

# THE AMAZONIAN LANGUAGES

*edited by*

R. M. W. DIXON

*and*

ALEXANDRA Y. AIKHENVALD

*Research Centre for Linguistic Typology*

*La Trobe University, Melbourne*



**CAMBRIDGE**  
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom <http://www.cup.ac.uk>  
40 West 20th Street, New York, NY 10011-4211, USA <http://www.cup.org>  
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1999

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1999

Printed in the United Kingdom at the University Press, Cambridge

Typeset in MT Times NR 9/13pt in QuarkXPress® [SE]

*A catalogue record for this book is available from the British Library*

*Library of Congress cataloguing in publication data*

Amazonian languages / edited by R. M. W. Dixon and Alexandra Y. Aikhenvald.

p. cm. – (Cambridge language surveys)

Includes bibliographical references.

ISBN 0 521 57021 2 (hardback)

1. Indians of South America – Amazon River Region – Languages.

I. Dixon, Robert M. W. II. Aikhenvald, A. Y. (Aleksandra

IUrevna) III. Series.

PM5099.8.A48 1999

498 – dc21 98-46270 CIP

ISBN 0 521 57021 2 hardback

# CONTENTS

---

<i>List of maps</i>	page xvii
<i>List of contributors</i>	xviii
<i>Acknowledgements</i>	xx
<i>List of abbreviations</i>	xxi
<i>Conventions followed</i>	xxiv
<b>1 Introduction</b> by R. M. W. Dixon and Alexandra Y. Aikhenvald	1
1 Cultural background	3
2 Linguistic diffusion	7
3 Genetic relationship	11
4 The punctuated equilibrium model	16
5 Organization of this book	19
Bibliography	20
<b>2 Carib</b> by Desmond C. Derbyshire	23
1 Introduction	23
2 Phonology	26
2.1 Segmental	26
2.2 Phonotactics and suprasegmentals	26
2.3 Morphophonology	28
3 Morphology, particles and pronouns	31
3.1 Inflectional morphology	31
3.1.1 Person-marking affixes on verbs, nouns, adverbials and postpositions	32
3.1.2 Tense, aspect, mode and number suffixes on verbs	37
3.1.3 Possession, tense and number suffixes on nouns	40
3.1.4 Inflectional suffixes on locative postpositions	42
3.2 Derivational morphology	43
3.2.1 Verb derivational affixes	44
3.2.2 Nominalizing affixes attached to verb stems to form nouns	45

3.2.3	Some adverbial derivations from noun and verb stems	50
3.3	The particle word class	53
3.4	Pronouns	53
4	Syntax	54
4.1	Main clause structuring	55
4.2	Subordinate clause constructions	56
4.2.1	Nominalizations	56
4.2.2	Adverbializations	57
4.2.3	Postpositions	58
4.3	Ergativity	60
	Bibliography	61
<b>3</b>	<b>The Arawak language family</b> by Alexandra Y. Aikhenvald	65
1	Introduction: the Arawak languages and their speakers	65
1.1	Comparative studies, genetic classification and subgrouping	73
2	Phonology	75
2.1	Segmental phonology	75
2.2	Syllable structure, suprasegmentals and phonotactics	78
2.3	Morphophonology	79
3	Typological profile	80
3.1	Word structure	80
3.2	Morphological processes	80
3.3	Word classes	81
4	Nominal morphology	82
4.1	Possession	82
4.2	Classifiers and genders	83
4.3	Number	84
5	Closed classes: personal pronouns, demonstratives, interrogatives, numbers, adpositions	85
6	Verbal morphology	85
6.1	Classification of verbs and predicate structure	86
6.2	Cross-referencing and split ergativity	87
6.3	Valency-changing derivations	90
6.4	Other verbal categories	93
6.5	Noun incorporation	95
6.6	Relative and negative markers	95
7	Syntax	96
7.1	Grammatical relations	96
7.2	Noun phrase structure	96
7.3	Complex predicates and serial verbs	97

<i>Contents</i>	ix
7.4 Constituent order	98
7.5 Non-verbal clauses	98
7.6 Complex sentences: relative clauses, complement clauses, coordination, pivot and switch-reference	99
7.7 Grammatical means used for marking discourse functions	101
Bibliography	102
<b>4 Tupí</b> by Aryon D. Rodrigues	107
1 Introduction	107
2 Phonology	110
2.1 Vowels	110
2.2 Consonants	112
2.3 Tone	114
3 Grammatical overview	114
4 Nouns	115
5 Pronouns	117
6 Verbs	118
7 Demonstratives	120
8 Questions	120
9 Subordinate clauses	121
10 Pivots	121
Bibliography	122
<b>5 Tupí-Guaraní</b> by Cheryl Jensen	125
1 Introduction	125
2 Identification of Tupí-Guaraní languages	128
2.1 Distinguishing characteristics of Tupí-Guaraní languages	128
2.2 Subgroups within Tupí-Guaraní	129
3 Proto-Tupí-Guaraní phonology	133
3.1 Proto-Tupí-Guaraní phonemes	133
3.2 Proto-Tupí-Guaraní allomorphs	134
3.3 Possible Tupí-Guaraní morphophonemic phenomena	135
3.3.1 Replacement by nasal consonants	135
3.3.2 Strategies for the loss of consonant clusters formed at morpheme juncture	136
4 Phonological changes within Tupí-Guaraní	137
4.1 Weakening of <i>*tʃ</i> and <i>*ts</i>	137
4.2 Palatalization	138
4.3 Labialization	140
4.4 Merger of <i>*β</i> with <i>*w</i>	141

4.5	Final-consonant phenomena	142
4.5.1	Devoicing	142
4.5.2	Nasalization	143
4.5.3	Loss	143
4.6	Vowel shift	144
5	Stems	146
5.1	Categories of stems	146
5.2	Stem classes	146
6	Person markers	146
7	Nouns	148
7.1	Case marking	148
7.2	Possession	150
7.3	Noun composition	151
7.4	Indication of number	151
7.5	Noun phrases	152
7.6	Possession classes	152
8	Postpositions	153
9	Verbs	154
9.1	Independent verbs	155
9.2	Oblique-topicalized verbs	156
9.3	Serial verb constructions	157
9.4	Temporal subordinate clause constructions	157
9.5	Valency-changing devices	158
9.5.1	Causatives	158
9.5.2	Detransitivizers	159
9.5.3	Object incorporation	159
10	Nominalizations	159
11	Semantics	160
	Bibliography	161
<b>6</b>	<b>Macro-Jê by Aryon D. Rodrigues</b>	<b>165</b>
1	Historical survey	165
2	Distribution	166
3	Linguistic scholarship	169
4	Phonology	171
4.1	Vocalic systems	171
4.2	Consonantal systems	174
4.3	Tone	180
5	Morphology	180
5.1	Inflection for contiguity of a determiner	180
5.2	Inflection for possession	182
5.3	Number	183

