

Oxford
LINGUISTICS

Areal Diffusion and Genetic Inheritance

*Problems in Comparative
Linguistics*

EDITED BY

Alexandra Y. Aikhenvald
and R. M. W. Dixon

Areal Diffusion and Genetic Inheritance

Explorations in Linguistic Typology

General Editors

Alexandra Y. Aikhenvald and R. M. W. Dixon

Research Centre for Linguistic Typology, La Trobe University

This series focuses on aspects of language that are of current theoretical interest and for which there has not previously or recently been any full-scale cross-linguistic study. Its books are for typologists, fieldworkers, and theory developers, and designed for use in advanced seminars and courses.

Published

1 Adjective Classes

edited by

R. M. W. Dixon and Alexandra Y. Aikhenvald

2 Serial Verb Constructions

edited by

Alexandra Y. Aikhenvald and R. M. W. Dixon

3 Complementation

edited by

R. M. W. Dixon and Alexandra Y. Aikhenvald

In preparation

4 Grammars in Contact

edited by

Alexandra Y. Aikhenvald and R. M. W. Dixon

Published in association with the series

Areal Diffusion and Genetic Inheritance

Problems in Comparative Linguistics

edited by

Alexandra Y. Aikhenvald and R. M. W. Dixon

Areal Diffusion and Genetic Inheritance

Problems in Comparative Linguistics

edited by

ALEXANDRA Y. AIKHENVALD

and

R. M. W. DIXON

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Singapore Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States

by Oxford University Press Inc., New York

© Editorial matter and organization: A.Y. Aikhenvald and R. M. W. Dixon 2001

© Individual chapters: their authors 2001

The moral rights of the authors have been asserted

Database right Oxford University Press (maker)

First published 2001

First published in paperback 2006

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organizations. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose this same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data applied for

Typeset by Cambrian Typesetters, Frimley, Surrey

Printed in Great Britain

on acid-free paper by

Biddles Ltd., King's Lynn, Norfolk

ISBN 0-19-829981-8 978-0-19-829981-3

ISBN 0-19-928308-7 (Pbk.) 978-0-19-928308-8 (Pbk.)

1 3 5 7 9 10 8 6 4 2

Contents

<i>Preface</i>	ix
<i>List of contributors</i>	x
<i>List of abbreviations</i>	xiii
1 INTRODUCTION	1
<i>Alexandra Y. Aikhenvald and R. M. W. Dixon</i>	
1 Types of similarity	1
2 Family trees	4
3 Punctuated equilibrium	9
4 Linguistic areas and areal diffusion	11
5 Overview of the volume	19
6 Prospects	23
References	23
2 ARCHAEOLOGY AND THE HISTORICAL DETERMINANTS OF PUNCTUATION IN LANGUAGE-FAMILY ORIGINS	27
<i>Peter Bellwood</i>	
1 Introduction	27
2 The archaeological record of early agriculture—some key points	34
3 A biological comment	40
4 Conclusions	41
References	42
3 AN INDO-EUROPEAN LINGUISTIC AREA AND ITS CHARACTERISTICS: ANCIENT ANATOLIA. AREAL DIFFUSION AS A CHALLENGE TO THE COMPARATIVE METHOD?	44
<i>Calvert Watkins</i>	
References	63
4 THE AUSTRALIAN LINGUISTIC AREA	64
<i>R. M. W. Dixon</i>	
1 Introduction	65
2 Characteristic features	70
3 Two parameters of variation and cyclic change	71
4 Low-level subgroups and small linguistic areas	83
5 Conclusions	87
Appendix—The ‘Pama-Nyungan’ idea	89
Summary list of languages (including likely low-level genetic subgroups)	99
References	102

5	DESCENT AND DIFFUSION: THE COMPLEXITY OF THE PILBARA SITUATION	105
	<i>Alan Dench</i>	
1	Introduction	105
2	Phonological innovations	114
3	Morphophonemic alternations	118
4	Case-marking patterns	120
5	Conclusion	130
	References	132
6	CONTACT-INDUCED CHANGE IN OCEANIC LANGUAGES IN NORTH-WEST MELANESIA	134
	<i>Malcolm Ross</i>	
1	Background	134
2	Equilibrium under the microscope	139
	References	161
7	AREAL DIFFUSION, GENETIC INHERITANCE, AND PROBLEMS OF SUBGROUPING: A NORTH ARAWAK CASE STUDY	167
	<i>Alexandra Y. Aikhenvald</i>	
1	Introduction	167
2	The Arawak family	170
3	Arawak languages north of the Amazon: grammatical and lexical comparisons	174
4	Case studies in restructuring north of the Amazon	176
5	Final remarks	190
	References	191
8	LINGUISTIC DIFFUSION IN PRESENT-DAY EAST ANATOLIA: FROM TOP TO BOTTOM	195
	<i>Geoffrey Haig</i>	
1	Introduction	195
2	The languages and the area	196
3	Pan-Anatolian structural parallels	200
4	Turkish–Laz contact and Turkish–Iranian contact: the issue of structural compatibility	210
5	Conclusions: patterns of borrowing and borrowing of patterns	217
	References	222
9	THE ROLE OF MIGRATION AND LANGUAGE CONTACT IN THE DEVELOPMENT OF THE SINO-TIBETAN LANGUAGE FAMILY	225
	<i>Randy J. LaPolla</i>	
1	Introduction	225
2	The migrations and their effects	227
3	Metatypy	242
4	Conclusion	245
	References	246

10	ON GENETIC AND AREAL LINGUISTICS IN MAINLAND SOUTH-EAST ASIA: PARALLEL POLYFUNCTIONALITY OF 'ACQUIRE'	255
	<i>N. J. Enfield</i>	
1	Introductory discussion: the Mainland South-East Asian area	256
2	Case study: polyfunctionality of ACQUIRE in Mainland South-East Asia	268
3	Discussion	279
4	Conclusion	287
	References	288
11	GENETIC VERSUS CONTACT RELATIONSHIP: PROSODIC DIFFUSIBILITY IN SOUTH-EAST ASIAN LANGUAGES	291
	<i>James A. Matisoff</i>	
1	Theoretical issues in establishing genetic relationship	292
2	Recognized language families of South-East Asia	295
3	Areal features in South-East Asia	298
4	Syllable structure and tone	303
5	Typology of Tibeto-Burman tone systems	306
6	Mono- versus polygenesis of tone in Sino-Tibetan and Tibeto-Burman	313
7	Tonogenetic parallels in South-East Asian languages: the Sinospheric <i>Tonbund</i>	315
8	Theoretical implications and desiderata for the future	322
	References	323
12	LANGUAGE CONTACT AND AREAL DIFFUSION IN SINITIC LANGUAGES	328
	<i>Hilary Chappell</i>	
1	The comparative method and reconstruction of Sinitic	328
2	Typological features of Sinitic	330
3	Language contact: stratification, hybridization, and convergence	337
4	Shared grammaticalization pathways in Sinitic, areal diffusion, and language universals	343
5	Conclusion	353
	References	354
13	AREAL DIFFUSION VERSUS GENETIC INHERITANCE: AN AFRICAN PERSPECTIVE	358
	<i>Gerrit J. Dimmendaal</i>	
1	Introduction	358
2	Two cases of areal diffusion	359
3	Diffusion versus genetic inheritance in Niger-Congo	365
4	Some answers and some further questions	387
	Appendix—The subclassification of Nilo-Saharan according to Greenberg (1963)	389
	References	389

14	CONVERGENCE AND DIVERGENCE IN THE DEVELOPMENT OF AFRICAN LANGUAGES	393
	<i>Bernd Heine and Tania Kuteva</i>	
1	Introduction	393
2	Language contact	394
3	Grammaticalizing metatypy	401
4	Conclusions	409
	References	410
15	WHAT LANGUAGE FEATURES CAN BE 'BORROWED'?	412
	<i>Timothy Jowan Curnow</i>	
1	Introduction	412
2	'Borrowing'	413
3	Scales of adoptability, hierarchies, and constraints	417
4	Impediments to the development of constraints on borrowing	419
5	What language units are borrowed? A summary	425
6	Conclusions	433
	References	434
	<i>Index of authors</i>	437
	<i>Index of languages and language families</i>	443
	<i>Subject index</i>	451

Preface

This volume includes revised versions of fourteen of the sixteen presentations at the International Workshop on ‘The connection between areal diffusion and the genetic model of language relationship’, held at the Research Centre for Linguistic Typology, at the Australian National University, 17–22 August 1998. (RCLT relocated to La Trobe University in Melbourne from January 2000.) The ‘position paper’ was Dixon’s essay *The Rise and Fall of Languages* (1997). In addition, participants were asked to address a number of questions, which are now incorporated into the Introduction to this volume.

All of the authors have experience in the intensive investigation of languages (in many cases, on the basis of fieldwork), as well as in dealing with historical comparative issues and problems of areal diffusion. They all work within the established methodology of historical linguistics. (We have omitted from this volume any discussion of unsubstantiated and unsubstantiable hypotheses of long-range comparison—Nostratic, Amerind, and the like.)

We thank all of the authors included here, for taking part in the Workshop, for getting their chapters in on time, for revising them according to recommendations of the editors and of the publisher’s referees, and for completing their revisions on schedule.

We are also grateful to Jennifer Elliott, Administrator of the Research Centre for Linguistic Typology, who organized the workshop with her normal care and efficiency. Jenny Bourne prepared a collated list of abbreviations. Anya Woods compiled the indices in an exemplary manner, and Tonya Stebbins checked the proofs with diligence.

Hilary Chappell thanks Professor S. A. Wurm for permission to reprint, in Chapter 12, a map from his *Language Atlas of China*.

Contributors

Alexandra Y. Aikhenvald

Research Centre for Linguistic Typology
Institute for Advanced Study
La Trobe University
Vic, 3086
Australia
e-mail: a.aikhenvald@latrobe.edu.au

Peter Bellwood

Department of Archaeology & Anthropology
Faculty of Arts
The Australian National University
Canberra, ACT, 0200
Australia
e-mail: peter.bellwood@anu.edu.au

Hilary Chappell

Linguistics Program
La Trobe University
Vic, 3086
Australia
e-mail: h.chappell@latrobe.edu.au

Timothy Jowan Curnow

Research Centre for Languages and Cultures Education
School of International Studies
University of South Australia
Magill SA 5072
Australia
e-mail: tjcurnow@ozemail.com.au

Alan Dench

Department of Linguistics
University of Western Australia
Nedlands, WA, 6907
Australia
e-mail: adench@cyllene.uwa.edu.au

Gerrit J. Dimmendaal

Institut für Afrikanistik
Universität zu Köln
Meister Eckehart-Strasse 7

D-50923 Köln
Germany
e-mail: gerrit.dimmendaal@uni-koeln.de

R. M. W. Dixon

Research Centre for Linguistic Typology
Institute for Advanced Study
La Trobe University
Vic, 3086
Australia
no e-mail

N. J. Enfield

Language and Cognition Group
Max Planck Institute for Psycholinguistics
PB310 AH-6500 Nijmegen
The Netherlands
e-mail: nick.enfield@mpi.nl

Geoffrey Haig

Seminar für Allgemeine und Vergleichende Sprachwissenschaft
Universität Kiel
Olshausenstrasse 40
D-24098 Kiel
Germany
e-mail: haig@email.uni-kiel.de

Bernd Heine

Institut für Afrikanistik
Universität zu Köln
Meister Eckehart-Strasse 7
D-50923 Köln
Germany
e-mail: bernd.heine@uni-koeln.de

Tania Kuteva

Heinrich-Heine-University Düsseldorf
Anglistics III
Geb. 23.21.01
Universitätsstrasse 1
40 225 Düsseldorf
e-mail: tanja.koutewa@uni-koeln.de

Randy J. LaPolla

Program of Linguistics
Faculty of Humanities and Social Sciences
La Trobe University

Vic, 3086
Australia
e-mail: R.LaPolla@latrobe.edu.au

James A. Matisoff
Department of Linguistics
University of California, Berkeley
Berkeley, CA, 94720
USA
e-mail: matisoff@socrates.berkeley.edu

Malcolm Ross
Department of Linguistics
Research School of Pacific and Asian Studies
The Australian National University
Canberra, ACT, 0200
Australia
e-mail: malcolm.ross@anu.edu.au

Calvert Watkins
Department of Linguistics
Harvard University
Boylston Hall 314
Cambridge, MA, 02138
USA
e-mail: watkins@fas.harvard.edu

Abbreviations

1	1st person
2	2nd person
3	3rd person
A	transitive subject
ABL	ablative
ABS	absolutive case
ACC	accusative
ADJ	adjective
ADP	adposition
ADV	adverb
AGT	agent
<i>AJL</i>	<i>Australian Journal of Linguistics</i>
AN	Austronesian
AOR	aurist
APPL	applicative
ART	article
ASP	aspectual marker
ATR	advanced tongue root
ATTR	attributive
aug	augmented
AUX	auxiliary
BENF	benefactive
BM	Bora-Muinane
BP	before present
BW	Bora-Witoto
C	consonant
CA	Common Anatolian
<i>CAAAL</i>	<i>Computational Analyses of Asian and African Languages</i>
CAUS	causative
CL	(numeral) classifier
CLNK	clause linker
COMP	complement clause
COND	conditional
CONJ	conjunction
CONS	consecutive marking
CONTIN	continuous
COP	copula
CRS	currently relevant state

DAT	dative
DEM	demonstrative
DET	determiner
DIMN	diminutive
du	dual
EC	extent complement
ECHO	echoic
ERG	ergative
exc	exclusive
EXP	experiential aspect
F, f	feminine
FP	focus particle
FUT	future
GEN	genitive (or associative) marker
HAB	habitable
IA	Indo-Aryan
IE	Indo-European
IMP	impersonal
IMPFV	imperfective
IMPV	imperative
INAN	inanimate
inc	inclusive
INDF	indefinite
INF	infinitive
INTERR	interrogative
IRR	irrealis
<i>JALL</i>	<i>Journal of African Languages and Linguistics</i>
<i>JCL</i>	<i>Journal of Chinese Linguistics</i>
L	linker (relative clause or attributive marker)
L	lateral consonant
LB	Lolo-Burmese
LOC	locative (marker)
LP	linking particle
<i>LTBA</i>	<i>Linguistics of the Tibeto-Burman Area</i>
m	masculine
MC	manner complement
min	minimal
MK	Mon-Khmer
MOD	modal
MSEA	Mainland South-east Asia
MSC	Modern Standard Chinese
N	nasal, noun
NEG	negative

NEGV	negative verb of possession and existence ('not have, there is/are not')
NE, nf	non-feminine
NOM	nominalizer
NP	noun phrase
NPOSS	non-posessed
n-sg	non-singular
O	direct object of transitive verb
o.bro	older brother
OBL	oblique
OL	<i>Oceanic Linguistics</i>
PAN	Proto-Austronesian
PAr	Proto-Arawak
PART	participle
PBM	Proto-Bora-Muinane
PCL	particle
PFV	perfective
PHO	Proto-Hipode-Witoto-Ociana
PIE	Proto-Indo-European
PK	Proto-Karen
PL, pl	plural
PLB	Proto-Lolo-Burmese
PMC	potential manner complement
POSS	possessive
POT	potentialis
PRC	potential result complement
PREP	prepositional
PRES	present
PRES.EYEW	present eyewitness
PRETR	pretransitive or disposal marker of patient noun
PROG	progressive
PRON	pronoun
PRT	particle
PTB	Proto-Tibeto-Burman
PW	Proto-Witoto
PWOc	Proto-Western Oceanic
Q	question
RDP	reduplicated
RECIP	reciprocal
REF	referential
REFL	reflexive
REM.PAST	remote past
RLS	realis
RPE	result of prior event

S	intransitive subject
S	stop consonant
S _a	subject of an intransitive active verb
SG, sg	singular
S _o	subject of an intransitive stative verb
ST	Sino-Tibetan
SUB	subordinate
SUBJ, Subj	subject
TAM	tense and/or aspect and/or mood
TB	Tibeto-Burman
TC	temporal complement
TOP	topic marker
UA	Uto-Aztecan
V	vowel, verb
VB	verb
VP	verb phrase
WB	Written Burmese
WSem	West Semitic
WT	Written Tibetan
y.sib	younger sibling

Introduction

Alexandra Y. Aikhenvald and R. M. W. Dixon

This volume consists of studies of the relationship between areal diffusion and the genetic development of languages from a number of critical parts of the world. In part, they follow up on some of the ideas in Dixon's essay *The Rise and Fall of Languages* (1997)—including the punctuated equilibrium model for language development—although in fact they range considerably beyond this.

The chapters cover Ancient Anatolia, Modern Anatolia, Australia, Amazonia, Oceania, South-East and East Asia, and Sub-Saharan Africa. We did not feel it necessary to commission specific discussions of South Asia, or North and Central America, or of the Balkans, since there are already excellent studies of these areas—in Masica (1976, 1991), Sherzer (1973, 1976), Campbell, Kaufman, and Smith-Stark (1986), Campbell (1997), and Joseph (1983). In addition, Chapter 2 gives an archaeologist's view on what may have triggered the punctuation of cultural and linguistic periods of equilibrium. The final chapter provides a conspectus on the kinds of linguistic feature that can be borrowed, drawing together the strands from earlier chapters. In this Introduction, we outline the parameters which underlie discussion in the volume, and comment on some of the recurrent conclusions of contributors; for example, the inadequacy of 'family tree' as the only (or as the main) means of describing relationships between languages.

1. Types of similarity

Two languages can resemble each other (a) in the categories, constructions and types of meaning they use; and (b) in the forms they employ to express these.

There are a number of kinds of explanation for similarities of types (a) and (b):

- (i) UNIVERSAL PROPERTIES OR TENDENCIES. Concerning (a), every language has a marker of clausal negation (but not every language has a distinct strategy for negating a predicate argument, for instance). With respect to (b), very many languages have a verb 'blow' that has iconic form, with a bilabial stop, often aspirated, plus a high back vowel (prototypically p^hu-).

We are grateful to Hilary Chappell, Alan Dench, and Nicholas J. Enfield for their comments on a draft of this chapter, which helped us to improve it.

- (ii) CHANCE. For (b), we can note that there are occasional coincidences of meaning between forms in different languages, which are notable by their very rarity. For instance, in the Australian language Mbabaram, the word for 'dog' is *dog*, [dɔk⁷]; the modern English form goes back to Old English *docga*, whereas the Mbabaram form goes back to *gudaga* (see Dixon 1991: 361–2 for an account of the regular sound correspondences involved). For (a), we can mention that shape-based gender is, coincidentally, found in languages from Africa and from New Guinea (see Aikhenvald 2000: 277).
- (iii) BORROWING OR DIFFUSION. Two languages in contact—where a significant proportion of the speakers of one also has some competence in the other—gradually become more like each other. The most pervasive borrowing generally involves (a), construction types, grammatical categories, and organization of lexical and grammatical meaning; these kinds of features steadily diffuse from one language to another. For example, if a language with no noun classes (or genders) moves into contact with one or more languages that have the category noun class, then it is likely to develop its own set of noun classes; most frequently it will achieve this not by borrowing the forms for marking noun classes from a neighbouring language, but by developing them from its own internal resources (see Aikhenvald 2000: 383–91). That is, it is just the category which is borrowed, not the forms used to mark it.

There can also, of course, be borrowing of (b) lexical forms, and—to a lesser extent—of some grammatical forms. Note though that this varies from culture to culture. Aikhenvald (in Chapter 7) describes how one Arawak language, Tariana, has a prohibition against borrowing forms from its neighbours, whereas Resígaro, an Arawak language spoken in a different region, borrows them freely.

Heine and Kuteva, in Chapter 14, examine comparative and reflexive constructions in languages from Africa and from other parts of the world. They conclude: 'it may happen that people borrow a comparative or reflexive morpheme from another language [our (b)] but . . . they are more likely to borrow conceptual templates [our (a)], like event schemas, to develop a new comparative or reflexive category.'

There may also be diffusion of phonetic and phonological characteristics. This often comes about through the borrowing of forms, but may not necessarily do so (see Aikhenvald 1996). Several of the contributors to this volume affirm that prosodic or suprasegmental features—such as tone and nasalization—are more likely to diffuse than segmental phonemes.

Another relevant point is that two languages may show similarities which are due to borrowing, but not from each other; they may each have borrowed from a third language. Tosco (2000) reports that a number of Semitic languages from Ethiopia show striking similarities due to their all

having a substantial substratum from a number of closely related languages of the Agaw subgroup of Cushitic.

- (iv) GENETIC RETENTION. If two languages descend from the same ancestor then they are likely to have similar categories, and meanings expressed by similar forms. In order for some point of similarity to be recognized as a mark of genetic affiliation it must be of type (b). That is, the forms and their meanings must be either identical or else easily relatable, through established rules for phonological change and semantic change in the languages (in terms of general theories of what types of phonological change and of semantic shift are possible).

Note the difference between (iii) and (iv). A similarity of type (a)—a construction, a category, or a way of organizing meaning—can be due to diffusion, as can (b), similarities concerning forms with the same meaning. But a similarity that is genetically significant must be of type (b); it must involve forms.

Some people have, in the past, noticed typological similarities—of type (a)—between a number of languages and taken them to be an indication of genetic affiliation. This is quite illicit. Dixon (1997: 31–2) reviews examples involving Japanese and Ural-Altaic, and similar cases involving African languages. Dimmendaal (in §3 of Chapter 13 below) discusses Greenberg's early classification of 'Ijoid, as well as several of the groups now classified as part of Benue-Congo' and comments: 'It is probably fair to state that their inclusion within Kwa was motivated to some extent by the observed typological similarity with languages still classified under Kwa today. For example, these various languages share such features as ATR-vowel harmony, nasalized vowels, reduced noun-class systems, and serial-verb constructions.' (See further comments in §2 below on the diffusibility of these features.)

- (v) PARALLEL DEVELOPMENT (OR CONVERGENT DEVELOPMENT). In §1 of Chapter 4, Dixon explains, with examples, how 'two languages (often, but not always, two languages of the same genetic group) may share an inner dynamic that propels them to change, independently, in the same way'. Sapir (1921: 171–8) discusses 'parallelism of drift', commenting that 'the momentum of the more fundamental, the pre-dialectic, drift is often such that languages long disconnected will pass through the same or strikingly similar phases'; he provides illustration from English and German. LaPolla (1994) discusses a number of examples from Tibeto-Burman languages, including the development of classificatory existential verbs, a set of similar innovations that have taken place independently in a number of genetically related languages. Another example would be the dissimilation of aspirated stops (Grassmann's Law) in both Greek and Sanskrit, which took place independently in each language, long after their genetic separation.

We have said that, associated with the idea of ‘parallel development’ is the fact that each of a group of languages may share an ‘inner dynamic’ that leads to the potentiality for a certain direction of change. If one language does change in this way, its neighbours are then likely to emulate the change; that is, the change diffuses through all languages in the group. Neighbouring languages were on the point of initiating the change in their own right, and are thus open to accept it by diffusion. (One example would be the development of bound pronouns in Australian languages, discussed by Dixon in Chapter 4.) Suppose that a certain change develops in languages A, B, and C. This might be taken as evidence that they are genetically related, with the change attributed to a common proto-language. In fact, the change might have developed in just one of A, B, and C, and then diffused into the others; such a shared change would provide no evidence of close genetic connection (e.g. subgrouping) for A, B, and C.

The ‘parallel development’ explanation for some kinds of similarity between languages is not always paid attention to (including by some of the contributors to this volume). As a consequence, similarities of this kind may—mistakenly—be taken to be markers of close genetic relationships.

The hardest task in comparative linguistics is to distinguish between these five kinds of similarity, and then to assess them. In Chapter 5 below, Dench provides a masterly summary of the problem. He states in §1.2: ‘Of course, making the argument for an innovation shared by virtue of a period of common development is never easy. I take it for granted that a statement of shared inheritance as explanation for a shared feature should only be made once all other possible explanations for the shared feature have been exhausted. These other possibilities will include accidental similarity in form, borrowing, and genetic drift.’

Dench then goes on to say: ‘We should leave open the possibility that all questions may turn out to be undecidable. It may not be possible to show conclusively for any particular innovation that it results from genetic inheritance rather than that it is motivated by contact with another language. If enough such cases occur, then the suspicion we might attach to any putative inherited innovation will mount and we should become increasingly sceptical of any suggested genetic classification.’

2. Family trees

The ‘family tree’, as a metaphor for relationship between languages, was popularized by August Schleicher (1861–2). It has, ever since, been the prevailing model within Indo-European linguistics, despite some opinions that it is not an appropriate model, or not a sufficient model. For example, Johannes Schmidt (1872) suggested a ‘wave’ metaphor to account for the spread of isoglosses across well-established language boundaries, and especially cases of modifying influence of foreign

languages on grammatical systems. (In contrast to this, Baudouin de Courtenay (1930) protested against the habit of treating languages as if they were ‘living organisms’ or phenomena of nature, independent of their speakers; that is, comparing them to a plant, as in family trees, or to a flowing liquid, as in wave theory.)

The family tree represents similarities of type (iv), due to genetic retention. But what about those of type (iii), borrowing or diffusion, or of type (v), parallel development? (We can leave aside (i) universal properties or tendencies, and (ii) chance, as either relatively uncommon or else easily discernible.) In Chapter 3 below, Watkins suggests ‘even the much-maligned family-tree model has a perfectly good notation for areal or other “influence”, the dotted line of the classical manuscript stemma which is the source of the family tree’. (See Emeneau 1967: 371 for an illuminating three-dimensional model of genetic and areal relationships involving the South Dravidian languages.)

As many of the contributors to this volume demonstrate, something more fine-grained is needed than a diagram with solid lines for genetic filiation and dotted lines for areal influence, if we are to model the multitude of diverse ways in which languages develop and influence each other. For example, Chappell concludes, at the close of a discussion of the situation in Sinitic languages in Chapter 12: ‘To reconstruct the history of a language family adequately, a model is needed which is significantly more sophisticated than the family tree based on the use of the comparative method. It needs to incorporate the diffusion and layering process as well as other language-contact phenomena such as convergence, metatypy, and hybridization. The desideratum is a synthesis of all the processes that affect language formation and development.’

Similar reservations are presented in half a dozen other chapters, relating to the language situations in South-East Asia, Africa, Amazonia, and Australia. In Chapter 5, Dench describes a close-knit group of languages in Western Australia where the distinction between inherited and diffused similarities is particularly hard to discern. Here the dotted lines can be drawn in but whether there are any solid ones is a matter for conjecture.

Other metaphors have been suggested. Watkins (Figure 3 in Chapter 3) suggests—as he says, ‘fancifully’—‘a sort of “cyclone” image of the diaspora of Indo-European languages’. Shevelov (1964: 611–12) verges on the poetic in opining that: ‘The disintegration of [Common Slavic] did not resemble the growth of a tree Nor can this disintegration be grasped in the traditional metaphor of waves spreading one after the other. If a metaphor is appropriate, the most suitable would be the image of clouds in the sky on a stormy day, with their constant changes in shape, their building up, overlapping, merging, separating, and their ability to vanish in an instant.’ However, no criteria are suggested for identifying types of linguistic ‘cloud’, and then plotting and quantifying their movement and interaction.

In Chapter 11, Matisoff declares that the family tree betokens ‘a vast oversimplification’. He goes on to say: ‘Languages rarely split off cleanly from their relatives.

A much more appropriate image for what one finds in linguistic areas like South-East Asia might be the “thicket”, an impenetrable maze of intertwined branches. Instead of clear-cut migrations of population groups, one finds slow “percolations” or “filtrations” of small groups of people.’ Matisoff is simply emphasizing that a simple branching tree is an inadequate model. He does not seriously suggest a ‘thicket diagram’ (with every twig identified) as a workable model. The point is that a family-tree-like diagram does not adequately demonstrate the many kinds of historical and current relationships between languages. It may well be suitable for some situations, but is simplistic and misleading for others and should not then be employed. The same point is well illustrated in Chapter 9, where LaPolla provides an illuminating account of migrations and shifting patterns of diffusion within the Sino-Tibetan family.

It is often easier to prove that a set of languages form a genetic unit (a language family) than it is to establish subsidiary genetic units (subgroups) within a family. In Chapter 7, Aikhenvald describes how languages of the Arawak family are spread over a dozen or so separate geographical regions. The small set of Arawak languages in a certain region are likely to share certain traits, but caution should be exercised before taking these as evidence of a subgrouping. The traits may also be found in other, non-Arawak, languages of the region and constitute areal features. As Dench emphasizes, all other possible explanations for a point of similarity should be examined, and dismissed, before concluding that it is a genetic retention. Thus, to obtain a full characterization of the genetic links within a language family, it is necessary to have information not only about all languages in the family, but also concerning all their present (and, ideally, all their past) neighbours.

LaPolla comments, at the end of Chapter 9: ‘Those who do subgrouping . . . often do not give the reasons for their groupings. In some cases there are clear isoglosses, but often subgrouping is affected by the author’s subjective “feel” of the language, shared features, or shared vocabulary, which are all often influenced by its geographical location.’

Indo-European is—in several respects—regarded as the prestige language family. For several centuries the languages and their speakers held a prestige position in the world. And Indo-European has attracted scholars of the highest quality, so that the results obtained have considerable scholarly worth. Work on relationships between Indo-European languages has—justly—been held up as a model of how to do things, and imitated by linguists working on languages from other parts of the world.

The dominant motif in Indo-European studies is the family tree, although generally hedged by caveats and annotations. The family-tree metaphor has been taken over for other parts of the world in stark form, often as the sole model for relationships between languages.

Rather than asking whether a form of family tree is appropriate to the language situation in some newly studied region, it has often been simply *assumed* that it is. What began as a metaphor has been ascribed reality, and has acted to constrain

enquiry along narrow lines. This can lead at best to a partial and at worst to a mistaken statement of language relationships.

Once it had been assumed that the family tree (in its simplest form) was everywhere applicable, attention turned to how to discover what the family tree is for a given group of languages—and how to do this as quickly as possible. Thus came into being ‘lexicostatistics’, where a ‘family tree’ could be inferred by simply comparing 100 or 200 words of ‘core vocabulary’. Terminology was now being used in a quite new way. Originally, a set of languages was recognized as a language family if a shared proto-language could be reconstructed, together with the systematic changes by which each modern language developed from this. In lexicostatistics, a set of languages is recognized as a ‘family’ if they share between 36% and 81% core vocabulary. And so it goes on, ever upwards: languages sharing 12–36% are a stock, those with 4–12% are a microphylum, those with 1–4% are a mesophylum, and those with less than 1% are a macrophylum. Associated with all this is ‘glottochronology’, which purports to supply dates for nodes on the ‘family tree’ (Gudschinsky 1956).

It did not take long to expose this as unsupportable. It depended on a set of premisses all of which are without foundation: that one can infer genetic relationship from lexicon alone (a careful study of Indo-European work reveals that similarities of grammatical form are of primary importance); that the lexicon of all languages is always changing at a constant rate (there is in fact considerable variation, depending on social attitudes, types of language contact, and so on); and that core vocabulary is always replaced at a slower rate than non-core (this applies in some parts of the world, but not in Australia, Amazonia, and New Guinea, for example—see Chapters 4 and 7). In most instances, dating the prehistory of languages is a speculative endeavour.

Only in Australia did lexicostatistics engender lasting damage. A ‘family tree’ was constructed, supposedly on lexicostatistic principles (although the sources used and percentage scores were never stated, and in fact the actual percentages of shared vocabulary between languages do not, in many cases, accord with the classification—see the Appendix to Chapter 4). This is now widely accepted, as something on a par with the Indo-European family tree, although no justification has been provided (and none could be, for most of the supposed genetic groups—see Dench’s incisive discussion in Chapter 5). In other parts of the world lexicostatistics has been discredited, in the way that false procedures always are (see Dixon 1997: 35–7 and references therein).¹

¹ In §2.2 of Chapter 14, Heine and Kuteva show how a lexicostatistic ‘family tree’ for Nubian differs significantly from a genetic classification by traditional means. They then add: ‘In general, lexicostatistics has turned out to be a fairly reliable tool for establishing first hypotheses on genetic relationship in Africa. In most cases where the comparative method and lexicostatistics have been employed they yielded similar results.’ We must bear in mind that few of the putative genetic groups suggested for African languages have been proved, by application of the comparative method. It will be interesting to see whether Heine and Kuteva’s opinion about lexicostatistics stands up when more detailed comparative reconstruction has been completed.

Discussing work on African languages, Dixon (1997: 21–3) states: ‘after reviewing the available literature an outsider is forced to conclude that the idea of genetic relationship and the term “language family” are used in quite different ways by Africanists and by scholars working on languages from other parts of the world’. Both Dimmendaal, in Chapter 13, and Heine and Kuteva, in Chapter 14, present the received opinion that there are in Africa four major ‘language families’ or ‘genetically defined units’: Niger-Congo, Nilo-Saharan, Afroasiatic, and Khoisan. In fact Khoisan is regarded—by those scholars who have studied it in most detail—as a linguistic area consisting of several distinct genetic families (see, among others, Westphal 1962, 1971; Köhler 1974). Afroasiatic does appear to be one genetic unit, although full justification for this has yet to be published in integrated form. There are wide divergences of opinion concerning the status of Nilo-Saharan.

For Niger-Congo all we can say is that there has so far been no principled justification for this as a language family. As Dimmendaal states, in his §3, ‘by the criteria of regular sound correspondences among these languages and of the reconstruction of proto-forms, Niger-Congo is not a proven genetic unity’. It may well be that, as Heine and Kuteva say, ‘Greenberg’s genetic classification of African languages is by now widely accepted’; but ‘being widely accepted’ (sc. among Africanists) does not equate with ‘has been scientifically justified’.

Dimmendaal presents an insightful evaluation of a number of features that have been said to characterize Niger-Congo languages: cross-height vowel harmony, nasalized vowels, labial-velar stops, serial verbs, and noun classes. Each is found in some, but not all, Niger-Congo languages, and in some languages from non-Niger-Congo groups. For all but one of these features Dimmendaal suggests that it could have spread by diffusion, rather than being a retention from a proto-language (to this list he adds tone, as an ‘ancient diffusional trait, covering major parts of the continent’). The exception is noun classes, which he suggests are a feature diagnostic of Niger-Congo as a genetic unit. But, as mentioned under (iii) in §1 above, the category of noun classes is one of the most easily diffusible in other parts of the world; indeed, Dimmendaal describes how the Nilo-Saharan language Luo has gained noun classes by diffusion from Niger-Congo languages. Noun classes in Niger-Congo may well be a genetic retention; but a fair number of other sure genetic retentions must also be found, if Niger-Congo is to be substantiated as a genetic entity.

At the end of §1 we quoted Dench’s methodological principle that only after ‘all other possible explanations for the shared features have been exhausted’ should they be taken to be a genetic inheritance. The opposite procedure appears to have been followed by Africanists. At a first stage, every sort of resemblance was taken as indicative of genetic connection; many of these are in fact typological or areal similarities.

In Chapter 14, Heine and Kuteva conclude that ‘contact-induced language change and the implications it has for language classification in Africa are still

largely a *terra incognita*?. Serious studies must be pursued, of the nature of similarities between African languages, in terms of the principles set out in §1 above. Ideally, received ideas on genetic classification should be put to one side, and only returned to once investigations of similarities due to diffusion, and those due to parallel development, are well advanced.

3. Punctuated equilibrium

Languages are always changing. Over time, two languages that have sprung from a single source become more and more dissimilar so that, eventually, it is not possible to discern that they did have a common ancestor. That is, there are limits on how far back one can discern a genetic link, and how far back one can reconstruct. In Chapter 3, Watkins suggests a maximum date of 8,000 or 10,000 years. In fact the dates commonly quoted for proto-languages are a little shorter than this: 6,000 or 7,000 years for Proto-Indo-European, about 5,000 years for Proto-Austronesian, and so on.

What happened before that? Were there family trees upon family trees? Have all languages been splitting, in the manner of Indo-European languages, ever since humankind developed language, which is generally acknowledged have been at least 100,000 years ago (many would prefer an earlier date)? The Indo-European family tree has produced just over 100 (10^2) languages from one in, say, 7,000 years. On this principle—since there are about 14 periods of 7,000 years within 100,000 years—a single language spoken about 100,000 years ago should have given rise to $10^{2 \times 14} = 10^{28}$ (that is, ten billion billion billion) modern descendents. This does not accord with the facts—something else must have happened.

Dixon's essay *The Rise and Fall of Languages* (1997) suggests that, in the history of the human race, there have been long periods of cultural and linguistic equilibrium, when the number of languages in a given geographical region would have remained relatively constant. Every so often, there will be a punctuation whereby one ethnic group (and its language) expands and spreads and splits. Bellwood, in Chapter 2, summarizes some of the main points of Dixon's hypothesis, and suggests that the introduction of agriculture is likely to have been the trigger for punctuations that led to many of the major language families in the world today.

A family-tree diagram, with the addition of various kinds of annotation, is an appropriate model for a period of punctuation. The family tree essentially shows how languages split (note that there is always some concomitant diffusion, both between languages within a certain family and between languages of different families). Eventually, after the punctuation slows to a halt, an equilibrium state comes into being, within a circumscribed geographical area. This is characterized by steady diffusion of cultural and linguistic traits across the area. During a period of punctuation (which is likely to last for just a few hundred or maybe a few thousand years) languages in a given family diverge from a common proto-language. During a period of equilibrium (which may prevail for thousands or even tens of

thousands of years) languages in a given area converge towards a common proto-type.

Dixon's idea was motivated partly by his lifelong preoccupation with trying to understand the Australian linguistic area, which may have been in existence for as long as Aboriginal people have been in the continent (at least 40,000 and perhaps 50,000 years). Dench, in Chapter 5, provides a fine-grained discussion of just one portion of Australia, demonstrating the pervasiveness of diffusion. Dixon's and Dench's point is that although many isoglosses can be recognized, they do not cluster, which would be needed if they were to have genetic significance.

If received ideas about genetic groupings in Africa could be temporarily set aside, it would be profitable to study the isoglosses across that continent, and to investigate their clustering. After all, humankind is believed to have originated in Africa, and languages have surely been spoken there for as long as (or longer than) anywhere else in the world.

There are various ways in which languages can become more like their neighbours, due to diffusion. These apply particularly in regions showing a state of linguistic equilibrium, but can also apply in other circumstances. Ross (1996, 1997) coined 'metatypy' in order to capture in a single term what had previously been described as gradual convergence of languages, characterized by a tendency towards structural and semantic isomorphism, and 'linear alignment' of morphological structure. Classic descriptions of this include Gumperz and Wilson (1971) and Nadkarni (1975), each a study of a community where both Indo-Aryan and Dravidian languages are spoken. The notion of metatypy is utilized by Ross in Chapter 6, by LaPolla in Chapter 9, by Chappell in Chapter 12, and by Heine and Kuteva in Chapter 14.

A diffusion area may be swamped by the punctuational expansion of some major family. But it is possible that enclaves of the original diffusion area may remain, in economically non-advantageous spots. These enclaves may share certain features of the erstwhile equilibrium zone; care should be taken before accepting these as markers of genetic affiliation. Dixon and Aikhenvald (1999a: 17) speculate: '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.' Dixon and Aikhenvald suggest that the scattered languages in north-west Amazonia which are labelled 'Makú' may be such relics; the Makú are hunters/gatherers and they live along small streams in the forest, while the major rivers are inhabited by agricultural people speaking languages from the large Tucano and Arawak families. In this type of circumstance it is virtually impossible to distinguish between similarities due to genetic retention and those which are the result of long-term diffusion.

Another point which recurs in several chapters is that one must have good quality descriptive data on languages in order to make confident decisions concerning

the relationships between them. For many languages—some now extinct, others still spoken—such data are unavailable. Work similar to that by Emeneau (1967) on Dravidian cannot be accomplished for the Arawak languages spoken north of the Amazon, for a simple reason—most of them are extinct, and even for those which are still spoken we will never be able to recover the substratum languages. Similarly, within the Afroasiatic language family, we can only suggest a hypothesis about a putative close genetic relationship between Berber, East-Numidian, and Guanche languages—these hypotheses can never be substantiated simply because East-Numidian and Guanche languages are now dead and were only poorly documented (see Aikhenvald and Militarev, 1991). Linguists should not be afraid to say ‘we don’t know (and perhaps never will)’. This ‘agnostic’ approach is justified by the current linguistic situation across the world, with languages dying quicker than anyone can record them.

4. Linguistic areas and areal diffusion

A linguistic area (or *Sprachbund*) is generally taken to be a geographically delimited area including languages from two or more language families, sharing significant traits (which are not found in languages from these families spoken outside the area). There must be a fair number of common traits and they should be reasonably distinctive (see 1 below).

Linguistic areas differ as to the relationships between the languages. We hypothesize that linguistic areas which arose as the result of equilibrium situations involve long-term language contact with multilateral diffusion and without any developed relationships of dominance. In contrast, areas which were formed as a result of sudden migrations or other punctuations tend to involve dominance of one group over other(s), and the diffusion is often unilateral. However, depending on the historical events, the direction of diffusion can suddenly change (see Johanson 1998, and Haig in Chapter 8 below, on changing directions in Turkic–Iranian influence within East Anatolia); this creates a ‘historically’ multilateral area, every synchronic ‘cut’ of which can be considered unilateral.

Dominance relations result in a severe areal impact from one language onto the other(s), and often leads to an extinction or severe reduction in use of a non-dominant language—see Sasse (1985) on Albanian varieties in a dominant Greek environment, or Aikhenvald (in Chapter 7) on the gradual ousting of Tariana in the Vaupés area and Resigaro in the Bora-Witoto region of South America. The distinction between linguistic areas with a clear-cut two-way contact and those with multilateral contact is artificial (see Matisoff in Chapter 11); what determines the characteristic features of an area is the relationships languages have within it. The chapters in this volume undoubtedly show that a neat linguistic area with unilateral diffusion is an abstraction and a simplification.

A number of questions related to the evaluation of areal phenomena as such are

discussed in this volume. It will be useful to articulate some of them here. Questions 1–3 are concerned with determining a linguistic area and its boundaries.

1. Which properties can be considered ‘area-defining’?
2. How many features do we need for an area to ‘qualify’ as such?
3. How quickly do areas form?

Recognizing the ‘diagnostic traits’ characteristic of languages within an area is obviously crucial. First, this is the only way to determine whether it is an area or not. Second, this helps to delineate which languages of the area are ‘central’ to it, or can be shown to ‘exemplify’ an areal type, and which are ‘marginal’ (see the attempt by van der Auwera (1998) to plot Meso-American features onto individual languages, and similarly for the Balkan situation).

Any typologically well-attested property cannot by itself be considered area-defining. However, the way properties cluster can be area-specific. In a classic paper, Campbell, Kaufman, and Smith-Stark (1986) single out four morphosyntactic features characteristic of the Meso-American area:

- (a) Nominal possession of the type *his-dog the man*.
- (b) Relational nouns (that is, body-part nouns used as spatial markers).
- (c) Vigesimal numeral systems.
- (d) Non-verb-final basic word order which possibly correlates with the absence of switch reference.

An additional feature includes numerous ‘pan-Meso-American’ formations of the type ‘knee’ = ‘head of the leg’, or ‘boa-constrictor’ as ‘deer-snake’.

None of these properties is restricted to Meso-America; it is their clustering that is area-specific. Similarly, none of the properties given for mainland South-East Asia as a linguistic area—see Chapters 10 and 11—is unique; however, the way they cluster and entail each other makes them ‘pan-South-East Asian’. A similar statement can be made about Amazonia (see the list of features given in Aikhenvald and Dixon (1998)); or mainland New Guinea (Ross in Chapter 6). The same holds for Chappell’s analysis, in Chapter 12, of typological properties of Sinitic languages, none of which is only found in Sinitic. Different areal features can be assigned varying weight with respect to their area-defining qualities, and typologically widespread or natural features or development paths have less weight than their ‘exotic’ counterparts (see the discussion of naturalness by Enfield in Chapter 10); however, the clustering of features may be exotic.

The question of how many features are sufficient to delimit an area (see Haig in Chapter 8) relates to the weight of each individual feature. Areas can be created on different levels; that is, in different linguistic situations one might expect varying degrees of diffusion of a given linguistic feature. The great majority of the shared features of Standard Average European are syntactic, while two of the most salient features of the South-East Asian languages are their ‘monosyllabicity’ and their ‘tone-proneness’ (see Matisoff in Chapter 11).

Another parameter for diffusional—or area-defining—properties involves evaluating tendencies for elaborating different mechanisms. Matisoff (in Chapter 11) concentrates on the ‘tone-proneness’ of the languages in Mainland South-East Asia, an area-specific tendency. As he points out (§3.2, and note 5), ‘changes in manner of initial consonants’ had ‘concomitant tonogenetic or registrogenic effect’, while ‘such mutations in the history of Indo-European (e.g. Grimm’s Law or the Second Germanic Sound Shift) have never led to tonogenesis’. And it seems, from Dimmendaal’s mention at the end of Chapter 13, that ‘tone-proneness’ is also a feature of Sub-Saharan Africa as a linguistic area.

Finally, how fast can a linguistic area form? And how fast can it disintegrate? Both questions probably have to be left open, in the present state of knowledge.

Linguistic areas can be relatively young or relatively old. (Note that, once we attempt to go back in time before the advent of written records, any kind of linguistic dating must, by its nature, be speculative.) In the Vaupés region of Amazonia, Tariana and East Tucano languages appear to have been in contact for no more than about four hundred years (see Chapter 7). Apparently, other known linguistic areas of the world—e.g., the Balkans, Arnhem Land in north Australia (see Heath 1978), Mesoamerica (Campbell, Kaufman, and Smith-Stark 1986), South Asia (Masica 1976), and linguistic areas for North American Indian languages north of Mexico (Sherzer 1976), such as the north-west coast (Bright and Sherzer 1976: 234)—are considerably older than this. The defining features of Standard Average European seem to have been in place by the end of the first millennium CE, while their formation in individual languages is hypothesized to go back several centuries earlier (Haspelmath 1998). In India, according to Gumperz and Wilson (1971: 153), the coexistence of Urdu, Marathi, and Kannaḍa goes back about three or four centuries, when the Urdu-speaking Muslims arrived in the region. This is an area with almost complete structural isomorphism (and occasional loans of morphemes). However, it is known that Kannaḍa-speaking and Marathi-speaking people had been in the region for more than six centuries; so the area could be older than three or four centuries.

A linguistic area created under an equilibrium situation would lack any dominance relations, and would, in consequence, be long-lasting. In case of a punctuation—and the ensuing relationship of dominance of one group over another—an area would not last. As a result of intensive contact one language simply ‘wins’ over another; the minor languages fall into disuse and die. In each case we need to know the social conditioning of an area, especially in relation to language attitudes and dominance relationships. As Watkins puts it, in Chapter 3, ‘both genetic families and diffusional areas would have their own distribution of rapid abrupt and slow gradual change’.

The next two questions concerning linguistic areas are related:

4. How do linguistic areas and areally diffused features correlate with social parameters and language attitudes?

5. Is there any 'hierarchy' with respect to which categories are more, and which are less, borrowable? That is, which aspects of linguistic structure are more stable and which are less stable, and which aspects of lexicon, phonology (segmental and suprasegmental), morphology, syntax and discourse structure are most likely—and which are least likely—to be borrowed, retained, or lost?

To answer question 4, we must address the correlations between a linguistic area and a culture area. It is well known that sharing cultural traits does not necessarily entail creation of a linguistic area. For instance, the Great Plains region in North America is recognized as a culture area, but not as a linguistic area, and it has been argued that the languages of the area have not had a long enough time to develop areal traits (Sherzer 1973, Bright and Sherzer 1976: 235). Linguistic borrowings presuppose some sort of interaction between peoples and some degree of knowledge of their languages; that is, at least some degree of bi- and/or multilingualism must be a condition for the creation of a linguistic area or for development of convergence phenomena (see Ross in Chapter 6 on different degrees of knowing a neighbour's language in New Guinea). Note that bilingualism was almost non-existent in the Great Plains, which must have been a major reason why the six language families spoken in the area do not form a linguistic area (Doug Parks, p.c.). The Upper Xingu area in Brazil (Seki 1999) is similar to the Great Plains region in that it is a recognized culture area, but not yet a linguistic one.

A drastically different case is that discussed in Chapter 10 by Enfield; Mainland South-East Asia is a linguistic area, but it is not a cultural area. This illustrates well the problems which might be caused by taking for granted 'that culture areas and linguistic areas will coincide' (see the discussion by Campbell 1997: 340).

An answer to question 5 is likely to depend on the following:

- (a) TYPE OF COMMUNITY. It is useful to distinguish between:

- (i) Communities that are internally tightly knit—bound together by linguistic and other types of solidarity—as opposed to loosely knit—involving a diversity of language and ethnic groups (Andersen 1988, Ross 1994, 1996). In some of the latter there may be an established lingua franca which can in time lead to the development of a more tightly-knit profile.
- (ii) Communities that are externally open (with plentiful social and economic interaction with their neighbours) as opposed to relatively closed.

It is also important to take note of the lifestyle of speakers (e.g. whether nomadic hunters/gatherers, village-dwelling agriculturalists, nomadic cattle herders, or largely urbanized groups); division of labour between sexes and between generations; social organization and kinship system; marriage and residence patterns; and religion/mythology.

-
- (b) **SIZE OF COMMUNITY.** The scale of diffusion within a small community and within a larger community may differ. Small communities are likelier to be more tightly knit than large ones, and as a result to show greater diffusion of linguistic features, or else diffusion at a faster rate.
- (c) **RELATIONS WITHIN A COMMUNITY.** These include hierarchies of prestige groups (castes, etc.) and relations of dominance among languages or dialects. There is, typically, borrowing from a prestige into a non-prestige language, e.g. from Turkish into the variety of Greek spoken in Asia Minor. In some societies, slaves taken during war generally come from another language group; the amount of borrowing they engender is likely to relate to their status within the society, e.g. whether they are allowed to marry non-slaves.
- (d) **CONTACT WITH OTHER COMMUNITIES.** Parameters here relate to whether contact is regular or sporadic; under what circumstances (e.g. trade, sport, religion, marriage patterns), and at what social levels. Interaction is sometimes restricted to written language, e.g. the influence of Classical Arabic on the vernacular languages of Moslem peoples, exclusively through the Koran.
- (e) **DEGREES OF 'LINGUALISM'.** Crucial factors in understanding types of language contact are whether there is multilingualism or simply bilingualism; involving what proportion of the community; and involving which social classes. The choice of which language to use may depend on social situation (this is diglossia, where—for example—one language, or dialect, may be used in the home and in religious observances, and another in all other circumstances) or on the individual (each person will speak their own first language, but be able to understand other languages used in the community). Different degrees of 'lingualism' can be connected to cultural practices—such as inter-marriage (i.e. endogamy or exogamy). In the Vaupés linguistic area, obligatory multilingualism appears to be 'conditioned' by obligatory exogamy: that is, marriage must be with someone who speaks a 'different language' (see Sorensen 1967, and Aikhenvald 1996).
- (f) **TYPES OF INTERACTION OF LANGUAGES WITHIN A PUTATIVE AREA.** There can be one-to-one language interaction, as appears to be the case in Swahili and Khoti (discussed by Dimmendaal in Chapter 13), or one language interacting with an already established group of areally close (and genetically related) languages—as in the case of Tariana and the Tucano languages in the Vaupés (in Chapter 7), or Baale and Tirma-Chai in Africa (also discussed in Chapter 13).
- (g) **LANGUAGE ATTITUDES.** Attitudes towards non-native languages may vary both between communities and within a given community. Speakers of Athapaskan languages preferred not to accept loan words from the languages with which they had contact but would instead create names for new objects

and ideas from their own lexical and grammatical resources. This also relates to questions of language planning (as when Kemal Atatürk resolved to rid Turkish of its Arabic loans—some of fair antiquity—replacing them with native coinings). At the opposite extreme, there has been forceful introduction of foreign elements from Chinese into the minority languages of China in order to ‘improve’ them (Matisoff 1991). Different language attitudes have conditioned different impacts of areal diffusion onto Adyghe (North-West Caucasian) spoken within Russia, and Adyghe spoken within Turkey (Höhlig 1997). And Pontius (1997) shows that social enmities (as in the case of Czech and German) can create an obstacle to structural borrowings. As a further example, lexical borrowings are condemned as culturally inappropriate in the Vaupés area.

Note that this is just a preliminary inventory of relevant parameters. The necessity of correlations between social groups and language attitudes, as well as patterns of multilingualism and diffusion, have been singled out as particularly important; however, only a few of the contributors in this volume discuss these at length (but see Curnow’s summary in Chapter 15). This represents a state-of-the-art situation—as study of the linguistic areas of the world progresses, we hope to learn more and more about these and other parameters, and their impact on areal studies.

Besides direct borrowings, we can note the following types of contact-induced change:

- (a) **SYSTEM-ALTERING CHANGES** involve the introduction of new categories, by analogy with other language(s) in the area—this can include metatypy, mentioned at the end of §3.
- (b) **SYSTEM-PRESERVING CHANGES** could be of two basic types (see Watkins in Chapter 3, and Heath 1997 and 1998): a ‘lost wax’ change (on the analogy of an ancient method for casting bronze artefacts) involves ‘upgrading a minor morpheme to a major morpheme if the latter is threatened’, while a ‘hermit crab’ change spreads fully functionally independent stems into morphology to preserve threatened functional categories (just as a soft-bodied crab can only survive by occupying an empty shell on a beach).
- (c) **LEXICAL ACCOMMODATION** involves adaptation of existing lexical roots in the language to those which are similar and possibly even cognate in a dominant contact language (see Dimmendaal, in Chapter 13, on the ‘borrowing’, or accommodation, in Baale, of the etymologically related term for water from Tirma-Chai).
- (d) **GRAMMATICAL ACCOMMODATION** involves morphosyntactic deployment of a native morpheme on the model of the syntactic function of a phonetically

similar morpheme in the diffusing language (that is, the language which is the source of diffusion). This is exemplified by the extension of imperfective in *-ske-* from Hittite to Greek (as described by Watkins in Chapter 3), in addition to some examples of similar extensions of (unrelated) Tariana morphemes under Tucano influence in the Vaupés (see §4.1.2 of Chapter 7).

- (e) GRADUAL CONVERGENCE AND ISOMORPHISM. Convergence can cover more than ‘just’ adopting some techniques, or discourse strategies (cf. the one-grammar-three-lexicons principle in Friedman (1997); Gumperz and Wilson (1971); or Nadkarni (1975)). It can result in considerable structural isomorphism, whereby the grammatical and semantic structure of one language is almost fully replicated in another.

How does diffusion start? According to Trubetzkoy (quoted by Watkins in Chapter 3), the first place to look in grammars for diffusional convergence is phonology. In Chapter 11, Matisoff discusses phonological salience as one of the criteria for diffusibility (and see Trask (1998) on how Basque phonology was affected ‘first’ by contact with Indo-European languages). However, this is not necessarily the case. The emblematicity of a salient feature (e.g. special phonemes, or sounds, or expressions) can make it resistant to borrowing. This is exemplified by Ross in Chapter 6. Other types of feature can also be emblematic. In Chapter 10, Enfield mentions that syntactic constructions such as patterns of periphrastic causatives can be recognizable as markers of ‘identity’, where a ‘foreign’ alternative construction has come into competition with a ‘native’ construction.

Emblematic features may not come from an official, ‘prestige’ dominating language. Chappell, in Chapter 12, shows that the Taiwanese variety of Mandarin underwent massive calquing and metatypy from Southern Min (a language which has only recently gained official recognition in some domains)—rather than applying in the opposite direction—probably, because Southern Min (and not Mandarin) is ‘emblematic of current loyalties’, serving as ‘a badge of being Taiwanese’.

Several chapters in this volume suggest that the ‘direction’ of borrowings could be ‘from top to bottom’, starting from larger discourse units, and clause coordination and subordination mechanisms, then extending to smaller syntactic units and finally to morphology. The isomorphism in discourse linking, and clause subordination and coordination, between different languages of modern East Anatolia is demonstrated by Haig in Chapter 8. This conclusion is corroborated by numerous studies of syntactic borrowings (see, for instance, Johanson (1998), for Irano-Turkic contacts). The discourse frequency of certain morphemes can also be conducive to their high borrowability—this explains the ready borrowability of classifiers—even if they refer to body parts etc.—in Resígaro, since classifiers (used as referent-tracking devices) are more frequent in discourse than nouns themselves.

The same principle accounts for the frequency of borrowing constituent order (not necessarily accompanied by borrowing word order within individual constituents), as well as connectives, and possibly, expressive morphology (see Haig in Chapter 8 and also Salmons (1990) and Brody (1995)). Not much is known about the way in which interjections are borrowed; Aikhenvald's personal experience with bilinguals in the Vaupés region indicates that they are indeed easily borrowed.

This principle also agrees with borrowing of stylistic patterns especially if one particular genre is 'borrowed' from one language into another—which might be the ultimate result of a process of loan-translation. In a fascinating study on distinct patterns of areal diffusion undergone by Adyghe spoken in Russia, and in Turkey, Höhlig (1997: 107) shows how Adyghe within Russia was forced to adopt new functional styles (such as fiction, newspaper articles, and scientific literature) which resulted in pervasive calquing of Russian stylistic patterns. This did not happen with Adyghe in Turkey where this language remains a vernacular (Höhlig 1997: 174).

This 'loan translation' principle may extend into changing the meaning of existing morphosyntactic patterns. Li and Thompson (1980: 496–7) describe how the adversity passive in Chinese is in the process of changing its semantics under the influence of 'translate' (in Chapter 12 Chappell gives the adversity passive as an areal feature shared by Sinitic languages, while the added features of an overt agent NP and a lexical source in verbs of giving or causative verbs may define Sinitic typologically). N. J. Enfield (p.c.) reports that the same phenomenon, ultimately under the influence of loan translations in radio programs etc., is pervasive across Mainland South-East Asia. In some cases two distinct syntactic structures—one native, and the other 'borrowed'—can be restricted to different stylistic registers of the same language; this has been called 'ditaxia' (see §3.2.4 in Chapter 12).

Finally, one of the most difficult questions to answer is the importance of typological compatibility between languages in facilitating structural borrowing and metatypy. Typological similarities between languages obviously make structural borrowing easier. However, this can hardly be considered a prerequisite for structural borrowing (if it had been so, no contact-induced typological changes would have occurred: see the discussion in Harris and Campbell (1995: 122–36)). A further problem is, however, that a 'shared' typological profile itself may be the result of some prior areal diffusion. This may be the case for structural compatibility between the languages spoken in modern East Anatolia which have been in contact for a long time (see Haig in Chapter 8), as well as in the Amazonian basin and in New Guinea (see Chapters 6 and 7).

In summary, we suggest that grammatical borrowing—which may start with simple calquing and end up with a total metatypy—typically proceeds from larger discourse units to smaller ones. Social conditions which favour and block borrowings are still a matter for further investigation. One of these conditions is the identity-preserving role of linguistic features considered 'emblematic'.

If different aspects of grammar, and of lexicon, differ in how easily they can be borrowed, can we establish a universal hierarchy of borrowability of linguistic elements? In Chapter 15, Curnow offers a detailed discussion of the problems which arise with any attempt to establish an absolute hierarchy of borrowing and of contact-induced change. However, the impossibility of postulating an overarching absolute hierarchy of borrowing—or a scale of borrowing, whereby the expectation for borrowing of different features or terms depends on the type and intensity of contact between the languages under consideration—does not preclude the existence of relative hierarchies and dependencies in borrowability. That is, one can hypothesize that if a language has borrowed (morphological) terms into its grammatical system, one would expect at least some borrowing of syntactic strategies. There could also be dependencies between the typological properties of a language and the borrowability of different categories; for instance, head-marking languages tend to borrow and restructure nominal morphology more easily than verbal morphology. This appears to be the case in Resígaro (see Chapter 7) and Michif (Bakker 1997). In contrast, both nominal and verbal morphology in Tariana—which combines head-marking and dependent-marking properties—underwent contact-induced change under pressure from Tucano languages. Isolating languages—such as Thai—seem to place few or no restrictions on the borrowability of different open word classes (Tony Diller, *p.c.*). A thorough investigation of universal and/or frequently attested dependencies between borrowings of different kinds is a topic for a separate empirical study.

5. Overview of the volume

Chapter 2, by Peter Bellwood, ‘Archaeology and the Historical Determinants of Punctuation in Language-Family Origins’, is an admirably clear and cogent exposition of an archaeologist’s position on the themes of the volume. He begins by summarizing some of the main points of Dixon (1997), and then suggests that the introduction of agriculture is likely to have been a trigger for the punctuational expansion of peoples and languages in the Middle East, China, and Meso-America. Bellwood’s final remarks seek to relate the expansion of Indo-European to the evolution of agriculture in Anatolia, a proposition to which Indo-Europeanists should devote careful attention.

In Chapter 3, ‘An Indo-European Linguistic Area and its Characteristics: Ancient Anatolia. Areal Diffusion as a Challenge to the Comparative Method?’, Calvert Watkins provides a fascinating description of ancient Anatolia as a linguistic area involving both Indo-European and non-Indo-European languages. He then discusses some of Heath’s recent stimulating ideas on language change. Quite independently of Dixon, Heath uses the idea of ‘punctuated equilibrium’. Oddly, he relates an equilibrium state to ‘relatively static monolingualism’, whereas for Dixon an equilibrium period would be characterized by multilingualism. Surely, in small tribal societies—which is all the world had, before the advent of agriculture—a

high proportion of people would have been able to understand (and, often, also to converse in) one or more languages besides their own. All in all, Watkins provides a penetrating summary of the utility of the comparative method, and of current ideas about the nature of language change and language contact.

In Chapter 4, 'The Australian Linguistic Area', Dixon shows how the linguistic situation in Australia is unlike any other in the world, because of its deep time depth (40,000 or 50,000 years) which has given rise to a well-established state of equilibrium. It constitutes one large diffusion area; within this can be recognized a number of low-level genetic groups (probably due to minor punctuations in the recent past), plus a number of small 'relic' linguistic areas. There is no way in which a 'family tree' can model the language situation in this continent. Dixon illustrates a number of the parameters of variation, around which Australian languages move in cyclic fashion—verbal organization, and free/bound pronouns. An appendix examines in some detail the 'Pama-Nyungan' idea, which persists as a species of belief, having no validity as either a genetic model or as a typological construct. Adherence to this idea has held back work on examining the nature of contact relationships in Australia.

Alan Dench, in Chapter 5, 'Descent and Diffusion: The Complexity of the Pilbara Situation', presents a careful and fine-grained discussion of the Pilbara region of Western Australia, demonstrating how Dixon may have been optimistic about the possibility of establishing even low-level genetic groups, with family-tree diagrams. After considering phonological innovations of various types, morphophonemic alternations, case marking patterns, and the shift from an ergative to an accusative type for main clauses, Dench concludes: 'none of the shared innovations ... can be considered, conclusively, to be innovations arising in a single ancestor' and 'our set of languages sharing both form and pattern might have as easily arrived at this similarity through contact rather than through shared inheritance'.

The next three chapters are concerned with intense contact situations and convergence, as well as their correlations with social aspects of language contact. In Chapter 6, 'Contact-Induced Change in Oceanic Languages in North-West Melanesia', Malcolm Ross describes the contact situation between Takia (Oceanic) and Waskia (Papuan) on Karkar island in Papua New Guinea as a paradigm case for intensive language contact. The prolonged contact has resulted in typological convergence, of the type labelled by Ross as 'metatypy'. This case study is supplemented with similar examples from other parts of Papua New Guinea, and provides a neat demonstration of the fact that contact-induced change can lead independently to similar results in different places. (Note that Ross uses the term 'period of equilibrium' in a quite different way from Dixon (1997). He also employs the term 'lect' on the assumption that 'there is no sharp boundary between the concepts of language and dialect', a proposition with which many linguists would disagree.)

In Chapter 7, 'Areal Diffusion, Genetic Inheritance, and Problems of

Subgrouping: A North Arawak Case Study', Aikhenvald considers the genetically inherited patterns of the Arawak language family—the largest in South America. She then shows, for the Arawak languages in northern Amazonia, how extensive and prolonged contact with genetically unrelated languages has obscured the subgrouping. This happened in different ways in sociolinguistic situations with distinct language attitudes. Thus, areal diffusion in the multilingual area of the Vaupés in Brazil, with a cultural inhibition against lexical borrowings, resulted in restructuring of grammar. In contrast, areal diffusion between Resígaro and Bora-Witoto in Peru—where no such inhibitions exist—resulted in massive borrowing of free and bound morphemes, as well as in drastic grammatical restructuring.

Geoffrey Haig, in Chapter 8, 'Linguistic Diffusion in Present-Day East Anatolia: From Top to Bottom', considers modern East Anatolia as a long-standing linguistic area. Languages spoken there—Turkish, Laz (from the Kartvelian family), and Kurmanji Kurdish and Zazaki (both Iranian languages)—display striking similarities in higher-level syntactic organization (strategies of clause linkage, relative clause formation, etc.). The areal influence of Turkish on Laz has resulted in the restructuring of its case system, and in numerous other traces of convergence. Haig shows that forms which occur on a clause boundary are particularly likely to be borrowed. (As Haig states, this study must be regarded as preliminary; further work should include detailed examination of Kartvelian languages spoken outside this area, to confirm that they do not show the areal features noted for Laz.)

The next four chapters discuss various issues concerning the Sino-Tibetan language family, and areal phenomena within South-East Asia as a linguistic area. In Chapter 9, 'The Role of Migration and Language Contact in the Development of the Sino-Tibetan Language Family', Randy LaPolla gives a general perspective on the classification of Sino-Tibetan languages, the applicability of the family-tree model, and the boundaries of genetic versus areal relations within Sinospheric languages. He overviews the extremely diverse patterns of population migrations—of Chinese into other parts of China, of non-Chinese peoples into China (or what later became part of China), and of speakers of Tibeto-Burman languages within and outside Burma. These migrations resulted in an overlay of areal features. He then compares diffusion patterns in what can be loosely defined as Indosphere and Sinosphere, showing how different sources of areal diffusion gave different structural results.

Nicholas J. Enfield, in Chapter 10, 'On Genetic and Areal Linguistics in Mainland South-East Asia: Parallel Polyfunctionality of "acquire"', gives a brief overview of Mainland South-East Asia as a linguistic area with a set of subareas. He then shows how a morpheme which he labels *ACQUIRE* displays a range of meanings and functions among the languages of mainland South-East Asia; in spite of the 'typological plausibility' of each of the individual developments, the multiple grammaticalizations of 'acquire' can be shown to constitute a pan-mainland South-East Asian feature. Enfield provides a judicious and fine-grained

discussion of the possible scenarios leading to syntactic and semantic isomorphism between languages, and criteria for deciding between these.

James A. Matisoff, in Chapter 11, 'Genetic versus Contact Relationship: Prosodic Diffusibility in South-East Asian Languages' provides a valuable list of areal features—grammatical, lexico-semantic, and phonological. His major focus is on tone systems. After presenting a typological characterization of types of tone systems, he discusses whether or not a tone system should be reconstructed for Proto-Tibeto-Burman or Proto-Sino-Tibetan, describing the difficulty of distinguishing between genetic retention and areal borrowing. He then investigates the ways in which tone systems shift from one type to another, often as the result of diffusional pressure.

In Chapter 12, 'Language Contact and Areal Diffusion in Sinitic Languages', Hilary Chappell weighs up several aspects of the grammar of Sinitic languages—the largest subgroup of Sino-Tibetan languages in terms of the number of speakers—in order to evaluate them as either the outcome of areal diffusion or simply as typologically plausible developments. She also includes an in-depth description of typologically unusual and hitherto underdescribed grammatical features of Sinitic languages, such as complementizers and double patient marking constructions, and offers a historical perspective as to the formation of the Sinitic subfamily.

The next two chapters consider the problems of genetic inheritance interrelating with areal diffusion in the African continent. Chapter 13, 'Areal Diffusion versus Genetic Inheritance: An African Perspective' by Gerrit J. Dimmendaal, begins with two case studies of intense areal diffusion between remotely related languages, which has obscured inherited patterns. One is between Swahili and coastal Bantu languages, within Niger-Congo; and the other is between Baale and other Surmic languages, within Nilo-Saharan. He then discusses some of the main features of Niger-Congo languages—nasal and oral vowels, seven-vowel systems versus nine-to-ten-vowel systems with vowel harmony, noun classes, and serial-verb constructions—showing the intertwining of genetically inherited and areally diffused patterns.

In Chapter 14, 'Convergence and Divergence in the Development of African Languages', Bernd Heine and Tania Kuteva present a general overview of areal diffusion, language mixing and genetic inheritance in African languages, paying particular attention to pan-African cognitive patterns, discernible in the ways reflexives and comparatives become grammaticalized. Heine and Kuteva show how the family-tree model is not sufficient to capture relationships in a situation of spaced migrations and the establishment of a small linguistic area. They then, in broad canvas, examine ways in which a typological profile can characterize a linguistic area, presumably as the result of contact-type diffusion.

Chapter 15, by Timothy Jowan Curnow, 'What Language Features Can Be "Borrowed"?' constitutes a partial epilogue. It summarizes the results of the papers in the volume in so far as they concern the borrowability of elements. Curnow

considers a variety of issues which include the notion of borrowing, the scales and hierarchies of borrowability or adoptability of linguistic elements, and various factors which either favour or block borrowings and contact-induced changes. He then discusses the units which can be borrowed, and concludes that it is probably impossible to establish a universally valid hierarchy of borrowability, at least in terms of our knowledge of the languages of the world at this point in time.

6. Prospects

Our knowledge about how and why human languages interact is still at a rather early stage. In order to know more about the development of languages and different types of language contact, we suggest the following fruitful lines for future inquiry.

- (a) Are there any limits to the borrowability of categories? If it is impossible to establish a universally valid 'hierarchy' of borrowability of linguistic elements, are there any regularities in (i) the order in which different elements are likely to be borrowed; (ii) dependencies between borrowings of distinct aspects of grammar, and/or lexicon; and (iii) correlations between 'gain' and 'loss' of categories and morphemes in the situation of areal diffusion and contact-induced change?
- (b) What is the role of code-switching, its socio-psychological motivations in various contact situations, and its role in diffusibility of features?
- (c) How can one determine the speed of language change? Assuming that languages change at different rates, what exactly determines the speed of language change? Among the relevant factors could be movements of population and spread; areal diffusion; and norm enforcement.
- (d) What social and linguistic factors can halt diffusion at a dialect boundary? (See Watkins, in Chapter 3, on how no further 'dialect' divisions occurred within Greek or Armenian.)
- (e) How does a language community adopt certain linguistic features as 'emblematic' of their ethnicity?
- (f) What are the possible linguistic consequences of equilibrium situations, alongside the impact of punctuations of various kinds?

These are but a few of the possible directions for future investigations in contact-induced change. In order to elaborate on these, and other issues (which have not been mentioned here), a large number of in-depth empirical studies of contact phenomena from across the world are required. This volume is but a start.

References

- AIKHENVALD, A. Y. 1996. 'Areal diffusion in north-west Amazonia: the case of Tariana', *Anthropological Linguistics* 38.73–116.

- AIKHENVALD, A. Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- and DIXON, R. M. W. 1998. 'Evidentials and areal typology: a case study from Amazonia', *Language Sciences* 20.241–57.
- and MILITAREV, A. Y. 1991. 'Livijsko-guanchskie jazyki' [Lybic-Guanche languages], pp. 148–266 of *Jazyki Azii i Afriki [Languages of Asia and Africa]*, vol. 4. Moscow: Nauka.
- ANDERSEN, H. 1988. 'Centre and periphery: adoption, diffusion and spread', pp. 39–85 of *Historical Dialectology*, edited by J. Fisiak. Berlin: Mouton de Gruyter.
- AUWERA, J. VAN DER. 1998. 'Revisiting the Balkan and Meso-American linguistic areas', *Language Sciences* 20.259–70.
- BAKKER, P. 1997. *A Language of Our Own: The Genesis of Michif, the Mixed Cree–French Language of the Canadian Metis*. New York: Oxford University Press.
- BAUDOUIIN DE COURTENAY, J. 1930. 'Zagadnienia pokrewieństwa językowego', *Biuletyn Polskiego towarzystwa językoznawczego* (Cracow) 2.104–18. (English translation, 'Problems of linguistic affinity', pp. 296–307 of *A Baudouin de Courtenay Anthology: The Beginnings of Structural Linguistics*, edited by E. Stankiewicz. Bloomington: Indiana University Press. 1972.)
- BRIGHT, W., and SHERZER, J. 1976. 'Areal phenomena in North-American Indian languages', pp. 228–68 of *Variation and Change in Language. Essays by W. Bright*. Stanford: Stanford University Press.
- BRODY, J. 1995. 'Lending the "unborrowable": Spanish discourse markers in indigenous American languages', pp. 132–47 of *Spanish in Four Continents: Studies in Language Contact and Bilingualism*, edited by C. Silva-Corvalan. Washington, DC: Georgetown University Press.
- CAMPBELL, L. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.
- KAUFMAN, T., and SMITH-STARK, T. 1986. 'Meso-America as a linguistic area', *Language* 62.530–70.
- DIXON, R. M. W. 1991. 'Mbabaram', pp. 348–402 of *The Handbook of Australian Languages*, vol. 4, edited by R. M. W. Dixon and B. J. Blake. Melbourne: Oxford University Press.
- 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- and AIKHENVALD, A. Y. 1999a. 'Introduction', pp. 1–21 of Dixon and Aikhenvald (1999b).
- 1999b. *Editors of The Amazonian Languages*. Cambridge: Cambridge University Press.
- EMENEAU, M. B. 1967. 'The South Dravidian languages', *Journal of the American Oriental Society* 87.365–413.
- FRIEDMAN, V. A. 1997. 'One grammar, three lexicons: ideological overtones and underpinnings in the Balkan Sprachbund', *Chicago Linguistic Society* 33: 2.23–44.
- GUDSCHINSKY, S. 1956. 'The ABC's of lexicostatistics', *Word* 12.175–210.
- GUMPERZ, J. J., and WILSON, R. 1971. 'Convergence and creolization: a case from the Indo-Aryan/Dravidian border in India', pp. 151–68 of *Pidginization and Creolization of Languages*, edited by D. Hymes. Cambridge: Cambridge University Press.
- HARRIS, A. C., and CAMPBELL, L. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- HASPELMATH, M. 1998. 'How young is Standard Average European?', *Language Sciences* 20.271–87.
- HEATH, J. 1978. *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.

- 1997. 'Lost wax: abrupt replacement of key morphemes in Australian agreement complexes', *Diachronica* 14.197–232.
- 1998. 'Hermit crabs: formal renewal of morphology by phonologically mediated affix substitution', *Language* 74.728–59.
- HÖHLIG, M. 1997. *Kontaktbedingter Sprachwandel in der adygeischen Umgangssprache im Kaukasus und in der Türkei*. Munich: Lincom Europa.
- JOHANSON, L. 1998. 'Code-copying in Irano-Turkic', *Language Sciences* 20.325–37.
- JOSEPH, B. D. 1983. *The Synchrony and Diachrony of the Balkan Infinitive. A Study in Areal, General and Historical Linguistics*. Cambridge: Cambridge University Press.
- KÖHLER, O. R. A. 1974. 'African languages, IV: Khoisan languages', *The New Encyclopaedia Britannica*. 15th edition. Chicago: Encyclopaedia Britannica. (Macropedia, vol. 1, pp. 228–31, in 1984 printing.)
- LAPOLLA, R. J. 1994. 'Parallel grammaticalizations in Tibeto-Burman: evidence of Sapir's "drift"', *LTBA* 17.61–80.
- LI, C. N., and THOMPSON, S. A. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley: University of California Press.
- MASICA, C. P. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- MATISOFF, J. A. 1991. 'Areal and universal dimensions of grammaticalization in Lahu', pp. 383–453 of *Approaches to Grammaticalization*, vol. 2, edited by E. C. Traugott and B. Heine. Amsterdam: John Benjamins.
- NADKARNI, M. V. 1975. 'Bilingualism and syntactic change in Konkani', *Language* 51.672–83.
- PONTIUS, J. 1997. 'Language codification and the perception of otherness: the case of Czech and German', *Chicago Linguistic Society* 33.23–44.
- ROSS, M. D. 1994. 'Areal phonological features in north central New Ireland', pp. 551–72 of *Language Contact and Change in the Austronesian World*, edited by T. Dutton and D. Tryon. Berlin: Mouton de Gruyter.
- 1996. 'Contact-induced change and the comparative method: cases from Papua New Guinea', pp. 180–217 of *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, edited by M. Durie and M. D. Ross. New York: Oxford University Press.
- 1997. 'Social networks and kinds of speech-community event', pp. 209–61 of *Archaeology and Language*, vol. 1: *Theoretical and Methodological Orientations*, edited by R. Blench and M. Spriggs. London: Routledge.
- SALMONS, J. 1990. 'Bilingual discourse marking: code switching, borrowing, and convergence in some German-American dialects', *Linguistics* 28.453–80.
- SAPIR, E. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and World.
- SASSE, H.-J. 1985. 'Sprachkontakt und Sprachwandel: die Gräzisierung der albanischen Mundarten Griechenlands', *Papiere zur Linguistik* 32.37–95.
- SCHLEICHER, A. 1861–2. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen: kurzer Abriss einer Laut- und Formenlehre der indogermanischen Ursprache*. Weimar.
- SCHMIDT, J. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.
- SEKI, L. 1999. 'The Upper Xingu as an incipient linguistic area', pp. 417–31 of Dixon and Aikhenvald (1999b).

