The Native Languages of South America

In South America indigenous languages are extremely diverse. There are over 100 language families in this region alone. Contributors from around the world explore the history and structure of these languages, combining insights from archaeology and genetics with innovative linguistic analysis. The book aims to uncover regional patterns and potential deeper genealogical relations between the languages.

Based on a large-scale database of features from sixty languages, the book analyzes major language families such as Tupian and Arawakan, as well as the Quechua/Aymara complex in the Andes, the Isthmo-Colombian region, and the Andean foothills. It explores the effects of historical change in different grammatical systems and fills gaps in The World Atlas of Language Structures (WALS) database, where South American languages are under-represented.

An important resource for students and researchers interested in linguistics, anthropology, and language evolution.

LORETTA O’CONNOR is a post-doctoral researcher in the South American languages research group of the Traces of Contact project, Radboud University Nijmegen.

PIETER MUYSKEN is Academy Professor of Linguistics at Radboud University Nijmegen.
The Native Languages of South America

Origins, Development, Typology

Edited by

Loretta O'Connor and Pieter Muysken
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of figures</td>
<td>vii</td>
</tr>
<tr>
<td>List of maps</td>
<td>ix</td>
</tr>
<tr>
<td>List of tables</td>
<td>x</td>
</tr>
<tr>
<td>List of contributors</td>
<td>xii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xiii</td>
</tr>
<tr>
<td>1 Introduction: South American indigenous languages;</td>
<td>1</td>
</tr>
<tr>
<td>genealogy, typology, contacts</td>
<td></td>
</tr>
<tr>
<td>PIETER MUYSKEN AND LORETTA O’CONNOR</td>
<td></td>
</tr>
<tr>
<td><strong>Part I</strong> Introduction to South America</td>
<td></td>
</tr>
<tr>
<td>2 Human migrations, dispersals, and contacts in South America</td>
<td>29</td>
</tr>
<tr>
<td>LORETTA O’CONNOR AND VISHNUPRIYA KOLIPAKAM</td>
<td></td>
</tr>
<tr>
<td>3 Basic vocabulary comparison in South American languages</td>
<td>56</td>
</tr>
<tr>
<td>HARALD HAMMARSTRÖM</td>
<td></td>
</tr>
<tr>
<td><strong>Part II</strong> Case studies in contact</td>
<td></td>
</tr>
<tr>
<td>4 Structural features and language contact in the Isthmo-Colombian area</td>
<td>73</td>
</tr>
<tr>
<td>LORETTA O’CONNOR</td>
<td></td>
</tr>
<tr>
<td>5 The Andean foothills and adjacent Amazonian fringe</td>
<td>102</td>
</tr>
<tr>
<td>RIK VAN GIJN</td>
<td></td>
</tr>
<tr>
<td>6 The Andean matrix</td>
<td>126</td>
</tr>
<tr>
<td>SIMON VAN DE KERKE AND PIETER MUYSKEN</td>
<td></td>
</tr>
<tr>
<td>7 The Arawakan matrix</td>
<td>152</td>
</tr>
<tr>
<td>LOVE ERIKSEN AND SWINTHA DANIELSEN</td>
<td></td>
</tr>
</tbody>
</table>
vi Contents

8 The Tupian expansion
   love eriksen and ana vilacy galucio 177

Part III Comparative perspectives on linguistic structures

9 Language internal and external factors in the development of the desiderative in South American indigenous languages
   neele müller 203

10 Verbal argument marking patterns in South American languages
   joshua birchall 223

11 The Noun Phrase: focus on demonstratives, redrawing the semantic map
   olga krasnoukhova 250

12 Subordination strategies in South America: nominalization
   rik van gijn 274

Part IV Major findings and conclusions

13 The languages of South America: deep families, areal relationships, and language contact
   pieter muysken, harald hammarström, joshua birchall, swintha danielsen, love eriksen, ana vilacy galucio, rik van gijn, simon van de kerke, vishnupraya kolipakam, olga krasnoukhova, neele müller, and loretta o’connor 299