5.4	Noun classification	184
5.5	Agreement marking on the verb	185
6	Syntax	187
6.1	Constituent order in declarative sentences	187
6.2	Adpositional phrases	188
6.3	Genitive phrases	190
6.4	Demonstrative phrases	192
6.5	Numeral phrases	192
6.6	Adjectival phrases	193
6.7	Ergativity	193
6.8	Valency-changing processes	195
6.8.1	Reflexives and reciprocals	195
6.8.2	Causativization	195
6.9	Switch-reference	197
7	About the consistency of Macro-Jê as a genetic group	198
7.1	A brief appraisal of the grammatical affinity	198
7.2	Phonological equations	198
	Bibliography	201
7	<b>Tucano</b> by Janet Barnes	207
1	Introduction	207
2	Phonology	209
2.1	Syllable structure	210
2.2	Segmental phonology	210
2.3	Nasalization	211
2.4	Accent	212
3	Morphology	212
3.1	The verb	213
3.1.1	Evidentials	213
3.1.2	Aspect	214
3.1.3	Mood and modality	216
3.2	Pronouns	217
3.2.1	Personal pronouns	217
3.2.2	Possessive pronouns	218
3.3	Classifiers	218
3.4	Specificity marker	219
3.5	Nouns	221
3.6	Adjectives	221
3.7	Negation	222
3.8	Nominalized verbs	222
3.9	Switch-reference	223
4	Syntax	224

4.1	Time and location	224
4.2	Noun phrase	225
	Bibliography	225
<b>8</b>	<b>Pano</b> by Eugene E. Loos	227
1	Introduction	227
2	Phonology	230
2.1	Segmental phonology	230
2.2	Common phonological variations	231
3	Grammar	234
3.1	Word classes and noun phrase formation	234
3.1.1	Noun phrase	235
3.1.2	Pronoun classes	235
3.1.3	Relative clauses	236
3.1.4	Noun phrase pluralization	236
3.2	Transitivity concord	236
3.2.1	Switch-reference	237
3.2.2	Adverbial suffix concord	239
3.2.3	Locative phrases	239
3.2.4	Ergative marking	240
3.3	Noun incorporation	243
3.4	Verb formation	243
3.4.1	Verb roots	243
3.4.2	Suffixes	244
3.5	Mood indicators	248
3.6	Deictics	248
	Bibliography	249
<b>9</b>	<b>Makú</b> by Silvana and Valteir Martins	251
1	Introduction	251
2	Genetic classification	253
3	Phonology	255
4	Word structure	257
5	Word classes	257
6	Nominal categories: possession, classifiers, number	258
7	Verb structure	259
8	Valency-changing derivations	260
9	Incorporation	261
10	Grammatical relations	263
11	Negation	264
12	Syntax	265
13	Lexicon	265
	Bibliography	266

<b>10 Nambiquara</b> by Ivan Lowe	269
1 Introduction	269
2 Phonology	271
2.1 The variants of phonemes	272
2.2 Syllable structure	272
2.3 Stress	272
2.4 Morphophonemics	272
3 Morphology	273
3.1 Verbs	274
3.1.1 Main verbs	274
3.1.2 Subordinate verbs	277
3.2 Adjectives	279
3.3 Nouns	280
3.4 Pronouns	283
3.5 Adverbs	283
3.6 Interjections	284
3.7 Ideophones	284
4 Syntax	284
4.1 Main clauses	284
4.2 Subordinate clauses	287
4.3 Clause coordination	288
4.4 Noun phrases	289
4.5 Nominalizations	290
Bibliography	291
<b>11 Arawá</b> by R. M. W. Dixon	293
1 Introduction	293
2 Phonology	295
3 Word classes	298
4 Nouns	298
5 Noun phrase structure	299
6 Verbs	300
7 Predicate structure	300
8 Pronouns	302
9 Demonstratives and interrogatives	304
10 Construction types	304
Bibliography	306
<b>12 Small language families and isolates in Peru</b> by Mary Ruth Wise	307
1 Introduction: the languages and families	307
2 Phonology	312
3 Morphology	318
3.1 Nominal morphology	319

3.1.1	Classifiers	319
3.1.2	Case	320
3.2	Pronouns	321
3.3	Adjectives	323
3.3.1	Comparative and superlative grades	323
3.4	Verbal morphology	324
4	Syntax	329
4.1	Constituent order	329
4.2	Relativization	332
4.3	Subordination and coreference/switch-reference	333
	Bibliography	335
<b>13</b>	<b>Other small families and isolates by Alexandra Y. Aikhenvald and R. M. W. Dixon</b>	<b>341</b>
1	Brazil	341
1.1	Yanomami	341
1.1.1	Phonology	345
1.1.2	Word structure and word classes	346
1.1.3	Nominal categories	346
1.1.4	Classifiers	347
1.1.5	Grammatical relations	348
1.1.6	Verb structure	349
1.1.7	Noun incorporation and verb compounding	350
1.1.8	Syntax	350
1.2	Trumai by Raquel Guirardello	351
1.2.1	Phonology	352
1.2.2	Features of the grammar	352
1.3	Mura-Pirahã	353
1.3.1	Phonology	354
1.3.2	Features of the grammar	355
1.4	Jabutí	357
1.4.1	Phonology	357
1.4.2	Features of the grammar	357
1.5	Chapacura family	358
1.5.1	Phonology	359
1.5.2	Features of the grammar	359
1.6	Máku	361
1.7	Aikaná and Koaia	362
2	Bolivia	364
2.1	The Tacana family	364
2.1.1	Phonology	365
2.1.2	Features of the grammar	366
2.2	Isolates	367

3	Colombia	369
3.1	Phonology	370
3.2	Morphology	372
3.2.1	Word structure and typological profile	372
3.2.2	Nominal morphology	373
3.2.3	Grammatical relations	375
3.2.4	Verbal morphology	375
3.3	Syntax	376
4	Venezuela	377
4.1	Phonology	377
4.2	Features of the grammar	378
	Bibliography	379
<b>14</b>	<b>Areal diffusion and language contact in the Içana-Vaupés basin, north-west Amazonia</b> by Alexandra Y. Aikhenvald	385
1	Areal diffusion in north-west Amazonia	385
2	Linguistic situation in the Içana-Vaupés basin	386
2.1	Languages spoken	386
2.2	Cultural setting and language attitudes	388
2.3	Historical evidence	390
3	The Vaupés region as a linguistic area	391
3.1	General observations	391
3.2	Phonological characteristics	394
3.3	Grammatical structure	396
3.3.1	Typological profile and word structure	396
3.3.2	Nominal morphology	397
3.3.3	Grammatical relations	403
3.3.4	Verbal morphology and predicate structure	404
3.4	Syntax and discourse techniques	405
3.5	Semantics	406
4	Properties shared by languages of the Içana and Vaupés region	406
4.1	Pitch accent	406
4.2	Topic-advancing verbal derivation	407
4.3	Possessive classifiers and <i>-ya-</i> possessive marker	409
4.4	Complex systems of classifiers	410
5	Conclusions	411
	Bibliography	413
<b>15</b>	<b>The Upper Xingu as an incipient linguistic area</b> by Lucy Seki	417
1	The Upper Xingu and its languages	417
1.1	Languages spoken	417
1.2	Historical background	419

1.3	The mobility of groups, language loss and further contacts	423
2	The Upper Xingu as a culture area	423
3	Linguistic situation	424
4	Incipient areal diffusion in the Upper Xingu	426
5	Conclusions	428
	Bibliography	428
	<i>Index of authors</i>	431
	<i>Index of languages and language families</i>	436
	<i>Subject index</i>	445

## MAPS

---

Map 1	Carib languages with approximate locations	<i>page 22</i>
Map 2	Arawak languages with approximate locations	66
Map 3	Tupí languages with approximate locations	108
Map 4	Tupí-Guaraní languages with approximate locations	126
Map 5	Macro-Jê languages with approximate locations	164
Map 6	Tucano languages with approximate locations	208
Map 7	Pano languages with approximate locations	228
Map 8	Makú languages and dialects with approximate locations	252
Map 9	Nambiquara languages and dialects with approximate locations	268
Map 10	Arawá languages and dialects with approximate locations	292
Map 11	Small language families and isolates of Peru with approximate locations	308
Map 12	Small language families and isolates of Brazil, Bolivia, Colombia and Venezuela with approximate locations	342
Map 13	Languages in the Içana-Vaupés Basin with approximate locations	384
Map 14	Languages of the Upper Xingu with approximate locations	418