- SHERZER, J. 1973. 'Areal linguistics in North America', pp. 749–95 of *Current Trends in Linguistics*, vol. 10, edited by T. A. Sebeok. The Hague: Mouton.
- 1976. *An Areal-Typological Study of American Indian Languages North of Mexico*. North-Holland Linguistic Series 20. Amsterdam: North-Holland.
- SHEVELOV, G. Y. 1964. *A Prehistory of Slavic*. Heidelberg: Winter.
- SORENSEN, A. P., JR. 1967. 'Multilingualism in the northwest Amazon', *American Anthropologist* 69.670–84 (reprinted as pp. 78–93 of *Sociolinguistics: Selected Readings*, edited by J. B. Pride and J. Holmes. Harmondsworth: Penguin. 1972).
- TOSCO, M. 2000. 'Is there an Ethiopian language area?', *Anthropological Linguistics* 42.329–65.
- TRASK, L. 1998. 'The typological position of Basque: then and now', *Language Sciences* 20.313–24.
- WESTPHAL, E. O. J. 1962. 'A re-classification of southern African non-Bantu languages', *Journal of African Languages* 1.1–8.
- 1971. 'The click languages of southern and eastern Africa', pp. 367–420 of *Trends in Linguistics*, edited by T. A. Sebeok, vol. 7: *Linguistics in Sub-Saharan Africa*. The Hague: Mouton.

Archaeology and the Historical Determinants of Punctuation in Language-Family Origins

Peter Bellwood

1. Introduction

As the sole archaeologist writing in a publication dominated by linguists, I hope to encourage a two-way examination of information across the two disciplines. Archaeologists and linguists are able to focus on the reasons for existence of large-scale and widely distributed aspects of human variation, especially major complexes of archaeological material culture and major language families/subgroups. The view taken in this chapter is that such large-scale entities track the evolution of ancestral configurations through processes of both genetic differentiation and contact-induced change. In simple terms, genetic differentiation requires dispersal from a homeland region. Contact-induced change cuts across the lines of genetic differentiation to impose significant aspects of regionality on the cultural pattern.

In situations where history reveals close correlations between the geographical extents of cultures and languages, as for instance in the recent centuries of European colonization of Australasia and the Americas, we see that linguistic change on a large scale is reflected in the record of change in material culture, and vice versa. The European colonizations were very rapid events from a long-term perspective, both in terms of language distribution and in the material culture records of the colonized territories. This chapter suggests that similar episodes of rapid change over very large areas also occurred on occasions in prehistory. These prehistoric episodes were revolutionary and world-changing in impact, at least by the standards of their day. They were also *quite abnormal* from the total, highly reticulative, perspective of human history. Such episodes have been allowed to occur because of specific and unusual concatenations of historical circumstances.

When Bob Dixon asked me to read an early typescript version of his book *The Rise and Fall of Languages* (1997), I immediately recognized a kindred spirit in the field of linguistics, one also stimulated by the possibility that episodes of punctuated evolution have occurred in human history. In Dixon's view, those major

language families which have reconstructable proto-languages and a coherent structure of genetic subgroups represent dispersive punctuations in the long-term flow of language through time. In my view, certain archaeological complexes located in the early agricultural phases of world prehistory represent similar punctuations. As far as interrelations between archaeology and linguistics are concerned, we need to ask if one can identify episodes of punctuation recorded in both disciplines which have *the same historical causation*. That is, do language families and archaeological complexes ‘square off’ in prehistory, or are the behaviour patterns that produce them entirely unrelated? The answer is, of course, partly ‘yes’, partly ‘no’. On the consciously interactive level of individual adjacent societies, processes such as language shift can ensure that specific languages need not always correlate with the material cultures one might expect on comparative grounds, and vice versa. Witness the complexity in this regard in Melanesia and Amazonia. But on the historically unconscious level of a major language family such as Indo-European or Austronesian, it is held that overall correlations do exist in early periods of dispersal, indeed *can be expected to exist*. This chapter focuses on such situations, examined on trans-continental and millennial scales.

1.1. PHYLOGENY AND RETICULATION

Linguists, plus those historians, archaeologists, and anthropologists who take an interest in historical comparative linguistics, have tended to view the past of language as reflecting predominantly one of two processes—either phylogenetic mother-language to daughter-language descent (the ‘family tree’ model, although the tree metaphor is often not the best one to use), or a coevolutionary process which stresses contemporary reticulative interaction between neighbouring languages (the ‘linguistic area’ model).¹ In Dixon’s view, as in mine, both processes occur; they simply reflect different kinds of historical trajectory.

Essentially, language families reflect both genetic and areal features. They have genetic structures of component subgroups as a result of relatively short-lived periods of expansion. They share areal features as a result of much longer-term processes of interaction and borrowing. In many major language families, such as Indo-European and Austronesian, the results of both processes are evident. Far-flung languages, often thousands of kilometres apart, share transparent phylogenetic relationships (e.g. English and Bengali, Malay and Tahitian, Navajo and other Athabaskan languages far to the north in Canada). On the other hand, languages within different families can share areal features within specific regions. Such ‘linguistic areas’ include the Indian subcontinent, Mesoamerica, the Balkans, and the Amazon basin. Yet—and this proviso requires stress—*the languages within these areas still retain traces of family-level phylogenetic relationship in spite of interaction*. In other words, we do not find the Indian subcontinent to be full of

¹ See Bellwood (1996a, 1998) for discussion of these models, and especially of the concept that any array of phylogenetically linked entities must imply a homeland and a dispersal/radiation.

phylogenetically unclassifiable languages which have all blended equal aspects of Indo-European, Dravidian, and Mundaic structure and vocabulary. Neither is this the case in Amazonia, where some groups practise linguistic exogamy, yet language families such as Arawak and Tucanoan remain essentially unblended and do not borrow lexical items frequently, despite structural convergence and widespread multilingualism (Aikhenvald, this volume).

Despite this, as Dixon notes, if such linguistic areas are allowed to develop for *tens* of millennia without external interference, rather than the few millennia presumably represented in the above examples, then phylogenetic relationships which reflect the history of dispersal of a language family might be erased completely and replaced by purely local patterns of language spread and areal interaction. He gives the example of Australia as such a region. Dixon suggests that the languages of the greater part of Australia, those classified by other linguists as 'Pama-Nyungan', represent not a genetic family with a relatively recent history of expansion, but simply reflect the results of areal interaction continuing perhaps since Australia was first settled over 40,000 years ago. The 'Pama-Nyungan' languages might indeed be genetically related in the final resort, but in Dixon's view they are not related by a coherent set of widely shared features which serve to separate them from the other Australian languages.

If this view is correct,² Australia becomes rather unusual in world terms, especially when compared to the hunter-gatherer languages of the Americas, which belong to, or even form exclusively, several well-defined families (e.g. some of Uto-Aztecan, much of Algonquian, and all of Athabaskan). In my view, the problem is not that areal convergence to the degree claimed by Dixon for Australia *cannot* occur, but that during the last few thousand years history has, in most other parts of the world, hardly ever allowed it to occur. In most parts of the world, humans have been too active and too competitive to allow such quiet conditions of interaction to continue for more than a few millennia.

Dixon goes on to make several other observations throughout his book, some of which are of direct relevance for what follows. Firstly, there is no constant rate of language change; all depends on the linguistic environment within which the change takes place, and so linguistic 'dating' methods such as glottochronology have inherent problems. Languages in intense bilingual situations of contact with other unrelated languages will change more quickly than languages that are isolated or only in contact with languages already closely related. For this reason, dates derived from non-written linguistic data alone are rather suspect. This need not mean that all dates derived from glottochronology are necessarily incorrect—the problem is to separate the good from the bad (archaeologists have similar context-related problems with their radiocarbon dates).

Secondly, each language will have a single parent in the normal course of

² It should be noted that the genesis of the Australian linguistic pattern over the past few thousand years is currently an issue of major debate (Dixon 1997, McConvell and Evans 1997).

linguistic evolution. The total 'merging' of two completely unrelated languages brought into propinquity would be a very unusual outcome, although merging of two languages *already related genetically* is a different matter and can occur. The major language families thus do not have histories of wholesale creolization—if they did, they would not exist.

Thirdly, in discussing proto-language reconstruction, Dixon (1997: 98) makes the point that any given 'language family may have emanated not from a single language, but from a small areal group of distinct languages, with similar structures and forms'. He believes such situations could help explain similarities noted between, for instance, Indo-European and Uralic, and Chinese and Tibeto-Burman. From an archaeologist's point of view this is very significant. If Early Chinese and Early Tibeto-Burman were spoken close together within a former linguistic area, this tells much about culture history. So too might the claimed relationships between early Indo-European and early Uralic and/or Semitic. It is not essential *for archaeological purposes* to distinguish precisely between relations of shared descent and relations of close geographical proximity at the proto-language levels. Much of the debate about the existence or otherwise of Nostratic, focused as it is upon genetic relationship, is tangential from an archaeologist's perspective to the main geographical observation that being in contact via some kind of areal relationship can be extremely significant. If speakers of Proto-Semitic and Proto-Indo-European really were located within a relatively circumscribed region this is an observation of great importance, regardless of whether or not such languages share a common genetic ancestry.

Thus, language families reflect two kinds of process—rapid punctuated origins and early spreads, and long-term 'settling-down' forms of interaction and convergence. The question for us to consider, one of the major questions of human history, is why the punctuation? Why isn't human history just one endless sequence of interaction and convergence?

1.2. SOME FURTHER OBSERVATIONS ABOUT LANGUAGES AND LANGUAGE FAMILIES IN TIME AND SPACE

Some further observations are important here as a background to the archaeological perspective which follows (Bellwood 1996c, 1997a). These derive from a general reading of the linguistic literature, looking into the historical dynamics of language spread, and observing the sociolinguistics of language use amongst small-scale tribal populations.

Firstly, many of the major language families had reached almost to, or entirely to, their pre-colonial geographical limits before the beginnings of literacy, empires, world religions, and so forth. Thus, they spread in social conditions of tribal society, perhaps at its most complex attaining the chiefdom or incipient state mode. This applies to Indo-European (Ireland to Bangladesh), Austronesian (Madagascar to Easter Island), Uto-Aztecan (Utah to El Salvador), and Bantu

(Cameroon to South Africa). In some other cases, e.g. Tai, Sino-Tibetan, and Afroasiatic (especially Semitic), a lot of the spread has taken place within the historical period, but actual extension of the limits of the whole family has not always been very great (except possibly in the cases of Thai and Chinese). From this we can state that a great deal of language-family spread took place under relatively non-authoritarian and certainly non-literate social conditions, in prehistoric circumstances of overarching historical unconsciousness.

Secondly, my understanding derived from careful examination of available historical records and various sociolinguistic studies suggests that under such social conditions, *extremely widespread* language dispersal at the vernacular level will require *substantial movement* of native speakers (Bellwood 1997a). On this I differ to some degree from Nichols (1998), who believes that major language families such as Indo-European have spread almost entirely by language shift amongst essentially unmoving populations, apart from the small groups necessary to introduce a new language into a new region in the first place. Language shift alone, in my view, is not sufficient to explain the pattern, since it is doubtful that a language family would hold together a coherent genetic structure if subjected to trans-continental spread through a mosaic of atomizing social structures of the kind presumably present before the rise of literate conquest states. Of course, language shift has occurred in *local* or even *regional* situations, but it is not sufficient as the sole explanation for the spread of any major language family. There is an important factor of scale here which should not be overlooked.

Thirdly, the great language-family spreads surely reflect a certain chronology which most linguists would place within the past 10,000 years. We cannot state ages for language families with any authority, but if the spreads of Indo-European or Austronesian since their primary proto-language stages had taken place over 20,000 instead of about 6,000 to 8,000 years, we probably would not expect such clear signs of genetic structure to survive as we witness in reality, especially if the regions of spread were already inhabited by other populations speaking unrelated languages. Historical observations of rates of erosion of cognates between paired languages surely have some validity on a general level. Glottochronology may not be an absolutely accurate measure of time depth, but at least it can give a general indication that most existing language families are Holocene rather than Pleistocene phenomena.

There are also possible implications of language-family subgroup definition for dispersal hypotheses. Subgroups which are defined by large numbers of unique innovations, e.g. the Polynesian subgroup within the larger Austronesian family, suggest geographical standstill periods during the homeland phase, a point recently made very clearly for Polynesian by Andrew Pawley (1996). But families which have two or more high-order subgroups which differ only slightly from each other on reconstructed vocabulary items (e.g. Blust's Austronesian subgroups of Malayo-Polynesian, Central Malayo-Polynesian, Central-Eastern Malayo-Polynesian, and also the Oceanic subgroups in Melanesia) are candidates

for rapid spread during the proto-language stage since nests of defining innovations have presumably had little time to form (Blust 1993). This need not be the only explanation for the rake-like shape of a 'family tree', but it is one which needs consideration, particularly if the subgroups are very widespread in geographical terms. Indo-European might also be a candidate for this kind of explanation, and so also might Uto-Aztecán (see below).

There is another observation, one which I think could be a major key to a historical explanation of the world linguistic map. In continental regions long settled by human populations, we may debate which particular archaeological cultures might have been associated with the dispersal of one or another language family, but we can never marshal *proof*.³ My answer to this problem is to compare world-wide patterns in the data of the two disciplines, both in space and through time, and to look for explanations for these patterns in the realms of human history and behaviour that can satisfy the demands of both disciplines. Demographic growth and dispersal of early agricultural populations is one such explanation. Here, we have a geographical cross-correlation between the two disciplines so strong that it seems to defy any explanations based on chance alone. Many of the major language family origin regions, as understood in terms of majority views amongst the linguistic community, correlate very well with agricultural-origin regions. The reverse is also true. Agricultural-origin regions (Middle East, China and Mainland South-East Asia, Mesoamerica, New Guinea Highlands, Sub-Saharan Africa) form zones of intersection of several major language families. Indeed, in terms of numbers of separate language families represented per unit of area, agricultural-origin zones can be regarded as linguistically very diverse. This, from my perspective, suggests a predominance over time of language and population outflow, rather than inflow (Bellwood 1997a).

The resulting hypothesis of an early agricultural stimulus for language family genesis (Renfrew 1991, Bellwood 1996b, c, 1997a) can be stated as follows. Within the past 10,000 years, human populations in many tropical and temperate regions have developed or acquired systematic forms of food production. Some of these food-producing populations, especially those involved in original and primary transitions to agriculture and animal husbandry at a time when the world was still very much the preserve of hunter-gatherers, were able to expand both in numbers and in geographical extent, often on remarkably large scales. The result of such expansions was the generation of a substantial proportion of the pattern of biological and ethnolinguistic human diversity present across the world today. Many of the world's major language families, plus the pre-AD 1500 distributions of several of the major geographical 'races' of mankind, owe their existences to these processes of primary agricultural dispersal.

³ Island Oceania may be different in this regard since it has only been settled by a single population beyond the Solomons, these being the Austronesians. At least, this is the received view, and in my opinion the most likely view apart from the possibility of late Polynesian contact with South America, a contact which evidently involved no major population movement.

There is a related observation which follows on from the above, based on observations of modern hunter-gatherers and on the strongly bimodal nature of food-taking (i.e. hunting and gathering) and food-producing economies as recorded in the *Ethnographic Atlas* (Murdock 1967). This observation suggests that one can hunt and gather, or one can produce food, but one cannot do both successfully on an equal basis in most environments. A half-way point is inherently unstable, a circumstance which renders any shift to 'professional' agriculture no small matter for a band of mobile foragers. A corollary of this is that agricultural economies surely spread mainly throughout the Neolithic/Formative world by means of demographic dispersal of the agriculturalists themselves, rather than through the adoption of agriculture by hunter-gatherers. My reading of the available evidence suggests to me that hunter-gatherers will resist agricultural shift unless they are already sedentary and part-way along the trajectory towards agriculture—a fairly rare state in the world hunting and gathering record, whether ethnographic or archaeological. Also, successful agriculturists with expanding populations who need new land will often not allow hunters and gatherers to 'join the club', although in frontier situations this might be counterbalanced by a need for personnel recruitment, often by the marriage of hunter-gatherer females into farming communities. Nevertheless, in my view, agriculture cannot have spread entirely because hunters and gatherers everywhere adopted it. Adoption obviously never worked for Australia or California, or even perhaps for Jomon Japan, even though opportunities provided by contact with farmers were not entirely lacking in these situations.

These comments obviously exclude those hunter-gatherers in many parts of the world amongst whom agricultural subsistence developed in the first place. They also exclude the known cases in which one-time hunters and gatherers did adopt agriculture in prehistory (for reasons which cannot be discussed in detail in this chapter), and added their own genes and languages to the trajectory of agricultural dispersal through the past ten millennia. But most hunter-gatherers of the prehistoric past, in my view, did not freely adopt agriculture, at least not to any significant degree.

Before going further, it must be made clear that agricultural dispersal cannot be the explanation for the *totality* of human macro-variation in all times and places since the end of the Pleistocene. It is of limited, even zero relevance for explaining the prehistory of hunter-gatherer regions such as Australia and the greater part of North America. In addition, many populations and languages (but not language families!) have spread over large distances in historical times, for instance native speakers of Thai, Chinese, Arabic, Spanish, and English. Many hunter-gatherer populations have also spread in recent millennia, especially in North America (Athabaskans, plus some Algonquians and some Uto-Aztecs). Conversely, by no means all food-producers in prehistory have spread on major scales; many, like the Sumerians, Ancient Egyptians, and New Guinea Highlanders, intensified food production 'at home' in order to feed growing

populations. Some environments allow this kind of *in situ* intensification, others do not, and such environmental differences can be of profound importance for understanding world history.

If early agricultural societies and the languages ancestral to many modern major language families spread together from overlapping homeland areas, as I am suggesting, and as Renfrew (1991) has suggested before for the Middle East, we should see traces clearly in the archaeological record. I believe we can, especially for the Middle East, China, Sub-Saharan Africa, and Mesoamerica. We probably can also for New Guinea and the northern and central Andes. But it is not possible here to examine all these areas in detail. I will examine the Middle East and China as the best-understood cases, and add some comments on northern Mesoamerica and the south-western United States.

2. The archaeological record of early agriculture—some key points

2.1. THE MIDDLE EAST

The Middle East is by far the best-known region in world prehistory for the transition to cereal agriculture. The Middle Eastern trajectory does not necessarily provide the model for all the other areas of transition, but it had, in terms of world significance, the greatest impact on human affairs, followed closely by that in China.

In the Middle East, the transition was closely related in timing with the oscillating amelioration and periodic re-cooling of post-glacial climate, focusing on the period between 16,000 and 11,000 BP (calibrated radiocarbon dates). It occurred in a relatively small region of the Levant, a kind of ‘proto-Fertile Crescent’ (as mapped by Hillman 1996: Figure 10.10) of very marked rainfall seasonality where wild cereals and legumes flourished. It involved firstly cereal and legume domestication, then secondly, after one or two millennia in most regions, animal domestication.⁴ The spread of pottery followed even later owing to the consumption of cereals in ground rather than whole-grain form—pottery in the Middle East was not so essential in the early days of agriculture as it was in China, where rice and millets were evidently prepared by whole-grain boiling rather than by grinding into flour.

The Middle Eastern transition evolved from a baseline of complex foraging in the Natufian cultural phase, with presumed sedentary or near-sedentary settlements existing before cereals became domesticated. Firstly, there was an initial switch, perhaps a very rapid one, c.11,000 BP, from the Natufian harvesting of

⁴ There is a possibility that some form of animal management preceded cereal domestication in eastern Turkey and the Zagros: see Rosenberg *et al.* (1998) for possible pre-agricultural pig husbandry in the site of Hallan Çemi in eastern Turkey.

slightly unripe wild grain (with few or no domesticatory side-effects) to a Pre-Pottery Neolithic harvesting and purposeful *planting* of ripe grain. This led rapidly to an increasing selection for the domesticated characters of wheat and barley—non-shattering habit, synchronous ripening, glume reduction, loss of obligatory summer dormancy. This switch in harvesting behaviour must have occurred in or very close to the region of reconstructed Late Pleistocene cereal and legume habitat (Hillman 1996). Exactly why the switch occurred is not clear and is a peripheral issue in terms of the present discussion, but climate change, increasing population and demands for food, and increasing social competition and feasting have all been suggested as possible stimuli.

As a result of the above shift to domesticated food production, we witness a massive increase in the size of the human population in the Levant from the Natufian into the Pre-Pottery Neolithic, and a corresponding increase in the sizes of villages. This increase occurred rapidly during the eleventh millennium BP and peaked during the tenth. By later Pre-Pottery Neolithic times, c.9,500 BP, the peoples of the Levant commenced animal domestication, surely in part as a necessity since growing populations meant lessening wild meat supplies due to over-hunting. The largest villages now reached up to twelve hectares in size, supported by a fully fledged mixed farming economy. Large quantities of timber were required for building construction and firing of lime mortar for floor and wall plaster.

By final Pre-Pottery Neolithic times, c.9,000–8,500 BP, many regions of the Levant were under environmental stress, probably due to increasing population and intensity of production. This is a region of low rainfall and of reluctant productivity outside the major river valleys such as Mesopotamia and the Nile, neither of which was intensively settled by agriculturalists until after this date. The results of this stress can be seen in the form of settlement abandonment, deforestation, decreasing settlement size, and moves into pastoralism in the southern Levant.

One ultimate result of this was probably an increasing interest in population movement to find better land—into the major irrigable river valleys (Nile, Mesopotamia, Indus) and eventually into Greece to commence the agricultural colonization of Europe via the Danube Basin and the Mediterranean coastline. The expansion of the western Eurasian mixed farming economy from the Middle East is a fairly well-defined and non-contested aspect of the archaeological record. Few archaeologists today would wish to deny that mud-brick architecture, female figurines, painted pottery, and the major domesticated plants and animals—wheat, barley, goats, sheep, perhaps even pigs and cattle—spread into Europe, North Africa, and Central Asia from a Middle Eastern Neolithic source region.⁵

⁵ Cattle herding in a Sahara much moister than today might have preceded cereal agriculture in north-eastern Africa, but this will not alter the overall significance of a Middle Eastern stimulus for the Nile Valley Neolithic in general.

But what is in contention amongst archaeologists is the question of *who* was responsible for the dispersal of these traits—existing agricultural populations with rapid demographic growth profiles, or *in situ* Mesolithic hunter gatherers who were keen to adopt agriculture?

Unfortunately, the archaeological record, being confined essentially to material culture, is not always going to provide suitable evidence for discussion of questions of population continuity versus immigration at a point of economic transition. Yet one must not overlook here the crucial matter of patterning. In many parts of the world, including the Middle East and Europe, late hunter-gatherer cultures on the eve of agriculture reveal considerable regional differentiation, but are then replaced by very widespread and well-integrated cultural patterns at the beginning of agriculture. In such circumstances, there is a very high chance that the pattern reflects some degree of population spread as opposed to mere diffusion, even if the archaeological record alone cannot provide an unambiguous answer.

It is at this point that the punctuated evolution approach to explaining *language-family origins and dispersals* becomes of great interest, because if the economic dispersals of agriculture occurred with movement of actual populations, then there must also have been some enormous changes in the language distribution maps of the regions concerned. This is perhaps where a punctuated-evolution view of agricultural origins can be matched against the punctuated-evolution view of language-family origins proposed by Dixon. Furthermore, if languages were spreading out of the Middle East with the agricultural economy, then they were spreading from relatively small homeland regions in and around the Levant. Is this the historical circumstance which has left the elusive traces of shadowy relationship between Indo-European, Afroasiatic, and Dravidian which many linguists identify as Nostratic (Renfrew 1991)? Did these language families spread in their ancestral forms, fuelled by agricultural population growth and environmental degradation to the rear, from around the edges of a Neolithic Middle Eastern linguistic area?

2.2. CHINA

In China we witness developments similar to those in the Middle East, but commencing perhaps a millennium or so later. By 9,000 BP, villages of millet and rice agriculturists were present in the middle and lower Yellow and Yangzi valleys. By 7,000 BP these regions supported archaeological cultures with a mixed farming economy and networks of large villages up to six hectares in size. The Chinese Neolithic has sufficient coherence in pottery forms and decoration, forms of stone adzes and reaping knives, domestic animals (pig, dog, chicken), and major crops (rice, foxtail millet) to suggest, as for the Middle East, that a generalized spread occurred from a relatively small homeland region, in this case encompassing the middle and lower Yellow and Yangzi valleys and the regions between. As with the

Middle East, this does not necessarily imply that one single ancestral population founded everything. It is more likely that the early developments took place in a region akin to a linguistic area, perhaps with two separate foci in the Yellow (millets) and Yangzi valleys (rice). Populations would have 'budded off' around the edges from time to time in order to move into new regions to colonize new agricultural territories.

By 6,500 BP the Chinese Neolithic had reached south coastal China, then Taiwan by 5,500 BP, where the oldest pottery-using culture (the Dapenkeng) has recently produced clear botanical evidence for rich agriculture from the Nanganli site in the vicinity of Tainan (Tsang Chang-hwa p.c.) As far as Taiwan generally is concerned, Dapenkeng sites are numerous, located in or on the edges of coastal lowland regions where rice cultivation is significant today, and well-linked in terms of pottery styles with contemporary sites in coastal Fujian. They are also very similar stylistically right around the island, as opposed to the later Neolithic cultures, which show much greater stylistic regionalization.

As I have argued elsewhere (Bellwood 1996c, 1997b), the expansion of an agricultural economy, involving particularly rice, from central China between 7,000 and 4,000 BP, has been the major factor behind the expansion of the biological population which we know as Southern Mongoloid (or simply 'Asian' in modern Australian media parlance). The onwards dispersal of this population into the Pacific, without the rice, puts the spotlight on the Austronesians. But any discussion of the homelands of the major continental eastern Asian language families—Tai, Sino-Tibetan, Austroasiatic, Hmong-Mien—which ignores the significance of the Chinese centre of demographic dispersal would in my view be incomplete.

2.3. OTHER AGRICULTURAL DISPERSALS

The Bantu languages, Japanese, and Dravidian can also be added to the list of languages or language families which have spread with agricultural societies moving into areas occupied mainly or entirely by hunters and gatherers. The Indian subcontinent is especially interesting in this regard because it is not a centre of primary agriculture. Its agricultural systems, and its Indo-European and Munda languages, entered from west and east respectively. Crops introduced from the west include sorghum from Africa, presumably via Harappan sea lanes to Arabia and the Persian Gulf, and wheat and domestic animals from the Middle East. Crops introduced from the east include rice and foxtail millet from China and South-East Asia. Given the archaeological and botanical data, a strong case can be made that the initial spreads of Indo-European and Munda languages into the subcontinent occurred with Neolithic archaeological spreads into northern India from both western and eastern directions. The exception here is of course the Dravidian language family, which may well have spread on linguistic grounds from the north-western part of the subcontinent (?), with native populations who had adopted crops and animals introduced during phases of dispersal (pre-Harappan and

onwards) into India from the Indus Valley (Possehl 1997). Some of this ex-Indus dispersal might represent another situation of population movement following environmental degradation, similar to that posited for the late PPNB/PPNC Levant. It underlay perhaps both the Indo-European and Dravidian colonizations of India, both processes commencing c.5,000 BP.

What of the Americas? Can we see agricultural dispersal in the linguistic record here? One problem, as noted by Jared Diamond (1997: 177), is that the Americas have a north–south axis, which means that vast territories were beyond the range of agriculture for prehistoric peoples. At European contact, the Americas had immense regions still occupied by hunters and gatherers, and agricultural economies lacked any major stockbreeding components. There was also only one major cereal—maize. Agricultural dispersal also faced bottleneck problems caused by geographical, climatic, and altitudinal constrictions in Mesoamerica and on the routes into North America, these being far more stressful than the routes out of the Old World agricultural homelands. Furthermore, some regions, such as Amazonia, have linguistic maps so mosaic-like and so contorted by post-colonial population movements that any attempts to relate language family dispersals to early agricultural populations will be fraught with difficulty. But there is still one excellent example, that of Uto-Aztecan, upon which I believe the linguistic and archaeological records can agree.

2.3.1. Uto-Aztecan

Uto-Aztecan is a very well-defined language family with over 100 solid Proto-Uto-Aztecan cognate sets reconstructed by Jane Hill (1999).⁶ But, like Indo-European, Uto-Aztecan has a slightly rake-like rather than tree-like basal structure to its phylogeny, with an early separation into two primary Northern Uto-Aztecan and Southern Uto-Aztecan subgroups with fairly similar levels of internal diversity (Campbell 1997: 134). This suggests that the proto-language spread fairly rapidly and quite far, a circumstance offering no obvious homeland area or centre of diversity. Past linguistic opinions on homeland have ranged from Oregon to northern Mexico, with apparent centres of gravity of opinion located in the Mexico–Arizona border region, or in eastern California or the Great Basin. Most scholars so far have favoured a placement of Proto-Uto-Aztecan firmly amongst archaic hunter-gatherer cultures, thereby assuming that early Uto-Aztecan populations adopted agriculture some time after their initial dispersal.

However, the most recent reworking of Uto-Aztecan lexical data by Jane Hill (in press) suggests that Proto-Uto-Aztecan had at least eight cognate sets related to maize cultivation, with some cognates occurring in Hopi and possibly even Numic. Hill suggests that Proto-Uto-Aztecan dates to between 5,600 BP, when

⁶ I wish to thank Jane Hill of the University of Arizona for providing me with unpublished manuscripts on her current Uto-Aztecan research.

domesticated maize first appears in Mesoamerica, and 4,500 BP. This is, admittedly, an archaeologically based date, but it is supported by new, and quite revolutionary, archaeological evidence which indicates that by at least 3,500 years ago an economy incorporating canal-irrigated maize cultivation was expanding through northern Mexico into southern Arizona. This represents almost a doubling of the previously accepted chronology for canal irrigation in the South-West. Recently discovered settlements of large size in the Tucson Basin of southern Arizona dating to between 3,200 and 2,500 BP indicate the strength of this economy in demographic terms (Mabry 1997, Muro 1998, 1998–9). The Santa Cruz Bend site near Tucson covered about eight hectares by soon after 2,800 BP, and this site and the nearby site of Las Capas in the Santa Cruz valley have circular post-walled houses and bell-shaped storage pits.⁷ Las Capas has well-dated irrigation canals from about 3,200 BP, the oldest in North America, and a remarkably early appearance of pottery at c.2,900 BP.

Although we cannot know for certain what languages these ancient maize-cultivators of northern Mexico and the South-West spoke, I know of no evidence to negate the possibility that they were mostly Uto-Aztecan, despite the existence of other highly localized modern pueblo populations who speak Tanoan and Keresan languages and Zuñi, all seemingly unrelated to Uto-Aztecan (Tanoan perhaps related at a very distant remove).

The hypothesis that Uto-Aztecan initial dispersal was fuelled by irrigated maize agriculture leaves one puzzling question. How can we explain all the Uto-Aztecan-speaking hunter-gatherers in eastern California and the Great Basin (the Numic, Tubatulabalic, and Takic speakers)? In my view, like the Southern Maoris, they presumably spread into environments where agriculture was not possible. Indeed, for the Numic speakers of the Great Basin the archaeological record actually provides some (albeit disputed) evidence for dispersal at about 1,000 BP (Bettinger and Baumhoff 1982). The Takic and Tubatulabalic speakers must have made similar adaptations in an earlier period, to judge from their higher levels of linguistic diversity. Ethnographic records of Paiute cultivation and irrigation of wild seed plants may thus reflect their ancestry in one-time agriculturalist populations, and not just the results of late borrowing.

This overall population-expansion-based scenario for Uto-Aztecan origins and history, centred on an expansion of agricultural populations from northern Mexico at about 3,500–4,000 BP, is not a new one. Romney published a version of it in 1957, and it has been favoured more recently in publications by Berry (1982, 1985). But it is interesting to note that most archaeologists and linguists alike have shied away from it during the past few decades, preferring instead on non-compelling linguistic grounds to regard Uto-Aztecan as a product of differentiation commencing with

⁷ Some of these observations I owe to a visit to the Las Capas site (Muro 1998–9) in the company of Jonathan Mabry of Desert Archaeology Inc. in Tucson, and Bill Longacre of the University of Arizona in Tucson.

in situ Archaic pre-agricultural populations in the South-West who are stated to have adopted maize and its cultivation from Mexico through diffusion (e.g. Plog 1997 for a recent archaeological perspective). When systematic maize agriculture was believed to have begun in the South-West only about 2,000 years ago, this was a viable suggestion. Now that maize cultivation is almost twice as old, and perhaps getting older (Mabry, p.c.), an agricultural background for Uto-Aztecan dispersal becomes an attractive hypothesis instead.

3. A biological comment

The general hypothesis supported in this chapter suggests that quite a large part of the biological patterning visible in modern populations in agricultural latitudes owes its existence to agriculturalist dispersal from ancient homeland regions such as the Middle East, China, and Mesoamerica. If the hypothesis is correct, then 'Caucasoid' population distributions reflect a high degree of Neolithic dispersal from the Middle East, likewise 'Mongoloids' from the Yellow and Yangzi river regions, 'Negroids' from West Africa, and 'Melanesians' from New Guinea. The Americas, settled by a Mongoloid population only about 12,000 to 15,000 years ago, have not had time to form such deep-seated biological divisions as populations in the Old World, so related observations do not hold there so clearly.

From this perspective, there should exist correlations between language families and racial populations on the fairly widespread level observed, for instance, by Cavalli-Sforza and colleagues (1994). But common-sense observations suggest that the reality will be complex. Inter-marriage between partners from two separate populations will give roughly a 50:50 genetic balance in the genes of their offspring, but the languages of the two parents will hardly ever blend into a 50:50 pidgin. We have only to note situations in which peoples of different biological origins speak languages which are closely related to understand that languages need not spread entirely by demographic dispersal of core populations of speakers. For instance, Javanese, Philippine Negritos, and most Island Melanesians speak Austronesian languages; Peninsular Malaysian Semang and Senoi, and Cambodians and Vietnamese, speak Austroasiatic languages; and so forth. Language shift, population bottlenecks, local factors of natural selection (e.g. malaria) have all played their roles in confusing the correlations of language and biology.

Major correlations and minor non-correlations between race and language are perhaps all we can expect from human history. If they are examined from broad rather than local perspectives, then it is my belief that the non-correlations become accessible to historical explanation. For instance, there are good reasons why so many Island Melanesians speak Austronesian languages, a family otherwise dominated by populations of clear and undoubted Asian genetic origin. Island Melanesia was a region of fairly dense existing population when the Austronesian dispersal began, partly a result of an independent growth of agriculture and arboriculture in and around New Guinea. Many Papuan speakers,

ever entrepreneurial, doubtless realized the advantages of learning an introduced language spoken right through the region (Proto-Oceanic and its immediate descendant dialects). Hence the significance of bilingualism and perhaps even language shift from a fairly early period. Superior Melanesian numbers in the intermarriage stakes would have turned the phenotype of the early Austronesians increasingly away from an Asian mode and towards a Melanesian mode. Indeed, it is interesting to note here a new analysis of a relatively early Lapita skeleton from Fiji, which is stated to be Asian in its affinities (Pietrusewsky, Hunt, and Ikehara-Quebral 1997). Later Lapita skeletons tend to be more Melanesian in phenotype.

The question here, of course, is why did not all the peoples of Indonesia eventuate as Melanesians as well, assuming that the original hunter-gatherer populations of Indonesia in the early Holocene were similar in phenotype to their equatorial neighbours in New Guinea and Island Melanesia? Archaeologically, Indonesia did not witness an independent development of agriculture, and many of the tree crops which supported the arboricultural trend in the western Pacific (sago, breadfruit, canarium) are essentially Wallacean/Melanesian rather than South-East Asian in origin. The pre-Austronesian peoples of Java, Borneo, and the Philippines perhaps had less interest in the Austronesian agricultural lifestyle than did the arboricultural Melanesians, and furthermore these South-East Asian islands lie much closer to the Asian mainland centre of gravity of agricultural population dispersal. Hence the replacement spread of former hunter-gatherers by an Asian agricultural population becomes stronger the further one moves north and west. The Philippine Negritos survived because of relative geographical isolation and, for some reason, several groups also adopted agriculture and Austronesian languages at an early stage (Headland and Reid 1989).

4. Conclusions

A chapter like this tends to be mainly one long conclusion. Agricultural dispersal surely mattered in prehistory, and its pattern-forming effects are still with us today. On a continental scale, archaeology and language correlations are eminently possible at the major language-family level. These correlations are not of the 'prehistoric pot style = prehistoric language' type which many have criticized in the past. I cannot guarantee that all the Lapita pots in the world were made and used by Austronesians, even though I am prepared to suggest that most were. I am essentially offering pan-continental generalizations. I do not believe one can understand the origins and dispersal patterns of major language families such as Indo-European or Austronesian if one only looks at restricted local data from within these families. A few trees do not equal the whole forest, and the total is not merely the sum of an infinitude of local parts.

In the case of Indo-European, a foundation dispersal into Europe commencing from the terminal Anatolian Pre-Pottery Neolithic, and progressing as an

agricultural horizon over a multitude of local Mesolithic cultures, makes far more sense to me (and to Renfrew 1987) than having the early Indo-Europeans as semi-nomads from the Ukraine overlording it over existing and long-established cultivators in the Early Bronze Age. This is because I cannot see why Indo-European, or any other major agricultural language family of a prehistoric antiquity, should be a total exception to those correlations between early *systematic* agriculture and language dispersal which appear to me to work well in other parts of the world. This does not mean that all the living subgroups of Indo-European were founded as separate entities in 7,000 BP. But it does suggest that Indo-European, like Austronesian, Bantu, and Uto-Aztecan, has its remote origins in an episode of punctuated population expansion which can be associated with the regional beginnings of systematic agriculture.

References

- BELLWOOD, P. 1996a. 'Phylogeny vs reticulation in prehistory', *Antiquity* 70.881–90.
- 1996b. 'The origins and spread of agriculture in the Asian-Pacific region', pp. 465–98 of *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, edited by D. Harris. London: UCL Press.
- 1996c. 'Early agriculture and the dispersal of the Southern Mongoloids', pp. 287–302 of *Prehistoric Mongoloid Dispersals*, edited by T. Akazawa and E. Szathmari. Oxford: Oxford University Press.
- 1997a. 'Prehistoric cultural explanations for the existence of widespread language families', pp. 123–34 of McConnell and Evans (1997).
- 1997b. *Prehistory of the Indo-Malaysian Archipelago*. 2nd edition. Honolulu: University of Hawaii Press.
- 1998. 'Human dispersals and colonizations in prehistory—the Southeast Asian data and their implications', pp. 188–205 of *The Origins and Past of Modern Humans—Towards Reconciliation*, edited by K. Omoto and P. V. Tobias. Singapore: World Scientific.
- BERRY, M. S. 1982. *Time, Space, and Transition in Anasazi Prehistory*. Salt Lake City: University of Utah Press.
- 1985. 'The age of maize in the Greater Southwest: a critical review', pp. 279–308 of *Prehistoric Food Production in North America*, edited by R. I. Ford. Ann Arbor: Museum of Anthropology, University of Michigan, Anthropological Papers 75.
- BETTINGER, R., and BAUMHOFF, M. 1982. 'The Numic spread', *American Antiquity* 47.485–503.
- BLUST, R. A. 1993. 'Central and Central-Eastern Malayo-Polynesian', *OL* 32.241–93.
- CAMPBELL, L. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.
- CAVALLI-SFORZA, L., MENOZZI, P., and PIAZZA, A. 1994. Editors of *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- DIAMOND, J. 1997. *Guns, Germs and Steel*. London: Cape.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- HEADLAND, T., and REID, L. 1989. 'Hunter-gatherers and their neighbours from prehistory to the present', *Current Anthropology* 30.43–66.

- HILL, J. In press. 'Why is Uto-Aztec so big?' *American Anthropologist*.
- HILLMAN, G. 1996. 'Late Pleistocene changes in wild plant-foods available to hunter-gatherers of the northern Fertile Crescent: possible preludes to cereal cultivation', pp. 159–203 of *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, edited by D. Harris. London: UCL Press.
- MABRY, J. 1997. 'Rewriting prehistory: recent discoveries at Cienega Phase sites in the Santa Cruz floodplain', *Archaeology in Tucson* 11: 3.1.
- MC CONVELL, P., and EVANS, N. 1997. Editors of *Archaeology and Linguistics*. Melbourne: Oxford University Press.
- MURDOCK, G. P. 1967. *Ethnographic Atlas*. Pittsburgh: University of Pittsburgh Press.
- MURO, M. 1998. 'New finds explode old views of the American Southwest', *Science* 279:653–4.
- 1998–9. 'Not just another roadside attraction', *American Archaeology* 2: 4.10–16.
- NICHOLS, J. 1998. 'The Eurasian spread zone and the Indo-European dispersal', pp. 220–66 of *Archaeology and Language*, vol. 2: *Archaeological Data and Linguistic Hypotheses*, edited by R. Blench and M. Spriggs. London: Routledge.
- PAWLEY, A. K. 1996. 'On the Polynesian subgroup as a problem for Irwin's continuous settlement hypothesis', pp. 387–410 of *Oceanic Culture History*, edited by J. Davidson. Dunedin: New Zealand Journal of Archaeology Special Publication.
- PIETRUSEWSKY, M., HUNT, T., and IKEHARA-QUEBRAL, R. 1997. 'A Lapita-associated skeleton from Waya Island, Fiji', *Micronesica* 30:355–88.
- PLOG, S. 1997. *Ancient Peoples of the American Southwest*. London: Thames and Hudson.
- POSSEHL, G. 1997. 'The transformation of the Indus Civilization', *Journal of World Prehistory* 11:425–72.
- RENFREW, C. 1987. *Archaeology and Language*. London: Cape.
- 1991. 'Before Babel: speculations on the origins of linguistic diversity', *Cambridge Archaeological Journal* 1.3–23.
- ROMNEY, A. K. 1957. 'The genetic model and Uto-Aztec time perspective', *Davidson Journal of Anthropology* 3.35–41.
- ROSENBERG, M., NESBITT, R., REDDING, R., and PEASNALL, B. 1998. 'Hallan Çemi, pig husbandry, and post Pleistocene adaptations along the Taurus-Zagros arc (Turkey)', *Paléorient* 24: 1.43–58.

An Indo-European Linguistic Area and its Characteristics: Ancient Anatolia. Areal Diffusion as a Challenge to the Comparative Method?

Calvert Watkins

I first examine in detail the characteristic features of a hitherto scarcely recognized ancient diffusional linguistic area, including both Indo-European languages of one subgroup and contiguous non-Indo-European languages. Chronological considerations are given particular attention, as well the apparent conjunction—and not disjunction—of areal development (diffusion towards a common prototype) and genetic development (language differentiation and the formation of species). I then present a classic case of diffusion of morphological or morphosyntactic features from one Indo-European subgroup (Anatolian) to another (Greek), geographically contiguous. The particulars involved are of theoretical significance: marked extension of the syntactic development of a native morpheme on the model of the syntactic function of a phonetically similar morpheme in the diffusing language. I claim that the comparative method can handle such phenomena. In support I examine some recent contributions by J. Heath to the theory of genetic morphological change in its historical dynamics, which present striking analogies to the cases of diffusional morphological change just examined. Heath invokes the biological model of punctuated equilibrium previously adopted by Dixon (1997), but with significant differences in ‘scenario’. These are then discussed and evaluated in the light of the Indo-European evidence presented earlier, and a fluid scenario presented combining genetic and areal development, both viewed as the legitimate purview of the comparative method and the comparativist historian of language.

My task in this chapter is to be the representative for Indo-European; on the one hand, as representative for some 75–100 related living languages (see Map 1), but on the other, as representative for ten or more related substocks (metaphorically called ‘branches’) (see Map 2), eight of which are still spoken (and only one

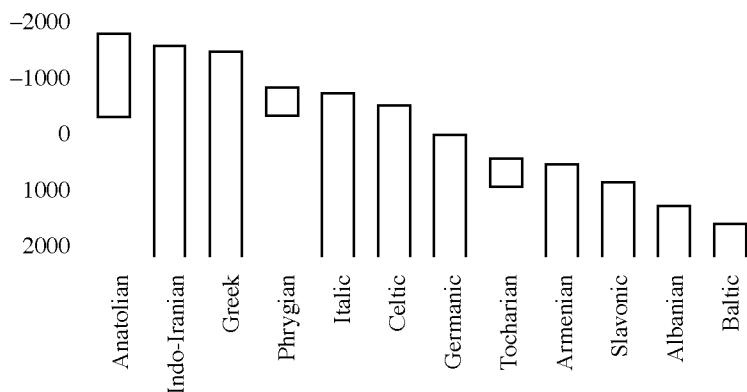
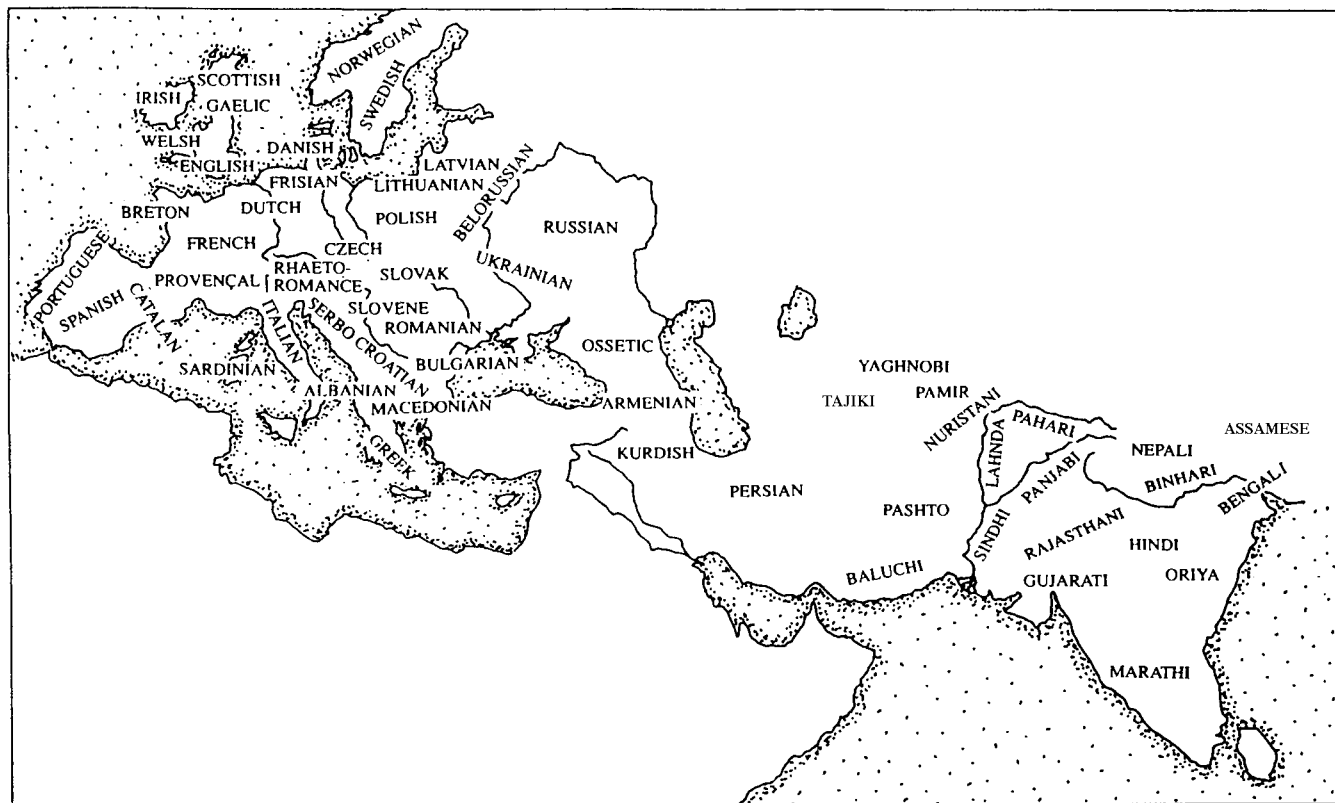


FIGURE 1. Branches of Indo-European by time of attestation

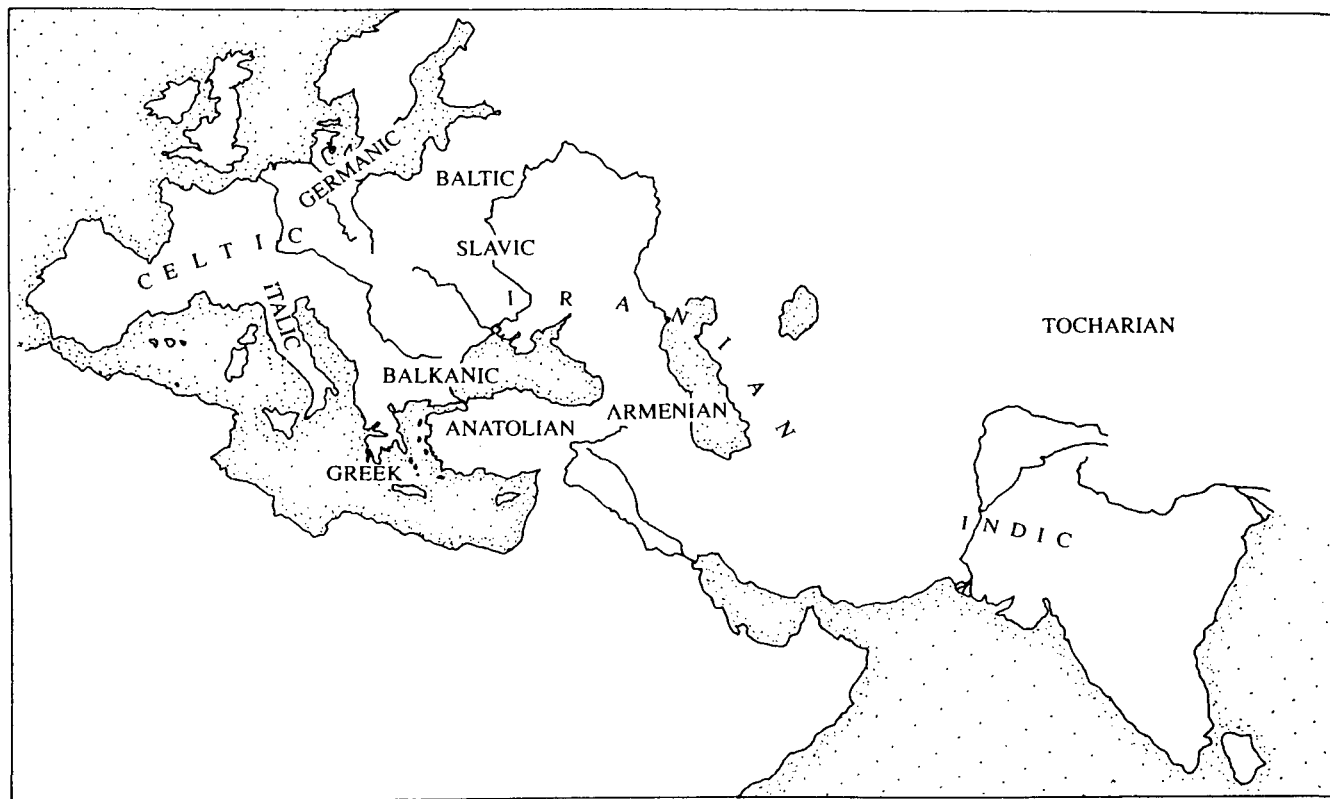
of which is endangered). Their collective documented history goes back nearly 4,000 years. A chart of these arranged in order of attestation is given in Figure 1. The value of this early and in some cases more or less continuous documentation is obvious and well known; by many reckonings, 4,000 years is more than half-way back to the time of the proto-language to which the whole family traces its ancestry. Equally well-known is the role of Indo-European in the development of the discipline of linguistics, both historical-comparative and theoretical. Linguistics became a science in the decade between 1870 and 1880, and it did so by figuring out (much of) the nature of the Indo-European proto-language and (much of) the process of its development into (most of) the several substocks. These facts and views are well known, if not wholly free of controversy; I will not discuss them further.

Far less familiar than the value of Indo-European as a laboratory for the traditional comparative method is the value of Indo-European as a laboratory for language contact and areal studies. Indo-European enters history as a contact phenomenon, viewed from the 'other' side: the two pre-Hittite loanwords in nineteenth-century BC clay tablets of the Old Assyrian (East Semitic) merchant colonies in central Anatolia. The borrowings are the words for 'contract' and 'night watchman' and their implications are clear enough: the Indo-European Anatolians are a law-abiding people, but watch your back. Similarly the earliest attestation of Celtic is on an Etruscan grave-inscription in a necropolis near Genoa from the sixth century BC, *mi nemetiēs*—'I am (the tomb) of Nemet(i)os'. It is culturally noteworthy that the name this Etruscanized Gaul chose to go by in Etruscan society was not his personal name, but an appellative giving his *status* in Celtic (Gaulish) society: Gaulish *nemeton* 'sanctuary' (neuter), Old Irish *nemed* 'privileged person'.

Regardless of the location of a putative 'homeland' or, as I have termed it, a 'staging area' for the speakers of Proto-Indo-European—which is a question I am



MAP 1. Indo-European languages in modern times



MAP 2. Indo-European languages at the beginning of the first millennium BC

not particularly interested in and will not further discuss—it is clear that none of the attested Indo-European subgroups originated where they are first historically attested. In each case, both in Europe and in Asia, they exemplify what Dixon (1997: 84) gives as an example of ‘punctuation’, *expansion into previously occupied territory*. Dixon goes on to say ‘the state of equilibrium in a certain area may be punctuated by invasion’. But it is clear that he is referring basically to modern times, the model of the Americas or Australia. We simply do not know (though we may sometimes guess) what the ‘state’ of the pre-Indo-European populations were in each case, in terms of equilibrium or punctuation. Certainly they differed considerably in such imponderables as ‘civilization’ or ‘cultural/artistic vigour’, if we think of, for example, the Helladic Mediterranean basin versus the Northern and Central European Neolithic. It is in any case gratuitous to assume either ‘equilibrium’ or ‘punctuation’ in the pre-Indo-European populations of which we know little or nothing. But what of the various post-Indo-European population groups? To speak of punctuation by ‘invasion’ prejudices the issue rather severely; the Indo-Europeanization of Italy and many other areas seems to have taken place both gradually and in dribbles. ‘Invasion’ may be an appropriate term for the coming of the Romans, then the Angles, Saxons, and Jutes, then the Vikings, and finally the Normans to the island of Britain; but so far as we can tell, ‘invasion’ was never an appropriate term for the coming of the Celts. Is it legitimate then to speak of a punctuation here? Compare what one of the leading Indo-Europeanists of this century, Emile Benveniste, wrote in 1939:

In their diversity these invasions have traits in common. They never involved vast movements of warriors. They are rather hardy little groups, strongly organized, founding their order on the ruin of established structures. They clearly knew neither the sea, nor cities. They have neither writing, nor a complicated religion, nor any sort of refinement. They will all preserve, along their individual destiny, the distinctive features of their first community: the patriarchal structure of the ‘extended family’, united in the cult of its ancestors, living from farming and animal husbandry; aristocratic style of a society of priests, warriors, farmers; ‘naturalistic’ worship and kingship sacrifice (of which the most significant was that of the horse, the Vedic *āsvamedha*); a conquering instinct and a taste for open spaces; a sense of authority and attachment to worldly goods. At the beginning they seem to be absorbed into the mass of often more civilized people which they have overwhelmed. A long silence follows their conquest. But by and by, from the new order which they found, there springs up a culture at first full of local elements, then developing in forms ever newer and bolder. An inventive power marks these creations, on which the language of masters confers the most perfect expression. The taking over of the land by ever newer invaders, but sprung from the same stock, thus creates the conditions for a supple and assimilative political organization, the home for a civilization vigorous enough to survive its founders, and original enough to influence permanently even what opposes it.

Even discounting the rhetoric of the eve of the Second World War, the picture to be gathered is more along the lines of relative stability, of ‘homoeostatic equilibrium’ as described by Dixon, than of punctuation. Clearly with each geographical

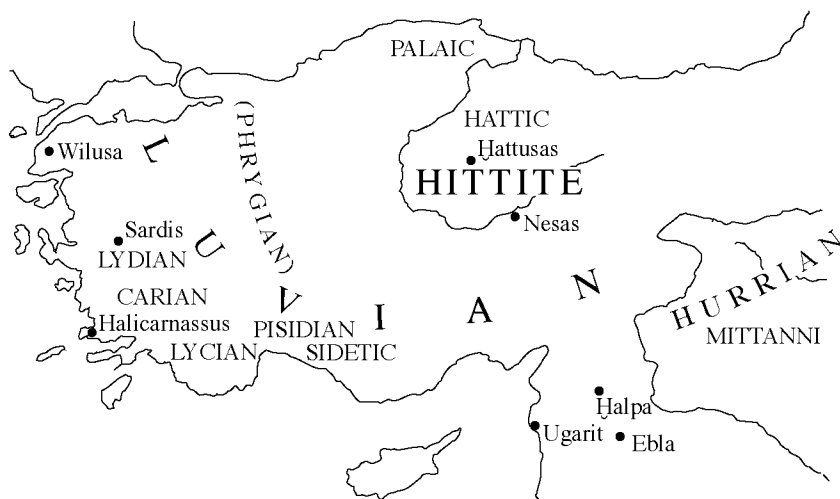
Indo-Europeanization a considerable amount of language contact must have taken place, with traces in occasional loanwords. In the south of Corsica the word (with definite article) *u djagaru*, from pre-Latin **īakar(u-)*, preserves a local Neolithic or Eneolithic word for ‘dog’. So for that matter English *dog* may preserve another. The Germanic word for ‘wife, woman’, which caught Sapir’s attention as a possible very early loanword, has now lost its isolation since a cognate has been identified in Tocharian.

But in most cases of the arrival of Indo-European speakers we cannot properly speak of the formation of linguistic areas, or of typological convergence towards a common prototype, which Dixon’s theory would associate with long periods of equilibrium. Nor for that matter are the time periods in question (e.g. 500–1,000 years between the ‘Indo-Europeanization’ of Italy and the actual attestation of Italic languages) long enough to ‘count’ as one of the kinds of equilibrium required to bridge 25,000, 50,000, or even 100,000 years of human history.

If the movement of Indo-European subgroups into the (already populated) territory they will be associated with (like Italy) counts as a punctuation, perhaps to be followed by a period of equilibrium à la Benveniste, it does not follow, nor is there any compelling evidence, that this punctuation leads to or results in the formation of species. The fragmentation of the proto-language probably preceded the movements in question, and the speciation of Italic into the Latino-Faliscan and Sabellic (Osco-Umbrian plus South Picene) sub-subgroups may well have taken place before the migration of either into the soil of Italy. In the case of Greek, and of the somehow closely related Armenian, no further speciation occurred at all, down to the present. Why and how this should be, no sociolinguist has ever explained to me. Perhaps this is how ‘language isolates’ were originally formed.

It would I think beg the question, and deprive the term of any meaning, to claim that each of the ten or more attested subgroups of Indo-European was a little linguistic area. But in at least one case that seems to be what happened. Yet the chronology is such that the linguistic area must have developed not over a long period of relative stability and ‘quiet’ diffusion, but quite rapidly, over less than a millennium at most, during which time the languages in question were *also* undergoing regular linguistic change and speciation. This seems directly counter to the Dixon hypothesis and deserves discussion.

The Indo-European subgroup in question is Anatolian. Since the facts about this extinct branch of the family are perhaps less familiar to non-specialists than most of Indo-European, let me briefly sketch them. Three languages written in cuneiform on clay tablets are attested in the second millennium BC: *Hittite*, the language of a great state (capital city Hattusas—Boğazköy) that flourished c.1600–1200 (periodicized into Old, Middle, and New/Neo-); *Palaic*, probably extinct by 1600 and preserved in a handful of cultic texts in Hittite context; and *Cuneiform Luvian*, in a few cultic texts in Hittite context 1600–1200. Geographically, compare Map 3. Hittite occupied the central Anatolian plateau;



MAP 3. The languages of ancient Anatolia

Palaic was spoken to the north in the Black Sea area; and Luvian was the language of Western and Southern Anatolia, from Wilusa (Troy) in the north-west to the Cilician gates in the south-east, and the most important second language of the Hittite empire, with a high degree of bilingualism.

It is likely that Indo-European languages of Anatolia entered from the west, across the Bosphorus, in the latter part of the third millennium; the different groups need not have entered at the same time. They are established *in situ* (see map) by the time of the Assyrian merchant colonies. (Phrygian, a different Indo-European subgroup, entered the same way at the end of the second millennium). The density of different languages in western Anatolia (like that of the Pacific North-West and California in North America) points to that region as the point of entry (or staging area) of these languages. The Luvian group includes the very similar *Hieroglyphic Luvian* written in a native pictorial syllabary (with logograms), used for monumental purposes and seals in Hittite context, but continuing in use in inscriptions from petty principalities in south-western Anatolia and northern Syria after the collapse of the Hittite Empire down to about 700 BC.

The remaining Indo-European Anatolian languages are known from a few inscriptions from classical times (sixth to fourth centuries BC) on the western and south-western coastal areas. *Lycian* and *Milyan* in Lycia in the south-west corner belong to the *Luvian* group. *Carian* (only recently read), the language of Halicarnassus to the north, and *Lydian* further on, the language of Sardis, may be independent descendants of Common Anatolian. For *Pisidian* and *Sidetic* on the south-western coast we have little more than names. Finally the position and

status of *Etruscan* (and the similar language of the Lemnos stele) remains controversial. It may be a heavily 'Anatolianized' non-Indo-European Asianic language, i.e. with areally diffused features. Note the Etruscan word for 'wine', *matu*, from a Luvian-like language: Cuneiform Luvian *maddu-*, Hieroglyphic Luvian *ma-tu-*, with a specifically Luvian sound change from Indo-European **médh-*, Greek *méthu* 'wine'. The semantics is a shared area feature of Greek and Western Anatolian; elsewhere the word means 'sweet; honey; mead', the English cognate. Both Greek and Common Anatolian also attest the widely diffused 'Mediterranean' wine word: Italic *uīnom*, Greek (*w*)*oinos*, Hittite and Luvian *wiyana-*.

To these languages of an Indo-European subgroup in Anatolia we must add several non-Indo-European languages with which the Indo-European groups were in intensive contact. In central Anatolia the autochthonous language was Hattic (Hittite *ḫattili* 'in Hattic', vs. *nesumnili* 'in Hittite', literally 'in the language of the inhabitants of Nesas = Kaneš = Kültepe') from whom the Hittites took their self-designation, as well as many cultural features of religion, the pantheon, and cult. Hattic is a language isolate; connection with some languages of the Caucasus has often been suggested (see Taracha 1995, 1998) but never proved.

From early in the second millennium the Hittites were in contact with various Semitic languages, beginning with the Old Assyrian of the merchant colonies. They learned to write Peripheral Akkadian, then Hittite on clay tablets not from the Assyrians but from Old Babylonian scribal schools, probably in northern Syria in the seventeenth to sixteenth centuries, with the first political expansion; Akkadian was the language of international correspondence and diplomatic relations in the second millennium, and the Hittite scribes were well versed in the East Semitic Sumero-Akkadian literary culture. The more southern Luvians were evidently in early contact with a West Semitic language, perhaps at Ebla, since the borrowing *ḫalal(i)* 'pure' (West Semitic *ḫll*, contrast Akkadian *ell[um]*) is thoroughly Luvianized by the beginning of our documentation. Later in their history the Hittites as administrators were in contact with another West Semitic literary language, Ugaritic of the north Syrian port city of Ugarit/Ras Šamra.

Finally we come to Hurrian, the language of the state of Mittanni in the east, with whom the Hittites were in close contact, first hostile and military and later cultural and religious, through most of their recorded history. Hittite religion and the pantheon underwent a profound Hurritization from the Middle Hittite period onwards, and the language played a major role in ritual and cult, with numerous loanwords of varying degrees of assimilation. Hurrian is an ergative language, with some thirteen cases. Transitive ergative verbs have ergative subject and absolutive object; transitive non-ergative ('antipassive') verbs have absolutive subject and 'essive' object; intransitive verbs have absolutive subject. Like Hattic it is a language isolate, save for the closely related Urartean of the Lake Van region in the first half of the first millennium BC. Recently C. Kühne at the July 1998 Rencontre Assyriologique Internationale has suggested that Hurrians migrated

from the east, perhaps from the Iranian plateau, towards the end of the third millennium. It was possibly in this region that the future Hurrian/Mittanni were first in contact with those Indo-Aryans who later appear as a superstratum (ruling class minority language) of the Mittanni state in the mid-second millennium.

Ancient Anatolia as a linguistic area is clear, and striking. We can observe remarkable convergences and innovations in all the languages of Anatolia, both Indo-European and non-Indo-European, both in phonology and in morphosyntax.

The first place to look in grammars for diffusional convergence is in the phonology, as Trubetzkoy noted long ago, and ancient Anatolian is no exception. Compare globally Melchert (1994). Consider the system of stop consonants. Proto-Indo-European had the traditional three series *t d dh*; already in Common Anatolian the latter two merged, yielding *t* and *d*. The correlation of voice was replaced by one of intensity (tense : lax), with the tense member realized with relative length, thus a tendency to an opposition geminate : simple. Word-finally there was probably since Indo-European times neutralization in favour of the voiced member. But more strikingly it appears that word-initially in Anatolian and there alone among all the Indo-European languages there was neutralization in favour of the unvoiced (tense) member. This explains why when the cuneiform syllabary was borrowed from Semitic, the Semitic voicing oppositions (e.g. TI vs. DI; the capitals denote values of syllabic signs) were ignored in favour of geminate versus simple: word-initial TI or DI to write the same word, but contrasting AT-TI or AD-DI vs. A-TI or A-DI. This system, and the same writing convention, is found in all the cuneiform languages, Indo-European and non-Indo-European alike. In Anatolia from the seventeenth century onwards Hittite at the centre, Palaic in the north, Luvian in the west and south, and Hurrian in the east showed the same distributional pattern of stops:

T-	-TT-	(-DD-)	
	-D-		-D

Only word-internally was there a contrast. It is a classic case of areal phonological convergence.

Yet the languages continued to evolve and change phonologically; there was no 'homoeostatic equilibrium'. A millennium later the alphabetically written Anatolian languages like Lydian and Lycian (the best documented and most clearly understood), while preserving the devoiced initial, had simplified the internal geminates and spirantized the voiced (lax) simplexes. Final stops were lost, and the resulting system was

T-	-T(-)
	-ð(-)

Compare the history of the Lycian word for 'son' *tideimi*, a participle of a reduplicated form of the Indo-European root **dheh₁i-* 'suckle, nurse' of Latin *filius*: Common Anatolian (third millennium) **di-dai-mnas* > Common Luvian

(second millennium) **tidaimmiz* > Lycian (first millennium) *tideimi* [tiðeimi]. This combination of simplification of geminates and spirantization of simple consonants is typologically quite similar to what happened in the history of the Romance languages and Celtic.

This is not the only area of convergence in the history of the consonants. From prehistoric times we can observe the effects of two opposing phonological processes which profoundly altered the distribution of the inherited obstruents: 'lenition' and 'fortition'. The lenition rule was that tense (geminate) stop became lax (simplex) after accented long vowel, and between unaccented vowels: or, more simply, in Adiego's recent formulation, between unaccented moras:

$$\begin{array}{l} T \rightarrow D / \check{V} \_ \\ T \rightarrow D / V \_ V \end{array}$$

In Luvian, and continued into Lycian, these rules generated important morphological variants, like *-ti* and *-di* (*-tti/-ddi* and *-t/-di*), *-ta* and *-da* (*-tta/-dda* and *-t/-da*) for the third singular present and preterite endings. In Hittite the effects were largely analogized away, but relic forms still attest it.

The opposite effect was fortition, which resulted in the multiplication of geminates. One such rule was 'Ćop's law': $\acute{e}CV \rightarrow \acute{a}C.CV$: Indo-European **médhu* > CA *médu* > Luvian *maddu* 'wine'. Indo-European, CA **mélit* > Luvian *mallid* 'honey'. In other cases the geminates reflect cluster assimilations, like $VC_1HV \rightarrow VC_1C_1V$: Indo-European **megh₂-* > Hittite *mekk(i)-* 'many', Indo-European **melh₂-o-* > Hittite *malla-* 'grind'.

A complex set of assimilation rules in the nominal morpheme chains in Hurrian similarly generated a large number of geminate (tense) consonants, especially continuants and sonorants, e.g. *-z* (ergative) + *nna* (enclitic 3sg object) → *-ssa*.

Another phonological development, with enormous consequences for the reconstruction of proto-Indo-European, probably has an areal explanation: the famous conservation of two of the three Indo-European laryngeals as consonants, tense *H* written *h-* *-h₂h-* and lax *h* written *h-* *-h-*. Their phonetic value is controversial. We find these in the three Indo-European languages of the second millennium, but also in Hattic and Hurrian. The different Semitic languages of culture with which Hittites and Luvians were in contact had also a rich repertory of laryngeals, which contributed to a favourable ambience for their conservation, wholly or in part, in Anatolian into the first millennium. On the other hand Cuneiform Luvian in the middle of the second millennium already shows sporadic laryngeal loss just as in other subgroups of Indo-European. (Laryngeal colouring took place already in the proto-language.)

The result of these is that the consonantal inventory and distribution in the three Indo-European cuneiform languages and both non-Indo-European, Hattic on the one hand and Hurrian on the other, is virtually identical, though from very different sources where we can know.

The development of the vowel system offers the same parallel and similar changes in the three Indo-European languages: the changes are posterior to Common Anatolian but fully accomplished by the time of their first attestation. These are basically to lengthen vowels under stress (both in open and in closed syllables, which is typologically rarer) and shorten unstressed long vowels. While we know little of the prehistory of Hurrian or Hattic, both show the same apparent correlation of length and stress, and show the same notation by the scribes of Hattusas. Our documentation of all five of these languages, Hittite, Luvian, Palaic, Hattic, and Hurrian is with few exceptions entirely from the archives of Boğazköy-Hattusas, written by Hittite and in some cases Luvian speaking scribes, whose spelling conventions are the same regardless of the language they are writing.

The resultant portmanteau inventory for all five languages of Anatolia, Indo-European and non-Indo-European, is

p	t	<u>ts</u>	k	(k ^w)	[+ tense, + long]	i	u	[+/- long]
b	d		g	(g ^w)	[- tense, - long]	(e) (o)		"
(f)		s	H		[+ tense, + long]	a		"
(v)		z	h		[- tense, - long]			
m	n				[+/- long]			
l	r				[+/- long]			
w	y							

The labiovelars are restricted to the Indo-European languages, and Hittite alone preserves both intact. The labial continuants *f* and *v* are found in Indo-European Palaic and non-Indo-European Hattic as well as in Hurrian. In Hittite they occur only in unassimilated loanwords. The vowel *e* is not found in Luvian, and *o* is apparently found only in Hurrian. Distribution and source of these speech sounds will vary from language to language, but the inventory is remarkably homogenous over Asia Minor throughout the second millennium.

In syntax as well as phonology the second millennium languages of Anatolia give the appearance of a partly convergent, diffusional linguistic area. Melchert (1994) identified three great syntactic isoglosses which set off the Anatolian subgroup from the rest of the Indo-European languages: (a) a split ergative system, with an ergative case for neuter nouns functioning as subject of a transitive verb and the development of enclitic subject pronouns used only in sentences with a subclass ('unaccusatives') of intransitive verbs; (b) development of enclitic 'chains' of particles and anaphoric pronouns after the first stressed word of the sentence; (c) the nearly obligatory use of phrase connectors (clause initial and enclitic) to link all the sentences of a discourse but the first. None of these features is found in any other early Indo-European languages; but to varying degrees they are present in both Hattic and Hurrian. Hurrian is an ergative language, as noted above, and Taracha (1995, 1998) claims the same for Hattic. In the new bilingual Hurrian-Middle Hittite texts, a Hittite ergative will translate a Hurrian ergative when the Hittite subject noun is of neuter gender; but when the Hittite subject

noun is of common (animate) gender, the nominative is used, which shows that the two systems are not superimposable.

For the enclitic pronoun and particle chains there are striking parallels both in Hattic (where they follow the sentence initial verb) and in Hurrian (where nominal forms can be followed by a 'Morphemkette' or morpheme chain). For the sentence connector note the semantic and syntactic identity of Hattic *pala/bala* and Hittite *nu*, both restricted to absolute initial position. While analogues both to the enclitic chain and to the sentence connectives and their mapping can be found in other early Indo-European languages, particularly Greek, the 'exuberant' development of these inherited materials is doubtless due to contact and diffusion: the Hittite usage of enclitic chains increases during the course of the second millennium.

The morphological consequences of these syntactic innovations, notably the system of split ergativity, was in all cases accomplished by reanalysis or reworking of inherited Indo-European material, not by diffusion of morphemes. The ergative case, Hittite *-anz* (regularly < **-anti*) and Luvian *-antis*, was probably extracted from the old ablative-instrumental of n-stems **-an-ti* by resegmentation. The enclitic subject pronouns were created just by substituting a nominative for the inherited third singular accusative pronouns: animate *-as* beside acc. *-an*; neuter *-ad* was both nominative and accusative. And the sentence-initial connectives represent syntactic redeployment of inherited deictic pronominal particles: Old Hittite *nu*, *ta*, *su* (replacing **sa* after *nu*), Luvian *a(-)*, of Indo-European **nu* 'now', **to-*, **so-*, **o/e-*.

Such then is the constitution of an Indo-European linguistic area: geographically bounded, and involving both three languages of an Indo-European subgroup and two further unrelated languages, each a language isolate. Of other non-Indo-European languages of Anatolia we simply lack enough information to say. The Indo-European subgroup Phrygian came in a thousand years after Anatolian, and is *not* a member of the second-millennium area. We find phonological and syntactic convergence toward a common type, to be sure, but all the languages preserve their individuality and their genetic identity. The 'common type' is in no sense actually achieved.

Furthermore it is significant that these convergences and parallel developments—in short, the formation of the linguistic area—all took place over a few hundred years at most. The convergent innovations are bounded by the arrival of four out of the five languages in Anatolia for contact to take place, say by 2000 BC, 2200 at the outside; and the convergent innovations are all completed by about 1700 BC, more likely 1900. That is hardly a long period of homoeostatic equilibrium, and would seem to me impressionistically to represent rather a period of punctuation, of rapid linguistic changes due to intensive language contact.

