphosyntax and argument structure*. Dordrecht: Springer.
- Wood, Jim, and Neil Myler. 2019. Argument structure and morphology. In *Oxford research encyclopedia of linguistics*, ed. Mark Aronoff. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.605>.

# Definiteness in A'ingae and Its Implications for Pragmatic Competition



Holly Zheng and Scott AnderBois

## 1 Introduction

### 1.1 Definiteness Cross-Linguistically

The contrast between definiteness and indefiniteness encodes the semantic feature of “uniqueness” and/or “familiarity” of noun phrases, although the exact definitions of these two types of noun phrases tend to have a blurry boundary. On an intuitive level, for example, in (1a), the entity “person” is considered new in discourse, as it has not been mentioned and is not familiar to the interlocutors in the discourse. On the other hand, “person” in (1b) needs to be already salient and familiar to the interlocutors for the definite marker “the” to be felicitous.

- (1) a. I met *a person* yesterday.
- b. I met *the person* yesterday.

For definite noun phrases such as (1b), analyses of English definiteness have identified two different kinds of definite uses, as in (2–3).

- (2) *Unique* definite: The content of a noun phrase can only be attributed to a single entity (in a given context).
  - a. *The professor* in our class
  - b. *The Queen* of England
- (3) *Anaphoric* definite: A noun phrase refers to an entity previously mentioned in the discourse.
  - a. I saw a movie yesterday. *The movie* was bad.

---

H. Zheng (✉) · S. AnderBois  
Brown University, Providence, RI, USA  
e-mail: [yuqi\\_zheng@alumni.brown.edu](mailto:yuqi_zheng@alumni.brown.edu); [scott\\_anderbois@brown.edu](mailto:scott_anderbois@brown.edu)

As shown in the above examples, English *the* allows for both uniqueness and anaphoric interpretations, and there has been extensive debate over the relationship between the two notions in English. While the English pattern appears to be not uncommon, recent works by Schwarz (2013) and others have shown that a great many languages encode uniqueness and anaphoricity in distinct ways. For example, in Fering, Schwarz (2009) shows that two distinct determiners correspond to unique and anaphoric noun phrases: a “weak” determiner is only available for uniqueness uses, while a “strong” determiner is used for anaphoric contexts. A possibly more prevalent pattern cross-linguistically is the one Jenks (2018) presents for Mandarin Chinese in which bare nouns are used in situations supporting uniqueness, and demonstratives are used in situations that call for anaphoricity.

Common to the English, Fering, and Mandarin Chinese patterns as presented by these authors is a strict complementarity between the forms used in indefinite, unique, and anaphoric contexts. Focusing on the definites, this is to say that in a language that makes such distinction between the unique and anaphoric definite forms, the form that is used for unique DPs is not also available for anaphoric DPs. At the level of analysis, a leading idea has therefore been to derive this complementarity from “hard” competition in which *Maximize Presupposition!* (or a similarly general pragmatic principle such as Jenks (2018)’s *Index!* principle) obliges the use of the anaphoric form where possible. Since complementarity is claimed to arise from pragmatic principles that are taken to be universal, such approaches therefore make a strong prediction that complementarity of this sort will also be universal.

In this chapter, we focus on the expression of definiteness in A’ingae, an endangered language isolate in the Amazonian Ecuador and Colombia. Empirically, we show that the A’ingae definiteness pattern quite systematically shows a lack of complementarity and therefore challenges theories relying on hard competition of this sort. Specifically, we show that the general pattern in A’ingae is that although bare noun phrases can be used in all of indefinite, unique, and anaphoric definite noun phrases, A’ingae also has a dedicated nominal anaphoric morpheme, *tsa*, that is felicitous only in anaphoric definite noun phrases.

The chapter is structured as follows: after a description of definite noun phrases in A’ingae (Sect. 3), we present a few pragmatic competition principles from previous work that have aimed to generalize over the structure of definiteness cross-linguistically, and we argue that none of them permits the empirical pattern of A’ingae definite NPs (Sect. 4). We then sketch an alternative semantic analysis for A’ingae that eschews hard pragmatic competition (Sect. 5). While applying this sort of approach more broadly is left to future work, we conclude with some thoughts on how this approach could extend to other languages and the diachronic role of pragmatic competition under such a view (Sect. 6).



## 2 Background on A'ingae and the Cofán People

The Cofán, or A'i, are an indigenous people of Amazonian Ecuador and Colombia. The language of the Cofán people is A'ingae (ISO: *con*), a language isolate spoken by around 1500 native speakers (Repetti-Ludlow et al. 2020). In the current literature, the autonym “A'ingae” and the exonym “Cofán” are commonly used as names for the language. A'ingae is an understudied and severely endangered language, especially in Colombia. In both Ecuador and Colombia, Cofán territory and therefore lifeways are under threat from extractive industries such as oil and mining, as well more indirect threat from relatively newly built roads and colonization (see Dąbkowski (2021) for a recent summary).

A'ingae is a SOV language with flexible word order in matrix clauses. A'ingae morphology is robust and complex—its set of suffixes and clitics encodes a large number of semantic categories, including aspect, subject person and number, switch reference, various modalities, and others. A more detailed overview of the A'ingae morphology can be found in Fischer and Hengeveld (2023), and a study on verbal morphology can be found in Dąbkowski (2019).

Of most direct relevance here is the structure of the nominal expressions, specifically in argument position. We provide a template for noun phrases in the language below in Table 1 where we point out the major components of a noun phrase in A'ingae. Determiner phrases in A'ingae have the order Det/Dem–Num–N. As shown in the table, adjectives and other modifiers may immediately precede or follow the head noun. Some example noun phrases are shown in (4).<sup>1</sup>

(4) *Example noun phrases in A'ingae:*

- a. *biani-'sû ande*  
far-ATTR country  
'a far-away country'
- b. *hashe'ye-ndekhû-'ye*  
old.man-PLH-HONR

**Table 1** Template of major components of A'ingae noun phrases

| –4                    | –3        | –2      | –1              | 0    | +1              | +2                               |
|-----------------------|-----------|---------|-----------------|------|-----------------|----------------------------------|
| Determiner            | Possessor | Numeral | Other modifiers | Head | Other modifiers | Enclitics for number, size, etc. |
| Demonstrative         |           |         | Adjective       | Noun | Adjective       | Assoc. Pl.                       |
| Quantifier            |           |         | RC              | ∅    | RC              | Augmentative Pl.                 |
| Definite <i>tsa</i>   |           |         |                 |      |                 |                                  |
| Indefinite <i>fae</i> |           |         |                 |      |                 |                                  |

<sup>1</sup> In addition to the Leipzig conventions, the following glossing abbreviations are used: ADD “additive,” ANA “anaphoric,” AND “andative,” ATTR “attributive,” CMP “comparative,” CT “contrastive topic,” DS “different subject,” FRST “frustrative,” HONR “honorific,” ITER “iterative,” PEJ “pejorative,” PLH “human plural,” PLS “plural subject,” PROP “property,” PRSP “prospective,” REP “reportative,” SBRD “subordinator,” SS “same subject,” and VER “veridical.”

- ‘the late elders’
- c. va kuchhi na kû’-a=ma  
 PROX pig meat red-ADJR=ACC  
 ‘this red pig meat’
- d. ñu-tshi-a  
 good-PROP-ADJR  
 ‘a good one’

(Fischer and Hengeveld 2023, (62, 96, 63, 69))

There are no numeral classifiers in A’ingae, though there is a robust group of classifier-like nominalizers that are noun-producing suffixes, most of which relate to the shape of an object (Fischer and Hengeveld 2023, p. 22). A’ingae has a small repository of numerals in regular use (*fae’khu* “1,” *khuangi* “2,” and *khuani fae* “3”), with higher numerals most typically expressed with Spanish borrowings or more rarely with morphologically complex native forms. The language also has a relatively small number-marking system: there is a morpheme *-ndekhû* for human plurality that attaches to the head noun and an associative plural suffix *-pa*. Outside of these plural morphemes, the number of an entity is not marked morphologically and is understood via context (or verbal agreement). There is an indefinite marker *fae* related to the numeral *fae’khu* “one.” The description and analysis of A’ingae indefinite and definite noun phrases will be the focus of Sect. 3.

