

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)

1 Introduction: South American indigenous languages; genealogy, typology, contacts

Pieter Muysken and Loretta O'Connor

This chapter discusses general issues concerning the relationships among the indigenous languages of South America and sketches the background for the rest of the volume. Why are there so many language families, and why so many isolates? What is the distribution of both larger families and isolates? Given the apparent genealogical diversity, why are there so many shared specific areal typological patterns, some characterizing most of the continent, and some, individual parts? Do shared patterns reflect older historical genealogical links, or are they the result of convergence? What can we learn about these issues from the perspective of language history (vertical transmission) and language contact (horizontal transmission)? We also describe and justify our research methodology and briefly outline the chapters in the volume.

1 Solving an intellectual puzzle

The relationships among the indigenous languages of South America pose complex intellectual problems that invite innovative research methodologies. Until recently these languages were relatively unknown. Not only were the vast majority not at all or poorly documented, but the complex relations among them also remained obscure. Our knowledge base was woefully inadequate with respect to classification, history, and typology. In recent years, our knowledge of these languages has grown enormously, but the puzzles remain.

With respect to *language classification*, using the comparative method there is a current consensus for some 108 separate language families on the continent, half of which are isolate languages. This total represents a large part of the overall inventory of language families of the world (420, according to Campbell (2012a: 59).

This introduction is based in part on material presented at the conference on methods in historical linguistics at Belém in September 2005, and a further developed version at the International Conference on Historical Linguistics (ICHL XIX) at the Radboud University Nijmegen in August 2009. We are grateful to the participants at these meetings, especially to Willem Adelaar, Rodolfo Cerrón-Palomino, Swintha Danielsen, Michael Dunn, Stephen Levinson, and Hein van der Voort for comments.

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)

2 *Pieter Muysken and Loretta O'Connor*

With respect to *language typology*, there is less diversity. While certainly there are considerable differences between the languages, there are many similarities as well, as shown below.

Our puzzle falls into three parts:

- (A) Why are there so many language families, and why so many isolates? What is the distribution of both larger families and isolates? Given the fact that South America was the most recently populated of all continents (15–13,000 BP, see O'Connor and Kolipakam, this volume), the genealogical diversity is surprising indeed.
- (B) Given the apparent genealogical diversity, why are there so many shared specific areal typological patterns, some characterizing most of the continent, and some, individual parts? Do shared patterns reflect older historical genealogical links, or are they the result of convergence? In either case, why do classical comparative methodologies based on lexical data not yield better results for language classification?
- (C) What can we learn about the relation between the issues in (A) and (B) from the perspective of language history (vertical transmission) and language contact (horizontal transmission)?

In our estimation, moving forward will require an extension of the historical analysis of South American languages to language typology and language contact. Auspiciously, the continent offers a unique opportunity, in that the relatively recent moment of its settlement roughly coincides with the upper limit of the time depth that can be reasonably investigated through language comparison.

2 Language relationships in South America

2.1 *The field of historical linguistics and the comparative method*

There is a long tradition of historical and comparative linguistics in South America, starting with Jesuit scholars such as Hervás (1800–1805), who identified Tupí-Guarani, and Gilij (1780–1784), who established the Arawakan and Cariban language families as a unit, two among nine *lenguas matrices* ('mother languages' or 'base languages') in Venezuela. Following these pioneers, there have been a large number of attempts at classification of the languages of South America and proposals for family groupings, from modest to grandiose.

Uhle (1890) identified the Chibchan family, and Davis (1968) argued that some more distant clusters together formed Macro-Jê. During the past century, numerous scholars have explored genealogical relationships and proposed different classifications, including Loukotka (1968), whose work is commented on in Hammarström (this volume). Swadesh (1959, 1962) proposed four large 'networks': MACRO-MAYAN, MACRO-CARIB, MACRO-ARAWAKAN,

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)

Introduction: genealogy, typology, contacts

3

MACRO-QUECHUAN. Kaufman (1990) is fairly conservative, arriving at 118 groups. This number is further specified in Campbell's report (2012a: 59) as 108 genealogical units, consisting of 53 families with more than one member and 55 one-member families, the so-called isolates.