---

# Introduction

R. M. W. DIXON AND  
ALEXANDRA Y. AIKHENVALD

The Amazon basin is the least known and least understood linguistic region in the world. Maps of the language families of South America (with one colour for each genetic group) purvey an impression of anarchy – there are dabs of yellow and blue and red and orange and brown mingled together like a painting by Jackson Pollock. And when one does get hold of a grammar of an Amazonian language it is likely to show strange properties – multiple sets of classifiers, oddly conditioned ergativity splits, and so on – that constitute exceptions to received ideas about typological universals. In other instances one finds the richest examples of categories that are weakly attested elsewhere. For instance, Tucano languages (chapter 7) have the most highly articulated systems of evidentiality in the world; this is an obligatory specification of the evidence a speaker has for making a statement – whether observed, or reported, or inferred, or assumed. However, a major difficulty is that a high proportion of available grammars are incomplete, affording a glimpse of some exotic grammatical property but with insufficient information to enable the reader to fully understand it, and to realize its overall typological significance.

In fact, the major language families all have markedly discontinuous distributions (more than is found in any other part of the world). The Tupí and Arawak families each comprise at least ten separate geographical regions and the Carib family at least five. Even Tucano, one of the most linguistically homogeneous of families, is spoken over three distinct areas. In addition, there has been a great deal of linguistic diffusion, over large and small regions within the Amazon River Basin. This can make it hard – sometimes impossible – to determine whether points of similarity between two languages in close proximity are evidence of shared genetic development, or simply of mutual borrowing.

One of the editors has devoted several decades to searching for substantive linguistic universals. In case after case, just as he thought he had achieved some significant typological statement, a counter-example popped up; and this was invariably from a language of Amazonia. He decided that the most sensible course of action was to learn Spanish and Portuguese and then go to South America – visit

universities and museums, read the published materials (much of it in obscure journals), consult unpublished theses, and talk to linguists who have worked on this or that language (many of whom will never get around to publishing a full grammar). In this way he achieved a degree of insight into the most complex linguistic area in the world today.

This book is put forward in a similar spirit. It attempts to describe – in as clear and accessible a manner as possible – the current linguistic situation in the Amazon Basin, involving at least 300 languages, belonging to 20 or so language families, plus more than a dozen genetic isolates. The volume should be regarded as a first and preliminary step towards elucidating the structural characteristics of Amazonian languages, and their genetic and areal relationships.

There are several unfortunate aspects to the linguistic situation in Amazonia today. Only a minority of scholars in linguistics departments in South American nations work on the indigenous languages, with the remainder preferring to confine themselves to looking at Spanish or Portuguese in terms of the latest type of formal theory to filter down from the north. (Seki's *Gramática da língua Kamaiurá*, currently in press with Editora da Unicamp, is the first full grammar of a Brazilian language, written by a Brazilian, since Anchieta's *Arte de grammatica* of 1595.) As a result, a high proportion of the descriptive materials published on South American languages over the past few decades have been by missionary linguists of the Summer Institute of Linguistics and similar organizations.

In other parts of the world (for instance, Australia) there is mutual respect and cooperation between missionary linguists and scholars from the local universities. In contrast, in most (although not all) South American countries there is antipathy – sometimes even open hostility – between the two groups. Malicious stories are spread that the missionaries are agents of the CIA or are engaged in gold exploration or drug trafficking. None of this is true. But many of the missionaries are evangelical Protestants (mostly from North America and Europe) operating in predominantly Catholic countries. They do have better aeroplanes and other support facilities (provided by their backers in the USA and elsewhere) than are available to most local linguists. And they are – in many cases – more prolific in writing grammars and compiling dictionaries, and publishing them.

There is good and bad in every group. Some of the missionaries do – as is alleged – attempt to destroy the native culture and religion and replace it with their own brand of fire-and-brimstone Christianity. (These people should be banned; they also tend to be those who do the poorest linguistic work.) But many of the missionaries do much more good than harm. They may help protect the lands of a native tribe from invasion by gold miners and the like. They often provide medicines. They will help a people adapt to the outside world that is gradually intruding

into their lives. And they can also (over and above the business of Bible translation) provide grammars, primers, vocabularies, volumes of traditional texts; and assist in literacy work.

In many places the lack of cooperation between the two groups is marked. One may find both a group of missionaries and a group of academic linguists (of similar quality) working on the same set of languages, but with neither referring to the work of the other in their publications. They sometimes won't attend the same conference; they decline to communicate and cooperate in a way that would be beneficial to all.

A separate point is that the standard of scholarship in South American linguistics is not high. Much of the amateur data from before about 1950 has only a limited usefulness, with the transcription often being poor. Many of the missionaries have had inadequate training and produce 'cookbook' descriptions (in the 1950s and 1960s these were often cast within the impenetrable formalism of tagmemics) that cannot do justice to the genius of a language. Linguists from universities may employ other kinds of formalisms, that will soon pass out of fashion. Having made these general observations, we must add that there are notable exceptions on both sides – a number of descriptive studies that achieve a high standard of clarity and explanation.

The *Handbook of Amazonian Languages*, edited by Desmond C. Derbyshire and Geoffrey K. Pullum and published by Mouton de Gruyter, is a most worthwhile enterprise that has so far run to four volumes that include ten grammars (ranging in quality from quite good to very good) together with a number of typological and historical studies. Yet the *Handbook* would be more useful if its contributors were not forced to follow an idiosyncratic scheme of organization: syntax, then phonology, then morphology. That this is basically unworkable is demonstrated by the fact that five of the eight grammars in the first three volumes have, as the whole of '23, Morphology', a single sentence along the lines 'This has been treated in earlier sections.' It is of course necessary to know the basic inflectional morphology in order to understand the syntax, so this information is slipped in early on in the description (but at different places in each grammar). The net result may be that nowhere is there any integrated morphological statement, e.g. of the structure of the verb.

## 1 CULTURAL BACKGROUND

The Amazon is the world's major river. Its drainage area is over 7 million square kilometres (more than twice that of the Congo). The volume of water carried is a fifth of all the flowing water in the world (more than four times that of the Congo). Its 7 major tributaries are each more than 1,600 kilometres in length (compare this

with the longest river in Europe, the Rhine, which is 1,320 kilometres long). There are distinct wet and dry seasons, at different times of year in different parts of the Amazon Basin, so that the water level may vary by up to 20 metres. Several hundred kilometres up a major tributary, such as the Purús, the river may be about 1 kilometre wide in the dry season but flood to a width of 15 kilometres in the wet season, inundating the adjoining forest lands. North of the Amazon is the Orinoco, flowing into the Caribbean sea, with a drainage area of almost a million square kilometres. In this book we attempt to cover languages spoken in the Amazon and Orinoco Basins – that is, from the north coast of South America, east to the mouth of the Amazon, west to the Andes, and south to the southernmost headwaters of the Amazon tributaries. If most of the languages in a family are spoken in the Amazon/Orinoco Basin (e.g. Arawak) then we cover that family. If most of the languages in a family are outside the region (e.g. Guaicuru) then we do not deal with that family.