Case marking is extensive in A’ingae, and case markers may be followed by additional morphemes related to information structure. A’ingae is a largely dependent-marking language with case clitics showing a nominative-accusative alignment. Argument roles, such as subject and direct object, are expressed through clitics that attach to the relevant NP and are not expressed on the verb.<sup>2</sup> For example in (5), the accusative marker *=ma* attaches to the object of the sentence *rande kuri-fi’ndi* “large (amount of) money,” and the dative marker *=nga* is attached to the recipient argument *ke* “2.sg.” More detailed discussion on case marking in A’ingae can be found in Fischer and Hengeveld (2023).

- (5) Rande kuri-fi’ndi=**ma**=ngi ke=**nga**=ja afe.  
 big gold-CLF=ACC=1 2.SG=DAT=CT give.  
 ‘I gave you big money (a large bill).’  
 (Fischer and Hengeveld 2023, (42))

<sup>2</sup> In this chapter, we will use the term “DP” to refer to maximal nominal expressions, since they can include a determiner. We leave to future work the precise syntactic analysis on the difference between “NP” and “DP.”

A'ingae also has marking for information structure and topic: *=ta* is tentatively analyzed as a marker for new topics, and *=ja* is for contrastive topic. Word order in matrix clauses and second position clitics seems to have a connection with information structure too, but details of each of these elements await more detailed investigation.

Data presented in this chapter without citation are gathered through elicitation, primarily with one native speaker, Shen Aguinda, from the Ecuadorian community of Dureno, and a small portion of the elicited examples come from native speakers of Zábalo and Dovuno. All elicitations were conducted remotely over Zoom video conferences, and we primarily communicated with the consultants in Spanish, which is a language that is commonly used in the Cofán communities besides A'ingae. Other data come from published sources as cited and otherwise come from texts from the A'ingae Language Documentation Project (AnderBois and de Lima Silva 2017), and these natural data primarily come from Zábalo. For naturalistic examples, most of their citations are hyperlinked and lead to video fragments of the example within a larger narrative as presented on the A'ingae Language Documentation Project (2022) website, powered by LingView (Pride et al. 2020). Despite regional differences in where the data and consultant come from, none of the main observations in this chapter differs across dialects to our knowledge.

### 3 Expressions of (In)Definiteness in A'ingae

In A'ingae, indefinite, unique definite, and anaphoric definite noun phrases can all be encoded as bare noun phrases. However, there is also a dedicated nominal anaphoric morpheme *tsa* that can also be used in anaphoric definite noun phrases (as well as a dedicated indefinite determiner *fae*). Neither bare NPs or *tsa* can be used in the “exophoric” contexts, where the reference is established through the referent being in the physical environment of the speakers—the proximal or distal demonstrative, *va* and *juva*, respectively, must instead be used, and *tsa* is not felicitous in these deictic contexts. The overall pattern for the uses of bare nouns and *tsa* is summarized in Table 2, and the remainder of this section justifies these claims in detail.

**Table 2** Summary of uses of bare noun and *tsa* in A'ingae

|            | Indefinite | Unique definite | Anaphoric definite | Bridging | Exophoric |
|------------|------------|-----------------|--------------------|----------|-----------|
| Bare noun  | ✓          | ✓               | ✓                  | ✓        |           |
| <i>tsa</i> |            |                 | ✓                  |          |           |

### 3.1 Indefinite Noun Phrases

Indefinite noun phrases are encoded as bare noun phrases across different syntactic positions. In (6), the indefinite *pandu* (“fox”) is in the subject position, while in (7), the indefinite *kusina* (“kitchen”) is in the object position. In (6), “fox” is introduced for the first time in the story, so it is an indefinite noun phrase whose referent’s existence is introduced. (7) represents a specific context in which indefinite noun phrases tend to appear: existential predicates and predicates of coming into existence, as here. In (7), the sentence includes the action of “building,” which leads to the existence of a new item as the result of the “building” action. The phrase for the kitchen *kusina* is an indefinite noun phrase, encoded as a bare noun. In general, however, existential predicates are not necessary for bare nouns to be used in indefinite noun phrases in the object position.

- (6) (*Context: Introducing the fox character in a story.*)

Pandu tsûifa’u jaiya.

**pandu** tsûi=fa’u      jai=’ya  
fox    walk=PEJ.ACC go.PRSP=VER

‘A fox walked by.’

(Kuke chiste FC 2:38)

- (7) Kusinavengi tsau’ña’je’fa.

**kusina=ve=ngi** tsau’ña-’je-’fa.  
kitchen=ACC2=1 build-IPFV-PLS

‘We’re building a kitchen.’

(Construir una casa de conambo

MM 1:54)

The nominal anaphor *tsa* is strictly limited to anaphoric contexts, so it is not felicitous in an indefinite noun phrase. In the context of (6), for example, adding *tsa* to the bare noun *pandu*, “fox” would suggest that the same fox has been mentioned in previous discourse, which contradicts the context of the sentence where the fox is introduced for the first time.

- (8) {\*Tsa} pandu tsûifa’u jaiya.

{\***tsa**} pandu tsûi=fa’u      jai=’ya  
{ANA} fox    walk=PEJ.ACC go.PRSP=VER

(*Intended: Introducing the fox character in a story*) ‘A fox walked by next to the hare.’

Finally, we note that the indefinite article *fae* can systematically be optionally inserted in indefinite contexts. To our knowledge, there is no interpretive difference between *fae* and bare nouns in such cases.

### 3.2 *Unique Definite Noun Phrase*

Unique definite noun phrases have referents that are unique within a certain context, i.e., for singulars, only one individual within a given situation meets the descriptive content. Previous work has often distinguished two categories of unique definites: “globally” unique noun phrases that have a unique reference because of our knowledge of the world or common sense and “locally” unique noun phrases that are unique given a narrower context, for example, the interlocutors’ surroundings, personal experiences, etc. In A'ingae, both types of unique definiteness are encoded as bare noun phrases. For example, *kue'je* (“sun”) in (9) is a globally unique noun phrase, and it is presented as a bare noun in the sentence. The referent of “house” is usually not unique, but *tsa'u* (“home”) in (10) refers to the only salient house in the story that the speaker is trying to tell, and here *tsa'u* is also in its bare form.

As shown in these two examples, *tsa* cannot be used in any of unique definite noun phrases, such as “the sun” in (9) and “the house” in (10), because it is not felicitous in contexts that lack the anaphoric interpretation.

- (9) {\*Tsa} *kue'jenga khûtsiansi tsaja aceite yaya'pave daya'ya.*

{\*tsa} *kue'je=nga khûtsû-an-si tsa=ja aceite yaya'pa=ve*  
 {\*ANA} *sun=DAT stand-CAUS-DS ANA=CT oil oil=ACC2*  
*da-ya-'ya.*  
 become-IRR-VER

‘Having been stood in the sun, it (mashed turtle egg) would turn into oil.’  
 ([Charapa proyecto 1:07](#))

- (10) (*Context: the house has not been mentioned before but is known to the speakers.*)

*Kuse vangakhe napi {\*tsa} tsa'unga.*

*kuse va=nga=khe napi {\*tsa} tsa'u=nga.*  
 night PROX=DAT=ADD arrive {\*ANA} house=DAT

‘I arrived at the house at night.’

Because *tsa* requires an antecedent to be present in prior discourse, it is also not felicitous in generic noun phrases, which have been argued to connect to the notion of uniqueness. In (11), for example, the noun phrase “snakes” is a generic noun phrase that refers to the snake species, and here *tsa* is not felicitous—bare *iyu* has to be used. Although the example here has the plural form for “snakes,” the infelicity of *tsa* here in this generic noun phrase is not dependent on the plural marker.

- (11) {\*Tsa} *iyundekhûtsû tsai'jefa.*

{\*tsa} *iyu-ndekhû=ta=tsû tsai-'je-'fa.*  
 {\*ANA} *snake-PLH=NEW=3 bite-IPFV-PLS*

‘Snakes bite.’

### 3.3 Anaphoric Definite Noun Phrase

Anaphoric definite noun phrases have a referent that is known to the speakers because the referent has been previously mentioned. In A'ingae, anaphoric definiteness can be encoded as either a bare noun phrase or with the nominal anaphor *tsa*. For example, (12) shows that the anaphoric definite “book” in the second sentence can be in either its bare form *tevaenjen* or in a noun phrase with *tsa*.