References 323
Subject and place index 366
Language index 371
Author index 376
Figures

4.2 Linguistic distances for Chibchan vs. non-Chibchan languages  
4.3 Linguistic distances for pre-defined areal groups  
4.4 Geographic distribution of feature values  
5.1 NeighborNet representation of the distances between the languages of the sample  
5.2 NeighborNet of distances between languages of the sample (phonological features only)  
5.3 NeighborNet of distances between languages of the sample (morphosyntactic features only)  
5.4 Correlation between linguistic distance and geographic distance  
5.5 Correlation between geographic elevation and linguistic distance  
5.6 Correlation between elevation and proximity to the Andean profile  
6.1 NeighborNet representation for the relative distances of the members of the Quechuan language family  
6.2 The distribution of languages per region, over time (Q = Quechua)  
6.3 NeighborNet representation for the relative distances of the different Andean languages discussed in this chapter  
7.1 The distribution of the personal paradigms in Arawakan languages  
7.2 Minimum spanning network of the Arawakan language family (also taken from the NeighborNet algorithm, Huson and Bryant 2006)  
7.3 Structural analysis of thirty-one Arawakan languages  
8.1 The branches of the Tupian language family  
8.2 NeighborNet representation of lexical distances among Tupian languages
List of figures

8.3 NeighborNet representation of structural distances among Tupian languages 186
11.1 Scale of semantic features encoded by demonstratives 269
12.1 NeighborNet of nominalizations as subordination strategies in the languages of the sample 283
13.1 Parsimony reconstruction for alignment in Tupian (Birchall 2014, based on the tree typology of Walker et al. 2012) 307
13.2 Correlation between Hamming distances and geographic distances for the pairs of families in the sample 312
Maps

1.1 South American language families, extensions from 1500 CE  
2.1 Archaeological sites from the Late Pleistocene–Early Holocene in South America  
4.1 The Isthmo-Colombian area, noting the position of languages in this study  
5.1 The greatest extent of the Quechuan, Aymaran, Panoan, Tupian, and Arawakan expansions  
5.2 The languages in the sample and their geographic distribution  
6.1 Approximate distribution of the indigenous languages in the Andes in the mid twentieth century  
7.1 The reconstructed geographic dispersal of the Arawakan and Tupian language families at the time of European contact. For complete references, see Eriksen 2011: 12  
8.1 The location of Tupí-speaking groups at the time of European contact  
8.2 Current location of the languages in our structural sample  
9.1 Desiderative/no desiderative marking in eighty-five South American languages  
10.1 Languages in the sample, with regions used in this chapter  
11.1 Geographic distribution of semantic features encoded by demonstratives  
12.1 The use of participant nominalization as a relativization strategy  
12.2 The encoding of notional subjects as possessors in subordinate clauses  
12.3 The encoding of notional objects as possessors in subordinate clauses  
12.4 The use of case marking to form adverbial clauses
# Tables

1.1 Schematic brief overview of some of the current families  
1.2 Preliminary overview of typological features mentioned in the literature  
1.3 Highly ranked stable features in the meta-analysis of Dediu and Cysouw (2013)  
3.1 Language families with their status in the work of Loukotka (1968), ASJP, and Campbell (2012a)  
4.1 Languages in this study  
4.2 Typological profiles, using features from Constenla (1991)  
4.3 Numbers of features in major categories and their intersections  
4.4 Ranges of features and types ranked by normalized median distance (\(N = 162\))  
4.5 Feature table  
5.1 The languages in the sample and their sources  
5.2 Areal studies of the Amazon and Andean regions used in this study  
5.3 The phonological features  
5.4 The morphosyntactic features  
5.5 The constituent order features  
5.6 The lexicon features  
5.7 The average linguistic distance per river system  
6.1 Languages of the study, compared to the varieties used by Heggarty (2005) and by Adelaar with Muysken (2004)  
6.2 Morphosyntactic and phonological features that distinguish Southern Peruvian Quechua from Ecuadorian Quichua  
6.3 Putative historical development of language use in Kallawayas' villages in the Charazani region  
6.4 The forms reconstructed for Proto-Quechua by Parker (1969)  
6.5 Features in more than twelve of the seventeen Quechuan varieties in our database, as compared to their occurrence in Aymaran and other Andean languages  
6.6 Quechua and Andean feature questionnaire
List of tables