Most of the Amazon and Orinoco Basins is lowland rain forest, rich in vegetable and animal foods. Archaeologists (by and large) consider that it was first populated about 12,000 years ago, by hunter/gatherer populations who crossed from Asia at the Bering Strait (which was then dry land) and travelled down the North American coast. It would have taken only a couple of thousand years for Amazonia to be fully populated. The major protein food is fish, supplemented by the hunting of peccaries, deer, tapir, rodents and monkeys. Domestication of plants – focusing first on bitter manioc – is dated to about 5,000 years ago. It is thought to have begun in the higher lands around the foothills of the Andes. People employing slash-and-burn agriculture soon spread out along all the major rivers, while populations of hunters and gatherers contracted to the more remote areas.

Each language family tends to have a characteristic profile in terms of the type of territory it is found in, methods of food procurement, and material culture. Almost all tribes speaking languages belonging to the Arawak, Carib and Tupí families are found in the rain forest, use agriculture, and manufacture canoes, hammocks and pottery. In contrast, Jê-speaking peoples are mostly found on grasslands; they have little agriculture and no canoes, hammocks or pottery; however, they do have the most complex systems of social organization. Scattered between the agricultural tribes, in the heart of the rain forest, are small tribes of hunters and gatherers, belonging to minor linguistic families such as Makú, Mura-Pirahã and Guahibo. Some of these may be the remainder of earlier populations that occupied larger tracts of land before the agricultural expansion. The languages of the seven Makú tribes, for instance, have been suggested to comprise a rather divergent language family, although full proof has still to be provided for this. They may instead constitute scattered relics of an olden-days linguistic area, whose lan-

guages had – through long-term diffusion – converged towards a common structural prototype. (This is a topic that requires further work.)

Multilingualism was (and is) the norm among the Indian tribes of Amazonia. Some groups – predominantly, those with agriculture – were rather warlike, given to raiding their neighbours. Within some tribes there was a caste of slaves, which originated as captives of war. There were also symbiotic relationships between different groups of people. For instance, the agricultural Tucano tribes (living along the Vaupés River) and the hunting/gathering Makú (living in the forest away from the river) are in a ‘master–underling’ relationship. The Makú supply their ‘masters’ with the meat of forest animals and with fish poison, receiving in exchange manioc flour from Tucano gardens and pottery artifacts. Makú people are considered socially inferior to the Tucano and do not enter into the system of exogamous intermarriage that characterizes the Vaupés region. There are some clans within Tucano tribes (and even one whole tribe) that are said to be of Makú ancestry. Although they now speak a Tucano language and practise agriculture they are considered socially inferior and to be less desirable as marriage partners. (In §1.1 of chapter 3 there is mention of the creation of a mixed language as the result of Carib conquest over an Arawak people.)

Then, in the sixteenth century, came the Europeans. White-skinned invaders quickly took over the coastal areas and major rivers as far as they were navigable. (Chapter 15 explains how rapids on the Xingu River proved a barrier to the invaders, preserving the Upper Xingu as a refuge area into which Indian tribes congregated.) The Indian population rapidly reduced until, by 1900, it was probably no more than one-tenth of what it had been in 1500.

The biggest killer was involuntary. White invaders brought with them diseases to which Indians had no immunity – smallpox, influenza and the like. European diseases spread ahead of the Europeans themselves; many tribes and languages are likely to have disappeared before even their name could be recorded. Others went a little later. The Arawá people (see chapter 11) are known only from a fifty-word vocabulary collected by the English explorer Chandless in 1867. Almost everyone in the tribe died a few years later from a measles epidemic; the few survivors took refuge with the neighbouring Kulina who are said to have massacred them (Rivet and Tastevin 1938: 72–3).

A typical situation was for a number of tribes, each reduced in numbers, to merge. The ethnologist Franz Caspar (1956: 221) lived for a while among the Tupari tribe and was told by them that ‘in the days of their grandfathers and great-grandfathers several small tribes had merged. Of every man and every woman Topto was able to say without hesitation of what extraction they were. There was only one man left out of each of the “Vaikorotá”, “Aumeh” and “Mensiató” tribes. Five were real

“Tupari”. All the rest . . . were “Vakarau” . . . Even the present language of the tribe, he said, was not the old Tupari, for the minorities had adopted the language of the Vakarau. But people could still recall how the various tribes had spoken and Topto told me a few words of the real Tupari language.’ Just one of the original languages had survived, almost certainly with considerable substrata from the others.

There was also enslavement by whites. Hemming (1978) tells how the European settlers at the mouth of the Amazon would travel upriver, capture a tribe, bring them back to work on their plantations under harsh conditions, then when they died out (often, within about ten years), go upriver again to capture and enslave another tribe. And there was a great deal of simple murder. A gun is more powerful than bows and arrows; if white people wanted some patch of land they simply took it, together – if necessary – with the lives of the people who had been living there.

Indian tribes invariably fought back; they might win in the short term but never in the long term. Some of them simply fled. At the end of the sixteenth century, eighty-four Tupinambá-speaking villages on the east coast of Brazil became exasperated by the treatment at the hands of the Portuguese and decided to migrate inland. They travelled – slowly – up the São Francisco River, across the grasslands of Mato Grosso, until they reached what is now Bolivia, where they were dismayed to encounter Spanish colonists. They then turned north-east and followed the Madeira River until it joined the Amazon, settling on a large island in the middle of the river, which is now called Tupinambarana. But there was no escape. In 1639 their island was invaded by Portuguese coming up the Amazon (Hemming 1978: 235).

Before Amazonia was invaded, relations between tribes in Amazonia were sometimes pacific, other times turbulent. We know, for instance, that speakers of Tupí-Guaraní languages spread out over a wide area, with considerable effect on languages of peoples they came into contact with (see §2.2 of chapter 5). There would of course have been instances of migration and of language death. But these happenings greatly increased after 1500. Invasion of the forest hinterlands intensified with the rubber boom at the end of the nineteenth century. This fell away from about 1910 (when cheaper rubber became available from Malaysia), and lands which had been occupied by non-Indians were again released to their original owners. There are, however, rather few Indians remaining, and these are fast being acculturated.

A typical example concerns the Yuquí, speakers of a Tupí-Guaraní language. This group lost agriculture and became hunters and gatherers (although they still retained a caste of slaves). For several hundred years they were able to keep out of the way of Spanish colonists, retreating further into the Bolivian forest. By the

1950s the colonial expansion left the Yuquí nowhere to hide. They were in danger of being wiped out when, in the early 1960s, they were persuaded to settle on a New Tribes Mission station (Spearman 1989). Having been saved from one fate, they are now pointed towards another. With their original life-style and culture stripped away (and with only Spanish names), the ethnic identity and language of the Yuquí are unlikely to survive for more than a couple of generations.

Estimates vary, but it is thought that there were, in 1500, somewhere between 2 and 5 million people living in Amazonia. The present Indian population is no more than 400,000. Of the 170 languages reported to be still spoken in Brazil, 115 have less than 1,000 speakers with only 4 having more than 10,000 speakers (none has more than 20,000); similar figures apply for other South American nations. The governments have been slow to extend roads into remote areas or to supply schools and medical posts. But these are coming, and with them assimilation into mainstream society. Every year the indigenous languages are used less and less, and Spanish and Portuguese more and more. Every year another few languages pass into oblivion. Of the estimated 300 languages now spoken in Amazonia, only a small fraction are likely to be still actively used in 100 years' time.