- (12) Chavangi fae tevaenjenma. {Tsa tevaenjen/Tevaenjen} panshaen karu.  
 chava=ngi fae tevaenjen=ma. {**tsa tevaenjen/tevaenjen**} panshaen karu.  
 buy=1 one book=ACC {ANA book/book} very expensive  
 ‘I bought a book. The book was very expensive.’

Specifically for the nominal anaphoric morpheme, *tsa* is available in referring to individual entities both adnominally and pronominally—in (13), *tsa* can either co-occur with the noun *tevaenjen* or replace the noun entirely. In addition to referring to individuals and objects, pronominal *tsa* can also anaphorically refer to propositions conveyed by previous discourse, such as the proposition “Red apples are delicious” in (14).

- (13) Chavangi fae tevaenjenma. {Tsa tevaenjen/Tsa} panshaen karu.  
 chava=ngi fae tevaenjen=ma. {**tsa tevaenjen/tsa**} panshaen karu.  
 buy=1 one book=ACC {ANA book/ANA} very expensive  
 ‘I bought a book. The book was very expensive.’

- (14) Kû’a manzanandekhûtatsû yayatshi’fa. Tsama atesûngi Juan ñanga kundasi.  
 kû’a manzana-ndekhû=ta=tsû yaya-tshi-’fa. **tsa**=ma atesû=ngi Juan  
 red apple-PLH=NEW=3 good-PROP-PLS ANA=ACC know=1.SG Juan  
 ña=nga kunda-si.  
 I=DAT tell-DS

‘Red apples are delicious. I know that because Juan told me.’

When *tsa* is referring to a proposition, the exact proposition does not need to be linguistically explicit for *tsa* to be felicitous. The proposition can be part of the implicature of the preceding discourse. In (15), for example, *tsa* picks up the *combined* content of the two preceding sentences. This shows the flexibility of *tsa* as a propositional anaphor that can pick up pragmatic content salient from prior discourse, also observed in Morvillo and AnderBois (2022), as well as complex discourse units, such as the conjunction of two propositions as shown in (15).<sup>3</sup>

<sup>3</sup> Many other works have investigated propositional anaphora more in depth, such as discussions on English “that” Snider (2017) and citations therein.

- (15) (Context: *My friend and my brother don't get along well.*)

Ña faengasûtsû ja'ñu ña'khû añã. Ñã antiankhe ña'khû añã. Tsatsû aiyepã.

ña faengasû=tsû ja'ñu ña=i'khû añã. ñã antian=khe ñã=i'khû añã. tsa=tsû  
 my friend=3 now 1.SG=INS eat my brother=ADD 1.SG=INS eat ANA=3  
 aiyepã.  
 difficult

'My friend is going to have dinner with me. My brother will, too. That is difficult.'

A special case of an anaphoric reference is situations where an anaphoric noun phrase corefers with a previously mentioned indefinite phrase despite not being able to be bound to that indefinite phrase in standard ways that quantifiers usually would allow. An instance of such example comes from a covarying situation, or a “donkey” sentence. In these sentences, anaphoric definites receive quantificationally bound interpretations despite the absence of a c-commanding antecedent in the same clause. In a donkey sentence in A'ingae, the anaphoric definite noun phrase can be encoded as either bare or with *tsa*, as shown in (16) and (17). In this way, *tsa* patterns with pronouns and anaphoric definites in other languages rather than, say with adjectives such as English  *aforementioned*  that more explicitly require prior mention as such.

- (16) Majan a'ima ke thû'senindangi (tsa) a'ima atheya.

majan a'i=ma ke thû'se=ni=ta=ngi (tsa) a'i=ma athe-ya.  
 who person=ACC 2.SG call=LOC=NEW=1.SG (ANA) person=ACC see-IRR

'Whoever you invite, I will see that person.'

- (17) Pûi afafanga dûshûkhû kuraga, (tsa) dûshûtsû dyu'je.

pûi afa<fa>- 'nga dûshû=i'khû kuraga, (tsa) dûshû=tsû dyu-'je.  
 each speak<ITER>-AND child=INS shaman (ANA) child=3 scare-IPFV

'Every time a shaman talks to a child, the child gets scared.'

Jenks (2018) points out that, in Mandarin, both bare nouns and demonstrative phrases can be felicitous in certain anaphoric definites in the subject position. He argues that the availability of bare noun as anaphoric definites here is because the bare noun phrase serves as a continuing topic, therefore being an exception to the fact that bare nouns are not available in anaphoric definites in Mandarin. He then argues that the pragmatic function of topic marking overrides and neutralizes the effect of an indexical constraint in such environments.

In A'ingae, we see a slight tendency for anaphoric definite NPs to occur with the contrastive topic marker =*ja*, such as the subject of the sentence *ainja* in (18). Despite this, we also see anaphoric bare nouns in non-topic position (such as (19)). In addition, *tsa* is also felicitous in both topic and non-topic positions, as long as the discourse satisfies the anaphoric context. This indicates that the possible patterns of

(in)definiteness in A'ingae are unaffected by any syntactic constraints, including a topic position.

- (18) *Context: The story is talking about a man and his dog in the previous sentences.*

Jata ainjan tayu kuankuan kan'jeni ja'ya.

ja=ta **ain=jan** tayu kuankuan kan'jen=ni ja-'ya.  
go=NEW dog=CT already Coancoan stay=LOC go-VER

'The dog went right where the Coancoan lived.'

(*Kuankuan kundasepa OCQ 4:44*)

- (19) Fae tsandie tuyakaen fae pûshesû kanje'fa tsa'uni, tsa'ma ñangi afa push-esûkhû.

fae tsandie tuya-'kan-e fae pûshesû kanje-'fa tsa'u=ni, tsa-'ma  
one man still-CMP-ADV one woman live-PLS house=LOC ANA-FRST  
ña=ngi afa **pushesû=i'khû**  
I=1.SG talk woman=INS

'There is a man and a woman in the house. I talked to the woman yesterday.'

Finally, while *tsa* looks in several respects more like an English demonstrative than a definite article (e.g., pronominal uses, propositional uses), it crucially differs from demonstratives in English and Mandarin, for example, in deictic contexts. Whereas demonstratives such as English *that* can be used exophorically, *tsa* is infelicitous in such uses. This pattern is not surprising given the evidence above that *tsa* requires an explicit anaphoric context to be felicitous, but contrasts clearly with demonstratives in many languages too. In an exophoric use, such as in (20), the speaker can use the noun phrase "that bird" without any previous mentioning of the bird, because some extra-linguistic cue (e.g., the speaker's pointing gesture) is sufficient in signaling the referent of the noun phrase. In such exophoric contexts, the exophoric demonstratives *va* PROX or *juva* DIST are used instead.

- (20) Kanjan, {juva/\*tsa} chhiririatsû vasia've chhaje

kan=jan, {juva/\*tsa} chhiriria=tsû vasia've chhaje.  
look=IMP {DIST/\*ANA} bird=3 slowly fly

'Look, that bird is flying slowly.'

### 3.4 Bridging Definites

A special case of anaphoric definite noun phrases, "bridging" anaphora, is first discussed by Clark (1975) (see also Hawkins (1978)'s "associative anaphora" and Prince (1981)'s "Inferrables"). In these constructions, as in (21–22), a definite noun phrase in the second clause refers back not to the entity mentioned in the first clause, but rather to some entity that is related to it in some way.



- (21) He drove his car down the street. *The steering wheel* was cold.  
 (22) I read a book yesterday, and *the author* was on TV this morning.

Bridging definites therefore can be thought of as having both an anaphoric component (the previously mentioned car/book above) and a uniqueness component (the relation that connects the definite's descriptive content to the prior referent must allow the address to uniquely pick out the referent).

For languages that express unique and anaphoric definites distinctly, the question therefore arises of which form is used in instances of bridging. Interestingly, Schwarz (2013) claims that, for German and a few other languages, the answer depends on the nature of the bridging relation. In some cases, such as part-whole relations as in (21) and (23a), the uniqueness form is used, intuitively because the relation itself is manifest in the previously established situation. In cases such as the producer-product relation in (22) and (23b), the bridging relation lies outside of the previously mentioned situation (i.e., the writing of the book occurred at another time and place).