It is clear from the first millennium facts that the languages of Anatolia continued to evolve and change in a manner consistent with the ordinary tenets of the comparative method. The common, portmanteau phonological system of the

second millennium is quite different from those of the languages of Western Anatolia in the middle of the first millennium, where we find such hitherto unknown features as nasalized vowels, the changes of the obstruent system already noted, and profound alterations due to wide-ranging and rather spectacular syncope of unstressed vowels, and aphaeresis. Some of these may reflect or attest the formation of yet another, successive linguistic area in Western Coastal Anatolian, which might include the putative ancestor of Etruscan, if that turned out to be a non-Indo-European Asianic language. But if Etruscan was brought to north-western coastal Italy by migration from Western Anatolia, before or around the turn of the millennium, it is at least curious that Etruscan itself underwent by the sixth century a massive set of syncope very reminiscent of what happened to the Indo-European languages of western Anatolia at about the same time. It would be a classic instance of *areal* drift.

The classical Indo-European linguistic areas are the Balkans and India. In both of these, as in the case of ancient Anatolia, we can observe that the formation of each linguistic area must be a relatively rapid one, on the one hand, and on the other, that the languages involved, while showing characteristic features spread by diffusion, continue to evolve genetically and to maintain their individual identities. The Balkan areal features like postposed definite article in Romanian (Latin), Bulgarian and Macedonian/Makedonski (South Slavic), and Albanian (an Indo-European subgroup we can term Balkanic) have developed only posterior to the arrival of each into the area: Latin brought with the Roman conquest, Slavic by migration from the north by around AD 600. Albanian was probably earlier spoken to the east of its present location. But the postposed article is securely established by the time Albanian is first attested in the fifteenth century.

In India the development of the area is necessarily posterior to the penetration of the Indo-Aryans into the subcontinent some 3,000 years ago, and many of the areal features like the spread of retroflex consonants and the two varieties of causative are basically of post-Vedic date, i.e. post-500 BC, when Middle Indic begins. In each case the languages continue to evolve during the period of formation of the area: the fragmentation of Indo-Aryan into the varieties of Middle Indic and the modern languages all took place at the same time as the formation of the linguistic area.

We find, in short, no evidence for a disjunction of areal development and genetic development. Both go hand in hand across the limited but still significant time span that we can observe directly.

Let me conclude this glimpse at some consequences of language contact in the Indo-European family with another, which is neither familiar nor even for the most part ever treated in the specialized literature. It involves diffusion from one Indo-European subgroup to another, without the ultimate development of a real linguistic area, but offers a useful methodological lesson. It is *Greek* and *Anatolian*, which were in geographical contact in Western Anatolia during the

second millennium and perhaps—though this is controversial—even on the mainland and islands in the late third and early second millennia.

Greek forms a large subgroup with Armenian (and in part Phrygian) and Indo-Iranian on the basis of shared grammatical features, like the ‘augment’-prefix, the prohibitive negation, and the whole structural organization of the verbal system. This group forms the basis on which the proto-language was first reconstructed, and it is probably the most recent in time of the various ‘branches’ or subgroups of the family. Greek and Indo-Iranian also share the largest number of ‘poetic’ features of any pair in the family: the largest number of shared formulas (common stock phrases), and a uniquely shared system of quantitative metrics based on the alternation of heavy and light syllables. If for the family-tree model we substitute the schematic branching diagram used by Uralicists and some Indo-Europeanists, the most recent or latest would be the group on the right edge of Figure 2. Other models are possible if we want to avoid the family tree: I once fancifully suggested a sort of ‘cyclone’ image of the diaspora of Indo-European languages, Figure 3. Again the bottom or touch-down would be Indo-Iranian, Armenian, and Greek. In this model the cyclone itself has a geographic trajectory.

Whatever model we adopt, the linguistically somewhat distant Greek and Anatolian end up geographically contiguous: across the well-travelled Aegean sea. Mycenaean Greek colonies dot the southern half of the western coast of Anatolia, and North Greek, Aeolic expansion on the northern half is doubtless very old. These regions were partly Luvian-speaking, but Hittite political hegemony was established in the fifteenth century, weakened, and later reinforced. There was

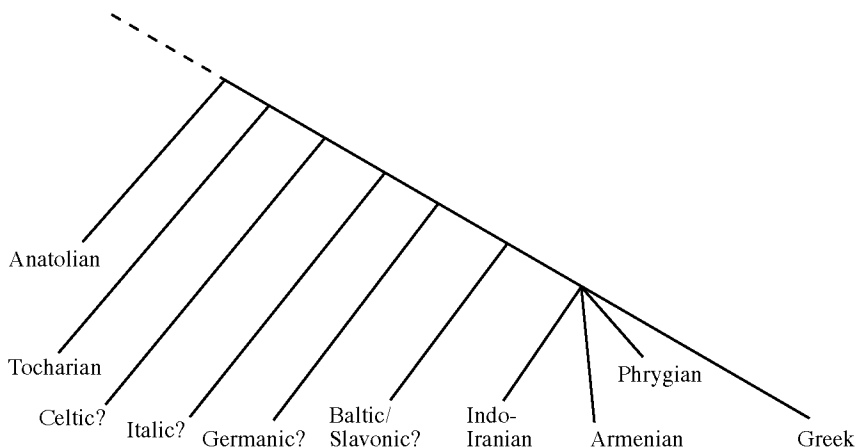


FIGURE 2. Schematic branching model of Indo-European subgroups

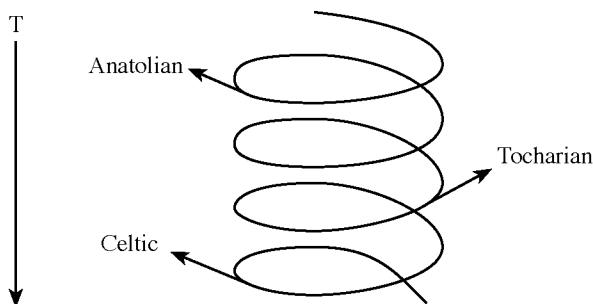


FIGURE 3. Fanciful 'cyclone' spin-off model of subgrouping

ample opportunity for intense local language and cultural contact. Mycenaean tablets from Crete and mainland Greece both attest the name *Aswiyo*s (feminine *Aswiya*): they are 'Aswians', refugees from the defeat of the Aswa (Hittite *Assuwa*) coalition of Western Anatolia by the Hittite king Tuthaliyas in the fifteenth century.

It should be noted that the ancestors of both Greek and of Anatolian, or of dialects of either or both, may have been in contact in the Balkans in the middle of the third millennium, on their way to their ultimate destinations. We simply cannot tell. Some scholars have suggested a prehistoric Luvian or Luvoid presence in Greece, and some even that the Cretan Linear A syllabary is Luvian writing. While the latter is so far unproven and unconvincing, it is superficially tempting to equate the mountain complex of *Parnassós* with the Luvian relational adjective genitive *parnassi/a-* to *parna-* 'house'.

Consider the following diffused morphological features. One dialect alone of Greek shows an iterative imperfective tense, marked by a suffix *-ske-* and the absence of the augment *e-*: the Ionic of Homer and the Asiatic coast (Western Anatolia). Hittite shows a semantically marked imperfective in *-ske-*, the same (inherited) morpheme. Luvian shows a cognate morpheme *-za-* in the same marked function. Either could have been diffused into Eastern Ionic Greek, which responded by extending the use of its cognate and phonologically similar native morpheme *-ske-*.

The Luvian languages mostly share the property that a derived inflected relational adjective fills the function of the genitive case in nouns. The derivational morphemes are Luvian *-assi/a-* or *-i/ya-*. Aeolic like the other dialects of Greek has a (cognate) relational adjective in *i(y)o-*; but only in Aeolic is the patronymic genitive of the father's name replaced by a relational adjective derived from the father's name.

These are only two diffused grammatical features, but they share the significant characteristic of showing marked extension of the syntactic deployment of a native morpheme on the model of that of the syntactic function of a phonetically

similar morpheme in the diffusing language. It is a type of diffusional grammaticalization. Both, I would submit, the diffusional and the genetic, may be discovered and handled by the comparative method. I thus take exception to Dixon's apparent claim that the comparative method is only applicable to genetic filiation. I see no principled reason to deny its applicability to areal diffusion, and suggest that is just what able and sophisticated practitioners of the method like Dixon himself are doing, when they speak of the fine-grained analysis required to discriminate between areal similarities and genetic similarities. Even the much-maligned family-tree model has a perfectly good notation for areal or other 'influence', the dotted line of the classical manuscript stemma which is the source of the family tree. And recall that Trubetzkoy observed sixty years ago that the competing 'wave theory' model was equally applicable to genetic filiation and areal diffusion. I believe that the resilience and the power of the comparative method lies in its sensitivity to similarity due both to genetic filiation and areal diffusion alike. Both are historical models, and the goal of comparison is history. This was demonstrated once and for all, as Stephanie Jamison reminds me, by Hübschmann (1875), when he proved that Armenian was a separate branch of Indo-European, and not a dialect of Iranian as previously thought. The Armenian language and its people had been Iranianized in language, culture, and religion for upwards of a thousand years by the time it was first reduced to writing.

My colleague Jay Jasanoff raises a further intriguing possibility of areal diffusion from Anatolian to Greek, which deserves to be sketched here. We know that the Western and South-Western Anatolian languages of the middle and second half of the first millennium BC, which as we saw constitute a new linguistic area phonologically, are not documented after the second century or so, though their onomastics lasted much longer. All of these languages, and indeed any others of most of Asia Minor, were sooner or later replaced by Greek, as is clear from the *Geography* of Strabo and the massive and extensive epigraphical and historical evidence of later antiquity and the Byzantine Empire. A preponderance of Greek speakers were residents of Anatolia. Now we can observe that a number of the phonological features which set off all dialects of later Greek from Classical Greek, like the spirantization of voiced stops and the voicing of unvoiced stops after nasal, are found also in the indigenous languages of Anatolia. The voiced spirants are discussed above, and for the voicing after nasal compare the spelling in Lycian of the Greek name Δημοκλ[ε]δης as *Ntemuxlida*, or that of the Persian emperor Dārayavauš as *Ntariyeusehe* (genitive). A similar change is first documented in Greek in the first half of the fourth century BC in the Pamphylian dialect, spoken on the south-eastern coastal area of Asia Minor where Anatolian Sidetic is found: Brixhe 3 πεδε (και δεκα Φετι[ι]α) 'fifteen years') beside Classical Greek everywhere πέντε. Perhaps the linguistic area of first-millennium Anatolia lived on in pan-Greek phonology. The first attestation of the Modern Greek spelling μπαρ 'bar' may thus be the Lycian abbreviation Mparahe (genitive) of the Iranian name Arttumpara 'Rta(m?)-bara'. Historians of the Greek language might take notice.

'The most important development in historical linguistics in the last decade or so has been a confluence between historical and typological lines of study.' So begins Heath (1997). The languages in question in Heath's paper are the non-Pama-Nyungan languages of Arnhem Land intensively studied by Heath and others in the 1970s and 1980s. They provided the material as well for his thesis, a study of linguistic diffusion in Arnhem Land, so he is well aware of the language-contact factor in historical development in the particular situation, and is careful to exclude it in this particular paper. His thesis is that '*in a stable sociolinguistic environment* [emphasis mine], the normal mechanism for renewal of a dysfunctional rich morphology is repair (formal renewal) of the weak links rather than the development of an entirely new morphology via grammaticalization'.

Intended as a companion-piece to this article is Heath (1998), dealing with the Takic subgroup of Northern Uto-Aztec languages. The two papers together offer not only 'colorful metaphors' (the author's phrase), but a profound and original contribution to the theory of genetic morphological change in the particulars of its historical dynamics.

This is not the place to discuss the details. Heath differentiates 'lost-wax' and 'hermit-crab' processes, in that 'lost-wax' upgrades a minor morpheme to a major morpheme if the latter is threatened, while 'hermit-crab' spreads fully functional independent stems into morphology to preserve threatened functional categories. But both are repair strategies whose goal is the preservation of the grammatical system; both are rapid and basically abrupt changes; and both involve creative redeployment—grammaticalization—of previously existing material whether bound morphemes of low function or fully functioning free forms. I recall what Ives Goddard referred to in the '70s at Harvard as 'Goddard's Law', something like 'a language can do whatever it wants to with whatever material it has to hand, if it wants to'. The two contact-induced grammatical changes in Greek discussed above show a number of similarities to Heath's processes. While 'system-altering' rather than 'system-preserving', they are both rapid and basically abrupt changes involving creative redeployment of previously existing material.

My real point in bringing up Heath's masterly applications of the comparative method—for they are that—is the following. In 'hermit-crabs' he adduces *exaptation*, introduced by Roger Lass from evolutionary biology, to reject it with the suggestion that *punctuated equilibrium* is 'a more useful borrowing'. Whether this is suggested independently or a direct or indirect response to Dixon (1997) I do not know. Heath's own formulation of the theory of punctuated equilibrium is that 'a biological species changes very little, except for an occasional burst of rapid genetic change as a new species is created. This model tries to account for both stability and change'. But then—and this is crucial for this volume—he goes on to add: 'A rough linguistic analogue might be rapid change in a short period of intense language contact, followed by a long era of continuity under monolingual conditions.'

It will be apparent that this scenario is in certain ways the inverse of Dixon's:

for Heath the ‘equilibrium’ is relatively static monolingualism, while the ‘punctuation’ is intense language contact, which is viewed as a catalyst for rapid language change and the formation of species. It is clear from the thrust of his paper that for Heath ‘(non-contact-induced) grammatical evolution’ can have the same result, and indeed lead to the formation of species, as in the case of his example of the Germanic dental preterite. Thus

	Dixon	Heath
punctuation:	rapid change due to non-linguistic causes	rapid change due to intense language contact or to rapid non-contact-induced grammatical evolution
equilibrium:	languages in contact converging towards a common prototype	relatively static monolingualism

I do not mean to suggest that Dixon’s and Heath’s are the only ways to apply the evolutionary biological theory to language history, nor that all the variables are so accountable. But if the theory of punctuated equilibrium is to be applied to language evolution—which seems to me a very promising suggestion, for which I for one am indebted to Bob Dixon—then we must consider what are the most plausible scenarios, in the light of whatever experience we have or can bring to bear on the question.

Towards that end let us consider some of the parameters involved in a comparison of Heath and Dixon. Such are

abrupt (rapid) change	:	gradual (slow) change
intense language contact and bilingualism	:	monolingualism

We have to inquire whether abrupt/gradual is the same as rapid/slow. We should inquire about degrees of language contact, intense vs. sporadic but steady. Other parameters include

language differentiation	:	convergence to a common type
formation of language families	:	formation of linguistic areas

which implies

genetic comparison	:	typological comparison

and the three-way distinction of

system-internal causality (linguistic)	:	contact-induced causality (linguistic)	:	external causality (non-linguistic)

Let us admit that the punctuated equilibrium model is a valid one for linguistic evolution, on which Heath, Dixon, and I are in complete agreement. I am perfectly

prepared to accept the impressionistic notion of relatively homoeostatic equilibrium, punctuated by periods of more rapid development and change of human language or languages, following the model of palaeobiology. The central problem is the combination of these parameters.

For Dixon, human linguistic history involved long periods of relative stability, 'homoeostatic equilibrium', with relatively minor changes, punctuated by periods of rapid and dramatic changes due to non-linguistic causes. During the periods of equilibrium linguistic areas develop with cross-language diffusion and convergence toward a common prototype, while during the periods of punctuation—and only then—the family-tree model was operative and valid.

As I stated in Paris (1997), while the biological model of punctuated equilibrium may be quite legitimately applied to language, I find some of Dixon's associated conclusions unconvincing. The reason lies in my own reading of the lessons of history in Indo-European, where the formation of diffusional linguistic areas is on the one hand relatively rapid (a matter of half a millennium or less), and on the other coexists with normal and relatively rapid genetic differentiations and the formation of species.

I believe the Indo-European examples show that both contact-induced linguistic change (i.e. diffusion) and system-internally driven linguistic change can occur with equal abruptness and rapidity—thus both counting as 'punctuation'. In some 4,000 years of attested Indo-European languages the only case known to me of something like 'equilibrium' is that of Iceland during the later Middle Ages and the early modern period, during which there was relatively little language contact, and monolingualism was the order of the day. But before the breakup of the subgroups of proto-Indo-European and before westward migration it can be easily imagined.

The language areas involving Indo-European languages have all been characterized by interdiffusion of grammatical features, but in none can we really speak of convergence to a common prototype, in the sense of loss of linguistic identity. I do not deny that this is possible, but it remains for me only a theoretical construct.

If biological equilibrium has an analogue in language, it is probably to be expected in the long stretches of the Upper Palaeolithic, where 'nothing much was going on' and human society and technology evolved at a snail's pace. Perhaps for these Dixon's model is entirely valid.

If pushed to the wall for an opinion—it is worth no more than that—I would picture the development of human language over the past 25,000 years—I would not want to go beyond that—as one involving the formation and development of genetic families and the formation and development of linguistic areas at the same time, with each having its own dynamic, its own history, and its own life and death. (Language areas do not necessarily last, cf. second-millennium Anatolia, and note that the features diffused from Western Anatolian to Greek were not 'areal' features. And both the family and the area were extinct by c.200 BC.) Both

genetic families and diffusional areas would have their own distribution of rapid abrupt and slow gradual change, and here we might see sequences of punctuation and equilibrium as well. 'Equilibrium' itself might not be so 'equi', if there is where we should look for the inherent imbalance that dictates the direction of language drift.

It remains the task of the comparativist-historian to sort all this out.

It may be that the 'classical' comparative method is not applicable beyond some 8,000 or 10,000 years, as Johanna Nichols has suggested; Kuryłowicz too once observed that we cannot reconstruct *ad infinitum* (which is the Nostratic fallacy, or one of them).

Among our parameters above was the traditional

genetic comparison : typological comparison

The *Comparative Method* is typically viewed as the technique of the former domain alone: but there is no principled reason to exclude it from the latter. The goal of genetic comparison is linguistic history, while that of typological comparison is often said to be linguistic universals. But one can and, I insist, must compare the components and manifestations of a linguistic area in order to draw *historical* conclusions. The comparative method, when properly handled, is sensitive enough to do both. It is linguistic comparison—comparative linguistics—which is the source of the distinction, the discrimination between areal similarities and genetic similarities. As I have said many times, the first principle of comparative linguistics is knowing what to compare.

References

- BENVENISTE, E. 1939. 'Les Indo-Européens et le peuplement de l'Europe'. Résumé. *Revue de synthèse, Synthèse historique*, p. 18.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- HEATH, J. 1997. 'Lost wax: abrupt replacement of key morphemes in Australian agreement complexes', *Diachronica* 14.197–232.
- 1998. 'Hermit crabs: formal renewal of morphology by phonologically mediated affix substitution', *Language* 74.728–59.
- HÜBSCHMANN, H. 1875. 'Über die Stellung des Armenischen im Kreise der indogermanischen Sprachen', *Zeitschrift für vergleichende Sprachforschung* 23.5–49.
- MELCHERT, H. C. 1994. *Anatolian Historical Phonology*. Amsterdam and Atlanta: Rodopi.
- TARACHA, P. 1995. 'Zum Stand der hattischen Studien', pp. 351–8 of *Atti del II Congresso Internazionale di Hittitologia* (Studia Mediterranea 9), edited by O. Carruba, M. Giorgieri, and C. Mora, Pavia: G. Iuculano.
- 1998. 'Neues zur Sprache und Kultur der Hattier', *Orientalistische Literaturzeitung* 93.9–18.
- WATKINS, C. 1998. 'La linguistique comparée en 1997: quelques réflexions', *Actes de 16e Congrès International des Linguistes*. Paris: CD-ROM.

The Australian Linguistic Area

R. M. W. Dixon

The two hundred and fifty or so languages of Australia make up a large linguistic area of considerable time depth.¹ We can recognize a number of low-level genetic groups (here called subgroups²), each due to recent expansion and split but on a small local scale. There is no clear evidence for higher-level genetic grouping.

Australian languages are characterized by a number of parameters of variation, most of which have an areal distribution. Each has its isogloss and the isoglosses do not bunch. This suggests that the distribution of such features is the result of separate processes of diffusion (and reinforces the impossibility of recognizing higher-level genetic links). Languages tend to move in cyclic fashion through the values of each parameter of variation.

There are a number of small relic areas, whose languages show archaic characteristics. As other languages move, and come into contact with these relic areas, widespread linguistic features are likely to diffuse into them.

The time depth is so great that there is insufficient evidence to help us decide whether all the languages come from one ancestor, or whether there were several genetic origins, with the original genetic groupings having become blurred through tens of millennia of diffusion, during what was more or less an equilibrium situation.

An appendix examines in some detail the two versions of the 'Pama-Nyungan' idea, demonstrating that it is without any scientific basis whatever (and can be followed only as an 'article of faith').

I thank Alexandra Aikhenvald, Lyle Campbell, and Alan Dench for providing constructive comments on a draft of this chapter.

¹ Digraphs are used as follows: *dh*, *nh*, and *lh* for lamino-dental, *dj*, *nj*, and *lj* for lamino-palatal, *rd*, *rn*, and *rl* for apico-postalveolar (retroflex) stop, nasal, and lateral. And *rr* for an apico-alveolar rhotic (generally a trill) with *r* used for an apico-postalveolar rhotic (generally a continuant). A glottal stop is shown as '. An affix boundary is indicated by '-' and a clitic boundary by '='.

² The term 'subgroup' is generally used for a low-level genetic grouping within an established language family. 'Subgroup' is employed here in a slightly different manner, for a low-level genetic grouping within the Australian linguistic area, which may or may not go back to one genetic family.

1. Introduction

Archaeologists have shown that people were in Australia at least 40,000 years ago, and probably 50,000 years ago. It might have only taken about 2,000 years for Aborigines to spread right across the continent (Birdsell 1957). At the end of this period of expansion and split—and for a few thousand years after—it would presumably have been possible to represent the relationship between the various languages through a family tree diagram. Once Australia was fully populated there would have been an equilibrium situation. So long as physical conditions (rainfall and the like) remained constant, the level of population would have stayed much the same, and probably also the number of languages. (See Dixon (1997) for an exposition of the Punctuated Equilibrium model of language development, in terms of which the discussion in this chapter is cast.)

But no human situation is ever static. We have no evidence of any major punctuation (for example, an aggressive and successful invasion from outside, or major conquests within the continent) but there would have been continual shifting around, with minor expansions and contractions of tribal groups, leading to some language splits, and some language extinctions.

During an equilibrium period many cultural features are likely to diffuse, eventually reaching every ethnic group making up a given geographical region. Technical and social features that have diffused over a continuous region (but not over the whole continent) include the boomerang, customs of circumcision and subincision, the section system, and—most recently of all—the subsection system (see McConvell 1985).

Many kinds of linguistic feature are particularly open to diffusion (in Australia and elsewhere). These include phonemic contrasts, syllable structure and the placement of stress at the level of phonology, plus structural profiles such as head- or dependent-marking, a system of noun classes, switch-reference marking, and ways of marking possession. Table 1 summarizes some of the features that are found in all or most of the languages of the continent, and also some that are in all or most of the languages of a specific region within Australia.

It would be mind-numbing to have to continually refer to each of the c.250 languages of Australia as an individual entity. For ease of reference I have organized them into fifty groups, labelled A–Y, WA–WM (where W stands for West) and NA–NL (where N stands for North); each group includes between one and twenty-three languages, shown by a number after the group name. For instance areal group W consists of two languages, W₁, Kalkatungu, and W₂, Yalarnnga. Subgroup B, North Cape York, consists of further subgroups Ba, Northern Paman, and Bc, Wik, and also Bb, which is a single language, Umpila. There are six languages in Bc—Bc₁, Wik-Ngathrr, Bc₂, Wik-Me'nh, and so on. A summary list of languages is included at the end of this chapter.

Some of these groups are low-level genetic subgroups, some are minor

diffusion zones (that is, small linguistic areas); others are simply grouped together on a geographical basis. Note that the quality of the material available varies. There are good to very good descriptions for about seventy languages with another twenty to thirty descriptions said to be in preparation. At the other extreme we have only word lists, and minimal grammatical information, for over forty languages.

There is one difficulty that dogs all work on comparative Australia. For any significant point of similarity between two languages there are three possible explanations; it is always difficult—and sometimes impossible—to decide between them. The similarity could be a genetic retention, something that was present in a common ancestor of these two languages and has been inherited by both. Or it could be something that has been borrowed from one language to another (one then needs to inquire into the direction of borrowing). Or it could be the result of parallel development (sometimes called ‘convergent development’). Two languages (often, but not always, two languages of the same genetic group) may share an inner dynamic that propels them to change, independently, in the same way. One example is the independent development of the second person singular verbal ending *-st* in English and in German (Greenberg 1957: 46).

There are many examples in Australia of parallel development. We encounter the dropping of the initial consonant of a word (sometimes with consequential paradigmatic augmentations) in several geographically distinct regions (see the map in Dixon 1980: 198). Most Australian languages have their initial syllable stressed, with the stress peak coming rather late in the syllable; the stress relates to the vowel and the syllable-closing consonant (if there is one) but not the syllable-initial consonant. As a consequence of this, word-initial consonants are at risk to be dropped; this has happened independently in about ten small regions across the continent (some involving just one dialect of a language, others several languages).

Australian languages appear to have an inner dynamic that propels them towards developing bound pronominal clitics or affixes (see Map 2 below). Once a development of this nature has actually taken place in one language it is likely to diffuse rapidly among neighbouring languages. They are simply borrowing something that they would have been likely, given time, to have developed for themselves.

With careful scholarly attention it is often possible to make an informed decision as to whether some similarity between languages X and Y (whether or not contiguous) is a mark of close genetic linkage, or is due to diffusion (recently, or further in the past) or is an instance of parallel development. But sometimes it is not possible to decide between these alternatives.

Table 1 summarizes a sample of recurrent features.

TABLE 1. Some characteristic areal features

A (for 'all')—the feature applies across the whole continent, not necessarily in every language, but in some languages from every region. In some cases a figure is given (e.g. *c.*80%) indicating the approximate proportion of languages the feature is found in.

R (for 'region')—the feature is found in one (or a few) continuous regions and is an areal feature for those regions.

D (for 'discontinuous')—the feature is found in a few languages or small groups of languages, spotted across the continent, rather than making up a solid geographical block.

Phonology

- 1 A, *c.*98% A nasal corresponding to every stop.
- 2 A, 100% At least four places of articulation for stops and nasals (best specified in terms of active articulator): labial, dorsal, apical, and laminal.
- 3 R A contrast between apico-alveolar and apico-postalveolar (retroflex) stops and nasals (and laterals)—everywhere except in a strip down the east coast and some languages in groups NH, NBL, and X.
- 4 R A contrast between lamino-dental and lamino-palatal stops and nasals (and laterals)—in two large areas, one on the west coast and the other comprising an east-central block, plus a couple of small areas elsewhere (see the maps in Dixon (forthcoming)).
- 5 A, *c.*98% Two semivowels, *w* and *y*. (Note that two contiguous languages in groups NF and NG, and a geographically separate language in WH, each have three semivowels: dorsal-labial *w*, lamino-palatal *y*, and lamino-dental *yh*.)
- 6 A, *c.*85% Two rhotics (grooved-tongue sounds), one articulated further forward in the mouth (generally an apico-alveolar trill or tap) and one further back in the mouth (apico-postalveolar or semi-retroflex, generally a continuant, sometimes a trill).
- 7 D Three rhotics, in *c.*7 distinct geographical regions.
- 8 A, *c.*75% A single series of stops; and no fricatives.
- 9 D Contrastive series of obstruents (either fortis and lenis stops, or stop and fricative, or two series of stops plus fricatives), found in *c.*60 languages in *c.*16 distinct geographical regions.
- 10 A, *c.*80% Basic syllable structure CV(C), all words beginning in a single consonant (not with a vowel or a consonant cluster).
- 11 D Initial dropping (of C or CV from the beginning of a word) has taken place in about ten distinct geographical regions (sometimes in just one dialect of a language, but at other times over all the languages in a small diffusion area). This leads to atypical word structures, e.g. CCVC(C) or V(C)CV(C) and to new paradigmatic distinctions, e.g. stop contrasts, fricative phonemes, additional vowel phonemes.
- 12 A, *c.*67% System of three vowels, *i*, *a*, and *u*. [All examples of different-size systems are the result of phonological change—just two vowels for some dialects in group WL (also suggested for NBd3, Aninhdhilyagwa) and more than three vowels in some languages of groups A, B, D, E, NB, ND, NG–NJ, NL (plus a few odd languages in J, M, T, U, WE, WM, NA, NE).]

Nouns

- 13 A, 90%+ Nouns take derivational suffixes between root and final inflection: typically, genitive, comitative, privative, dual, plural (may all be followed by case inflection).
- 14 A, c90% Nouns take final case inflections to distinguish core functions; generally ergative case for A, absolutive case (with zero realization) for S and O functions.
- 15 R Two groups of languages (in WH and NA) have shifted to a case system with nominative for S and A and accusative for O.
- 16 R Many languages with well-developed head-marking have lost case-marking of NPs in core functions—in WJ, NB, ND, NG, NI, NK, NL.
- 17 A, c.98% Nouns have case inflections for dative, purposive, and instrumental functions and also for local functions including locative, allative, and ablative.
- 18 A, c.85% Instrumental function marked by the same suffix as ergative (where there is an ergative suffix).
- 19 D If in a non-prefixing language instrumental is not the same as ergative, then it coincides with locative—in eight languages, scattered across the continent.

Pronouns

- 20 A, c.85% Singular/dual/plural number system in pronouns.
- 21 R Minimal/unit-augmented/augmented pronoun system, where the minimal terms are 1st person, 2nd person, and 1st-plus-2nd person ('me and you') plus, in some languages, 3rd person; unit-augmented involves one participant added to the minimal set and augmented more than one—in three areas: (i) NE in the north-west; (ii) B in the north-east; and (iii) some languages from WJ, NB, NH, NI and NL in the central north.
- 22 D Inclusive/exclusive distinction for 1st person dual and/or plural; found in about two-thirds of the languages with a singular/dual/plural system
- 23 A, c.75% Free and bound non-singular pronouns show a nominative (S and A)/accusative (O) case system.
- 24 D Singular free pronouns have different forms for each of the three core functions, S, A, and O; this is an archaic feature found in about eight geographical enclaves across the continent.
- 25 D Free pronouns follow an absolutive/ergative system, like nouns (all languages with this feature have bound pronouns, with nominative/accusative forms)—in a geographical block encompassing WI, WJ, some dialects of WD; plus scattered languages elsewhere (including P, W, WA/WB).

Verbs

- 26 A, c.98% Final suffixal inflection, indicating tense and/or aspect plus imperative mood.
- 27 A, c.80% Derivational suffixes between root and final inflection, generally including valency-decreasing derivation(s) (covering reciprocal and usually also reflexive).

- 28 D Some languages have developed either a reflexive/reciprocal pronoun, in place of verbal derivations; or just a reflexive pronoun, retaining a verbal suffix for reciprocal (but never the reverse, with reciprocal pronoun but reflexive derivational suffix to the verb)—these pronouns are all individual developments, in languages spotted around the continent.
- 29 R Prefixes to verbs (and sometimes also to nouns), always including bound pronominal prefix for S and A functions, and generally also one for O function; in most cases there is also a TAM prefix, often fused with the pronominal prefixes—in WMa, NB–NL.

FORMS
Lexemes

- 30 A *mayi* ‘vegetable food’—in 17 of the 38 groups A–Y, WA–WM and in 6 of the 12 groups NA–NL.
- 31 A *dhalanj* ‘tongue’—in 29 of the 38 groups A–Y, WA–WM and in 6 of the 12 groups NA–NL.
- 32 A *dirra*, *lirra*, *rirra* or *yirra* ‘tooth’ (sometimes extended to ‘mouth’)—in 23 of the 38 groups A–Y, WA–WM, and in NB and NH.
- 33 A *bu(-m)* ‘hit’—in 28 of the 38 groups A–Y, WA–WM and in 7 of the 12 groups NA–NL.
- 34 A *na(-ŋ)* ‘see’—in 4 of the 12 groups NA–NL and in W and X; and *nha(-ŋ)*—in 29 of the 36 groups A–V, Y, WA–WM.

Case allomorphs

- 35 A Ergative allomorph related to **-lu*, on demonstratives, interrogatives, proper nouns, kin terms, generic nouns and pronouns; allomorph related to **-dhu*, on other nouns—one or both of these is found in about 80% of languages with ergative case marking (see Sands 1996).
- 36 R ergative allomorph *-ŋgu* (developed from **-dhu*, probably by different routes in different regions)—in (i) c.25 languages in WD, WE, WG–WM; (ii) in c.30 languages in B–K plus the adjoining Nd and W; (iii) in Mgi, Gumbaynggirr. (See the Appendix.)

Pronoun forms

- 37 A 2sg *ŋinj-* in about half the languages in NA–NL; 2sg based on **ŋin* in c.95% of the languages in A–Y, WA–WM.
- 38 A 2n-sg *nu-* in c.70% of the languages in NA–NL (and in X); 2n-sg *nhu-* in c.60% of the languages in A–W, Y, WA–WM.
- R 1du(inc) *ŋali* in c.80% of the languages in A–Y, WA–WM, but in none from NA–NL. (See the Appendix.)

This is just a selection of the recurrent features in the Australian linguistic area, or in sub-areas within it. Others that could be added to the list include: lateral consonants, all words in a language ending in a vowel or all ending in a consonant, vowel length, glottal stop as a syllable prosody, stress placement, aversive case marking (‘for fear of’) on nouns, classifiers and noun classes, kin-determined pronouns, transitivity classes of verbs, nominal incorporation in verb stems, associated motion derivational suffixes to verbs (‘done while going/coming’, etc.), number and person neutralization in bound pronouns, interrogatives/indefinites, deictics, verbless and copula constructions, marking of negation, switch-reference marking, types of subordinate clauses. Under recurrent forms we could add nominal/verbal purposive suffix *-ga*, reflexive/reciprocal verbal suffix **-dharri-*, imperative *-ga*, several more personal pronoun forms and about 130 lexemes (over 60 verbs, about 60 nouns, and about 6 adjectives).

2. Characteristic features

Within a linguistic area one gets strong diffusion of categories, structures, and construction types, but lesser diffusion of actual forms. This is illustrated in the Table, where the sample includes fewer examples of forms than of categories.

Australian languages show a pervasive tendency to assimilate contiguous segments. One typical assimilation is of an initial consonant to a following vowel. The morphemes which retain a consistent form across the continent include those where initial C and V have the same place of articulation (e.g. purposive suffix *-gu*, verbal root *bu-* 'hit', item 33 in the Table) or where the first vowel is *a* (e.g. *dhalanj* 'tongue', item 31 in the Table). In addition, an initial stop or nasal (or a medial stop) may lenite to the corresponding semi-vowel (*g* or *ŋ* to *w*; *dj*, *nj*, *dh*, or *nh* to *y*).

Examples of assimilation and lenition include (all forms are attested in modern languages):

(1) 'give'	<i>nju-</i> > <i>yu-</i>	(2) 2sg pronoun	<i>ŋin</i>
	V		V
	<i>ŋu-</i> > <i>wu-</i>		<i>njin-</i> > <i>yin-</i>

The second singular pronoun can take ergative suffix *-du* and then we also find vowel assimilation: *ŋindu* > *ŋundu*, *njindu* > *njundu* and *yindu* > *yundu*.

In a non-homorganic nasal-plus-stop cluster the nasal often assimilates to the place of articulation of the stop, or vice versa. Thus, the verb 'fall' is *bunga-* in languages from group WD; *bunda-* in group N, and *bugga-* in groups B, P, U, WB, WH, WI, and NB. We can recognize one form as original, inasmuch as it is plausible for the other forms to have developed from it by phonological changes. It is likely that the original form here was *bunga-* with assimilation of *-ng-* either to *-ŋg-* or to *-nd-*. This word must have diffused widely, with the assimilations applying sometimes before it was borrowed into another language, sometimes after.

One set of striking cognates which illustrates all of these assimilations is the verb which means 'laugh', 'play', or 'dance'. The original form is most likely to have been *ginga-*. Attested forms in modern languages are:

- (3) *ginga-* in (one or more languages from groups) E, J, WM
 gingi- in WJ
 ganga- in WG (and *gangi* or *gangi-* in U)
 ginda- in J, M, N, V, WA
 gindi- in M
 gingga- in WA
 ginggi- in WA
 djinga- in K, WA
 djingga- in WA, WJ (*yiga-* in WD may also be cognate)

We here get *-ng- > -nd-*, *-ng- > -ngg-*, *gi- > dji-*, and—in the indented examples—vowel assimilations *i-a > i-i* and *i-a > a-a*.

3. Two parameters of variation and cyclic change

We will here briefly comment on a parameter concerning verbal organization, in §3.1, and one concerning bound pronouns, in §3.2.

3.1. VERBAL ORGANIZATION

Many, but not all, Australian languages (those in the shaded areas of Map 1) have two kinds of verbal element: (i) a simple verb, which takes TAM and other suffixes (and also prefixes, in prefixing languages); and (ii) a coverb, which generally takes no affixes at all. A clause may include just a simple verb, or else a simple verb plus a coverb, each contributing something to the meaning of this ‘complex verb’ constituent. (Generally, a coverb cannot occur alone but requires an accompanying simple verb.) Basically, a simple verb has a broad, general meaning, and a coverb adds further specification to this.

For example:

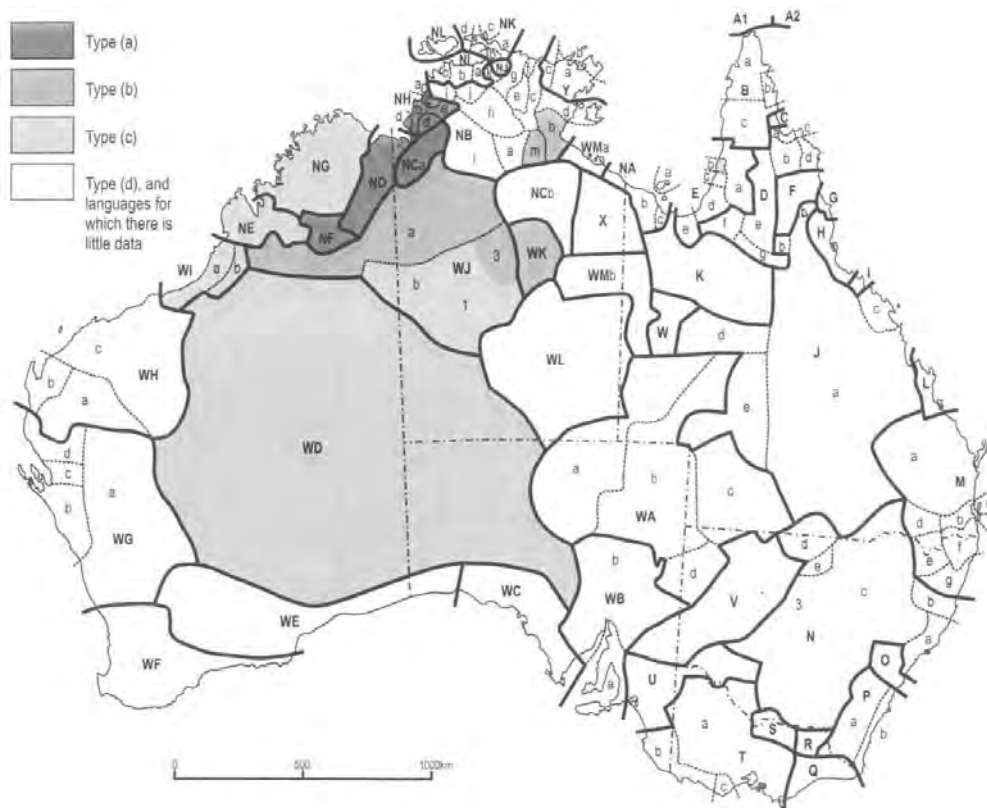
(4) NBl2, Wardaman (Merlan 1994)

complex verbs				simple verbs	
(coverb plus simple verb)					
(i)	<i>gabgab</i> -bewe-	‘wobble about’		-bewe-	‘tread’
	<i>gabgab</i> -bu-	‘waver, shoot and miss’		-bu-	‘hit’
(ii)	<i>wirrinjma</i> -gi-	‘turn’		-gi-	‘put’
	<i>wirrinjma</i> -ya-	‘be/get dizzy’		-ya-	‘go’

One can perceive a meaning element common to each pair of complex verbs. In (4i), with coverb *gabgab*, there is the idea of unsteadiness; and in (4ii), with coverb *wirrinjma*, there is the idea of rotation.

Languages can roughly be divided into four types:

- Just a few simple verbs (generally from five to about thirty) and many coverbs. All simple verbs can occur with coverbs, making up complex verbs, which are much more common than simple verbs in texts.
- Between about thirty and sixty simple verbs and many coverbs; some simple verbs can occur with coverbs, making up complex verbs, which are much more common than simple verbs in texts.
- A hundred or more simple verbs; just a selection of the simple verbs can occur with coverbs, making up complex verbs, which are more common in texts than simple verbs used alone.
- A large number of monomorphemic verbs (for instance, I have recorded over six hundred for H1, Dyirbal) and no coverbs (and also rather few compound verbs, no more than about 10% of the total).



MAP 1. Types of verbal organization

The occurrence of types (a), (b), and (c) is shown on Map 1; it will be seen that the types are distributed in an areal pattern. Languages of type (a) form a solid block, type (b) is next to it and then type (c); we have here a prototypical diffusion gradient. Subgroups NBb and NBm are type (b) and they are a little separated from other type (b) languages. There is also a language of type (c) in group Eb, in North Queensland, far away from the other languages of types (a–c).

Of course, nothing is static. There is evidence that Australian languages shift from one type of verbal organization to another and that this tends to take place in a cyclic pattern, roughly:



In type (c) there are hundreds of simple verbs, and generally only about a dozen of them occur with coverbs. These coverb-plus-simple-verb combinations are used much more freely than simple verbs. The simple verbs that do not occur in complex verbs are likely to gradually drop out of use, so that a language goes from type (c), with hundreds of simple verbs, to type (b), with thirty to sixty simple verbs, to type (a), where the only simple verbs remaining are the few—between about five and about thirty—occurring with coverbs.

In a type (a) language each verb is clearly analysable into two components, coverb and simple verb. These parts will in time become phonologically fused and semantically blended so that it will not then be possible to analyse them into two components. Each verb will consist of a single morpheme, with an irreducible meaning. We would go from type (a), where coverb and simple verb are distinct elements, to type (d) where almost all the verbs are, synchronically, monomorphemic.

This direction of shift in verbal organization is particularly evident in some languages from the prefixing area. There are two kinds of change:

- (i) A coverb-plus-simple-verb combination coalesces into a single verb.
- (ii) Bound pronominal clitics develop into pronominal prefixes to the verb.

The original structure would have been:

- (6) coverb [simple verb]-plus-suffixes

In some languages change (ii) has applied but not change (i), so that we get:

- (7) coverb prefixes-plus-[simple verb]-plus-suffixes

This is exemplified by (8) from NBl2, Wardaman (Merlan 1994: 265) where the coverb *worlag* bears no affix but simple verb *-bu-* (meaning ‘hit’ when used) has a pronominal prefix and a tense suffix:

- (8) *wongo worlag ya-bu-n gunga*
 NOT wash 3sg-'hit'-PRES 3sg+DAT
 'She is not washing for her.'

In other languages change (i) has applied—with coverb and simple verb becoming a compound—before change (ii), giving:

- (9) prefixes-plus-[coverb-fused-with-simple verb]-plus-suffixes

We can give an example here from NBg1, Mayali:

- (10) *ḡaban-dulu+bu-n*
 1minA+3augO-shoot-NON.PAST
 'I am shooting them.'

In this language the simple verb *-bu-* (meaning 'hit' when used alone) is fused with what we can assume was an original coverb *dulu*, to form a verb *-dulubu-* 'shoot', to which prefixes and suffixes are added.

Languages in groups NBC, NBe-i, and N1b-c (to the north of the shaded region in Map 1) are of this kind. They have moved from type (a) to or towards type (d).

The other change shown in (5) is from type (d) to type (c). This would involve a language with many monomorphemic verbs (each with a rather specific meaning) investing a handful of the verbs with a general meaning, and using them in compound constructions with a coverb. There is evidence that this is happening in Nc3, Ngiyambaa. An adverbial constituent has as its first element a manner adverbial morpheme and as second element one of thirteen generic verbs, some of which are cognate with simple verbs in the language (see Donaldson 1980: 201–24). This illustrates how simple verbs may have their meanings generalized, and be used in a verbal combination (in this instance, with an adverbial element) which could be the first stage in the development from a type (d) to a type (c) system of verbal organization.

The cyclic changes set out in (5) are essentially due to the inner dynamic of languages. There can also be changes due to geographical diffusion of a structural profile. WJb is a low-level subgrouping including WJb3, Warlmanpa, which has about forty-three simple verbs, and WJb1, Warlpiri, with about a hundred and twenty simple verbs (both languages have many complex verbs). It is likely that Proto-WJb was of type (b), like Warlmanpa, with forty or so simple verbs. Warlpiri has increased the inventory of simple verbs, moving into type (c). Nash (1982) suggests that this may have happened through Warlpiri having (i) reanalysed some coverb-plus-simple-verb combinations as new simple verbs; and (ii) accorded simple verb status to what were coverbs, so that they now take the suffixes associated with simple verbs. It is likely that Warlpiri increased the number of simple verbs under areal pressure from its south-westerly neighbour WD, the Western Desert language, which is of type (c), with a couple of hundred simple verbs.

We thus get movement (c) > (b) > (a) > (d) > (c) due to the internal dynamic of languages, and movement (b) > (c)—and presumably also (a) > (b)—due to the areal diffusion of a structural profile.

It is likely that Australian languages have been shifting around the cyclic parameter in (5)—in either direction—for a very long time, perhaps tens of millennia.

(A full discussion of this topic will be found in Dixon (2002: 183–201). The summary here is considerably truncated and simplified, omitting a number of intermediate types of verbal organization. It simply presents the essence of the matter.)

3.2. DEVELOPMENT OF BOUND PRONOUNS

It seems clear that Australian languages were originally all of the dependent-marking type. But they do exhibit a marked tendency to develop bound pronouns. These can be clitics to the verb, or to some other constituent in the clause, or they can be affixes to the verb. The distribution of bound pronouns (shown on Map 2) suggests that there have been parallel developments in a number of distinct areas; and that when bound pronouns are innovated, they then tend to diffuse.

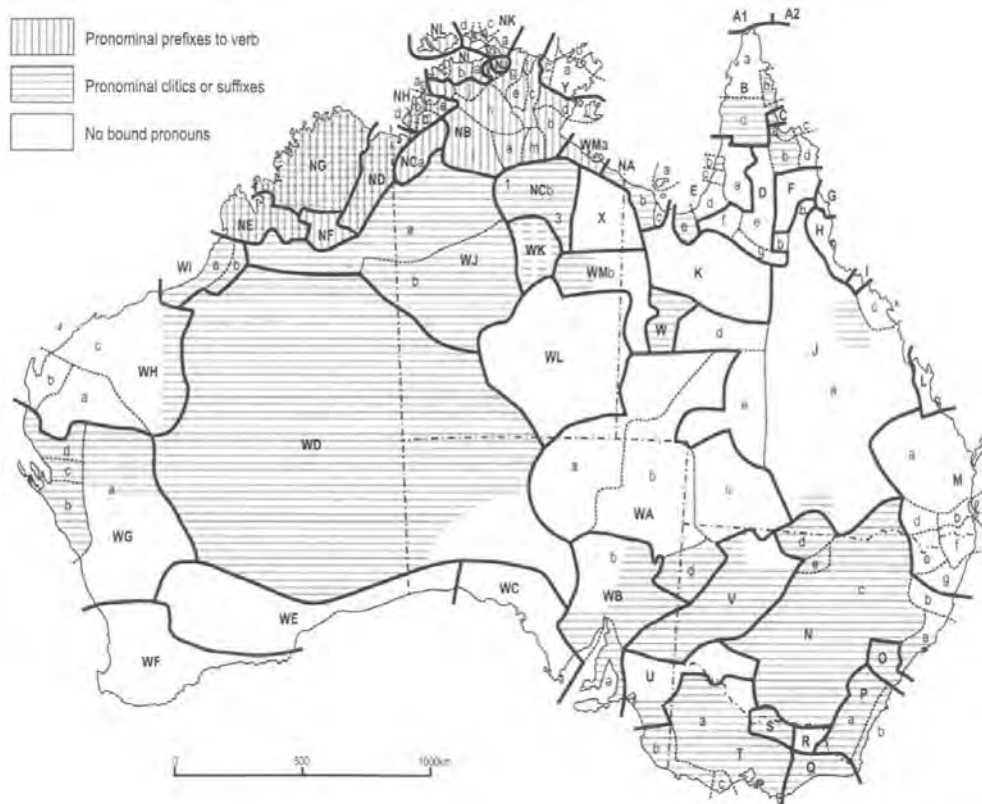
The stages of development are clear:

- (a) No bound pronouns at all, just dependent-marking.
- (b) Bound pronouns as enclitics are at first transparent reductions of free forms. They may be added to the first constituent of the clause, or to the end of the verb, or (in languages from subgroups Yc and Bc—see the discussion of Bc under (ii) below) to the end of the word immediately preceding the verb. In some languages bound pronouns are added to an ‘auxiliary’ element, which generally bears information about TAM; the auxiliary-plus-bound-pronoun constituent may go into any of the three positions just described.
- (c) When pronominal enclitics are added to the verb, after TAM, they may develop into suffixes and may fuse with the TAM suffixes.
- (d) From functioning as enclitics to the word immediately preceding the verb, bound pronouns (or TAM-auxiliary-plus-bound-pronouns) may develop to be pronominal prefixes (or pronominal-prefixes-fused-with-TAM) to the verb.

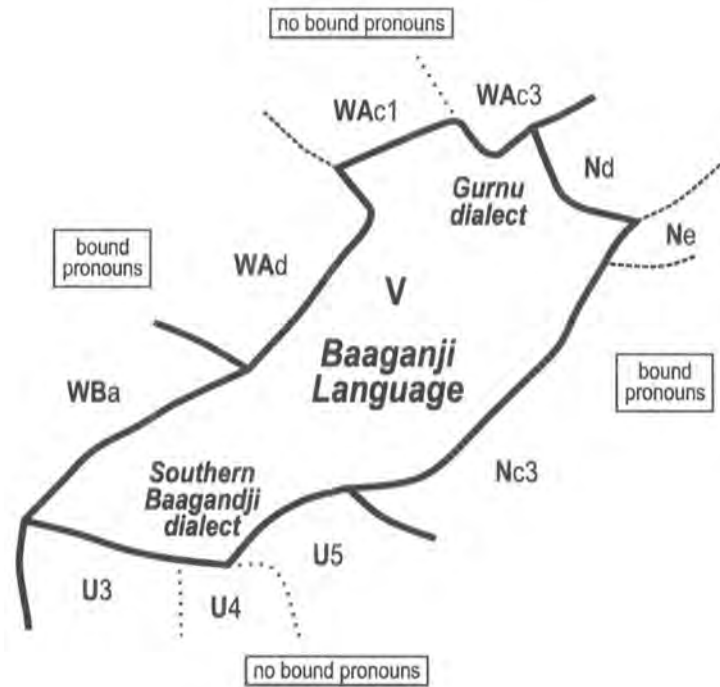
We tend to get developments:

$$\begin{array}{l} (a) > (b) > (c) \\ \text{or} \quad (a) > (b) > (d) \end{array}$$

There are, however, variations on these themes. We will briefly describe: (i) the development of bound pronouns back into free pronouns, i.e. (a) > (b) > (a); (ii) the development of pronominal enclitics to be verbal suffixes, fusion with tense, followed by phonological reduction with loss of some information content, and then the evolution of a second set of bound pronouns, i.e. (a) > (b) > (c) and then



MAP 2. Distribution of bound pronouns



MAP 3. V, Baagandji, and neighbours (identified on summary list of languages)

(a) > (b) once again, all in the same language; (iii) the development of bound pronouns from being enclitics, to being prefixes to the verb, to becoming enclitics again, i.e. (a) > (b) > (d) > (b).

- (i) **BOUND PRONOUNS BECOMING FREE PRONOUNS** Baagandji is spoken over a considerable area on both sides of the Darling River in New South Wales. This is V on Map 2, shown in greater detail on Map 3. Its dialects are lexically very close but differ in a number of grammatical features, one of these being bound pronouns. By comparing dialects we can trace the evolution of bound pronouns in Baagandji, their fusion with tense suffixes, and then the reanalysis of tense-plus-bound-pronoun as a new set of free pronouns.

We can surmise that there were originally no bound pronouns. In the Southern Baagandji dialect, bound pronominal enclitics are generally added to a verb after tense inflection. As shown in (13), the pronominal enclitics are derived from free pronouns by omitting the initial consonant. Tense inflections include $\emptyset$ for present, *-d* for future and *-ngu* for perfect. Thus (using ‘-’ for an affix boundary and ‘=’ for a clitic boundary):

- (11) V, Southern Baagandji dialect (Hercus 1982: 198)

ŋiŋga-ŋgu=aba

sit-PERFECT=1sgS

'I sat (there, in the past but never sit there now).'

In the Gurnu dialect tense and bound pronouns have fused. We get, for instance, past-tense-plus-1sgS form *w-aba*. What is more, the fused constituent (erstwhile tense suffix plus bound pronoun) is now recognized as a separate word. That is

verb-tense+bound.pronoun

has become:

verb tense-bound.pronoun

The tense-plus-bound pronoun generally follows the verb, but it does not always do so. It can occur clause-initially, for emphasis, as in:

- (12) V, Gurnu dialect (Hercus 1982: 124)

w-adhu gaandi barlubarlu

PAST-3sgA carry small.children

'It was him that carried the small children.'

Note that a verb in Gurnu generally does not show any tense inflection (it may, just occasionally, include past marker *-dji*).

It is interesting to compare, in (13), a representative sample of free and bound pronouns in Southern Baagandji with the free pronouns (fused with tense) of Gurnu.

(13)	Southern Baagandji		Gurnu free pronouns		
	free pronouns	bound pronouns	present	past	future
1sg, SO form	<i>ŋaba</i>	<i>-aba</i>	<i>ŋ-aba</i>	<i>w-aba</i>	<i>g-aba</i>
2sg, SO form	<i>ŋimba</i>	<i>-imba</i>	<i>ŋ-imba</i>	<i>w-imba</i>	<i>g-imba</i>
3sg, AS form	<i>ŋadhu</i>	<i>-adhu</i>	<i>ŋ-adhu</i>	<i>w-adhu</i>	<i>g-adhu</i>

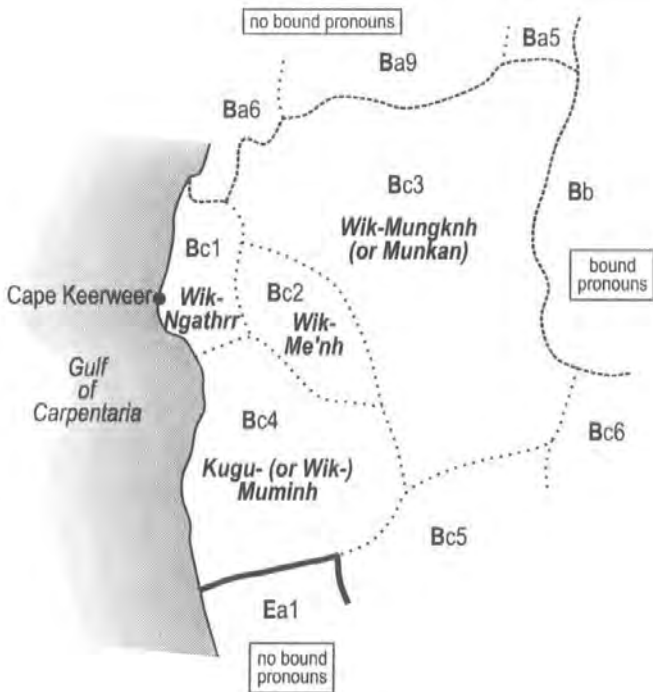
The original pronominal form is now the present tense pronoun in Gurnu, corresponding to zero inflection on the verb for present tense in Southern Baagandji. Past tense pronouns in Gurnu begin with *w-*, which may relate to the perfect inflection *-ŋgu* in Southern Baagandji. Future tense pronouns in Gurnu begin with *g-*; this is quite different from the future tense suffix *-d* in Southern Baagandji. However, this is unsurprising. The typical situation in Australia is for related languages (or even dialects) to have similar pronominal forms, and similar nominal affixes, but to show some differences in verbal inflection.

We thus have an example of free pronouns developing into bound pronouns and then back into free forms, i.e. (a) > (b) > (a). During this diachronic journey the pronouns picked up what was a tense suffix to verbs,

which has now become a tense prefix to pronouns. In fact, it is this which provides the clue that these there-and-back-again changes have taken place.

The changes may all have been due to areal influence. Languages to the east and west of Baagandji have bound pronouns and there may have been areal pressure from these directions for their innovation into Baagandji. Gurnu is spoken in the north-west of the language area, bordering languages that lack bound pronouns; the reinterpretation of bound pronouns (linked with tense) as free forms in Gurnu may have been due to areal diffusion from the north.

- (ii) REPEATED DEVELOPMENT OF BOUND PRONOUNS The six languages of the Wik group, Bc, constitute a low-level genetic subgroup. They have very similar grammatical categories and forms. All show bound pronouns, but of different types, indicating that these must have developed independently in each language. The locations of Bc1–4 (on which there is good data) are shown in Map 4.



MAP 4. Languages of the Wik subgroup, Bc

Bc1, Wik-Ngathrr, has pronominal enclitics which may be added to the verb, or to the word immediately preceding the verb, or (less commonly) to any word in the clause. These enclitics are transparent reductions from free-form pronouns, e.g. 2sg SA free form *nhunta*, enclitic *-nta*; 3pl SA free form *thana*, enclitic *-ána* (Sutton 1978: 244–5).

In Bc3, Wik-Mungknh, there is a full set of SA bound pronouns which have become suffixes to the verb, and are fused with tense. Verbal suffixes are a portmanteau of one of the ten pronominal categories plus one of the four TAM categories (present, past, future, or irrealis). In (14) we illustrate with six bound pronouns and two tenses, also giving the corresponding free-form pronouns for comparison.

(14) Bc3, Wik-Mungknh (Kilham *et al.* 1986: 406–7, Hale, MS)

SA argument	verbal inflections		free pronouns
	present	past	
1sg	<i>-aŋ</i>	<i>-aŋ(-an)</i>	<i>ŋay</i>
1n-sg.exc	<i>-an-an</i>	<i>-an</i>	<i>ŋan</i>
2sg	<i>-an-an</i>	<i>-an</i>	<i>nhint</i>
2pl	<i>-an-iy</i>	<i>-an</i>	<i>nhiiy</i>
3sg	<i>-an</i>	Ø	<i>nhil</i>
3pl	<i>-an-than</i>	<i>-(iy)in</i>	<i>than</i>

We can roughly recognize present-tense *-an* (except in 1sg) and past-tense Ø. But note that some of the bound forms show considerable differences from the corresponding free forms, and there is neutralization between 1n-sg.exc and 2sg in present, both being shown by *-an-an*. And *-an* covers 1n-sg.exc, 2sg, and 2pl in past as well as 3sg in present.

In Bc4, Kugu-Muminh (also known as Wik-Muminh, or Kugu-Nganhcara), there has been further phonological fusion and reduction of the tense/pronominal portmanteau suffixes to verbs, so that now only four pronominal categories are distinguished: 1sg, 2sg, 3pl, and an unmarked choice covering all other person/number combinations. There are three TAM choices—present, past, and irrealis (past and irrealis fall together for 1sg). The full paradigm of verbal inflections is given in (15).

(15) Verbal inflections in Bc4, Kugu-Muminh (Smith and Johnson 2000)

SA argument	present	past	irrealis
1sg	<i>-ŋ</i>	<i>-aŋ</i>	<i>-aŋ</i>
2sg	<i>-ŋan</i>	<i>-an</i>	<i>-nhun</i>
3pl	<i>-yin</i>	<i>-adhan</i>	<i>-nhin</i>
unmarked	Ø ~ <i>-an</i> ~ <i>-en</i>	<i>-a</i>	<i>-nha</i>

The phonological reduction of the bound pronominal forms has led to loss of information—only three specific person/number categories are marked, whereas free pronouns show eleven categories. As a response to this loss, the

language has evolved bound pronominal categories all over again. There is a full paradigm of pronominal enclitics, almost exactly mirroring the free pronoun paradigm and transparently reduced from it. The pronominal enclitics are generally added to the end of the word immediately preceding the verb (whatever that may be), although they can alternatively be added to the verb.

Free and bound pronouns each exist in three case forms (SA, O, and dative) in the singular and in two forms (SA and O/dative) in the non-singular. The pronouns show eleven categories—1st, 2nd, and 3rd person; singular, dual, and plural number; and inclusive/exclusive for 1du and 1pl. We illustrate free and bound forms in (16), with the same sample of person/number categories as in (14), in SA and in O function. (Note that Wik-Mungknh, shown in (14), differs from Kugu-Muminh in having a single 1n-sg.exc term, whereas Kugu-Muminh has separate 1du.exc and 1pl.exc.)

(16) Bc4, Kugu-Muminh—free and enclitic pronouns

	free		bound	
	SA	O	SA	O
1sg	<i>ɣaya</i>	<i>ɣanji</i>	<none>	- <i>nji</i>
1du.exc	<i>ɣana</i>	<i>ɣanana</i>	- <i>na</i>	- <i>nan</i>
1pl.exc	<i>ɣanhɬja</i>	<i>ɣanhɬjara</i>	- <i>nhɬja</i>	- <i>nhɬjara</i>
2sg	<i>nhinta</i>	<i>nina</i>	- <i>nta</i>	- <i>na</i>
2pl	<i>nhiya</i>	<i>nhiyana</i>	- <i>ya</i>	- <i>yara</i>
3sg	<i>nhila</i>	<i>nhunha</i>	- <i>la</i>	- <i>nha</i>
3pl	<i>thana</i>	<i>thaarana</i>	<none>	- <i>ran</i>

It can be seen from (15) that verbal suffixes indicate just 1sg, 2sg, and 3pl arguments in SA function. Pronominal enclitics lack any form for 1sg and 3pl in SA function. That is, they avoid repeating information that is already provided by the verbal suffix system. Only for 2sg do we find double specification.

These pronominal enclitics in Kugu-Muminh are similar in form and placement to those in Bc1, Wik-Ngathrr, which is in fact its north-westerly neighbour. We can ascribe the recent development of a second set of bound pronouns in Kugu-Muminh to two factors: (a) the need to replace the information lost by reduction of the original set of bound pronouns, now suffixes to the verb and fused with tense; and (b) diffusional pressure from a neighbouring language, Wik-Ngathrr, to have a full set of bound pronominal enclitics (in each language they are generally added to the word immediately preceding the verb, or to the verb itself).

- (iii) FROM PRONOMINAL ENCLITIC, TO PRONOMINAL PREFIX, AND BACK TO PRONOMINAL ENCLITIC The five languages in NC, the Mindi group, make up a genetic subgroup. This is one of only two examples in Australia of a

discontinuous subgroup (the other is WM). The western set of Mindi languages, NCa, is separated from the eastern set, NCb, by non-prefixing languages of group WJa and by prefixing languages of group NBl (see Green 1995).

NCa languages have a verbal organization of type (a), as described in §3.1, with between fifteen and twenty-two simple verbs that take pronominal prefixes and TAM suffixes. Simple verbs can occur alone, or with an immediately preceding coverb. Thus, to say 'die' one uses a non-inflecting coverb *digiridj* plus the simple verb with underlying root *ga-* (used alone this means 'go'). The simple verb takes a 3sg S pronominal prefix (underlying form *ga-*) and a TAM suffix. In (17), prefix, root, and suffix are fused together.

- (17) NCa1, Ngaliwuru dialect (Bolt, Hoddinott, and Kofod 1971: 126, 95)

<i>digiridj</i>	<i>gaydganj</i>
die	3sgS+'GO'+PAST
'He/she/it died.'	

The eastern block of Mindi languages, NCb, maintain the same basic structure with lexical verb plus simple verb constituent. But the number of simple verb roots has effectively been reduced to three—one indicating 'going', one indicating 'coming', and a neutral choice used in all other circumstances. What is more, the three roots have fused with tense suffixes. We thus get a 'simple verb' constituent (perhaps now better called an 'auxiliary' constituent) which effectively consists of a pronominal prefix (cognate with the prefixes in NCa languages) and a direction/tense suffix. The suffix forms in NCb3, Wambaya, are given in (18). Note that the suffixes in the 'going' column may have developed from a simple verb root *-ga-* 'go', as illustrated in (17) for NCa languages.

- (18) NCb3, Wambaya directional/tense suffixes in the auxiliary constituent (Nordlinger 1998: 146, 151)

	'going'	'coming'	neutral
future } - (g)uba	-ulama	$\left\{ \begin{array}{l} -u \sim \emptyset \\ \emptyset \sim -a \\ -a \end{array} \right.$	
present }			
past - (g)anj	-amanj		

In NCb1, Djingulu, the auxiliary constituent is encliticized to the verb (the original coverb). Thus, the original pronominal prefix to a simple verb could now be described as a prefix to a zero auxiliary root (which takes a directional/tense suffix), the whole functioning as an enclitic to the lexical verb. Alternatively, we could say that the bound pronoun—which was a prefix to a simple verb in Proto-NC (and still is in the NCa languages)—is now an enclitic to the only verb in its clause (the old coverb).

In Djingulu this pronominal enclitic must be added to the verb, but in Wambaya (Pensalfini 1997) the auxiliary follows the first constituent of the

clause, reminiscent of a typical position for pronominal clitics in non-prefixing languages. For example:

(19) Wambaya (Nordlinger 1998: 250)

<i>igima</i>	<i>g-amanj</i>	<i>yarru</i>	<i>nanga</i>
THAT.ONE	3sgS-‘COMING’+PAST	move	3sgm+OBL
‘That one came to him.’			

(Note that in Wambaya a monosyllabic auxiliary is an enclitic to the preceding word, while a polysyllabic auxiliary—as in (19)—constitutes a separate word, and bears its own stress.)

The NCb languages have thus come full circle. We hypothesize an earlier stage, (b), in which there were pronominal enclitics, probably added to the word immediately preceding the verb (as in Wik-Ngathrr and Kugu-Muminh, described under (ii) above). These would have developed into (d), prefixes to the simple verb, as in the modern-day NCa languages. Then most simple verb roots were lost and the two remaining were fused with tense; the old simple verb constituent (the new auxiliary) became encliticized to the old coverb (now the sole verbal element). (Not only do we get an affix becoming a clitic, but a *prefix* becomes an *enclitic*.) The old pronominal prefix to a simple verb has become an enclitic to the verb in Djingulu, back to stage (b). In Wambaya the bound pronominal element has detached itself from the verb, and now follows the first constituent of the clause (which is often a verb but need not be, as it is not in (19)).

The final change is undoubtedly due—at least in part—to areal pressure. Wambaya’s southerly neighbour is the non-prefixing language WK, Warumungu, which places its pronominal clitic complex after the first constituent of the clause. As we have said before, the movement from one structural type to another—within the parameters of variation that characterize the Australian linguistic area—may be motivated in part by the inner dynamic of a language (involving tendencies for parallel development) and in part by areal pressure to become more like neighbouring languages.

There are other examples of cyclic movement within the Australian linguistic area. These include: from an ergative to an accusative profile and then back again to ergative; and the change from classifiers to noun classes, and then loss of noun classes (under areal pressure). Note that there is continual reanalysis and remodelling, especially in pronominal systems (see, for example, Heath 1997).

4. Low-level subgroups and small linguistic areas

A preliminary point needs to be made. Unlike in some other parts of the world, there is in Australia no fundamental difference in the rate of replacement between core vocabulary and non-core vocabulary. If one compares 100 or 200 or 500 or

2,000 words between two Australian languages, one gets approximately the same percentage of similar forms (the variation will be no more than about 5%) so long as about the same mix of nouns and verbs is maintained in each size sample. A difference arises with respect to items from different word classes—nouns tend to get replaced at a faster rate than verbs.

When one language splits into two, the percentage of vocabulary shared between the new languages will gradually fall until it reaches about 50%. Since verbs are replaced at a slower rate than nouns, the verb score will generally be higher than that for nouns. And since grammatical forms tend to change more slowly than lexemes, the similarity of grammatical forms will be greater still. When two rather different languages come into geographical contact, they will borrow lexemes back and forth until they share about 50% of their vocabularies. However, the verb score is likely to be lower than this, and the similarity of grammatical forms lower still.

That is, once two distinct languages have been in contact for sufficient time—whether or not they come from a low-level common ancestor—their percentage of common vocabulary will tend to level out at around 50% (in practice, between 40% and 60%). Details of the calculations underlying this figure are in Dixon (1972: 331–7) and will not be repeated here. Alpher and Nash (1999) provide further discussion.

It is useful to calculate the percentages of shared lexemes (for general vocabulary and for verbs) between Australian languages. These figures can be suggestive of two kinds of groupings: one is low-level genetic subgroups; the other is small linguistic areas (within the larger Australian linguistic area).

Percentages of shared vocabulary can, at best, lead to a hypothesis concerning low-level subgroups. Proof of genetic relationship must then be provided by reconstruction of parts of the putative proto-language and of the systematic changes which have led to the development of each modern language within the subgroup. This work has been completed in some instances; in others it remains to be done (but there is every expectation that it will be possible to achieve this).

4.1. LOW-LEVEL GENETIC SUBGROUPS

The suggestion that Australia has been an equilibrium area for some tens of thousands of years implies that there has not been—within this period—any major punctuation, with one ethnic group expanding and splitting, leading to a family of languages (all descended from the language of the original ethnic group) spreading over all or most of the continent, and presumably replacing languages that were originally spoken there.

But the equilibrium hypothesis does not imply any sort of static situation. There would have been continual flux—between and within languages. Ethnic groups would have moved location, and some would have merged or split. There has been continual diffusion of cultural and linguistic features, across a given geographical region.

Note that the geographical environment has not been static. For instance, water resources have varied. Geographers believe that the land which supported perhaps one million Aborigines at the time of the white invasion (in 1788) would have provided for substantially less than that number 20,000 years ago (when it was drier, colder, and windier than today), but it would have supported more than the 1788 population 10,000 years before that.³ The Aboriginal population may have spread out over the whole continent; then, as water resources diminished, they might have contracted to the coast and major rivers. At a later period, with increased rainfall, they could have again populated the interior. And once new territory becomes available for occupation this naturally leads to an expansion of population with the likelihood of split of ethnic groups and languages—that is, a minor punctuation.

There are a number of clearly-defined low-level genetic subgroups among the c.250 languages of Australia. I have assessed these according to conservative criteria—considerable correspondence of grammatical and lexical forms such that it should be possible to reconstruct a good deal of a proto-language (but, of course, this needs actually to be done, in order to really prove the genetic relationship). Likely subgroups are marked with a star in the summary list of languages at the end of this chapter. There appear to me to be about thirty-seven low-level subgroups (about twelve in the NA–NL area, and about twenty-five outside this area), each consisting of between two and about seventeen languages. (Twenty-three of the subgroups each consist of just two languages.) Just on half of the languages can be assigned to one of these low-level subgroups. For some of the rest there is insufficient data on which to base a decision (or else there is insufficient data on a neighbouring language which is a possible genetic congener). For other languages it is clear that they have no close genetic relationship with any other language—these include C, Ubindhamu; Q, Muk-Thang (Gaanay); V, Baagandji; WK, Warumungu; and NL, Tiwi. There are just two subgroups that are geographically discontinuous—NC (discussed under (iii) in §3.2) and WM.

There are a number of subgroups which cover a significant extent of territory. The largest area is that of WD (which comprises a single language, the Western Desert language) but this is mostly desert and the population is sparse. Subgroups Ja (five languages), Nc (three languages), and Ta (three languages) are each spoken in fertile country and some of their languages relate to a dozen or more tribal groups, speaking mutually intelligible dialects of one language.

These low-level subgroups have plainly arisen as a result of minor punctuations in the fairly recent past.⁴ In some cases (e.g. WD) the expansion may have

³ Note also that the coastline was then further out, with Australia joined to both Tasmania and New Guinea until about 10,000 years ago.

⁴ I offer no suggestion whatsoever as to the periods of time that might be involved here (note that there could be a different period for each subgroup). There is nothing extra-linguistic against which the development of Australian languages can be calibrated, and any suggestions of an actual date can only be fanciful.

been in response to an increase in water resources, with territory becoming newly inhabitable. In other instances (e.g. Ta, Nc) there may have been no geographical trigger, but simply expansion of tribes and languages—either pushing other tribes out of their territory, or else assimilating them. But, as will be seen from the groups marked on Maps 1 and 2, these minor punctuations and expansions never encompassed more than a small part of the continent.

4.2. SMALL LINGUISTIC AREAS

There are a number of instances of the languages in a small region showing significant similarities to each other and considerable differences from languages outside the region. However, the similarities are not such as would permit the reconstruction of a common proto-language. That is, these languages do not comprise a low-level genetic subgroup. Rather, they make up a small linguistic area—the languages have probably been in their present locations, and in contact with each other, for a considerable period, so that a number of area-specific linguistic features have diffused across the region. As other languages move, and come into contact with these relic areas, linguistic features that are widespread outside the small areas are likely to diffuse into them.

One small linguistic area is group U, spoken on both sides of the lower Murray River. Another is group NH, spoken in the Daly River region of the Northern Territory. Other small areas include group E, on the western side of the Cape York peninsula in North Queensland; group T, in western Victoria; group WA in the Lake Eyre Basin and the Arandic languages, group WL, in central Australia. Alan Dench, in his chapter in this volume, discusses WH as a small diffusion area.

One fascinating small area is group W, consisting of W₁, Kalkatungu (for which there is good information, in Blake 1979) and W₂, Yalarnga (for which the data available is rather slender—see Blake 1971, 1989). These languages share around 43% general vocabulary, but only about 10% of their verbs are cognate and few grammatical forms are similar. Each is more similar to the other than to any neighbour (lexical scores with neighbours vary between 2% and 20%). It is clear that Kalkatungu and Yalarnga do not comprise a low-level genetic subgroup but they do appear to make up a small linguistic area. They have probably been in their present locations (and in contact with each other) for a considerable time.

There are a number of reasons why we suggest this. One of these is that Kalkatungu has bound pronouns, which are present in none of the neighbouring languages (save in Yalarnga, where there is just a trace). There are in fact three paradigms of bound pronouns, which are today used only in subordinate clauses (two sets) and for anaphora in main clauses (one set). By their form (some of them comprising just a syllable-closing consonant) the bound pronouns appear to be an ancient feature of the language. It is likely that they are gradually being lost—and have already been lost from the normal head-marking function in main clauses—under diffusional pressure from surrounding languages.

Another bit of evidence is the way in which other Aborigines regarded the Kalkatungu. W. Turnbull, a perceptive white settler who lived to the north of Kalkatungu country, wrote (1903: 10): ‘Now “Kalkadoon” is used as a term of reproach among the blacks, or rather I should say a term of contempt. A white man will call a “low” white a blackfellow, while the blacks call a low black a “Kalkadoon”’. Typically, the original inhabitants of a region are looked down upon by later arrivals. Note also that Kalkatungu territory is mountainous and relatively inhospitable, on the watershed between rivers that flow north to the Gulf of Carpentaria, and those that flow south to the inland lakes of South Australia. All this is consistent with the hypothesis that the Kalkatungu have been in their present region for a considerable time: indeed, they may originally have occupied a larger territory and then been pushed up into the mountains by other Aboriginal groups when they came into the region (compare with Basque, which used to be spread over a good deal of northern Spain but is now confined to the vicinity of the Pyrenees).