We should round off this cultural overview with a short comment on trade languages. Several creoles evolved during the early years of colonization, based on Spanish, French and Portuguese. But there was one lingua franca of major importance. What was called *Língua Geral* ('language [that is] universal' in Portuguese) or *Nheen-gatu* ('speech [that is] good' in *Nheengatu* itself) evolved on the east coast of Brazil in the sixteenth century. Its morphology was simplified from *Tupinambá* – a language of the *Tupí-Guaraní* subgroup that was spoken all along the east coast – but the syntax is similar to Portuguese. *Língua Geral* soon spread up the Amazon and had dialectal variants in different regions. Indeed, it began to compete with Portuguese as the major language of Brazil. In 1727 King John V (back in Lisbon) banned the use of *Língua Geral* and it began gradually to fade. *Língua Geral* was the trade language to which most tribes in the Brazilian Amazon were first exposed and it was only replaced by Portuguese in the twentieth century. This creole is not quite extinct, still being spoken as first language by a small number of people in the Upper Rio Negro region.

## 2 LINGUISTIC DIFFUSION

There are a number of cultural traits that recur throughout Amazonia. These include female initiation rites (which are much commoner than puberty rites for boys). There are generally a number of shamans who control spirits that can both cause and cure diseases. (See, for example, Steward and Faron 1959: 284–318.) The

word for ‘dog’ is often either the same as the word for ‘jaguar’ or else related to it. Probably the most pervasive lexeme is *kuku* or *koko* for ‘mother’s brother’ or ‘father-in-law’ (in a system where one can marry a cross-cousin). If a man marries a woman from another tribe he must know how to address his wife’s father; this has undoubtedly led to the widespread borrowing of the term for ‘wife’s father’.

Linguistic traits, by and large, diffuse more slowly than other cultural traits. However, given sufficient time, languages from several genetic groups that are located in the same geographical area will gradually come to share certain linguistic features and will, as a consequence, make up a ‘linguistic area’. This can be defined as: a region including languages from several different genetic groups, with the languages sharing certain symptomatic features which can be inferred to have diffused across the area. The features will not, as a rule, be found in languages from these genetic groups which are located outside the area. It should be noted that very common properties (those found in very many of the languages of the world, e.g. a tense system) are less significant as diagnostic markers of a linguistic area than are more unusual properties (e.g. an evidentiality system).

Amazonia can be recognized as a linguistic area in terms of features like the following, which are shared by all (or most) languages in the area.

- (a) The majority of languages are polysynthetic and head marking; agglutinating with little fusion.
- (b) There is typically one liquid phoneme, which is frequently a flap. There are usually more affricates than fricatives. The high unrounded central vowel *ɨ* is frequent. A typical Amazonian vowel system has five members: *i, e, a, ɨ, u/o*. There is typically contrastive nasalization of vowels.
- (c) Many languages have extensive classifier and/or gender systems. Gender assignment is often semantically transparent, and is not overtly marked on the head noun.
- (d) There are very few oblique cases – often just a locative and an instrumental/comitative.
- (e) Possession (either alienable or inalienable) is typically marked on the possessed noun, not on the possessor; the most widespread word order is ‘possessor possessed’ (e.g. ‘John his-canoe’).
- (f) Often, just one core argument is cross-referenced on the verb. There may be different bound pronominal paradigms depending on which core argument is being cross-referenced in each particular instance.
- (g) The rules for which core argument is cross-referenced can be complex (relating to the meaning of the verb, clause type, etc.) often giving rise

to a 'split-ergative' system. Fully accusative systems of marking for predicate arguments are rarely encountered.

- (h) The bound pronominal forms marking a possessor within an NP are typically the same as one of the bound pronominal paradigms for marking core arguments of a clause (sometimes the same as the A or A/S<sub>a</sub> series, other times the same as the O or O/S<sub>o</sub> series).
- (i) Most (although not all) languages have prefixes; there are typically fewer prefix than suffix positions.
- (j) If there are several prefix positions, the bound pronominal prefix(es) will typically appear further from the root than prefixes that mark valency-changing derivations (e.g. causative, applicative). (Tucano languages are entirely suffixing; the bound pronominal suffixes appear further from the root than suffixes that mark valency-changing derivations.)
- (k) Most verbal categories (e.g. tense, aspect, modality, direction) are expressed through optional suffixes.
- (l) Subordinate clauses typically involve nominalized verbs, with the type of subordination being marked on the verb.
- (m) If there is noun incorporation, typically only those nouns which are obligatorily possessed can be incorporated, and they typically precede the verb root.
- (n) In many languages adverbs and adpositions may be incorporated into the verb, typically following the verb root.
- (o) There is generally only a small class of lexical numbers.

There are of course a few exceptions to these pan-Amazonian tendencies. For instance, Tupí-Guaraní languages are the only languages in Amazonia which allow incorporation of unpossessed nouns; however, in modern languages this technique is falling out of use (Seki forthcoming, Kakumasu 1986). In Nadëb, a Makú language from the Middle Rio Negro, incorporated adverbs and adpositions precede the verb root instead of following it. In Palikur, a North Arawak language from Brazil and French Guiana, incorporated body parts follow the root (Aikhenvald and Green 1998).

It is interesting to compare typological characteristics of the Amazonian linguistic area, in lowlands South America, with those of the Andean linguistic area in the adjacent mountains, which comprises the Quechua and Aymara families. The Andean area is clearly different in almost all of the characteristics just listed. (b) There are two or three liquids; fricatives rather than affricates; and a three vowel system, *i*, *a* and *u*, with no contrastive nasalization. (c–e) There are no classifier or

gender systems; there is an extensive set of case markers; possession is marked both on possessor and on possessed. (f–h) Two core arguments are marked on the verb, in an entirely accusative system; bound pronominal markers of possession show some similarity to, but are not identical with, the forms marking core arguments on the verb. (i–k) There are no prefixes; and there is an obligatory suffixal system for tense and aspect. (l) Subordination does not involve nominalization. (m–n) There is no incorporation of nouns, adverbs or prepositions (Cole 1982: 161 mentions incorporation within nominalization, but this is just a type of compounding, as in English). (o) There is a full set of lexical numbers. As regards characteristic (a), Andean languages are synthetic, and combine head and dependent marking; they are basically agglutinating with some fusion (subject, object and tense suffixes to the verb may be fused).

There is no sharp boundary between the Andean and Amazonian linguistic areas – they tend to flow into each other. For instance, Andean features such as lack of prefixes and an accusative technique for marking syntactic function are found in languages of the Tucano family, which are in Amazonia but fairly close to the Andes.

There are a number of grammatical properties which are not shared by all Amazonian languages but are found in the languages in certain regions and help to define these as linguistic subareas within a wider linguistic area. These include:

- (1) contrastive lexical tones are found in two regions – one in southern Amazonia (overlapping the states of Rondônia and Mato Grosso) and one in the north-west (along the Vaupés river, from Brazil into Peru).
- (2) switch-reference marking is found in a group of languages in western Amazonia.
- (3) gender assignment is not semantically transparent in a region of southern Amazonia centred on the Purús river basin (where Bolivia, Brazil and Peru meet) which includes languages from the Arawá and Chapacura families and the Peruvian and Pre-andine subgroups of Arawak.
- (4) classifiers are used in different morphosyntactic contexts in different regions – for instance, they are used in possessive constructions in the Upper Rio Negro region but not in Peruvian Arawak languages.

The limited regional distribution of these features can be important for hypothesizing whether certain features were more widespread at some time in the past, and are being progressively lost, or whether they are recent innovations that are currently diffusing more and more widely.