- (23) a. Der Kühlschrank war so groß, dass der Kürbis problemlos  
 the fridge was so big that the pumpkin without.a.problem  
**im/#in dem Gemüsefach** untergebracht werden konnte.  
**in-the<sub>weak</sub>/#in the<sub>strong</sub> crisper** stowed be could  
 'The fridge was so big that the pumpkin could easily be stowed in the crisper.'
- b. Das Theaterstück missfiel dem Kritiker so sehr, dass er in seiner  
 the play displeased the critic so much that he in his  
 Besprechung kein gutes Haar **#am/an dem Autor** ließ.  
 review no good hair **#on-the<sub>weak</sub>/on the<sub>strong</sub> author** left  
 'The play displeased the critic so much that he tore the author to pieces in his review.'
- (Schwarz 2013, (16ab))

There remain many empirical and analytical details to be further understood on such splits in bridging in German and elsewhere. In A'ingae, however, we find that bridging shows no such split. Bridging of both types of intuitive relations makes use of bare nouns: (24) is an example of part-whole bridging definite, where the nouns "head" *tsuve* and "body" *ai'vu* are parts of the musk hog that was introduced in previous discourse; (25) is an example of producer-product bridging definite, where "singer" *sethapuen'sû* in the second half of the sentence is the producer of its antecedent "song" *sethapuen'chu*.

- (24) *Context: A story about hunting a musk hog.*

Ma'the pu'taeña tsuveyeti pu'taeña asi'thaemba ai'vuye pu'taeña.

ma'the pu'ta-en-ña **tsuve=ye=ti** pu'ta-en-ña asi'thaen-pa  
 where pierce-CAUS-IRR head=ELAT=INT shoot-CAUS-IRR think-SS  
**ai'vu=ye** pu'ta-en-ña  
 body=ELAT pierce-CAUS-IRR

‘“Where should I shoot it? Should I shoot through the head?” I thought, “should I shoot through the body?”’  
(Caza y pesca OCQ 1:42)

- (25) Sethapuen’chutsû mendetshi, tsa’ma sethapuen’sûma atesûmbi.

Sethapuen-’chu=tsû mende-tshi, tsa-’ma **sethapuen-’sû**=ma atesû-mbi.  
sing-SBRD=3 beautiful-PROP ANA-FRST sing-ATTR=ACC know-NEG

‘The song is beautiful, but I don’t know the singer.’

Given that we have already seen that bare nouns have both unique and anaphoric uses, it is perhaps unsurprising that they may be used in both cases. More strikingly, however, we find that *tša* is unavailable in either type of bridging. Of specific interest are the producer–product cases, which have often patterned with the anaphoric definite morpheme in other languages. Intuitively, the singer has not been explicitly mentioned before, and so *tša* is thus not available, as seen in (26).

- (26) #Sethapuenchutsû mendetshi, tsa’ma tša sethapuen’sûma atesûmbi.

#Sethapuen-’chu=tsû mende-tshi, tsa’ma **tša** sethapuen-’sû=ma atesû-mbi.  
sing-SBRD=3 beautiful-PROP but ANA sing-ATTR=ACC know-NEG

Intended: ‘The song is beautiful, but I don’t know the singer.’

In summary, we have presented empirical evidence for the pattern of definiteness in A’ingae: bare nouns are available in indefinite, unique, and anaphoric definite noun phrases, while the nominal anaphor *tša* is only felicitous in anaphoric definite noun phrases.

### 3.5 *A Lack of Complementarity Between Unique and Anaphoric Forms*

As shown previously in this section, the unique definite noun phrases in A’ingae can only be presented as a bare noun phrase, while the anaphoric definite noun phrases can be presented either as a bare noun phrase or a noun phrase with *tša*. The crucial implication of this empirical pattern is the lack of complementarity between the unique and anaphoric definite forms in A’ingae.

As we discuss in detail below, this lack of complementarity presents a challenge to many recent cross-linguistic accounts, many of which propose “weaker” and “stronger” determiners meanings along with pragmatic competition between them obliging the use of the stronger anaphoric form where possible. In A’ingae, however, we have seen that there is a “strong” exclusively anaphoric form *tša*, yet bare nouns are freely available in anaphoric contexts.

Before examining this challenge in more detail, we first note that A’ingae is not alone in having this empirical pattern of a lack of complementarity between the

unique and anaphoric forms. Recent cross-linguistic works covering more languages have revealed that several other languages present similar empirical pictures: a lack of complementarity between bare nouns that also have uniqueness uses and dedicated anaphoric determiners. Among these languages are: Shan (Tai-Kadai family, Moroney 2021), Tumbalá Ch'ol (Mayan family, Vásquez Martínez and Little 2020, Little 2020), San Pedro Mixtepec Zapotec (Vásquez Martínez 2020), San Pedro Güilá Zapotec (Arrieta Zamudio 2020), and Tsotsil Sureño (Mendoza 2021). More detailed examples supporting the non-complementarity can be found in these cited works as well as Chapter 2 of Zheng (2022).

Common to A'ingae and the other languages cited here is the lack of complementarity between the unique and anaphoric forms of each language. This wide cross-linguistic empirical evidence on the lack of complementarity leads to a challenge to a commonly used strategy based on pragmatic blocking to analyze and predict the structure of definiteness cross-linguistically. In the next section, we will elaborate on some of these pragmatic-based theoretical frameworks. We will argue that such frameworks based on pragmatic competition do not account for the empirical pattern presented in A'ingae and the languages we have cited above, and then we will provide a preliminary alternative analysis for bare nouns and *tsa* in A'ingae that does not rely on pragmatic blocking but is instead rooted in the semantic (anti-)presuppositions of both forms.

## 4 Pragmatic Blocking Is Incompatible with A'ingae Definiteness

In the previous section, we have described the inventory of (in)definite expressions in A'ingae. Here, we turn to consider several recent pragmatic accounts that aim to explain aspects of such inventories and argue that for each account, patterns such as the A'ingae one are unexpected.

### 4.1 *Competition Between Covert and Overt Determiner Form*

One prominent framework using pragmatic competition to analyze definiteness patterns is that proposed in Chierchia (1998) and extended in subsequent work in Dayal (2004). These works focus on languages with definite uses of bare nouns, proposing that such languages utilize a semantic type-shifting operation to produce definite uses of bare nouns. In order to explain why such a type shifter is only available in certain languages, they propose that this type shifting is subject to a blocking principle:

- (27) *Blocking Principle:*  
Don't do covertly what you can do overtly!

This blocking mechanism, then, specifically applies to the availability of overt articles vs. bare noun phrases in a language: if there are overt determiners, use this overt determiner instead of performing the type shifting on a bare form. This is a competition in the sense of the Gricean maxim of manner, since the competition mainly applies to the form of determiners. The maxim of quantity is of secondary relevance here, because the competing forms must first be both usable in a given utterance to have the competition arise, but the primary explanatory force is manner-based.

In the A'ingae pattern, a direct counterexample to the predicted pattern under the blocking principle is the dedicated anaphoric marker *tsa*. In definite noun phrases, both bare nouns and *tsa* are available in anaphoric definite noun phrases; therefore, the overt form, *tsa*, does not block the bare form.

Another potential instance that challenges whether blocking principle can be fully applied in A'ingae is the overt indefinite article, *fae*, which is related to the numeral *faekhu* “one.” *Fae* is limited to indefinite contexts, while at the same time bare nouns are also available as indefinite noun phrases. In this chapter, we do not have detailed semantic analysis of *fae*, so the assumption here is that *fae* and indefinite bare noun phrases both only assert existence of the referent. If this is the case, then the availability of *fae* also contradicts the blocking principle, because the existence of the overt indefinite form *fae* does not block the bare nouns in indefinite noun phrases. However, if the semantics of *fae* is more complicated, such as *fae* contributing specificity or an epistemic meaning of some sort on top of existence of the referent, then the blocking principle from Chierchia and Dayal would be compatible. In this case, since the overt *fae* would be doing different work overtly than the covert operator, the principle would not be violated. However, we have not seen empirical evidence of such a “specificity” requirement in contexts where *fae* occurs, but careful scrutiny of the semantics of *fae* is necessary for future work.

## 4.2 Maximize Presupposition!

Another type of pragmatic competition that has been argued to apply to not only the domain of definiteness but also many other linguistic features is *Maximize Presupposition!* (Heim 1991; henceforth MP). MP is proposed as a general economy principle that chooses the form with stronger presuppositions among otherwise equivalent competing forms (see Bade (2021) for a recent survey of MP and its proposed application in other domains besides definiteness). MP directly concerns pairs of forms that differ minimally in each form’s presupposition, where the “stronger” form triggers a semantic presupposition that the “weaker” form lacks. MP predicts that the weaker form is infelicitous in a context where the presupposition in question is already part of the common ground. On the other hand, in contexts where the presupposition is *not* part of the common ground, using the weaker form implicates that the presupposition is false or unknown.