5. Conclusions

Australia constitutes a linguistic diffusion area, involving about two hundred and fifty languages, with some of them being related in low-level subgroups (many of these consisting of just two languages). There are many parameters of variation and many isoglosses may be drawn—for structural features such as bound pronouns or prefixing or noun classes or switch-reference; and for various recurrent lexical and grammatical forms; and for various phonological parameters. Many features have a continuous geographical distribution; others are scattered across the continent.

The point to note is that there is absolutely no bunching of isoglosses, which would be needed for high-level subgrouping within a fully articulated family tree.

The question as to whether all the modern-day Australian languages come from a single ancestor—with family-tree-like splitting and then tens of millennia of equilibrium and splitting—is an interesting one. It can be likened to the question of whether human language evolved just once (monogenesis) or independently in several different locations (polygenesis). We just don’t know. Some people say they feel that it must have been monogenesis while others report a hunch in favour of polygenesis. All that we really have—or ever can have—are feelings and hunches.

It is much the same for Australia. There could well have been a single original ancestor. But, if so, one would hesitate to say what it was like (in its grammar or in its forms) since so many recurrent features of modern languages are plainly the result of diffusion—that is, diffusion of some characteristic that may have originated in just one language at some indeterminate time during the last 40,000 or so years. Or there may have been several original languages—and original

language families—with the distinctions between them becoming blurred through aeons of diffusion.

Evidence can be put forward for each alternative. We will here just give one that might be taken to incline towards polygenesis.

Most Australian languages have distinct forms for 'who' and 'what' (only in a minority of languages is there a single form covering the two meanings), and also a form for 'where' (this is sometimes synchronically based on 'what', but is often a distinct form). Quite a few languages have nonce forms, but there are five forms which each recur in a fair number of languages. For each of them the approximate number of languages it occurs in is given:

(20) 'who'	'what'	'where'
<i>ŋa(:)n-</i> , c.70	<i>ŋa(:)n-</i> , c.35	<i>ŋan-</i> , c.4
<i>wanh-</i> , c.20	<i>wanh-</i> , c.4	<i>wanhdha-</i> , c.100
<i>wa:r(r)-</i> , c.20		<i>warr-</i> , c.2
<i>nha(n)-</i> , c.5	<i>nha:-</i> , c.25	
	<i>minha</i> , c.45	

Note that a recurrent locative inflection is a stop homorganic with the last consonant of the stem, plus *a*. Thus *wanh-dha* would be the expected locative of *wanh-*. Note though that *wanh-* 'who' is found in about twenty languages, *wanh-* 'what' in about four, but *wanhdha-* 'where' is in about one hundred languages.

The difficulty with the forms in (20) is that there are five of them. *Minha* 'what' probably comes from grammaticalization of a general noun 'edible animal'. But that still leaves four forms for two meanings (assuming that 'where' was originally based on 'what'). This is too many for one language but just right for two languages. This must not be taken to provide evidence that modern-day Australian languages do share between them precisely two ancestors, but simply to indicate that the question concerning the monogenesis or polygenesis of Australian languages should be left open (rather than being provided with a glib but unprovable answer).

If it were possible to peel off the layers of diffusion, would it be possible to tell whether the languages of Australia originally made up one genetic family (modelled by a family tree diagram)? We don't know. Is it sensible to try to establish a family-tree diagram for the c.250 modern languages of Australia? I'd say 'no'. (I've tried this, without success. I've experimented with many sorts of hypotheses during thirty-five years of trying to understand the relationships between Australian languages.)

One thing which is certain is that we have everything to learn. The possibilities for research on the Australian linguistic area are boundless. But to make progress one must approach the matter with an open mind, and with the realization that this is a completely different linguistic situation from those reported from anywhere else in the world.

Appendix—The ‘Pama-Nyungan’ idea

It is a received idea that Australian languages divide into two genetically based groups, ‘Pama-Nyungan’ and ‘non-Pama-Nyungan’. When a typologist quotes data from an Australian language they invariably include after it ‘(Pama-Nyungan)’ or ‘(non-Pama-Nyungan)’. In fact there is no principled scientific basis to the ‘Pama-Nyungan’ idea; it has the same order of validity as Greenberg’s (1987) ‘Amerind’. In this appendix I review the history of the ‘Pama-Nyungan’ idea, in its two incarnations.

Schmidt (1919) divided Australian languages into a ‘southern group’ and a ‘northern group’,⁵ and provided further classification within these groups. However, this was based on somewhat superficial features, including which sounds can appear at the end of a word. Then Capell (1956) put forward a classification in terms of morphological type—between prefixing and non-prefixing languages (he used the terms ‘prefixing’ and ‘suffixing’ but in fact all the languages employ suffixes). See the maps in Dixon (1980: 20).

A lexicostatistic classification of Australian languages (the work of Hale, O’Grady, and Wurm) was published in O’Grady, Voegelin, and Voegelin (1966), with a slightly revised version in Wurm (1972). The criterion for grouping was said to be a mechanical comparison of core vocabulary (a list that was of unspecified length and composition). Thus (O’Grady, Voegelin, and Voegelin 1966: 24–5, Wurm 1972: 110):

COGNATE DENSITY OF	INDICATES
less than 15%	different phylic families
16–25%	different groups of the same phylic family
26–50%	different subgroups of the same group
51–70%	different languages or family-like languages of the same subgroup
over 71%	different dialects of the same language

(No information was given—in either source—on what should be inferred if the cognate density were exactly 15% or exactly 71%.)

In this classification, the languages of Australian were said to comprise a ‘macro-phylum’ (a supposed genetic unit) which was divided into twenty-nine ‘phylic families’. One of these has become well-known in the literature: ‘Pama-Nyungan’ (named after the words for ‘person’ or ‘man’ in the extreme north-east and the extreme south-west) covers about three-quarters of the languages and more than three-quarters of the geographical area.

However, all that was published was the classification. The data on which it was based were not specified, nor were the cognate densities between languages. A different publication, O’Grady (1966: 121), did include a ‘cognate density matrix’ for a number of western languages and dialects. The percentages presented there do not fully accord with the lexicostatistic classification. Thus, the cognate density between ‘Wadjeri’ (my WG_{a1}, Watjarri) and ‘Nanda’ (my WG_b, Nhanta) is given as 42%, which should indicate ‘different subgroups of same group’. However, Wadjeri and Nanda are placed in the same subgroup (the ‘Kardu

⁵ In a series of superb pieces of reconstruction, Hale showed that the languages of North Queensland and of the Centre (our groups B and WL), which appear on the surface to have unusual word structures—and were placed by Schmidt in his ‘northern group’—have developed from languages of normal profile by a series of extensive phonological changes. (See Dixon 1980: 195–207, 487 for a summary and references.)

subgroup') in O'Grady, Voegelin, and Voegelin (1966: 37). (My calculation of shared vocabulary between them is 34%.) The percentage given by O'Grady for cognate density between 'Targari' and 'Warienga' is 45%; Austin (1988: 7) gives a score of 80%. O'Grady, Voegelin, and Voegelin place 'Targari' and 'Warienga' in different subgroups whereas in fact they constitute mutually intelligible dialects of a single language.

The examples quoted in the last paragraph are relatively minor; others are more serious. I have calculated percentages of shared vocabulary using the data available on a range of languages and a high proportion of the figures would—applying the lexicostatistic criteria—give strikingly different classifications from those in O'Grady, Voegelin, and Voegelin (1966) and Wurm (1972). For instance:

- (a) Between the 'Nyulnyulan phylic family' (my NE) and the 'Marngu subgroup of the South-west group of the Pama-Nyungan phylic family' (WI) there is a *c.*40% cognate density. On the lexicostatistic criterion these should be different subgroups of the same group; they were classified as different phylic families.
- (b) Between the 'Wororan phylic family' (NG) and the 'Bunaban phylic family' (NF) there is a cognate density of about 24%, indicating that they should be different groups of the one phylic family, rather than distinct phylic families.
- (c) Between the 'Bunaban phylic family' and the 'Djeragan phylic family' (ND) there is *c.*38% cognate density, which should indicate different subgroups of one group, rather than different phylic families.
- (d) Between the 'Wambaya phylic family' (NCb) and the 'Ngumbin subgroup of the South-west group of the Pama-Nyungan phylic family' (WJa) there is a *c.*30% cognate density; this should indicate different subgroups of the same group, rather than different phylic families.
- (e) Between the 'Wambaya phylic family' and the 'Karwan phylic family' (X) the cognate density is *c.*34% which should again indicate different subgroups of the same group, rather than distinct phylic families.
- (f) The 'Narrinyeric group of the Pama-Nyungan phylic family' (U) has a cognate density of no more than 15% with any neighbour and should, on the criteria stated, be considered a distinct phylic family.

This is only a sample of the instances where actual cognate densities do not support the 1966 classification (for which no percentage scores were in fact quoted, and the sources used were not indicated).

There were a number of untenable assumptions underlying this work: (i) that all relations between languages can be shown through a family-tree-type genetic model; (ii) that we can infer genetic relationships from lexicon alone, (iii) that the lexicon of all languages is always replaced at a constant rate; and (iv) that core vocabulary always behaves in a different way from non-core. This led to the invalid inference that a family-tree-type model can be discovered by comparing short lists of core vocabulary. Note also that even if lexicostatistics were a solid method for other parts of the world it would not be for Australia, where cognate scores are roughly the same for both core and non-core vocabulary.

The fact that erroneous lexical scores were obtained in many cases—as illustrated in (a)–(f) above—would have made the results unsound even if the method and the assumptions behind it had validity (which they did not have).

Cognate scores between contiguous languages are in fact useful as an indication of the degree of contact between the languages, and of how much borrowing there has been.

Figures such as 24% for NG/NE, 38% for NE/ND, and 34% for NCb/X are useful as indicators of degree of borrowing and relative time-depth of geographical contact. Note that verb scores and similarities of grammatical forms between all of these groups are very low. Each of NE, ND, NC, and X is a low-level subgroup, and no higher-level genetic links can be established between them. (The three languages in group NG comprise a small linguistic area.)

The lexicostatistic classification has been accepted by the majority of people working on Australian languages, and by many people outside Australia. In particular, great emphasis is attached to the 'Pama-Nyungan'/'non-Pama-Nyungan' distinction (where 'non-Pama-Nyungan' is used as a cover label for the other twenty-eight phyletic families in the 1966 classification). One leading Australianist was heard to say that he had little interest in attending a workshop on 'non-Pama-Nyungan languages' since he was 'just a Pama-Nyunganist'.

There is a rough correlation between the 'non-Pama-Nyungan groups' and prefixing—twenty-five of the 'non-Pama-Nyungan groups' (all save Wambayan, Karwan, and Minkinian) use prefixes. If 'Pama-Nyungan' were a valid genetic group (as suggested by the 1966 lexicostatistic work) one might as a consequence posit a 'proto-Pama-Nyungan' ancestor language. But some Australianists have gone further. Heath (1978: 10)—in a study of diffusion between Australian languages—works in terms of 'proto-prefixing', while Heath (1997: 200) has 'proto-non-Pama-Nyungan' (although this is 'Pama-Nyungan Mark II'—see below). The development of prefixing is in fact an areal phenomenon. Languages in the prefixing region have pronominal prefixes referring to core arguments of the clause but there is considerable variation in the actual forms of the prefixes and also in their ordering. In some languages the A (transitive subject) prefix precedes the O (transitive object) prefix, in some O precedes A, and in some a non-third person argument precedes a third-person argument (irrespective of their syntactic functions). In some A and S are marked by pronominal prefixes but O by enclitics to the verb. In view of this variety it would be impracticable to essay any suggestion as to what the prefixal forms (and their ordering) might be in Heath's 'proto-prefixing'. It is instead clear that the structural type 'prefixing' has diffused over a continuous area, with each language developing pronominal and other prefixes in an individual way, from its own internal resources.

Although no proper justification had been provided for 'Pama-Nyungan', it came to be accepted. People accepted it because it was accepted—as a species of belief. Associated with the belief came a body of lore. One part of this is that there is a sharp linguistic division along the 'Pama-Nyungan'/'non-Pama-Nyungan' geographical boundary. That this is untrue can be seen from a selection of cognate percentage figures (some were given earlier). From west to east across the 'Pama-Nyungan'/'non-Pama-Nyungan' boundary the lexical scores include (groups whose code letters begin with N are 'non-Pama-Nyungan'): WI/NE, c.40%; WJa/NE, c.22%; WJa/ND, c.29%; WJa/NCa, c.30% and X/NCb, c.34%. In the west there is a gradual shading in verb structure: NG has pronominal prefixes to the verb for both subject and object; NE has a pronominal prefix for subject but an enclitic for object; and WI has pronominal enclitics to the verb for both subject and object.

The appropriate question to ask was: 'what is the justification for "Pama-Nyungan"?' But many Australianists accepted—as an article of faith—that 'Pama-Nyungan' was a valid and useful idea. They simply asked: 'What is the nature of "Pama-Nyungan"?' The answer to this question involved reassessment of what languages should be taken to belong to 'Pama-Nyungan'. Thus, 'Pama-Nyungan Mark II' came into being; it differed from 'Pama-Nyungan Mark I' in the subtraction of NA and the addition of WMa. (It seems that the status of group X has not yet been decided on.)

As already mentioned, there are many linguistic parameters in terms of which Australian languages can be classified. One involves whether or not non-singular pronouns have number-segmentable forms; i.e. whether there is a single stem for each of 1n-sg and 2n-sg, with dual and plural (and sometimes also trial or paucal) number suffixes being added to them. This type of structure applies to most of the prefixing languages (Wma is a notable exception) and to the non-prefixing group NA. 'Pama-Nyungan Mark II' was effectively defined as those languages with number-segmentable non-singular pronouns. Lexicostatistic figures, which had been the justification for 'Pama-Nyungan Mark I', were no longer mentioned.

'Pama-Nyungan Mark II' covers my groups A–Y, WA–WM while 'non-Pama-Nyungan Mark II' covers NA–NL. The convention of using a first letter 'N' for all the groups assigned to 'non-Pama-Nyungan Mark II' was adopted purposefully, as a way of demonstrating that no other parameter coincides with that of having number-segmentable non-singular pronouns. It almost coincides with the prefixing/non-prefixing distinction shown in Map 2. It does not correlate with type of verbal organization, shown in Map 1; nor with the distinction between pronominal systems organized on a singular/dual/plural and those organized on a minimal/unit-augmented/augmented basis. It does not correlate with the distinction between languages with ergative case-marking, those with accusative case-marking, and those with no case-marking at all for core functions. It does not correlate with the distinction between languages with noun classes and those without. It does not correlate with any phonological distinction. Dixon (2002) discusses many further parameters of variation and for almost every one of them there are some languages from groups NA–NL on each side of the isogloss.

Another piece of 'Pama-Nyungan' lore is that there is a stock of lexemes found all over the 'Pama-Nyungan' area but not in 'non-Pama-Nyungan' languages. This is without foundation. To illustrate this, we can divide Australian languages (omitting the Papuan languages, in group A) into four sets of approximately equal size:

groups B–J, 64 languages	groups WA–WM, 59 languages
groups K–Y, 61 languages	groups NA–NL, 61 languages

I have investigated 112 lexemes each of which occurs in at least two of these sets (full details are in Dixon, 2002: 96–129). The number in each set is:⁶

groups B–J, 90 lexemes	groups WA–WM, 98 lexemes
groups K–Y, 89 lexemes	groups NA–NL, 83 lexemes

(A sample of five items is given at 30–4 in the Table.)

It will be seen that there are a few less instances of recurrent lexemes in the set consisting of groups NA–NL than in other sets, but not significantly less.⁷

The revamping of 'Pama-Nyungan' into Mark II is due in large part to Blake (1988) and Evans (1988). They support the idea of all Australian languages constituting one language

⁶ The figures are: lexemes in all four sets, 52; in all except NA–NL, 13; in all except WA–WM, 6; in all except K–Y, 5; in all except B–J, 8; just in B–J and K–Y, 3; just in B–J and WA–WM, 8; just in B–J and NA–NL, 3; just in K–Y and WA–WM, 5; just in K–Y and NA–NL, 2; just in WA–WM and NA–NL, 7.

⁷ Many of the prefixing languages have noun-class prefixes (which have to be peeled off, to get to the root) and many have undergone considerable phonological changes. This has sometimes made it difficult to recognize, in the prefixing languages, cognates for lexemes that recur in languages over the remainder of the continent.

family, and of 'Pama-Nyungan' being a high-level genetic subgroup within this family. They suggest a number of innovations that are purported to have taken place between 'Proto-Australian' and 'Proto-Pama-Nyungan'.

Evans (1988) presents a number of cognate sets where an initial apical stop or nasal in 'non-Pama-Nyungan' languages corresponds to a laminal stop or nasal in 'Pama-Nyungan' languages, e.g. (see items 34 and 38 in the table):

'see' is *na(-ŋ)* in four of the groups NA–NL (and in W and X) and *nha(-ŋ)* in 29 of the 36 groups A–V, Y, WA–WM

2n-sg pronoun is *nu-* in c.70% of the languages in NA–NL (and also in X) and *nhu-* in c.60% of the languages in A–W, Y, WA–WM

He suggests that Proto-Australian had an apical in these words (which is continued in the 'non-Pama-Nyungan' groups) but that in Proto-Pama-Nyungan this apical became a laminal. However, all but one of Evans' cognate sets show exceptions—an initial apical in some 'Pama-Nyungan' languages or an initial laminal in some 'non-Pama-Nyungan' languages; see Dixon (forthcoming).

We can add a correspondence in the opposite direction, involving the final segment of a stem where a laminal nasal in 'non-Pama-Nyungan' languages corresponds to an apical nasal in 'Pama-Nyungan' languages (see item 37 in Table 1):

2sg pronoun is *ŋinj-* in about half the languages of NA–NL, and is based on **ŋin-* in c.95% of the languages in A–Y, WA–WM. (Note that 2sg is *ninj-* in group X.)

The initial apical/laminal isogloss almost coincides with the criterion adopted for 'non-Pama-Nyungan'/'Pama-Nyungan' (Mark II), i.e. whether or not languages have number-segmentable non-singular pronouns.

The apical/laminal distinction is a possible piece of evidence in favour of 'Pama-Nyungan' as a genetic group. However, it would need to be supported by a number of other 'innovations', and none is forthcoming. An alternative explanation would be that the apical/laminal distinction is the result of areal diffusion, like so many other parameters of variation in Australia.

Evans and Blake do suggest two other bits of evidence for 'Pama-Nyungan Mark II' as a genetic group: the ergative allomorph *-ŋgu*, and the 1du.inc pronominal form *ŋali*. Neither of these stands up under careful scrutiny.

Following Hale (1976), I suggested in Dixon (1980) that the original form of the ergative case would have been *-du* after a consonant and *-ŋgu* or *-lu* after a vowel. None of these allomorphs is found in 'non-Pama-Nyungan' languages, leading Evans (1988: 93) to suggest that ergative allomorphs *-ŋgu* and *-lu* are 'Pama-Nyungan' innovations.

In an outstanding contribution to Australian linguistics, Sands (1996) showed that the most appropriate reconstruction for ergative case consists in two forms: **-lu*, on demonstratives, interrogatives, proper nouns, kin terms, generic nouns, and pronouns; and **-dhu* on other nouns (see item 35 in the Table). Reflexes of **-dhu* are found right across the continent, in prefixing and in non-prefixing languages, in 'Pama-Nyungan' and in 'non-Pama-Nyungan' languages.

The idea of ergative allomorph *-ŋgu* as an innovation in 'Proto-Pama-Nyungan' is not sustainable. This form occurs in only about one third of the 'Pama-Nyungan' languages (shown in Map 5). It is plainly an areal feature, being found in around twenty-five



MAP 5. Distribution of the ergative allomorph -*ngu*

languages in a western block and in around thirty languages in an eastern block⁸ plus one language separated from either block, Mgi, Gumbaynggirr (see item 36 in the Table). It seems clear that *-ngu* developed out of **-dhu* in at least three distinct places (and possibly by a different path of change in each place—see Dixon (2002: 157–66) and that in two instances the *-ngu* then diffused over a continuous geographical area.

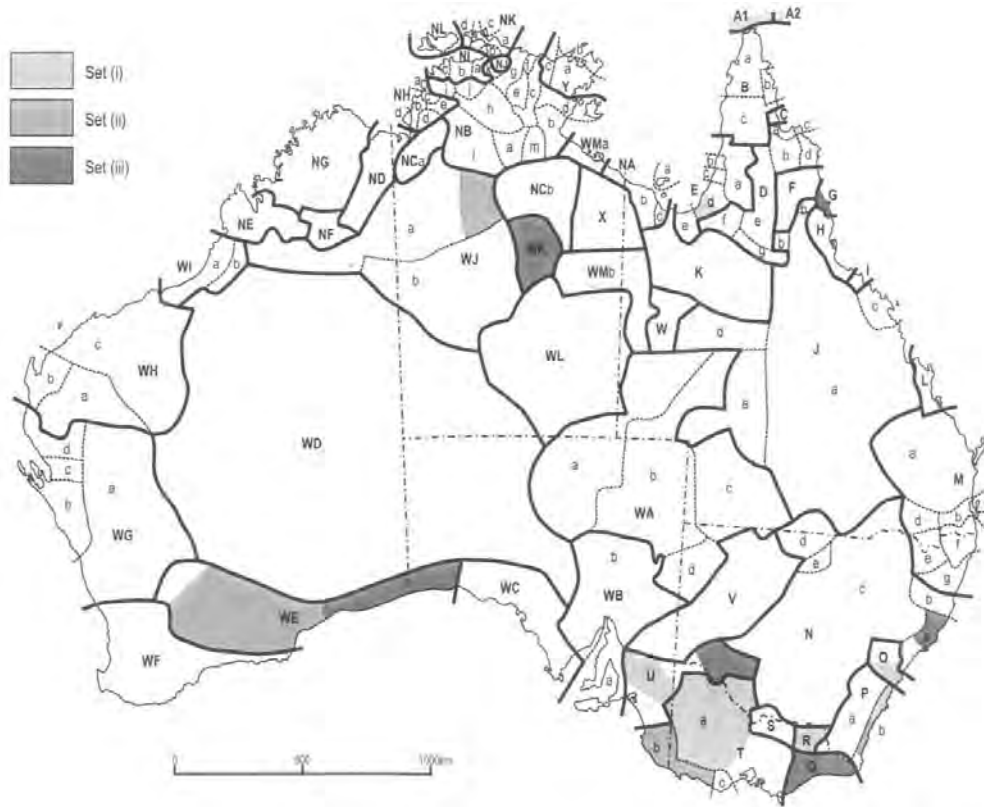
Blake (1988) presents two series of pronouns, one ‘Pama-Nyungan’ (based in part on Dixon 1980) and the other ‘Northern’. (Note that Blake does not state that these relate to ‘Proto-Pama-Nyungan’ and ‘Proto-non-Pama-Nyungan’ respectively, although there is an implication in this direction.) Blake’s ‘Northern’ pronouns are discussed in detail in Dixon (2002: 253–62), where some but not all forms are shown to be supportable. Of the ‘Pama-Nyungan’ forms he gives none occur in more than about half the ‘Pama-Nyungan’ languages and most have an areal distribution. For instance, Blake’s 2du **nyuNpalV* is not found in any languages of groups L–V or WA in the south-east, nor in WE in the south-west (nor in Y). His 3pl **tyana* is missing from almost all languages in groups M–V, WC, WE, WF and WJ–WK.

A persistent nugget of ‘Pama-Nyungan’ myth is that the 1du or 1du.inc pronominal form *ɲali* is found in all and only the ‘Pama-Nyungan’ languages (and is thus an unimpeachable candidate to be an innovation in ‘proto-Pama-Nyungan’). It is true that *ɲali* is found in none of the languages with number-segmentable non-singular pronouns (the ‘non-Pama-Nyungan groups’). But it occurs in only about 80% of the Pama-Nyungan languages. (The distribution is shown in Map 6.)

Languages from groups A–Y, WA–WM which lack *ɲali* fall into three sets.

- (i) Some languages lack *ɲali* but have one or more 1n-sg pronominal forms beginning with *ɲal-*; some or all of them could be based on an earlier form *ɲali* (or, equally well, there could be some alternative origin for the *ɲal-* portions of these forms). In Dc1, the Flinders Island language, for instance, we find 1du.inc *ɲaluntu*; 1du.exc *ɲalulu*; 1pl.inc *ɲalapal* and 1pl.exc *ɲalada*. Languages in this set belong to groups A, Dc, Ed, O, Pb, R, Ta, U.
- (ii) In a few languages all non-singular pronouns involve number increments to singular forms; this applies to some languages from groups Pb, Tb, WE, and WJb. It is possible that some or all of these languages had 1du.inc *ɲali* at an earlier stage, and that this was replaced when the pronoun paradigm was restructured. It is equally possible that in some or all of these languages there never was any pronominal form *ɲali*.
- (iii) There are six languages with full pronominal paradigms (not involving number-segmentable forms) and no trace of *ɲali* or *ɲal-*. Interestingly, five of the six languages of set (iii) are on the fringes of the ‘Pama-Nyungan’ area. Four of them are on the coast (from groups G, Na, Q, and WE) while WK is on the inland boundary of the area. The sixth language is U5, Yitha-Yitha, spoken some way up the Murray River (within the small linguistic area, U). Note that of the languages in set (ii), those in groups Pb, Tb, and WE are also spoken on the coast.

⁸ Ergative *-ngu* is also in WMa, Yanyuwa, which is currently separated from the other languages. This language is genetically related to Wmb languages and must have been in contact with them in the past. Proto-WM probably had ergative *-ngu*, which is retained in Yanyuwa but has become *-gu* or *-əg* in the Wmb languages.



MAP 6. Languages in groups A–Y, WA–WM lacking 1 du(inc) ṇali (this is lacking from all languages in groups NA–NL)

We can compare two competing hypotheses:

- (i) THE GENETIC ('PAMA-NYUNGAN') HYPOTHESIS. A proto-language ancestral to A–Y, WA–WM had 1du.inc pronoun *ɲali*. This has been retained in most of the modern languages. One would then have to explain why there is no trace of *ɲali* in languages of set (iii), or in those of set (ii). It has presumably been lost. It must have been lost from at least nine distinct areas, all but one of them on the fringe of the region that has *ɲali*.
- (ii) THE DIFFUSIONAL HYPOTHESIS. A 1du.inc form *ɲali* has simply diffused over a large continuous area. It also occurs in Y, the Yolngu⁹ subgroup, from north-east Arnhem Land; this implies that the Yolngu languages were part of the *ɲali* diffusion area at some time in the past but have recently become separated from it.

The pronoun *ɲali* covers almost all the region occupied by groups A–Y, WA–WM. However, it has not yet reached about nine areas, all but one of them on the fringe of the region (with all but one of these being on the coast). The only non-fringe language to lack *ɲali* is in group U which, as mentioned in §4, lies in a small linguistic area which shows a number of archaic features; the languages of group U have probably been in their present location for a considerable period, and appear to have been relatively resistant to diffusional influences from other languages.

Hypothesis (i) involves about nine separate losses of *ɲali*, almost all on the coast. Hypothesis (ii) involves a steady diffusion of this form over a continuous area. (Dixon (2002: 277–82) describes two instances of the continuing diffusion of *ɲali*, into coastal areas into which it had not previously penetrated.) The second alternative is simpler and plainly to be preferred. Thus, the 'ɲali argument' for 'Pama-Nyungan' as a genetic group is not strong.

We have shown that 'Pama-Nyungan' cannot be supported as a genetic group. Nor is it a useful typological grouping in that it relates to just one typological parameter (that of number-segmentable non-singular pronouns). This almost, but not quite, correlates with the parameter of prefixing. It has little or no correlation with other typological parameters.

The putative division between 'Pama-Nyungan' and 'non-Pama-Nyungan' (either Mark

⁹ Note that 'having the 1du(inc) pronoun *ɲali*' is one of the few features said to characterize 'Pama-Nyungan languages' that the Yolngu subgroup does possess. It does not have ergative allomorph *-ŋgu*, for instance. Of ten 'Pama-Nyungan pronouns' in Blake (1988: 6) it shows only three: 1du.inc *ɲali*, 1pl.exc *ɲanaa* and 3pl *than-*.

From the list of recurring lexemes that I have compiled, forty-three occur in Yolngu. Thirty-three of these are in both 'Pama-Nyungan' and 'non-Pama-Nyungan' groups, seven just in 'Pama-Nyungan' and three just in 'non-Pama-Nyungan'. Taking into account that there are about three times as many 'Pama-Nyungan' as 'non-Pama-Nyungan' languages, it will be seen that there is no significant lexical association between Yolngu and the other 'Pama-Nyungan groups', A–X, WA–WM.

The Yolngu languages are likely to have been in contact with some languages from groups A–Y, WA–WM in the past, in order for *ɲali* and a few more forms to have diffused into them. That is, a genetic connection between Yolngu and other non-prefixing groups cannot be sustained, but a previous areal connection is most likely.

Note that Yc, the Inland Yolngu subgroup, has recently developed bound pronouns which generally come immediately before the verb, either encliticized to the word before the verb or as free forms. It is likely that the next step will be for these to become pronominal prefixes to the verb, as the areal feature of prefixing continues to expand.

I or Mark II) has had a deleterious effect on the study of Australian languages. Too often, students are assigned to study a certain topic 'within Pama-Nyungan' or 'within non-Pama-Nyungan' when the feature under study is found in languages from all over the continent. (Dixon (2002) provides instance after instance of this.)

I mentioned at the beginning of this appendix that the 'Pama-Nyungan' idea is similar in some respects to Greenberg's 'Amerind' idea.¹⁰ But the historical order of the hypothesizing is different. Scholars of American Indian linguistics by and large agree on what are provable genetic groupings—see, among others, Campbell and Mithun (1979), Goddard (1996), Campbell (1997), and Mithun (1999). Greenberg (1987) proposed that all American Indian languages except for the Eskimo-Aleut family, the Athapaskan-Eyak-Tlingit family, and the isolate Haida make up one genetic 'stock', which he called 'Amerind'; this comprises several score language families and several dozen isolates from North, Central, and South America.

In Australia, things happened the other way around. First of all, we had the Greenbergian-type idea of an 'Australian macro-phylum' which consisted of twenty-nine 'phylic families', one of them 'Pama-Nyungan'. After this some people worked in terms of all of the 'non-Pama-Nyungan' making up one genetic grouping (but with no justification provided for this), implying that 'Proto-Australian' had a binary split into 'Proto-Pama-Nyungan' and 'Proto-non-Pama-Nyungan'. In this view of things, every Australian language has its place within a fully-articulated family tree, just as every American Indian language does in Greenberg's scheme. There is no sustained attention to distinguishing between those similarities which are due to genetic retention, those due to areal diffusion, and those due to parallel development (see §1 above). Only now are a few scholars attempting to assess the relationships between languages, and to distinguish those groups which can probably be shown to constitute genetic subgroups from those which comprise small linguistic areas.

Scholars outside Australia who quote a bit of data from, say, Watjarri (my WGa1) tend to look up the lexicostatistic classification and state that it belongs to 'the Kardu subgroup of the Southwest group of the Pama-Nyungan phylic family of the Australian macro-phylum' (O'Grady, Voegelin, and Voegelin 1966: 37). This is on a par with saying that Tarascan (recognized by almost all scholars as an isolate—see Campbell 1997: 166) belongs to the 'Chibchan' grouping within the 'Chibchan-Paezan group', within the 'Central Amerind stock' of the 'Amerind family' (Greenberg 1987: 382).

It is satisfying to have a set of pigeon-holes into which to place things, and it can be frustrating when one is told that a neat and tidy, all-encompassing scheme of classification has no validity. But, if progress is to be made in understanding the types of relationships between Australian languages, we must start at the bottom, provide proof for those low-level genetic subgroups which can be recognized, and study the multifaceted patterns of diffusion that have flowed forwards and backwards across the continent for the past several tens of millennia.

¹⁰ The methodologies involved are, of course, different. Greenberg employed what he calls the 'mass comparison' technique whereas 'Pama-Nyungan' was suggested on the basis of lexicostatistic counts (although without the sources used or percentages obtained being made available) and then redefined on the basis of occurrence of number-segmentable non-singular pronouns. My point is each system of classification lacks a scientific basis and tends to hold back scientific work on language relationships.

Summary list of languages (including likely low-level genetic subgroups)

This list—which should be regarded as a draft—contains the names of all of the known indigenous languages of Australia (excluding Tasmania); it is likely that there were further languages, which have been lost without trace. There are many alternative names for languages, and for dialects within languages; in this summary list only a few of these are included.

A * after a letter indicates the likelihood that all the languages in this group can be shown to make up a low-level subgroup, e.g. B* shows that B is probably a subgroup and Ba* that Ba is probably a subgroup within B. If two languages within a group are probably genetically related then * is included after each of their numbers, e.g. 1* and 2* within De (there is insufficient information on De3 to be able to decide whether this belongs in the subgroup with De1 and De2).

- A 1, West Torres (Papuan, with Australian substratum); 2, East Torres (Papuan)
- B* Ba*: 1, Gudang; 2, Uradhi; 3, Wuthati; 4, Luthigh/Mpalitjanh; 5, Yinwum; 6, Anguthimri/Mpakwithi/Awnghim/Ntra'angith/Alngith/Linngithigh; 7, Ngkoth; 8, Aritinngithigh; 9, Mbiywom; 10, Andyingit
Bb: Umpila/Kuuku-Ya'u/Kaantju
Bc*: 1, Wik-Ngathrr; 2, Wik-Me'nh/Wik-Ep; 3, Wik-Mungsknh (Wik-Munkan); 4, Kugu-Muminh (or Wik Muminh or Kugu/Wik Nganhcara); 5, Bakanha; 6, Ayabadhu
- C Umbindhamu
- D Da*: 1, Morroba-Lama (or Umbuygamu); 2, Lama-Lama (or Mba Rumbathama)
Db: 1, Rimang-Gudinhma; 2, Kuku-Wara
Dc: 1, Flinders Island language (or Yalgawarra); 2, Marrett River language (or Tartalli)
Dd: 1, Guugu Yimidhirr; 2, Barrow Point language
De: 1*, Kuku-Thaypan; 2*, Kuku-Mini/Ikarranggal/Aghu Tharrnggala; 3, Takalak
Df: Walangama
Dg: Mbara
- E Ea: 1, Kuuk Thaayorre; 2, Oy kangand/Olgolo; 3, Ogh-Undjan
Eb: 1, Yirr Yoront/Yir Thangedl; 2*, Koko Bera (or Kok Kaber); 3*, Kok Thawa
Ec: Kok Narr (or Kok Nhang or Kundara)
Ed*: 1, Kurtjar (or Gunggara); 2, Kuthant
Ee: Kukatj (or Galibamu)
- F Kuku-Yalanji/Kuku-Njungkul/Wakura/Wakaman/Jangun/Muluridji
- G* 1, Djabugay; 2, Yidinj/Gunggay/Wanjurru
- H 1, Dyirbal/Girramay/Djiru/Gulngay/Mamu/Ngadjan; 2, Warrgamay/Biyay; 3, Nyawaygi; 4, Manbara/Wulgurukaba/Nhawalgaba
- I 1, Cunningham; 2, Gorton; 3, O'Connor (all only from 'Lower Burdekin' and 'Mouths of Burdekin' vocabularies in Curr)

- J** **Ja***: 1, Bidjara/Marrganj/Gayiri/Dharawala/Mandandanjdji/Guwamu/Gunggari/Nguri; 2, Biri/Gangulu/Wirri/Yilba/Baradha/Yambina/Yetimarala/Garanjbal/Yangga; 3, Warungu/Gugu-Badhun/Gudjala; 4, Ngayungu; 5, Yirandhali
Jb: 1, Mbabaram; 2, Agwamin (or Wamin)
Jc: 1, Ngaro; 2, Giya
Jd: 1, Guwa; 2, Yanda
Je: 1, Kunggari; 2, Pirriya (or Bidia)
- K*** 1, Ngawun/Wunamara/Mayi-Thakurti/Mayi-Yapi/Mayi-Kulan; 2, Mayi-Kutuna
- L** 1, Darambal; 2, Bayali
- M** **Ma**: 1, Dappil; 2, Gureng-Gureng; 3, Gabi-Gabi/Badjala; 4, Waga-Waga/Duungidjauw
Mb: Yagara
Mc: Guwar
Md: Bigambal
Me: Yugambal/Ngarraabul (Ngarrbal)
Mf: Bandjalang/Yugumbir/Minjangbal/Gidabal/Wudjeebal
Mg*: 1, Gumbaynggirr/Baanbay/Gambalamam; 2, Yaygirr
- N** **Na***: 1, Awabagal/Cameeragal/Wonarua; 2, Gadjang/Warimi/Birbay
Nb*: 1, Djan-gadi; 2, Nganjaywana (Aniwan)
Nc*: 1, Gamilaraay (Kamilaroi)/Yuwaalaraay/Yuwaaliyaay (Euahlayi); 2, Wiradhurri; 3, Ngiyambaa/Wangaaybuwan/Wayilwan
Nd: Muruwarri
Ne: Barranbinja
- O*** 1, Dharuk/Gamaraygal; 2, Darkinjung
- P** **Pa**: 1, Gundungurra/Ngunawal; 2, Ngarigo
Pb: 1, Dharawal; 2, Dhurga/Dharamba; 3, Djirringanj; 4, Thawa
- Q** Muk-thang (Gaanay, Kurnai, Kunnai)/Bidhawal
- R** 1, Pallanmanganmiddang; 2, Dhudhuroa/Yaithmathang
- S** 1, Yota-Yota (Bangerang); 2, Yabala-Yabala
- T** **Ta***: 1, Wemba-Wemba/Baraba-Baraba/Madhi-Madhi/Ladji-Ladji/Wergaya/Djadjala/Jab-wurrong/Pirt-Koopen-Noot/Jaja-wurrong; 2, Wadha-wurrong; 3, Wuy-wurrong/Bun-wurrong/Dhagung-wurrong
Tb*: 1, Bungandik (or Bundanditj); 2, Kuurn-Kopan-Noot/Peek-Whurrong/Dhautgart/Tjarcote (misnamed Gournditch-Mara)
Tc: Kolakngat (or Kolijon)
- U** 1, Yaralde (or Ngarrindjeri or Narrinyeri); 2, Ngayawang; 3, Yuyu (or Ngarrket); 4, Keramin; 5, Yitha-Yitha/Dardi-Dardi
- V** Baagandji/Gurnu/Baarrundji/Barrindji/Marrawarra (Maruara)
- W** 1, Kalkatungu; 2, Yalarnnga
- X*** 1, Waanji; 2, Garrwa (Garawa)
- Y*** **Ya***: 1, Dhuwal/Dhuwala (including Gupapuyngu, Gumatj, Djambarrpuynngu); 2, Dhay'yi; 3, Ritharngu (or Dhiyakuy)
Yb*: 1, Nhangu; 2, Dhangu; 3, Djangu
Yc*: 1, Djinang; 2, Djinba

- WA** **WAa:** 1*, Pitta-Pitta (Pitha-Pitha); 2*, Wangka-Yutjuru; 3, Arabana-Wangkangurru
WAb: 1, Yandruwanhdha/Yawarawarga; 2*, Diyari/Dhirari/Biladaba; 3*, Ngamini/Yarluyandi/Karangura; 4, Midhaga/Karuwarli/Marulta
WAc: 1, Wangkumara/Punthamara; 2, Galali; 3, Badjiri
WAd: Maljangapa/Yardliyawara/Wardikali
- WB** **WBa:** Kadli (Kurna, Nantuware, Ngadjuri, Narangka, Nukunu)
WBb*: 1, Parnkala; 2, Adnjamathanha/Guyani/Wailpi
- WC** Wirangu/Nhawu
- WD** The Western Desert language (dialects: Warnman, Yulparitja, Manjtjiltjara, Kartutjarra, Kukatja, Pintupi, Luritja, Ngaatjatjarra, Nganajatjarra, Wangkatha, Wangatja, Ngaliya, Pitjantjatjarra, Yankunjtjatjarra, Kukarta)
- WE** 1, Mirning; 2, Kalaaku (Ngadjunmaya); 3, Karlamay
- WF** Nyungar (including Pipalman, Pindjarup, Whadjuk)
- WG** **WGa*:** 1, Watjarri; 2, Parti-maya; 3, Cheangwa language; 4, Nana-karti; 5, Natingero; 6, Witjaari
WGb: Nhanta/Watchandi/Amangu
WGc: Malkana
WGd: Yingkarta
- WH** **WHa:** Tharrkari/Warriyangka/Tjiwarli/Thiin
WHb*: 1, Payungu/Purduna; 2, Thalantji/Pinikura
WHc: 1, Nhuwala; 2, Martuthunira; 3, Panjtjima; 4, Yinjtjipartnti/Kurama; 5, Ngarluma; 6, Kariyara (Kariera); 7, Tjururu; 8, Palyku/Niyiyapali; 9, Nyamal; 10, Ngarla
- WI** **WIa*:** 1, Njangumarta; 2, Karatjarri
Wib: Mangala
- WJ*** **WJa*:** 1, Walmatjari/Tjuwalinj/Pililuna; 2, Djaru/Wawari/Njininj; 3, Gurindji/Wanjdjirra/Malngin/Wurlayi/Ngarinman/Pilinara; 4, Mudbura/Karranga/Pinkangarna
WJb*: 1, Warlpiri/Ngaliya/Walmala/Ngardilpa; 2, Ngardi; 3, Warlmanpa
- WK** Warumungu
- WL** 1, Arrernte (Aranda) (including: Anmatjirra, Aljarwarra, Ayerrereenge, Antekerrepenhe, Ikngerripenhe, Pertami, Alenjerntarrpe); 2, Kaytetj
- WM*** **WMa:** Yanyuwa (or Yanyula)
WMb*: 1, Wagaya/Yindjilandji; 2, Bularnu/Dhidhanu; 3, Warluwara/Kapula/Parnkarra
- NA*** **NAa:** Lardil
NAb*: 1, Kayardild/Yangkaal; 2, Yukulta (Kangkallita)/Nguburindi
NAc: Minkin
- NB** **NBa:** Mangarrayi (Ngarrabadji)
NBb*: 1, Marra; 2, Warndarrang (Wuyarrawala)
NBc*: 1, Rembarrnga/Kaltuy; 2, Ngalkan
NBd: 1, Ngandi; 2, Nunggubuyu (Wubuy, Yingkwira); 3, Aninhdhilyagwa
NBe: Dalabon (Dangbon)/Buwun/Ngalkbun (Ngalabun)

- NBf***: 1, Burarra/Gidjjangaliya/Anbarra/Gun-nartpa; 2, Gurrgoni; 3, Nakkara; 4, Ndjebbana (Kunibidji/Gunavidji)
NBg: 1, Gunwinjgu (Mayali, Bininj-gun-wok, Neinggu); 2, Gunbarlang
NBh: 1, Jawoyn; 2, Warray
NBi: Gungarakanj
NBj: Uwinjmil (Awunjmil, Winjmil)
NBk: Gaagudju
NBl*: 1, Wagiman; 2, Wardaman/Dagoman/Yangman
NBm: Alawa
NC* **NCa***: 1, Djamindjung/Ngaliwuru; 2, Nungali
NCb*: 1, Djingulu (Djingili); 2, Ngarga (Ngardji); 3, Wambaya/Gudandji/Binbinka
ND* 1, Kitja (Lunga); 2, Miriwung/Gajirrawung
NE* 1, Njigina/Warrwa/Yawuru/Jukun; 2, Baardi/Jawi/Njul-Njul/Jabirr-Jabirr/Ngumbarl/Nimanburru
NF* 1, Bunuba; 2, Guniyandi
NG 1, Worrorra/Unggumi; 2, Ungarrinyin; 3, Wunambal/Gamberre/Kwini (Gunin)
NH **NHa**: Patitjamalh/Kandjerramal (Pungu-Pungu)
NHb*: 1, Emmi/Merranunggu(Warrgat); 2, Marrithiyel/Marri-Ammu/Marritjevin/Marridan/Marramanindjdji; 3, Mari Ngarr/Magati-ge
NHc: Malak-Malak
NHd: 1, Murrinh-patha; 2, Ngan.gi-tjemerri (Ngan.gi-kurunggurr, Ngan.gi-wumeri)
NHe*: 1, Matngele; 2, Kamu
NI **NIa**: Umbugarla/Bugurndidja/Ngumbur
NIb: 1, Limilngan; 2, Wuna
NIc: Larrakiya (Gulumirrgin)
NJ Giimbiyu (including: Urningank, Mengerrdji, Erre)
NK* **NKa***: 1, Mawung (Maung, Gun-marung)/Mananggari; 2, Iwaydja/Ilgar/Garik
NKb: Amurdag (Wardadjbak, A'mooridiyu)/Urrirk/Didjurma
NKc: Marrgu (Terrutong, Yaako, Raffles Bay language, Croker Island language)
NKd: Popham Bay language (Iyi, Limpapiu)
NL Tiwi

References

- ALPHER, B., and NASH, D. 1999. 'Lexical replacement and cognate equilibrium in Australia', *AJL* 19.5–56.
 AUSTIN, P. 1988. 'Classification of Southern Pilbara languages', pp. 1–17 of *Papers in Australian Linguistics*, No. 17. Canberra: Pacific Linguistics.
 BIRDELL, J. B. 1957. 'Some population problems concerning pleistocene man', *Cold Spring Harbor Symposia on Quantitative Biology* 22.47–69.
 BLAKE, B. J. 1971. 'Jalanga: an outline morphology', pp. 1–11 of *Papers on the Languages of Australian Aboriginals*. Canberra: Australian Institute of Aboriginal Studies.
 ——— 1979. *A Kalkatungu Grammar*. Canberra: Pacific Linguistics.

- 1988. 'Redefining Pama-Nyungan: towards the prehistory of Australian languages', pp. 1–90 of *Aboriginal Linguistics* 1, edited by N. Evans and S. Johnson. Armidale: University of New England.
- 1989. 'Notes towards a Grammar of Yalarnnga'. Typescript.
- BOLT, J. E., HODDINOTT, W. G., and KOFOD, F. M. 1971. 'An Elementary Grammar of the Ngaliwuru Language of the Northern Territory'. Duplicated by Australian Institute of Aboriginal Studies, Canberra.
- CAMPBELL, L. 1997. *American Indian Languages: The Historical Linguistics of Native America*. New York: Oxford University Press.
- and MITHUN, M. 1979. Editors of *The Languages of Native America: Historical and Comparative Assessment*. Austin: University of Texas Press.
- CAPELL, A. 1956. *A New Approach to Australian Linguistics*. Sydney: Oceania Linguistic Monographs.
- DIXON, R. M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- 1980. *The Languages of Australia*. Cambridge: Cambridge University Press.
- 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- 2002. *Australian Languages: Their Nature and Development*. Cambridge: Cambridge University Press.
- DONALDSON, T. 1980. *Ngiyambaa: The Language of the Wangaaybuwan*. Cambridge: Cambridge University Press.
- EVANS, N. 1988. 'Arguments for Pama-Nyungan as a genetic subgroup, with particular reference to initial laminalization', pp. 91–110 of *Aboriginal Linguistics* 1, edited by N. Evans and S. Johnson. Armidale: University of New England.
- GODDARD, I. (ed.). 1996. *Handbook of North American Indian*, vol. 17: *Languages*. Washington: Smithsonian Institution.
- GREEN, I. 1995. 'The death of "prefixing": contact induced change in northern Australia', *Berkeley Linguistics Society Papers* 21.414–25.
- GREENBERG, J. H. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press.
- 1987. *Language in the Americas*. Stanford: Stanford University Press.
- HALE, K. L. 1976. 'On ergative and locative suffixal alternations in Australian languages', pp. 414–17 of *Grammatical Categories in Australian Languages*, edited by R. M. W. Dixon. Canberra: Australian Institute of Aboriginal Studies.
- 'Field Notes on Wik-Mungkanh'. Manuscript.
- HEATH, J. 1978. *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- 1997. 'Lost wax: abrupt replacement of key morphemes in Australian agreement complexes', *Diachronica* 14.197–232.
- HERCUS, L. A. 1982. *The Bāgandji Language*. Canberra: Pacific Linguistics.
- KILHAM, C., PAMULKAN, M., POOTCHEMUNKA, J., and WOLMBY, T. 1986. *Dictionary and Source Book of the Wik-Mungkan Language*. Darwin: Summer Institute of Linguistics, Australian Aborigines Branch.
- MC CONVELL, P. 1985. 'The origin of subsections in northern Australia', *Oceania* 56.1–33.
- MERLAN, F. C. 1994. *A Grammar of Wardaman, a Language of the Northern Territory of Australia*. Berlin: Mouton de Gruyter.
- MITHUN, M. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.

- NASH, D. 1982. 'Warlpiri verb roots and preverbs', pp. 165–216 of *Papers in Warlpiri Grammar in Memory of Lothar Jagst* (Workpapers of SIL-AAB, A6), edited by S. Swartz. Darwin: Summer Institute of Linguistics.
- NORDLINGER, R. 1998. *A Grammar of Wambaya, Northern Territory (Australia)*. Canberra: Pacific Linguistics.
- O'GRADY, G. N. 1966. 'Proto-Ngayarda phonology', *OL* 5.71–130.
- , VOEGELIN, C. F., and VOEGELIN, F. M. 1966, *Languages of the World: Indo-Pacific, Fascicle 6 (Anthropological Linguistics 8, ii)*.
- PENSALFINI, R. T. 1997. 'Jingulu Grammar, Dictionary and Texts'. Ph.D. dissertation, MIT.
- SANDS, K. 1996. *The Ergative in Proto-Australian*. Munich: Lincom Europa.
- SCHMIDT, W. 1919. *Die Gliederung der australischen Sprachen*. Vienna: Druck und Verlag der Mechitharisten-Buchdruckerei.
- SMITH, I., and JOHNSON, S. 2000. 'Kugu Nganhcara', pp. 357–489 of *Handbook of Australian Languages*, vol. 5, edited by R. M. W. Dixon and B. J. Blake. Melbourne: Oxford University Press.
- SUTTON, P. 1978. 'Wik: Aboriginal Society, Territory and Language at Cape Keerweer, Cape York Peninsula, Australia'. Ph.D. dissertation, University of Queensland.
- TURNBULL, W. 1903. Letter to the editor. *Science of Man*, vol. 6, no. 1 (21 February), pp. 9–11.
- WURM, S. A. 1972. *Languages of Australia and Tasmania*. The Hague: Mouton.

Descent and Diffusion: The Complexity of the Pilbara Situation

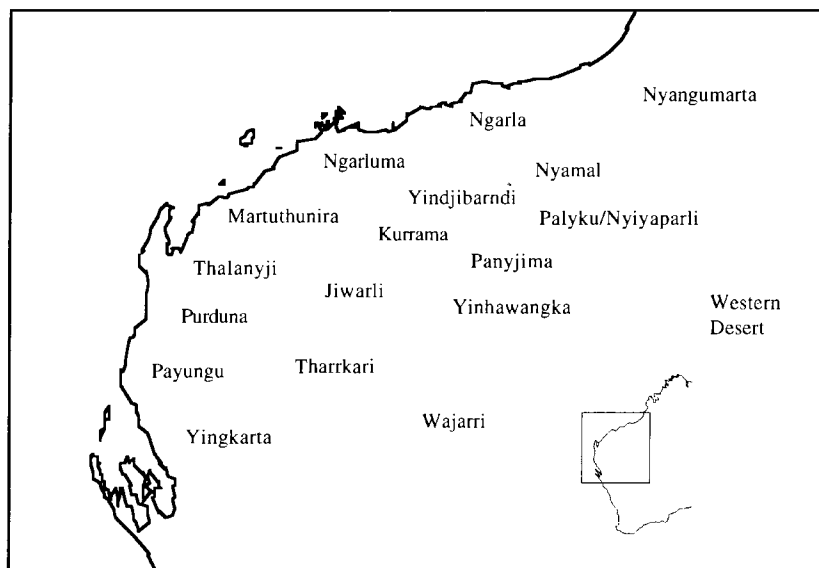
Alan Dench

1. Introduction

This chapter presents a case study of a group of Australian languages for which it is especially difficult to determine whether shared innovations are the result of genetic inheritance from a common ancestor or are the result of contact. In the preceding chapter of this volume, Dixon demonstrates the difficulty of determining clear genetic groupings across the Australian continent as a whole. This chapter shows that similar problems of indeterminacy can hold at the lowest level of language comparison in the Australian context.

The Pilbara and adjacent regions of Western Australia are home to a relatively wide diversity of Australian languages. While the languages of the Pilbara are likely to be genetically related, at some level, there is a good deal of morphosyntactic variety within the region: the Southern Pilbara languages have an extensive tripartite case-marking system, the Central Pilbara languages have innovated a nominative/accusative case-marking system, and the Northern Pilbara languages unlike either of the other groups have a split ergative case-marking system very like that of languages to their north and east, and an agreement system in the verb. This diversity of morphosyntactic type is found in few places in Australia and so the Pilbara provides a useful laboratory in which to investigate genetic versus diffusional relationships. This chapter explores some aspects of the languages of the region and considers the extent to which, on the basis of these few examples, we may ultimately be able to decide between genetic explanations and diffusional explanations in accounting for similarities among different languages.

The chapter is organized as follows. After introducing the languages in their geographical and social context (§1.1), I approach the question of classification from a brief review of previous attempts to determine genetic relationships within the area and a discussion of methodological principles (§1.2). The review points to a set of phonological and morphosyntactic features which appear most likely to be useful in testing for genetic versus diffusional accounts of similarity and these are discussed in sections which follow: phonological innovations including lenition of stops, fortition of liquids, and simplification of sonorant + obstruent clusters (§2);



MAP 1. Approximate locations of languages in the (wider) Pilbara region

morphophonemic alternations dependent on stem length (§3); patterns of case-marking, in particular the nature of split systems and alignment shifts (§4). For each of the patterns discussed, I will consider to what extent the construction of a genetic account of shared similarities might be compromised by patterns which appear to require an explanation involving some diffusion of form and/or pattern. Section §5 summarizes the evidence and considers its wider implications.

1.1. GEOGRAPHICAL AND SOCIAL CONTEXT

The Pilbara is generally recognized as a distinct ecological region with a characteristic climate, geology, fauna, and flora. It lies within the Australian Arid Zone and is bordered by the Great Sandy Desert in the north-east, the Little Sandy Desert in the east, and the Carnarvon and Gascoyne regions to the west and south (see Beard 1990).

There are around twenty named languages¹ recognized for the area. Table 1 lists

¹ I use the term 'language' somewhat loosely here. I make use of the local identification of what constitutes a language even though these divisions in some cases do not correspond to what a linguist might choose to distinguish by some measure of relative similarity. Speakers of 'languages' in the region recognize the difference between sociopolitical groups who share a named language, and local groups identified for a particular geographically defined region and who may have a distinct 'dialect' of a 'language'. Thus Kurrama and Yindjibarndi, for example, are linguistically very similar to one another, yet are recognized as the languages of different peoples. The Kurrama recognize local groups, such as *Yartira*, *Ngamangamara*, etc., who are understood to speak identifiably different varieties of Kurrama.

TABLE 1: Languages and received/standard classification

Sources		Group	Regional Label
Nyangumarta	Sharp 1998	Marrngu	<i>Northern Desert</i>
Nyamal	Dench fieldnotes	Ngayarta	<i>Northern Pilbara</i>
Ngarla	Dench fieldnotes		
Palyku/Nyiyaparli	Kohn 1996, Dench fieldnotes		<i>Central Pilbara</i>
Panyjima	Dench 1991		
Yinhawangka	Dench fieldnotes		
Yindjibarndi	Wordick 1982		
Kurrama	Dench fieldnotes		<i>Southern Pilbara</i>
Ngarluma	Simpson 1983, Kohn 1994		
Martuthunira	Dench 1995		
Thalanyji	Austin 1981a, b, 1994b	Kanyara	
Purduna	Austin 1994b		
Payungu	Austin 1994b	Mantharta	
Tharrkari	Austin 1981a, b, 1994a		
Jiwarli	Austin 1994a	Kardu	
Yingkarta	Dench 1998c		
Wajarri	Douglas 1981, Marmion 1996		<i>Murchison</i>

those which are explicitly mentioned in this study together with references to the principal sources. I have also given the received classifications of these languages following O'Grady, Voegelin, and Voegelin (1966), O'Grady (1966), and as revised by Austin (1988). These labels are well established in the Australian literature and so are given here to provide general accessibility, even though the arguments presented in this chapter suggest that they may not be supportable. I have also given geographical labels and will make some use of these throughout the chapter. Map 1 shows the approximate location of the languages.

There is a close fit between the different ecological regions and what have traditionally been treated as distinct linguistic subgroups. Thus, the Pilbara region (proper) is inhabited by speakers of Ngayarta languages; speakers of Nyangumarta and of other Marrngu languages inhabit the western part of the Great Sandy Desert, the westernmost speakers of Western Desert languages inhabit the Little Sandy Desert, Mantharta languages were spoken in the Gascoyne, Kanyara languages in the Carnarvon region, Yingkarta also falls into the Carnarvon region, and Wajarri was spoken in the Murchison and Gascoyne.

The language communities of the region can also be grouped by some broad cultural criteria—though it is fair to say that there is little detailed ethnographic

work for the region. The most significant features here are perhaps the westward extent of the circumcision and subincision rites (their linguistic importance first being suggested by O'Grady (1958)), and the patterning of the kinship systems and section systems.

The western boundary of the circumcision rite separated the Ngarluma, and Martuthunira, together with the Thalanyji, Purduna, Payungu, Tharrkari, and Jiwarli, from the peoples to the east. In the west, male initiation involved, instead, a practice of binding the biceps with a tourniquet. This arm-tying practice united the peoples of the Southern Pilbara, and was shared with the Ngarluma. My Martuthunira informant maintained that the Martuthunira did not practise this, but that eastern Martuthunira sent young men to the Kurrama and Yindjibarndi for circumcision. For the south, the Wajarri practise circumcision but it is not certain that the Yingkarta ever did. What is clear is that the eastern rites have been gradually progressing westwards and that this was the case well before European contact.

The communities of the Pilbara mainly had a four-section system for classifying kin though there are some local differences in the arrangement of the terms of that system. However, the southernmost communities appear not to have had a well developed section system. Tindale (1974) classes such groups as of the Nhanta type. The Wajarri appear to have been of this type, and Austin's (1996) comments on the Payungu suggest that they too did not have a well-established section system. I was not able to elicit information from Yingkarta speakers, and assume that they too fall into this category.

The kinship systems are, in the northern areas, of the prototypically Kariera type with two patri/matrilineal lines of descent and the (theoretical) possibility of cross-cousin marriage. Radcliffe-Brown (1913) distinguished a Martuthunira type similar to the Aranda type (with four lines of descent and marriage dependent on cross-cousin links in the parent's generation), but Scheffler (1978) has argued that the Martuthunira system is essentially of the Kariera type. Austin (1996) maintains the position that the Mantharta and Kanyara language groups all had the Radcliffe-Brown Talaindji type of organization, also a variant of the Arandic type. There are similarities amongst the different systems, but the details of patterns of similarity and difference, and their implications for a unified view of the cultural diversity of the region, have not been worked out. What we can say from the geographical extent of these different cultural traits—patterns of male initiation, the section system, and the Aranda-type versus Kariera-type kinship system—is that they do not coincide to define distinct cultural boundaries.

There are similarities between the traditional mythologies of the Southern Pilbara and those of their neighbours in the Central Pilbara but where themes are shared with groups in the desert, these turn out to be pan-Australian (Austin 1996). The few origin myths (legends) I collected for Martuthunira point to close connections between the Martuthunira and their coastal neighbours. Clearly there were

well-established north(east)–south(west) trade routes in manufactured items, just as there were established trade relations between coastal peoples and hinterland peoples (for example in the distribution of pearl-shell ornament and shell water-carriers).

These different cultural patterns point to long and well-established contact amongst groups in the region. Indeed, there are enough similarities that we cannot immediately establish any strong evidence that originally culturally distinct groups have come into contact, rather than that the diversity reflects a gradual differentiation as different cultural innovations have diffused into and across the area.

It is important to consider how such contact is maintained and what might be its implications for the linguistic ecology of the region. Exogamy between different language groups is common and is encouraged by a system of promised marriage arising through male initiation practices. Multilingualism is correspondingly common and people pride themselves on the number of languages they know, though they do not necessarily proclaim rights to speak them all. Initiation meetings, amongst others, bring people from different language groups together. Men and women travel into the country of other language groups to further their knowledge of traditional law and to learn new songs, stories, and occasionally languages. Thus the contact between speakers of different languages and the prevalence of multilingualism provides every opportunity for diffusion and language ‘convergence’.

Nevertheless, there is a strong tradition of linguistic integrity. Languages are appropriate to particular areas of land and speakers identify both with a language and with country. While language shift does occur at an individual level, regions of country do not as easily change their linguistic affiliation. Even where a language has become extinct, the country its speakers called their own remains affiliated with that forgotten language. Traditional mythology in the area describes patterns of succession—one group replacing another (usually depicted as ‘devils’) through conquest—but both are described as speaking the same language.

Folk descriptions of dialect and language differences occasionally make reference to lexical differences (essentially differences in lexification; for discussion see Ross in Chapter 6) but these are not always the first point of comparison. Very often, such descriptions allude to phonological differences and speakers recognize instinctively the sometimes subtle phonotactic differences among languages with strikingly similar phonetic inventories. While speakers are aware of the morphosyntactic differences among languages (speakers certainly know the difference between languages that have a passive and those that do not), this does not prevent some calquing of morphosyntactic patterns. I know of recorded instances where, for example, a Ngarla speaker calqued the Ngarla subordinate clause case marking patterns onto Nyamal, a Nyamal speaker made an exact mirror image adjustment in her Ngarla, and a Martuthunira/Kurrama speaker

levelled the tripartite case-marking system of Thalanyji towards something more closely resembling the consistent accusative alignment of his primary lects. All of these examples can be described as transfer resulting from the imperfect learning of a secondary lect.

Linguistic differences do serve as emblems of identity, usually through linking speakers associated with a common area of country, but it is not so clear that language identity serves to define a group of people independently of their shared affiliation to land. In Ross's terms (see Chapter 6) the Pilbara communities can be characterized as open and loose-knit, and his model would allow the possibility of such communities shifting from a primary to a secondary lect. But in the Pilbara there is no shared community secondary lect (thus no common target) nor does there appear to be any strong motivation to change. The traditional social context provides ample opportunity for the diffusion of linguistic patterns but the fact of contact does not provide a motivation to change.

1.2. PREVIOUS CLASSIFICATIONS

As noted in the introduction, it is likely that all the Pilbara languages are related, though there is no evidence that they form a genetic subgroup of some higher grouping of Australian languages. Their relatedness is demonstrated by O'Grady's (1966) and Austin's (1981b) reconstruction of vocabulary and by their reflexes of wider reconstructions of phonology (Dixon 1980), pronoun systems (Dixon 1980, Dench 1994), verb morphology (Dixon 1980, Dench 1998b), and nominal morphology (Dixon 1980, Sands 1996). Ultimately, however, the assumption of relatedness is only as safe as these reconstructions and as Dixon points out in Chapter 4 of this volume, and as this chapter serves to illustrate, no reconstruction that does not seriously consider the diffusional propensities of Australian languages can be considered entirely safe.

The linguistic classification of languages in the wider Pilbara area has been discussed a number of times in the literature. It is possible to recognize three broad approaches represented in the relevant studies. The first, 'classical' lexicostatistics, provides the initial classification into groups based on lexical similarity. The second, which can be described as broadly typological, proceeds by identifying similarities in type while making no detailed attempt at reconstruction or explanation of similarity in genetic terms. The approach has a long history in Australia: the common classification of Australian languages into the two broad groups, prefixing and non-prefixing, is just such a typological classification. The third approach is classical subgrouping by the comparative method, proceeding through the reconstruction of aspects of a proto-language and the subsequent identification of shared innovations in some subset of daughters. However, while reference is made to this third method none of the existing attempts to classify the Pilbara languages is strictly faithful to it.

The first detailed classification of Western languages was presented by O'Grady, Voegelin, and Voegelin (1966), based on the lexicostatistical analysis of

TABLE 2: Distribution of shared grammatical features, following O'Grady (1966)

	1	2	3	4	5	6	7	8	9	10	11	12
<i>Western Desert</i>									•	•	•	
Nyangumarta									•	•	•	
Ngarla		•	I	I					•		•	
Nyamal		•	I	I					•	•	•	I
Palyku/Nyiyaparli	?	•	I	I					•		•	I
Panyjima	•	•	I	I	I	I	I	I	(•)		•	I
Yinhawangka	•	•	I	I	I	I	I				•	I
Yindjibarndi/Kurrama	•	•	I	I	I	I	I	I			•	I
Ngarluma	•	•	I	I	I	I	I	I			•	I
Martuthunira	•	•	I		I	I	I				•	I
Thalanyji	•	•	I								•	I
Purduna	•	•	I								•	I
Payungu	•	•	I									I
Tharrkari	•	•	I								•	I
Jiwarli	•	•	I								•	I
Yingkarta	•	•	(I)						•			
Wajarri	•	•	I						•		•	

1. phonemic laminal contrast
2. lack of initial apicals
3. ergative allomorphs conditioned by length of nominal stem
4. nasal dissimilation of *-ngkV* suffixes
5. active/passive voice distinction
6. loss of ergative marking for transitive subjects
7. generalization of 'dative' to general 'objective'
8. shift of future/purposive to present tense
9. use of bound person markers
10. negative particle + irrealis marking in negative clauses
11. inclusive/exclusive contrast in pronoun paradigm
12. loss (reanalysis) of monosyllabic verb stems

The shaded band is O'Grady's Ngayarta group, the symbol I indicates a suspected innovation, and features in parentheses have marginal status in the particular language.

core vocabulary partially mediated by O'Grady's additional knowledge of the languages and a view of their typological similarity. O'Grady's (1966) paper is a detailed reconstruction of the phonology of his Ngayarta group. It does not provide historical phonological evidence for the Ngayarta group as such—the reconstruction focuses on a few innovating languages within the group—but O'Grady does present grammatical evidence supporting the earlier lexicostatistical classification. The evidence is given in the form of shared features, with some reference to innovations (though these are not argued). Some discussion of O'Grady's diagnostic features and attendant methodology is presented in Dench (1998b), in reply to O'Grady and Laughren (1997).

O'Grady's three groups of four features are summarized in Table 2. The details

here have been updated in the light of new information about the languages and so do not correspond exactly to O'Grady's (1966) findings.²

As this summary shows, the Ngayarda group is far from homogeneous with respect to the distribution of the features. The strongest *prima facie* evidence for some kind of genetic grouping is provided by features 4–8 (three of which are interconnected and involve a shift in alignment from split-ergative to accusative). Yet even here, we do not find a completely uniform distribution nor any features which correspond exactly to the Ngayarda group. But beyond this, it is the status of these shared innovations—as diffusional or as arising through shared inheritance—which is most at issue here. We will return to some detailed consideration of these features in sections which follow.

Like O'Grady's (1966) classification, Austin's (1988) classification of the Southern Pilbara languages uses mainly morphosyntactic criteria to define, most importantly, the Kanyara and Mantharta groups. Austin describes the classification as a hypothesis of genetic relationship yet makes no determined effort to demonstrate that particular features are shared innovations. In fact, he lists known retentions as symptomatic of close relationship where found in a restricted group of contiguous languages in a particular (even conservative) configuration. Thus, the approach identifies clusters of properties which might be considered akin to bundles of historical isoglosses while nevertheless using the terminology of classical subgrouping. More recently, Austin (1996) has provided additional discussion of the relationships between Mantharta and Kanyara languages. He provides a list of grammatical similarities shared by the two groups of languages and again emphasizes typological pattern rather than strictly identifiable form–function correspondence.

So far there have been no attempts at a classification of the languages of the Pilbara which can be seriously defended as genetic classifications despite intentions in this direction. There are a number of problems with the existing studies. First, shared innovations cannot be recognized without the prior recognition of a deeper reconstruction, and in most cases this reconstruction is lacking. Second, there has been an expectation that morphosyntactic criteria (including typological features) will ultimately align with the original lexicostatistical classification and will thus support it as a *bona fide* genetic classification. Third, similarity in pattern has sometimes been given importance over form in deciding questions of relationship. There has also been insufficient attention paid to the possibility that shared features might be explained as diffusional. There has been no overt suggestion that rather than constituting a set of closely related languages, the Pilbara should be seen as one or more diffusional zones. This chapter represents a first, necessarily brief, attempt to consider this aspect of the problem.

² This does not alter the basic point of the exercise. O'Grady did not have complete data sets but was well aware of this. Thus he writes, for example, 'the better known languages of the Ngayarda . . . group'. This neither explicitly assumes nor denies the possibility that the feature might be more widely distributed amongst languages in this group.

Of course, making the argument for an innovation shared by virtue of a period of common development is never easy. I take it for granted that a statement of shared inheritance as explanation for a shared feature should only be made once all other possible explanations for the shared feature have been exhausted. These other possibilities will include accidental similarity in form, borrowing, and genetic drift.

Where a group of languages are *a priori* quite likely to have a common ancestor, remain typologically similar, and remain in close contact, as is the case in the Pilbara, deciding which shared features have resulted from a shared inheritance and which from contact can never be an easy task. Standard application of the comparative method assumes that borrowing is identifiable (since borrowings may form classes of definable exceptions, or have an abbreviated history) and that it can be factored out in the procedure of reconstruction. Yet there are circumstances in which this can be especially difficult and in some cases may not be possible.

The alternative is to seek to explain patterns of similarity as the result of contact. After all, at the micro-level all linguistic change is diffusional, both within a linguistic system and within a speech community. We do recognize barriers to diffusion—geographical, typological, and social—yet so far we have no fully articulated theory of the relative diffusibility, in absolute universal terms, of different grammatical subsystems.³ Where languages spring from a common source their split must be seen as the eventual result of an accumulation of blocked diffusions—waves breaking against geographical and, most importantly, social barriers. In just the same way, linguistic areas meet their limits at some barrier to diffusion.

Choosing explanations for shared innovative features amongst related languages in a possible diffusion zone may be a matter of taste. While it may be wise to assign an important role to diffusion, if the languages *are* ultimately related then surely some things are inherited, and perhaps the innovations arose in a period of shared history (that is, a history involving a single speech community). How do we decide amongst the alternatives? One possibility is to begin by attempting to identify patterns which are most clearly the results of diffusion and attempting to distinguish these from patterns which are most clearly the result of a shared innovative inheritance. If this can be done, then the results of such a categorization might then be used to assist in deciding less clear cases. Essentially then, we would reconstruct a history of contact and inheritance and allow it to lead us to particular solutions for changes which cannot, on their own merits, decide the question.

However, we should leave open the possibility that all questions may turn out to be undecidable. It may not be possible to show conclusively for any particular

³ Steps in this direction are taken in a number of the studies presented in this volume, and especially by Curnow in Chapter 15.

innovation that it results from genetic inheritance rather than that it is motivated by contact with another language. If enough such cases occur, then the suspicion we might attach to any putative inherited innovation will mount and we should become increasingly sceptical of any suggested genetic classification. The present case study leads to this conclusion for the Pilbara languages, and supports Dixon's broader characterization of the Australian linguistic situation presented in the preceding chapter.

2. Phonological innovations

The greater part of O'Grady (1966) is a detailed reconstruction of phonological changes in Yindjibarndi and Kurrama in comparison with a set of reconstructed Pilbara vocabulary items. Austin (1981b) describes quite similar changes in two Southern Pilbara languages, Purduna and Tharrkari, and also provides a list of reconstructed vocabulary for languages of this area. Austin (1982) shows that phonetic tendencies in the conservative languages of the Mantharta group parallel the changes in Purduna and Tharrkari, suggesting some areal distribution. I have also commented on the areal nature of the changes in describing patterns in Martuthunira (Dench 1995), and similar observations are made in Dench (1998c) in considering allophonic patterns in Yingkarta, and in comparison with Wajarri and Nhanta further to the south.

What is suggested by these studies is that none of the changes described serve to uniquely identify any particular grouping of languages. Instead, there appear to be regional tendencies. The tendencies include the lenition of intervocalic stops, the 'simplification' of sonorant + obstruent clusters, and the more general fortition of laterals and rhotics. These patterns are described in the following subsections.

2.1. LENITION OF INTERVOCALIC STOPS

The following tables provide summary details of the lenition of intervocalic stops occurring for selected environments in those languages which show the changes. Details are summarized from O'Grady (1966) (with some refinements to the

TABLE 3. Lenition and loss of intervocalic peripheral stops

	Martuthunira		Yindji/Kurr.		Purduna		Tharrkari	
	*k	*p	*k	*p	*k	*p	*k	*p
V _i —V _i	Ø	w	Ø	Ø	Ø	Ø	g/w	w
V _i —V _j	w	w	w	w	w	w	g/w	w
wV _i —V _i	Ø	w	k	p	Ø/g	b	g	?
wV _i —V _j	w	?	k	p	g	b	?	?

TABLE 4. Lenition and loss of intervocalic laminal stops

	Martuthunira		Yindji/Kurr.		Purduna		Tharrkari	
	*j	*th	*j	*th	*j	*th	*j	*th
$V_i \text{---} V_i$	<i>y</i>	<i>th/Ø</i>	Ø	Ø	<i>y</i>	<i>y</i>	<i>j/y</i>	<i>dh</i>
$V_i \text{---} V_j$	<i>y</i>	<i>th</i>	<i>y</i>	<i>yh</i>	<i>y</i>	<i>y</i>	<i>j/y</i>	<i>dh</i>
$wV_i \text{---} V_i$	<i>y</i>	<i>th</i>	<i>j</i>	<i>yh</i>	?	?	<i>j/y</i>	<i>dh</i>
$wV_i \text{---} V_j$	<i>y</i>	<i>th</i>	<i>j</i>	<i>yh</i>	?	?	<i>j/y</i>	<i>dh</i>

Kurrama patterns based on more recent data collections), Austin (1981b), and Dench (1995). The details of some conditioned splits are not given here as they become a little complicated and do not alter the case; see O'Grady (1966) and Austin (1981b) for details.

In addition to these general patterns of lenition, there are instances of more restricted lenitions within paradigms. The most pervasive exemplar of this is the genitive/dative suffix, **-ku*, which has a lenited form, *-wu*, or *-yu* in a number of languages which otherwise show little evidence of regular lenition processes (see Dench 1998a, and §3 below). Further, we should recognize some phonetic lenition even where this does not lead to phonemic split. For example, the lamino-dental stop /th/ in Martuthunira has lenited variants consistent with the phonetic patterns of Yindjibarndi and Kurrama. However, in these latter two languages other changes have led to a split such that a new lamino-dental glide phoneme, /yh/, has arisen.

The patterns of lenition shown here are very similar, especially in that, with the exception of Martuthunira, the extremes of lenition are blocked where the preceding consonant is already a glide. This might suggest that we look for a sequence of ordered changes with the variation resulting from a late differentiation of the languages. However, intervocalic lenitions are extremely common and may arise independently (perhaps as an instance of drift). No one should be tempted to posit subgrouping on this kind of evidence alone.

2.2. CLUSTER SIMPLIFICATION AND LATERAL/RHOTIC FORTITION

More interestingly, a number of languages in the Pilbara region show a general tendency to avoid liquid + stop and nasal + stop clusters. The patterns of change affecting such clusters range from loss or fortition of the nasal or lateral preceding the stop, to lenition of the stop following the sonorant. The changes might be viewed as a general conspiracy against the mixing of manners.

Austin (1981b) describes the history of N+S clusters in Purduna and Tharrkari, languages which as we have seen already suffer a degree of intervocalic consonant lenition. In these two languages, nasals in homorganic N+S clusters are simply deleted and the cluster surfaces as a voiceless stop. It is this change which, paired

with the allophonic tendency to voice intervocalic stops, has led to the development of a voicing contrast. In heterorganic N+S clusters, the nasal is realized as a stop.