### 3 GENETIC RELATIONSHIP

There can be a number of explanations for some perceived similarity between two (or more) languages. It can be an indication of genetic relationship. Or it can be the result of areal diffusion, borrowing from one language into another (in one or in both directions) or borrowing by both languages from a third language. Or it can be some universal feature, e.g. the word for 'blow' typically imitates the activity, having a form something like *p<sup>h</sup>u* (the actual form will relate to the phonetic and phonological resources of each language). Or it can be chance, as where the Australian language Mbabaram and the Indo-European language English both have *dog* as the name for an animal of the species *Canis*.

To say that a group of languages is genetically related (as a language family) is to say only one thing. It is to assert that they go back to a common ancestor, each having developed from this 'proto-language' by its own set of historical changes. Concordant with this, there is only one way to prove that a group of languages is genetically related. This is to propose what the proto-language was like (in some detail) and to describe how each of the modern languages developed, by systematic changes, from this common ancestor.

If two (or more) languages show a certain set of similarities, these need not necessarily be evidence for genetic relationship. They may, alternatively, be due to borrowing, to universal tendencies or just to chance. It is true that certain kinds of similarity (a pronoun paradigm, or a suppletive set) are strongly suggestive of genetic relationship. But suggestion is not proof. And the only proof is to *demonstrate* the genetic relationship by showing the regular changes which each individual language has undergone from a postulated proto-language (dealing with phonology, verbal and nominal morphology, pronouns, markers of negation, etc., as well as all kinds of lexemes).

In only a few instances do we have written records of sufficient time depth to identify the proto-language and details of the changes modern languages have undergone – Sanskrit and the modern Indic languages; Latin and Romance languages; Old Irish and Irish Gaelic, Scots Gaelic and Manx. In other instances, large portions of a putative proto-language have been reconstructed, together with the changes modern languages have undergone.

A number of proto-languages have been proved in this way, so that all scholars are agreed on their genetic unity; they include Indo-European, Uralic, Dravidian, Austronesian, Algonquian and Mayan. Although most of these language families have been recognized for a fair period of time, it has not proved possible to establish any higher-level genetic linkage between them.

However, some people are not prepared to rest upon what can be scientifically proved and want to go on – to attempt to relate families together as higher-level families, and so on back (sometimes, even back to a putative proto-World language). Such suggestions of higher-level genetic groupings are generally based on a few odd correspondences (often, between one or two languages in family X and a few languages in family Y, rather than between proto-X and proto-Y) and thus cannot be accepted as serious scholarship.

The suggestions that have been made are mutually incompatible, which is itself not a good portent. For instance, Schuller (1919/20) proposed a genetic link of Carib with Arawak (also throwing in Chibcha and Mayan); Greenberg (1960) linked Carib with Jê, Pano and Nambiquara; Greenberg (1987) kept Carib, Jê and Pano together but now placed Nambiquara in a group with Tucano and Makú, among others; and Rodrigues (1985) suggested a genetic link between Carib and Tupí. Thus Carib has been linked with Arawak, Maya, Chibcha, Jê, Pano, Nambiquara and Tupí at one time or another (although not all at once). Kaufman (1990) surveys a number of higher-level genetic suggestions, showing how these disagree; he has lists of languages that are ‘Macro-Jê for some [investigators], Macro-Kariban for others’, and of languages that are ‘Macro-Tupian for some, Macro-Jê for others’, etc.

A perceptive comment on suggestions of this type concerning higher-level genetic links comes from Mason (1950: 162): ‘It is a truism of linguistic research that, given large enough vocabularies to compare, and making allowances for all possible changes in the form of a word or stem, as well as in its meaning, a number of apparent similarities, convincing to the uncritical, can be found between any two languages.’ The lack of scientific basis for posited long-distance genetic links can be illustrated with a couple of examples (these could be multiplied a hundred times over).

Ehrenreich (1897) first suggested a genetic link between the Arawá and Arawak families. He gave 17 ‘correspondence sets’ with examples for each drawn from 1 or more of 3 Arawá languages, and 1 or more of 29 Arawak languages. A typical set is:

‘bow’ (as in ‘bow and arrow’)

Arawá languages	Arawak languages
Paumari: <i>kudaii</i>	Barē: <i>dauidaja</i>
Jamamadi: <i>didiša</i>	Kustenaú: <i>tuti</i>
Arawá: <i>bigauaha</i>	Tariana: <i>shidoa</i> (‘arrow’)

No proto-form was suggested (nor is it easy to see one that could be suggested), for proto-Arawá, or for proto-Arawak, or for proto-Arawá-Arawak. In fact Arawá lan-

guages all have different (and unrelated) words for ‘bow’. For proto-Arawak, Payne (1991: 396) reconstructs \**tapo* ‘bow’ (a form that is in fact reflected in Tariana *yawithepu* ‘bow’). Yet on the basis of 17 sets of ‘correspondences’ like this, it was suggested that the Arawá and Arawak families are genetically related. Rivet and Tastevin (1940) followed a similar ‘method’, drawing odd correspondences between some forms in a few Arawá languages and some forms in a few Arawak languages (and they drew on a pool of 86 Arawak languages). A typical (short) entry from their list of ‘cognate’ sets is (Rivet and Tastevin 1940: 8):

appeler [‘to call’]:

*wawana-he*, il appelle [‘he calls’], Kulina [Arawá]  
*pi-uāna*, *bi-uāna*, Tariana [Arawak]

The idea that Arawá (now called Madi) and Arawak are related in a family called ‘Arawakan’ was repeated by Matteson (1972). She believed that she was demonstrating genetic relationships by ‘rigorous application of standard techniques of the comparative method’ (Matteson *et al.* 1972: 21). The first of her 353 putative cognate sets of proto-Arawakan is (Matteson 1972: 172):

1. ‘abdomen’ \**tia-ri* [in proto-Arawakan]
  - \**tika-telko*, proto-Piro-Apuriná
  - \**tso-mon-tí-a*, proto-Ashaninka
  - \**ato-li*, proto-Madi [i.e. proto-Arawá]
  - \**tee-IV*, proto-Newiki

Note that the only segment that the proposed proto-Arawakan reconstruction and the four intermediate reconstructions have in common is ‘t’.

Further on in the paper, Matteson (1972: 203) gives the Piro and Apuriná forms:

1. ‘abdomen’ \**tiká-telko* [*sic*, accent included here but not above]
  - [proto-Piro-Apuriná]
  - čkete*, Piro
  - tikáko*, Apuriná

and then the forms in her Ashaninka subgroup (Matteson 1972: 211):

1. ‘abdomen’ \**tso-mon-tí-a* [proto-Ashaninka]
  - motía*, Machiguenga
  - tsomonte*, Campa
  - tsomoné*, Nomatsiguenga

In the proto-Madi section there is no word for ‘abdomen’. However, the first cognate set given is (Matteson 1972: 219):

1. ‘bark’ \**ato-li* [proto-Madi]  
*ete-ru*, Culina  
*\*atu-li*, pre-Jamamadi

There are a number of possible explanations here. One is that Matteson just copied out the first item from her proto-Madi list, assuming that (as in the other lists) it was ‘abdomen’, overlooking the fact that the proto-Madi list has no term ‘abdomen’ but its first item is ‘bark’. Or she may have imagined that ‘abdomen’ could be cognate with ‘bark’, but neglected to mention this.

There is no ‘abdomen’ given in the proto-Newiki list, but we do find this further on, under proto-Western-Newiki (Matteson 1972: 231):

1. ‘abdomen’ \**tée-IV* [*sic*, accent included here but not above] [proto-Western-Newiki]  
*dé-*, Piapoco  
*tere*, Cabiari

This example is typical of Matteson’s methods. It shows a lack of understanding of the established methodology of language comparison and reconstruction (and an extreme sloppiness). Indeed, the volume begins with ‘reconstructions’ of lexemes in ‘proto-Amerindian’, comparing odd forms from odd modern languages across two continents.