MP was originally proposed to account for data such as (28), where these sentences are not necessarily *false* but *infelicitous*. This infelicity is analyzed to *not* arise from a presupposition failure of the indefinite determiner “a” because of observations such as (29). The empirical generalization from data such as (28) is that these utterances already satisfy the presuppositions of “the” (the uniqueness of the sun in (28a) and the weight of the tent in (28b)), so using “a” is infelicitous because it is the weaker form without the uniqueness presupposition.

- (28) a. # A sun is shining cf. The sun is shining.  
 b. # A weight of the tent is 5 kg. cf. The weight of the tent is 5 kg.
- (29) a. Robert caught a 20-foot catfish.  
*does not presuppose:* There is more than one 20-foot-long catfish.

In addition to its use in understanding the relationship between definites and indefinite in languages such as English, MP-like reasoning has more recently been applied to understanding the relationship between uniqueness and anaphoric definites in languages making more fine-grained definiteness distinctions.

To get such an account off the ground requires a particular semantics of uniqueness and anaphoricity, one in which there is a relationship of asymmetric entailment between their respective presuppositions. While not explicitly adopting MP-based competition, Schwarz (2009) provides such a semantics in his account of “weak” and “strong” determiners in German (which mostly show complementarity as noted above). Specifically, Schwarz proposes that both unique and anaphoric definites presuppose the existence of a unique individual to which they refer, but that anaphoric forms contain an additional index argument that is not present in the unique definite form; therefore, the anaphoric form’s presupposition entails that of the unique form.

Deriving from *Maximize Presupposition!* and building upon Schwarz’s analysis of the asymmetrical entailment relationship between unique and anaphoric definites, Jenks (2018) focuses on Mandarin Chinese and Thai and proposes a more specialized competition strategy based on pragmatic blocking, which he calls *Index!*. Jenks’ analysis for the unique and anaphoric forms in a language adopts the part of the analysis from Schwarz (2009) that treats the anaphoric form as having an additional index variable, and *Index!* states that an index should be represented explicitly whenever possible.

- (30) *Index!*  
 Represent and bind all possible indices. (Jenks 2018, (53))

*Index!* builds off of MP by connecting the “index” semantics of the anaphoric form with an asymmetrical entailment between the unique and anaphoric forms. Both forms presuppose the existence of a unique individual, but because the anaphoric form contains the additional index variable that makes the presuppositions of the anaphoric form stronger than that of the unique form, the anaphoric form should be used whenever possible (Jenks 2018, p. 14). The anaphoric form entails the uniqueness form, but not the other way around.

In Mandarin, for example, Jenks argues bare nouns are—setting aside existential uses—allowed only in uniqueness contexts, while demonstrative phrases such as *zhe ge* and *na ge* are obligatory in anaphoric definite phrases. *Index!* blocks the use of a uniqueness definite in anaphoric contexts due to the stronger presupposition of the competing anaphoric demonstrative. Jenks then argues that these demonstrative phrases are the dedicated anaphoric morphemes in Mandarin. In (31), for example, Jenks argues that the demonstrative phrase “na ge” is obligatory, because the noun phrase in (31b) is an anaphoric definite in non-subject position.<sup>4</sup>

- (31) a. Jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng,  
classroom inside sit-PROG one CLF boy and one CLF girl,

‘There is a boy and a girl sitting in the classroom...’

- b. Wo zuotian yudao #(na ge) nansheng  
I yesterday meet that CLF boy

‘I met the boy yesterday.’

(Jenks 2018, (16))

Returning to A’ingae, we see that Jenks (2018)’s MP-based approach similarly faces a challenge. In particular, in anaphoric contexts, the Jenks/Schwarz semantics holds that the “weak” definite bare nouns contribute uniqueness and therefore semantically are compatible with uniqueness of any kind. The anaphoric determiner *tsa*, by hypothesis, contributes uniqueness as well as an index argument. It therefore has a stronger presupposition and according to MP or the more specific *Index!* ought to be used when possible. While this blocking effect has a virtue in explaining the data presented by Schwarz (2009) and Jenks (2018) in which the uniqueness form is infelicitous in anaphoric contexts, here it incorrectly predicts that we ought to find blocking, such that the bare noun is not available in anaphoric contexts. We have seen this illustrated above in Sect. 3, and we see it again in (32). The preceding line in the story discusses a pack of peccaries, and then we see anaphoric reference made to the aforementioned pack in consecutive lines, once with the bare noun, in (32a), and then once with *tsa* in (32b).

- (32) Context: A story about a man hunting peccaries

- a. Napisi sūya tayuti ja vaeyitsû mûnda ja khende sūya.  
napi-si sū-’ya tayu=ti ja vae=yi=tsû mûnda  
arrive-DS say-VER already=INT go already=EXCL=3 peccary  
ja khen=te sū-’ya.  
go QUOT=REP say-VER

<sup>4</sup> Jenks does point out one exception to the generalization of *Index!*, part-whole bridging cases, because the prior mention of an argument of the noun licenses the anaphoric form. See Jenks (2018) for detailed discussion.

‘When he came, he asked if the pack had already gone. “Just a moment ago,” they said.’

- b. Tsete tsa kuenza ûfambe pasaya tsumbate tse umbaemba jaya tsa mûndai'khû.

tse=te            tsa kuenza ûfa=mbe            pasa-'ya tsun-pa=te  
 ANA.LOC=REP ANA old    blow=NEG.ADV pass-VER do-SS=REP  
 tse            umbuen-pa ja-'ya    **tsa mûnda**=i'khû.  
 ANA.LOC follow-SS    go-VER ANA peccary=INS

‘Since he hadn’t hunted anything, he decided to follow the peccaries.’  
 (Kuankuan kundasepa OCQ 1:32)

Moreover, as noted above in Sect. 3.5, a number of other recent works present similar patterns of non-complementarity in which a dedicated anaphoric form and a more general uniqueness form are both available in anaphoric contexts. While the accounts of Schwarz (2009) and Jenks (2018) provide various ways to capture cross-linguistic variation, the hard competition of MP/*Index!* proposed by Jenks makes a strong cross-linguistic prediction that we will not find these sorts of instances of non-complementarity. Languages may differ in what denotations they have, but the pragmatic competition is precisely the part that is taken to be universal.

To summarize, in this section, we have seen that the lack of complementarity shown by A'ingae bare nouns and *tsa* in anaphoric contexts presents a challenge to pragmatic competition based on *Maximize Presupposition* or its specific instantiation as *Index!*. We turn now to consider one other more recent pragmatic account in Sect. 4.3.

### 4.3 Bare Noun Blocking

We have seen thus far that the pattern presented by A'ingae definiteness poses a challenge to two different sorts of pragmatic competition-based accounts: Chierchia (1998) and Dayal (2004)'s manner-based *blocking principle* and Jenks (2018)'s maximize presupposition-based *Index!*. In this section, we turn to consider one further competition-based proposal from Ahn (2019).

Motivated by recent empirical evidence from languages that do not present complementarity, this proposal essentially modifies the situation under which *Index!* occurs to be a condition that depends on the existence of a morphologically simplex pronoun in a language. Ahn claims that *Index!*-like blocking is found in all and only languages with morphologically simplex pronouns, a principle she calls “Bare Noun Blocking”:

(33) *Bare Noun Blocking*

If a bare argument language has morphologically simplex pronouns (“simplex pronouns”) for 3rd person reference, bare nouns are blocked from intersentential anaphora involving one salient entity (Ahn 2019, (25)).

Ahn further proposes that the basis for this blocking principle originates from a scale of anaphoricity for all languages, which states that languages have different lexicalizations of definite features that result in different anaphoricity scales. Then, a *Don’t Overdeterminate!* principle chooses the form lowest on the scale whenever possible and blocks any redundant expressions when a simpler form is available. This competition mechanism, therefore, concerns not only the meaning of determiner forms in a language but also its internal syntactic structure.