Writing a history of the classificatory efforts would require a separate paper. Adelaar with Muysken (2004) provide a concise summary, and a detailed treatment is also given in Campbell (1997, 2012a) and Adelaar (2012a). Hammarström (2010: Appendix) provides an exhaustive meta-analysis of the published sources, returning to this issue in this volume.

Efforts to investigate existing language families and reconstruct hypothetical earlier forms continue, as in the work on Tupian, carried out in Brasília (Rodrigues 1999; Rodrigues and Cabral 2012) and at the Goeldi Museum in Belem, and the work on Nambikwara at the Free University Amsterdam (Telles and Wezels 2009), coupled with efforts to relate Tupian, Cariban, and Jêan as TuCaJê (Rodrigues 1985). There are also efforts to link individual unclassified languages to well-established language families, such as the recent linkage of Jabutí (Ribeiro and van der Voort 2010) and Chiquitano (Adelaar 2008b) to Macro-Jê, and smaller families have been proposed, such as Harakmbet and Katukina (Adelaar 2000). In Argentina, Viegas Barros (2005) has argued for the Chonan family.

However, the overall picture is not one of unification, and this raises issues about the universal applicability of the comparative method, argued by Sapir (1929: 208):

There can be no doubt that the methods first developed in the field of Indo-European linguistics are destined to play a consistently important role in the study of all other groups of languages, and that it is through them and through their gradual extension that we can hope to arrive at significant historical inferences as to the remoter relations between groups of languages that show few superficial signs of a common origin.

In contrast, scholars such as Thurston (1987) have argued that the traditional comparative method may not always be the right model, on two grounds: Wave and network models are sometimes better than tree models, and structural features may play a role next to lexical and morphological features. In this book we will stick to the middle ground and adopt a plurality of methods for studying the historical linguistics of South American languages, using both published materials based on the comparative method and other techniques that rely on structural features. We also systematically explore the possible role of language contact, which can obscure genealogical relationships.

Like other historical linguists, we depart from the Uniformitarian assumption, namely that processes of language change are uniform in different periods of human history. However, the precise implications of this need to be specified, since in the prehistory of the South American continent different conditions

held, which must have had a profound impact on processes of change and contact (see O'Connor and Kolipakam, this volume). In the early period, we are dealing with relatively isolated hunter-gatherer societies, with low population density, in probably intermittent contact with other groups. Later, after the intensification of agriculture, more complex exchange networks were built up between communities in areas with much higher population density. Finally, in the colonial period, the tremendous demographic decline and ethnic destruction following the Iberian invasion led to much lower indigenous population densities once again, the destruction of social and commercial networks, and the restructuring of communities. All these changes must be taken into account when we start modeling language change in the South American continent.

Comparative-historical linguistics in the South American context faces a number of problems. There are few reconstructed proto-languages for comparisons at the level of families, and the coverage and quality of documentation is very uneven for many languages. In addition, information about word frequency, needed to apply some lexicostatistic methods, is non-existent. Similarly, lexical semantic information is incomplete and highly skewed. However, through enormous documentation efforts currently underway, both lexical and grammatical information on a large range of languages has become available. The challenge is to bring this material to bear on the issue of language relationships.

2.2 *Greenberg's Amerind*

In an influential but controversial 1987 book, Joseph Greenberg assumed, based on mass comparisons of lexical material, that all the languages of South America belonged to a single language family dubbed Amerind. This hypothesis has met with little support among South American language specialists, even though it was accepted by outsiders, both linguists and others. Greenberg was criticized on four counts:

- The idea of a single overall family may not be incorrect, but it cannot be demonstrated empirically.
- The intermediate macro-families or stocks postulated have been criticized; alternative macro-families postulated such as the already mentioned Tupian, Cariban, and Jêan family cluster (TuCaJê) have more support.
- Many data used by Greenberg were inaccurate; the work is riddled with mistakes in data interpretation and reproduction.
- His method, mass lexical comparison, is superficial and cannot replace the traditional comparative method based on reconstructions.