The Arawak family is well established (going back to 1783). Recently, the term ‘Arawakan’ has been used for an unproven higher-level unit which includes Arawak (now called Maipuran), Arawá, Chapacura, Guamo, Harakmbet and Uro-Chipaya (the details vary a little from author to author). This is part of the reason why we avoid the term ‘Arawakan’ (along with other terms ending in ‘an’) in this volume, preferring ‘Arawak’, a label consistently used by scholars from South American countries for the well-established genetic grouping (corresponding to the Maipuran of the Arawakanists).

The first satisfactory reconstruction of a part of proto-Arawak vocabulary is by Payne (1991); this includes 203 forms. Dixon has reconstructed over 400 lexemes for proto-Arawá. Three possible cognates can be recognized between these lists, none of them fully convincing. The grammatical forms of proto-Arawá and proto-Arawak are also quite different. It must be concluded that there is no evidence whatsoever that (despite their similar names and geographical proximity) the Arawá and Arawak language families are genetically related.

In fact, the most perceptive commentators on South American languages do not suggest any genetic connection here, e.g. Mason (1950), Loukotka (1968), Tovar and De Tovar (1984) and Rodrigues (1986). However, the eccentric suggestion from

Ehrenreich, Rivet and Tastevin, and Matteson has been repeated by others, e.g. Métraux (1948), Noble (1965) and Greenberg (1960, 1987). Even so eminent and (generally) reliable a scholar as Derbyshire uncritically accepts this, mingling data from Arawá with that from Arawak languages in his 'Comparative study of morphology and syntax in Brazilian Arawakan' (Derbyshire 1986).

There can be a tendency to accept what is put forward as an orthodoxy, without pausing to question it. When a missionary linguist from Peru, who had spent over thirty years studying Kulina (an Arawá language), received a copy of Dixon (1995), she wrote to Dixon: 'I became so intrigued by your declaration of no kinship with the Arawak family that I read and re-read that part [of the paper] . . . No wonder we never did seem to "fit the pattern" among Arawaks here.' That is, this linguist had been told that Kulina was an Arawak language. But there are no significant cognates and entirely different grammars, which puzzled her.

We have dwelt at some length upon the Arawak/Arawá suggestion to illustrate its vacuous nature. A similar refutation could be provided for almost all other (perhaps for all other) suggestions of higher-level genetic relationships between the established language families.

There have been a number of commentators who are sound scholars and confine themselves to summarizing the established families and the isolates – as already mentioned, these include Mason (1950), Loukotka (1968), Tovar and De Tovar (1984) and Rodrigues (1986). Rivet has considerable reputation but his work on South American comparative linguistics is deeply flawed (he also suggested a genetic connection between South American Indian and Australian Aboriginal languages). Greenberg's (1960, 1987) 'Amerind' has attracted considerable publicity but is without scholarly foundation (see the critical comments in Dixon 1997: 54–5 and references quoted there).

It should be borne in mind that just because something has been published in a book does not necessarily mean that it has any value. People writing a grammar of language X from family Y sometimes feel impelled to say something like 'family Y belongs to the Macro-Tucanoan stock within the Equatorial-Tucanoan stock within Greenberg's Amerind'. Such a statement should be avoided. It simply adds a veneer of fantasy to what may well be a sound and useful grammar of language X. (Contributors to this volume have confined themselves to just mentioning those genetic groupings that are well accepted and have been proved.)

As a rough rule of thumb, readers should be on their guard against any name beginning with 'macro-' or any unit labelled as a 'stock'. With the possible exception of Macro-Jê (which, if proved to be a genetic group, ought to be relabelled) all macro-X's are intensely speculative. A similar comment applies for almost all stocks. (Note that in this book we refer to the Tupí family and the Tupí-Guaraní

branch or subgroup within it – just as one talks of the Indo-European family and the Germanic branch or subgroup within it – in preference to Tupí stock and Tupí-Guaraní family.)

#### 4 THE PUNCTUATED EQUILIBRIUM MODEL

It is generally recognized among historical linguists that one can only recognize historical relationships between languages (and reconstruct shared proto-languages) to a time-depth of 5,000–8,000 years. Given any longer time, individual languages are likely to have changed too much for their original genetic connection to be apparent. Thus, Greenberg's (1987) views about a relationship between all the languages of South America and most of those of North America are not only unproven but unprovable.

But if one views the 'family tree' as the basic model of language relationship, one must presume that there probably was a family tree joining together family trees. That is, that many of the language families (and isolates) of the Americas presumably are genetically related, only one would need a time machine to discover exactly how. In this view, Greenberg's scheme is a perfectly possible one. But there are many alternative scenarios that are equally plausible, and there is no way of deciding between them.

We believe that this is a mistaken view. The family-tree model is appropriate for certain kinds of language development but not for all of them. Humankind is believed to have had language for at least 100,000 years (many people would opt for a longer time-span). Consider the Indo-European family, which has over 100 modern languages and for which a time-depth of about 6,000 years is posited. There are about 17 periods of 6,000 years in 100,000 years. If one language spawns  $10^2$  descendants in 6,000 years, then over 100,000 years it should give rise to  $10^{2 \times 17} = 10^{34}$  or 10 million billion billion billion languages. But how many languages do we have in the world today? About 5,000–6,000. This suggests that we should re-think the idea that a family-tree type of language split is the universal model of language development.

Dixon's essay *The rise and fall of languages* (1997) puts forward an alternative idea, the Punctuated Equilibrium model of language development. He suggests that during most of the history of humankind there have been long periods of equilibrium, interrupted by short periods of punctuation. During an equilibrium period a given geographical area would have been inhabited by a number of small political groups of similar size, each with its own traditions, religion, laws and language. No one group or language would have substantially greater prestige than any other. Cultural traits, including linguistic features, would steadily diffuse across the geo-

graphical area. There would always be gradual changes; some languages would split and some would cease to be spoken, but this would be on a relatively minor scale. In due course, as a result of linguistic diffusion, all the languages in the geographical area would become more similar in structural profile – they would converge towards a common prototype.

An equilibrium situation may be punctuated by some cataclysmic happening. This could be a natural event (flood or drought, rising or falling sea levels); or some material innovation (say, a new method of food production); or the rise of some military or religious leader, intent on building an empire; or entry into new territory. In a period of punctuation there will be expansion and split of peoples and of languages. It is during punctuation that a family-tree model is appropriate; we get the divergence of a steadily increasing set of languages, all emanating from a common proto-language. Eventually, the punctuation mode will gradually lose power, and merge into a new equilibrium period.

When the first people came into South America – 12,000 or so years ago – they would quickly have expanded to fill the continent. Population is likely to roughly double each generation if there is unlimited food and land available. This was a period of punctuation, describable by a family-tree diagram. Once the land was fully populated (which may only have taken 2,000–3,000 years), a period of equilibrium would have commenced within each geographical zone – in the tropical forests, on the grasslands, on the mountains, and so on. There would have been many small groups of hunters and gatherers living in a state of relative equilibrium with each other. Linguistic traits would have diffused across the languages within each region.