(34) English Scale of Anaphoricity: pronoun < definite description < demonstrative description (Ahn 2019, (82))

(35) *Don’t Overdeterminate!*: a principle that chooses the lowest element in the scale of anaphoric expressions that can successfully resolve the referent. (Ahn 2019, (90))

If we take *Maximize Presupposition!* and the subsequent *Index!* as derivatives of the maxim of manner in informational status, the *Don’t Overdeterminate!* principle could represent a different type of maxim of manner in the morpho-phonological forms of referring expressions: whenever there is a simpler form, *Don’t Overdeterminate!* chooses the simpler form instead of the more complex one that holds the same level of determinacy. Although different from *Index!*, the *Don’t Overdeterminate!* principle and the *Bare Noun Blocking* prediction still do not result in a correct depiction of the pattern in A’ingae.

As shown in Sect. 3, *tsa* can be used pronominally, which means it has the function of a morphologically simplex pronoun in A’ingae.<sup>5</sup> The existence of a simplex pronoun *tsa* (such as in (36) and (37)) does not block the existence of anaphoric bare nouns.

(36) Chavangi fae tevaenjenma. {Tsa tevaenjen/Tsa/Tevaenjen} panshaen karu.

chava=ngi fae tevaenjen=ma. {tsa tevaenjen/tsa/tevaenjen} panshaen  
buy=1 one book=ACC {ANA book/ANA/book} very  
karu.  
expensive

‘I bought a book. The book was very expensive.’

<sup>5</sup> An alternative possible analysis raised by Andrés Saab would be to analyze pronominal *tsa* as involving nominal ellipsis. However, we are not aware of any specific evidence that suggests that such examples do involve ellipsis, and more generally, the existence of demonstrative-like elements with pronominal and adnominal uses is cross-linguistically common and not necessarily attributed to ellipsis generally. Finally, it is somewhat unclear whether/how an elliptical analysis would impact Ahn (2019)’s predictions.





## 5 Toward a Semantic Alternative to Pragmatic Competition

Thus far, we have seen that anaphoric contexts in A'ingae freely allow for the use of bare nouns, adnominal/determiner *tsa*, or pronominal *tsa*. The language (and seemingly the others noted in Sect. 3.5), therefore, appears to differ with the languages described by Jenks (2018) and Ahn (2019) precisely in the felicity of a more general uniqueness form alongside the dedicated anaphoric *tsa*. While their exact accounts differ in the details of their pragmatic motivations, they each propose pragmatic mechanisms that are rooted in quite general principles. While many kinds of cross-linguistic variation can be captured with such frameworks, this particular kind therefore poses a seemingly intractable challenge.

Moreover, even for languages with apparent blocking such as Mandarin and German, several recent papers have suggested the empirical picture to be more complicated in ways that appear problematic. Dayal and Jiang (2021), for example, provide evidence that the complementarity between bare nouns and demonstratives in Mandarin may be less strict than *Index!* would dictate. In addition, Bremmers et al. (2021)'s corpus work on translated texts between Mandarin and German also shows that the distributions of German weak/strong articles and Mandarin bare nouns/demonstratives do not coincide precisely, which provides evidence from a different perspective that the split between Mandarin bare nouns and demonstrative phrases is perhaps not the same as the split in the German determiners.

If blocking-based pragmatic theories cannot capture this variation, then it would seem we are left with the following picture: synchronically, grammars do not have “hard” pragmatic competition in the domain of definiteness (though we presume they may have various forms of “soft” competition); rather, it is the *semantics* that determines such patterns more or less in their entirety. For example, A'ingae *tsa*, Mandarin demonstratives, and German strong determiners must have a semantics that restricts their use to only anaphoric situations (as in the accounts in Sect. 4). However, we suggest the same must be true for weak definite articles and bare nouns as well: A'ingae bare nouns must have a semantics that allows them to be compatible with either unique or anaphoric situations (again setting aside existential/indefinite uses for the moment). German weak definites and Mandarin bare nouns, on the other hand, must also have a *semantics* that determines their distribution, i.e., one that is compatible with uniqueness uses but not with anaphoric ones (the complications just noted above notwithstanding).

Theoretically, the only potential downside of this semantically based analysis is (arguably) a lack of parsimony. If the semantics of German weak articles is in some sense complementary to that of the strong articles, perhaps there is the intuition that this complicates the grammar more than necessary. The extent to which this is so, however, depends in part on how far one extends this approach, specifically if there are true cases of synchronic *Maximize Presupposition!* producing hard competition/blocking. Moreover, we may still understand a cross-linguistic tendency toward such complementarity between definite forms as the diachronic result of grammaticalization driven by soft competition synchronically—the crucial point here is that the only synchronic competition is a soft one, a preference of one definite

form over another in certain contexts, instead of a hard exclusivity between definite forms as would be dictated by *Maximize Presupposition!*-related analyses. As the grammar of a language evolves over time, these soft preferences ultimately lead certain definite forms to surface only in certain contexts. We leave it to future work to develop such accounts for German, Mandarin, and other cases, though we note that Chapter 5 from Schwarz (2009) for German and Dayal and Jiang (2021) for Mandarin are arguably steps in this direction.

Although applying such an approach cross-linguistically may present substantial challenges in some cases, for A'ingae, things appear relatively straightforward, with a few possible analytical paths. First, one could imagine adopting the semantics of Jenks (2018) for Mandarin. The overt determiner *tsa* would have an anaphoric presupposition,  $t^x$ , following Schwarz (2009) and Jenks (2018). Bare nouns in their definite uses presuppose uniqueness due to a covert type shifter,  $t$ , that contributes a uniqueness presupposition. They therefore are predicted to be usable both in uniqueness situations and in anaphoric ones (since no blocking principle obtains).

While adopting Jenks (2018)'s semantic proposal does work, there is another, arguably more parsimonious analytical option available. In Jenks (2018)'s analysis, as in many works on bare nouns, the overt noun is consistently combined with either an existential or uniqueness type-shifting operator. Having eschewed MP-style competition in the synchronic grammar, however, a simpler option appears possible. In this alternative, which we adopt here, we treat A'ingae bare nouns as always having existential semantics (i.e., asserting existence), so bare nouns only assert the existence of an entity having the appropriate nominal property. Crucially, though, bare nouns in A'ingae lack any sort of anti-uniqueness/novelty condition, so nothing prevents the existential claim from being met by a previously mentioned or present entity (indeed this is often the most natural interpretation when available). Unique and anaphoric uses, then, are simply particular ways in which the existential claim may be true, and bare nouns in A'ingae on this view uniformly lack any presuppositions or covert operators can therefore occur in both indefinite and definite uses. We can further note that cases of donkey anaphora similarly are predicted to be fine assuming that the existential's contribution takes narrow scope (as independently seems to be the case). Finally, besides lacking any anti-familiarity or anti-uniqueness semantics, bare nouns also do not show any hard syntactic constraints for where they can be indefinite or definite.

This analysis of A'ingae bare nouns, where bare nouns have no presuppositions, is similar to the analysis given by Matthewson (1996) for Salish languages, where she argues that Salish determiners do not encode definiteness or specificity (see also Heim (2011) and Šimik and Demian (2020) for similar approaches to bare nouns). In Salish, the same determiner can be used in the reference to a novel or a familiar object, but this determiner cannot be considered as homophonous between the indefinite and definite forms. Salish languages also lack quantificational determiners that presuppose existence.<sup>7</sup>

---

<sup>7</sup> One crucial difference between Salish determiner system with the A'ingae one is that A'ingae has an additional indefinite marker, *fae*, that is only felicitous in indefinite uses. In order to capture

For *tsa*, on the other hand, the same analysis as above applies: it presupposes a referent that is unique and familiar in prior discourse, similar to Jenks'  $t^x$  rather than  $t$ . There are of course thorny details about how to capture the micro-variation in bridging data as we presented in Sect. 3.4, and we will leave such analysis of bridging definiteness cross-linguistically to future work.

We illustrate our analysis of A'ingae bare nouns and *tsa* with (39) as an example. This example shows that all of pronominal *tsa*, adnominal *tsa tevaenjen*, and bare noun *tevaenjen* are available in the anaphoric noun phrase in the second sentence.

(39) Chavangi fae tevaenjenma. {Tsa tevaenjen/Tsa/Tevaenjen} panshaen karu.

chava=ngi fae tevaenjen=ma. {**tsa tevaenjen/tsa/tevaenjen**} panshaen  
 buy=1 one book=ACC {ANA book/ANA/book} very  
 karu.  
 expensive

'I bought a book. The book was very expensive.'