Nonetheless, Greenberg's ideas loom large in the background, for several reasons. First, there are odd resemblances, some very specific, between languages assumed to be entirely unrelated. There are a number of "pan-Americanisms"

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)

Introduction: genealogy, typology, contacts

5

Table 1.1 *Schematic brief overview of some of the current families*

Type	Name	Location
<i>Larger families</i>	Arawakan	Widespread, from Belize and Honduras to Bolivia and Paraguay
	Tupian	Central Amazon, from Paraguay to French Guyana and Peru
	Cariban	Northern Amazon
	Macro-Jéan	Central and southern Amazon
	Chibchan	Central America, from Honduras through northwestern Colombia
	Tucanoan	Western Amazon and Pacific coast
	Pano-Takanan	Western Amazon
<i>Smaller families</i>	Chon	Southern Cone
	Barbacoan, Arawan, Chapacuran, Makúan, Nambikwaran, Witotoan	In the Andean foothills, central and southern Amazon basin
	<i>Clusters</i>	Aymaran, Quechuan
<i>Isolates and small groupings</i>		Many in the Amazonian fringe

(Kaufman 1990: 26; Campbell 1997: 257–259), even though they do not allow us to reliably create larger family units. Second, there is a sense of “Amerindian language type,” even if it is not very precise. Third, as our knowledge of the individual languages increases and more comparative work is done, the evidence for macro-families grows, even if these do not correspond to the ones postulated by Greenberg. Without undue optimism, we can predict that further links will be established in the coming years. Currently, the mass comparison method propagated by Greenberg has been replaced by automated similarity judgment on basic lexicon in the ASJP (Automatic Similarity Judgment Program; Holman et al. 2008); see Hammarström (this volume).

2.3 *Current distribution of the language families and isolates*

Table 1.1 presents some larger families in South America, as well as a number of smaller families and language isolates, and two entities we could label clusters, relatively shallow families with a wider spread. The categories are illustrated in Table 1.1.

The Andean Foothills and the Amazonian Fringe include foci of extreme linguistic diversity, such as the Vaupés region in the northwest and the Mamoré-Guaporé and Chaco regions in the southwest (see van Gijn, this volume). Some languages in these regions are Arawakan and Tupian, but many represent minor families and isolates. This fragmentation contrasts with the more homogeneous

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)6 *Pieter Muysken and Loretta O'Connor*

Map 1.1 South American language families, extensions from 1500 CE

central and eastern plains, where most languages belong to stocks like Macro-Jê, Arawakan, Cariban, and Tupí-Guarani. Map 1.1 gives an impression.

2.4 *Explanations for the current diversity*

Why is there such diversity of languages in the continent, concentrated mostly on the Andean fringe area? Consider five possible hypotheses:

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)

Introduction: genealogy, typology, contacts

7

- (a) Genealogical fragmentation is the original situation for all of South America. The ethnically and linguistically more homogeneous areas simply result from more recent expansions. Notice that we can date the expansion of some of the larger families fairly precisely. Aikhenvald (2012: 4) writes: “Amazonia and the Americas in general have a large number of recognizably distinct language families. This is compatible with the relatively late peopling of the continents. In areas of deeper antiquity of settlement such as Australia (estimated at 50,000 years), the time-depth promoted intensive convergence of languages towards a common prototype.” This explanation links the very diversity to the late date of settlement. However, the original population may not have been large enough to include the full number of 100+ families (plus ones that have already disappeared). We must assume that a substantial amount of the diversity is more recent than the moment of spread across the continent, and is subsequently shallower.
- (b) Fragmentation in the fringe zones arose because these areas functioned as zones of refuge, to which many smaller groups fled when pushed out of richer areas by stronger groups. Recent work on ethnogenesis (e.g. Hornborg 2005, Eriksen 2011) casts doubt on this, citing linguistic diversity as a possible outcome of contact and not necessarily of isolation. Certainly, isolation cannot have been the sole cause of linguistic fragmentation, as speakers of many language isolates were far from isolated. Rather, they participated in intensive trading networks. The areas along the fringe of the Amazon were a contact zone for lowland–highland exchanges.