While the general loss of nasals from homorganic N+S clusters is restricted to Southern Pilbara languages, there is a similar though more restricted morphophonemic alternation in languages of the Northern Pilbara and Central Pilbara. In all of these languages except Martuthunira, the dimoraic allomorphs of the locative (*-ngka*) and (old) ergative (*-ngku*) suffixes are affected by a rule of nasal dissimilation—the N+S cluster is simplified to a stop if the preceding syllable boundary also involves a N+S cluster (this is feature 4 in Table 2). The following examples involving the locative suffix are from Panyjima:

<i>yurlu-ngka</i>	<i>jinyji-ka</i>
<i>munma-ngka</i>	<i>mungka-ka</i>
<i>pulku-ngka</i>	<i>yinti-ka</i>

The rule appears to be restricted to just reflexes of the ergative and locative in most of the languages affected, though there is evidence from surviving irregular dative pronoun forms in Nyiyaparli that the rule also affected a pronominal dative formative, **-mpa*. Wordick (1982) suggests that in Yindjibarndi the rule also affects the clitic particle *-mpa*. More generally, sequences of homorganic N+S clusters are rare in these languages and it may be that this reflects an earlier general phonotactic constraint. However, I have not identified any clear cognates which might establish loss of the nasal as a widespread phonological change.

The ‘simplification’ of lateral+stop clusters is more general. First, in Tharrkari these are strengthened just as are the N+S clusters—the heterorganic lateral is realized as a stop. This fortition of laterals is more extensive in Tharrkari to the extent that in one dialect all laterals have merged with the corresponding stop. Austin (1982) reports similar replacement of intervocalic laterals with stops by one of the two then remaining Purduna speakers, a pattern described by the other as a feature of a southern dialect. In Martuthunira, the response to L+S clusters is quite different—a bilabial or velar stop following a lateral or the apical tap/trill is lenited to a glide.

In Yindjibarndi and Kurrama, the patterns are more complex, and both patterns—the fortition of the lateral or lenition of the stop—are found. In Kurrama, all syllable-final laterals have merged with stops (including those in word-final position). Table 5 presents the reflexes of clusters involving the laterals and trill and the peripheral stops for the four languages.

More generally, we can see phonetic tendencies in a range of languages which parallel the phonological changes in the four languages described here. First, Austin (1982) describes phonetic tendencies in Jiwari and in Yingkarta which parallel changes in Tharrkari. In Jiwari, nasals are lost from homorganic clusters and are strengthened to stops in heterorganic clusters, in fast speech. Laterals are also occasionally realized as stops in preconsonantal position and in intervocalic

TABLE 5. Reflexes of L + S and *rr* + S clusters in innovating languages

	Martuthunira	Yindjibarndi	Kurrama	Tharrkari
*lp	<i>lw</i>	<i>tp</i>	<i>tp</i>	<i>tp</i>
*lk	<i>l.y</i>	<i>rrk</i>	<i>tk</i>	<i>tk</i>
*rlp	<i>rlw</i>	<i>rp</i>	<i>rtp</i>	<i>rtp</i>
*rlk	<i>l.y</i>	<i>rk</i>	<i>rtk</i>	<i>rtk</i>
*lyp	<i>lyw</i>	<i>jp</i>	<i>jp</i>	<i>jp</i>
*lyk	<i>l.y</i>	<i>yk</i>	<i>jk</i>	<i>jk</i>
*rrp	<i>rrw</i>	<i>rrw</i>	<i>rrw</i>	<i>rrp</i>
*rrk	<i>rry</i>	<i>rr/rrw</i>	<i>rrw</i>	<i>rrk</i>

position. In Yingkarta too, laterals may be stopped in intervocalic position. Phonetic pre-stopping of laterals in syllable-final position occurs in Martuthunira (Dench 1995), and phonetic glottal closure of syllable-final laterals occurs in Panyjima (Dench 1991).

A tendency to realize the rhotic trill /*rr*/ as a stop is also widespread in the area. The pattern occurs in Yingkarta (Dench 1998c) and is reported for Wajarri (Marmion 1996). In Nhanta, much further to the south, both lateral-stop and rhotic-stop clusters descend as stop-stop clusters. In this language, other changes have meant these phonetic tendencies have become phonological changes (see Blevins and Marmion 1996 for details).

We might attempt to write ordered rules which account for this variation. For example, if the lateral fortitions are unpacked into a set of changes which first strengthened the laterals preceding an obstruent, then in other syllable-final environments, and finally intervocalically, we could see Tharrkari, Yindjibarndi, and Kurrama as a group from which Yindjibarndi split first (before the second change and with a subsequent lenition of clusters involving, say, a stop preceding a velar stop) and Kurrama later (before the third change). But such an analysis fails to say anything about the changes in Martuthunira L + S clusters and the possibility that these are motivated by similar factors which led to both the lateral fortitions and the more restricted loss or fortition of nasals—an apparent conspiracy to ‘simplify’ clusters consisting of consonants with distinct manners.

Attempts to write rules which capture the shared innovations amongst the innovating languages in terms of family trees would thus fail to capture a number of interesting and apparently areal features. Table 5 gives the clearest example: changes which avoid liquid + stop clusters and where different languages have found different solutions to essentially the same problem. The patterns of change include lenition, loss, and fortition but in different combinations.

It should be pointed out that speakers of these languages are apparently well aware of patterns of correspondence among languages. Algy Paterson once offered a Martuthunira form *ngal.yu* for ‘wild onion’, later remembering that the

Martuthunira word was *partunya*. He explained that he had constructed the form by analogy with the following set: Panyjima *ngarlku*, Yindjibarndi *ngarku*, Kurrama *ngartku*, hence Martuthunira *ngal.yu*. This explicit correspondence mimicry has a number of consequences. First, it is clear that where speakers are able to perform this kind of feat happily, we use regional patterns of phonological variation as evidence for subgrouping at some peril. But we must also wonder whether the language-engineering revealed in this correspondence mimicry is not more widespread. Yindjibarndi and Kurrama are remarkably similar lects differing most obviously in their phonotactic patterns. Similarly, Tharrkari differs from Jiwarli mainly in its phonological patterns. It is at least conceivable that these differences are consciously maintained in order to preserve some distinction between the different lects.

The discussion in this section presents our first example of relatively involved variation amongst lects but which cannot be conclusively described as the result of differentiation through inheritance of successive innovations. While models of this kind might be constructed from the data, they would fail to recognize some areal patterns in the kinds of phonological changes that have occurred and which argue, on the other hand, that the changes have not arisen in isolation from one another. Of course not all of the changes described here have arisen through contact; there had to be initial innovations in some lect or lects which may then have diffused or triggered similar changes in neighbouring languages. The problem is the undecidability of the issue: it is very difficult to determine where the innovations arose and where and how they have influenced patterns in other lects.

3. Morphophonemic alternations

As one of the features distinguishing western languages, O'Grady (1966: 75) noted that these languages share a rule of 'morphophonemic alternation in the form of the "agent-instrumental" suffix **-lu/-ngku*, conditioned by the length of the word stem'. The patterning of this agent/instrumental suffix (commonly the ergative) is paralleled by the patterning of the locative—it is common in Australian languages for the two suffixes to show near-identical patterns of conditioning.

In many of the Pilbara languages, the allomorphs on vowel-final stems are split depending on whether the stem contains two morae, or more than two morae. For the locative, the forms are *-ngka* on dimoraic stems and *-la* on polymoraic stems. By contrast, in languages of the Western Desert the allomorphs are semantically conditioned. The suffix is *-ngka* on stems denoting lower animates and inanimates and *-la* on proper names and stems denoting humans.

In none of the Pilbara languages is the semantic basis of the conditioning completely absent. Thus even where a language has length conditioning of the allomorphs, dimoraic pronouns and proper names may consistently select the *-la* allomorph. O'Grady saw the length-conditioned pattern as a retention (and see Hale 1976, Dixon 1980), yet this assumption is now in doubt following Sands'

TABLE 6. Allomorphs of (*)ergative, locative, and (*)dative

	Locative		(*)Ergative		(*)Dative	
	2mora	2+mora	2mora	2+mora	2mora	2+mora
Nyangumarta	-ngka	-la	-ngku	-lu	-ku	-ku
Ngarla	-ngura	-la	-ngku	-lu	-rra	-ku
Nyamal	-ngka	-la	-ngku	-lu	-yu	-ku
Palyku/Niyiyaparli	-ngka	-la	-ngku	-lu	-yu	-ku
Panyjima	-ngka	-la	-ngku	-lu	-yu	-ku
Yinhawangka	-ngka	-la	-ngku	-lu	-yu	-ku
Yindji/Kurrama	-ngka	-la	-ngku	-lu	*-yu	*-ku
Ngarluma	-ngka	-la	—	—	-yu	-ku
Martuthunira	-ngka	-la	-ngku	-lu	*-ku	*-ku
Thalanyji	-ngka	-la	-ngku	-lu	-ku	-ku
Purduna	-ngka	-la	-ngku	-lu	-ku	-ku
Payungu	-ngura	-la	-ngku	-lu	-ku	-ku
Tharrkari	-ngka	-la	-ngku	-lu	-ku	-ku
Jiwarli	-ngka	-la	-ngku	-lu	-ku	-ku
Yingkarta	-ngka	-la	-ngku	-lu	-ku	-ku
Wajarri	-ngka	-la	-ngku	-lu	-ku	-ku

Shading shows the retention of original allomorphs of the dative

(1996) work on the problem. We can view the remaining semantic conditioning as a retention and the length-conditioned allomorphy as an innovation.

Length-conditioning of allomorphs is not restricted to the locative and old ergative suffixes—a number of Central Pilbara languages also have length conditioned allomorphs of the old dative (reanalysed in these languages as the accusative). Here the allomorph selected by dimoraic stems, *-yu*, is a lenited form of the allomorph selected by polymoraic stems, *-ku*, possibly arising as a result of its weakly stressed environment in simple trisyllabic forms (see Dench 1998a). For most languages, the patterns of alternation involve the identical set of suffix allomorphs, but there are exceptions, as Table 6 shows.

We see that the innovation of a length conditioned pattern for the old dative falls within the wider region showing this conditioning for the locative and old ergative. While it may be that the conditioning of the dative allomorphs has an independent motivation, it is not unreasonable to suppose that it was aided and abetted by analogy with the conditioning patterns of the locative and ergative.

The exceptional forms in Ngarla and Payungu (shown in bold face) raise another question. While their provenance remains unclear (and it seems reasonable given wider patterns to suppose that they are innovations) there appears to be no immediately good reason why the alternations in these languages should conform to the dimoraic template, unless by analogy to the patterns established internally by the ergative and/or by the patterns established by alternations in

neighbouring languages. If the latter, then these examples provide the suggestion of an indirect diffusion—regional pressure to conform to a length-conditioned pattern of allomorphy.

Finally, while Yingkarta and Wajarri are listed in Table 6 as sharing the pattern of length-conditioning, these patterns are not as robust as they are in other languages. Marmion (1996) describes such a pattern for Wajarri but Douglas (1981), and based on data recorded at least twenty years earlier than Marmion's, describes a rule of semantic conditioning very similar to that found in the Western Desert languages. It is not inconceivable that the difference results from a recent change in Wajarri such that it now conforms more to languages to its north than to languages to its east. It is difficult to say very much about the Yingkarta patterns given the nature of the data. If the variability does not simply reflect a loss of morphophonemic integrity in a dying language, then it too may represent an interrupted shift towards a length-conditioned allomorphy (see Dench 1998c: 20).

Once again, the comparison reveals some areal tendencies in a set of patterns which must be assumed to have arisen originally from innovations in one or more of the languages. The length-conditioned allomorphy of the dative is an innovation shared by a subset of the languages, but it is likely that at least some of this commonality has arisen through the borrowing of a pattern. In the face of this evidence, it cannot be certain that all of the instances in which form and pattern are shared arose in a common ancestor. Determining what is shared through descent and what through diffusion may turn out to be impossible.

4. Case-marking patterns

Turning from the morphophonemic patterns of case suffixes, we can consider the range in case-marking patterns. The most obvious point of variation is due to an alignment shift in Central Pilbara languages such that an old nominative–dative case frame now serves to mark all transitive type clauses. But rather than focus on this pattern alone, it is necessary to consider the factors which contribute to the determination of case marking, to varying degrees, in each of the languages. In doing this, we can consider both the diachronic and synchronic features of the alignment shift and its seeds in the parameters determining the extent of original split-ergative-type marking systems within the region.

In general terms, the choice of case-marking in any clause is dependent on three main parameters: predicate type, nominal type, and clause type (and see Silverstein 1976, Austin 1981a). To begin with, we need to identify a set of predicate types against which different case-marking choices can be mapped. In Table 7, the case choices for representative predicates are shown for languages of the different regions.

In the Northern Pilbara languages, intransitive subjects (S) and transitive objects (O) are typically unmarked absolutive, transitive subjects (A) are ergative. In the Southern Pilbara languages, the transitive object (O) is typically marked

TABLE 7. Predicate subcategorization frame

Predicate type	Example	Northern	Southern	Central
intransitive nominal	'be tall', 'be human'	S	S	Subj
intransitive verb	'sit', 'go'	S	S	Subj
extended nominal	'like', 'fear', 'know'	S DAT	S DAT	Subj O
extended intransitive	'wait for'	S DAT	S DAT	Subj O
transitive verb	'hit', 'see', 'eat'	A O	A O	Subj O
ditransitive verb	'give', 'show'	A O DAT	A O O	Subj O O
	'tell'	A O LOC	A O O	Subj O O

with a distinct accusative case suffix. In the Central Pilbara languages, all subjects are unmarked nominative, all objects are marked accusative. However, this accusative case is the old dative suffix (Dench 1982) and thus the transitive frame in these languages effectively corresponds to the extended intransitive frame of the northern and southern languages. We will have more to say about this in §4.3. First, though, it is necessary to look in more detail at transitive clauses in the Northern Pilbara and Southern Pilbara languages in an attempt to come to terms with the case-marking splits which occur here. Section 4.1 considers splits based on nominal class, §4.2 considers splits based on clause type.

4.1. CASE-MARKING PATTERNS DETERMINED BY NOMINAL CLASS

The split-ergative systems of the Northern Pilbara and Southern Pilbara languages can first be described in terms of nominal class. Since for most classes case is indicated by a clearly segmentable suffix, the splits can in turn be described in terms of the domain (over classes) of ergative (on the one hand) and accusative (on the other) case suffixes. Thus an ergative pattern will exist for a class which falls within the domain of the ergative suffix, but not of the accusative suffix. An accusative pattern will exist for a class which falls within the domain of the accusative suffix but not of the ergative. And a tripartite pattern will exist where a class falls within the domains of both the ergative and accusative suffixes.

The following figure shows the extents of these different case-marking patterns for the Northern Pilbara and Southern Pilbara languages. The accusative suffix, some reflex of **-nha*, is shown with a domain extending downwards from the top of the nominal class hierarchy: the ergative, cognate across the set of languages, extends upwards from the bottom of the hierarchy. Where these overlap, the marking is tripartite.

Two patterns are worth noting here. First, the Southern Pilbara languages (Yingkarta through to Thalanyji) show a degree of tripartite marking in the middle of the hierarchy, but the extent of this differs across the languages. On the one hand, regular ergative marking extends into the pronoun class leaving only the first person singular pronoun, in four of the languages, with an accusative

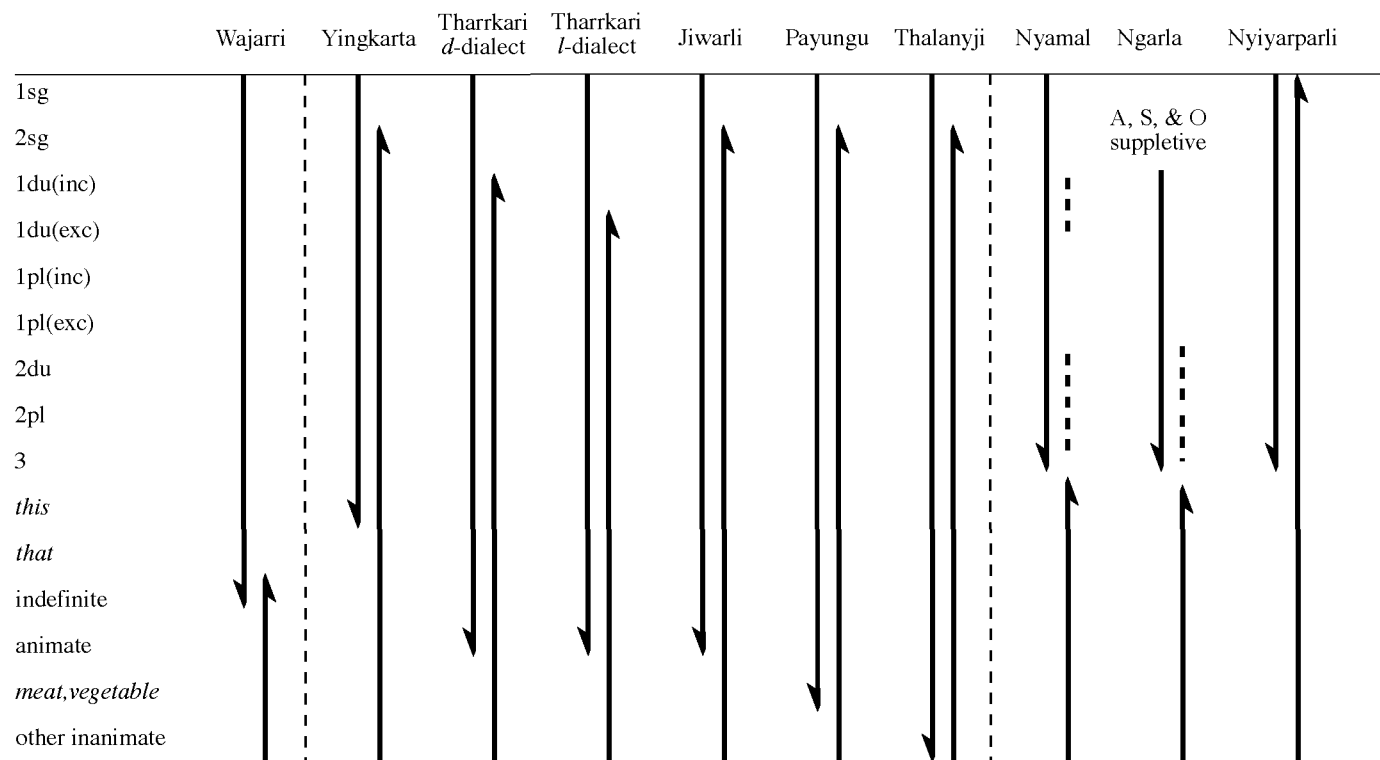


FIGURE 1. Case splits by nominal class

pattern. On the other hand, regular accusative marking extends into the nominal paradigm. In Jiwari and in the two dialects of Tharrkari, all animates have an accusative form. In Payungu, accusative marking also extends to the words for 'meat' and 'vegetable'. In Thalanyji, all nominals take the accusative suffix. In Yingkarta, by contrast, only the proximal demonstrative, 'this', outside of the pronoun class has a distinct accusative form.

The pattern in the Northern Pilbara languages is a little different. The accusative suffix is here restricted to the pronoun paradigm and nominals are consistently marked in an ergative pattern. The extent of ergative marking in the pronoun paradigm varies. In Nyiyaparli, all pronouns take the ergative suffix in transitive subject (A) function and as a result Nyiyaparli has a tripartite pattern in the pronoun class. While Ngarla retains⁴ an original tripartite pattern in the singular paradigm, pronouns in Nyamal and Ngarla otherwise operate in an accusative pattern. However, some forms in the Nyamal and Ngarla nominative paradigm appear to involve an ergative suffix (the bold, dashed lines in Figure 1).

We can make some comparisons with languages further to the east. First, Nyangumarta is closest to Nyiyaparli in that we find apparently regular accusative and ergative forms of pronouns. But there is a complication here. The Nyangumarta accusative pronoun forms which do occur have a very specific function and are syntactically (and in some cases phonologically) bound to the verb. Elsewhere, pronouns in object function are unmarked (Sharp 1998). Thus, leaving the bound forms aside, pronouns and nominals are consistently ergative/absolutive in Nyangumarta. Western Desert languages more closely resemble Nyamal in that pronouns are consistently inflected on an accusative pattern, nominals are consistently ergative. Wajarri, which also borders the Western Desert language area is similar: pronouns and demonstratives are nominative/accusative while nominals are ergative/absolutive, with the single exception of the indefinite/interrogative forms which take both accusative (O) and ergative (A) suffixes (see Figure 1).

Reconstructing an original system of case-marking patterns out of this variation is not a simple task. First, we can be reasonably confident that the first and second person singular pronouns originally inflected on a tripartite pattern, as in Ngarla. But we cannot be certain that this pattern extended beyond the singulars. Second, the nominative (A/S) form of the first person singular pronoun in all languages which no longer have the original tripartite forms is the old ergative. This suggests a collapse of the tripartite pattern into a nominative-accusative pattern. Third, whatever explanation we suggest for the ergative pronoun forms in nominative function in Nyamal, we need to recognize some historical stage (either earlier or later) at which there was some pressure to have pronouns operate in an accusative pattern in contrast to nominals operating in an ergative pattern. This is also the broad system we find in the Western Desert languages.

⁴ For arguments that this is indeed a deep retention, see Dench (1994).

One possibility is a proto-system very like that which continues in Ngarla: distinct nominative, ergative, and accusative forms of singular pronouns, other pronouns taking an accusative suffix in O function, and nominals selecting an ergative suffix in A function. Ngarla is the only local relic of the original case marking system. Changes to this system involved first the loss of distinct nominative singular pronoun forms and thus the creation of a simple pronoun versus nominal split-ergative type. The geographical extent of this simplification is immense and includes all other Pilbara languages and those of the Western Desert. However, it would be a mistake to separate these as a first-order subgroup in contrast to Ngarla. The analogical simplification represented by this change could have occurred independently in any number of languages and could most certainly have been diffused.

Two other changes would then be involved; the spreading of the ergative suffix into the pronoun paradigm, and the extension of the accusative suffix beyond the pronoun paradigm. Both changes have occurred in the Southern Pilbara languages but each to varying degrees, as we have seen. All except Yingkarta have extended the accusative suffix to animates; Payungu and Thalanyji have gone beyond this. Yingkarta has extended the accusative suffix just to the proximal demonstrative. Further south, Wajarri includes both the demonstratives and the indefinite/interrogatives within the domain of the accusative, and thus effectively within a wider pronominal class.

All Southern Pilbara languages except the two dialects of Tharrkari have extended ergative marking to all but the first person singular pronoun. However, the spread of ergative marking into the pronoun paradigm in Nyiyaparli and Nyangumarta is both geographically separated from the Southern Pilbara changes and is different in that there are no surviving accusative patterns within the free pronoun paradigm. The intrusion of ergative-like forms in nominative function into the pronoun paradigms of Nyamal and Ngarla may be the result of either of two processes. The pattern may be a simple analogy of form partly prompted by a trisyllabic template for non-singular pronoun stems (Dench 1994) in other parts of the paradigm, and formed by analogy to (or by the borrowing of) ergative forms in Nyangumarta and Nyiyaparli. Alternatively, the forms may be relics of an earlier tripartite pattern (like that of Nyiyaparli) subsequent to which the original ergative forms extended into intransitive subject (S) function. This latter change then represents the (re)establishment of the pattern keeping pronominal case marking distinct from nominal case marking.

As an alternative history, however, we might consider the wider proto-system to have been somewhere between Nyiyaparli and Ngarla; consistent tripartite marking in the pronoun paradigm but with irregular singular forms, and a consistent ergative pattern in the nominal paradigm. Regularization of the singulars involved selecting the ergative form as a new stem, and the variation in ergative marking in the Southern Pilbara languages is then seen as the (beginnings of a) collapse of tripartite marking at the top of the hierarchy in favour of an accusative

pattern. This has gone still further in the Western Desert region. In Nyamal and Ngarla, the loss of tripartite marking in the pronoun paradigm favoured the retention of ergative forms as nominatives where these fit a trisyllabic template. Accounts of the extension of the accusative suffix into the nominal class are essentially the same under both scenarios.

Whichever scenario we choose, neither presents strong arguments for shared innovations arising from unique inheritance. All changes are simple analogical extensions or levelling and may have occurred either independently within each language, under diffusional pressure from patterns in neighbouring languages, or in one (or more) languages preceding a further splitting into separate speech communities. Certainly sets of languages share similar patterns; the staggered variation most suggests waves of analogical change operating partly independently in each language.

4.2. CASE-MARKING PATTERNS DETERMINED BY CLAUSE TYPE

Case-marking choice in the Northern Pilbara and Southern Pilbara languages also depends on clause type. In those languages for which this is a factor, main clauses involving a transitive verb are typically 'plain' ergative, while a variety of nominalized clauses and dependent clauses are 'normalized' to a (nominative-)dative pattern, or have other patterns of marking for arguments.

The patterns for Nyamal, Ngarla, and Jiwari (which is quite representative of the Southern Pilbara languages) are shown in Table 8. For each clause type, the table indicates whether the type can be used as a main or as a subordinate clause and the case-marking pattern. For the sake of simplicity, Table 8 presents an elaborated set of subordinate clauses as indicated by their distinct verbal inflections. The declarative and nominalized clause types cover a range of inflectional categories.

All languages share the feature that declarative main clauses select a plain ergative pattern of case-marking in which the object is either unmarked absolutive or is marked accusative (in accordance with considerations of nominal class, as described in the preceding section). For at least some subordinate clauses, however, the object is marked dative/genitive. The main clause-subordinate clause split is clearest for the purposive (in Nyamal and Ngarla) and intensive (in Jiwari) clauses. Verbs bearing these inflections select different case-marking patterns according to the dependency status of the clause. While the tendency to suspend the plain ergative pattern is found in all subordinate clauses in the Northern Pilbara languages, in Jiwari this tendency typically does not extend to clauses which are controlled by other than the subject of the matrix clause. Nominalized clauses (and the privative in Ngarla and Nyamal) may be used both as main clauses and as subordinate clauses, with identical case-marking patterns. In all of these, the subject is in the unmarked (absolutive or nominative case) if it appears.

It is important to stress that a reconstruction of patterns for clause types cannot be undertaken without close reference to the forms (and functions) of

TABLE 8. Case patterns by clause type in Nyamal, Ngarla and Jiwarli

Nyamal			Ngarla			Jiwarli		
	Cl	Case		Cl	Case		Cl	Ca
<i>declarative</i>	M	ERG	<i>declarative</i>	M	ERG	<i>declarative</i>	M	ERG
<i>purposive(d)</i>	M	ERG	<i>purposive</i>	M	ERG	<i>intentive</i>	M	ERG
<i>purposive(i)</i>	S	DAT	<i>purposive</i>	S	DAT	<i>intentive</i>	S	DAT
						<i>purposiveSS</i>	S	ALL
						<i>purposiveDS</i>	S	ERG
<i>relativeSS</i>	S	STAT	<i>relativeSS</i>	S	DAT	<i>relativeSS</i>	S	DAT
<i>relativeDS</i>	S	SCE	<i>relativeDS</i>	S	SCE	<i>relativeDS</i>	S	ERG
<i>past relative</i>	S	VInfl	<i>past relative</i>	S	ABL	<i>perfectiveSS</i>	S	DAT
						<i>perfectiveDS</i>	S	ERG
<i>nominalized</i>	VInfl		<i>nominalized</i>	DAT		<i>nominalized</i>	DAT	
<i>privative</i>	DAT		<i>privative</i>	DAT				

(d)	(desiderative)	Case	case-marking
(i)	(implicated)	ERG	A ergative
DS	different subject	DAT	O dative/genitive
SS	same subject	ABL	O ablative
Cl	clause status	SCE	O source
M	main clause	ALL	O allative
S	subordinate clause		
Vinfl	verb inflection		

specific verbal inflections. While Table 8 suggests clear similarities of pattern amongst particular clause types across the languages, unless we consider the verbal forms involved here we cannot speculate about cognacy and cannot begin to reconstruct. This work has still to be done. What is clear is that the suspension of declarative clause case-marking patterns in subordinate and nominalized clauses is typical of the languages of the area and that the most widespread pattern is for this to involve marking the object as dative/genitive. It is this point which leads us to the next section and a consideration of the alignment shift in the Central Pilbara languages.

As a final word, just as we should be wary of drawing too many conclusions about close genetic relationship from instances of shared intervocalic lenition, so should we be wary of placing too much importance on the use of a dative/genitive (and similarly an ‘ablative’ or ‘source’ case) in coding the objects of nominalized and/or subordinate verbs. Such tendencies are widespread in the world’s languages and will be natural choices for analogical extensions where a language seeks to simplify a complex system of case-marking.

4.3. ALIGNMENT SHIFT IN THE CENTRAL PILBARA LANGUAGES

As noted in §1.3, the alignment shift from split-ergative to consistently accusative links a set of Central Pilbara languages and involves the sharing of a number of

features: the loss of the ergative suffix as a marker of A function, the shift of dative to general 'objective' (both direct and indirect object), and the innovation of a passive. At least the first two changes are connected, though how exactly the development of the passive should be integrated with the shift in case-marking remains less clear. The important question to be considered is whether the set of changes is a clear instance of an innovation occurring once in the shared history of the set of languages and thus defining them as a genetic subgroup, or whether the changes might have occurred independently and possibly as a result of regional diffusional pressures.

A solution to the problem depends partly on the mechanism or mechanisms identified as responsible for the shift, and whether or not these can be found to have been the same in each of the modern accusative languages. My current preferred scenario for the change is that it involves a relaxation of the clear formal distinction between dependent and independent clause types that we see in the Northern Pilbara and Southern Pilbara languages. Thus, the nominative–dative case-marking patterns of nominalized and some subordinate clauses in southern and northern languages generalized to all transitive-clause types resulting in a consistent nominative–'accusative' pattern. Perhaps more subordinate-clause types came to be used as independent clauses (through a pragmatically motivated process of 'insubordination': the shifting of clauses from dependent to independent status) taking their case-marking patterns with them into the declarative domain. Perhaps a range of declarative clauses came to be used as subordinate clauses (thus expanding the range of relative clause TAM categories) and through conforming with subordinate case-normalization patterns provided the basis for an analogical extension of these patterns back into main-clause functions. The change may have occurred by either one or other, or a combination, of these shifts.

The proof of the hypothesis lies, once again, in a detailed reconstruction of verbal inflections across the languages of the area in an attempt to identify which may have shifted their dependency status: however, this detailed work has not yet been completed. Nevertheless, there is some circumstantial evidence supporting the hypothesis. As Table 8 partly shows, Northern Pilbara and Southern Pilbara languages make a clear distinction between main and subordinate-clause types. Subordinate clauses typically involve specific verbal inflections, may involve a distinct case-marking pattern, and their subordinate status is also usually indicated by a complementizing suffix attached to the subordinate verb (see Dench and Evans 1988). In Nyamal and to a lesser extent in Ngarla the verb in a declarative is marked in agreement with the person and number of the subject. This pattern does not extend to subordinate-clause types.

By contrast, there are relatively few distinct main-clause versus subordinate-clause inflections in any of the accusative languages; dependency status is clear only from the presence or absence of the complementizing case suffix on the subordinate verb (and in some languages also on the arguments of this verb).

Martuthunira is a case in point. There are just two inflections which quite clearly can occur only in independent clauses (the imperative and the present tense) and just two which must occur in dependent clauses (the ‘contemporaneous relative’ and the morphologically related ‘sequential relative’). The latter inflections must have the same subject as their controlling clause and hence the complementizer is zero (in agreement with the nominative case of the matrix clause subject). Thus despite their dependency status, there is little formal indication, other than this syntactic constraint, of their dependency status. Ten other verbal inflections have both dependent and independent uses, some more marked than others. Similar patterns hold in other accusative languages.

Detailed reconstruction may ultimately reveal the directions of shift for each inflection—from dependent-clause to increased independent-clause use, or vice versa. For the time being, all that is clear is that the problem is not a simple one. There is little clear cognacy amongst verbal inflections across the Pilbara area and so no immediately obvious solution to the problem is available. What is more, we might expect that since the shift in alignment involves the extension of a pre-existing case-marking pattern into a range of new domains, the analogical changes may have been gradual and cumulative. Some initial shifts may have occurred in a common ancestor, but subsequent steps might have continued independently following the break up of this putative common ancestor into a set of distinct language communities. It is also quite possible that an analogical change conceived in just one language provided a model for similar restructuring in its neighbours. This question may ultimately turn out to be undecidable.

The development of the passive voice may be a little more concrete than the general alignment shift and so more amenable to a reconstruction producing results useful for subgrouping. The development involves at least two steps. First, an ‘inflectional’ passive has arisen through the reinterpretation of old ergative clause patterns (in which the subject was ergative and the nominal object was unmarked) as passives (with an unmarked nominative subject and an oblique agent bearing the reflex of the ergative). The passive was thus restricted to clauses bearing particular verbal inflections. Subsequently, a ‘derivational’ passive arose through the extension of a patient oriented inchoative stem-forming suffix, **-nguli*, to transitive verbs. These verbs then select regular active inflections and have a case-frame equivalent to the inflectional passives (Dench 1982). While all of the accusative languages use the same derivational passive suffix—some reflex of **-nguli*—there is a variety of inflectional passive suffixes. Table 9 gives the forms of the inflectional passive perfectives across the set of accusative languages together with apparently related active forms. For each suffix there are two forms, determined by verb conjugation.⁵

⁵ Most of the accusative languages also have passive inflections used to mark what is variously described as admonitive, apprehensional, or ‘lest’ clauses. However, there is little similarity amongst the forms used to mark this category across the different languages.

TABLE 9. Passive perfective inflections in the accusative languages

	passive perfective	related forms	category
Yinhawangka	<i>-jangujangu/-rnujangu</i>	<i>-jangu/-rnu(jangu)</i>	(perfective) relative
Panyjima	<i>-jangaanu/-rnaanu</i>	<i>-jangu/-rnu</i>	imperfective relative
Yindji/Kurrama	<i>-yangaanu/-rnaanu</i>	<i>-(ya)ngu/-rnu</i>	imperfective (relative)
Ngarluma	<i>-nhakurla/-rnakurla</i>	<i>-nha/-rna</i>	past
Martuthunira	<i>-yangu/-rnu</i>	—	

Each of the passive inflections in Yinhawangka, Panyjima, Yindjibarndi, and Kurrama involves an increment to the active relative inflection, **-jangu/-rnu*. In Yinhawangka this appears to involve the further addition of the *-jangu* relative suffix, though it is fair to say that with an as yet very limited corpus, the use of these suffixes is not well understood. For the other three languages, it seems likely that the increment may descend from an ablative form, **-janu*, found in Western Desert dialects to the east of the Central Pilbara group. The formation (both synchronic and diachronic) of a perfective relative clause with the ablative suffix is common to languages in the area more generally. In Ngarluma, the passive perfective also involves an increment, though this time to the active past tense. The source of the increment is not known. In each of these cases, then, the passive inflection involves an increment to an active inflection, though the increment itself does not appear to be inherently passive. By contrast, the passive perfective in Martuthunira is identical to the imperfective relative inflection in Panyjima, Yindjibarndi, and Kurrama and yet here there is no increment. What is an active inflection in one language is strictly passive in another.

With Ngarluma as the exception, there is a clear relationship between the passive perfective and a general relative clause marker with forms **-jangu/-rnu*. It might be suggested that this inflection originally had both active and passive-like functions. That is, it was involved in the construction of both subject and object relative clauses (and possibly perfective passive nominalizations), and it was the object relative and nominalizing functions which allowed the development of the passive. But the forms in Table 9 suggest that this development occurred independently in the different languages. The Panyjima, Yindjibarndi, and Kurrama forms point to a common origin for the passive, at least in these three languages. But there is a problem even here. The long vowel in the Panyjima passive perfective suffixes is very unusual—it represents the only instance of an non-initial long vowel in the language—suggesting that the form may in fact be calqued from neighbouring Yindjibarndi and Kurrama. The differences amongst the forms in the other languages—Yinhawangka, Ngarluma, and Martuthunira—suggest the independent development of the passive in each of these.

If the inflectional passives are independent developments, some patterned on innovated constructions in neighbouring languages perhaps, then the derivational

passive is possibly an even more direct diffusional phenomenon. As noted above, the same morpheme is used in all of the accusative languages, but is also present in an inchoative function shared also with the non-accusative Southern Pilbara languages. The same shift in function appears to have taken place in each of the accusative languages, but we cannot be sure if this involves a pattern of diffusion or indeed the simple borrowing of the derivational passive morpheme along with its inchoative functions from some original innovator.

5. Conclusion

In this chapter, I have given a brief overview comparison of a number of features of the grammatical systems of the languages of the Pilbara region, as an Australian case study. I have chosen to consider features which, given their accessibility and our current state of knowledge of these languages, might best reveal some evidence for grouping the languages, either genetically or areally. My main concern has been to consider whether having identified shared innovations, these innovations can be claimed to have arisen in a single common ancestor or whether they may have diffused from one language into another, either directly (through the borrowing of forms) or indirectly (through the borrowing of patterns).

None of the shared innovations described in this chapter can be considered, conclusively, to be innovations arising in a single ancestor. For each change which appears to allow a single reconstruction, we find a pattern in a neighbouring language which parallels that change and which, since it involves distinct forms, must have arisen independently. This raises the suspicion that our set of languages sharing both form and pattern might have as easily arrived at this similarity through contact rather than through shared inheritance. I have reached similar conclusions elsewhere in a detailed study of the pronoun paradigms of the Pilbara languages (Dench 1994) and have made similar suggestions with regard to some aspects of verb morphology (see Dench 1996, 1998b).

If given the tendency towards indirect diffusion we may be unable to arrive at a clear subgrouping for the languages of the Pilbara, might we instead look for evidence of one or more linguistic areas? If any case is to be made then it is perhaps that the set of accusative Central Pilbara languages comprise a linguistic area. However, aside from the collection of features which coincide with the alignment shift, there is little that brings them together, uniquely, as a group. The set of morphophonemic alternations described in §3, specifically the pattern of length-conditioned allomorphy extended to reflexes of the old dative, and the nasal dissimilation rule for the locative and ergative, includes the non-accusative Northern Pilbara languages with the Central Pilbara group, but also excludes the accusative language Martuthunira. Similarly, the various innovative phonological patterns group some of the Central Pilbara languages with Southern Pilbara languages, but exclude others. On balance, there is little evidence from this study to suggest clearly defined linguistic areas within the region. Instead, we find that

different features have different ranges. Some of these do coincide, but perhaps more by virtue of being in related grammatical subsystems than because speakers of the different languages form a defined areal community. The lack of clear linguistic areas parallels the lack of distinct cultural blocks.

It remains to be seen whether further detailed work in the historical comparison of the Pilbara languages will allow any clearer evidence of genetic or areal groupings to emerge. And it also remains to be seen how the Pilbara language situation might compare with that of other specific regions within Australia—there has as yet been very little detailed historical comparison of Australian languages at the lowest level. It is worth emphasizing that the complexity of the Pilbara situation, as demonstrated in this chapter, has become increasingly clear the more detailed our analyses and comparisons of the languages have become. And it would not be surprising to find detailed comparative studies of languages in other parts of Australia showing similar patterns.

The case study presented here reinforces Dixon's more general description of the Australian situation, presented in the preceding chapter, as one in which determining deep genetic relationships is especially difficult and may ultimately prove to be impossible. Indeed this case study suggests that the problem is even more complex. Where Dixon states that we *are* able to recognize a number of low-level subgroups yet cannot build higher level genetic groupings, this study shows that even building low level subgroups may be impossible in some cases.

What is it that makes the Australian linguistic situation so complex in this respect? Dixon hypothesises that the situation is a result of history—specifically a very long period of equilibrium on the Australian continent. This explains why low-level groups may be identifiable (as recent low-level punctuations) while deep genetic connections cannot be determined. But it does not explain the micro-level complexity of something like the Pilbara situation. Given the time depths for which Dixon's (1997) punctuated-equilibrium model is posited, the linguistic variation in the Pilbara might be seen as 'noise' against a general equilibrium-state background. But if the indeterminacy we find at this low level is equivalent in kind to the larger indeterminacy, an explanation for the Australian situation may lie elsewhere; it may lie at the micro-level.

I noted in §1.2 that barriers to diffusion may be typological, geographical, or social and that the Pilbara situation appears to offer very few social barriers (and there are few typological or geographical barriers either). To use Ross's typology (Chapter 6 of this volume, and references therein), these are open and loose-knit communities and so the possibilities for diffusion are high. Simplistically, why then is there not a single language spoken across the whole region? Clearly, the different language communities choose to remain distinct to some degree and choose to speak different languages. This level of commitment does not extend to the point that they choose to share nothing with their neighbours. Their unique identity may be a unique combination of linguistic features many of which are nevertheless shared with one or more neighbours.

Explanations for particular language-contact situations, like that in the Pilbara, must ultimately be couched in socio-historical terms (Thomason and Kaufman 1988, Ross, this volume). However, given the levels of destruction of traditional Australian patterns of community interaction following European settlement, the extinction of languages and of whole peoples, we simply do not have the information necessary to do this. Ironically, it may be that the closest we can come is to reconstruct aspects of the original social situation using the linguistic data itself. As we continue to develop models of how different sociolinguistic settings may result in different kinds of language contact and contact-induced change, of what may or may not diffuse in different contact situations, and based on case studies of diverse language communities around the world, we may be in a position to make more sense of the complex interacting patterns of descent and diffusion in cases like that described here for the Pilbara.

References

- AUSTIN, P. 1981a. 'Case-marking in Southern Pilbara languages', *AJL* 1.211–26.
 — 1981b. 'Proto-Kanyara and proto-Mantharta historical phonology', *Lingua* 54.295–333.
 — 1982. 'Southern Pilbara Phonology'. Manuscript, La Trobe University.
 — 1988. 'Classification of Southern Pilbara Languages', pp. 1–17 of *Papers in Australian Linguistics*, vol. 17. Canberra: Pacific Linguistics.
 — 1994a. 'A Reference Grammar of the Mantharta Languages, Western Australia'. Manuscript, La Trobe University.
 — 1994b. 'A Reference Grammar of the Kanyara Languages, Western Australia'. Manuscript, La Trobe University.
 — 1996. 'The Kanyara and Mantharta languages'. Paper presented at the *Where do Western Desert Languages Come From?* workshop, 3rd Australian Linguistics Institute, Canberra.
 BEARD, J. 1990. *Plant Life of Western Australia*. Kangaroo Press: Kenthurst.
 BLEVINS, J., and MARMION, D. 1996. 'Is Nhanda a Kartu language'. Paper presented at the *Where do Western Desert Languages Come From?* workshop, 3rd Australian Linguistics Institute, Canberra.
 DENCH, A. 1982. 'The development of an accusative case marking pattern in the Ngayarda languages of Western Australia', *AJL* 2.43–59.
 — 1991. 'Panjyima', pp. 124–243 of Dixon and Blake (1991).
 — 1994. 'The historical development of pronoun paradigms in the Pilbara region of Western Australia', *AJL* 14.155–91.
 — 1995. *Martuthunira: A Language of the Pilbara Region of Western Australia*. Canberra: Pacific Linguistics.
 — 1996. 'Pilbara verbal morphology and the Western Desert: some first steps towards a comparative reconstruction'. Paper presented at the *Where do Western Desert Languages Come From?* workshop, 3rd Australian Linguistics Institute, Canberra.
 — 1998a. 'The *dative split in Martuthunira', *Proceedings of the Australian Linguistics Society Meeting 1998*. <http://www.cltr.uq.edu.au/als98/>
 — 1998b. 'What is a Ngayarta language? A reply to O'Grady and Laughren', *AJL* 18.91–107.

- 1998c. *Yingkarta*. Munich: Lincom Europa.
- and EVANS, N. 1988. 'Multiple case-marking in Australian languages', *AJL* 8.1–47.
- DIXON, R. M. W. 1976. Editor of *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies.
- 1980. *The Languages of Australia*. Cambridge: Cambridge University Press.
- 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- and BLAKE, B. 1981. Editors of *The Handbook of Australian Languages*, vol. 2. Canberra: Australian National University Press.
- — 1991. Editors of *The Handbook of Australian Languages*, vol. 4. Oxford: Oxford University Press.
- DOUGLAS, W. 1981. 'Watjarri', pp. 196–272 of Dixon and Blake (1981).
- HALE, K. L. 1976. 'On ergative and locative suffixal alternations in Australian languages', pp. 414–17 of Dixon (1976).
- KOHN, A. 1994. 'A Morphological Description of Ngarluma'. BA thesis, University of Western Australia.
- 1996. 'Case-marking and agreement in Nyiyaparli, an Eastern Pilbara language'. Handout prepared for linguistics seminar, University of Western Australia.
- MARMION, D. 1996. 'A Description of the Morphology of Wajarri'. BA thesis, University of New England.
- O'GRADY, G. N. 1958. 'The Significance of the Circumcision Boundary in Western Australia'. BA thesis, Sydney University.
- 1966. 'Proto-Ngayarda phonology', *OL* 5.71–130.
- and LAUGHREN, M. 1997. 'Palyku is a Ngayarta language', *AJL* 17.129–54.
- VOEGELIN, C. F. and VOEGELIN, F. M. 1966. *Languages of the World: Indo-Pacific, Fascicle 6* (*Anthropological Linguistics* 8, ii).
- RADCLIFFE-BROWN, A. 1913. 'Three Tribes of Western Australia', *Journal of the Royal Anthropological Institute* 43.141–95.
- SANDS, K. 1996. *The Ergative in Proto-Australian*. Munich: Lincom Europa.
- SCHEFFLER, H. 1978. *Australian Kin Classification*. Cambridge: Cambridge University Press.
- SHARP, J. 1998. 'A Grammar of the Nyangumarta Language of the Pilbara'. Ph.D. dissertation, University of Western Australia.
- SILVERSTEIN, M. 1976. 'Hierarchy of features and ergativity', pp. 112–71 of Dixon (1976).
- SIMPSON, J. 1983. 'Ngarluma as a W* language'. Manuscript, MIT.
- THOMASON, S. G. and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- TINDALE, N. 1974. *Aboriginal Tribes of Australia*. Canberra: Australian National University Press.
- WORDICK, F. 1982. *The Yindjibarndi Language*. Canberra: Pacific Linguistics.

Contact-Induced Change in Oceanic Languages in North-West Melanesia

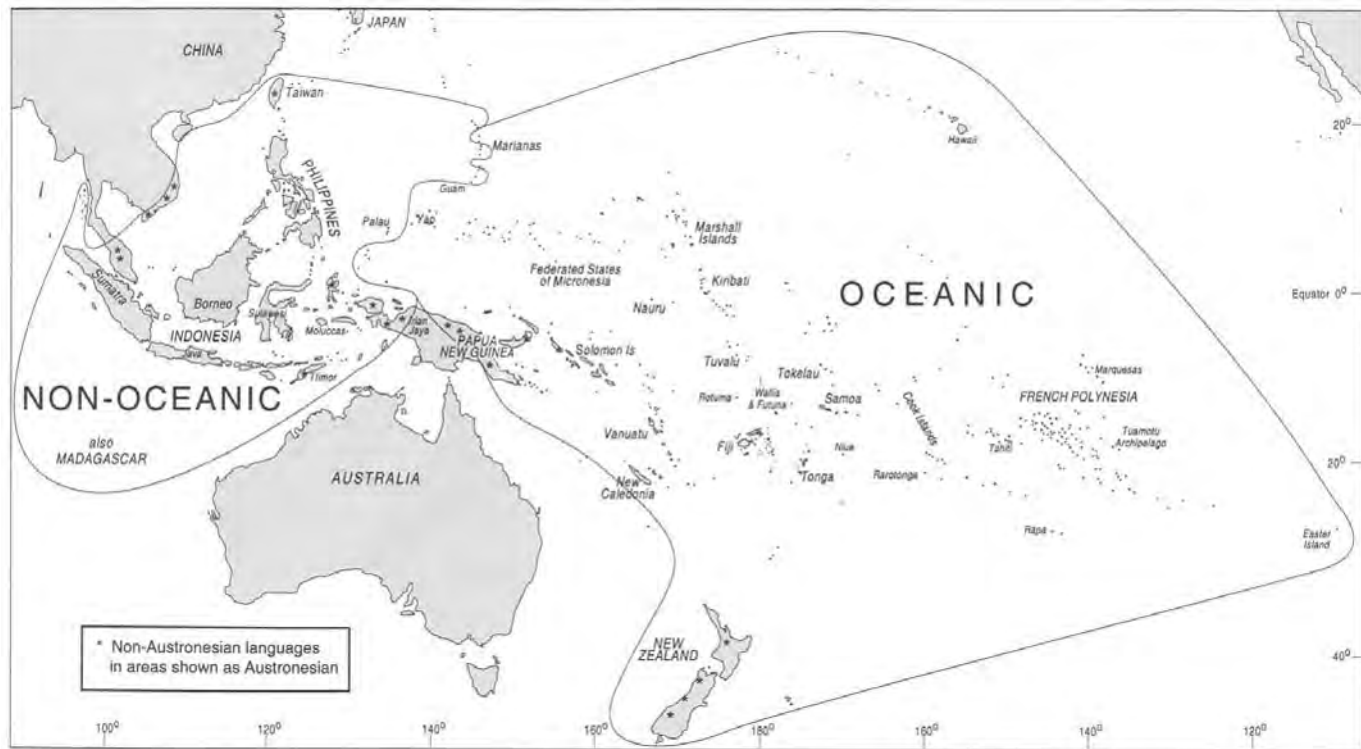
Malcolm Ross

1. Background

The Oceanic language family forms a subgroup within the larger Austronesian family. The Austronesian family, with perhaps as many as 1,500 languages, embraces the large portion of the world shown in Map 1 and has been formally recognized by scholars since the 1840s (Ross 1996d). The Oceanic subgroup of Austronesian was first recognized by Dempwolff (1927, 1937), although the exact definition of its western boundary (Map 1) awaited the work of Grace (1971), Blust (1978), and Ross (1996b).

In any discussion of contact-induced change, we need to distinguish between those features that are shared between languages as the result of contact and those shared by inheritance, and so it is important to describe briefly the criteria by which both Austronesian and its Oceanic subgroup are recognized as genetic groupings. The integrity of the Austronesian family is supported by a large quantity of lexical cognates with regular sound correspondences. Otto Dempwolff's (1934) first major listing of these has since been expanded enormously, especially in the work of Robert Blust (1980, 1983–4, 1986, 1989, 1995). There is considerable typological variation within Austronesian, but there is a small but significant collection of derivational morphemes that are reflected in all major branches of the family.

Dempwolff recognized the Oceanic subgroup of Austronesian on the basis of a set of phonological innovations shared by languages across a huge area embracing most of Melanesia and Micronesia and all of Polynesia. Whilst scholars have made a few modifications and additions to the innovations he originally listed (Ross (1995) provides a summary), his Oceanic hypothesis stands unchallenged. It is supported by an even larger quantity of lexical cognates with regular sound correspondences than the Austronesian hypothesis, as well as a larger quantity of bound morphemes, both derivational and inflectional. Typological variation within Oceanic is arguably less than in Austronesian as a whole, but there is still



MAP 1. Oceanic and non-Oceanic Austronesian languages

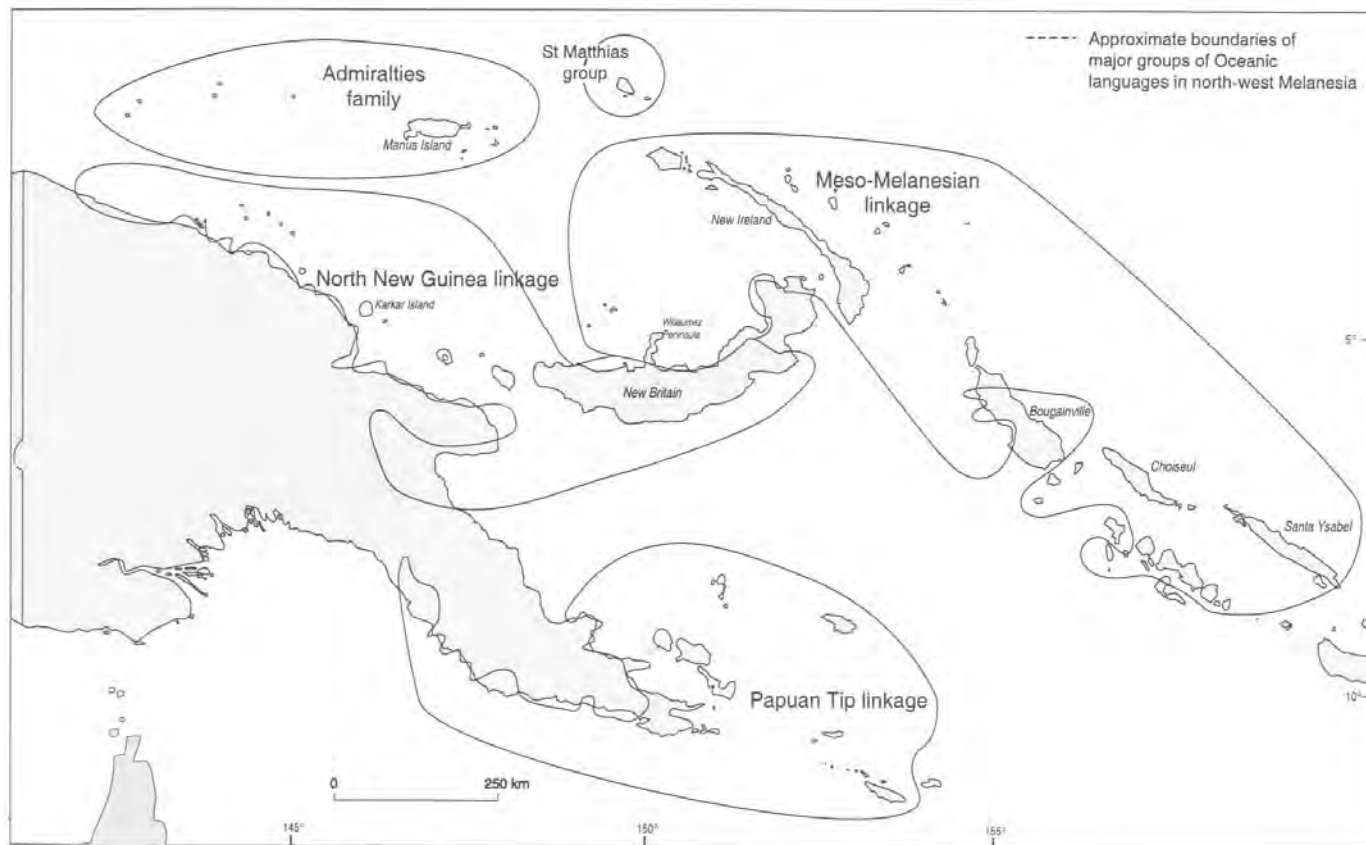
considerable variation in constituent order, as well as a number of ergative islands in an ocean of accusativity.

Such are the innovations that define Oceanic that we can be reasonably confident that its languages are descended from a single reconstructable interstage, Proto-Oceanic, spoken in the Bismarck Archipelago of north-west Melanesia¹ about 3,500 years ago (this date is based on a correlation of linguistic and archaeological evidence; Pawley and Ross 1993, 1995, Ross 1995). I say this because, in the literature from Ray (1926) to Capell (1976) there are suggestions that Oceanic languages are due to multiple incursions of Austronesian speakers into north-west Melanesia. This theory was most clearly articulated by Capell (1943), and he intends us to believe that different Oceanic languages in south-east Papua are descended from different Austronesian incursions. However, the innovations shared by all Oceanic languages and the additional innovations shared by the languages of south-east Papua (Ross 1992) make nonsense of this theory.

In order to understand the sociolinguistic situation in northwest Melanesia both in Proto-Oceanic times and now, however, it is necessary to describe something of what must have happened *before* the genesis of Proto-Oceanic. As Austronesian speakers first spread eastwards across Indonesia, perhaps sometime in the third millennium BC, they seem to have encountered little resistance in their search for agricultural land until they reached the island of New Guinea (Pawley and Ross 1993). Here they found a land of steep mountains peopled by speakers of the precursors of today's Papuan languages, many of whom practised—and had practised for millennia—an agriculture based on taro. The term 'Papuan', incidentally, does not denote a single genetic grouping of languages, but rather a number of genetic groups characterized by the fact that they are spoken in north-west Melanesia and are not Austronesian. The linguistic evidence suggests that Austronesian speakers gained some footing in the north-west of what is now Irian Jaya around the edges of the Bird's Head and Cenderawasih Bay and on offshore islands, and that speakers of a precursor of Proto-Oceanic must have reached New Britain from somewhere in this area at a date sometime in the first half of the second millennium BC, voyaging east along the north coast of New Guinea (Ross 1988: 19–21). It seems probable that there was initially very little Austronesian settlement on mainland New Guinea, and even today, as Map 2 shows, Oceanic speakers occupy only relatively small areas of it. The rest is peopled by Papuan speakers.

Instead, the original Austronesian-speaking immigrants found land on the less densely populated Bismarcks, where they evidently practised trade, among other things. It seems a reasonable inference that the innovations which turned their speech into Proto-Oceanic came about through their contact with Papuan speakers, although the evidence for this is circumstantial. It is very likely, for example,

¹ The Bismarck Archipelago comprises New Britain, New Ireland, the St Matthias Group, and the Admiralty Islands, whilst north-west Melanesia is roughly the area shown in Map 2.



MAP 2. Groups of Oceanic languages in north-west Melanesia

that the Proto-Oceanic term **m^wapo(q)* ‘taro’ (i.e. the genera of the *Araceae* family) was borrowed from a Papuan language (Ross 1996c). It is also probable that the formal distinction between alienable and inalienable possession entered Proto-Oceanic or an immediate precursor through Papuan contact. Within a period of two or three centuries, Oceanic speakers had peopled the Pacific from New Britain to Samoa. Almost all the Pacific islands beyond north-west Melanesia had previously been uninhabited.

The spread of Oceanic speakers within north-west Melanesia itself, and especially onto mainland New Guinea, seems to have been a much slower process (we do not yet have enough archaeological evidence to be certain), but not necessarily one which entailed much conflict. In general, Papuan speakers seem to have lived inland, away from the malaria-ridden coasts, pursuing a mixture of agriculture and hunting and gathering. Oceanic speakers possessed ocean-going and often potting technology, and were traders. They evidently also practised agriculture, as well as fishing and gathering reef products. Only in a few low-lying areas did they penetrate inland. Dutton, writing about the Port Moresby area of central Papua, suggests that the arrival of Oceanic speakers with their new technology drew Papuan speakers down to the coast, where a symbiotic relationship developed between the two groups (1994, 1971). Further east on the same coast, he has described Magori, a dying Oceanic language containing a heavy Papuan admixture (Dutton 1976, 1982). Lynch (1981) remarks that the verb-final and postpositional structures of the Papuan Tip cluster of Western Oceanic languages (Map 2) are best attributed to contact with Papuan languages. Thurston (1982, 1987, 1989, 1994) has described the outcomes of contact between several Oceanic languages and Anêm, a Papuan language of north-western New Britain. In my own work I have described the phonological effects of an inferred language shift from Papuan to Oceanic on New Ireland (Ross 1994a), the genealogy of the so-called ‘mixed’ language Maisin in eastern Papua (Ross 1996a), and the Papuanization of Takia (Ross 1987, 1994b, 1996a).

These case studies, quite widely distributed within north-west Melanesia, all point to what Dixon (1997: 68–73) would call a period of equilibrium: ‘During a period of equilibrium, languages in contact will diffuse features between each other, becoming more and more similar’ (70–1). However, where Dixon’s canvas is vast—he speaks of ‘fifty languages’ and of ‘millennia’—mine is tiny, depicting what happens to a single language during an equilibrium period.

Most of my examples are drawn from Takia, the Oceanic language spoken on Karkar Island, a volcanic island off the north coast of Papua New Guinea (Map 2). Karkar has a population of about 45,000, just over half of whom speak Takia. The remainder speak the Papuan language Waskia. Takia speakers occupy the southern half of the island, Waskia the northern half. Despite the striking linguistic difference between them, however, the two groups display no other discernible cultural differences (McSwain 1977), an outcome which is not surprising after an equilibrium period (Dixon 1997: 70).

2. Equilibrium under the microscope

There are seven points that I would like to make about contact-induced change viewed from the perspective of a single language, but I will devote the lion's share of my space to the first point:

- (a) The 'syntactic borrowing' which occurs as a result of contact is part of a larger process whereby semantic structures are also 'borrowed'.
- (b) Lexical borrowing is independent of 'syntactic borrowing'.
- (c) Syntax and phonology may behave quite differently under contact conditions.
- (d) If 'converge' means 'change to become more like each other', then languages do not usually converge. Instead, one language becomes more like a second, while the second may be relatively unaffected by the contact.
- (e) Processes of change which Dixon (1997) associates with equilibrium and with punctuation are not necessarily mutually exclusive.
- (f) Within a given linguistic area, changes which are similar not only syntactically but also formally may occur independently in related languages, complicating the reconstruction of linguistic prehistory.
- (g) The kind of change described in this chapter is only one member of a tentative paradigm of changes which can be exemplified in northwest Melanesia.

2.1. 'METATYPY' OR 'SEMANTICO-SYNTACTIC BORROWING'

This clumsy subtitle is intended to draw attention to a problem: what is commonly labelled 'syntactic borrowing', for example by Harris and Campbell (1995: 120–50), is in fact a larger process than this label suggests, and one for which so far there is no generally accepted label, in spite of the fact that the process itself is well described by Grace (1981).

We can gain some insight into this larger process by examining what has happened in Takia, apparently as the result of its contact with Waskia.² I write 'apparently', because some of Takia's closest relatives on the mainland have undergone similar structural changes to Takia, implying that some of the contact-induced change in Takia may date from a period before Takia split from its mainland relatives.

Takia is obviously Oceanic in its lexicon and in its bound morphology, and these show regular sound correspondences with cognates in other Oceanic languages. But any linguist familiar with Oceanic languages spoken across Melanesia will be struck by the fact that Takia is very un-Oceanic in its syntactic structure. In this it more closely resembles its Papuan neighbour Waskia. The examples in (1) illustrate these points.

² Published sources containing Takia material are Ross (Ross, 1987, 1994b, 2002). I have also drawn on Waters, Tuominen, and Rehburg (1993) and on my own fieldnotes. Published sources for Waskia are Ross with Paol (1978) and Barker and Lee (1985).

- (1) Takia: *tamol* *tubun* *uraru* *en*
Waskia: *kadi* *bi-biga* *itelala* *pamu*
man (PL-) big two this
'these two big men'
- Takia: *Waskia tamol an*
Waskia: *Waskia kadi mu*
Waskia man DET
'the Waskia man'
- Takia: *Kai sa-n ab*
Waskia: *Kai ko kawam*
Kai CLASSIFIER-his house
POSTPOSITION
'Kai's house'
- Takia: *ŋai tamol an ida*
Waskia: *ane kadi mu ili*
I man DET with.him
'the man and I'
- Takia: *tamol an ŋai i-fun-ag=da*
man DET me he-hit-me=IMPFV
Waskia: *kadi mu aga umo-so*
man DET me hit-PRES.he
'The man is hitting me.'

Table 1 gives a typological comparison between Proto-Western Oceanic, Takia, and Waskia, using shading to show whether Takia correlates typologically with Proto-Western Oceanic or with Waskia (or, in the case of adjective syntax, with both). Proto-Western Oceanic is the language ancestral to Takia and all other Oceanic languages of the western Solomon and Papua New Guinea, except for the Admiralties.³ Lighter shading shows an incomplete shift by Takia towards Waskia, i.e. the acquisition by Takia of enclitics corresponding to functions of the Waskia portmanteau verbal suffix. Table 1 shows that Takia largely follows Proto-Western Oceanic in features that relate to bound morphology, that is, to word structure, but it follows Waskia in matters of phrasal and clausal syntax. Thus in the last example in (1), the Takia verb *ifunagda* is characteristically Oceanic Austronesian in having a preposed subject marker (glossed as 'he') and a postposed object marker (glossed as 'me'), whilst the Waskia verb is typically Papuan with its postposed portmanteau marking of tense/aspect and subject. However, the Takia clause follows the syntax of Waskia. The clause as a whole is

³ It could be argued that, since the Western Oceanic languages probably result from increasing differentiation within a dialect network, there never was a unitary Proto-Western Oceanic. However, the degree of early dialectal differentiation within it was not such as to undermine the reconstructions or arguments presented here.

TABLE 1. A typological comparison of Proto-Western Oceanic, Takia, and Waskia

	Proto-Western Oceanic [PWOc]	Takia (Oceanic Austronesian)	Waskia (Papuan)
<i>Unmarked clause order</i>	SVO	SOV	SOV
<i>Noun phrase</i>			
non-deictic determiner	preposed article distinguishing common from personal	postposed determiner	postposed determiner
adjective syntax	postposed	postposed	postposed
adjective agreement	ADJ-PRON, where PRON = possessive pronoun agreeing in person and number with noun	as PWOc	none
attributive noun	postposed	preposed	preposed
<i>Possession system</i>			
possessor NP	alienable vs. inalienable with subcategories of alienable	alienable vs. inalienable	alienable vs. inalienable
possessor pronoun: with inalienables	postposed	preposed	preposed
possessor pronoun: with inalienables	suffixes to possessed noun	suffixes to possessed noun	prefixed or infixes to possessed noun
	suffixes to classifier	suffixes to classifier	independent, preposed
<i>Verb complex</i>			
subject referencing pronoun tense/aspect/mood	prefix or proclitic	prefix or proclitic	
	prefix or proclitic; reduplication for continuative	enclitic	portmanteau suffix
object referencing pronoun	suffix or enclitic	suffix	independent
<i>Pronoun system</i>	inclusive/exclusive distinction	inclusive/exclusive distinction	no inclusive/exclusive distinction
<i>Adpositional phrases</i>	prepositional	postpositional	postpositional
<i>Clause linkage</i>	coordinate, subordinate	coordinate, cosubordinate, subordinate	coordinate, cosubordinate, subordinate
clause-linking devices	parataxis, conjunctions	enclitic to verb	part of portmanteau suffix to verb

AOV, rather than Proto-Western Oceanic AVO. The characteristic Western Oceanic shape of this clause is captured in the rough Proto-Western Oceanic reconstruction in (2)).

- (2) **a tam^wata i-punu-punuqi-au*
 DET man he-CONTIN-hit-me
 'The man is hitting me.'

The tense/aspect of the Takia clause is expressed by an enclitic (= *da* IMPERFECTIVE) to the verbal complex, apparently following the Waskia model of expressing tense/aspect within a portmanteau suffix on the verbal complex. A typical Western Oceanic language would express tense/aspect either by a particle inserted between the subject noun phrase and the verb, or, in this case (to form the continuative), by reduplicating the verb stem. The Takia phrase *tamol an* 'the man' follows the pattern of Waskia *kadi mu*.

What is noteworthy is that although Takia syntax follows Waskia rather closely, it uses inherited Western Oceanic forms. Sometimes the Papuanized syntax of Takia has apparently been achieved by simply altering the sequence of elements. But more often it has been effected by changing the function of an element that happens to appear in the right position by Waskia standards. For example, Takia did not achieve the noun phrase sequence of HEAD NOUN + DETERMINER by simply reversing the sequence of Proto-Western Oceanic **a tam^wata*. What happened was more subtle. Proto-Western Oceanic had a set of three deictic morphemes, shown in (3):

- (3) **i~e* 'this, near speaker'
 **a* 'that, near hearer'
 **o* 'that, near neither speaker nor hearer'

When one of these was used attributively, it followed the Proto-Western Oceanic adjective pattern (as indicated in Table 1), taking a pronominal suffix agreeing in person and number with the head noun, so that 'that man' was expressed as in (4):

- (4) **a tam^wata a-ñā*
 DET man that-3sg
 'that man'

The steps which separate this from Takia *tamol an* 'the man' are three: loss of the deictic force of **a-ñā*, deletion of the preposed article **a*, and a few phonological changes.

Apparently by a similar process, Takia has developed a set of postpositions where Western Oceanic languages commonly have prepositions. Takia and Waskia postpositions are shown in (5):

(5)	Takia	Waskia
location	<i>na, te</i>	<i>se, te</i>
location 'in'	<i>lo</i>	<i>i, nuŋi</i>
location 'on'	<i>fo, fufo</i>	<i>kuali</i>
ablative	—	<i>ko</i>
instrument	<i>nam</i> (= <i>na-mi</i>)	<i>se</i>
referential	<i>o</i>	<i>ko</i>
manner	<i>mi</i>	<i>wam</i>

With the possible exception of *te*, no Takia postposition is borrowed from Waskia. At least two, *lo* 'in' and *fo/fufo* 'on', are derived from inalienably possessed Proto-Western Oceanic relational nouns.

A possessive noun phrase with an inalienable head had possessed–possessor order in Proto-Western Oceanic, the head bearing a suffix coreferencing the person and number of the possessor, as in (6):

(6)	<i>*mata-ña</i>	<i>a</i>	<i>boRok</i>
	eye-its	DET	pig
	'the eye of a/the pig'		

In Proto-Western Oceanic, locations were often expressed by a prepositional phrase, and the governee noun phrase often had the structure of (6) with a relational noun as its head:

(7)	<i>*i</i>	<i>lalo-ña</i>	<i>a</i>	<i>Rumaq</i>	<i>*i</i>	<i>papo-ña</i>	<i>a</i>	<i>Rumaq</i>
	PREP	inside-its	ART	house	PREP	top-its	ART	house
	'inside the house'				'on top of the house'			

The head of **lalo-ña a Rumaq* '(the) inside of the house' is **lalo-ña* 'its inside', and the possessor is **a Rumaq* 'the house'. This eventually became the Takia structure in (8):

(8)	<i>ab</i>	<i>lo</i>	<i>ab</i>	<i>[fu]fo</i>
	house	in	house	on
	'in the house'		'on top of the house'	

The developments which led from (7) to (8) were (a) loss of the preposed article and preposing of the possessor, giving **i Rumaq lalo-ña* or, more probably, **Rumaq i lalo-ña*; (b) loss of the preposition and concomitant grammaticization of the relational noun as a postposition (Ross 1996a: 188–90). Presumably, (b) occurred as part of the restructuring of the clause from AVO to AOV. The reduction of **lalo-* to *lo* and optionally of **papo-* to *fo* is typical of the phonological erosion which accompanies grammaticization.

Another application of Oceanic forms to serve 'Papuan' purposes is the Takia use of predicate enclitics which reflect Proto-Western Oceanic conjunctions. As noted in Table 1, Takia shares with Waskia a category of clause linkage not usually

found in Oceanic languages but common in Papuan languages of the New Guinea linguistic area. This is the category which Foley and van Valin (1984) call 'cosubordination', more recently labelled 'coordinate dependence' by Foley (1986). In this kind of linkage, only the verb of the last clause in a chain of otherwise coordinate clauses is a fully finite verb. Chain-medial verbs are morphologically encoded for less information and are dependent on (cosubordinate with) the chain-final verb for the full specification of tense, aspect, mood, and sometimes subject coreference. Every cosubordinate clause in Takia ends with a predicate enclitic, usually *-go/-g* 'realis' or *-pe/-p* 'irrealis', derived respectively from Proto-Western Oceanic **ga* 'realis conjunction' and **be* 'irrealis conjunction' (Ross 1987).

However, an account of contact-induced change in Takia is incomplete if it deals only with syntax. Indeed, the syntactic change is simply part of a more profound restructuring of the language as Takia speakers have increasingly come to construe the world around them in the same way as the Waskia. This has entailed restructuring the *semantic* organization of Takia on the Waskia model, so that equivalent lexical items in Takia and Waskia have the same range of meaning, closed sets of morphemes have similar membership and semantic structure, and complex lexical items, whether compound words, phrases, or larger formulae have been reformulated so that their component morphemes are the same as their Waskia equivalents.

For example, where Western Oceanic languages commonly have perhaps two or three prepositions, and express more complex relationships either with relational nouns or by serialized verbs, Takia has developed a set of postpositions that is remarkably similar semantically (as well as syntactically) to the Waskia set. These were listed in (5).

Somewhat similarly, where Proto-Western Oceanic distinguished at least two categories of alienable possession (consumable and neutral), Takia has followed Waskia in reducing the two categories to one. The morphemes representing the two categories are still reflected in Takia, but there is no longer any semantic distinction between them in most dialects, and they are more or less interchangeable. Curiously, however, Takia has retained the distinction between inclusive and exclusive in first person plural pronouns, although Waskia lacks it.

Examples of parallel compounds are given in (9):

(9)	'literal' meaning	Takia	Waskia
'person'	'man-woman'	<i>tamol-pein</i>	<i>kadi-imet</i>
'animal'	'pig-dog'	<i>bor-goun</i>	<i>buruk-kasik</i>
'his parents'	'his mother-his father'	<i>tinan-taman</i>	<i>niam-niet</i>
'(do) first'	'his eye-his eye'	<i>malan-malan</i>	<i>motam-motam</i>

The Proto-Western Oceanic lexeme for 'person' was **tau*, but this has been replaced in Takia by *tamol-pein*, following the Waskia *pattern* but derived from the Takia words reflecting Proto-Western Oceanic **tam*^w*ata* 'man' and **papine* 'woman'.

Some examples of parallel formulae are listed in (10):

(10)		'literal' meaning			
	'the palm of my hand'	'my hand's liver'		Takia:	<i>bani-g ate-n</i>
					hand-1sg liver-3sg
				Waskia:	<i>a-gitiŋ gomaŋ</i>
					1sg-hand 3sg.liver
	'I am dizzy'	'my eye goes round'		Takia:	<i>mala-g i-kilani</i>
					eye-1sg 3sg-go.round
				Waskia:	<i>motam gerago-so</i>
					eye go.round-3sg
	'I disobey him'	'I cut his mouth'		Takia:	<i>awa-n ŋu-tale</i>
					mouth-3sg 1sg-cut
				Waskia:	<i>kuritŋ batugar-so</i>
					3sg.mouth cut-1sg
	'I am angry'	'my guts are bad'		Takia:	<i>ilo-g saen</i>
					inside-1sg bad
				Waskia:	<i>a-gemaŋ memek</i>
					1sg-liver bad
	'I am waiting'	'I am putting my eye'		Takia:	<i>mala-g ŋi-ga</i>
					eye-1sg 1sg-put
				Waskia:	<i>motam bete-so</i>
					eye put-1sg

Takia *-ga* and Waskia *bete-*, which occur in the last example, have the same range of meaning, 'put, do, make'.

Clearly, the package of changes that Takia has undergone is not captured by the term 'syntactic borrowing', because it also entails semantic reorganization. It seems to me, however, that it is a *single* package and needs a single term. The semantic manifestations of this kind of contact-induced change, illustrated in (9) and (10), are traditionally described as 'calques' and 'loan translations', and one might stretch these terms to include (5), but they hardly capture the syntactic aspects of the larger process that were illustrated in (1). I have considered elaborating Harris and Campbell's term 'syntactic borrowing' to 'semantico-syntactic borrowing', this is clumsy. I also think 'borrowing' is more felicitously reserved for the copying of lexical forms from one language to another, and I am unhappy about applying it to the wholesale restructuring of a language's semantic and syntactic structures (especially when forms are not copied), even though Thomason and Kaufman (1988) use it in this way.

For this reason, I have coined the noun 'metatypy' and the adjective 'metatypic' (Ross 1996a, 1997) to refer to the larger process which is manifested in the package of changes I have briefly described. The model for the coinage was the terms 'metamorph' and 'metamorphic', which apply to changes of form (although not in linguistics!). 'Metatypy' accordingly refers to a change of linguistic type, since, in the terms of Greenberg (1966), this is what Takia has undergone. Metatypy is

thus the process whereby the language of a group of bi- or multilingual speakers is restructured on the model of a language they use to communicate with people outside their group. In its fullest manifestation, the process includes:

- (a) the reorganization of the language's semantic patterns and 'ways of saying things';
- (b) the restructuring of its syntax, i.e. the patterns in which morphemes are concatenated to form (i) sentences and clauses, (ii) phrases, and (iii) words.

Apparently (a) precedes (b), and (i), (ii) and (iii) are restructured in this order. More often than not, however, this sequence remains incomplete.