In the bare noun case, we treat the bare noun phrase as not having any presuppositions about its referent "book," so the truth condition of the bare *tevaenjen* in the second sentence is met as long as a book exists that is expensive. This condition is met due to the first sentence, where the indefinite noun phrase *fae tevaenjen* asserts the book's existence, so using bare *tevaenjen* in the second sentence is felicitous. That is to say, with the bare noun, the coreference is not semantically specified, but rather is purely a matter of pragmatics. However, the pragmatic mechanism needed is nothing that is not already needed in any language to resolve which prior discourse referent an anaphoric form refers to.

In the case of the *tsa* noun phrases, *tsa* returns a unique entity that both satisfies the noun predicate "book" and is familiar. In this case, the existence presupposition is satisfied by the indefinite *fae tevaenjen*'s assertion of existence. The uniqueness presupposition is also valid because the previous sentence is focusing on one particular book. Lastly, the familiarity presupposition is also fulfilled as the book from the first sentence is the same book as the one in the second sentence.

It is interesting to note that this analysis of A'ingae bare nouns has as its core an existential quantifier, which resembles the analysis of indefinite noun phrases in English with the determiner "a." The crucial difference between English indefinite NPs and A'ingae bare nouns, however, is that English indefinite NPs have an additional anti-presuppositional feature that dictates that the use of an indefinite form presupposes the *non*-uniqueness of the referent. For A'ingae bare nouns, such anti-presupposition does not exist.

---

the more limited distribution of *fae*, then, we would need to posit that unlike bare nouns, it has a lexically specified constraint of anti-uniqueness and anti-familiarity. We leave it to future work to flesh out such an account in detail, but note here that regardless of the approach to definiteness, this appears necessary since the competing overt form, *tsa*, only has anaphoric uses, and therefore, reasoning based on its non-use would derive too weak of an anti-presupposition.

While there remain many important details to work out formally for such an account, our goal here is to give a general sense of what a semantic account could look like, having argued against accounts based on pragmatic blocking/hard competition. Whereas pragmatic accounts predict universal constraints on cross-linguistic variation, we have seen that A'ingae (and other languages) does not fit these proposed universals. The semantic account, on the other hand, is more permissive, allowing enough flexibility to capture the attested patterns with soft pragmatic competition providing a diachronic motivation for certain kinds of systems to be more common than others.

## 6 Conclusions and Future Directions

In this chapter, we have examined the expression of indefiniteness and definiteness in A'ingae. Empirically, we have shown that A'ingae bare nouns can be freely used in indefinite, unique, and anaphoric uses. Despite this, A'ingae has dedicated indefinite and anaphoric definite morphemes, *fae* and *tsa*, respectively.

While the recent literature has uncovered a range of variations in the expression of definiteness cross-linguistically, the A'ingae pattern is striking in particular for the lack of complementarity between forms. Schwarz (2013), Jenks (2018), and other recent works have established the existence of dedicated anaphoric determiners not unlike *tsa*.<sup>8</sup> However, the uniqueness forms in these cases often cannot be used in anaphoric cases.

Previous authors have often proposed to account for the apparent complementarity found in other languages through a sort of “hard” pragmatic competition: Chierchia (1998) and Dayal (2004)’s Blocking Principle, Heim (1991)’s *Maximize Presupposition!* and Jenks (2018)’s instantiation of it as *Index!*, and Ahn (2019)’s *Bare Noun Blocking*. While the details of these analyses differ, we have argued that A'ingae (along with other languages cited in Sect. 3.5) violates the predictions of such accounts. Moreover, as they are based on putatively universal pragmatic principles, it seems unlikely that their presence or absence would itself be a matter of grammar.

Instead, we have suggested an analysis in which “hard” pragmatic competition does not play a synchronic role in the expression of (in)definiteness. “Soft” pragmatic competition based on similar principles may of course exert pressure diachronically toward determiner systems that display complementarity. However, hard constraints on where various determiners and other DP forms are synchronically due to the semantics, not pragmatic competition. Finally, we have briefly

---

<sup>8</sup> Though to reiterate, we have also found two important differences between *tsa* and the other such markers. First, *tsa* is never felicitous in bridging uses regardless of the specific bridging relation, whereas previously described anaphoric forms are often usable in a subset of bridging environments. Second, *tsa* lacks the deictic/exophoric uses that are possible with many similar forms (e.g., Jenks (2018) work on Mandarin).

sketched what such a semantic account would look for bare nouns, indefinite *fae*, and anaphoric *tsa* in A'ingae.

Finally, we have argued that the lack of complementarity of (in)definiteness in A'ingae calls into question the role of *Maximize Presupposition*-like reasoning in this domain. In particular, we have argued that hard constraints on the use of specific forms in this domain are not the result of pragmatic competition, but rather semantics. Parallel arguments have been made in some other putative MP cases such as Bade (2016), who argues that apparently obligatory additive particles such as English *too* are also not due to hard MP-based hard competition but other sources. While we leave a full consideration of the nature and scope of MP to future work (see also Bade (2021) and references therein), we hope to have shown that there is good reason to doubt that competition between definite expressions should be taken as an instance of general MP pragmatics rather than cross-linguistically variable semantics.

**Acknowledgments** First of all, our heartfelt thanks to the A'i who have welcomed us into their community and shared their language with us. Thanks especially to our A'ingae-speaking consultants—Shen Aguinda, Hugo Lucitante, and Raúl Quieta—who have graciously spent their time thinking carefully about the data and ideas discussed here. We also gratefully acknowledge the support of the NSF DEL/DLI #BCS-1911348/1911428 “Collaborative Research: Perspective Taking and Reported Speech in an Evidentially Rich Language” to Scott AnderBois and Wilson Silva. Finally, thanks to Carol-Rose Little, Peter Jenks, and audiences at Brown and WSCLA 25 and of course to the editors of this volume, Andrés Saab and Cilene Rodrigues.

## References

- Ahn, Dorothy. 2019. THAT thesis: A competition mechanism for anaphoric expressions. Ph.D. thesis, Harvard University.
- A'ingae Language Documentation Project. 2022. Proyecto de documentación del idioma A'ingae. <https://cofan-aldp.github.io/LingView/#/index>. Accessed 25 July 2022.
- AnderBois, Scott, and Wilson de Lima Silva. 2017. Kofán Collaborative Project: collection of audio-video materials and texts. In *Endangered languages archive*. <https://www.elararchive.org/dk0466/>.
- Bade, Nadine. 2016. Obligatory presupposition triggers in discourse—Empirical foundations of the theories Maximize Presupposition and obligatory implicatures. Ph.D. thesis, Tübingen.
- Bade, Nadine. 2021. On the scope and nature of Maximise Presupposition. *Language and Linguistics Compass* 15(6): 1–16.
- Bremmers, David, Jianan Liu, Martijn van der Klis, and Bert Le Bruyn. 2022. Translation mining: Definiteness across languages (a reply to Jenks 2018). *Linguistic Inquiry* 53 (4): 735–752. [https://doi.org/10.1162/ling\\_a\\_00423](https://doi.org/10.1162/ling_a_00423).
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6: 339–405.
- Clark, Herbert H. 1975. Bridging. In *Theoretical issues in natural language processing*, ed. R. C. Schank and B. L. Nash-Webber. New York: Association for Computing Machinery.
- Dayal, Veneta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27: 393–450.
- Dayal, Veneta, and Li Julie Jiang. 2021. The puzzle of anaphoric bare nouns in mandarin: a counterpoint to *Index!*. *Linguistic Inquiry*: 1–42.