7.1 The Arawakan language family 165
9.1 Desiderative markers in three Nambikwaran languages 213
9.2 Desiderative markers in seven Cariban languages 215
9.3 Desiderative markers in ten Tupian languages 217
9.4 Desiderative marking in eighty-five SA languages 220
10.1 Comparison of characteristic features for Amazon vs. Eastern and Andes vs. Western 244
10.2 Language sample with additional information (by region) 247
11.1 Distribution of semantic features in the sample 267
11.2 Language sample (ordered by language family) 272
12.1 Semantic relations considered for subordination strategies 276
12.2 Questions on nominalization 278
12.3 Comparison of global and South American distributions of nominalized structures 281
12.4 Comparison of distribution of nominalized structures per semantic relation type 282
12.5 Overlap of semantic relation types 284
12.6 Non-core case markers and adpositions used to form adverbial relations 294
13.1 WALS features for which South American languages show a significantly distinct profile as a group 306
13.2 Comparing the top 200 language family pairs in the sample 310
13.3 Rank order correlation between the language pairs for TAME, SubOrd, ArgMar, and NP 313
13.4 The variance in the four domains 314
13.5 Average stability of our different sets of features for the language families 314
13.6 Language pairs ranked highest on the domains of NP, TAME, SubOrd, and ArgMar (languages from the western region bold, Tupian-Cariban-Macro-Jê languages italic) 315
13.7 Some well-known expansion varieties in South America 318
Contributors

Joshua Birchall, PhD student, Radboud University Nijmegen, Netherlands; postdoctoral researcher, Goeldi Museum, Belem, Brazil

Swintha Danielsen, postdoctoral researcher, University of Leipzig, Germany

Love Eriksen, postdoctoral researcher, Lund University, Sweden

Ana Vilacy Galucio, senior researcher, Goeldi Museum, Belem, Brazil

Rik van Gijn, postdoctoral researcher, Radboud University Nijmegen, Netherlands, and University of Zurich, Switzerland

Harald Hammarström, postdoctoral researcher, Radboud University Nijmegen and Max Planck Institute for Psycholinguistics, Netherlands

Simon van de Kerke, lecturer, Leiden University, Netherlands

Vishnupriya Kolipakam, PhD student, Max Planck Institute for Psycholinguistics, Netherlands

Olga Krasnoukhova, PhD student, Radboud University Nijmegen, Netherlands

Neele Müller, PhD student, Radboud University Nijmegen, Netherlands; postdoctoral researcher, University of Marburg, Germany

Pieter Muysken, Professor, Radboud University Nijmegen, Netherlands

Loretta O’Connor, postdoctoral researcher, Radboud University Nijmegen, Netherlands
We would like to acknowledge the support from various institutions and individuals that helped us in putting together this volume.

The research for this book was funded by the European Research Council Advanced Grant 230310 “Traces of Contact” awarded to Pieter Muysken, the Veni Grant to Rik van Gijn from the Netherlands Organization for Research (NWO), as well as the Academy Chair awarded to Pieter Muysken by the Royal Netherlands Academy of Sciences (KNAW).

We would also like to acknowledge the advice and support of a number of colleagues working on South American indigenous languages and on phylogenetic methods: Willem Adelaar, Mily Crevels, Dan Dediu, Michael Dunn, Fiona Jordan, Sergio Meira, Lev Michael, Annemarie Verkerk, and Hein van der Voort. Needless to say, none of these people can be held accountable for the claims made in this volume.

Our student assistant Ellen van den Broek of Radboud University Nijmegen compiled and checked the references and helped prepare the manuscript for publication. All maps but one were produced by Thijs Hermsen, using ArcGIS 9.2, and we thank Love Eriksen of Lund University for providing the polygon shape file used in four of the maps.