There are a number of accounts in recent publications of languages which have evidently undergone metatypy on the model of a second language in much the way that I have just described. Examples that come to mind are the Oceanic language Maisin, which also underwent metatypy on the model of a neighbouring Papuan language (Ross 1996a), the Papuan language Anêm, on the model of Oceanic Lusi (Thurston 1982), Austronesian Phan Rang Cham on the model of Vietnamese (Thurgood 1996), Tariana in Brazil on the model of Tucanoan languages (Aikhenvald 1996), Bantu Ilwana on the model of Cushitic Orma (Nurse 1994), Arvanitic on the model of Greek (Sasse 1985), the Greek dialects of Asia Minor and Western Armenian on the model of Turkish (Thomason and Kaufman 1988: 215–23, Sasse 1992), the Turkish dialects of western Macedonia and Kosovo on the model of Macedonian and Albanian (Friedman 1996), Rhaeto-Romance dialects on the model of German or Italian (Haiman 1988), Sauris German on the model of a Rhaeto-Romance dialect and of standard Italian (Denison 1968, 1977, 1988), and the Mixe dialect of Basque on the model of Gascon (Haase 1992).

However, it is one thing to describe the linguistic manifestations of metatypy and to list examples. It is another to understand the sociolinguistic conditions in which it occurs and the psycholinguistic process which is at its root. Sociolinguistically, it is clear that metatypy only occurs where a group's speakers are polylectal: we can recognize among their lects an ingroup lect and one or more outgroup lects. The ingroup lect is the one which is peculiar to the group and which is often emblematic of its speakers' identity (Grace 1975, 1981: 155–6), whilst outgroup lects are used for external communication. It is important to note, however, that in some groups many speakers will speak an outgroup lect more often than their emblematic ingroup lect. Reasons may be that an outgroup lect is commonly used within the group, or the speakers spend a large portion of their time interacting with outsiders, or both.

Several terminological decisions are implicit in the previous paragraph. Since there is no sharp boundary between the concepts of language and dialect, I refer to both simply as 'lects' and to speakers who speak two or more lects as 'polylectal' (etymologically correct 'dilectal' is unusable because of its near-homophony with 'dialectal'). I use 'group' for a social network of speakers who share the same

repertoire and usage of lects. The terms 'ingroup lect' and 'outgroup lect' follow from this definition of a group.

In modern Papua New Guinea, the general pattern is that a village has its own emblematic vernacular as its ingroup lect, and a lingua franca, most often Tok Pisin, as the outgroup lect of most or all of its speakers. In pre-modern times, the outgroup lect would normally not have been a pidgin, but the lect of a neighbouring village or a lect functioning as a local trade language (Laycock 1982). Often there would have been more than one outgroup lect spoken in the village. This pattern has no doubt been repeated thousands of times in pre-modern agricultural communities around the world, and also survives, for example, in many parts of Europe where the ingroup lect is a so-called 'local dialect' and the outgroup lect the 'standard language'.

The present-day situation on Karkar Island is that Tok Pisin is the *lingua franca* between the two halves of the island, and bilingualism in Takia and Waskia is not particularly common. But we can make inferences about earlier conditions from ethnography. In traditional times, the males of the island were linked by trading partnerships which were passed on from father to son, and these partnerships often crossed the Takia–Waskia boundary. If a village conducted hostilities with its neighbour, men would call on their kinsmen and their trading partners to come to their aid. As a result, although hostilities were usually between villages which spoke the same ingroup lect, the men fighting on *behalf* of a particular village could readily include speakers of both languages, who must have used one of the two lects in order to communicate (McSwain 1977: 17–21). The linguistic data imply that it was Takia men who were bilingual in Waskia, rather than vice versa, and that they spoke it so often and were so at home in it that over generations they gradually restructured their ingroup lect on the Waskia model.

Of course, what I have just said is largely inference. Because metatypy does not in itself entail the borrowing of *forms*, we cannot always be sure what language was the metatypic model. It is possible that Takia underwent a measure of metatypy on the model of some other Papuan language before its speakers ever migrated to Karkar.

The classic case of metatypy in the literature is Gumperz and Wilson's work (Gumperz 1969, Gumperz and Wilson 1971) on the Indian village of Kupwar, which lies on the Indo-Aryan–Dravidian border. Here, varieties of Indo-Aryan Urdu and Dravidian Kannaḍa have been remodelled, mainly on the basis of Indo-Aryan Marathi, whilst all three have largely retained their own lexical forms, so that text from one of the lects can be translated morpheme by morpheme into either of the others.

Observing the situation at Kupwar, Labov (1971: 459–60) likens it to a 'dotted line'—a perforation—running between the grammar and the lexicon of each lect, as if the lexicon of one lect can be 'torn out' and replaced by the lexicon of another without affecting the grammar. But George Grace (1981: 23–32), in his book *An Essay on Language*, takes issue with Labov and argues that it is not the lexicon which is

torn out and replaced, but just the forms. In conventional terminology, 'lexicon' refers to both form and *meaning*, and Grace's point is that these polylectal speakers construe reality in the same way and have the same semantic organization in each lect they speak. It is only the forms, which Grace calls 'lexification', that change.

The situation with Takia and Waskia is a little more complicated. We can say of the examples in (1) and (10) that the semantic organization of both members of each pair is the same, but not the syntax. Their clausal and phrasal syntax *are* almost the same, but their word-internal morphosyntax differs markedly. Indeed, when we look at the various recorded cases of metatypy, we see that what they all have in common is an identity or close similarity between the semantic organizations of the speakers' two lects, but varying degrees of morphosyntactic similarity, a matter to which I return below.

Interestingly, the components into which Grace's perforation divides language are similar to those used in recent models of speech production. These models use the term 'lemma' for the semantic and syntactic storage unit that underlies a word but excludes its form. If the idea that a word is stored as an abstract unit independent of its form seems odd to you, let me draw your attention to the common enough experience, at least at my age, of knowing what I want to say but not being able to find the 'word' to say it. In speech-production terms, I have retrieved the lemma, but not its form. Levelt (1992) describes the process of speaking as having two primary stages. Stage 1 is 'lemma access', when the speaker generates a 'message' with one or more lemmas. This in turn drives morphosyntax, since each lemma has a limited range of combinatorial possibilities. Stage 2 is 'phonological encoding', when the message is fleshed out into phonological form. The correspondence between Levelt's, Grace's, and my terminologies is shown in (11):

(11)	Levelt (1992)	Grace (1981) ⁴	This chapter
	lemma access	content substance	semantic organization
	morphosyntactic encoding	content form	syntax
	phonological encoding	lexification	lexification

The fact that linguists dealing with quite different aspects of language have arrived at more or less the same model implies rather strongly that this model provides a good (if gross) representation of what happens in speakers' minds.

The semantic correspondences in (10) draw attention to another point. Grace's model, presented in his 1981 book and developed further in *The Linguistic Construction of Reality* (1987), suggests that there are (at least) three main steps from thought to utterance, corresponding to those in (11). First comes the construal of the speaker's perceptions in terms of the language's semantic organization (Grace 1987: 31), second the morphosyntactic encoding of that construal, and third the lexification. This would seem to place an enormous processing burden on the speaker's cognitive/linguistic faculties. But the potential burden is

⁴ Grace borrows the terms 'content form' and 'content substance' from Hjelmslev (1961: 50–2).

lightened by a huge collection of entrenched collocations and structures which appear to be stored as wholes, as Pawley has shown in various publications (Pawley and Syder 1983, Pawley 1985, 1987, 1991). In other words, a lemma is not necessarily a single word: it may be a word, a phrase, or a clause (and the phrase or clause may be complete or may have a slot or slots into which other lemmas are inserted), and any of these may be affected by metatypy.

It is only a short step from here to formulate a hypothesis about what drives metatypy. It seems reasonably clear that for a speaker of, say, Takia and English, cognitive and linguistic processing must impose a substantial burden since the two lects have markedly different semantic organizations and morphosyntax, as well as different lexifications. It appears that there is a strong tendency for the polylectal speaker to reduce the burden by making the two semantic organizations one, and this unification is liable to bring with it a progressive restructuring of syntax as well. The end result—after a number of generations—is a speaker with just one semantic organization, increasingly similar syntactic systems, but two lexifications. In Levelt's terms, this means just one set of lemmas but two phonological encodings. In Weinreich's (1963) once popular terminology, speakers move from coordinate to compound bilingualism. In Sasse's (1985) words, 'bei fortgestztem Sprachkontakt entsteht die Tendenz, *eine Sprache (qua langue) mit unterschiedlichem Wortschatz zu entwickeln*' [with advanced language contact there arises the tendency to develop a single language (= *langue*) with different vocabularies].

The effect of metatypy on syntax appears to follow a regular sequence. The first stage is probably that some semantic reorganization occurs without affecting morphosyntax. At stage two, morphosyntactic restructuring begins. Speakers seek to express the same message in both lects, and this pushes them to reorganize discourse and clause linkage in the ingroup lect in ways which approximate to the target of the outgroup lect. The sentence or clause-chain is presumably the syntactic carrier of a message, so its structure in the ingroup lect is progressively modified towards the outgroup lect target. This has happened in Takia, where clause linkage has been restructured on a Papuan model. At the same time, clause-internal structure is reorganized. This is roughly the stage reached by Anêm and Lusi of north-western New Britain, described by Thurston (1987). Here, speakers of Papuan Anêm are bilingual in the Oceanic intergroup language Lusi. Semantic organization and clause structure are very close. In (12) an English construal would be roughly 'Hand me some tobacco to smoke', but in both Anêm and Lusi one says roughly 'Let some tobacco come (and) I will eat it':

(12) Anêm:	<i>uas</i>	<i>gox</i>	<i>o-mên</i>	<i>da-t</i>
	tobacco	some	HORTATIVE.it-come	IRR. I-(eat-)it ⁵
Lusi:	<i>uasi</i>	<i>eta</i>	<i>i-nama</i>	<i>ŋa-ani-Ø</i>
	tobacco	some	it-come	I-eat-it

⁵ The example is from Thurston (1987: 69), with additional glossing provided by Thurston (p.c.). The verb 'eat' in Anêm has no segmentable stem (Thurston 1987: 57).

Weinreich (1963 [1953]: 50) provides an example of the same kind from the Balkans, where Aromanian, Albanian, Greek, Bulgarian, and Serbo-Croatian all express 'may God punish you' with their equivalents of 'may you find it from God'.

Unlike Takia and Waskia, however, Anêm and Lusi have different phrase structures, apparently because Anêm has not proceeded as far along the metatypic road as Takia has. The examples in (13) (from Thurston 1987: 82) show these differences, but the shared semantic organization remains obvious:

- (13) Anêm: *gêt-i* *ia*
 ear-his fish
 Lusi: *iha* *ai-taŋa*
 fish his-ear
 'lateral fin of a fish'
 Anêm: *eil-îm* *te*
 eye-his knife
 Lusi: *uzage* *ai-mata*
 knife his-eye
 'knife blade'
 Anêm: *agîm-k-i* *tiga*
 neck-LIGATURE-his foot
 Lusi: *ahe-gu* *ai-gauli*
 foot-my his-neck
 'my ankle'

The third stage of metatypy is phrasal restructuring, which was observed above in Takia and Waskia. Finally, word-internal structure is also reorganized, and this is what we apparently see at Kupwar. Sasse (1985) notes that one consequence of reorganizing word-internal structure is that the derivational morphology of the ingroup lect is eroded, and speakers may compensate for the loss of this resource by borrowing lexical forms from the outgroup lect. In Takia, it is certainly true that the language has lost some of its derivational possibilities (for example, it has no valency-changing devices), but its speakers have not resorted to widespread borrowing.

The basis of syntactic restructuring seems to be that polylectal speakers try to equate each construction in their ingroup lect with one in their outgroup lect (cf. Prince 1998). (By a 'construction', I mean the pairing of a morphosyntactic structure with a discourse function.) They progressively adopt a number of strategies to make each construction in the ingroup lect syntactically more similar to the corresponding construction in the outgroup lect. Sometimes, however, speakers are unable to find an equivalent construction in their ingroup lect. In these circumstances, they may resort to outright borrowing of the construction, complete with its grammatical forms. Thus a number of languages in Papua New Guinea have adopted not only the Tok Pisin modal verb construction but also the forms of the modals themselves (Ross 1985). The same languages have also

borrowed Tok Pisin conjunctions. Similarly, Tagalog and Chamorro (both western Austronesian languages) have borrowed Spanish subordinating conjunctions. This does not mean that the discourse function performed by the borrowed construction could not previously be expressed in the ingroup lect, but simply that the constructions used for the function in the two lects were so different that speakers did not equate them.

A special case of this kind of borrowing is the adoption of discourse markers from the outgroup into the ingroup lect. Both Takia and Waskia, and many languages along the north coast of Papua New Guinea, use the marker *aria* 'all right' to mark a 'fresh start' or shift in the topic of conversation. Welsh speakers use English *well* as a discourse boundary marker just as they would in English. This borrowing apparently occurs because here the pairing is not between a morphosyntactic structure and a discourse function but between a form and a discourse function: there is nothing to copy except the form itself. Because it is the larger discourse units that appear to be affected first in metatypy, and this kind of borrowing makes no structural demands on the speaker, I assume that it happens very early in the metatypic process.

I have written above as if metatypy will always occur in a group where the majority of speakers share an ingroup and an outgroup lect. Obviously, this is not true. If a group is tight-knit enough to maintain its ingroup lect but open enough to use an outgroup lect regularly, then metatypy is likely to occur unless there are countervailing factors. What are these potentially countervailing factors? A major sociolinguistic factor is the internal structure of the group's social network. If it is such that there is strong norm enforcement in the speakers' ingroup lect, then metatypy will be slowed or stopped. Norm enforcement is decidedly weak in Papua New Guinea languages, as they rarely correspond to an ethnic group with a strong authority structure, and metatypy appears to be correspondingly common.

Whether and how metatypy is affected by the relative complexity of the two languages or their degree of structural difference is a matter for further research, but I would expect that radical differences in structure would prevent speakers from recognizing the constructional equivalences which are the starting point for metatypy. Maltese, for example, is an Arabic dialect which has undergone metatypy on the model of Italian. From Drewes' (1994) account it seems that Maltese semantic organization reflects reshaping on the Italian model, but the vast differences in structure have hindered syntactic restructuring.

Where speakers' ingroup and outgroup lects are related and structurally similar, the issue of what counts as emblematic in the ingroup lect is important. Where emblematicity is carried largely by the lexicon and speakers can readily establish equivalence at the morpheme level between the two lects, as Sasse describes for Arvanitic and Greek, then they may begin to perceive the two lects as variants of each other, and borrow bound morphemes from a paradigm in the outgroup lect into the corresponding paradigm of the ingroup lect. This is what seems to have

happened in the oft-mentioned case of Meglenite Romanian (spoken to the north of the Greek city of Salonika), where Bulgarian person/number suffixes replaced their Romanian equivalents on the Meglenite verb (Weinreich 1963 [1953]: 32; Thomason and Kaufman 1988: 98).

2.2. LEXICAL BORROWING AND METATYPY

Lexical borrowing is only an intrinsic part of contact-induced change when that change entails abrupt creolization or pidginization (cf. §2.7). Otherwise, it is *not* a necessary condition or concomitant of change: lexical borrowing can occur without metatypy and vice versa. For all the *reorganization* of the Takia lexicon, illustrated in (9) and (10), there has apparently not been a great deal of borrowing of lexical forms from Waskia (or from any other Papuan language). Aikhenvald similarly reports that Tariana speakers avoid lexical borrowing because emblematicity is carried by the lexicon. And, on the other hand, Japanese has borrowed quite extensively from English without widespread polylectalism. Lexical borrowing may *accompany* contact-induced change, but it is not a necessary part of it.

2.3. PHONOLOGY AND METATYPY

Some writers note that metatypy may be accompanied by phonological assimilation of the ingroup lect to the outgroup, e.g. Aikhenvald (1996, this volume) for Tariana, Sasse (1985: 57–61) for Arvanitic, and Thurgood (1996) for Phan Rang Cham. Yet Haase (1992: ch. 8) records with equal conviction that metatypy is not accompanied by phonological assimilation in Mixe Basque, and the historical evidence shows that Takia has moved *away* from Waskia phonologically.

What should we do with this seemingly contradictory evidence? First, we may assume as the default that metatypy is *not* accompanied by phonological assimilation, since polylectal speakers are more likely to have a ‘foreign accent’ in their outgroup than their ingroup lect (cf. §2.7). If the ingroup lect does assimilate phonologically to the outgroup, it is reasonable to infer that there is a special reason for this, and indeed the available case studies do offer such reasons. Tariana speakers always marry outside their own group, so every child has a parent who is not a native speaker and who is likely to speak a phonologically modified form of Tariana. A Takia speaker, on the other hand, usually marries another Takia. In Arvanitic, there have been two stages of phonological innovation. First, there was substantial borrowing of Greek vocabulary, complete with Greek phonemes. And later, as speakers began to abandon Arvanitic in favour of Greek, Greek phonological patterns took over. Note that none of these factors—intermarriage, lexical borrowing, or language death—is a necessary condition or concomitant of metatypy. Indeed, as noted in §2.2, lexical borrowing can occur without bilingualism. In Tagalog, for example, extensive lexical borrowing from Spanish resulted in changes in the vowel system, complicating the language’s verbal morphology—all without a majority of Tagalog speakers being bilingual in

Spanish. So we can reasonably conclude that metatypy and phonological assimilation are not causally related.

2.4. HOW CONVERGENCE IS EFFECTED

Dixon (1997: 70–1) suggests that during a period of equilibrium, the languages of an area will converge grammatically, and this is true, for example, of the Papuan languages occupying much of mainland New Guinea. Many of these languages have remarkably similar categories and structures, generally those shown for Waskia in Table 1. However, neighbouring languages often use non-cognate forms for similar tasks, and this combination of typological likeness but formal non-cognacy is exactly what is symptomatic of metatypy. We have seen, however, that metatypy is not a reciprocal or convergent process, but one whereby an ingroup lect is unilaterally reorganized on the model of an outgroup lect. The convergence which manifests itself in a linguistic area like New Guinea is the outcome of this unilateral process repeated over and over again with different language pairs.

This is not, of course, to say that metatypy is never reciprocal. Haase (1997) shows that whilst Basque has undergone metatypy on the model of Gascon, Gascon has also undergone metatypy on the model of Basque. But this convergence is the outcome of two unilateral applications of the metatypic process.

2.5. EQUILIBRIUM AND PUNCTUATION PROCESSES ARE NOT MUTUALLY EXCLUSIVE

Dixon (1997: ch. 6) describes the history of human languages in terms of long periods of equilibrium disturbed by short periods of punctuation. A period of punctuation may entail ‘a multiple “split and expansion” (which would be appropriately modelled by a family tree diagram)’, whilst a period of equilibrium is characterized by gradual convergence.

If a large enough span of human history is viewed, as it were, from a distance, then this model has much to recommend it. From a closer viewpoint, however, things look rather different, as we see in the case of Takia. Metatypy has, over time, brought Takia closer to Waskia, as I illustrated in §2.1, and it is legitimate to describe this as a period of equilibrium which probably lasted from the arrival of Oceanic speakers on the north coast of Papua New Guinea (probably more than a millennium ago) until German colonization in the 1880s. But as Takia has undergone metatypy it has also slowly diverged from its closest Oceanic relatives, and we can also readily depict this divergence in the form of a family tree (as shown in Ross 1988: 122 and 161). Similarly, we can easily recognize genetic relationships among various European languages, even though contact—apparently during the great migrations which preceded the Middle Ages (Haspelmath 1998)—has resulted in a typological grouping (‘Standard Average European’) which cuts across family boundaries.

The point here is not so much that the family tree only models what happens in a period of punctuation, but that it can be used to model the two ends of a

cline. At one end is language fissure (Dixon's 'split and expansion'), at the other the gradual growth and differentiation of a dialect network, which may involve metatypy in some member lects. Thus, if our view is close enough, we can also use the family tree to model events during an equilibrium period, with two caveats. First, the family tree may only tell half the truth, as it usually omits contact phenomena. Second, the tree should use different representations for language fissure and lectal differentiation (among other processes) (see Ross 1997 for a discussion). All too frequently, linguists have drawn family trees as if they were oblivious to this difference.

2.6. HOW INDEPENDENT BUT FORMALLY SIMILAR CHANGES MAY OCCUR

I noted in §2.4 that the Papuan languages of much of New Guinea form a linguistic area. I also showed in §2.1 that metatypy entails the reemployment of inherited forms in new functions. One outcome of these facts is that Oceanic languages sometimes share an innovation not because of shared inheritance but because they have independently undergone metatypy on the model of languages belonging to the same linguistic area.

Maisin is, like Takia, an Oceanic language that has undergone metatypy on the model of a Papuan language (it has also undergone other processes, but these need not concern us here; Ross 1996a). In both Takia and Maisin, metatypy has entailed the creation of a class of cosubordinate clauses (§2.1), and in both languages there is a distinction between realis and irrealis cosubordinate markers, which are encliticized to the (clause-final) predicate. In Takia, the form of the irrealis marker is *-p* or *-pe*; in Maisin it is *-fe*, both reflecting the Proto-Western Oceanic conjunction **pe*. If we did not have ample evidence that Takia and Maisin have quite long separate histories and have undergone metatypy separately, we might be tempted to assume that these forms represent a shared innovation. What has happened, however, is that the two languages have independently undergone metatypy, but on models that both belonged to the New Guinea linguistic area and provided similar structural templates. In both languages, interclausal conjunctions have been reanalysed as cosubordinate markers, and the conjunctions reanalysed as the irrealis enclitic happened to be cognate.

Whilst there is nothing particularly surprising about the changes described above, they do provide an instantiation of Campbell's (1997) warning that areal features may interfere with reconstruction.

2.7. TOWARDS A TYPOLOGY OF CONTACT-INDUCED CHANGE

It is all too easy to talk about contact-induced change as if it were a unitary phenomenon. However, if one accepts Thomason and Kaufman's (1988: 35) dictum that 'it is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact', then one needs to position any given case of contact-induced

change within a relevant sociolinguistic paradigm. I have presented parts of such a paradigm elsewhere (Ross 1991, 1997), and will repeat the relevant parameters here without supporting argument.

All parameters in the paradigm are gradient rather than binary, but for the sake of brevity I will refer to them as if they were binary. The first parameter is that contact-induced change is either gradual or catastrophic, but I have been concerned here only with gradual change. We can identify various kinds of gradual contact-induced change, and the occurrence of one or other of these evidently depends on the structure of a group's social network. We can conveniently define types of group by structural parameters, as follows (similar, but not identical, parameters have been proposed by Andersen 1988: 70-4 and Thurston 1987: 55-60, 1989: 556):

- (14) Types of group defined by structural parameters:
 - (a) closed
 - (b) open . . .
 - (i) and loose-knit
 - (ii) and tight-knit

Essentially, a closed group is one with few relationship links to speakers in other groups, and an open group is one with many such links. In other words, the closed/open distinction describes a group's *external* relationships. An open group, in turn, may be tight-knit or loose-knit. This distinction refers to its *internal* relationships. A loose-knit group is one where speakers are not bound together by tight bonds of linguistic solidarity, whilst a tight-knit group is one where they are.

There are three kinds of gradual change in polylectal groups which correlate with these parameters, but only the last involves metatypy:

- (15) Types of gradual change in polylectal groups:
 - (a) If a group becomes more *closed*, its members may modify their version of the lect they share with their neighbours in one (or both) of two ways:
 - (i) They may make their lect harder to learn and understand, complicating it with phonological compactness, morphological opacity, and suppletion. Thurston (1987) labels this process 'esoterogeny' (the generation of an 'esoteric' language) and illustrates it from Oceanic languages of north-west New Britain. Other examples are described by Andersen (1988).
 - (ii) They may relexify their lect with the lexicon of an outgroup language so that it becomes unique to them. Thus people of mixed ancestry in Java created Javindo by relexifying Javanese with Dutch lexicon (de Gruiter 1994) and Ecuadoran descendants of Quechua speakers who are culturally alienated from both rural Quechua and

urban Spanish-speaking cultures speak Media Lengua, Quechua relexified with Spanish lexicon (Muysken 1994).

- (b) If a group is *open* and becomes *loose-knit*, its speakers may shift from the ingroup to the outgroup lect. Often, I assume, no trace of the ingroup lect remains, and so the shift evades detection by historical linguists. Sometimes, a few lexical items survive which retain some emblematic significance. For example, one clan of speakers of the Arop-Sissano language of the North New Guinea cluster retained the terms for 'dog' and 'coconut' from (Papuan) One, their ingroup language before they shifted to Arop-Sissano (Laycock 1973). If the outgroup lect becomes inaccessible to the shifting group before they have acquired native-like mastery, then at least the ingroup 'accent' survives, reshaping the phonology of the outgroup lect on the ingroup model. The Oceanic language Madak, spoken on New Ireland, reflects phonological reshaping, with phonological rules imported probably by speakers of Kuot, the neighbouring Papuan language (Ross 1994a).
- (c) If a group is *open* and *tight-knit*, metatypy may occur, as illustrated in §2.1, restructuring the ingroup lect's semantic organization and at least part of its syntax (starting at the level of the clause) on the model of the outgroup lect.

Cutting across these sociolinguistic parameters is the parameter of linguistic similarity, ranging from close relationship and mutual intelligibility at one extreme to absence of relationship and a minimum of structural isomorphism at the other. Esoterogeny, for example, will only occur if the ingroup lect and at least one relevant outgroup lect are similar; if they are unrelated and structurally unlike, there is no reason for speakers to make their lect harder to understand. Metatypy, as I noted earlier, may be fostered by close structural similarity (but not necessarily a close relationship) between lects, whilst great structural difference may impede it.

It is clear that the parameters above need further refinement. For example, Grace (1996) reports on a situation in a collection of closely related Oceanic languages in southern New Caledonia where the lexicon shows an astounding quantity of multiple reflexes of Proto-Oceanic phonemes. This finding can be interpreted somewhat as follows. At some time in the past, metatypy had probably already eliminated such semantic and syntactic differences as existed among the languages and any phonological differences had also been levelled, but the diverse *forms* of words had been retained, presumably because of their emblematic value. For reasons now lost, speakers evidently came to spend more of their time with speakers of one or more of the other lects than of their own and, as a result, acquired an intuitive grasp of sound correspondences among lects, using them to convert the phonological shapes of words from one lect to another. At times, the speakers' internalized correspondences and the actual historical correspondences

differed because the word had been the object of a lexically diffused change or reflected a proto-phoneme which had merged with another proto-phoneme in the 'donor' but not in the 'recipient' language. Consequently, the output of the speaker's conversion was not the same as if the word had been directly inherited and over centuries the 'original' sound correspondences were completely lost and reflexes were multiplied.

Although we can describe this situation in terms of the network model implicit in the paradigm above, the parameters in (14) are not really sufficient to capture it, as we are confronted by a situation in which speaker group (i.e. speakers of the ingroup lect) and social group have ceased to match. Johnson (1990) describes a situation in Australia's Cape York Peninsula where the data are patterned in a way similar to Grace's, and where the mismatch between the two groups could still be recorded. Here, each person belongs to an exogamous patriclan with its own lect (the speaker group), but people move about in a hunter-gatherer band (their social group) with a shifting membership drawn from various patriclans and speaking their own lects because the lect is emblematic of their clan's right to use a particular area of land. This is important, since Australia is Dixon's (1997) paradigm case for a long period of equilibrium (see also Dixon, this volume, and Dench, this volume), and Grace's New Caledonia analysis perhaps provides one starting point for improving our understanding of Australia. Since Australian speaker groups are clearly open and tight-knit—remembering that 'tight-knit' means 'linked by tight bonds of linguistic solidarity'—it seems that we need a parameter subordinate to 'tight-knit' which we might label with the binary pair *inwardly associating*–*outwardly associating*. Speaker groups like those described by Grace and Johnson would be outwardly associating, i.e. speakers participate in a social group which is not coterminous with the speaker group.

The parameters involved in catastrophic change are less clear to me, partly because they are less well illustrated in north-west Melanesia. But catastrophic change is relevant here because it sometimes gives rise to phenomena which resemble metatypy. 'Catastrophe' seems always to entail the enforced melding of groups with different ingroup lects into a new larger group, where enforcement is either by human intervention or by natural disaster. A new social network is abruptly created or rearranged, so that old groups are compelled to become more *open*, establishing multiplex relationship links with each other.

At the extreme of catastrophe, old groups also become *loose-knit*, abandoning their identity in favour of the identity of the new larger group. They are faced with the problem of inter-communication between members of the old groups, and there appear to be three broad types of solution, listed in (16), depending (a) on the degree of similarity among their lects and (b) the availability and accessibility of a (potentially) shared outgroup language. Note that the listed solutions are those which a historical linguist can detect. If one or more old groups shifts fully to the lect of another old group, or all old groups shift fully to a shared outgroup language, then, as noted under (15b) above, the shift most often leaves no trace of

the ingroup lect(s). Exceptions occur when the outgroup lect is derived from an unstable jargon or is not accessible to the shifting group to a sufficient degree or for a sufficient time for its members to acquire native-like mastery. That is, cases of lectal shift are detectable only when shift is imperfect, meaning that what is reconstructable probably underplays the importance of catastrophic change in linguistic prehistory.

(16) Strategies of intercommunication in a new social network:

- (a) Where there is a degree of mutual intelligibility among the ingroup lects of the old groups, a new lect may arise out of the fusion of the old (Ross 1997: 228–31). There appear to be two versions of the fusion process:
 - (i) Where speakers are conscious of their membership of the new group rather than the old, features in which the old lects differ are suppressed, especially where these are emblematic of a particular old group. Sometimes this levelling has only minor effects. For example, certain dialects of Ukrainian and of north Russian appear to have simplified their vowel systems as a result of being used as a lingua franca (Jakobson 1962 [1929]: 82; Andersen 1988: 49–51). In more extreme cases, the outcome is koineization, i.e. the levelling of differences (Ross 1997: 236–8).
 - (ii) Where speakers lack this awareness, features in which the old lects differ may co-occur, resulting in irregularity in the new lect. A south Melanesian example is Anejom, where, among other things, every noun has a prefix reflecting the Proto-Oceanic article **na*, but the prefix has two forms and there is no way of predicting which prefix will occur on which noun. The two forms seem to come from different lects which fused after the population was drastically reduced by European-introduced disease (Lynch and Tepahae 1999).
- (b) Where there is little or no mutual intelligibility, a shared outgroup lect may be adopted. A new lect emerges only where full shift does not occur, and this happens in two circumstances (Thomason and Kaufman 1988: 147–99):
 - (i) Where the outgroup lect is not fully accessible to speakers for one reason or another, ‘abrupt creolization’ occurs: the new lect is an imperfectly learned version of the outgroup lect.
 - (ii) Where the outgroup lect is not a fully fledged language but an unstable intergroup jargon or a functionally restricted intergroup lect (a pidgin), the old groups of speakers expand it into a fully fledged lect (a creole).

The question of how speakers of a creole replace structural features which are missing from the imperfectly learned or restricted intergroup lect has long been a matter of controversy, summarized by Thomason and Kaufman.

- (c) Where there is no mutual intelligibility, no potentially shared outgroup lect, and the new group is a small one created by intermarriage between two groups, the process which Bakker calls 'language intertwining' sometimes occurs, so that a fairly consistent mixture of the two lects arises. For example, Michif (Thomason and Kaufman 1988: 228–33, Bakker 1994) mixes French and Cree: most nouns are French, verbs are almost all Cree, possessive pronouns are French, and so on. A celebrated example is Copper Island Aleut, which uses Russian finite verb inflections in an otherwise largely Aleut matrix (Thomason and Kaufman 1988: 233–8, Golovko 1994, Golovko and Vakhtin 1990). Note that this is not the same outcome as relexification (15a, ii).

In less extreme cases of catastrophe, especially those involving invasion by a foreign power, old groups remain *tight-knit* and retain their identities whilst also being part of the new larger group. In the cases I am aware of, the component groups have retained their lects but share a new 'outgroup' lect, often the invaders' language, which becomes the 'ingroup' lect of the larger group. In each of these cases, this lingua franca is or was a language also spoken outside the domain of the larger group: Mandarin in Taiwan, English in Singapore, Dutch in the Cape. However, in each case large segments of the component groups lack direct access to the standard version of the lingua franca, learn it imperfectly, and a more or less stable variety emerges which serves as the shared outgroup lect. This variety simplifies and regularizes the original, and introduces phonological and syntactic features as well as features of semantic organization from one or more of the component groups' lects (see Chappell on Taiwan Mandarin in this volume). Platt (1975) coins the term 'creoloid' to describe a stable variety of this kind.

In Taiwan, for example, Mandarin replaced Japanese as the official lingua franca in 1945, and the Communist takeover on the mainland in 1949 caused an exodus of refugees to Taiwan. Seventy per cent of Taiwan's population, however, speaks a Southern Min dialect of Chinese, mutually incomprehensible with Mandarin, and this has had a profound phonological, morphosyntactic, and lexical effect on the Taiwan Mandarin creoloid which has emerged since 1949 (Kubler 1985). Sometimes the process of emergence is complicated by the fact that some, often elite, members of component groups do have full access to the lingua franca and learn it well, so that what emerges is something resembling a creole continuum ranging from the creoloid extreme to the acrolectal standard, as Platt (1975) describes for Singapore Colloquial English. Roberge (1993) argues that Afrikaans has emerged from somewhere in the middle of an eighteenth-century continuum ranging from Cape Dutch Creole through a creoloid to a Cape Dutch dialect of Netherlands Dutch.

I have attempted to add catastrophic changes to this paradigm of contact-induced change types because I would like to be able to distinguish between metatypy and those types of change whose diagnostic features overlap with it.

This overlap is obvious in the cases of creolization, whether abrupt or via pidginization (16b), and creoloid formation, discussed immediately above. There has also been some confusion in the literature between the features diagnostic of metatypy and of imperfect shift (15b), as I have noted elsewhere (Ross 1997: 246–7), although their outcomes seem to me to differ sharply. The features diagnostic of esoterogeny and relexification (15a), fusion (including koineization) (16a), and language intertwining (16b) are so different from those resulting from metatypy that there is no need to consider them further here.

When we recognize that a certain lect shows signs of contact-induced change, often we cannot immediately identify the kind of change that has occurred. What we can often identify, however, is its genetic inheritance, manifested in its lexicon, in its regular sound correspondences with its genetic relatives (unless it manifests the extreme irregularity reflected in lects spoken by outwardly associating groups), and in the forms of its grammatical morphemes (especially its bound morphemes). Thus Takia, which has undergone metatypy, has grammatical morphemes composed of inherited Oceanic forms and a largely Oceanic lexicon with regular reflexes of Proto-Oceanic phonemes. The same is true of Madak, a putative outcome of imperfect shift. In Tok Pisin, a creole, grammatical morphemes are composed of English material, from which the language also has most of its lexicon and with which it has regular sound correspondences. The same holds for Singapore Colloquial English, a creoloid. There is thus no point in examining inherited features if we wish to diagnose what kind of contact-induced change has occurred. Instead, we need to identify the features which show a mismatch with genetic inheritance, asking whether there has been semantic reorganization, syntactic restructuring, phonological modification, and/or simplification. The diagnostic values of these features are tabulated in (17):

(17)	semantic reorganization	syntactic restructuring	phonological modification	simplification and regularization
metatypy	yes	yes	not necessarily	no
creolization	yes	yes	yes	yes
creoloid	yes	yes	yes	some
imperfect shift	no	not necessarily	yes	not necessarily

This tabulation shows that the diagnosis of contact history is not necessarily simple: it is, for example, potentially difficult to distinguish a creole and a creoloid. Whilst it is reasonably clear on inspection that, by the diagnostics in (17), Afrikaans is a creoloid, it may well be that there are creoloids that have undergone so much simplification that we cannot distinguish them from creoles like Tok Pisin. Basilectal Singapore English would be such a case if it were divorced from the basilect-to-acrolect continuum: it shows all the diagnostic features of a creole, including regularization that introduces features attributable to none of the languages of Singapore (Ho and Platt 1993, Gil, forthcoming).

Our interest here, however, is to check whether we can distinguish metatypy

from other kinds of contact-induced change, and the answer is a tentative 'yes'. The symptoms of metatypy and imperfect shift differ most from each other, and the worst-case scenario is that a case of metatypy might differ from one of imperfect shift only in the feature of semantic reorganization—present in metatypy but absent in imperfect shift. The distinction between metatypy and creolization or creoloid formation is potentially rather small. A creole like Tok Pisin has its semantic organization and syntactic structure not from its lexifier language, English, but from the Oceanic languages of its first speakers. Hence if we compare Takia and Tok Pisin with regard to semantic organization and syntactic structure, we will find that neither has acquired them from its apparent genetic source. If, however, we compare them with regard to phonological modification and to simplification and regularization, we find that Takia has undergone neither, whilst Tok Pisin has undergone both in quite a radical way. This is enough to tell us that Takia is the outcome of metatypy and Tok Pisin of creolization.

It thus seems that metatypy is not only the outcome of a paradigmatically distinct set of social conditions. It is also an outcome which can generally be distinguished from other outcomes of contact-induced change and which can be used to reconstruct something of the sociolinguistic history of its speakers.

Finally, it is worth noting that there is a fundamental difference between the methodological roles of innovations identified by historical linguists for subgrouping purposes and those identified in the diagnosis of contact-induced change. In the case of contact-induced change, the *kind* of innovation is crucial to diagnosis. In the case of subgrouping, it is the fact that innovations are *shared* by several languages which have inherited them from a common ancestor that matters, not especially what kind of innovations they are (although some kinds of innovation are more likely to occur independently and are therefore less reliable subgrouping indicators). In fact, an innovation could be used for both purposes: to define a subgroup and to diagnose contact change that had occurred in their common ancestor.

References

- AIKHENVALD, A. Y. 1996. 'Areal diffusion in north-west Amazonia: the case of Tariana', *Anthropological Linguistics*, 38.73–116.
- ANDERSEN, H. 1988. 'Centre and periphery: adoption, diffusion and spread', pp. 39–85 of *Historical Dialectology*, edited by J. Fisiak. Berlin: Mouton de Gruyter.
- BAKKER, P. 1994. 'Michif, the Cree-French mixed language of the Métis buffalo hunters in Canada', pp. 13–33 of Bakker and Mous (1994).
- and MOUS, M., 1994. Editors of *Mixed Languages: 15 Case Studies in Language Intertwining*. Studies in Language and Language Use 13. Amsterdam: Institute for Functional Research into Language and Language Use (IFOTT).
- BARKER, F., and LEE, J. 1985. *Waskia diksenari*. Dictionaries of Papua New Guinea 7. Ukarumpa: Summer Institute of Linguistics.

- BLUST, R. A. 1978. 'Eastern Malayo-Polynesian: a subgrouping argument', pp. 181–234 of *Second International Conference on Austronesian Linguistics: Proceedings* (Pacific Linguistics C-61), edited by S. A. Wurm and L. Carrington. Canberra: Pacific Linguistics.
- 1980. 'Austronesian etymologies', *OL*, 19.1–182.
- 1983–4. 'Austronesian etymologies II', *OL*, 22–3.29–149.
- 1986. 'Austronesian etymologies III', *OL*, 25.1–123.
- 1989. 'Austronesian etymologies IV', *OL*, 28.111–80.
- 1995. 'Austronesian comparative dictionary'. Computer files, University of Hawaii, Honolulu.
- CAMPBELL, L. 1997. 'Typological and areal issues in reconstruction', pp. 49–72 of *Linguistic Reconstruction and Typology*, edited by J. Fisiak. Berlin: Mouton de Gruyter.
- CAPELL, A. 1943. *The Linguistic Position of South-Eastern Papua*. Sydney: Australasian Medical Publishing Company.
- 1976. 'General picture of Austronesian languages, New Guinea area', pp. 5–52 of Wurm (1976).
- CHAPIN, P. G. 1974. 'Proto-Polynesian *ai'. Mimeo.
- DEMPWOLFF, O. 1927. 'Das Austronesische Sprachgut in den melanesischen Sprachen', *Folia Ethnographica* 3.32–43.
- 1934. *Vergleichende Lautlehre des austronesischen Wortschatzes*, vol. 1: *Induktiver Aufbau einer indonesischen Ursprache*. Beihefte zur Zeitschrift für Eingeborenen-Sprachen 15. Berlin: Dietrich Reimer.
- 1937. *Vergleichende Lautlehre des austronesischen Wortschatzes*, vol. 2: *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen*. Beihefte zur Zeitschrift für Eingeborenen-Sprachen 17. Berlin: Dietrich Reimer.
- DENISON, N. 1968. 'Sauris: a trilingual community in diatypic perspective', *Man*, n.s. 3.578–92.
- 1977. 'Language death or language suicide?', *International Journal of the Sociology of Language* 12.13–22.
- 1988. 'Language contact and language norm', *Folia Linguistica* 22.11–35.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DREWES, A. J. 1994. 'Borrowing in Maltese', pp. 83–111 of Bakker and Mous (1994).
- DURIE, M., and ROSS, M. D. 1996. Editors of *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York: Oxford University Press.
- DUTTON, T. E. 1971. 'Languages of south-east Papua: a preliminary report', *Papers in New Guinea Linguistics* 14 (Pacific Linguistics A-28: 1–46), edited by T. E. Dutton, C. L. Voorhoeve, and S. A. Wurm. Canberra: Pacific Linguistics.
- 1976. 'Magori and similar languages of south-east Papua', pp. 581–636 of Wurm (1976).
- 1982. 'Borrowing in Austronesian and non-Austronesian languages of coastal south-east mainland Papua New Guinea', pp. 109–77 of *Papers from the Third International Conference on Austronesian Linguistics*, vol. 1: *Currents in Oceanic* (Pacific Linguistics C-74), edited by A. Halim, L. Carrington, and S. A. Wurm. Canberra: Pacific Linguistics.
- 1994. 'Motu–Koiarian contact, Papua New Guinea', pp. 181–232 of Dutton and Tryon (1994).
- and TRYON, D. 1994. Editors of *Language Contact and Change in the Austronesian World*. Berlin: Mouton de Gruyter.

- FOLEY, W. A. 1986. *The Papuan Languages of New Guinea*. Cambridge: Cambridge University Press.
- and VALIN JR, R. D. VAN 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- FRIEDMAN, V. 1996. 'West Rumelian Turkish in Macedonia and adjacent areas'. Paper presented to the Conference on Turkish in Contact, Netherlands Institute for Advanced Studies, Wassenaar.
- GIL, D. Forthcoming. 'English goes Asian: number and (in)definiteness in the Singlish noun-phrase', *Noun Phrases in European Languages: Empirical Approaches to Language Typology*, edited by F. Plank. Berlin: Mouton de Gruyter.
- GOLOVKO, E. 1994. 'Mednyj Aleut or Cooper Island Aleut: an Aleut-Russian mixed language', pp. 113–21 of Bakker and Mous (1994).
- and VAKHTIN, N. B. 1990. 'Aleut in contact: the CIA enigma', *Acta Linguistica Hafniensia* 22.97–125.
- GOULDEN, R. J. 1996. 'The Maleu and Bariai languages of West New Britain', pp. 63–144 of *Studies in Languages of New Britain and New Ireland*, vol. 1: *Austronesian Languages of the North New Guinea Cluster in Northwestern New Britain* (Pacific Linguistics C-135), edited by M. D. Ross. Canberra: Pacific Linguistics.
- GRACE, G. W. 1971. 'Notes on the phonological history of the Austronesian languages of the Sarani coast', *Oceanic Linguistics* 10.11–37.
- 1975. 'Linguistic diversity in the Pacific: on the sources of diversity'. Paper presented to the Thirteenth Pacific Science Congress, Vancouver.
- 1981. *An Essay on Language*. Columbia, SC: Hornbeam.
- 1987. *The Linguistic Construction of Reality*. London: Croom Helm.
- 1996. 'Regularity of change in what?', pp. 157–79 of Durie and Ross (1996).
- GREENBERG, J. H. 1966. *Language Universals*. The Hague: Mouton.
- GRUITER, M. de. 1994. 'Javindo, a contact language in pre-war Semarang', pp. 151–9 of Bakker and Mous (1994).
- GUMPERZ, J. J. 1969. 'Communication in multilingual communities', in *Cognitive Anthropology*, edited by S. Tyler. New York: Holt, Rinehart, and Winston.
- and WILSON, R. 1971. 'Convergence and creolization: a case from the Indo-Aryan/Dravidian border', pp. 151–68 of *Pidginization and Creolization of Languages*, edited by D. Hymes. Cambridge: Cambridge University Press.
- HAASE, M. 1992. *Sprachkontakt und Sprachwandel im Baskenland: Einflüsse des Gaskognischen und Französischen auf das Baskische*. Hamburg: Buske.
- 1997. 'Gascon et Basque: bilinguisme et substrats', *Sprachtypologie und Universalienforschung* 50.189–228.
- HAIMAN, J. 1988. 'Rhaeto-Romance', pp. 351–90 of *The Romance Languages*, edited by M. Harris and N. Vincent. London: Croom Helm.
- HARRIS, A. C., and CAMPBELL, L. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- HASPELMATH, M. 1998. 'How young is Standard Average European?', *Language Sciences* 20.271–87.
- HJELMSLEV, L. 1961. *Prolegomena to a Theory of Language*, trans. F. J. Whitfield. 2nd edition. Madison: University of Wisconsin Press.
- HO, M.-L., and PLATT, J. T. 1993. *Dynamics of a Contact Continuum: Singaporean English*. Oxford: Clarendon Press.

- JAKOBSON, R. 1962 [1929]. 'Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slaves', pp. 7–116 of *Selected Writings*, vol. 1, edited by R. Jakobson. The Hague: Mouton.
- JOHNSON, S. 1990. 'Social parameters of linguistic change in an unstratified Aboriginal society', pp. 419–33 of *Linguistic Change and Reconstruction Methodology*, edited by P. Baldi. Berlin: Mouton de Gruyter.
- KUBLER, C. C. 1985. *The Development of Mandarin in Taiwan: A Case Study of Language Contact*. Taipei: Student Book Co.
- LABOV, W. 1971. 'The notion of "system" in creole studies', pp. 447–72 of *Pidginization and Creolization of Languages*, edited by D. Hymes. Cambridge: Cambridge University Press.
- LAYCOCK, D. C. 1973. 'Sissano, Warapu, and Melanesian pidginization', *OL*, 12.245–78.
- 1982. 'Melanesian linguistic diversity: a Melanesian choice?', pp. 33–8 of *Melanesia beyond Diversity*, vol. 1, edited by R. J. May and H. Nelson. Canberra: Research School of Pacific Studies, Pacific Linguistics.
- LEVELT, W. J. M. 1992. 'Accessing words in speech production: stages, processes and representations', *Cognition* 42.1–22.
- LYNCH, J. 1981. 'Melanesian diversity and Polynesian homogeneity: the other side of the coin', *OL*, 20.95–129.
- and TEPAHAE, P. 1999. 'Digging up the linguistic past: the lost language(s) of Aneityum, Vanuatu', in *Archaeology and Language*, vol. 3, edited by R. Blench and M. Spriggs. London: Routledge.
- MC SWAIN, R. 1977. *The Past and Future People*. Melbourne: Oxford University Press.
- MUYSKEN, P. 1994. 'Media Lengua', pp. 207–11 of Bakker and Mous (1994).
- NURSE, D. 1994. 'South meets north: Ilwana = Bantu + Cushitic on Kenya's Tana River', pp. 213–22 of Bakker and Mous (1994).
- PAWLEY, A. K. 1973. 'Some problems in Proto Oceanic grammar', *OL*, 12.103–88.
- 1985. 'On speech formulas and linguistic competence', *Lenguas modernas (Universidad de Chile)* 12.84–104.
- 1987. 'Encoding events in Kalam and English: different logics for reporting experience', pp. 329–60 of *Coherence and Grounding in Discourse*, edited by R. S. Tomlin. Amsterdam: John Benjamins.
- 1991. 'Saying things in Kalam: reflections on language and translation', pp. 432–44 of *Man and a Half: Essays in Pacific Anthropology and Ethnobiology in Honour of Ralph Bulmer* (Polynesian Society Memoirs 48), edited by A. Pawley. Auckland: The Polynesian Society.
- and ROSS, M. D. 1993. 'Austronesian historical linguistics and culture history', *Annual Review of Anthropology* 22.425–59.
- 1995. 'The prehistory of Oceanic languages: a current view', pp. 39–74 of *The Austronesians: Historical and Comparative Perspectives*, edited by P. Bellwood, J. Fox, and D. Tryon. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, Australian National University.
- and SYDER, F. 1983. 'Two puzzles for linguistic theory: nativelike selection and nativelike fluency', pp. 191–225 of *Language and Communication*, edited by J. Richards and R. Schmidt. London: Longman.
- PLATT, J. T. 1975. 'The Singapore English speech continuum', *Anthropological Linguistics* 17.363–74.
- PRINCE, E. F. 1998. 'The borrowing of meaning as a cause of internal syntactic change', pp.

- 339–62 of *Historical Linguistics 1997: Selected Papers from the 13th International Conference on Historical Linguistics, Düsseldorf, 10–17 August 1997*, edited by Monika S. Schmid, Jennifer R. Austin, and Dieter Stein. Amsterdam: John Benjamins.
- RAY, S. H. 1926. *A Comparative Study of the Melanesian Island Languages*. Cambridge: Cambridge University Press.
- ROBERGE, P. T. 1993. *The Formation of Afrikaans*. Stellenbosch Papers in Linguistics 27. Stellenbosch: Department of General Linguistics, University of Stellenbosch.
- ROSS, M. D. 1985. 'Current use and expansion of Tok Pisin: effects of Tok Pisin on some vernacular languages', pp. 539–56 of *Handbook of Tok Pisin (New Guinea Pidgin)* (Pacific Linguistics C-70), edited by S. A. Wurm and P. Mühlhäusler. Canberra: Pacific Linguistics.
- 1987. 'A contact-induced morphosyntactic change in the Bel languages of Papua New Guinea', pp. 583–601 of *A World of Language: Papers Presented to Professor S. A. Wurm on his 65th Birthday* (Pacific Linguistics C-100), edited by D. C. Laycock and W. Winter. Canberra: Pacific Linguistics.
- 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Pacific Linguistics C-98. Canberra: Pacific Linguistics.
- 1991. 'Refining Guy's sociolinguistic types of language change', *Diachronica* 8.119–29.
- 1992. 'The position of Gumawana among the languages of the Papuan Tip cluster', *Language and Linguistics in Melanesia* 23.139–65.
- 1994a. 'Areal phonological features in north central New Ireland', pp. 551–72 of Dutton and Tryon (1994).
- 1994b. 'Describing inter-clausal relations in 'Takia'', pp. 40–85 of *Topics in Descriptive Austronesian linguistics* (Semaian 11), edited by G. P. Reesink. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Rijksuniversiteit te Leiden.
- 1995. 'Some current issues in Austronesian linguistics', pp. 45–120 of *Comparative Austronesian Dictionary*, vol. 1, edited by D. Tryon. Berlin: Mouton de Gruyter.
- 1996a. 'Contact-induced change and the comparative method: cases from Papua New Guinea', pp. 180–217 of Durie and Ross (1996).
- 1996b. 'Is Yapese Oceanic?', pp. 121–66 of *Reconstruction, Classification, Description: Festschrift in Honor of Isidore Dyen*, edited by B. Nothofer. Hamburg: Abera.
- 1996c. 'Reconstructing food plant terms and associated terminologies in Proto Oceanic', pp. 163–221 of *Oceanic Studies: Proceedings of the First International Conference on Oceanic Linguistics* (Pacific Linguistics C-133), edited by J. Lynch and F. Pat. Canberra: Pacific Linguistics.
- 1996d. Compiler of 'On the origin of the term "Malayo-Polynesian"', *OL*, 35.143–5.
- 1997. 'Social networks and kinds of speech-community event', pp. 209–61 of *Archaeology and Language*, vol. 1: *Theoretical and Methodological Orientations*, edited by R. Blench and M. Spriggs. London: Routledge.
- 2002. 'Takia', in *The Oceanic Languages*, edited by J. Lynch, M. D. Ross, and T. Crowley. London: Curzon Press.
- with PAOL, J. N. 1978. *A Waskia Grammar Sketch and Vocabulary*. Pacific Linguistics B-56. Canberra: Pacific Linguistics.
- SASSE, H.-J. 1985. 'Sprachkontakt und Sprachwandel: die Gräzisierung der albanischen Mundarten Griechenlands', *Papiere zur Linguistik* 32.37–95.
- 1992. 'Language decay and contact-induced change: similarities and differences', pp. 59–80 of *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*, edited by M. Brenzinger. Berlin: Mouton de Gruyter.

- THOMASON, S. G., and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- THURGOOD, G. 1996. 'Language contact and the direction of internal drift: the development of tones and registers in Chamic', *Language* 72.1–31.
- THURSTON, W. R. 1982. *A Comparative Study of Anêm and Lusi*. Pacific Linguistics B-83. Canberra: Pacific Linguistics.
- 1987. *Processes of Change in the Languages of North-Western New Britain*. Pacific Linguistics B-99. Canberra: Pacific Linguistics.
- 1989. 'How exoteric languages build a lexicon: esoterogeny in West New Britain', pp. 555–79 of *VICAL 1, Oceanic Languages: Papers from the Fifth International Conference on Austronesian Linguistics*, edited by R. Harlow and R. Hooper. Auckland: Linguistic Society of New Zealand.
- 1994. 'Renovation and innovation in the languages of north-western New Britain', pp. 573–609 of Dutton and Tryon (1994).
- WATERS, B., TUOMINEN, S., and REHBURG, J. 1993. 'Takia, a Western Oceanic Language of Papua New Guinea'. Manuscript, Summer Institute of Linguistics, Ukarumpa.
- WEINREICH, U. 1963 [1953]. *Languages in Contact*. The Hague: Mouton.
- WURM, S. A. 1976. Editor of *New Guinea Area Languages and Language Study*, vol. 2: *Austronesian Languages*. Pacific Linguistics C-39. Canberra: Pacific Linguistics.

Areal Diffusion, Genetic Inheritance, and Problems of Subgrouping: A North Arawak Case Study

Alexandra Y. Aikhenvald

1. Introduction

1.1. THE PROBLEM OF LINGUISTIC SUBGROUPINGS

To establish subgroupings within a language family one needs to recognize ‘a set of changes common to a particular subgroup which has occurred between the period of divergences of the family as a whole and that of the subgroup in question’ (Greenberg 1953: 49). The shared changes should be significant and fairly unusual, not the sorts of changes that recur all over the world and could well have happened independently in each of the languages (e.g. palatalizing a velar consonant next to a front vowel). Note that the criterial features must be shared *innovations*. If a number of languages within a given family share retentions from the proto-language this does not require a period of shared development and does not constitute evidence for subgrouping. Shared loss (e.g. loss of a system of possession markers, pronominal suffixes, or tense inflections) is also not a good criterion for subgrouping. What may be significant is the way the loss is replaced; if each of the languages loses tense inflection on verbs but then makes reference to time through an innovated system of auxiliaries (with the auxiliaries being cognate in form) then this shared innovation does provide evidence for their constituting a subgroup.

If two languages show certain similarities one must decide whether these are

I am grateful to all my teachers of the Arawak languages—Cândido, Graciliano, Olívia, Jovino, and José Brito (Tariana); Humberto Baltazar and Pedro Ângelo Tomas (Warekena); the late Candelário da Silva (Bare); Afonso, Albino, and João Fontes, Celestino da Silva and Cecília and Laureano da Silva, Januário Paiva and the late Marcília Rodrigues (Baniwa); and to Raimunda Palikur (Palikur). Special thanks go to R. M. W. Dixon, to Silvana and Valteir Martins, to Elias and Lenita Coelho, and to Diana and Harold Green.

due to chance (or to universal tendencies of sound symbolism) or to borrowing in a contact situation, or to genetic inheritance. It is only the last kind of similarity that can provide evidence for subgrouping.

The kinds of similarities that may provide evidence for subgrouping include sharing a whole series of structural changes—such as a series of phonological shifts, unusual patterns of analogical restructuring, common development of a morphological construction, and especially common morphological and lexical innovations. As Greenberg (1953: 54) pointed out, ‘the mere counting of the number of cognates shared, without attention to morphological or phonological evidence and without consideration of the general distribution of each form for its bearing on the question of innovation, is a relatively crude method which disregards much relevant evidence’.

The problem of subgrouping is exceptionally difficult in the context of Amazonian languages, due to frequent migrations and language contacts which resulted in extensive borrowing and grammatical change. The latter means restructuring of grammar in agreement with areally spread patterns, reinterpreting existing morphemes, and introducing new morphology (often by grammaticalizing lexical items).

The Amazon basin comprises around three hundred languages which include around fifteen language families and a fair number of isolates. Six major linguistic families spoken there are Arawak, Tupí, Carib, Pano, Tucano, and Jê; among the most well-known smaller families are Makú, Bora-Witoto, Harakmbet, Arawá, and Guahibo.

Since all the major language families are highly discontinuous, so that the language map of South America resembles a patchwork quilt where half a dozen colours appear to be interspersed at random, this produces a linguistic situation unlike those found in the other parts of the world, creating unusual and extraordinary difficulties for distinguishing between similarities due to genetic retention and those due to areal diffusion. This poses particular problems for the recognition of subgroups within the major families.

Languages of Amazonia share a number of structural features, enough to be considered a large linguistic area, which includes several distinct subareas (cf. Payne 1990). Features of the ‘Amazonian’ linguistic type are summarized in Aikhenvald and Dixon (1998). Quite a few grammatical phenomena are shared by some but not all Amazonian languages. Among such phenomena are phonological tones, evidentiality, lack of rhotic or lateral phoneme, complex classifier systems, and nominative–accusative patterns. Instead, they can be detected as characteristic of several unrelated languages in certain regions, and can help to establish areal characteristics of each of these. The Amazon can be compared to a set of Chinese boxes of linguistic areas and subareas, included within each other. The distribution of these traits indicates, for instance, that there are sufficient reasons to consider the area covering north-eastern Peru and adjacent regions of Colombia, Brazil, and Venezuela a linguistic ‘subarea’ (see §4.1 and §4.3). One of

the dangers for a student of Amazonian languages lies in a number of shared monosyllabic morphemes spread all over the continent (see Payne 1990), e.g. negative *ma* or *na*, oblique case *-ri* or *-li*, *-ne*, possessive *-ri*, *-ni*, or *-i*, subordinate or relative marker *-ka*, or *ka-*, valency changing *-ta*, *-ka*, *-sa*, *-na*, or *-ma*, etc. For these morphemes, it is hard to determine whether their spread is due to borrowing, genetic affinity, diffusion of any sort, or just coincidence. See Aikhenvald (2002). Shared innovations can often be interpreted as the result of areal diffusion once speakers of a language migrate into a certain linguistic area.

The main purpose of this chapter is to analyse the problem of subgrouping within the Arawak language family, geographically the most extensive in South America. There is by now hardly any doubt as to the limits of the family, and one can 'easily' discern a smallish archaic 'nucleus' of Proto-Arawak grammar and lexicon (§2). However, individual languages spoken in separate (but often quite close) locations show an amazing degree of structural and formal divergence in the area of grammatical morphemes, even if they share 50%, 60%, or even 70–80% lexicon. The shared morphology hardly goes beyond the archaic 'nucleus' reconstructable for Proto-Arawak. A number of Arawak languages spoken to the north of the Amazon are looked at from this point of view in §3.

The differences between languages spoken in distinct locations can in many cases be explained by convergence with neighbouring languages that are not genetically related. However, this may happen in different ways partly depending on the sociolinguistic situation and the way languages 'treat' foreign material; in other words, whether they favour actual borrowing or not (cf. Sapir 1921: 196–7). The core of this chapter lies in the two case studies in §4 for which §2 and §3 provide a background.

In §4.1 I discuss restructuring of an Arawak language under areal pressure without lexical borrowing. Tariana, spoken in the linguistic area of the Vaupés basin (on the border between Colombia and Brazil) and known for its obligatory multilingualism due to linguistic exogamy, underwent drastic restructuring under areal pressure from the genetically unrelated Tucano languages. There is a strong cultural inhibition against lexical borrowing which is viewed as 'mixing languages'. Thus, the areal influences involve just the calquing of Tucano patterns.

In §4.2 I consider a rather different case—that of Resígaro, now spoken in north-eastern Peru. Resígaro has been restructured beyond recognition under pressure from genetically unrelated languages of the Bora-Witoto family, all of whose speakers are bi- or trilingual. There are no inhibitions about any type of borrowing; as a result, loans into Resígaro include 'core' and 'non-core' lexicon, some pronouns, and bound nominal morphology.

At the same time, Resígaro shares around 60% lexicon with Tariana and with Baniwa, an Arawak language which shares a large number of cognates with Tariana but is spoken outside the Vaupés area. These features can be attributed to the fact that Tariana, Tucano, Baniwa, Bora-Witoto, and Resígaro are all spoken in a larger linguistic area. There are also a number of shared

innovations between Tariana, Baniwa, and Resígaro which may point to a period of common development. This agrees with the information (albeit scarce) which we have concerning the recent migrations of Resígaro, Bora, Ocaina, and Witoto from Colombia to Peru. It is possible that Tariana, Baniwa, and Resígaro did form a subgroup at one time; however, massive borrowing and areal diffusion from the dominant languages has contributed to 'obscure' the actual subgrouping. In all these cases, we have little, if any, idea about the time depth of language contact; we suspect that the depth of Tariana–Tucano contact must have been fairly shallow.

The ultimate explanation for this lies in the existence of a limited stock of genetically inherited morphemes, overlaid by vast influxes of areally diffused patterns, due to intensive and prolonged contact with neighbouring languages. A similar case can be made for some other families within the North Amazon, for example, the Makú language family (§5).

This example from South America shows how massive multilingualism and language contact under distinct sociolinguistic conditions result in different kinds of restructuring of language systems.

2. The Arawak family

2.1. GENERAL FACTS ABOUT THE FAMILY

Languages of the Arawak family are spoken in at least six locations south of the Amazon, and in over eleven locations in the north (see the list of Arawak languages in Aikhenvald 1999b). This family spans four countries of Central America—Belize, Honduras, Guatemala, Nicaragua—and eight of South America—Bolivia, Guyana, French Guiana, Suriname, Venezuela, Colombia, Peru, Brazil (and formerly Argentina and Paraguay). There are about forty living Arawak languages. The genetic unity of Arawak languages was first recognized by Father Gilij in 1783, three years before Sir William Jones's famous statement about Indo-European (see Aikhenvald 1999b for further details). The recognition of the family was based on a comparison of pronominal cross-referencing prefixes in Maipure, a now extinct language from the Orinoco Valley, and Moxo from Bolivia. Gilij named the family Maipure. Later, it was 'renamed' Arawak by Brinton (1892) after one of the most important languages of the family, Arawak (or Lokono), spoken in the Guianas. This name gained wide acceptance during the following decades (see Aikhenvald 1999b). The limits of the family were established by the early twentieth century.

Comparative and historical studies of the Arawak family have a long history. The first truly scientific reconstruction of Proto-Arawak phonology—over two hundred lexical items and a few grammatical morphemes—was published by Payne (1991). However, his tentative subgrouping of Arawak languages—which is based on lexical retentions, rather than on innovations—remains open to debate.

Reconstruction, internal classification, and subgrouping of Arawak languages are still a matter for discussion.¹

2.2. WHAT WE KNOW ABOUT PROTO-ARAWAK

There is a common Arawak morphological nucleus shared by all or almost all languages. See §2.2.1. Common Arawak lexicon is discussed in §2.2.2.

Arawak nouns preserve more common Arawak categories than do verbs. In most cases, nouns are also simpler than verbs, and more nouns than verbs can be reconstructed for the proto-language.

In reconstructing Proto-Arawak morphology and lexicon, it is difficult to go beyond the common ‘nucleus’ outlined here since most other categories differ greatly in their realization and in their meaning.

2.2.1. *Common Arawak grammar*

All Arawak languages are predominantly head-marking, polysynthetic to varying extents, and predominantly agglutinating, with some fusion. There are no case markers for core syntactic functions (but see §4.1 and Aikhenvald 1999b). They are mostly suffixing with only a few prefixes. Prefixes are rather uniform across the family, while suffixes are not. Free morphemes often get grammaticalized as bound morphemes (e.g. adpositions become applicative markers, and verbal roots become aspect markers; see Aikhenvald 1999b). This creates difficulties for reconstructing morphology.

Arawak languages south of the Amazon (‘South Arawak’) have a more complex predicate structure than those north of the Amazon (‘North Arawak’) (see Aikhenvald 1999b). This difference may be due to areal diffusion since most Arawak and non-Arawak languages of south Amazonia are more polysynthetic than those in the north.

The common Arawak archaic ‘nucleus’ consists of (a) cross-referencing affixes and personal pronouns, (b) gender, (c) number, (d) possession markers on nouns, (e) attribution and negation. There is also one shared suppletive form (see f). Other categories show different marking (see discussion in Aikhenvald 1999b).

- (a) CROSS-REFERENCING AFFIXES. Cross-referencing affixes are cognate across the family—see Table 1, and Aikhenvald (1999b: 88). All the verbs are divided into transitive, active intransitive, and stative intransitive. Suffixes or enclitics are used to cross reference O/S_o. Prefixes are used for cross-referencing A/S_a and possessor (unlike Tupí and Carib languages where the possessor is the

¹ The putative studies of ‘Arawakan’ by Matteson (1972), Noble (1965), Oliver (1989), and others are deeply flawed. Unfortunately, these have been adopted as the standard reference for the classification of Arawak languages, especially among anthropologists, archaeologists and geneticists, influencing ideas on a putative proto-home and migration routes for ‘Proto-Arawakan’ (cf. Tovar and de Tovar 1984).

TABLE 1. A/S_a/possessor prefixes and O/S_o suffixes/enclitics

person	prefixes		suffixes	
	sg	pl	sg	pl
1	<i>nu-</i> or <i>ta-</i>	<i>wa-</i>	<i>-na, -te</i>	<i>-wa</i>
2	<i>(p)i-</i>	<i>(h)i-</i>	<i>-pi</i>	<i>-hi</i>
3nf	<i>ri-</i> , <i>i-</i>	<i>na-</i>	<i>-ri, -i</i>	<i>-na</i>
3f	<i>thu-</i> , <i>ru-</i>	<i>na-</i>	<i>-thu, -ru, -u</i>	<i>-na</i>
'impersonal'	<i>pa-</i>	—	—	—
dummy O/S _o	—	—	<i>-ni</i>	—
non-focused A/S _a	<i>i-</i> / <i>(a-?)</i>	—	—	—

same as O/S_o). The Proto-Arawak morphological split-ergativity marked with cross-referencing affixes involves the following:

A = S_a—cross-referencing prefixes

O = S_o—cross-referencing suffixes or enclitics

The form of the first person pronoun *nu-* vs. *ta-* provides a division of languages into Ta-Arawak in the Caribbean (Lokono, Guajiro, Añun, Taino) and the rest (known as Nu-Arawak).

A four-person system can be reconstructed for Proto-Arawak. A prefix for non-focused or indefinite possessor and A/S_a is found mostly in the languages north of the Amazon, e.g. Palikur *i-wan-ti* (INDEF-arm-NPOSS) 'an arm', *i-nar-ti* 'a mother' (Green and Green 1972: 52), Baniwa *i-hwida-ti* 'a head'. However, the prefix *i-* may well be a shared archaism rather than an innovation (cf. the nominalizing prefix *i-* in Waurá, spoken south of the Amazon in Xingú park). Exclusive/inclusive is atypical (see §4.2, on Resígaro).

- (b) GENDER. Most Arawak languages distinguish two genders—masculine and feminine—in cross-referencing affixes, in personal pronouns, in demonstratives, and in nominalizations (e.g. Palikur *amepi-yo* 'thief' (woman), *amepi-ye* 'thief (man)', Tariana *nu-phe-ri* 'my elder brother', *nu-phe-ru* 'my elder sister'). Typical pronominal genders are feminine and non-feminine. No genders are distinguished in the plural. The markers go back to Proto-Arawak third person singular cross-referencing: feminine *-(r)u*, masculine *-(r)i* (see Aikhenvald 1999b: 83–4 for further details).

A number of languages also have complicated systems of classifiers (Aikhenvald 1994c). Unlike genders, these show great diversity from one language to another in semantics and form and appear to have developed on the level of individual languages.

- (c) NUMBER. All Arawak languages distinguish singular and plural; plural is optional (unless the referent is human), with the markers: **-na/-ni* 'animate/human plural', **-pe* 'inanimate/animate non-human plural'.
- (d) POSSESSION. Nouns divide into inalienably and alienably possessed. Inalienably possessed nouns include body parts, kinship nouns, and a few other nouns, e.g. house, louse. In some North Arawak languages deverbal nominalizations belong to this class. Both types of possession are marked with prefixes (A/S_a). INALIENABLY POSSESSED NOUNS have an 'unpossessed' form (called 'absolute' by Payne 1991: 379) marked with the suffix **-ŋi* or **-hV*, e.g. Pareci *no-tiho* 'my face', *tiho-ti* '(someone's) face' (Rowan and Burgess 1979: 79); Baniwa *nu-hwida* 'my head', *i-hwida-ti* (INDF-head-NPOSS) 'someone's head'. ALIENABLY POSSESSED NOUNS take one of the suffixes **-ne/ni*, **-te*, **-re*, **-i/-e* (Payne 1991: 378), or **-na*. These suffixes are also used as nominalizers.
- (e) ATTRIBUTION AND NEGATION. Most Arawak languages have a negative prefix *ma-* and an attributive-relative prefix *ka-*, e.g. Piro *ka-yhi* (ATTR-tooth) 'having teeth', *ma-yhi* (NEG-tooth) 'toothless' (Matteson 1965: 119), Bare *ka-witi-w* (ATTR-eye-F) 'a woman with good eyes', *ma-witi-w* 'a woman with bad eyes, blind'.
- (f) SHARED IRREGULAR FORMS OF THE NOUN 'HOUSE'. Most Arawak languages show irregular alternations between cognate non-possessed and possessed forms of the word 'house'. These are either (i) different stems: **pe* 'house: non-possessed' (either with a non-possessed marker **-ŋi*, or without it), **pana/i* 'house: possessed' (Payne 1991: 408), as in Waurá *pāi*, *-pina*, Palikur *pai-t*, *gi-vin* (his-house); or (ii) vowel alternations: **pan(i)* 'non-possessed', **-pana* 'house of', as in Warekena *pani-fī*, *pane*, Bare *phani*, *-bana*, Baniwa *pan-ti*, *-pana*, Tariana *pani-si*, *-pana*, Resígaro *panii-tsi*, *-paanu/paana* and others.

2.2.2. Common Arawak lexicon

The common Arawak lexicon (cf. Payne 1991) includes mostly nouns. There are quite a few body parts, fauna, flora, and artefacts, e.g. **maka* 'hammock', frequently grammaticalized as a classifier for stretched things (borrowed into many European languages as a word for 'hammock'). Only a few verbs can be reconstructed since most verb roots are monosyllabic and have undergone numerous phonological changes, e.g. **kau* 'arrive' (Payne 1991: 394), **pi(da)* 'sweep', **po* 'give', or **da* 'give' (the latter could have given rise to Tariana, Baniwa, Piapoco *-a* 'give, say, go'). The reliable disyllabic verbal reconstructions are **(i)ya* 'cry', **kama* 'be sick, die', **itha* 'drink', **ara* 'fly', **kema* 'hear, understand', **kiba* 'wash', **nika* 'eat', **dima* 'stand' (and possibly **kika* 'dig'). In the history of individual languages a thematic syllable (often of the same origin as a valency changing affix) is often added to a monosyllabic root and gets fused with it, obscuring

the reconstruction. Most languages have just the numbers ‘one’ (PAr **pa-*; also used to mean ‘someone, another’) and ‘two’ (PAr **(a)pi* and **yama*: Payne 1991).

Lexical items reconstructable as synonyms are scattered within the family; they can hardly give any information as to subgrouping. For instance, two terms for ‘fish’ can be reconstructed for Proto-Arawak: **kopaki* is found in Terêna, Waurá, Yawalapiti, and Pareci south of the Amazon and in Baniwa, Tariana, Achagua, Piapoco, and Wapishana north of the Amazon, while **hima* is found in Chamicuro, Campa, and Ignaciano south of the Amazon, and in Lokono, Guajiro, Palikur, and Yavitero north of the Amazon. The term **(a)pi* ‘two’ is found in Ignaciano, Moxo, Campa, and Aputina south of the Amazon, and in Palikur, Yavitero, Piapoco, and Bare north of the Amazon, while the other term for ‘two’, **yama*, is found in Waurá and Yawalapiti south of the Amazon, and in Achagua, Yucuna, Tariana and Baniwa north of the Amazon.

3. Arawak languages north of the Amazon: grammatical and lexical comparisons

Arawak languages spoken in eleven distinct locations in northern Amazonia (see Map 2 in Aikhenvald 1999a) are: Wapishana, Palikur, Achagua, Piapoco, Yucuna, Baniwa, Tariana, Warekena, Bare, Bahwana, and Resigaro.

Table 2 shows shared vocabulary percentages between these languages, as well as vocabulary shared with Guajiro (spoken on the Guajiro Peninsula on the

TABLE 2. Shared vocabulary percentages

Tariana														
77–80 Baniwa*														
55	52–6	Resigaro												
56	53	48	Piapoco											
43	50	45	47	Yucuna										
53	53	39	72	43	Achagua									
41	41	27	30	30	28	Bare								
32	36	26	25	25	24	36	Warekena							
33	32	35	30	36	29	39	39	Bahwana						
33	30	26	30	30	25	29	22	33	Palikur					
35	35	34	32	24	30	35	29	31	43	Wapishana				
28	27	25	25	30	25	33	31	33	34	30	Pareci			
22	23	20	22	21	20	22	17	21	30	23	25	Ignaciano		
30	26	20	26	17	24	22	25	29	34	24	31	29	Iñapari	
25	23	21	20	25	23	29	24	23	23	24	20	16	17	Guajiro

*Baniwa of Içana (also known as Baniwa-Kurripako) is a dialect continuum. All the dialects are mutually intelligible and share 90–95% vocabulary and most grammar; the figures in Table 2 reflect the average number of cognates. The closest dialect to Tariana is that of Hohôdene.

Caribbean coast) and with three languages spoken in southern Amazonia—Pareci, Iñapari, and Ignaciano. These percentages were calculated on the basis of Swadesh's hundred-word list, as well as a standardized list of 375 words used by the Summer Institute of Linguistics (see Huber and Reed 1992, Allin 1975, and Aschmann 1993). The results obtained were basically the same, thus indicating that the 'core' vs. 'non-core' vocabulary distinction is not crucial for Arawak languages (cf. similar results for Australian languages: Dixon, this volume, and for Papuan languages: Comrie, p.c.). Shared vocabulary indicates some connection between Tariana, Baniwa, Resígaro, Piapoco, Yucuna, and Achagua. However, all the Arawak languages spoken north of the Amazon display significant differences in grammatical forms and categories.

All the North Arawak languages preserve a certain amount of the common Arawak morphological 'nucleus'. Cross-referencing suffixes (Table 1 and (a) in §2.2.1) is where the languages differ most: Baniwa, Warekena, and Bahwana have a full set of cross-referencing suffixes and prefixes and preserve the Arawak split-ergative pattern, while Wapishana, Achagua, Yucuna and Piapoco use suffixes only for third person. Of these, only Piapoco is split-ergative. Bare, Resígaro, and Tariana have no cross-referencing suffixes, and no split-ergativity of the Arawak type. Baniwa and Piapoco share one innovation in cross-referencing suffixes: Baniwa innovated third person singular masculine *-ni*, feminine *-nu*; and Piapoco innovated third person singular *-ni* (cf. PAr **-ni* 'dummy O/S_o' in Table 1). Piapoco is unusual in that it lost second person plural *i-*, replacing it with second person singular *pi-* and a plural marker *-cue*. It lost the word-initial *r-* in subject cross-referencing third person masculine singular **ri-*, feminine **ru-*, which give *i-* and *u-* respectively, since there are no rhotics or laterals in word-initial position. Baniwa also has elements of fluid S marking.