Lexical diversity may go hand in hand with contact and indeed be a product of language contact. Bower et al. (2011) compared lexical borrowing in hunter gatherer populations across the globe and state that although “loan levels varied both within and among regions, they were generally low in all regions . . . , despite substantial demographic, ecological, and social variation. Amazonian levels were uniformly very low, with no language exhibiting more than 4%. Rates were low but more variable in the other two study regions, in part because of several outlier languages where rates of borrowing were especially high.” The authors interpret this result as showing “an association between language and group identity that is relatively strong compared to many other parts of the world, and pertains widely within Amazonia.”

- (c) The fragmentation along the eastern flanks of the Andes, the so-called *montañas*, is due to the fact that this was the oldest inhabited zone of South America, the path along which groups moved southward as they came through the Isthmus of Panama (Dahl et al. 2011). This is not likely to be the right scenario, given the time depth involved. A key problem is that the paleoarcheological evidence does not necessarily point to the *montañas* as the oldest inhabited zones (see O'Connor and Kolipakam, this volume, and van Gijn on the Andean foothills, this volume).

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)8 *Pieter Muysken and Loretta O'Connor*

- (d) Certain features of South American language systems may themselves contribute to the diversity puzzle. Nichols (1992) tentatively argues that in head-marking languages, language change tends to destroy the information needed for the reconstruction of deeper genealogical units, while in dependent-marking languages this information is more likely to be preserved over time. This argument merits further investigation.
- (e) Social factors also play a role (and see 4.1). Thurston (1987) argues that there are group dynamics that strengthen specific historical developments. These can be grouped under the general rubric of *exotericity* and *esoterogeny*. Exotericity is involved in intergroup relations and may lead to the indigenization of a variety by a particular group. In esoterogeny, language is treated as an internal emblem of group membership, leading to elaboration, the emergence of morphologically complex forms involving suppletion and irregularity, the creation of idioms, and new vocabulary. On the whole, the norm-enforcing dynamic of such a group process is considerable, and could lead to differentiation. A similar line of argumentation has been developed in Nettle (1999) and Trudgill (2011).

3 Linguistic typology and the areal distribution of features

South American languages tend to have a number of recurrent typological features that distinguish them from the languages of the Old World (Wichmann et al. 2010c). There are fewer differences with the languages of Central and North America, but this has been investigated even less. Table 1.2 gives a very preliminary overview of some of the features mentioned in the literature so far, organized in terms of the four major categories of elements studied in this volume.

Notice that the noun phrase and verbal argument marking have been more thoroughly explored than Tense-Aspect-Mood-Evidentiality (TAME). Similarly, certain geographic areas have been more thoroughly explored than others.

Typologically, the languages involved are very interesting, but information about patterns of typological markedness is skewed by earlier under-representations of South America in typological surveys. South American languages have been disproportionately absent in large language samples, in large part because few descriptive grammars of the type that typologists like to work with were available. The aforementioned classification by Greenberg (1987) of all South American languages into a single Amerind family did not help, as this justified the relative under-representation in the samples, on the basis of which some typologists have drawn statistical conclusions.

From the qualitative perspective, the types of unusual patterns found in the languages of South America were insufficiently known. The recognition of new typological properties of the languages of South America will change our

Table 1.2 *Preliminary overview of typological features mentioned in the literature*