- Daḅkowski, Maksymilian. 2019. The Morphophonology of A'ingae Verbal Stress. Bachelor's Thesis, Brown University.
- Daḅkowski, Maksymilian. 2021. A'ingae (Ecuador and Colombia) – language snapshot. In *Language documentation and description*, Vol. 20. ed. Peter K. Austin, 1–12.
- Fischer, Rafael and Kees Hengeveld. 2023. A'ingae (Cofán). In *Volume 1 language isolates I: Aikanā to Kandozi-Shapra: An international handbook*, ed. Patience Epps and Lev Michael, 65–124. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110419405-002>
- Hawkins, John A. 1978. *Definiteness and indefiniteness*. London: Croom Helm.
- Heim, Irene. 1991. Artikel und definitheit. In *Semantics: An international handbook of contemporary research*, ed. Arnim von Stechow and Dieter Wunderlich. Berlin: de Gruyter.
- Heim, Irene. 2011. Definiteness and indefiniteness. In *Semantics: An international handbook of natural language meaning*, Vol. 1, ed. Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Berlin: Walter de Gruyter.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49: 501–536.
- Little, Carol Roseman. 2020. Mutual dependencies of nominal and clausal syntax in Ch'ol. Ph.D. thesis, Cornell University.
- Matthewson, Lisa. 1996. Determiner systems and quantificational strategies: Evidence from Salish. Ph.D. thesis, The University of British Columbia.
- Mendoza, Eduardo López. 2021. Marcación de la definitud en Tsotsil Sureño. Presentation at FAMLi 6 meeting.
- Moroney, Mary. 2021. Updating the typology of definiteness: Evidence from bare nouns in Shan. Glossa: A journal of general. *Linguistics* 6 (1): 56. <https://doi.org/10.5334/gjgl.1221>.
- Morvillo, Sabrina, and Scott AnderBois. 2022. The inner workings of contrast: decomposing A'ingae tsa'ma. In *Proceedings of SULA 11*.
- Pride, Kalinda, Nicholas Tomlin, and Scott AnderBois. 2020. LingView: a web interface for viewing FLEx and ELAN files. *Language Documentation & Conservation* 14: 87–107. <http://hdl.handle.net/10125/24916>.
- Pride, Ellen F. (1981) Toward a taxonomy of given/new information. In *Syntax and semantics: Vol. 14. Radical pragmatics*. ed. P. Cole, 223–255. New York: Academic Press.
- Repetti-Ludlow, Chiara, Haoru Zhang, Hugo Lucitante, Scott AnderBois, and Chelsea Sanker. 2020. A'ingae (Cofán). *Journal of the International Phonetic Association* 50 (3): 431–444.
- Schwarz, Florian. 2009. Two types of definites in natural language. Ph.D. thesis, University of Massachusetts, Amherst.
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistic Compass* 7/10: 534–559.
- Šimfík, Radek, and Christoph Demian. 2020. Definiteness, uniqueness, and maximality in languages with and without articles. *Journal of Semantics* 37: 311–366.
- Snider, Todd Nathaniel. 2017. Anaphoric reference to propositions. Ph.D. thesis, Cornell University.
- Vásquez Martínez, Norma Leticia. 2020. Definitud e indefinitud en el zapoteco de San Pedro Mixtepec, unpublished Ms. CIESAS.
- Vásquez Martínez, Morelia, and Carol-Rose Little (2020) Dimensions of definiteness in Ch'ol: A dialectal comparison. Presentation at the 2020 SSILA Annual Meeting.
- Zamuel, Arrieta, and Ana Laura. 2020. *La expresión de la cardinalidad definida en español y en zapoteco de San Pedro Güilá*. Master's thesis, UNAM.
- Zheng, Yuqi Holly. 2022. Anaphoric Expressions in A'ingae. Bachelor's Thesis, Brown University.

# Index

## A

Acceptability variables, 7, 107–128  
Affix order, 315, 319, 320, 323, 341–342  
Agglutinating languages, 289  
Agree, 1, 4, 7, 8, 65, 66, 71, 86, 88, 93, 94, 97–99, 101, 102, 173, 187, 235, 250, 272, 287, 318  
Amazonian languages, 13  
A-movement, 173, 183, 186–188, 192, 196  
Anaphora, 100, 356, 364, 367  
Applicative, 16, 269, 312–315, 322, 327, 331–342  
Argentinean Spanish, 11, 12, 133–167  
Argentinian Sign Language (LSA), 25–53  
Argument structure, 6, 15–17, 79–102, 111, 303, 311–342  
Atomicity, 10, 11, 235, 236, 245–251

## B

Bare nominals, 10, 354, 367  
Bare nouns, 10, 16, 227, 239, 240, 348, 351–355, 357–360, 362–370  
Bimodal bilingualism, 107–112  
Blending languages, 107, 110, 111, 127, 128  
Brazilian Portuguese, 2, 3, 6, 8–11, 108, 171–198, 226–228, 235–251, 264  
Brazilian Sign Language, 3, 6, 57

## C

Causative, 16, 266–268, 300, 302, 307, 311–319, 321–324, 327, 336, 341  
Chilean Spanish, 11, 203–230  
Classifiers, 6, 26, 68, 79, 109, 236, 308, 350

Code-blending, 8, 107–113, 124, 128  
Code-switching, 112, 126, 128  
Cofan, 16, 349–351  
Co-localization, 7, 61, 67–70  
Compounds, 15, 49, 257, 258, 261, 262, 270–282  
Concessive marker, 146–155, 159, 165, 166

## D

Definiteness, 16, 17, 347–370  
Determiner phrases (DP), 10, 65, 66, 119, 139, 173, 177–184, 186–188, 198, 210, 236, 248, 299, 318, 320, 321, 323, 324, 349, 369  
Dialect distance, 206, 207  
Direct objects, 288, 290, 295, 298–300, 307, 316, 321, 323, 330, 331, 350

## E

Endocentric, 279, 281  
Epenthesis, 6, 32, 36, 38, 49, 53  
Epistemic marker, 11, 141–146, 159, 163, 166  
Exocentric, 258, 279–281

## G

Grammaticalization, 11, 12, 133–167, 305, 308–309, 366

## H

Handshape insertion, 27, 32–36, 41, 46, 49, 51, 53



**I**

Indigenous Languages, 2, 14, 15, 311  
 Inherent Case, 10, 179–192, 195–198

**J**

Judgment task, 8, 108, 112–114, 116, 121,  
 126–128, 206, 229, 251, 365

**K**

Krenak, 15, 258–264, 268–272, 275–281

**L**

Languages of South America, 1, 2, 17  
 Language synthesis, 110  
 Lengua de Señas Peruana, 6, 7, 79–102  
 Libras (Brazilian Sign Language), 2, 6–8,  
 57–75, 107–128  
 LSP, 6, 7, 79–102

**M**

Maxakalí, 15, 258–263, 265–268, 270, 271,  
 273, 275–281  
 Maximize Presupposition, 16, 348, 360–364,  
 366, 367, 370  
 Morphology, 3, 9, 16, 17, 58, 60, 71, 74, 94,  
 112, 182, 195, 196, 206, 208, 209, 215,  
 226, 235, 238, 250, 251, 260, 267, 268,  
 270, 278, 299, 311–342, 349

**N**

Nominal plurality, 6, 25–53  
 Noun incorporation, 15, 259, 278, 287, 288,  
 291, 306–309  
 Null Subject Parameter, 11, 203, 207, 208, 229

**P**

Peruvian Sign Language, 6, 7, 79–102  
 Polysemy, 245, 247  
 Portuguese Codas (Children of deaf adults), 8,  
 107, 115  
 Possessor raising, 171, 194, 196, 198  
 Pragmatic competition, 16, 17, 347–370

Prefixation, 15, 16, 285–309  
 Pro-drop, 11, 12, 173, 177, 198, 203–230

**Q**

Quechua, 3, 13, 15–17, 311–342

**R**

Reflexive, 16, 133, 189, 311–313, 315, 318,  
 319, 323, 324, 327–331, 338–341, 365

**S**

Semantic variation, 236  
 Semi-lexicality, 155–167  
 Sign Languages, 2–8, 25, 57, 79, 111  
 Spanish syntax, 206  
 Synthesis model, 8, 109, 110, 112, 126, 128

**T**

Topic-subject constructions, 9, 171–198  
 Transitivity, 119, 289, 299–302

**U**

Uniqueness, 17, 347, 348, 353, 357, 359,  
 361–363, 366–369

**V**

Valency, 16, 298, 311, 312, 314  
 Variation, 6, 7, 14, 17, 53, 57, 80, 97, 134, 135,  
 141, 155–167, 193, 197, 207, 229, 236,  
 241–251, 264, 273, 288, 313, 315, 318,  
 321, 335, 339, 363, 366, 368, 369  
 Verb agreement, 7, 57, 60, 61, 66–68, 72, 73,  
 119  
 Voice, 8, 15, 285, 286, 288, 290, 294, 308, 309,  
 314, 316–320, 324, 325, 328, 329, 331,  
 339, 341, 342

**X**

Xavante, 15, 258–261, 264–266, 269–270,  
 273–278, 280, 281  
 Xerente, 15, 258, 260, 261, 264–265, 269–270,  
 273–278, 280, 281