North Arawak languages show considerable divergence in most grammatical categories. For instance, classifiers vary considerably; the differences in other categories are discussed in Aikhenvald (2002). The number of cognates in classifiers is shown in Table 3. In calculating cognates we do not consider gender markers and feminine and masculine classifiers which contain Proto-Arawak gender markers found in all the languages.

Numeral classifiers are found everywhere except Piapoco, Bare, and Wapishana. In Yucuna, Achagua, Bahwana, and Warekena there are only numeral classifiers. Tariana, Baniwa, Resígaro, and Palikur have noun classes, while verbal classifiers are found just in Tariana, Baniwa, and Palikur (see Aikhenvald 1994b). Possessive classifiers are found in Tariana and in Baniwa (where they are restricted to predicative possession). Tariana and Resígaro are the only Arawak languages of the region to use classifiers with demonstratives, while Palikur is unique in having locative classifiers (see the typological parameters set out in Aikhenvald 2000a).

Most classifiers shared by more than two languages typically have cognates in Arawak languages south of the Amazon, e.g. Baniwa *-na/-ne, -nay* 'vertical objects, mammals', Tariana *-na* 'vertical objects', Achagua *-na* 'four-legged animals',

TABLE 3. Number of cognates in classifiers in North Arawak

Tariana (over 60 classifiers in the language)

31	Baniwa (44)						
5	7	Resígaro (56)					
5	5	0	Yucuna (8)				
5	5	3	2	Achagua (12)			
0	0	0	0	0	Warekena (6)		
4	3	2	1	2	0	Bahwana (20)	
1	1	2	0	0	0	0	Palikur (12)

Yucuna *-na* ‘mammal, tree’, Bahwana *-na* ‘long’ (bottle, pineapple), cf. Arawak south of the Amazon Yawalapiti *-na* ‘vertical objects’, or Baniwa, Tariana *-pi* ‘long thinnish objects’; cf. Waurá *-pi* ‘linear’, Terêna, Pareci *-hi* ‘long, thin objects’, Baure *-pi* ‘long objects’, Ignaciano *-pi* ‘long and thin objects’, Amuesha *-py* ‘long objects’ (PAR **-pi* ‘long thin objects’ from **api* ‘snake’). See Aikhenvald (2002) for further examples.

Thus, apart from the common Proto-Arawak ‘nucleus’, even languages which are lexically close share very few grammatical cognates. As a result, grammatical categories and their marking can hardly be used as a basis for subgrouping since they probably developed on the level of individual languages. In the next section we show how two Arawak languages north of the Amazon—which share a considerably high percentage of common lexicon—underwent restructuring of different kinds under different areal influences.

4. Case-studies in restructuring north of the Amazon

Drastic differences in morphology and grammatical structure between lexically close languages can often be explained by convergence with neighbouring genetically unrelated languages. Languages differ with respect to how they ‘treat’ loans: some favour the borrowing of actual forms, and some do not.

The two case studies in this section represent the two extremes. In §4.1 I discuss restructuring without any loans. Tariana spoken in the multilingual Vaupés basin underwent influence from Tucano languages, following Tucano-type grammaticalization paths and developing new morphology to match the categories found in Tucano. There is a cultural inhibition against lexical loans in the Tariana-Tucano speaking area.

In §4.2 I consider restructuring with heavy borrowing. Resígaro, dominated by Bora-Witoto languages for a long time, has borrowed quite a few ‘core’ and ‘non-core’ vocabulary items, grammatical markers, and a pronoun. Its grammatical system was restructured under Bora-Witoto influence.

Tariana, Baniwa, and Resígaro share a number of innovations; there are also a few structural similarities between Resígaro-Bora-Witoto, on the one hand, and

Tucano-Tariana and Baniwa on the other. These features may be indicative of areal diffusion in the past. Thus, evidence in favour of old diffusion areas helps obscure the genetic relationships among languages.

4.1. RESTRUCTURING WITHOUT LEXICAL BORROWING: TARIANA

4.1.1. *General remarks*

The Vaupés basin in north-west Amazonia is a linguistic area, with a convincing number of structural features shared by languages from two genetically unrelated families—Tariana (Arawak) and East Tucano (see Map 13 in Aikhenvald 1999c). These features are not found in Arawak languages spoken outside the area, and thus can be considered as diagnostic for areal diffusion. In some cases we are able to establish the direction of diffusion (see Sorensen 1967, Jackson 1974, Aikhenvald 1996a, b, 1999b, c). Baniwa is an Arawak language which shares quite a few lexical cognates with Tariana (see Table 2) but it is spoken outside the Vaupés region. It is instructive to compare Tucano, Tariana, and Baniwa.

Tariana is spoken in a very peculiar linguistic situation of obligatory multilingualism of the Vaupés basin, dictated by the principles of linguistic exogamy ('those who speak the same language with us are our brothers, and we do not marry our sisters'). The distinctive feature of the Vaupés linguistic area is the absence of lexical borrowings due to a strong cultural inhibition: 'language mixing' viewed in terms of lexical loans is condemned as culturally inappropriate, and is tolerated only as a 'linguistic joke'.

East Tucano languages spoken in the Vaupés basin are structurally and formally very similar. They share from 60% to 90% vocabulary (see Sorensen 1967 on the East Tucano linguistic type). However, whether this East Tucano profile is due to areal diffusion patterns or to the common genetic origin of East Tucano languages remains a problem which goes beyond the scope of the present discussion. (Solving this problem would involve a full reconstruction of Proto-East Tucano and its comparison with Proto-West Tucano and Proto-Tucano.)

The Tucano influence on Tariana phonology, grammatical structure, syntax, discourse organization, and semantics has been discussed at length in Aikhenvald (1996a, 1999b). Some of the points are illustrated in §4.1.2. The 'time depth' of the Vaupés area is discussed in §4.1.3. In §4.1.4 I present evidence in favour of the existence of a larger linguistic area comprising the Içana and the Vaupés basins.

4.1.2. *Tucano influence on Tariana*

Areal diffusion from East Tucano to Tariana involves:

- (a) emergence of new categories present in East Tucano but absent from Arawak, e.g. case-marking connected with topicality and the use of just one locative case, evidentials, verb compounding, and switch-reference;

TABLE 4. Grammatical relations in Tariana, Tucano, and Baniwa

Languages	Head or dependent marking	Ergative or accusative	Grammatical relations making
Tucano	head and dependent	accusative	cross-referencing and core cases
Tariana		traces of split S, but basically accusative	
Baniwa	head	split S, some fluid S	cross-referencing

TABLE 5. Core cases in Tariana and in the Tucano languages

Grammatical function	Discourse status	Tariana		Tucano	
		nouns	pronouns	nouns	pronouns
subject (A/S)	non-focused	- $\emptyset$		- $\emptyset$	
	focused	-ne/-nhe		-pe	
non-subject (non A/S)	non-topical	- $\emptyset$	-na	- $\emptyset$	-fe
	topical	-naku, -nuku		-fe	

(b) structural levelling of Tariana to agree with East Tucano syntactic structures and discourse techniques, and also obsolescence and subsequent loss of some categories that are not present in East Tucano languages.

As the result of diffusion, the typological profile of Tariana and the ways grammatical relations are marked are much more similar to the Tucano languages than to Baniwa—see Table 4.

Table 5 shows striking structural similarities in the core-case marking in Tariana, and in the Tucano languages. The forms are completely different (Tariana topical non-subject marker *-naku/-nuku* is most probably cognate with Baniwa locative case *-naku*).

The classifier system in Tariana also underwent restructuring to accord with Tucano patterns. Baniwa has a large closed system of forty-four classifiers, while Tariana has sixty plus a potentially unlimited set of repeaters (Aikhenvald 1994b). Tucano has a similar set. Tariana and Baniwa share only thirty-one classifiers (see Table 3); some of the Baniwa classifiers ‘survive’ in Tariana only as derivational affixes.

Unlike in Baniwa, classifiers in Tariana and Tucano are used with demonstratives and in possessive constructions, as in (1) and (2) below. Baniwa uses genders marked on demonstratives; classifiers are only used in predicative possession, as

in (3). Examples (1) and (2) illustrate structural isomorphism between Tariana and Tucano; the interlinear glosses are the same.

- (1) Tucano *ati-wi'i* *numio-ya-wi'i*
- (2) Tariana *ha-panisi* *inaru-ya-panisi*
 DEM:INAN-HOUSE woman-POSS-HOUSE
 'a woman's house'
- (3) Baniwa *hliehē* *panti* *inazu* *i-dza-dapana*
 DEM+NF house woman INDF-POSS-CL:HAB
 'This house is (a) woman's.'

Along similar lines, the system of oblique cases in Tariana was restructured following the Tucano model—see Table 6.

Other areas of morphosyntax which underwent a particularly strong Tucano impact are possession marking (see Aikhenvald 1999c, 1996a), verb compounding (see Aikhenvald 2000b), evidentiality and tense, constituent order and the obligatory use of overt noun phrases, switch-reference, complementation techniques, and discourse markers.

The constituent order in Tariana and Tucano is verb-final. In Baniwa and most other Arawak languages of the area it is either verb-medial or verb-initial. Most

TABLE 6. Oblique cases in Tariana, Tucano and Baniwa

Meanings	Tariana	Tucano	Baniwa	Reflexes of Baniwa markers in Tariana
Locative: general Locative: on surface of	-se 'locative'	-pi	-riku	-riku 'derivational suffix', -riku-se 'different subject switch-reference marker'
Directional			-naku	-naku, -nuku 'topical non-subject'
Ablative			-hre	-se 'locative' corresponds to both Baniwa -hre or -te by phonological rules
Perlative	none		-wa	-wa derivational suffix, e.g. kada-wa 'get dark'
Comitative/ instrumental	-(i)ne	-mẽ'ra	no case: adposition inai 'with'	-(i)ne 'with'
Double case marking	locative co-occurs with topical non-subject marker Ta -nuku, Tu -re		only locative cases co-occur	

North Arawak languages—like most other South American Indian languages—avoid sentences with two full noun phrases, especially when one of them is a free pronoun. Free pronouns are mainly restricted to emphatic function. In contrast, Tariana—similarly to Tucano languages—makes wide use of free pronouns. Also like the Tucano languages, Tariana has a well-developed system of switch-reference (same subject or different subject) in subordination.

However, it would be wrong to say that Tariana has absolutely no loans from Tucano, or any other neighbouring language (see discussion and examples of the rare loans from Tucano in Aikhenvald 1999a). Tucano influence on Tariana involves mostly calquing of patterns, sometimes accompanied by grammatical accommodation, that is, ‘syntactic deployment of a native morpheme on the model of a phonetically similar morpheme in the diffusing language’ (that is, the language which is the source of diffusion), as illustrated by Watkins (this volume) for the possible extension of native morpheme *-ske-* in Ionic Greek to mark iterative imperfective under the influence of the morpheme with the same shape in Hittite and Luvian. For instance, Tucano and other Tucano languages use *-ya* as a marker of imperative. Tariana has a phonologically similar morpheme *-ya* ‘emphatic’ which is used on imperative verbs. Similarly, Tucano has the marker *-ri* used for commands with a tinge of a ‘warning’ (make sure you don’t fall). Tariana has a relativizer *-ri* used in a wide variety of functions (this morpheme goes back to Proto-Arawak) which is also used in commands, with a similar meaning (see Aikhenvald (2002) for further discussion). ‘Grammatical accommodation’ of this sort can be explained by the fact that the native morphemes which undergo extension under Tucano influence do not sound as foreign, and so their existence does not go against the prohibition against borrowing and ‘language mixing’.

4.1.3. *The time depth of the Vaupés linguistic area*

There are reasons to believe that the Vaupés is a relatively young area. Tariana and East Tucano languages have probably been in contact for no more than about four hundred years. The settlement of East Tucano tribes on the Vaupés goes back somewhat further (cf. Nimuendajú 1982: 169–70). The other reasonably well-described linguistic areas of the world, e.g., the Balkans, Eastern Arnhem Land in Australia (see Heath 1978, 1981), Mesoamerica (Campbell, Kaufman, and Smith-Stark 1986), South Asia (Masica 1976), and the linguistic areas of North America north of Mexico (Sherzer 1976), such as the north-west coast, are considerably older than this.

As I have argued elsewhere (Aikhenvald 1996c), a study of types of Tariana place names shows that two of these types are predominantly monolingual—‘historical’ names which refer to places where the Tariana used to live in the remote past, and ‘mythological’ names which refer to the adventures of characters in origin myths. In contrast, place names which refer to actual dwelling sites are multilingual, and are usually calqued into several languages. Even when

TABLE 7. Properties shared by languages of the Içana-Vaupés

Properties shared	Vaupés		Içana	Outside this area	
	East Tucano	Tariana	Baniwa	Piapoco	Warekena Bare
pitch accent	yes	yes	yes	yes	no
topic advancing	yes	yes	yes	no	no
derivation					
possessive classifiers	yes	yes	yes (with possessive predicates)	no	no
possessive <i>-ya-</i> to which classifiers are attached	yes	yes	yes	no	no
classifiers with demonstratives	yes	yes	yes	no	no
several types of classifiers	yes	yes	yes	no	no

‘historical’ places also have names in languages other than Tariana they are never calque translations from one language into another. These properties of ‘historical’ and ‘mythological’ place names, unexpected in an environment of obligatory multilingualism, suggest that the Tariana might have arrived in the Vaupés from a mostly monolingual context, and that they have adopted multilingualism fairly recently.

4.1.4. *Içana-Vaupés as a linguistic area?*

We can now look at the larger area consisting of the Içana and the Vaupés basins. Besides cultural similarities widespread across this area, there are a number of linguistic properties shared by Baniwa and the languages of the Vaupés discussed above, but absent from other North Arawak languages. Table 7 summarizes these properties. Only one of these properties—pitch accent—is also shared by Piapoco (which is lexically close; see Table 2).

According to the Tariana origin stories, they came to the Vaupés from a tributary of the Içana river, probably the Aiary, where they lived together with the Baniwa (and where Baniwa is still spoken; see Aikhenvald 1996a, 1999c).

The existence of structural and even formal similarities shared by Tariana, Tucano languages, and Baniwa—but absent from Arawak languages of the area—indicates a certain amount of diffusion in an area which goes beyond the Vaupés into the basin of Içana and its tributaries. It is impossible to decide whether Tariana and Baniwa share a relatively high percentage of morphemes and lexicon due to areal diffusion or to genetic affinity, or to both.

4.2. RESTRUCTURING WITH LEXICAL BORROWING: RESÍGARO

Resígaro is a highly endangered language now spoken by just a few people in the far north-eastern corner of Peru in Puerto Isango and Brillo Nuevo, on the river Yaguasyacu, a tributary of Ampiyacu (which flows into the Amazon at Pebas) (Allin 1975, Loukotka 1968: 137). All the speakers of Resígaro have as their main language Bora and/or Ocaina (from the Bora-Witoto family). The linguistic situation is that of unilateral diffusion from Bora to Resígaro. (Little can be said about the diffusion from Ocaina, since there are no data available.)

The Bora-Witoto family consists of two main branches: Bora-Muinane and Witoto-Ocaina. A phonological and lexical reconstruction of Proto-Bora-Witoto has been published by Aschmann (1993). Grammatical information on Bora comes from Thiesen (1996), and on Witoto from Minor and Loos (1963) (also see Wise (1999) and references therein). Almost nothing is known about Ocaina. Grammatical and lexical data on Resígaro come from Allin (1975) (there are some additional data in Rivet and de Wavrin (1951)).

The information about the history of Bora, Resígaro, Witoto, and Ocaina is extremely scanty (see the overview in Allin 1975: 3–5). Whiffen (1915) encountered them on the banks of Japurá (Caquetá) to the north of the Cahuinari in Colombia. Even at that time the Resígaro were a minority (of 1,000 people, while the Bora had about 15,000 people). Presumably, they then moved from that location—which is notably closer to the Içana-Vaupés area than their modern one—further south to Peru. The Map shows the current location of Resígaro and their location according to Whiffen (1915) (the likely direction of their migration is indicated with an arrow).

Lexical borrowings from Bora are discussed in §4.2.1, and the borrowed grammar in §4.2.2. Resígaro shares a high percentage of lexicon with Tariana and with Baniwa (see §4.3).

4.2.1. Lexical borrowings in Resígaro

Resígaro shows a large number of loans from Bora. A lexical comparison of 100 ‘core vocabulary’ items and of 218 non-core items between Resígaro and Bora and Witoto shows that about 24% are loans.

Table 2 showed the shared lexicon between Resígaro and a number of other Arawak languages. Notably, the lexical percentage between Bora and Resígaro (approximately 24–6%) is the same or slightly higher than that between Resígaro and Palikur (26%), Pareci (25%), Ignaciano (20%), Iñapari (20%), or Guajiro (21%). Allin (1975) even classified Resígaro as related to Bora. However, Table 2 shows that Resígaro shares over 50% with Tariana and with Baniwa.

Lexical loans contain a few verbs and body parts, plus words for ‘fish’, ‘hill’, etc. (As I will show in §4.2.2, nouns denoting body parts are systematically borrowed as classifiers.) Some examples of ‘core’ vocabulary borrowed from Bora into Resígaro are given below.



MAP. Geographical distribution of Arawak and Tucano languages discussed in this chapter

- (i) Resígaro *-eʔhépe* ‘teeth’, Muinane Witoto *ípe*, Proto-Witoto **pe* (cf. PAR **nene*; note Resígaro *-onéné* ‘front teeth’);
- (ii) Resígaro *ámoogí* ‘fish’, Bora *amóópe*, Proto-Bora-Muinane **amóóbe* (cf. PAR **kopaki*, Tariana, Baniwa *kuphe*, Piapoco *cubái*).

Examples of non-core loans include

- (a) NATURAL PHENOMENA, e.g.
 - (i) Resígaro *kočhívu* ‘wind’, Bora *kʰítʰe-pa*, Proto-Bora-Muinane **kítʰe-ba* (cf. Achagua, Piapoco *kauli*, Tariana, Baniwa *kare*);
 - (ii) Resígaro *teéʔi* ‘river’, Bora *tʰeé-ʔi*, Proto-Bora-Muinane **teé-ʔi* (cf. PAR **huni* ‘water, river’, Tariana, Baniwa, Piapoco *uni* ‘water, river’);
- (b) names for INSECTS and ANIMALS, e.g.
 - (i) Resígaro *heété* ‘fly’, Bora *éétʰepa*, Proto-Bora-Muinane **ééteba* (Tariana, Baniwa *pupu*, Piapoco *pulederi*);
 - (ii) Resígaro *paagáu* ‘spider’, Bora *paáwájí* (Allin 1975), Proto-Bora-Muinane **páaga-xi* (Aschmann 1993: 139) (Achagua, Piapoco, Baniwa, Tariana *e:ni*);
 - (iii) Resígaro *hoʔbu* ‘capybara’, Bora *óhbá* (Allin 1975), Proto-Bora-Muinane *ʔoba* (Aschmann 1993: 137) (Tariana *hemasiéere*);
 - (iv) Resígaro *piíme* ‘ant’, Bora *piimyebá*, Muinane *ʔiimo*, Proto-Bora-Muinane **ptimeba* (Aschmann 1993: 139) (Tariana has many terms for ‘ant’, none of which is a cognate with this form in Resígaro);
- (c) ARTEFACTS, e.g.
 - (i) Resígaro *maáníʔúmi* ‘mask’, Bora *máhníi*, Proto-Bora-Muinane *mááʔnií* (*ʔimo*) (Aschmann 1993: 143) (Tariana, Baniwa *-maka* ‘classifier: cloth-like’).

In all these cases the direction of loans is from Bora to Resígaro, and not the other way round. This is confirmed by the existence of Proto-Bora-Muinane and sometimes Proto-Bora-Witoto forms (see Aschmann 1993).

Resígaro *sa-* ‘one’ was borrowed from Bora *tsa* (cf. PAR **pa-* whose reflexes are found in Tariana and Baniwa), and *migaa-* ‘two’ was borrowed from Bora *mípnéé/mihaa-CL* (cf. PAR **-yama*).

Earlier materials (collected by de Wavrin, possibly, in the early 1930s: see Rivet and de Wavrin 1951: 238, Loukotka 1968: 137) contain different forms, of an Arawak origin, for numbers ‘one’—Resígaro *apa(ba)pene*, cf. PAR **pa*—and ‘two’—Resígaro *ʔe(i)tza:mo*, *itsā(a)mə*, cf. PAR **yama* (the other Proto-Arawak term for ‘two’, **(a)pi*, is unrelated). This may indicate either that the Resígaro borrowed the Bora numbers after the 1930s, or that de Wavrin collected his data from another dialect of Resígaro (probably now extinct).

4.2.2. Borrowed grammar in Resígaro

Resígaro phonology has been influenced by Bora in the following ways:

- (a) Resígaro has two phonological tones (high and low), just like Bora (most other Arawak languages do not have tones);
- (b) there is a phonemic glottal stop, like Bora, Ocaina, and Witoto but unlike most Arawak languages north of the Amazon;
- (c) Resígaro has the syllable structure $(C_1)V(C_2)$ with only *h* and *ʔ* in C_2 position (similarly to Bora, Ocaina, and Witoto, but unlike Arawak).

Bora influence on Resígaro grammar involves the emergence of new categories found in Bora but atypical of an Arawak language and expressed with borrowed morphemes (see A below); and structural levelling of Resígaro and Bora which involves the calquing of categories (see B).

A. BORROWED MORPHEMES: NOMINAL MORPHOLOGY AND PRONOUNS.

Borrowed bound morphemes include one pronoun, number markers, classifiers, and oblique cases. The independent pronouns and cross-referencing prefixes in Resígaro (where they are mostly used to mark A/S_a and possessors of inalienably possessed nouns) are compared to Bora in Table 8 (data from Allin 1975: 116–17, Thiesen 1996: 33; borrowed morphemes are in bold type). Borrowing of a pronominal form is quite unusual (though attested elsewhere—see Campbell 1997).

TABLE 8. Pronouns in Resígaro and in Bora

	Resígaro		Bora		
	pronouns	prefixes	pronouns	prefixes (poss)	prefixes (subject)
1sg	<i>nó</i>	<i>no-</i>	<i>oó</i>	<i>ta-</i>	
2sg	<i>phú, pha</i>	<i>p-</i>	<i>uú</i>	<i>di-</i>	
3sg m	<i>tsú, tsá</i>	<i>gi-</i>	<i>diibye</i>	<i>i-, áádi-</i>	
3sg f	<i>tsó, tsó</i>	<i>do-</i>	<i>diille</i>		
iinc du m	<i>fa-musi</i>	<i>f-/_h</i>	<i>meé</i>	<i>me-</i>	<i>me-</i>
		<i>va/- elsewhere</i>			
iinc du f	<i>fa-mupi</i>				
1exc du m	<i>muu-musi</i>	<i>múu-</i>	<i>muhtsi</i>		
1exc du f	<i>muu-mupi</i>		<i>muhpi</i>		
2du m	<i>ha-musi</i>	<i>hú-, i-</i>	<i>á-muhtsi</i>	<i>amuʔá-</i>	<i>me-</i>
2du f	<i>ha-mupi</i>		<i>á-muhpi</i>		
3du m	<i>na-musi</i>	<i>n/_h</i>	<i>diityé-tsi</i>	<i>aát^hje-</i>	
3du f	<i>na-mupi</i>	<i>na-</i>	<i>diityé-pi</i>		
iinc	<i>faʔa, fú, fa</i>	<i>f/ua-</i>	<i>meé</i>	<i>me-</i>	<i>me-</i>
1exc	<i>muu-ʔa, muu</i>	—	<i>muúha</i>		
2pl	<i>ha-ʔa, hu</i>	<i>i-</i> (imperative)	<i>ámuúha</i>	<i>amuʔá-</i>	<i>me-</i>
3pl	<i>na-ʔa, hná</i>	<i>na-</i>	<i>diitye, aátyé</i>	<i>aát^hje-</i>	

Unlike most other Arawak languages but like the Bora-Witoto group, Resígaro has inclusive versus exclusive opposition in first person non-singular, and dual number. The first person exclusive pronoun *muuʔa* has been borrowed from Bora, and subsequently reanalysed as consisting of a prefix *muu-* and a particle *-ʔa*, following the analogy of other non-singular pronouns, such as *na-ʔa* 'third person plural' and *fa-ʔa* 'first person inclusive'. The dual markers feminine *-mupi*, masculine *-musi* (also from Bora) combine with *muu-* reanalysed as a bound form. Unlike other pronouns, the first person exclusive has no corresponding prefix used with nouns and with verbs.

Resígaro borrowed dual number markers used with human nouns (m. *-musi*, f. *-mupi*), with body parts and with classifiers, and also the marker for animate plural (Resígaro, Bora *-mu*) (Allin 1975: 164 and Thiesen 1996: 123–9).

Resígaro and Bora have masculine and feminine distinctions in the first, second, and third person dual, but not in the plural. This typologically rare pattern is a feature of Bora and most Witoto languages which diffused into Resígaro (see Aikhenvald 2000a: 246, 387).

The Bora influence on the Resígaro CLASSIFIERS involves borrowing of bound morphemes and grammaticalization of borrowed free morphemes as classifiers. Semantic and formal principles of noun classification in Resígaro are very similar to those in Bora, Witoto, Tucano, and Tariana (see §4.4). While two genders are distinguished in verbal cross-referencing, classifiers are used with demonstratives, numerals, and adjectival modifiers, and on head nouns as singulative markers (Allin 1975: 153 ff.).

Bora has over four hundred classifiers (Thiesen 1996: 102). Resígaro has around fifty-six (Allin 1975: 154 ff.); only eight or nine of these have an etymology in Arawak languages (mostly Baniwa); and thirty-six have been borrowed from Bora. These fall into three groups.

- (a) CLASSIFIERS WHICH CORRESPOND TO BOUND MORPHEMES IN BORA. Twenty of the borrowed classifiers are used only as classifiers in Bora. They categorize nouns in terms of their shape and form, e.g.
- (i) Resígaro *-gú* 'long and flat', Bora *-kʷaá* (classifier which appears in words for 'finger', 'toe'), Proto-Bora-Muinane *-gai* (Aschmann 1993: 131);
 - (ii) Resígaro *-hí* 'round and flat', Bora *-ji* 'round, elongated, circular, like a disc' (Thiesen 1996: 102), e.g. Resígaro *hípo-hí* 'earth (a combination of a reflex of the PAr **hipa-γ* and a classifier), Bora *únú-jí*, Proto-Bora-Muinane *xíni-xi* (Aschmann 1993: 134);
 - (iii) Resígaro *-hu* 'long and flat, horizontal', Bora *-iʔkhi*, *-íxi* (Aschmann 1993: 135), Proto-Bora-Muinane *-íxi* (Aschmann 1993: 135);
 - (iv) Resígaro *-í* 'stick-like', Bora *-i*, Proto-Bora-Muinane *-i* (Aschmann 1993: 140);

- (v) Resígaro -*kó* 'classifier for thick stick', Bora -*ko* 'classifier for stick-like objects' (Thiesen 1996: 102).
- (b) CLASSIFIERS WHICH CORRESPOND TO BOUND AND TO FREE MORPHEMES IN BORA. A few specific classifiers are used as classifiers and as nouns in Bora, and only as classifiers in Resígaro. A classifier of a Bora origin is attached to a noun of Arawak origin, e.g.
- (i) Resígaro classifier -*mi* 'canoe' in *hiítá-mi* 'canoe' (cf. Bare *isa*, Achagua *tída*, Tariana *ita(-whya)*, Baniwa *ita*, Piapoco *ída*, Yucuna *hiíta* 'canoe') also used as a classifier in Bora: -*mi* 'canoe, other transport' (Thiesen 1996: 102), and as a root in Bora *mü-ne* 'canoe' (cf. Proto-Bora-Muinane **mü-ne*: Aschmann 1993: 136);
 - (ii) Resígaro classifier -*ʔaamí* 'leaflike', cf. Resígaro singulative *apána-ʔaamí* 'leaf' (from *apánú* 'leaves', PAR **pana* 'leaf': Payne 1991: 410), Bora (-)*háamí* 'leaf-like, leaf', Muinane *áame*, Proto-Bora-Muinane (*ina*)-*ʔáamí* (Aschmann 1993: 140) 'leaf'.
- (c) CLASSIFIERS WHICH CORRESPOND TO FREE MORPHEMES IN BORA AND WITOTO. Resígaro classifiers which correspond to free nouns in Bora include four body parts, the word for 'village' and a few nouns which refer to natural phenomena ('uninhabited part of the jungle', 'honey', 'day, period of day', 'cotton', 'path' and 'field'). Some attach to an item of Arawak origin with the same semantics and are also used as agreement markers with numbers and classifiers, e.g.
- (i) Resígaro -*ʔosí* 'classifier: hand', singulative -*ké-ʔosí* 'hand' (Resígaro -*ké* from PAR **kapi*: Payne 1991); Bora *hojtsü* 'hand', Proto-Bora-Muinane -*ʔóxtsü* 'hand';
 - (ii) Resígaro -*tuʔá* 'classifier: foot', singulative -*hiiʔá-tuʔá* (Resígaro -*hiiʔpu*, with *ú > á* is a regular morphophonological process) from PAR **kihti-ba* (cf. Tariana, Baniwa *hipa*), Bora *tuháá* 'foot', Proto-Bora-Muinane -*ttü-ʔaái* 'foot' (Aschmann 1993: 132);
 - (iii) Resígaro -*kuba* 'leg', cf. Miraña *khúpaá* (Huber and Reed 1992: 25) in Resígaro -*íphi-kuba* 'leg', -*hiiʔá-kuba* 'leg'; where *íphi* is cognate with Tariana -*phi-na* and Achagua -*húi* 'upper leg, thigh', from PAR **boki* (Payne 1991: 421).

Others occur with a 'dummy' root *té* (from Bora *te/tee* (Thiesen 1996: 34) 'something mentioned before'), e.g.

- (iv) Resígaro -*bahú* 'uninhabited part of the jungle' in *té-bahú* 'uninhabited part of the jungle' (cf. Bora *bájú-pa né* 'bush': Allin 1975: 508), *páxiä*, Proto-Bora-Witoto **báxiä* (Aschmann 1993: 140);
- (v) Resígaro *té-baké* 'root', Bora *bájkýéé* (Allin 1975: 509), *páxkhyéé*, Muinane *bakó-* (Aschmann 1993: 140), Proto-Bora-Muinane **ba(i)(k)ke* (Aschmann 1993: 140).

One classifier, *-koomí* 'village', used with Resígaro *panítítsí* 'house' (from PAR **pani* 'house'), as in *paní-tsí-mú-koomí* (house-NPOSS-PL-CL:VILLAGE) 'village', also occurs with the dummy root *te*, as in *té-koomí* 'village'. It has been borrowed from Bora *cóomíí*, Proto-Bora-Muinane *kóomíí* (Aschmann 1993: 135).

The classifiers of groups (a) and (b) must have been borrowed as bound morphemes, while classifiers of the group (c) were most probably borrowed as free morphemes and then grammaticalized as classifiers.

This extensive borrowing of classifiers in almost all the semantic fields can be explained by their role in discourse: once the referent is established it is referred to with a classifier, so that classifiers appear to be more frequent in discourse than nouns themselves. In a number of Amazonian languages—including Tucano, Arawak, and Bora-Witoto (see Aikhenvald 2000a: 287)—a full noun is almost always omitted from a noun phrase: classifiers are used instead. As a result, classifiers are more frequent in discourse than full nouns.

Three of the seventeen oblique case markers in Resígaro borrowed from Bora are *-maʔ* 'without', from Bora *-ma* 'without', *-gi* 'instrumental', from Bora *-ri* 'instrumental' (Thiesen 1996: 96; from Proto-Bora-Muinane **-ri*: Aschmann 1993: 152), and *-ké* 'dative' (also 'while') (Allin 1975: 238), from Bora *-ki* 'purposive' (Thiesen 1996: 96).

- B. STRUCTURAL LEVELLING: BORA INFLUENCE ON RESÍGARO VERBAL MORPHOLOGY AND SYNTAX. The verbal morphology of Resígaro has been restructured to fit the dominant Bora patterns; however, there is no evidence of direct borrowing of morphemes. Table 9 illustrates the structural matching of Bora tense distinctions onto Resígaro; the actual morphemes in Resígaro have cognates in other Arawak languages.

Resígaro pronouns and nominal morphology underwent restructuring with the borrowing of free and bound morphemes. In contrast to nominal morphology, the verbal morphology of Resígaro has been restructured to fit the dominant Bora patterns, with hardly any borrowing of morphemes.

TABLE 9. Tense in Bora and Resígaro

Tense	Bora	Resígaro	Cognates in Arawak
remote past	<i>-pe</i> or lengthening of final vowel	<i>-ʔpe</i>	Tariana <i>-ka-pe</i> 'habitual', Baniwa <i>-ka-pe</i> 'remote past'
recent past	<i>-ne/hne</i>	<i>-mi</i>	cf. Baniwa <i>-mi</i> 'past'
future	<i>-iíkye/-ii/-i</i>	<i>-vá</i>	cf. Baniwa <i>-wa</i> 'incomplete'

4.3. PROPERTIES SHARED BY TARIANA, BANIWA, RESÍGARO, TUCANO, AND BORA-WITOTO

Table 2 shows that Resígaro shares a high percentage of lexicon with Tariana (55%) and with Baniwa (52–6%). A number of grammatical morphemes are also shared with Tariana and with Baniwa, but not with other Arawak languages.

The following oblique case markers in Resígaro have cognates with Tariana and/or with Baniwa:

- (i) Resígaro *-neé* ‘with’ (Allin 1975: 252), Tariana *-(i)ne* ‘instrumental’; Baniwa adposition *-inai* ‘with’;
- (ii) Resígaro *-giko* ‘in’ (Allin 1975: 276), Baniwa *-riku* ‘locative’, Tariana *-riku* ‘derivational suffix’ (note that *g* in Resígaro regularly corresponds to *r* in other Arawak languages).
- (iii) Resígaro *-ípe* ‘in front of’, Tariana, Baniwa verb and adposition *-pe* ‘be in front, in front’.

Resígaro, Tariana and Baniwa share a marker of remote past (see Table 9), and a nominalizer *-mi* (Allin 1975: 111). Baniwa interrogative particles *hapha* and *-pha* are cognates to Resígaro interrogative particle *kapha*. Both Tariana and Resígaro use the suffixes *-se* and *-thé* (Allin 1975: 117) on distal demonstratives, where Resígaro *th* is a regular correspondent of Tariana *s*.

A number of lexemes are shared exclusively by Tariana, Baniwa, and Resígaro, e.g. Resígaro *hee?ko* ‘day’, Tariana, Baniwa *hekwapi* ‘day’; Resígaro *-dápee* (Allin 1975: 146) ‘sing’, Tariana, Baniwa *-rapa* ‘sing, dance’; Resígaro *poo?gi* ‘oven’, Tariana, Baniwa *puari* ‘oven’; Resígaro *patshá-nu?* ‘wet’, Tariana, Baniwa *puffa* ‘wet’; Resígaro *va?nu* ‘command’, Tariana, Baniwa *-wana* ‘call, order’.

A few items are shared just by Resígaro and Tariana, e.g. Resígaro *epítshí* ‘axe’, Tariana *episi* ‘axe’; Resígaro *kedáví?* ‘red’, Tariana *kerawiki* ‘snuff’ (powder which is reddish in colour).

These shared morphemes may be indicative either of some genetic relationships between Tariana, Baniwa, and Resígaro, or areal diffusion (possibly prior to the Resígaro migration into Peru).

Bora-Witoto-Resígaro share a number of features with the languages of the Vaupés (Tariana and Tucano), but not with Baniwa. These include nominative–accusative profile, the use of classifiers with demonstratives, and the use of classifiers as individualizing and singulativizing markers (cf. Allin 1975: 151, Thiesen 1996: 122). These features could be indicative of some areal diffusion between the languages of the Vaupés and Bora-Witoto-Resígaro linguistic areas. I mentioned in §4.2 that Bora, Witoto, and Resígaro groups were located on the Caquetá river over eighty years ago much closer to the Vaupés (the putative migration is shown on the Map). Thus, in the past Resígaro was much closer to the Vaupés than it is now. The similarities may be due either to older contacts between these languages, or to shared independent contacts between the Tucano

languages with Bora-Witoto and Resígaro, on the one hand, and with Tariana, on the other hand. In any case, we cannot decide whether Tariana, Resígaro, and Baniwa share a largish number of lexical (and also grammatical) morphemes due to areal diffusion, or to genetic affinity, or an interaction of both.

5. Final remarks

The divergences between Arawak languages spoken north of the Amazon make it almost impossible to go beyond very low-level subgroupings, such as Tariana-Baniwa.

There is definitely not enough evidence to justify taking all the North Arawak languages to be a genetic group. The ultimate explanation for this lies in the existence of a limited stock of genetically inherited morphemes, overlaid by vast influxes of areally diffused patterns, due to intensive and prolonged contacts with neighbouring languages.

In many cases languages show a high percentage of shared lexicon but differ significantly in terms of their grammatical morphemes. This is the case with Tariana, Baniwa, Achagua, and Piapoco. While Tariana and Baniwa are spoken within distinct subareas of the large linguistic area of the Içana-Vaupés basin, Piapoco and Achagua (which are now in close contact) belong to a different area. In this case, the different areal diffusion patterns could be assumed to have obscured the erstwhile subgrouping.

The number of lexical and grammatical cognates depends on how languages get restructured within distinct linguistic areas. Tariana and Resígaro illustrate contact-induced language change in different sociolinguistic situations. Tariana, the only Arawak language spoken in the Vaupés area dominated by East Tucano groups, illustrates a drastic restructuring of grammar without any borrowings of morphemes. In contrast, Resígaro, the only Arawak language spoken among the Bora and Witoto groups, combines restructuring with an unusually large amount of borrowed free and bound nominal morphemes and at least one pronoun.

Additional structural similarities between the languages of the Içana-Vaupés area (Tariana, Baniwa, and the East Tucano) and Bora-Witoto-Resígaro may be indicative of older contacts.

The example of Resígaro, Tariana, and Baniwa shows that languages—which still show some evidence of their common origin—can be restructured beyond recognition. This is especially dramatic in the case of Resígaro where even bound morphemes get borrowed. However, since Resígaro, Tariana, and Baniwa also share a number of forms as well as structural features, it is impossible to decide whether their similarities are ultimately due to a common genetic origin, or are the result of a long-term coexistence with each other in one linguistic area, or both.

It is known that ‘extensive and prolonged contact, as it is frequently found in areas long settled by speakers of the same language, may cause considerable difficulties for the historical linguist’, making it ‘next to impossible to classify dialects

in terms of their genetic relationship' (Hock 1991: 447). In the case of the Arawak languages in the North Amazon, their extensive and prolonged contact with genetically unrelated languages has obscured the subgrouping—to a different extent in different sociolinguistic situations of multilingualism.

Other language families north of the Amazon face similar problems. East Tucano languages share numerous features in common which distinguish them from West Tucano; however, it is not at all clear whether this is due to the fact that East Tucano languages have been spoken in the Vaupés linguistic area for a long time, or that they are indeed a genetic subgroup. The same holds for the similarities between Bora-Muinane and Witoto languages. At present, we cannot decide whether strong similarities between Witoto languages are due just to the fact that they form a genetic subgroup, or are partially conditioned by other factors, such as areal diffusion (or parallel development).

In one case areal diffusion patterns have even obscured the actual genetic relationships. The Makú languages spoken near the Upper Rio Negro and in the Vaupés area share a number of lexical morphemes—all of which are monosyllabic—and some parts of pronominal paradigms with their putative relatives, Nadëb and Shiriwe, spoken on the Middle Rio Negro; their grammars show drastic differences. There are two equally valid possibilities: either the Makú languages constitute an old family obscured by areal diffusion; or they constitute a relic of an old linguistic area where we cannot distinguish old borrowings from genetic inheritance (see Martins and Martins 1999). The ways in which these languages developed make it impossible to decide whether their similarities are due to intensive areal diffusion or to the fact that they formed a closely related subgroup, or both.

According to one hypothesis, the Arawak languages spread from the Orinoco headwaters, which has been suggested as one of the places where agriculture developed. This punctuation could have given rise to the emergence of the Arawak family. Subsequent periods of minor punctuation (such as the move of the Tariana to 'join' the Tucano groups on the Vaupés; or the contact between Resígaro and Bora-Witoto), and of intermediate equilibrium periods, contributed to the areal diffusion of patterns and forms, which has helped to obscure the original genetic relationships.

However, in many cases we do not have enough information on exactly what areal influences come from where (as in the case of Achagua, Piapoco or Yucuna, or Palikur: we simply have no information about what languages might have become extinct in these regions).

Thus, numerous migrations and long equilibrium processes followed by intermittent punctuations contribute to the difficulties encountered by those who want to distinguish genetic inheritance from areal diffusion in Amazonia (cf. Dixon 1997).

References

- AIKHENVALD, A. Y. 1994a. 'Grammatical relations in Tariana', *Nordic Journal of Linguistics* 17.201–17.

- AIKHENVALD, A. Y. 1994b. 'Classifiers in Tariana', *Anthropological Linguistics* 36.407–65.
- 1994c. 'Classe nominal e gênero em línguas Aruák', *Boletim do museu paraense Emílio Goeldi* 10.137–259.
- 1995a. *Bare*. Languages of the World: Materials 100. Munich: Lincom Europa.
- 1995b. 'Person marking and discourse in North Arawak languages', *Studia Linguistica* 49.153–95.
- 1996a. 'Areal diffusion in north-west Amazonia: the case of Tariana', *Anthropological Linguistics* 38.73–116.
- 1996b. 'Words, phrases, pauses and boundaries: evidence from South American Indian Languages', *Studies in Language* 20.487–517.
- 1996c. 'Multilingual and monolingual placenames in Tariana', *Names: A Journal of Onomastics* 44.272–90.
- 1998. 'A Grammar of Warekena', pp. 215–439 of *Handbook of Amazonian Languages*, vol. 4, edited by D. C. Derbyshire and G. K. Pullum. Berlin: Mouton de Gruyter.
- 1999a. *Tariana Texts and Cultural Context*. Lincom Europa Language Materials 7. Munich: Lincom Europa.
- 1999b. 'The Arawak Language Family', pp. 65–105 of Dixon and Aikhenvald (1999).
- 1999c. 'Areal diffusion and language contact in the Içana-Vaupés area, north-west Amazonia', pp. 385–415 of Dixon and Aikhenvald (1999).
- 2000a. *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- 2000b. 'Areal typology and grammaticalization: the emergence of new verbal morphology in an obsolescent language', pp. 1–37 of *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*, edited by Spike Gildea. Amsterdam: John Benjamins.
- 2002. *Language Contact in Amazonia*. Oxford: Oxford University Press.
- 2003. *A grammar of Tariana, from northwest Amazonia*. Cambridge: Cambridge University Press.
- and DIXON, R. M. W. 1998. 'Evidentials and areal typology: a case study from Amazonia', *Language Sciences* 20.241–57.
- and GREEN, D. 1998. 'Palikur and the typology of classifiers', *Anthropological Linguistics* 40.429–80.
- ALLIN, T. R. 1975. 'A grammar of Resígaro'. Ph.D. dissertation, University of St Andrews, UK.
- ASCHMANN, R. P. 1993. *Proto-Witotoan*. Arlington: Summer Institute of Linguistics and the University of Texas.
- BARNES, J. 1999. 'Tucano', pp. 207–26 of Dixon and Aikhenvald (1999).
- BRINTON, D. G. 1892. 'Studies in South American native languages', *Proceedings of the American Philosophical Society* 30.45–105.
- CAMPBELL, L. 1997. 'Amerindian personal pronouns: a second opinion', *Language* 73.339–51.
- , KAUFMAN, T., and SMITH-STARK, T. 1986. 'Meso-America as a linguistic area', *Language* 62.530–70.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- and AIKHENVALD, A. Y. 1999. Editors of *The Amazonian Languages*. Cambridge: Cambridge University Press.
- GREEN, H., and GREEN, D. 1972. *Surface Structure of Palikur Grammar*. Brasília: Summer Institute of Linguistics.

- GREENBERG, J. H. 1953. *Essays in Linguistics*. Chicago: The University of Chicago Press.
- HEATH, J. 1978. *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- 1981. 'A case of intensive lexical diffusion', *Language* 57.335–67.
- HOCK, H. H. 1991. *Principles of Historical Linguistics*. 2nd edition. Berlin: Mouton de Gruyter.
- HUBER, R. Q., and REED, R. B. 1992. *Comparative Vocabulary. Selected Words in Indigenous Languages of Colombia*. Santafé de Bogotá: Asociación Instituto Lingüístico de Verano.
- JACKSON, J. 1974. 'Language identity of the Colombian Vaupés Indians', pp. 50–64 of *Explorations in the Ethnography of Speaking*, edited by R. Bauman and J. Sherzer. Cambridge: Cambridge University Press.
- LOUKOTKA, C. 1968. *Classification of South American Indian Languages*. Los Angeles: Latin American Studies Centre, UCLA.
- MARTINS, S. A., and MARTINS, V. 1999. 'Makú', pp. 251–68 of Dixon and Aikhenvald (1999).
- MASICA, C. P. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- MATTESON, E. 1965. *The Piro (Arawakan) Language*. Berkeley and Los Angeles: University of California Press.
- 1972. 'Proto-Arawakan', pp. 160–242 of *Comparative Studies in Amerindian Languages*, edited by E. Matteson, A. Wheeler, F. L. Jackson, N. E. Waltz, and D. R. Christian. The Hague: Mouton.
- MINOR, E. E., and LOOS, E. E. 1963. 'The structure and contexts of Witoto predicates in narrative speech', pp. 37–66 of *Studies in Peruvian Indian Languages* 1. SIL Publications in Linguistics 9. Norman: Summer Institute of Linguistics and University of Oklahoma.
- NIMUENDAJÜ, C. 1982. *Textos indigenistas*. São Paulo: Ed. Loyola.
- NOBLE, G. K. 1965. 'Proto-Arawakan and its descendants', *International Journal of American Linguistics* 31: 3, part II.
- OLIVER, J. R. 1989. 'The Archaeological, Linguistic and Ethnohistorical Evidence for the Expansion of Arawakan into Northwestern Venezuela and Northeastern Colombia'. Ph.D. dissertation, University of Illinois.
- PAYNE, DAVID L. 1985. 'The genetic classification of Resígaro', *International Journal of American Linguistics* 51.222–31.
- 1990. 'Some widespread grammatical forms in South American languages', pp. 75–88 of Doris Payne (1990).
- 1991. 'A classification of Maipuran (Arawakan) languages based on shared lexical retentions', pp. 355–499 of *Handbook of Amazonian Languages*, vol. 3, edited by D. C. Derbyshire and G. K. Pullum. Berlin: Mouton de Gruyter.
- PAYNE, DORIS L. 1990. Editor of *Amazonian Linguistics. Studies in Lowland South American Indian Languages*. University of Texas Press: Austin.
- RAMIREZ, H. 1992. *Le Bahuana, une nouvelle langue de la famille Arawak*. Chantiers Amerindia. Supplement 1 to no. 17 of *Amerindia*, Paris.
- 1997. *A fala Tukano dos Yepá-masa*, vol. 1. *Gramática*. Manaus: Inspetoria Salesiana.
- RIVET, P., and WAVRIN, R. de 1951. 'Un nouveau dialecte Arawak: le Resígaro', *Journal de la Société des Américanistes de Paris* 40.203–38.
- ROWAN, O., and BURGESS, E. 1979. *Parecis Grammar*. Arquivo Lingüístico. Brasília: Summer Institute of Linguistics.
- SAPIR, E. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace,

and World.

SEKI, L., and AIKHENVALD, A. Y. Forthcoming. 'Proto-Arawak of Xingu'.

SHERZER, J. 1976. *An Areal-Typological Study of American Indian Languages North of Mexico*. North-Holland Linguistic Series 20. Amsterdam: North-Holland.

SORENSEN, A. P., JR. 1967. 'Multilingualism in the northwest Amazon', *American Anthropologist* 69.670–84 (reprinted as pp. 78–93 of *Sociolinguistics: Selected Readings*, edited by J. B. Pride and J. Holmes. Harmondsworth: Penguin 1972).

THIESEN, W. 1996. *Gramática del Idioma Bora*. Série Lingüística Peruana 38. Yarinacocha, Pucallpa: Instituto Lingüístico de Verano.

TOVAR, A., and LARRUCEA DE TOVAR, C. 1984. *Catálogo de las lenguas de América del Sur*. Madrid: Editorial Gredos.

WHIFFEN, T. W. 1915. *The North-West Amazonas: Notes of Some Months Spent among Cannibal Tribes*. London.

WILSON, P. J. 1992. *Una descripción preliminar de la gramática del Achagua (Arawak)*. Bogotá: Summer Institute of Linguistics.

WISE, M. 1999. 'Small language families and isolates in Peru', pp. 307–40 of Dixon and Aikhenvald (1999).

ZAMPONI, R. 2003. *Maipure*. Munich: Lincom Europa.

Linguistic Diffusion in Present-Day East Anatolia: From Top to Bottom

Geoffrey Haig

1. Introduction

For centuries, East Anatolia has been host to representatives of four distinct language families: Indo-European, Kartvelian, Semitic, and Turkic. Although this degree of linguistic diversity pales in comparison with, say, Papua New Guinea or the Amazon basin, by Eurasian standards it is extremely high, making East Anatolia one of the best linguistic laboratories for investigating language contact in the Old World. But due to the repressive Turkish policies on Anatolian minorities, language contact has until very recently attracted little scholarly attention.¹

In this chapter I will attempt a preliminary synthesis on language contact in East Anatolia, based primarily on data from four languages. I say preliminary because given the size of the area and the lack of reliable sources for some of the languages, any conclusions can be no more than tentative at this stage. Nevertheless, I believe it is worth taking a shot at the broader view in order to identify recurrent patterns and to formulate hypotheses which can be tested in sorely needed local case studies of language contact in the area. My objectives are threefold: first, to present and analyse a considerable amount of data; second, to discuss the question of whether East Anatolia constitutes a linguistic area. Finally, I will formulate a more general hypothesis regarding the mechanisms of contact-induced linguistic change, namely that it begins at larger syntactic units, e.g.

I wrote most of this chapter during a fellowship at the Research Centre for Linguistic Typology at the Australian National University, and it owes a great deal to the spirit of that institution. In particular I would like to thank Sasha Aikhenvald and Tim Curnow for extensive comments on earlier drafts. The following people also generously contributed their time and expertise: Winfried Boeder, Friederike Braun, Nick Enfield, Sevim Genç, Bernd Heine, Yaron Matras, Ludwig Paul, Malcolm Ross, Christoph Schroeder, Kevin Tuite, and two anonymous referees. The usual disclaimers apply.

¹ The recent upsurge in interest in the topic is documented in Johanson (1992), Johanson (1998), Dorleijn (1996), Matras (1998), Bulut (2000, 2005), and Haig (forthcoming). As this chapter was going to press the recent work of Uwe Blasing on Armenian and of Bernt Brendemoen on Pontus Greek came to my attention, but unfortunately could not be incorporated into the present chapter.

techniques of clause linkage, before filtering down to affect lower levels of grammatical organization. I should stress, however, that I am concerned solely with morphosyntax, not with the lexicon or phonology.

The chapter is organized along the following lines: in §2 I give background information on the area and the languages. The primary language data is concentrated in §3 and §4: in §3 a number of structural parallels across all languages are presented, and the question of whether East Anatolia qualifies as a linguistic area is briefly addressed; in §4 the Anatolian data are examined against the backdrop of the structural compatibility issue. Finally, in §5 I return to broader issues and formulate some generalizations on the mechanisms of contact-induced language change.

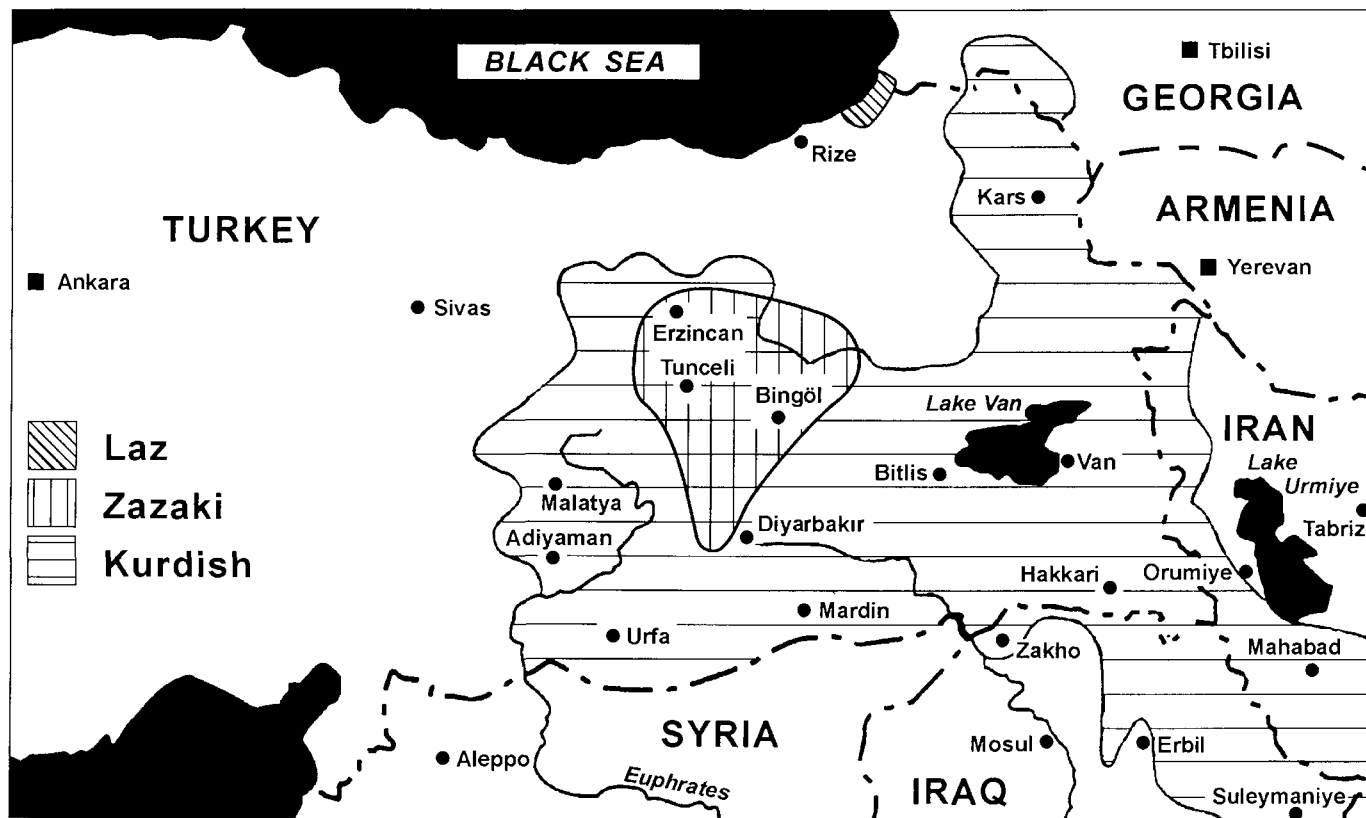
2. The languages and the area

East Anatolia, for the purposes of this chapter, is that portion of modern Turkey roughly east of a line drawn north-south from the town of Sivas. It constitutes both linguistically and ethnically a transitional zone, between Afroasiatic in the South, the western members of Indo-European in the west, Iranian and Indo-Aryan in the East, and the three indigenous Caucasian language families in the north-west. Politically it is currently under Turkish dominance, but it has retained a far higher degree of linguistic diversity than the western parts of Turkey, where Turkification is almost complete; in a large part of East Anatolia only 25–50% of the rural population know Turkish, with the figure dropping to a reported 5% in the far south-east (Nestmann 1989: 551).

The major languages currently spoken in the area are Turkish throughout, Laz (Kartvelian) in the north-west, Zazaki (Iranian), and Kurmanji Kurdish (Iranian) in central and south-east Anatolia, and Aramaic and Arabic in the south-east. There are also scattered remnants of the Indo-European languages Armenian and Greek, which were spoken by large speech communities prior to their forced exodus (see below). Finally, there are isolated villages where Circassian and Kabardian (North-West Caucasian) are spoken by Muslims who emigrated to Anatolia in the nineteenth century. I will be concentrating on the four best-documented languages spoken in the area: Turkish (Turkic), Laz (Kartvelian), Kurmanji Kurdish, and Zazaki (both Iranian).² Although Zazaki is closely related to Kurmanji, they are not mutually intelligible. The distribution of the three minority languages is shown in the Map.

Note that Laz is geographically isolated from the other minority languages.

² The sources for the individual languages are the following: For Turkish, my own knowledge supported by a variety of sources; for Laz, Dumézil and Ensenç (1972), Holisky (1991), Kutscher, Mattissen, and Wodarg (1995), my own fieldwork, and consultation with Sevim Genç, a native speaker; for Kurmanji, Bedir Khan and Lescot (1986), MacKenzie (1961), Barnas and Salzer (1994), Dorleijn (1996), Bulut (2000) and my own fieldwork; for Zazaki, Paul (1998) and personal communication with Ludwig Paul.



MAP. The distribution of Kurdish, Zazaki, and Laz in East Anatolia

Turkish, as the language of administration, education, and broadcasting is spoken in all larger towns throughout Anatolia, hence cross-cutting the areas where the minority languages are spoken. The uniting factor across all the minority languages is therefore long-standing contact with Turkish.

Current estimates of the numbers of speakers vary considerably: for Laz, between 50,000 and 500,000 (Feuerstein 1994, Holisky 1991: 397, Vanilişi and Tandilava 1992: 83, and Andrews 1989: 176); for Zazaki between 1.5 and 2.5 million (Paul 1998: xiii). Estimates for the number of Kurmanji speakers in Turkey range from 8 to 15 million. There are no reliable figures on minority populations prior to this century because Ottoman records did not distinguish ethnic minorities but rather religious minorities (McCarthy 1983: 7).

Contact between Turkish and the minority languages of East Anatolia goes back at least five hundred years. However, the status of the minority languages changed abruptly at the beginning of the twentieth century. Under Ottoman rule (up to the end of the First World War), the use of languages other than Ottoman Turkish was perfectly acceptable; the numerous minority-language communities within the empire's boundaries were under no pressure to abandon their languages. In fact, the Turkish of the common people had no favoured status over other languages, and up until the final years of the Empire it was actually held in low esteem; the languages of high-status were Arabic and Persian. Although the official language of the Ottoman Empire, Ottoman Turkish, was based on Turkish grammar, it had become so elaborated with Arabic and Persian elements that it was incomprehensible to all but the educated elite. Thus within the Ottoman Empire, minorities were not generally discriminated against on linguistic grounds, although the minorities of East Anatolia, particularly non-Muslim minorities, suffered in other ways.

Around the First World War the status of minority languages in Anatolia deteriorated radically. First there was the enforced and bloody deportation of thousands of Armenians from Anatolia. After the war, the Turkish Republic was founded, based on an ethno-nationalist ideology which made little provision for accommodating linguistic minorities, and indeed denied their very existence. The deportation of thousands of ethnic Greeks was a further step in bleaching the colour out of Anatolia's ethnic tapestry. The other minorities remained, but the use of languages other than Turkish was officially repressed. Compulsory schooling and military service, massive urban drift, and the recent scorched-earth policies of the Turkish army in their efforts to combat militant Kurds have all contributed to the erosion of East Anatolian rural culture, and to the destruction of the linguistic equilibrium. Van Bruinessen (1992: 66) notes that between one third and one quarter of the Kurds have left their homeland over the last fifty years.

The events of the past hundred years have had a devastating impact on East Anatolia, essentially redrawing the ethnic and linguistic map of the area. However, when considering the contact situation as a whole it is important to bear in mind

that the current economic and political pressure on minority-language speakers to adopt Turkish is by no means typical for the language-contact situation over the previous five centuries.

The historical background in Anatolia makes it difficult to evaluate the genesis of many of the evident contact-induced changes: are they due to centuries of gradual assimilation, or are they the result of imperfect learning due to sudden recent interruption of transmission of minority languages? In extreme cases, it can be difficult to distinguish what is probably sporadic code-switching from what are entrenched patterns due to systematic borrowing. A second difficulty is that Anatolia itself is a transitional zone, at the intersection of several higher-level diffusion areas. For example, northern Kurmanji dialects have an additional row of voiceless stops, thought to have been borrowed from Armenian; this is a feature typical of the Caucasus. The southern dialects on the other hand have additional emphatic consonants, clearly under Arabic influence (see Kahn 1976). Thus Kurmanji straddles an intermediary zone between Semitic and the languages of the Caucasus. The task of teasing out the local Anatolian contact phenomena from the broader Eurasian–Transcaucasian contact zone is unfortunately beyond the scope of this chapter.

2.1. TYPOLOGICAL PROFILES OF THE LANGUAGES

The four languages under consideration here differ from each other structurally in several respects, which I will briefly summarize here. However, the two Iranian languages, Zazaki and Kurdish, are structurally similar enough to be treated as a single unit, ‘Anatolian Iranian’, across most typological parameters.

In terms of morphological typology, Turkish is an exclusively suffixing, agglutinative language, while the others are mixed prefixing/suffixing, and have some fusional characteristics. The Iranian languages have nominal gender, which Turkish and Laz lack.

The verb systems of the languages differ considerably: Laz has a system of distinct conjugational classes, as do the Iranian languages. Turkish on the other hand has a single conjugation class, with one type of inflection across the board. Turkish has productive morphological passive and causative formation. Laz has morphological causatives and middles, but no passive. Kurmanji has neither morphological passives nor causatives; they are expressed periphrastically (Zazaki has a semi-productive morphological passive). As far as cross-referencing of core arguments on the predicate is concerned, Turkish patterns with the Iranian languages in that it cross-references maximally a single core argument, while Laz allows up to two arguments to be cross-referenced.

Turkish and Laz are postpositional, while the Iranian languages are mixed post- and prepositional. All languages have verb-final order as the pragmatically unmarked constituent order in the simple clause. Within the NP, Turkish and Laz are head-final, while the Iranian languages are head-initial.

3. Pan-Anatolian structural parallels

In this section I present a selection of structural parallels in the four languages, including examples of actual borrowing of morphemes, as well as calquing of structural patterns. The examples are intended as a representative sample; it would be quite possible to extend this list further.

As Campbell, Kaufman, and Smith-Stark (1968: 534) point out, the mere existence of structural parallels in neighbouring languages is in itself no evidence that the languages concerned have affected each other. In order to develop a case for contact-induced change, we also require supporting evidence that the parallels represent developments otherwise unlikely given the genetic predisposition of the individual languages, and unlikely from the point of view of known universal tendencies. Although for some of the structural parallels discussed below, such supportive evidence is readily available, for others, matters are less straightforward. I would therefore like to make it quite clear that I am not claiming contact influence is necessarily the source of all these similarities. Nevertheless, cataloguing potential candidates remains the prerequisite for a later more detailed analysis; thus the main thrust of this section is concerned with presenting some of the more likely candidates. The data are admittedly heavy-going in parts; a summary is given in §3.6 under Table 1.

Finally, it is undeniable that all four languages share a considerable body of common cultural vocabulary, idioms, certain categories such as evidentiality, formulaic expressions in traditional narratives, and many situation-bound expressions (greetings, expressions of thanking and requesting, etc.), i.e. what one could broadly characterize as ‘ways of saying things’. Such aspects have tended to be ignored in contact linguistics, but as Ross (this volume) points out, they are part and parcel of contact-induced change. For reasons of brevity, however, the present chapter is restricted to an investigation of grammatical features.

3.1. THE COMPLEMENTIZER *KI*

Many languages of the area use a complementizer, variously realized as *ku*, *ki*, or *ko*, all of which go back to an original Iranian word (cf. modern Persian *ke*). I will refer to these elements collectively as *KI*. In East Anatolia *KI* occurs in the four languages discussed here (and perhaps in all languages of East Anatolia?), where it fulfils a variety of functions. The following two functions are covered by *KI* in all four languages: linking verbs of speech or thought to their complements, and in ‘so that’-constructions. In the examples it is glossed simply ‘*KI*’.³

³ The transcription used in the examples generally follows that of the respective source, with some minor exceptions. In the interests of consistency, a simplified morphological gloss has been applied across the board: morpheme boundaries are largely ignored in the source text, but the presence of individual morphemes, and their order, is recorded in the gloss. In the Laz examples, only a single core argument, that corresponding to syntactic subject, has been recorded in the glosses of the predicates.

3.1.1. *KI with verbs of speech, thought, etc.*

A Turkish example of the complementizer *KI* is the following:

(1) Turkish:

anladım	ki	onun	bir	derdi	var
understand.PAST.1sg	KI	3sg.GEN	a	problem.POSS3sg	exist.3sg

'I realized that he had a problem.'

This is of course not the only type of complement clause in Turkish—we also find a nominalized complement clause without any complementizer. But *KI*-clauses such as (1) are certainly a regular feature of Turkish discourse. Note that *ki* in Turkish is not a native morpheme, but a loan from Persian (cf. Persian *ke*). The Turkish form may have involved contamination with the Turkish interrogative pronoun *kim* 'who'.

In Laz, *KI* is common as a subordinator (in Laz /k/ is often palatalized to /ç/):

(2) Laz (Wodarg 1995: 130):

Nana	muşı	užomei=çi	ma	hui	ma	hui	bulur
mother	POSS.3SG	say.3sg.PRES=KI	1sg	now	1sg	now	go.1sg.PRES

'Her mother says: "now I, now I go [...]".'

(3) Laz (Wodarg 1995: 116):

dvažonu=ki	šku	didamangisa	domažonanen
think.3sg.PFV=KI	1pl	witch	believe.1pl.FUT

'(She) thought we would think that (she) is a witch.'

The following examples illustrate the use of *KI* in Kurmanjî and Zazaki:

(4) Kurmanjî (Barnas and Salzer 1994: 104):

wî	got	ku	biray-ê	wî	nexweş-e
3sg.OBL	say.PAST.3sg	KI	brother-of	3sg.OBL	sick-COP.3sg

'He said that his brother is sick.'

(5) Zazaki (Paul 1998: 134):

fâhm	keno	ki	derdêndê	ney	esto
understanding	do.PRES.3sg	KI	suffering-of	3sg.OBL	exist.3sg

'(He) realizes that he has a problem.'

Although *KI* looks superficially alike in all these languages, it has actually undergone some rather subtle changes: whereas Persian *ke* is a subordinating conjunction *introducing* a complement clause, to which it is generally considered to belong (see Behzad and Divshali (1994: 212), Alavi and Lorenz (1988: 123)), in Turkish and Laz, *KI* is more or less enclitic on the *main* clause, i.e. is not a constituent of the complement clause. This development brings *KI* into line with the typical Turkish pattern of marking syntactic relations at the right-hand boundaries of constituents, rather than at the left-hand boundaries. But note that

this development has not altered the linear order of elements in Turkish and in Persian: *KI* appears in both cases between main and subordinate clause.

I presume that the shift to clausal enclitic was accomplished in Turkish, and *KI* was borrowed in this function into Laz.⁴ In the Iranian languages Zazaki and Kurmanjî, *KI* is a native morpheme, but whereas Kurmanjî *ku* actually represents a historically older stage (cf. Middle Persian *ku*), the Zazaki form *ki* may have been a result of secondary contamination with Turkish *ki*. The case of *KI* is instructive in that it illustrates how boundary markers of large constituents—in this case clauses—are readily accommodated into the grammars of typologically different languages.

3.1.2. *KI* in ‘so . . . that’ constructions

The second use of *KI* common to all four languages is in constructions corresponding to English sentences of the type *she is so clever that no one can match her*. The second clause is usually negated. Examples from the four languages are the following:

(6) Turkish

Sınav-da	o kadar	heyecanlandım	ki
exam-LOC	that much	get excited.PAST.1sg	KI
tek	kelime	bile	yazamadım
single	word	even	write.POT.NEG.1sg

‘I was so nervous in the exam I couldn’t write a single word.’

(7) Laz (Wodarg 1995: 109):

ma	hik’u	zabuni	borti-çi	va-momalu
1sg	so	ill	be.PAST.1sg-KI	NEG-come.POT.PAST.1sg

‘I was so ill I couldn’t come.’

(8) Kurmanjî (Bedir Khan and Lescot 1986: 294):

hertişt	ewqas	giran	bûye	ko
everything	so.much	expensive	be.PFV.3sg	KI
êdi	qûneta	perê	ne	maye
no.more	value-of	money	NEG	remain.PFV.3sg

‘Everything has got so expensive that money no longer has any value.’

(9) Zazaki (Paul 1998: 163)

hendi	rind	benā	ki	kes [. . .]
so	beautiful	be.F.3sg	KI	someone
nêşeno	wesfān-ê	ji	bi-do	
NEG.can.3sg	praise-of	3sg.F	MOD-give.3sg	

‘(She) is so beautiful that no one is able to praise her (adequately).’

⁴ In fact in Turkish it is not always clear which clause *KI* is associated with—see the discussion of this point in Soper (1996: 236–8), and also for etymologically related *ki* in Hindi in Hock (1991: 479–80).

3.2. CLAUSAL ENCLITIC CONDITIONAL MARKER

In all four languages, the protasis of a conditional precedes the apodosis. In Turkish, the verb of the protasis is marked with an enclitic *-sE*. The same enclitic marker has been borrowed into some varieties of Kurmanjî and Zazaki. In Laz, the same type of construction is used, but a native Kartvelian morpheme marks the verb of the protasis:

- (10) Turkish:
 şehir-de iş bul-sa köy-e dönmez
 town-LOC work find-COND village-DAT return.NEG.AOR.3sg
 'If she finds work in town she won't return to the village.'
- (11) Laz (Holisky 1991: 435):
 ʒir-u -k'on
 see.3sg.AOR.COND
 'If he saw (it)'

Although traditional grammars of Kurmanjî, such as Bedir Khan and Lescot (1986), ignore conditionals marked with *-sE*, they are a regular feature of many spoken varieties, and are also found in Zazaki. The following examples are illustrative:

- (12) Kurmanjî (Dorleijn 1996: 54):
 eger bi te re hey-e-se
 if ADP 2sg.OBL ADP exist-COND.3sg
 'If there is with you (if you have (some) with you)'
- (13) Zazaki (Paul 1998: 155):
 bikewo-se dāḥā weş nē-beno
 fall.3sg-COND again healthy NEG-become.3sg
 'If he falls, he won't get well again.'

3.3. AFTER-CLAUSES

In Turkish, Laz, and Zazaki, sequences of clauses linked temporally, which could be expressed in English with 'X happened then Y happened' take the form [X happened]-after [Y happened]. In this construction the element glossed as 'after' is part of the first clause, i.e. differs clearly from English *then*; in Laz and Turkish, it is arguably a postposition. Note, however, that in all three languages using this pattern, a native morpheme is used. Kurmanjî differs from the others in that it adheres to the standard Iranian pattern of using a clause-initial conjunction *piştî* (*ku*) 'after' introducing the first clause.

Turkish uses a non-finite verb form and a postposition:

- (14) Turkish:
 giyin-dik-ten sonra gitti
 get dressed-NOM-ABL after go.PAST.3sg
 'After (he) had got dressed he left.'

Laz uses a verb form followed by *şuk'ule*. This form can be analysed as *şu + k'ule*, genitive suffix + postposition, i.e. as a postposition governing the genitive (see Holisky 1991: 459). That makes it structurally parallel to the Turkish construction, i.e. case-marked, non-finite predicate plus postposition. The difference is that the Laz predicate cannot be readily classified as non-finite. However, case-marked finite verbs are a feature of Laz, as we shall see.

- (15) Laz (Wodarg 1995: 108):

ham	çitabi	golobioni=şuk'ule	omçiru-şa	bidi
DEM	book	read.1sg.PFV=after	swim.INF-LOC	go.1sg.PFV

'After I had read this book I went swimming.'

Zazaki appears to have calqued the pattern from Turkish, but it uses its own postposition, *tepeyā*, to mark the first clause:

- (16) Zazaki (Paul 1998: 151):

ti	merdī	tepeyā,	ez	se	kerā?
2sg	die.PAST.2sg	after	1sg	what	do.MOD.1sg

'After you have died, what should I do?'

The *after*-clauses in Zazaki and Laz are good examples of the calquing of a grammatical pattern without any borrowing of actual material (as opposed to the case of KI discussed in the previous section). What is common to both types of contact influence is that the linear order of comparable elements aligns across the contact languages.

3.4. 'NEVERTHELESS' TYPE CLAUSE LINKER

All four languages use a similar means of linking two clauses, where the second clause expresses something that runs contrary to the expectations raised in the first clause. In all four languages, the marker used introduces the second of the two clauses; in all four languages it has an identical composition, albeit with etymologically distinct morphemes. The first element is the word for 'again' in the respective language, the second is the enclitic topic-switch marker (see §3.7.1). The parallels in both structural composition and the semantics of the source elements are certainly striking, and I am unaware of this particular composition of elements used for the same function in other languages.

The elements in the individual languages are: Turkish *yine de*, Laz *xolo ti*,⁵ Kurmanjî *dîsa jî*, Zazaki *finā žī*. The pattern is exemplified with the following Zazaki example:

⁵ There is another expression in Laz, *do xolo*, which fulfils a similar function. But according to Sevim Genç (p.c.), for her dialect of Laz *xolo ti* is the more usual.

(17) Zazaki (Paul 1998: 157):

hewtāy dewān werdi,
seventy villages eat.PAST

finā žī nēšā bi-qedēno
nevertheless NEG-be able.3pl COND-finish.3pl

‘Seventy villages ate (the melon), but they were still unable to finish (it).’
(i.e. ‘although seventy villages ate the melon . . .’)

The comma (also in the original text) makes it clear that *finā žī* introduces the second clause, just as its equivalents *yine de* and *dîsa jî* do in Turkish and Kurmanjî.

3.5. OTHER CONSTRUCTIONS

For *either-or* constructions, all four languages use the same pattern, presumably based on Turkish:

ya [clause 1] *ya da/yan jî* [clause 2]

Laz calques the construction completely from Turkish, Zazaki and Kurmanjî substitute their own enclitic topic switch markers (cf. §3.7.1) for Turkish *da*:

Turkish: *ya . . . ya da*
Laz: *ya . . . ya da* (Holisky 1991: 454)
Kurmanjî: *ya . . . yan jî*
Zazaki: *yā(n) . . . yā(n) (žī)* (Paul 1998: 120)

For *neither-nor* constructions, all four languages use *ne . . . ne*—for Laz see Holisky (1991: 454), for Kurmanjî see Bedir Khan and Lescot (1986: 234), for Zazaki see Paul (1998: 12).

Another feature common to all four languages is the use of the Turkish type of comparative construction. The features common to all four languages are: (a) order of elements; (b) the standard of comparison is marked with a local case marker or adposition; (c) the adjective lacks an obligatory comparative form; (d) a word meaning ‘more’, Turkish *daha*, is optionally used. Schematically we have:

he from/at-me (*daha*) big is
‘He is bigger than me.’

(On Laz see Kutscher, Matissen, and Wodarg (1995: 27), for Zazaki see Paul (1998: 58).) In Kurmanjî a special comparative form of the adjective is mentioned in grammars, but in many spoken dialects the form given above is regularly used. Note that in languages genetically related to the Anatolian languages the adjective does take a special comparative affix, e.g. in Persian, in Georgian (Tschenkéli 1958: 224–5), and in Uzbek. This suggests that the lack of a special comparative form of the adjective is an Anatolian areal feature. In superlative constructions, the Turkish superlative particle *en* is an option in all four languages (though it is not

mentioned in the more normative grammars of Kurmanji). Thus there are good grounds for assuming a common East Anatolian comparative and superlative type.