Feature	Region or language	References
<i>Noun Phrase</i>		
A rich nominal determiner system, including nominal tense or aspect	Cariban, Tupian, Rondônia	Nordlinger and Sadler (2004); Muysken (2008c)
Positional deictics (standing, sitting, lying, (in)visible, etc.)	Southern Cone, Chaco, Rondônia	Kirtchuk (1996); Krasnoukhova (2012); Campbell and Grondona (2012)
Both nominal and verbal classifier systems	Some languages in Bolivia and Rondônia, southern Amazon in Brazil, western Amazon in Brazil and Colombia	Aikhenvald (2012); Crevels and van der Voort (2008); Payne, Doris L. (1987); Seifart and Payne (2007)
Inclusive/exclusive Genitive classifier for possessed domestic animals	Diverse spread Chaco	Crevels and Muysken (2005a) Campbell and Grondona (2012)
Gender in demonstratives and pronouns	Chaco	Campbell and Grondona (2012)
Lack of classifiers	Andes	Adelaar (2008a); Torero (2002); Adelaar (2012b)
Lack of nominal number	Guaporé-Mamoré	Crevels and van der Voort (2008)
<i>Argument marking and verbal morphology</i>		
Complex verbal morphology, (poly)synthesis and concatenative syntax-like morphology	Lowland South America	Payne, D. (1990); Crevels and van der Voort (2008)
A high incidence of prefixes	Guaporé-Mamoré	Crevels and van der Voort (2008)
Contrast between active/stative/inverse alignment	Chaco, Guaporé-Mamoré	Crevels and van der Voort (2008); Campbell and Grondona (2012)
Do-verbs	Northwestern South America	Jäger (2006)
Verb affixes marking direction, location, position, orientation; meaning extension to mark tense, aspect, mood	Western Amazon, Chaco, Guaporé-Mamoré	Payne, D. (1990); Crevels and van der Voort (2008); Campbell and Grondona (2012)
Sociative causatives	Tupian, Tacanan, Arawak, Panoan, Mosestén	Guillaume and Rose (2010); Aikhenvald (2012)
Serial verbs	Northern Amazon in Brazil, parts of Paraguay and adjacent Bolivia and Brazil	Aikhenvald (2012)
Verbal number	Guaporé-Mamoré	Crevels and van der Voort (2008)
<i>Tense-Aspect-Mood-Evidentiality (TAME)</i>		
Rich evidential systems	Rio Negro, Rondônia/Bolivia, south-central Brazil	(Aikhenvald 2004, 2012); Crevels and van der Voort (2008)
<i>Subordination</i>		
Switch-reference and head-tail linking	North central Andes and foothills: Quechua, Panoan, Jivaroan, Barbacoan, Tucanoan, and Uru-Chipaya	Adelaar (2008a)

Cambridge University Press

978-1-107-04428-9 - The Native Languages of South America: Origins,
Development, Typology

Edited by Loretta O'Connor and Pieter Muysken

Excerpt

[More information](#)10 *Pieter Muysken and Loretta O'Connor*

perspective considerably. The promising typological features for phonology include nasal spreading, nasal harmony, and various tonal patterns; for the lexicon, we find reduplication, ideophones and sound symbolism, noun incorporation (especially involving the distinction between noun incorporation and affixation), nominal classification systems, and both nominal and verbal classifiers (see Krasnoukhova, this volume, for properties of the noun phrase). In morphosyntax, features of interest include possession (especially on the types of nouns that may be considered inalienable and alienable) as well as number, positional deictics (e.g. sitting, standing, lying), directional markers on the verb (e.g. grammaticalized verbal morphology for 'upriver,' 'downriver,' etc.), reflexive and reciprocal relations, verb compounding, and serialization.

With respect to the variables focused on in this project, TAME has not been studied systematically, but among the striking features are the marking of tense on the noun in several families, complex evidentiality systems, and fine-grained tense marking (e.g. future beyond tonight) (see Müller, this volume). Argument realization in the Amazonian Fringe is highly complex and involves reference systems marked on the verb, differential object marking and other animacy effects (where object markers occur on nouns depending on their animacy), and active/stative/inverse alignment (see Birchall, this volume). The occurrence, precise properties, and distributions of these features across the Amazonian fringe still need to be explored. This line of research has already proven to be highly promising (cf. e.g. Grinevald and Seifart 2004, Crevels and van der Voort 2008). In the field of subordination, Everett (2005) has argued that Pirahã lacks recursion in the clausal domain, and hence true subordination.

To take but one example of the link between typology and classification, as far as is known, nominal tense or aspect is strongly rooted in the Cariban language family and may have spread from there to individual members of other families, such as Tariana and Chamicuro (Arawakan); Wari' (Chapacuran); Nambikwaran; Movima (isolate); Mosetén (Mosetenan); Cofán (isolate); Weenhayek' (Matacoan); as well as Guaraní, Sirionó, and Yuki (Tupian). However, it is also quite possible that it is an original feature of other language families, including Tupian. Further study of the precise geographic distribution of the features involved is urgently needed, also in the light of the possible grammaticalization of lexical suffixes referring to 'deceased' and 'future' (see Müller, this volume).

4 Language contact

The topic of contact between indigenous languages in South America is vast and almost intractable; a first general exploration is presented in Muysken (2012b). We still know little about the history of the languages of the continent,