A major punctuation would have been triggered by the adoption of agriculture, believed to have taken place about 5,000 years ago. The peoples with agriculture had a distinct advantage in food production. As a result they expanded and split, and so did their languages. This serves to explain the readily provable genetic unity of the Arawak, Carib and Tupí families. The agriculturalists took over the best land, along major rivers. Scattered between them are hunters and gatherers. Genetic relationships are here less clear. As mentioned in §1, it is possible that modern-day Makú tribes are relics of an earlier equilibrium situation, with their similarities being the result of having belonged to an olden-days diffusion area (rather than being indicators of close genetic relationship). The point being made is that there is no sure way to distinguish between similarities that are due to areal diffusion and those that reflect common inheritance. Suppose Europe came to be invaded and settled by the Chinese, leaving just small pockets of people speaking Italian and Basque and Hungarian. A later-day linguist might well take the similarities between these three relic languages (their ‘Standard Average European’ features) as evidence

of genetic relationship. It is possible that the history of the Makú follows a similar pattern.

Chapter 6, on Macro-Jê, discusses the Jê family and 11 other language families, basically all spoken in the grasslands region. Rodrigues presents 39 possible cognates and a number of points of grammatical similarity. All of these are of a typological nature, relating to similarities in construction type, constituent order and grammatical categories between the languages, rather than shared grammatical forms. (Pronominal forms, for instance, appear to be rather different between Macro-Jê families.) Rather than all the Macro-Jê families being related in a higher-level family tree, it seems to us that they could constitute a long-term linguistic area; this would account for their considerable typological similarities. (There may, of course, be genetic links between *some* of the established families, within the linguistic area.)

It will be seen that – in terms of the Punctuated Equilibrium model of language development – the quest for a family of language families (a tree of trees) is misconceived. Each modern language family probably had its origin in the end of a period of equilibrium. Similarities that are noted between proto-languages may well be areal features which had – during the equilibrium period – spread to all or most of the languages in a given geographical region.

In summary, at the end of the initial human expansion across the whole of South America a family tree would have appropriately mapped the period of punctuation that was drawing to an end. A long period of equilibrium then ensued, with roughly stable population and a great deal of cultural and linguistic diffusion within each ecological zone. The genetic relationships between languages, which had been clear at the end of the period of punctuation, would have gradually become blurred and finally lost as more and more features diffused. Then the equilibrium would have been punctuated and just a few of the languages (those whose peoples had some distinct advantage in living and winning, e.g. agriculture) would expand and split, each starting its own family tree (and obliterating other languages, whose speakers did not have this point of advantage).

Establishing that a group of languages is genetically related, as a language family, is generally an easier matter than deciding on the internal constitution of the family tree, i.e. subgroupings. To assign a set of languages to a subgroup there must be evidence that they have shared some historical development (and it should be some rather distinctive change, not anything that commonly recurs all over the world). The difficulty here is to distinguish between similarities that are due to shared genetic development and those which are due to diffusion. Consider the Arawak family, for instance. A number of subgroups have been tentatively established, each of them being located in a certain region. The languages in a given subgroup do

share certain developments. But these may well be characteristic features of that geographical region, and found in both the Arawak and the non-Arawak languages that are spoken there. It is first necessary to examine the areal linguistics of the region, factor out the areal features, and then see whether the remaining similarities between the Arawak languages spoken there constitute sufficient evidence for subgrouping. None of this has yet been done. A full investigation of subgrouping – within an areal perspective – is an important topic for future research.

The European invasion, commencing in 1500, acted as an abrupt punctuation of linguistic areas across all of South America. The prestige languages – Spanish, Portuguese and a few creoles – are continually expanding their domains. It is likely that more than half (perhaps much more than half) of the languages spoken in 1500 have already passed into oblivion, and the remainder are following at a steady rate. Quechua is currently estimated to have about 8.5 million speakers; however, in central Peru (and probably in other regions as well), most children of Quechua-speaking parents are preferring to speak just Spanish (Adelaar 1991: 50). Only Guaraní appears to be safe, in the medium term, because it is one of the two national languages of Paraguay (and is in fact spoken in Paraguay by more people than is Spanish). In lowland Amazonia there is little hope for even medium-term survival of any language. The tentacles of European-style civilization have been slow to penetrate the rain forest, and it is this that has helped some cultures and languages to maintain their autonomy. But the outside world is now creeping in. Every decade, each indigenous language is spoken a little less (sometimes, a lot less) and Spanish or Portuguese a little more (or a lot more).

Describing these languages, before they disappear, is an urgent task. If everyone who calls himself a linguist – from South American countries and from overseas – were to devote a year or so to fieldwork, and then write and publish a grammar, dictionary and volume of texts for some previously undescribed (or scarcely described) language, then most of the rich linguistic and cultural heritage would be preserved, for posterity. (This would also lead to substantial enrichment of Basic Linguistic Theory.)

## 5 ORGANIZATION OF THIS BOOK

Basically, we have devoted a chapter to each of the major language families – a long chapter for a large family (with several dozen languages) and a short chapter for a small family (with just a few members). Tupí-Guaraní (a subgroup within the Tupí family) is the best-known group of languages and we have accorded it a chapter of its own.

Then, in chapter 12, Mary Ruth Wise surveys the small language families of Peru. In chapter 13, we provide some information on the small families and language isolates of Brazil, Bolivia, Colombia and Venezuela. The final two chapters look at linguistic areas – the established area in the Içana-Vaupés River Basin, and the incipient area on the Upper Xingu River.

There are a few languages that we say nothing about, for the simple reason that almost nothing is known about them. These include: Awaké, Hoti, Irantxe, Kanoé, Puinave, Sape and the small Katukina family.

Editing the present book has not been an easy task. The contributors come from differing backgrounds and have different kinds of linguistic training. In a couple of instances one of the editors had personal contact with contributors and she was able to work closely with them on their chapters. Chapters 9 and 15 were written in Portuguese and translated into English by the editors. We have not in all cases been able to achieve the degree of coverage that we had hoped for. The reader will be able to see for themselves, from this volume, what the state of the art is – at the end of the twentieth century – with respect to knowledge of the various language families in Amazonia.

#### BIBLIOGRAPHY

- Adelaar, W. F. H. 1991. 'The endangered languages problem: South America', pp. 45–91 of *Endangered languages*, ed. R. H. Robins and E. H. Uhlenbeck. Oxford: Berg.
- Aikhenvald, A. Y. and Green, D. 1998. 'Palikur and the typology of classifiers', *Anthropological Linguistics*, 40.429–80.
- Anchieta, J. de. 1595. *Arte de grammatica da lingoa mais usada na costa do Brasil*. Coimbra: Antonio de Mariz.
- Caspar, F. 1956. *Tupari*. London: Bell.
- Cole, P. 1982. *Imbabura Quechua*, Lingua Descriptive Series. Amsterdam: North-Holland.
- Derbyshire, D. C. 1986. 'Comparative survey of morphology and syntax in Brazilian Arawakan', pp. 469–566 of *Handbook of Amazonian languages*, vol. I, ed. D. C. Derbyshire and G. K. Pullum. Berlin: Mouton de Gruyter.
- Dixon, R. M. W. 1995. 'Fusional development of gender marking in Jarawara possessed nouns', *IJAL*, 61.263–94.
1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Ehrenreich, P. 1897. 'Materialien zur Sprachenkunde Brasiliens: Vokabuläre von Purus-Stämmen', *Zeitschrift für Ethnologie*, 24.59–71.
- Greenberg, J. 1960. 'The general classification of Central and South American languages', pp. 791–4 of *Men and cultures: Selected papers of the 5th International Congress of Anthropological and Ethnographic Sciences*, ed. A. Wallace. Philadelphia: University of Pennsylvania Press.
1987. *Language in the Americas*. Stanford: Stanford University Press.
- Hemming, J. 1978. *Red gold: The conquest of the Brazilian Indians*. Cambridge, Mass.: Harvard University Press.