In fact, comparative constructions are generally highly diffusible (see e.g. Campbell 1987) and it is almost certainly no coincidence that the distribution of different types of comparative constructions world-wide is largely areal. This fact was ignored in typological literature on comparatives (e.g. Stassen 1985) but is taken up in Heine and Kuteva (this volume).

Along with the constructions noted above, the four languages share a number of conjunctions, adverbs, and discourse markers (given in Turkish orthography): *ama* 'but' (< Arabic); *eğer* 'if' (< Persian); *yani* 'that is, that means' (< Arabic); *daha* 'more' (< Turkish); *hele* 'certainly', used for emphasis (< ?); *işte* 'well, so' (< Turkish); *peki* 'well, good' (< Turkish *pek iyi* 'quite good', not attested in Zazaki); *ve* 'and' (< Arabic, found in Turkish and Laz, but Kurmanji and Zazaki use Iranian *u*). Most of these, as well as several others, have also been borrowed from Turkish into Asia Minor Greek—see the list in Thomason and Kaufman (1988: 217), quoting Dawkins (1916).

The question of the origin of these words in a strict etymological sense is less relevant in the present context. I would prefer to see them as part of a common Anatolian repertoire of discourse particles. The case of *κί*, discussed in §3.1, is instructive: from its Iranian origin it has been borrowed into representatives of at least four other language families: Turkic, Kartvelian, Nakho-Daghestanian (Lezgian, see Haspelmath 1993: 370–1), and Dravidian (Brahui, from Iranian Baluchi, see Emeneau (1980: 345) and Elfenbein (1989: 360). Once fully integrated into a language, the fact that *κί* is etymologically of non-native origin is fairly unimportant; recipient languages can readily become donors, as for example Turkish, which has passed *κί* on to at least two other unrelated languages, namely Laz and Asia Minor Greek.

3.6. SUMMARY OF CLAUSE LINKAGE AND BOUNDARY MARKERS

Table 1 sums up the constructions considered so far across three parameters: linear order and positions of the boundary marker; etymological origin of boundary markers, and composition of boundary markers (only valid when the marker is composed of more than one morpheme). A language is indicated as having a particular construction if that construction is solidly attested in at least some of its dialects. Note that in some cases the construction listed exists alongside native constructions.

3.7. OTHER PARALLELS

In this section I document just two of many structural parallels found at levels not readily captured in the traditional tripartite division phonology–morphology–syntax. In fact, many of the most pervasive similarities found are phenomena which tend to be ignored in grammars, only becoming evident when longer texts are scrutinized.

TABLE 1. Selected structural commonalities in languages of East Anatolia

Construction	Common linear order of major constituents; common position of boundary marker	Source of boundary marker	
		Common etymological source	Different etymological source, but identical composition
verbs of speech + complement	T, L, K, Z (P)	T, L, K, Z	
'so . . . that' constructions	T, L, K, Z	T, L, K, Z	
conditional with -sE	T, L, K, Z	T, K, Z	
'after'-clauses	T, L, Z		
'nevertheless'	T, L, K, Z		T, L, K, Z
'neither . . . nor'	T, L, K, Z	T, L, K, Z	
'either . . . or'	T, L, K, Z	T, L	K, Z
comparatives (with Turkish <i>daha</i>)	T, L, K, Z	Turkish <i>daha</i> : T, L, K, Z	
superlatives (with Turkish <i>en</i>)	T, L, K, Z	Turkish <i>en</i> : T, L, K, Z	

T Turkish, L Laz, K Kurmanji Kurdish, Z Zazaki

3.7.1. Enclitic topic-switch marker

All four languages have an enclitic particle that marks, among other things, the reintroduction of a previously established topic. The NP so marked is not a completely new topic, but one which has been introduced in the broader discourse setting, then drops out of topic status, and is subsequently recalled using the enclitic marker. It is difficult to render in English; it is a little like English *as for* . . ., but not as stylistically marked and not as emphatic.

The enclitic topic marker is one of the most pervasive features of narrative texts in all four languages. There is a remarkable feeling of similarity across the four languages in this regard, which is difficult to convey without providing more extensive examples with supporting context, something notably lacking in many studies of language contact. That is unfortunate because it is arguably in precisely the intermediary zone between syntax and discourse that structural convergence begins. In the following examples, the enclitic topic marker is glossed TOP.

TURKISH DA/DE

The preceding context describes the bargaining process that has taken place between the narrator and a villager (*köylü*). An agreement has finally been reached, and the narrator is giving the villager the money (from a short story in Nesin 1995: 96):

- (18) paral-ları eline say-dık. Köylü de ...
 money-PL.ACC to.his.hand count-PAST.2pl. Villager TOP
 ‘we gave him the money (lit. ‘counted into his hand’). The villager (for his part) ...’

LAZ *TI*

The preceding context describes how the narrator and her brother are fleeing from a swarm of bees. The brother runs into their house and closes the door behind him (Wodarg 1995: 121):

- (19) ma ti himuşi peşine nek’na gomži
 1sg TOP 3sg.-GEN behind door open.PFV.1sg
 ‘I too ran behind him and opened the door’

KURMANJĠ *JĠ*

The narrator is recounting his grandfather’s life. As background he sketches the feudal system whereby the Agha, the landowner, had power over life and death, and everyone did as the Agha commanded. He redirects the narrative back towards his grandfather with the following words (taken from my own transcribed Kurmanjî data, Tunceli dialect—see Haig, forthcoming):

- (20) bapir-ê mî jî ne-kîriye
 grandfather-of 1sg.OBL TOP NEG-do.PFV.3sg
 ‘(but) my grandfather, he did not (do as the Agha told him)’

ZAZAKI *ZÎ/ŽÎ*

The preceding context describes how the friends of a young boy used to call him by the nickname of Gukulah (Paul 1998: 232):

- (21) Läjikî zî enā leqām-dā xwi-rā
 boy TOP DEM:OBL nickname-of REFL-about
 zāf xūy kerdînî
 much annoyance do.PAST.3sg
 ‘The boy, for his part, was very annoyed about this nickname of his.’

The etymological origin of the marker in the four languages is unclear. Turkish *da/de* would appear to be native Turkic, as the same marker crops up in Turkic outside Anatolia (e.g. in Uzbek). But it is unclear whether Zazaki *ti*, Kurmanjî *jî*, and Zazaki *ži* are borrowed from the same source, and if so, what that source might be.

3.7.2. *Echoic expressives*

The next example of difficult-to-classify structural parallels is a type of expressive reduplication by which a word is repeated for expressive effect, but its initial segment is replaced by [m]. The construction, almost certainly originally Turkic, conveys a sense of ‘and so on’—see Lewis (1967: 237–8) for a fuller discussion:

(22) Turkish (Lewis 1967: 237):

dergi	mergi	okumuyor
magazine	ECHO	read.NEG.PRES.3sg

‘(He) doesn’t read magazines, journals or anything like it.’

Examples from the other languages are: Laz *tolî moli* ‘eyes and stuff = the face generally’ (< *tolî* ‘eye’, Wodarg 1995: 124); Kurmanjî *hestî mestî* ‘bones and stuff’ (< *hestî* ‘bone’, Dorleijn 1996: 170); Zazaki *dew mew* ‘village or anything like one’ (< *dew* ‘village’, Paul 1998: 55). Note that in these examples the base word is of native origin, which suggests that the technique is a genuinely productive one and not simply copying of complete Turkish expressions. This type of echoic expressive is widely attested outside Anatolia: throughout Turkic (whence it presumably originates), but also in non-Turkic languages in the Balkans (Grannes 1978), in Iran (Persian), and in the Caucasus (Armenian, Georgian; Kevin Tuite, p.c.). It seems that expressive techniques of this type are among the most readily diffusible linguistic features: Emeneau (1980: 114) considers a similar phenomenon to be a characteristic of the Indian linguistic area (though here it does not involve an initial *m*-segment and is presumably not directly related to the Anatolian echoics).

3.8. EAST ANATOLIA AS A LINGUISTIC AREA?

Let us briefly return to the areal perspective with the question of whether East Anatolia qualifies as a linguistic area. The short answer is ‘we don’t know yet’. Establishing a linguistic area involves more than just cataloguing similarities among the languages of a particular area. We must go on to demonstrate that similarities are not due to chance typological similarity, or genetic inheritance, and then establish the areal delineation of the features concerned. For East Anatolia, a number of additional steps remain to be taken: (a) A thorough investigation of the other languages of the area (e.g. Semitic languages, Armenian). (b) A comparison with areally disjunct, but genetically related languages. (c) An examination of languages spoken on the periphery of East Anatolia, for example Asia Minor Greek, spoken further west, which shares many of the features discussed here, or the minority languages of Iran.

In fact, the areal delineation of the linguistic area, if indeed it is one, requires considerable refinement. Anatolia is a transition zone, surrounded by other linguistic contact zones (the Balkans, the Arabian peninsula, the Caucasus), and several common features of Anatolia also extend into these areas (e.g. echoic expressives, use of *ki* as a complementizer). Thus we are unlikely to find a neat bundling of isoglosses defining a well-defined geographic area. The case of Turkish and the Iranian languages in East Anatolia is further complicated by the fact that Turkic and Iranian have been in contact for centuries (Johanson 1998); some of the convergence phenomena discussed here almost certainly predate Turkish settlement in Anatolia, and are found in related languages outside East Anatolia (see Soper 1996 and Dehghani 1998b). Thus teasing out the ancient

convergence from the local effects, particularly in the case of Turkic-Iranian contact, is an extremely difficult matter.

We may have to expand the original area to an Anatolian-Transcaucasian linguistic area, or to abandon it as an areal unit altogether. The present study has, however, uncovered some promising examples of common contact-induced innovations within the area which can be subjected to scrutiny against additional data from related languages outside the area, and from the less well-documented minority languages within East Anatolia.

4. Turkish–Laz contact and Turkish–Iranian contact: the issue of structural compatibility

As the previous sections have demonstrated, both the Kartvelian language Laz and the two Iranian languages Kurmanji and Zazaki have been affected by Turkish. However, the results of Turkish influence are by no means identical for each language. Among the many possible reasons for the different outcomes of the contact situations is the fact that Laz and the Iranian languages are structurally very different from each other. In other words, different contact outcomes may be due to differing grades of structural compatibility between Turkish and the minority languages. In this section I will explore the issue of structural compatibility as a determining factor in shaping contact-induced change by examining Turkish influence on specific grammatical domains in the individual languages, closing with a brief case study of extreme Turkish influence, the Ardeşen dialect of Laz.

Although I concentrate on structural features here, I should emphasize that extra-linguistic factors such as relative size of speech communities in contact, time depth of contact, degree of bilingualism, etc, must also be considered when evaluating the overall extent of Turkish influence. However, in this section I look at fairly narrowly defined grammatical domains for which, as it turns out, fairly plausible explanations in terms of structural compatibility present themselves.

4.1. SUBORDINATE CLAUSES

For three types of subordinate clauses, Laz displays, in terms of linear order of the constituents, an identical structure to Turkish. Examples of these are given below. The Iranian languages on the other hand have quite different strategies, which I will briefly discuss at the end of the section.

The first subordinate clause type is temporal clauses expressing roughly ‘during’. The Laz strategy involves what appears to be a case-marked sentence: a finite verb form marked with a nominalizing suffix identical in form to the general locative case marker (and assumed by ‘some authors’ to be etymologically related to it—see Holisky (1991: 460) for references).

In the example (23), the subordinated clause ‘as I entered the water’ is preposed to the main clause and subordinated with the nominalizing suffix *-şa* which gives

it the reading 'as/during'. The translationally equivalent Turkish structure could also be a nominalization with a locative case marker, shown in (24), or a converb in *-ken* (see (34) below):

(23) Laz:

žari-ša	amafti-ša
water-LOC	enter.1sg-NOM
'As I entered the water'	

(24) Turkish (same meaning as (23))

su-ya	gir-diğ-im-de
water-DAT	enter-NOM-POSS1sg-LOC

This type of subordination is extremely common in Laz, and, more significantly, appears to be unparalleled in the other Kartvelian languages. Therefore it appears reasonable to assume Turkish influence (see Harris and Campbell (1995: 145) for further discussion).

The second type of subordinate clause I wish to look at is relative clauses. The Kartvelian languages Georgian, Svan, and Mingrelian all have—as one of their major relative clause strategies—relative clauses which follow their head nouns, and which are introduced by relative pronouns (for Svan see Tuite (1997: 42), for Georgian see Aronson (1991: 284–5), for Mingrelian Harris (1991b: 382–4)).

Laz also has post-head relative clauses, but unlike in other Kartvelian languages (e.g. Svan), post-head relatives in Laz are apparently only very rarely used (Holisky 1991: 457). In the Ardeşen dialect of Laz, however, there appear to be no post-head relative clauses at all; the sole strategy involves a nominalized clause which is preposed to the head noun:

(25) Laz, Ardeşen dialect (Dumézil and Ensenç 1972: 33):

na	golulun	k'oči
NOM	pass by.3sg	man
'the man who passed by'		

Compare this to the Turkish structure, which also uses a participle in the relative clause:

(26) önümnden geç-en adam
in.front.of.me pass-PART man
'the man who passed by me'

Holisky (1991: 459) points out the 'very interesting' fact that when such relative clauses are headless, the nominalized verb form 'can also bear the plural suffix *pe*'. Again, this is a regular feature of headless relative clauses in Turkish; the example quoted by Holisky would have had an identical structure and meaning in Turkish. Although right-headed relative clauses are by no means unknown in Kartvelian generally, their use as the sole strategy in Ardeşen Laz is highly suggestive of Turkish influence.

In causal subordinate clauses Turkish influence also appears likely. In Laz the predicate of such clauses is nominalized with the particle *na*, and followed by the postposition *şeni* ‘for’:

- (27) baba şiki but’uce-pi-şi zade
 father POSS1sg bee-PL-GEN much
 na nugnamtu şeni dido ğunne-pe miğurtei
 NOM understand.3sg.PAST for many beehive-PL have.1pl.PAST
 ‘Because my father knew a lot about bees we had many beehives.’

This is exactly parallel to Turkish, which uses a nominalized verb form followed by the postposition *için* ‘for’, also preposed in front of the main clause:

- (28) baba-m arıcılık-tan anla-dığı için ...
 father-POSS1sg beekeeping-ABL understand-NOM.POSS3s for ...
 ‘because my father knew about beekeeping ...’

All three types of subordinate clause are, in terms of linear order of the clauses and position of the boundary marker, identical to their Turkish translational equivalents. None of these three constructions plays a prominent role in clause subordination in Mingrelian, Laz’s closest genetic relative (at least in so far as the brief description in Harris (1991b) allows any conclusions). It would appear then that Laz, and in particular the Ardeşen dialect, has brought its techniques of clause linkage into line with Turkish patterns.

Turning now to Kurmanji or Zazaki, we find that neither of them has comparable constructions: *when*-clauses are usually introduced by some form of conjunction, relative clauses are post-head, introduced by *ku/ke*, and *because*-clauses are also introduced by a conjunction. I suggest that one reason behind the differences between Laz and the two Iranian languages is a structural feature of the Iranian languages: the Iranian languages of Anatolia have virtually no non-finite verb forms. For example, there simply is no active participle which could be used in participial relative clauses equivalent to (26). This is a fundamental feature of the Zazaki and Kurmanji verbal lexicon, with ramifications for the entire syntax (e.g. a lack of non-finite complements of verbs such as ‘want’); I believe that it has been a major obstacle to developing Turkish-type patterns of subordination. I would not, however, claim that this structural obstacle is insurmountable. Turkic-like constructions are found in Tajik, an Iranian language closely related to Kurdish, due to prolonged Uzbek influence—see Soper (1996). But it appears that this type of change requires longer and more intense contact than has yet been the case in East Anatolia.

4.2. BORROWING VERBS

It is often claimed that verbs are less readily borrowed than nouns (cf. Dixon 1997: 20). In Anatolia, all three minority languages have borrowed verbs from Turkish,

TABLE 2. Borrowed Turkish verb forms in Kurmanji (examples from Dorleijn 1996: 51 and Haig, forthcoming)

Kurmanji form	Meaning of Turkish verb + meaning of Kurmanji finite verb	Total meaning in Kurmanji
<i>tanişmiş bûn</i>	get to know + be	get to know
<i>nişanlanmiş bûn</i>	get engaged + be	get engaged
<i>başlamış kirin</i>	begin + do	begin
<i>sömürmüş kirkin</i>	exploit + do	exploit

TABLE 3. Borrowed Turkish verb forms in Zazaki (Paul 1998: 100)

Zazaki form	Meaning of Turkish verb + meaning of Zazaki finite verb	Total meaning in Zazaki
<i>dāmiş biyāyîş</i>	endure + be	endure
<i>qārmîş biyāyîş</i>	interfere + be	interfere
<i>dîşmiş biyāyîş</i>	think + do	think

but the means for doing so differ radically. The two Iranian languages borrow Turkish verb forms ending in *-mîş*, a suffix with a perfective meaning (and in some contexts, an evidential component). In Turkish, these verb forms can be used both as finite verbs and as participles. The strategy in Kurmanji and Zazaki (and indeed other Iranian languages in contact with Turkic, e.g. Tajik) is to combine a Turkish *mîş*-verb form with the native verb for 'be' or 'do'. Examples from Kurmanji and Zazaki are given in Tables 2 and 3. The predominant verb-borrowing strategy is thus:

Turkish *mîş*-verb form+Iranian *do/be*.

Now Iranian languages, like Indo-Aryan languages, make extensive use of combinations of often borrowed nominal elements plus a semantically bleached native 'light verb' to extend their verb lexicons. Typical examples are Kurmanji *qebûl kirin* 'accept' (lit. 'do acceptance', *qebûl* is borrowed from Arabic), or Zazaki *qezenj kerdîş* 'earn' (lit. 'earning do', *qezenj* is borrowed from Turkish). It seems reasonable that the prior existence of light verb constructions meant that there was a slot available into which Turkish verb forms could be fitted. What is unclear is why it should be almost exclusively Turkish *mîş*-forms that are borrowed into this particular structure, and not, say, Turkish infinitives.

Yet when we turn to Laz, there seems to be no regular pattern of borrowing *mîş*-verb forms; at least I found none in the available data. In fact, very few Turkish verbal lexemes are borrowed into Laz at all. The form that is borrowed is the Turkish root, which in Laz is treated like a Laz verb root. I have found three Turkish verbs in the Laz material: *düşün-* 'think', *şaş-* 'be bewildered', and *çalış-* 'work'. They are borrowed as roots and conjugated in the normal manner of Laz verbs, as in:

- (29) *iduşunai* ‘think.PRES.3sg’ (Wodarg 1995: 119)
 (30) *goişaşı-yi* ‘go crazy.2sg.INTERR’ (Dumézil and Ensenç 1972: 34)
 (31) *čalış-ap-s* ‘work.PRES.3sg’ (Holisky 1991: 439)

Laz, unlike Zazaki and Kurmanji, makes very little use of light verb constructions. I conjecture that this is one of the reasons for the difference in strategies of borrowing verbs.

The lack of borrowed verbs in Laz is certainly conspicuous when one considers the number of other borrowed lexical items and the high degree of structural influence. It is reminiscent of French influence on Algonquian languages, which has resulted in large numbers of borrowed French nouns but apparently no borrowed verbs (Bakker and Papen 1997: 354–5), or of Resígaro, which also borrows nominal elements freely, but no verbs (Aikhenvald, this volume). Both these languages, like Kartvelian languages, have extremely complex verb structures, where most of the grammatical information for the clause is indexed; it may be that head-marking languages of this type are generally more resistant to borrowing verbs.

4.3. EXTREME CONVERGENCE: ARDEŞEN LAZ

Finally, I would like to look at one Laz dialect in some detail, namely Ardeşen Laz, where structural parallels with Turkish have penetrated deeply into the morphosyntax. It has long been noted that the Pazar dialect group, to which Ardeşen Laz belongs, is the variety of Laz most heavily influenced by Turkish (Vanilişi and Tandilava 1992: 59, 73). However, as yet there has been no systematic treatment of Turkish influence on these dialects.

One of the most remarkable changes in Ardeşen Laz has been the restructuring of the case system. In Kartvelian languages, case-marking of core arguments is generally determined by the class of the predicate: predicates from different classes require different valency patterns. Furthermore, case-marking patterns are also dependent on the tense-aspect of the governing predicate, with ergative alignment with aorist tenses. Whether the resulting systems are really ergative, or split-intransitive, or what, is of little concern here; what is important in the present context is that the argument which corresponds roughly to the subject in Standard Average European, and in Turkish, may take a variety of different case forms depending on the class and the tense of its governing verb.

Laz differs from the other Kartvelian languages in that there is no tense/aspect split in the nominal case-marking, but most varieties of Laz retain the complexities of nominal case marking determined by the different verb classes. The distribution of nominal case across syntactic functions for most Laz dialects is given in Table 4. Ardeşen Laz on the other hand has undergone radical restructuring of its nominal case system, resulting in the distribution seen in Table 5. The most striking difference between the system shown in Table 4 and that of Table 5 is that in

TABLE 4. Case marking of core arguments in Laz (Holisky 1991: 446)

S	A	O	Indirect object	Goal
-Ø	-k	-Ø	-s	-še ~ ša
-k	-s		-še~-ša	
-s				

TABLE 5. Case-marking of core arguments in the Ardeşen dialect of Laz

S	A	O	Indirect object	Goal
-Ø	-Ø	-Ø	-Ø	-š

TABLE 6. Case-marking of core arguments in Turkish

S	A	O	Indirect object	Goal
-Ø	-Ø	-Ø~- (y)I	- (y)A	- (y)A

Table 5, the S/A category is given unified treatment. In other words, there is no mismatch between nominal morphology and syntactic status, i.e. syntactic subjects, or pivots, are given unified treatment in the morphology. Direct objects on the other hand remain unmarked.

The net result of these changes is to bring Ardeşen Laz nominal morphology much closer to Turkish than the other Laz dialects are, albeit without any borrowing of actual forms. Compare the Ardeşen Laz system with the Turkish one, given in Table 6.

The zero-marking of indirect objects in Ardeşen Laz remains of course a major difference. However, in my own field work with three young Turkish/Laz bilinguals now living in Ankara I found that they consistently use the goal-marker for indirect objects, illustrated in the following example, with the Turkish equivalent given below:

- (32) Young urban Turkish-Laz bilinguals:
 k'oç'i laçi-şa xordži meçai
 man dog-DAT meat give.3sg.PRES
 'The man gives meat to the dog.'

- (33) Standard Turkish
 adam köpeğ-e et veriyor
 man dog-DAT meat give.PRES.3sg

The case system for these young speakers is given in Table 7.

TABLE 7. Case marking of core arguments in the speech of young urban Laz/Turkish bilinguals when speaking Laz

S	A	O	Indirect object	Goal
-Ø	-Ø	-Ø	-ša	-ša

Comparing the case systems in different varieties of Laz (Tables 4, 5, and 7), we can readily discern an overall direction of change: the system of case-marking of core arguments is progressively approaching that found in Turkish (Table 6), albeit without borrowing a single piece of morphological material. At present it is impossible to say whether the system given in Table 7 will remain restricted to the speech of a few urban semi-speakers, or whether it is a precursor of future systematic developments in the core of the Laz speech community. But it certainly shows without doubt one potential path of development in the case system, and it shows that such paths can be contact-driven.

As a final illustration of the parallels between Laz and Turkish consider the following short text extract, taken from Dumézil and Ensenç (1972: 33). I have provided a Turkish translation in the form of a word-for-word gloss, which results in a perfectly natural Turkish rendering of the same story. Thus we find almost complete isomorphism in the order of constituents down to word level, and in some cases morpheme level. The English gloss is a simplified one that ignores some of the sub-word-level morphological distinctions. Where the Laz and Turkish texts require different glosses, these are separated by a slash (/) in the glosses.

- (34) L. xoj'a Nusrettin a ndya çarşı-ša it't'u-ša koçe-pe him uc'vey çi:
 T. Hoca Nasreddin bir gün çarşı-ya gider-ken insan-lar ona:
 Hoca Nasreddin one day market-DAT go-SUB person-PL 3sg.DAT say that
- (35) L. xoj'a, si iri-tulli kogişkun
 T. Hoca sen her şeyi biliyorsun
 Hoca, you every thing you.know
- (36) L. ham yat't'ile gazit't'asen-i
 T. bunu bakalım bilebilecekmisin
 this let's.see you.can.say.INTERR /T: you.can.know.INTERR
- (37) L. xoj'a: 'peki, mic'vitu' t'k'u
 T. Hoca: 'peki sorun' dedi
 Hoca: 'alright say.IMPV/T: ask.IMPV:' he.said
- (38) L. 'xoj'a, dunya nak'k'u metre onu?' diye ç'itxey
 T. 'Hoca, dünya kaç metre dir?' diye sordular
 Hoja, world how.many meter is' saying they-asked

Translation:

'One day, as Hoca Nasreddin was going to market, the people said to him

"Hoca, you know everything.

Let's see if you can tell (= answer) this."

"Alright, ask!" he said.

"Hoca, how many metres (long) is the world?" they asked.'

In terms of linear ordering of free constituents, Turkish and Ardeşen Laz approach full isomorphism. In the morphology too there are considerable parallels. However, the Kartvelian polypersonal predicate complex has remained impenetrable to foreign influence, largely retaining its common Kartvelian profile.

4.4. TURKISH INFLUENCE ON EAST ANATOLIAN MINORITY LANGUAGES: SUMMING UP THE DIFFERENCES

Turkish influence has affected the minority languages in rather different ways. Generally, it seems that certain dialects of Laz have been more strongly influenced than the Iranian languages. I suggested that a greater degree of initial structural compatibility between Laz and Turkish may have contributed to this. In fact, Kurmanjî and Zazaki have, despite a heavy influx of loanwords, retained a remarkably stable structural profile (e.g. retention of gender, retention of head-modifier order in the NP, conjunctions in subordinate clauses, ergative syntactic alignment with past tenses).

But although inherited Iranian structural features may have inhibited closer structural convergence with Turkish in Anatolia, they cannot be considered obstacles in an absolute sense, for elsewhere Iranian languages have moved further towards Turkish, e.g. the adoption of Turkic verb serialization in Tajik, and modifier-head order in the NP (adjective-head order in Baluchi, and possessor-possessed order in Tajik (Windfuhr 1987: 544). Therefore, extra-linguistic factors must also be involved in the Anatolian case, for example the greater sizes of the Kurmanjî and Zazaki speech communities compared to Laz. But again, this remains pure speculation in the absence of any comparative data on patterns of multilingualism in the area.

5. Conclusions: patterns of borrowing and borrowing of patterns

The literature on language contact is littered with disproved and discarded 'structural universals of borrowing' (see critical discussion in Johanson (1992), Campbell (1993), Harris and Campbell (1995: 122–36) and Curnow, this volume). I will nevertheless risk drawing some more general conclusions regarding patterns of borrowing, based primarily on the Anatolian data but supplemented by data from further afield. I should note that similar conclusions have been reached by other scholars (e.g. Ross, this volume), though some differences in detail remain, and I am unaware of any more explicit statement than the present one.

Two much-neglected aspects of contact-induced change are highlighted by Ross (Ch. 6), namely the 'reorganization of the language's semantic patterns' and 'the ways of saying things'. Both are, as Ross points out, intrinsic components of structural convergence. Together with the restructuring of the syntax, they make up the package collectively referred to by Ross as *metatypy*, i.e. a change of type. The important point is that structural convergence does not proceed in isolation, but is accompanied by a restructuring of underlying cognitive patterns. However, in this section I will be dealing solely with observable features of morphosyntax. Although it is undoubtedly true that the languages of East Anatolia also display remarkable parallels in semantic structuring, they must remain a topic for a future study.

A striking characteristic of much contact-driven structural change is that what is almost invariably affected is the surface linear order of constituents. In many contact situations we find a clearly discernible drift towards structural isomorphism, a realignment of various constituents to bring them into line with comparable elements in the contact language. In what follows I will refer to the process of bringing semantically and functionally comparable constituents into comparable positions, of creating what Johanson (1992: 15) calls 'equivalence positions', as 'linear alignment'.

Unlike the term *metatypy* (Ross, Chapter 6), this term does not necessarily imply a matching of semantic structure, although like Ross I believe linear alignment is almost certainly accompanied by semantic restructuring. Rather, linear alignment is, at least in principle, independent of both semantic restructuring, and in principle independent of whether actual morphological material is copied. That in many cases linear alignment, semantic restructuring, and borrowing of actual material go hand in hand is undeniable, but the three can nevertheless be kept distinct. The following simple example of linear alignment from the languages of Anatolia should illustrate this. In written Kurmanji Kurdish, the numerals 11–19 have the form 'x-ten'. For instance, 14 is *çardeh*, where *çar* is the numeral 4, and *deh* is the numeral 10. Furthermore, we find typically Indo-European irregular forms for the numerals 11 and 12, as we do in Persian, so this is certainly the original Kurdish pattern. But in strongly Turkish-influenced varieties, e.g. that of the Tunceli region (see Haig, forthcoming, for details), the numerals from 11 to 19 all have the reverse order of digits. For instance, the numeral 14 is *dehuçar*, i.e. ten-four. Furthermore, the irregular forms for 11 and 12 are now fully transparent, and conform with the new pattern. The model for the inversion of the order of digits is undoubtedly Turkish. In Turkish, the numerals 11 to 19 are all fully regular and follow the pattern '10-x' (*on-bir* 'ten-one' = 11, and so on). Note that beyond 19, Turkish and Kurdish numerals have the same order. What this simple example demonstrates is that realignment of linear order, bringing comparable elements into comparable positions, is a factor operative at various levels, and one that is potentially independent of both semantic restructuring, and of actual transfer of morphological material.

Given that some degree of linear alignment is an outcome of many contact situations, it would be desirable to articulate some more general predictions on how it may, or may not, proceed. In what follows, I will suggest that the relevant structural parameter in terms of which such predictions can be stated is level of grammatical organization. By this I mean a rough hierarchy running from the highest level, namely that of clause linkage, through basic phrase structure, i.e. major constituents of the simple clause, down through the structure of the NP, and finally on to the internal structure of the word. At the top of the scale, the grammar of clause linkage shades gradually into the grammar of discourse, with no clear dividing line.

My specific claim is that linear alignment will proceed from larger to smaller units, starting perhaps with the narrative organization, means of expressing direct speech, topic introduction and tracking, and progressing down through clause coordination, subordination, and constituent order in the clause. I would, however, exclude grammatical subsystems such as the numerals from the domain of these generalizations.

Why should linear alignment start with the largest constituents? There are a number of reasons: first, and most important, there is usually greater positional freedom of elements at this level. For example, many languages allow both possible ordering of main and subordinate clause for a large number of constructions. Therefore, the order found in the contact language is often available as at least a secondary option in the affected language anyway. Secondly, clauses are probably universal units, marked by intonational contours in all languages, i.e. they are perceptually easier units to recognize and to match to one's own language. Finally, realigning the order of clauses is relatively independent of the more rigid parts of the grammar, i.e. the order of morphemes in the word, etc. Below the NP, and in morphology generally, language structure is more tightly regimented by the typological profiles of the languages concerned.

The claim that linear alignment proceeds from larger to smaller units leads to certain empirically testable predictions. For example, we would predict that if a language A influences a language B through prolonged interaction, patterns of clause linkage will be the first items of B which will realign to match the order of A, followed by basic clause constituents, i.e. subject, object, and verb (or possibly relative clauses—see below). But—and this is the crucial point—linear alignment at lower levels will not precede linear alignment at higher levels. I am aware of a few cases which seem to confirm this prediction: Meso-America contains only non-verb-final languages, but at least one, Mixe-Zoquean, has retained traces of an earlier verb final order, e.g. postpositions. This suggests that its basic constituent order has adapted to the areal profile, but its adpositional order has not (Campbell, Kaufman, and Smith-Stark 1986: 547–8). Another example which at least superficially seems to confirm the expectations is Cushitic influence on Amharic (Harris and Campbell 1995: 138). The above claim would be falsified if we were to find a language which had, for example, through contact with another

language shifted from prepositions to postpositions, but its basic word order remained distinct from the contact language.

Of course it is well known that the order of elements within different types of constituents appears to be subject to certain universal constraints. For example, verb-initial order in the clause correlates strongly with prepositions as opposed to postpositions (Greenberg's Universal no. 3). It has been suggested that certain combinations of constituent order are universally favoured, or less marked, than others. If that is the case, one could expect changes in constituent order at one level to trigger a chain of changes at other levels as the language strives to comply with one of the supposed universally favoured combinations. This possibility is discussed in Harris and Campbell (1995: 140–1). However, Comrie (1989: 100) notes that more than half the world's languages do not conform to ideal types, so the internal pressures for such wholesale shift cannot be that great. Whatever the effects of language-internal pressures in affecting changes in constituent order, they do not actually impinge on the claims made here, which are concerned solely with constraints on the relative ordering of contact-induced shifts in constituent order.

If linear alignment of higher-level constituents really does occur relatively early, then we would expect to find that patterns of clause linkage, and of basic constituent order are features which diffuse quickly across large areas, cross-cutting genetic groupings. Again, there is considerable support for this in the literature. Large and linguistically diverse areas such as Africa and Papua New Guinea show remarkable parallels in techniques of clause linkage. Soper (1996) has documented linear alignment at this level in two case studies of Iranian–Turkic language contact. East Siberia is also a linguistically diverse area united by the use of converb clause chaining (Nedjalkov 1998). Campbell (1987) notes how clause coordination in Pipil (Uto Aztecan) has adapted to Spanish, and the data presented above document a number of striking similarities among genetically diverse languages in East Anatolia. Similarly, basic constituent order is also prone to spread—see the discussion with references in Harris and Campbell (1995: 136–7). As a consequence, a particular constituent order is very often a shared trait across large geographically contiguous areas, as in for instance Africa (see Heine and Kuteva, this volume), the Indian subcontinent, Meso-America, the Baltic, Ethiopia, and of course Western Europe (Campbell 1998: 301–6). There are also well-documented examples of languages shifting their basic constituent order under areal pressure, for example the Khamti dialect of Thai, spoken in Assam, which has shifted from Thai SVO to 'Brahmaputra areal' SOV order (Diller 1992: 20). There seems then little doubt that the larger constituents show a marked tendency to align under contact conditions. These facts have an important implication for the topic of the present volume: commonalities in clause linkage, and in basic constituent order, are, taken in themselves, poor evidence for genetic relationships, but good indicators of language contact (see Bisang 1998 on clause linkage in so-called Altaic languages).

As far as smaller units are concerned, e.g. constituents of noun phrases and

prepositional phrases, I make no specific claim on the relative order of alignment. I do not know whether, for example, the order of nouns and adjectives is likely to align before, say, that of genitive attribute and head noun, and it is quite possible that no preferences will be found. It is however worth noting that as far as noun/adposition order is concerned, languages tolerate more than one order. Thus borrowed adpositions tend to retain the order of the donor language rather than adapt to the recipient language (Harris and Campbell 1995: 136). For example Iranian Azeri, normally postpositional, has borrowed Persian prepositions and uses them as prepositions, not as postpositions (Dehghani 1998a: 219). However, one counter-example to this tendency is found in Basque, which has borrowed the Spanish preposition *contra* but uses it as a postposition (Trask 1998: 320).

The above generalizations on the relative order of alignment run into difficulties with relative clauses. In terms of hierarchical status, they are, like adjectives, constituents of NP, and should therefore be expected to align after basic constituent order has done so. In terms of syntactic weight, however, they are potentially of the same order as subordinate clauses. Now if we assume that the guiding principle behind the sequence of linear alignment is 'higher-level constituent before lower-level constituent', then relative clauses would be expected to pattern like adjectives, i.e. to realign relatively late. If, however, the guiding principle is 'heavy before light', then we would expect relative clauses to align earlier. The available evidence suggests that the latter is more likely: there are cases of languages realigning the relative order of head noun and relative clause with that of a contact language. For example relative clauses of the Turkish type are attested in Asia Minor Greek and Armenian (Johanson 1992: 112–13), while post-head Persian relative clauses occur regularly in the Turkic language Azeri (Dehghani 1998a: 225–6). The most striking evidence for the relative ease of alignment in relative clauses comes from Basque. According to Trask (1998), Basque has recently developed a type of post-head relative clause quite distinct from the inherited pre-head type, and clearly modelled on the surrounding Romance languages. While this development, given the geographic and historic situation of Basque, would be in itself not particularly surprising, it is remarkable because Basque has otherwise resisted linear alignment with its neighbours, remaining for example stubbornly SOV. It should be noted, however, that none of the languages discussed above has abandoned its inherited relative-clause type entirely; rather, they now have an additional relative-clause strategy, based on that of the contact language.

The evidence from relative clauses thus suggests that the primary factor determining ease of linear alignment is not level of syntactic organization, but syntactic weight. Now, generally, high syntactic level and syntactic weight correlate quite closely, so the two factors will mutually reinforce each other. In the case of relative clauses, however, we have conflicting motivations, and here it seems that syntactic weight is ultimately the more powerful determinant.

The claim that heavy syntactic constituents are most susceptible to realignment has a correlate in patterns of code-switching. Several researchers have noted that

code-switching is most likely to occur at the boundaries of higher-level constituents (see e.g. Romaine 1995: 124, Appel and Muysken 1988: 172). The link to the notion of linear alignment developed above is clear, and is explicitly discussed by Appel and Muysken (1988: 172): what is common to both is that maximally manipulable constituents are affected first. This is not a surprising result; linear alignment is driven by multilingual discourse involving frequent code-switching, and is therefore common to very specific types of contact situation.

Before closing, let us briefly touch on another type of contact-driven alignment of linguistic elements, neatly illustrated by the convergence of the case system of Ardeşen Laz with that of Turkish (see §4.3). Here it is not the linear order of elements in actual discourse, but the underlying systems that converge, the number and type of grammatical distinctions that together make up a particular paradigm. But like linear alignment, paradigmatic alignment does not necessarily involve any actual borrowing of morphological material (cf. 'indirect diffusion', Heath 1978: 119). Further examples are readily available: the restructuring of verbal categories in Tariana on a Tucanoan model (see Aikhenvald, this volume) and of nominal morphology in Asia Minor Greek on a Turkish model (Sasse 1992: 65–6). It is fairly clear that this type of 'paradigmatic alignment' should be kept distinct from the linear alignment discussed in detail above, but it is also clear that the two are not fully independent. Just how they interact remains one of the most fascinating issues for future research in contact linguistics.

References

- ALAVI, B., and LORENZ, M. 1988. *Lehrbuch der persischen Sprache*. Leipzig: VEB Verlag.
- ANDREWS, P. 1989. Editor of *Ethnic Groups in the Republic of Turkey*. Wiesbaden: Reichert.
- APPEL, R., and MUYSKEN, P. 1988. *Language Contact and Bilingualism*. London: Edward Arnold.
- ARONSON, H. 1991. 'Modern Georgian', pp. 219–312 of Harris (1991).
- BAKKER, E. 1993. 'Boundaries, topics, and the structure of discourse. An investigation of the Ancient Greek particle *DE*', *Studies in Language* 17:275–305.
- BAKKER, P., and PAPEN, R. 1997. 'Michif: a mixed language based on Cree and French', pp. 295–363 of *Contact Languages: A Wider Perspective*, edited by S. Thomason. Amsterdam: John Benjamins.
- BARNAS, U., and SALZER, J. 1994. *Lehrbuch der kurdischen Sprache*. Berlin: İnstitutya Kurdî.
- BEDİR KHAN, E. D., and LESCOT, R. 1986. *Kurdische Grammatik Kurmanji-Dialekt*. Bonn: Verlag für Kultur und Wissenschaft.
- BEHZAD, F., and DIVSHALI, S. 1994. *Sprachkurs Persisch. Eine Einführung in die persische Sprache der Gegenwart*. Bamberg: Edition Zypressen.
- BISANG, W. 1998. 'Structural similarities of clause combining in Turkic, Mongolian, Manchu-Tungusic and Japanese—a typological alternative to the hypothesis of genetic relatedness', pp. 199–223 of *The Mainz Meeting. Proceedings of the Seventh International Conference on Turkish Linguistics, August 3–6, 1994*, edited by L. Johanson. Wiesbaden: Harrassowitz.
- BULUT, C. 2000. 'Indirectivity in Kurmanji', pp. 147–84 of *Evidentials. Turkic, Iranian and neighbouring languages*, edited by L. Johanson and B. Utas. Berlin: Mouton de Gruyter.

- . 2005. 'Turkish elements in spoken Kurmanji', pp. 95–121 of *Turkic languages in contact*, edited by H. Boeschoten and L. Johanson. Wiesbaden: Harrassowitz.
- . 'Indirectivity in Kurmanji'. Manuscript.
- CAMPBELL, L. 1987. 'Syntactic change in Pipil', *Journal of Anthropological Linguistics* 53:3.253–80.
- . 1993. 'On proposed universals of grammatical borrowing', pp. 91–109 of *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, Rutgers University, 14–18 August 1989*, edited by H. Aertsen and R. J. Jeffers. Amsterdam: John Benjamins.
- . 1998. *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- KAUFMAN, T., and SMITH-STARK, T. 1986. 'Meso-America as a linguistic area', *Language* 62.530–70.
- COMRIE, B. 1989. *Language Universals and Linguistic Typology*. 2nd edition. Chicago: University of Chicago Press.
- DAWKINS, R. M. 1916. *Modern Greek in Asia Minor: A Study of the Dialects of Silli, Cappadocia and Phárasa with Grammars, Texts, and Glossary*. Cambridge: Cambridge University Press.
- DEGHANI, Y. 1998a. 'A Grammar of Iranian Azari Including Comparisons with Persian'. Ph.D. dissertation. School of Languages, Cultures and Linguistics, La Trobe University, Melbourne.
- . 1998b. 'Morphological and syntactic borrowing in Azari', *Melbourne Papers in Linguistics* 17.1–10.
- DILLER, A. V. N. 1992. 'Tai languages in Assam: daughters or ghosts?', pp. 5–43 of *Papers on Tai Languages, Linguistics, and Literatures, in Honor of William J. Gedney on his 77th Birthday*, edited by C. J. Compton and J. E. Hartmann. Northern Illinois University: Center for Southeast Asian Studies.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DORLEIJN, M. 1996. *The Decay of Ergativity in Kurmanji. Language Internal or Contact Induced?* Tilburg: Tilburg University Press.
- DUMÉZIL, G., and ENSENC, T. 1972. 'Textes en Laze d'Ardeşen', *Bedi Kartlisa* 29–30.32–41.
- ELFENBEIN, J. 1989. 'Balōčī', pp. 350–62 of *Compendium Linguarum Iranicarum*, edited by R. Schmitt. Wiesbaden: Reichert.
- EMENEAU, M. 1980. *Language and Linguistic Area. Essays by Murray B. Emeneau, selected and introduced by Anwar S. Dil*. Stanford: Stanford University Press.
- FEUERSTEIN, W. 1994. 'Bir alman gözüyle: Lazlar', *Ogni* 1:2.19–22.
- GRANNES, A. 1978. 'Le redoublement turk à M initial en bulgare', *Linguistique Balkanique* 21.37–50.
- HAIG, G. Forthcoming. 'Turkish influence on Kurmanji Kurdish: evidence from the Tunceli dialect', to appear in *Studies in Turkic-Iranian contiguity*, edited by C. Bulut and F. Kiral. Wiesbaden: Harrassowitz.
- HARRIS, A. 1991a. *The Indigenous Languages of the Caucasus*, vol. 1: *The Kartvelian Languages*. New York: Caravan.
- . 1991b. 'Mingrelian', pp. 313–94 of Harris (1991a).
- and CAMPBELL, L. 1995. *Historical Syntax in a Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- HASPELMATH, M. 1993. *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.

- HEATH, J. 1978. *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- HOCK, H. H. 1991. *Principles of Historical Linguistics*. 2nd edition. Berlin: Mouton de Gruyter.
- HOLISKY, D. A. 1991. 'Laz', pp. 395–472 of Harris (1991a).
- JOHANSON, L. 1992. *Strukturelle Faktoren in türkischen Sprachkontakten*. Stuttgart: Franz Steiner.
- 1998. 'Code-copying in Irano-Turkic', *Language Sciences* 20.325–37.
- KAHN, M. 1976. *Borrowing and Regional Variation in a Phonological Description of Kurdish*. Ann Arbor, Michigan: Phonetics Laboratory of the University of Michigan.
- KUTSCHER, S. 1995. 'Nomen und nominales Syntagma', pp. 21–35 of Kutscher, Mattissen, and Wodarg (1995).
- MATTISSEN, J., and WODARG, A. 1995. Editors of *Das Mutafi-Lazische*. Köln: Universität Köln.
- LEWIS, G. 1967. *Turkish Grammar*. Oxford: Oxford University Press.
- MCCARTHY, J. 1983. *Muslims and Minorities*. New York: New York University Press.
- MACKENZIE, D. 1961. *Kurdish Dialect Studies*, vol. 1. London: Oxford University Press.
- MATRAS, Y. 1988. 'Utterance modifiers and universals of borrowing', *Linguistics* 36.281–331.
- MORAVCSIK, E. 1978. 'Language contact', pp. 93–123 of *Universals of Human Language*, vol. 1: *Methods and Theory*, edited by J. H. Greenberg, C. Ferguson, and E. Moravcsik. Stanford: Stanford University Press.
- NEDJALOV, I. 1998. 'Converbs in the languages of Eastern Siberia', *Language Sciences* 20.339–51.
- NESIN, AZİZ. 1995. *Sizin memlekette eşek yok mu?* Istanbul: Doğan Yayın Holding.
- NESTMANN, L. 1989. 'Die ethnische Differenzierung der Bevölkerung der Osttürkei in ihren sozialen Bezügen', pp. 543–81 of Andrews (1989).
- PAUL, L. 1998. *Zazaki. Grammatik und Versuch einer Dialektologie*. Wiesbaden: Reichert.
- ROMAINE, S. 1995. *Bilingualism*. 2nd edition. Oxford: Blackwell.
- SASSE, H.-J. 1992. 'Language decay and contact-induced change: similarities and differences', pp. 59–80 of *Language Death: Factual and Theoretical Explorations with Special Reference to East Africa*, edited by M. Brenzinger. Berlin: Mouton.
- SOPER, J. 1996. *Loan Syntax in Turkic and Iranian*. Bloomington, IN: Eurolingua.
- STASSEN, L. 1985. *Comparison and Universal Grammar: An Essay in Universal Grammar*. Oxford: Blackwell.
- THOMASON, S. G., and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- TRASK, L. 1998. 'The typological position of Basque: then and now', *Language Sciences* 20.313–24.
- TSCHENKEL, K. 1958. *Einführung in die georgische Sprache*, vol. 1: *Theoretischer Teil*. Zurich: Amirani.
- TUITE, K. 1997. *Svan*. Munich: Lincom Europa.
- VAN BRUINSEN, M. 1992. 'Kurdish society', pp. 33–67 of *The Kurds: A Contemporary Overview*, edited by P. Kreyenbroek and S. Sperl. London: Routledge/SOAS Politics and Culture in the Middle East Series.
- VANILIŞI, M., and TANDILAVA, A. 1992. *Lazların tarihi*. Istanbul: Ant Yayınları.
- WINDFUHR, G. 1987. 'Persian', pp. 523–46 of *The World's Major Languages*, edited by B. Comrie. London: Croom Helm.
- WODARG, A. 1995. 'Textsammlung', pp. 113–32 of Kutscher, Mattissen, and Wodarg (1995).

The Role of Migration and Language Contact in the Development of the Sino-Tibetan Language Family

Randy J. LaPolla

1. Introduction

A strong case can be made for a genetic linking between the Sinitic languages (the Chinese ‘dialects’) and the Tibeto-Burman languages. There are hundreds of clear cognates of basic vocabulary (Benedict 1972, Matisoff 1978, Baxter 1995; see LaPolla 1994a for a list of two hundred of the most uncontroversial) as well as some derivational morphology that can be reconstructed to the Proto-Sino-Tibetan level.¹ Within Tibeto-Burman again we find hundreds of cognates of basic vocabulary, and there are some relatively uncontroversial groupings based on shared innovations, such as Lolo-Burmese, Bodish, Qiangic, and Karenish, but subgrouping within Tibeto-Burman (and to some extent within Sinitic) is quite problematic. Benedict (1972; see the Figure) had Tibeto-Karen as one of two branches of Sino-Tibetan (the other being Chinese), with Tibeto-Burman and Karen being the two highest branches of Tibeto-Karen. Karen was given this position because it has verb-medial word order rather than the usual verb-final order of Tibeto-Burman. However, most linguists working on Tibeto-Burman now consider Karen to be a branch within Tibeto-Burman, as they assume that Karen word order changed due to contact with Mon and Tai, and therefore is not an important factor to be used in genetic grouping.² As can be seen from the

¹ What can be reconstructed is an *-s- causative and denominative prefix (Mei 1989), possibly alternation of voicing and/or aspiration of initials for causatives, a *-t suffix for transitivity (Benedict 1972: 98–102, Michailovsky 1985, van Driem 1988), and a nominalizing *-n suffix (see LaPolla 1994a, Jin 1998). There is no evidence of relational morphology at the Proto-Sino-Tibetan or Proto-Tibeto-Burman levels (for discussion see LaPolla 1992a, 1992b, 1994b, 1995).

² Forrest 1973 had suggested Karen was so similar to Mon that it could just as easily be a Mon language influenced by Tibeto-Burman as a Tibeto-Burman language influenced by Mon. Luce (1976: 33) states that Karen is neither a Tibeto-Burman language nor a Mon-Khmer language, though it has been heavily influenced by both Tibeto-Burman and Mon-Khmer. He says it is ‘pre-Tibeto-Burman’. Not many scholars working on Karen would agree with these assessments now.

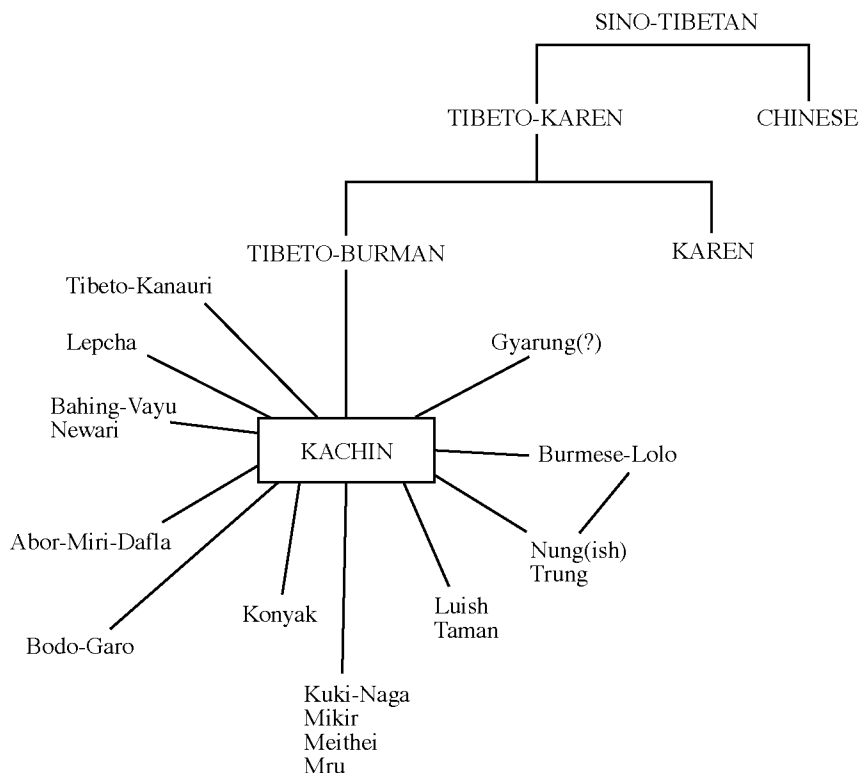


FIGURE. Schematic chart of Sino-Tibetan relations (from Benedict 1972: 6)

Figure, Benedict's model for relationships within Tibeto-Burman is not a family tree, as it represents 'an interlocking network of fuzzy-edged clots of languages, emitting waves of mutual influence from their various nuclear ganglia' (Matisoff 1978: 2). Matisoff (1978) shows that the evidence from Tibeto-Burman does not support a clear tree model. Rather there are waves of mutual influence, particularly in the spread of word families.³ On the Sinitic side, Pulleyblank (1991) has argued that the traditional Stammbaum model is also inappropriate for the Chinese dialects. He argues instead for 'some kind of network model, with provincial and regional centers of influence as well as successive national centers of influence in the form of standard languages based on imperial capitals' (Pulleyblank 1991: 442).

A major problem is the relationship of the Tai and Hmong-Mien (Miao-Yao)

³ Several other proposals on the subgrouping of Sino-Tibetan and/or Tibeto-Burman are Bradley 1997, Burling 1983, Dai, Liu, and Fu 1989, DeLancey 1987, 1991, Grierson 1909, Li Fang-kuei 1939, Shafer 1955, 1966, and Sun 1988.

languages to Chinese or Sino-Tibetan as a whole, that is, whether we consider the similarities among Chinese, Tai, and Hmong-Mien to be due to contact or due to genetic inheritance. Many scholars in China argue that the languages are related, but most linguists outside China feel the shared words are very old loans, and the other features, such as the similarities in the tone systems, the use of the classifier for definite marking, etc. spread areally. This makes it similar to the case of Vietnamese, which at one time was also thought to be related to Chinese, due to its many Chinese-like features and words, but is now thought to be a Mon-Khmer language heavily influenced by Chinese.

Three main factors have been involved in the formation of the present-day Sino-Tibetan language family: a shared genetic origin, divergent population movements (i.e. innovations appearing after these splits), and language contact. Population movements and language contact have in fact generally been two aspects of a single phenomenon. It is this fact that is the link between Dixon's (1997) view of rapid change due to non-linguistic causes and Heath's (1997, 1998) view of rapid change due to intense language contact, discussed by Watkins (this volume). The present chapter will look at the history of the development of this family from the point of view of population movements and language contact, to show the role language contact has had in the formation of the family as we know it today.

2. The migrations and their effects

From what we can piece together from the archeological and linguistic evidence (see for example Chang 1986, Treistman 1972, Pulleyblank 1983, Fairbank, Reischauer, and Craig 1989, Xing 1996, Ran and Zhou 1983), it seems the Sino-Tibetan-speaking people (if we associate the Neolithic Yang-shao culture with the Sino-Tibetans) originated in the central plains of what is now north China, in the valley of the Yellow River. At least 6,500 years ago, some members of the original group moved largely south and east, while others moved largely westerly at first, then moved in a southerly or south-westerly direction. Differences in identity and possibly language were evident at the time of the earliest Chinese writing, about 3,000 years ago, but there continued to be contact between the two related groups and others that surrounded them in the early period (see, for example, Wang Huiyin 1989), and frequent mixing of peoples (for example, the ancestors of some early Chinese rulers are said to have been from the western group—Ran and Zhou 1983, Ran, Li, and Zhou 1984, FitzGerald 1961). The group that stayed in the central plains, including those members of the western group that stayed in the central plains and nearby areas, as well as those who moved south-easterly, eventually became what we think of as the Chinese, while the group that moved south-westerly became what we think of today as the Tibeto-Burmans.

2.1. SINITIC

The movements in both directions were not single movements, but consisted of larger or smaller waves of movement, often into the same areas. Government-encouraged migration was practised as early as the Yin dynasty (roughly 1600–1027 BC), and has been practised by all Chinese governments up to the present one. There have also been massive private migrations and shifts of national or regional capitals due to natural disasters, war, and the pull of new economic opportunities (Ge, Wu, and Cao 1997).

The movement of the Chinese has almost never been to an area where there were no people. Splitting of the language by migration almost always involved language contact, either with non-Chinese languages or other Chinese dialects, and very often in government-sponsored migrations there was purposeful mixing of peoples. What we now think of as the Han Chinese have from very early on continually absorbed other peoples into the race (Wang Ming-ke 1992, Wiens 1967, Xu 1989). As the Chinese moved into new areas, they often absorbed the peoples there into the Han (Chinese) nationality, or, in some cases, were absorbed by the local nationalities (see, for example, Dai, Liu, and Fu 1987 and He 1989, 1998 for a case of Mongolian soldiers and settlers sent to the south-west in the Yuan dynasty (1234–1368) being absorbed into the Yi culture and developing a new language).

Table 1 summarizes the major movements, giving the time period, the place the population moved from and the place they moved to, the number of people who moved, if it is known from government records, and the original inhabitants of the area they moved to (data mainly from Lee 1978, 1982, Lee and Wong 1991, Zhou 1991, Ge, Wu, and Cao 1997).

It can be seen from Table 1 that many of the movements were chain movements. For example, the movement of over two million non-Chinese people into the central plains from the northern steppes in the second and third century caused at least three million Chinese to flee south. To give one example of how drastically these movements affected the populations, according to Lee (1978: 29), in one county (Bingzhou in Shanxi), two-thirds of the population emigrated between 289 and 312. This not only affected the population of the north, but also of the south, as one out of every six people in the south was a displaced northerner after the movement. Nanjing became the capital of the Eastern Jin (317–420) and Southern (420–589) dynasties; it attracted over 200,000 migrants, a figure greater than the original local population. The form of speech in the area then changed from a Wu dialect to a northern dialect. The speech of another Wu area, Hangzhou, became what Zhou and You (1986: 19) call a ‘half-Guanhua (Mandarin)’ area because of the shift of the Song dynasty capital from the north to Hangzhou in 1127 and the resulting massive influx of northerners. While the phonology is basically that of a Wu dialect, it is lexically and grammatically more similar to the northern dialects, and does not have the usual literary/colloquial reading distinction of characters that the other Wu dialects have.

TABLE 1. Some Major Population Movements in China

Century	Moved from	Number	Moved to	Original inhabitants
BC: 7th	Wei River valley (Shaanxi)	—	lower Yangtze	Bai Yue
6th	central plains (between Yellow and Yangtze rivers)	—	Han River and middle Yangtze (Hubei)	Bai Yue
3rd	Han/Middle Yangtze	—	Xiang River (Hunan)	Bai Yue
3rd–2nd	central plains	1.9 million	Hunan/Jiangxi/Guangdong/Guangxi/northern Vietnam	Bai Yue
2nd	Henan/Hebei/Shandong	155,000	Jiangsu/Zhejiang	Wu Chinese/Bai Yue
2nd	Henan/Hebei/Shandong	580,000	Gansu/Ningxia/Mongolia	Tungusic/Mongol
AD: 213	Fujian (Min-Yue people)	—	Yangtze/Huai River	Chinese
1st–2nd	Jiangsu/Zhejiang (orig. Wu speakers)	—	Fujian	Yue/Min-Yue
3rd–4th	Jiangsu/Zhejiang (later Wu speakers)	—	Fujian	early Wu speakers/Min-Yue
2nd–3rd	northern steppes (Tungusic)	2 million	central plains	Chinese
2nd–4th	central plains	3 million	Jiangxi/Zhejiang/Jiangsu	Wu/Chu Chinese
3rd	Shanxi	—	Hebei	
3rd	Hebei	200,000	north-east China	Altaic
4th	Shaanxi (Di and Chinese)	several hundred thousand	Sichuan/Yunnan	
4th	Sichuan	tens of thousands	Hunan/Hubei	
8th	central plains	hundreds of thousands	Hunan/Hubei/Jiangsu/Jiangxi	
9th	central Jiangxi	—	Fujian/Guangdong/eastern Jiangxi	
11th–13th	central plains	millions	all areas of south	
13th	Fujian/Guangdong/eastern Jiangxi	—	north-eastern Guangdong	
13th	all over China	50,000 soldiers and families	Yunnan	Tai/TB
14th–17th	all over China	1 million	Yunnan/Sichuan	Tai/TB
17th–18th	Fujian (Min and Hakka)	—	Taiwan	Austronesians
17th–18th	northeastern Guangdong	—	Sichuan/Guangxi	Tai/TB/Chinese
17th	Hunan/Hubei	—	Sichuan	
18th–19th	Sichuan/Jiangxi/Hunan	2.5 million	Yunnan/Guizhou	
18th–20th	Hebei/Shandong	tens of millions	north-east China	Altaic
20th	all areas of China	millions	Inner Mongolia	Altaic
20th	lower Yangtse/Shandong	1.4 million	Taiwan	Southern Min/Hakka

The movements were often so massive that they caused major shifts in the overall demographics and language distribution of the entire country. For example, in the seventeenth century, north-east China, south-west China, and the upper Yangtze comprised only about five per cent of the population of China and ten per cent of the Mandarin speaking population, but the movement of people from the middle Yangtze and north China was so massive that by 1982 these three areas included one third of China's population and about half of the Mandarin speaking population (Lee and Wong 1991: 55). In some areas the movements have meant almost an entire displacement of the original population. For example, since 1949 there has been massive Government-orchestrated movement of Han Chinese people into the minority areas of Inner Mongolia, Xinjiang, and Tibet. In Inner Mongolia the population is now less than twenty per cent Mongolian, and the capital, Huerhot, is less than two per cent Mongolian. This of course had a drastic effect on the use of Mongolian in the capital.

Aside from migrations of Chinese into other parts of China (or what later became part of China), there was also quite a bit of influence from non-Chinese people moving into areas of China, particularly north China, where for more than half of the last thousand years the Chinese were under the control of Altaic invaders. Beijing, for example (see Lin Tao 1991), was a secondary capital of the Liao dynasty (Khitan people; 907–1125) and the early Jin dynasty (Jurchen; 1115–1234), and was capital of the Jin from 1153 to 1234. Beijing was again the capital of the Yuan (Mongol; 1234–1368), Ming (Han; 1368–1644), and Qing (Manchu; 1644–1911) dynasties. Except for three hundred years during the Ming dynasty, Beijing was a political centre of non-Chinese peoples for the last thousand years. The populations changed, though, as the Jin government almost emptied the city in 1123, moving the people to the north-east. In 1368, the Ming government moved large numbers of people mainly from Shanxi and Shandong into Beijing to populate the city. In 1644, the Manchu rulers moved most of the original inhabitants out of the inner city and moved the Eight Banner army and their family members into the inner city. While many of the invaders assimilated, they also had an effect on the language of the north. Mantaro Hashimoto (e.g. 1976, 1980, 1986) has talked about this as 'the Altaicization of Northern Chinese', and has argued that a continuum of features from north to south, such as the northern dialects having fewer tones, less complex classifier systems, and an inclusive/exclusive distinction in the first plural pronoun, while the southern dialects have more tones, more complex classifier systems, and other features similar to the Tai and Hmong-Mien languages (You 1982, 1995, Zhou and You 1986, Wang Jun 1991), is due to Altaic influence in the north, and Tai/Hmong-Mien influence in the south. He also suggests (Hashimoto 1976, 1992: 18) that the preservation of final *-n* and *-ŋ* in Mandarin while all the stop endings and *-m* were lost might be due to the fact that these two finals are found in Manchu. Li Wen-Chao (1995) argues that the inventory of vowels and the syllable structure of Chinese changed after the Tang period due to the Altaicization of the

language, that is, the adoption of the Chinese lexicon and grammar by Altaic speakers, but with Altaic phonology.

The resulting mixtures of the people from these migrations with the people originally in the areas they moved into are what give us the dialects we have today (cf. Zhou and You 1986, Wang Jun 1991). For example, the early Wu dialect had formed from a south-eastern migration into an Austroasiatic area,⁴ and the Chu dialect (a precursor to the Xiang dialect) formed from a very early southern migration into a Tai/Hmong-Mien area,⁵ and then the Gan dialect formed in the area where the Wu and Chu dialects had contact with each other in central and northern Jiangxi because of a later migration during the Han dynasty (206 BC–AD 220). Later migrations brought successive waves of immigrants into the area from the north, and then there was a split of this dialect into the Gan and Hakka dialects by migration of what became the Hakka to the east and south, and then later to the west. Contact with languages in each area where the Hakka migrated to resulted in varieties of Hakka that reflect features of those languages (see Hashimoto 1992). In Fujian (Bielenstein 1959, Norman 1991) the language was that of the Min-Yue (a subgroup of the Bai Yue) before any Chinese came into the area, and then the first Chinese settlers in the Eastern Han Dynasty (25–220) brought with them the older dialect of the Wu area, as colonization was from Zhejiang in the north. The original Wu dialect in Zhejiang changed quite a bit after that period due to the massive immigration from the north after the fall of the Western Jin Dynasty in the fourth century. Many of these latter Wu speakers again migrated south into Fujian, and so now the Fujian (Min) dialect shows evidence of influence from at least the following languages: the Min-Yue language, the Chinese language of the Han period, a post-Han stratum brought in by later immigrants, a Tang dynasty (post-eighth century) literary form of the Tang koine, and Modern Mandarin (Norman 1988, 1991). Lien (1987; discussed in W. S.-Y. Wang 1991b, ch. 4) has discussed the complicated interactions of these various strata, and has shown how these interactions led to an ongoing gradual bidirectional diffusion of features (of tones and segments) among the different strata, which has been creating forms that are not identifiable as originating from one particular source language, such as the word for ‘thank’ in the Chaozhou dialect, which has a segmental form, [sia], which derives from the Tang dynasty literary layer, but a tone that the form would have in the colloquial layer. There are also cases of different combinations, such as colloquial initial with literary final and tone, and literary initial with colloquial final and tone (see also Lien 1993, 1997,

⁴ See Zhao and Lee (1989) for genetic evidence that ‘the modern Chinese nation originated from two distinct populations, one originating in the Yellow River valley and the other originating in the Yangtze River valley during early Neolithic times (3,000–7,000 years ago)’ (p. 101), and Mountain *et al.* (1992), Du *et al.* (1992) on the correspondences among surname distribution, genetic diversity, and linguistic diversity in China.

⁵ For linguistic evidence that Chu was a Tai/Hmong-Mien area, see Li Jingzhong (1994). See also Tian (1989) on the ethnic diversity of Chu and the affiliations of the different peoples.

Wang and Lien 1993). The initial discovery of this phenomenon led to the development of the theory of lexical diffusion (see, for example, Chen and Wang 1975), of which Lien's work is an extension. An important point to note is that while the initial strata were the result of language contact (massive borrowing of literary forms or substrate/superstrate influence), the gradual bidirectional diffusion of features has been occurring over a long period of time and is a language-internal phenomenon (though one which of course may be influenced by other factors, such as new superstrate influence).

While in Chaozhou there was a mixing of pre-existing phonemes to create new morphemic forms, there are also cases of the creation of new phones or phonemes because of contact influence, such as in the creation of voiced aspirates for morphemes in a particular tone category in the Yongxing form of the Xiang dialect spoken in Sichuan. Ho (1988; also discussed in W. S.-Y. Wang 1991b) suggests that these voiced aspirates arose because of contact between this dialect and the surrounding Mandarin dialects. In these Mandarin dialects, words that formerly had voiced initials and were in the level-tone category became voiceless aspirates, while in the Xiang dialect in general they continued to be voiced. In Yongxiang, due to the competing influences of the Mandarin feature of aspiration and the Xiang feature of voicing, about 80% of the initials of morphemes in that tone class have become voiced aspirates, a new type of initial for that language.

Compare these phenomena with Dixon's (1997) discussion of the gradual diffusion of linguistic features in a linguistic area. This same sort of bidirectional diffusion among different languages of a bilingual population (rather than strata within a single language) can lead to the areal similarities associated with a linguistic area. Chen Baoya (1996) is a careful study of the bidirectional diffusion of features between Chinese and Tai in Dehong Prefecture of Yunnan Province in China. Chen has shown that in some cases there has been simplification of the sound system of a native language due to the influence of the contact language, e.g. the loss of the distinctions between /l/ and /n/ and between /ts/ and /tʂ/ in the Chinese spoken by ethnic Chinese, as these distinctions do not exist in Tai, and the loss of certain vowel distinctions in the Tai of ethnic Tai (e.g. between /u/ and /y/) because these sounds are not distinguished in Chinese. In other cases there has been an increase in phonemes due to the influence of loanwords in the language, e.g. the development of /kh, tsh, tʂh/ in the Tai of Luxi county. Chen argues that much of the influence is through an interlanguage he calls 'Tai-Chinese', so in a sense there is a tridirectional diffusion in this context.

In Table 1 it is stated that many of the early movements were into areas inhabited by the Bai Yue (Hundred Yue). From the linguistic evidence, it seems there were at least two subgroups of the Hundred Yue, one which spoke Austroasiatic-related languages (mostly along the coast from possibly as far north as Shandong), and another that spoke Tai and Hmong-Mien-related languages (mostly the interior of the south up to the Yangtze and as far west as present-day Sichuan province) (Pulleyblank 1983, Li Jingzhong 1994, Bellwood 1992, Tong 1998).

Norman and Mei (1976; see also Norman 1988) give words for 'die', 'dog', 'child', and others that seem to be cognate with words in Austroasiatic rather than Sino-Tibetan. Yue-Hashimoto (1967, 1991) and others (e.g. Baron 1973, You 1982, 1995, Zhou and You 1986, Huang 1990, Cao 1997, and Meng 1998) give evidence of contact influence between Cantonese and the Tai and Hmong-Mien languages, including not only lexical evidence, but structural evidence, such as word order, the specifics of the tone system, marked phonetic patterns, special uses of the classifiers, etc. In the prehistoric period, the Hundred Yue may have included speakers of the precursors of Austroasiatic, Tai, Hmong-Mien, and possibly Austronesian (see, for example, Blust 1984–5, 1994).

There has also been influence from national and provincial prestige dialects on other dialects throughout Chinese history. Centres of population concentration developed, and languages in those centres came to be quite distinct from each other, with each having prestige within its own area, and then spread out from those centres. The result is languages forming something like prototype categories rather than areas with sharp boundaries (see, for example, Iwata 1995). For example, comparing Guangzhou city Yue with Xiamen city Southern Min (each the prototype of its category), the differences are quite clear, and the languages are easily distinguishable, but in the areas of Guangdong where the two languages meet, there are many forms of each dialect that to different degrees differ from the prototype of their category while having characteristics of the other category. In some cases it is difficult to distinguish whether a certain form of speech is a Yue dialect or a Southern Min dialect, as the two have leached into each other to form something that cannot be uncontroversially put into either category. These major centres have also influenced each other in various ways. See for example Yue-Hashimoto (1993) on the spread of certain patterns of interrogative syntax and other constructions among the Yue, Min, and Beijing dialects, Chappell, this volume, on the creation of 'syntactic hybrids' in the southern dialects due to the influence of Mandarin, and Chang Kuang-yu (1994) on the spread of features of the Wu dialect.

In Modern times there has been quite a bit of influence on the dialects from the Common Language (Mandarin).⁶ There has been a strong effort to teach the Common Language, and this has been very successful in some areas, with the result often being influence on the local dialect. For example, children in Shanghai often speak Mandarin amongst themselves, as that is what they speak in school, even if they speak Shanghainese with their parents. This has caused some changes within Shanghainese, such as the levelling of vocabulary and phonology in terms of becoming more like Mandarin (see, for example, Qian 1991, 1997). In Taiwan, many young people of Taiwanese descent do not learn Taiwanese well (if at all),

⁶ The Common Language (Putonghua) is a dialect created in the early twentieth century by a group of linguists to be the national language of China. It takes the phonology of the Beijing dialect as the basis of its phonology, but the lexicon and grammar represent a more generalized levelling of northern dialects.

and even when they speak it, it is often a somewhat levelled form, where, for example, a Mandarin-based compound word will be pronounced in Taiwanese rather than using the traditional Taiwanese form (e.g. instead of [sin³³ ku⁵⁵] for 'body', you often hear [sin³³ te⁵³], based on Mandarin *shēntǐ*). There is also loss of distinctions in some semantic areas, such as the differentiation of verbs used for the sounds animals make.

In areas where Mandarin is a well-established second language, regional varieties are forming, such as the many varieties of Mandarin developing in the north-west of China because of influence from various Altaic, Turkic, or Tibeto-Burman languages (e.g. Dwyer 1992, Chen 1982). Another interesting example is Taiwanese Mandarin, which can be said to have creolized to some extent out of an interlanguage. After 1949, there was a large influx of people from the mainland because of the Communist takeover of the mainland. These people were mostly from Wu dialect areas, and spoke Mandarin only imperfectly as a second language. The Wu speakers attempted to teach the Taiwanese population Mandarin, and forced the Taiwanese to speak it even amongst themselves. The Taiwanese did not generally have access to native speakers, and so did the best they could with what they had, and often added pieces from their native language, Japanese, and English, forming an interlanguage heavily influenced by Taiwanese (see Kubler 1985, Hansell 1989 for examples). For the Taiwanese this remained a second language, but for the sons and daughters of the mainlanders, who generally did not learn their parents' dialects, and did not learn Taiwanese, this interlanguage became their first language. This group then became the first generation of native Taiwanese Mandarin speakers. There may eventually be a coalescence of the Taiwanized Mandarin and the Mandarinized Taiwanese.

2.2. TIBETO-BURMAN

Turning to Tibeto-Burman, the major migrations were west into Tibet and south-west into Burma, but there were also minor movements into northern Thailand, Laos, and Vietnam. Two large subgroupings formed by areal contact can be distinguished within Tibeto-Burman: the 'Sinosphere' and the 'Indosphere' (these terms from Matisoff, e.g. 1990). One reason for the differences between the two spheres is the objective dominance of Chinese or Indic languages over different subsets of the Tibeto-Burman languages; another is the subjective analysis of those languages falling within the scope of work by Chinese-trained or Indic-trained linguists. There are certain features that we frequently find in languages in the Indosphere that we do not find in the Sinosphere. In phonology we find, for example, the development of retroflex stop consonants. In syntax we find, for example, post-head relatives of the Indic type (relative clauses are generally pre-head and without relative pronouns in Sino-Tibetan languages). For example, in Chaudangsi (Shree Krishan, 2001a: 412), of the Pithoragarh District of Uttar Pradesh, India, a relative clause is formed using

one of two borrowed (Indo-Aryan) relative pronouns, /jo/ (with human subjects) or /jəi/ (with non-human subjects).

- (1) (a) hidi əti siri hle jo nyarə ra-s
 this that boy is who yesterday come-PAST
 'He is the same boy who came yesterday.'
 (b) hidi əti hrəŋ hle jəi be ər ɡwən-cə
 this that horse is which mountain from fall-PAST
 'It is the same horse which fell from the mountain.'

Another feature of the Indosphere, discussed by Saxena (1988a, b), is the frequent grammaticalization of a verb meaning 'say' into a quotative, causal, purpose, or conditional marker, a complementizer, or an evidential particle, due to areal influence from the Indic and Dravidian languages, whereas languages in the Sinosphere are less likely to do this.⁷ In (2) are examples of the use of the verb for 'say' as part of a causal connective in Nepali, an Indo-European language, and Newari, a Tibeto-Burman language of Nepal.

- (2) (a) Nepali (Saxena 1988a: 376):
 timiharumadh-e ek jana murkh ho kinabhane yo dhoro-horo hoina
 you(pl) among-LOC one CL fool is
 why+say+PART this tower-be+NEG
 'One of you is a fool because this is not a tower.'
 (b) Newari (Saxena 1988a: 379):
 chi-pi cho-mho murkho kho chae-dha-e-satho dhoro-hora mo-khu
 you-pl one-CL fool are why-say-INF-if this tower NEG-is
 'One of you is a fool because this is not a tower.'

In Sino-spheric languages we often find the development of tones. For example, among the Qiang dialects of north-western Sichuan, there is a north-west to south-east cline in the degree to which tones are a stable and important part of the phonological system: the closer the dialect is to the Chinese areas, generally the stronger and more developed the tone system is (Sun 1981, Liu 1998, Evans 1999).^{8,9} Contact with Chinese can also result in monosyllabicity and an isolating

⁷ There is a complementizer derived from a verb meaning 'say' in the Southern Min dialect of Chinese (and now also in Taiwanese Mandarin), and this has been discussed as a South-East Asian areal feature (Matisoff 1991, Chappell, this volume), but in the Sinosphere it is usually simply a complementizer, and does not usually develop into a cause or purpose-marking connective, as in (2).

⁸ There are actually two different types of situation related to the development of tone systems in Asia. There is the simple contact type, as in Qiang, where the tones have not developed out of (or been influenced by) segmental features in the language, and there is the type where the language is in contact with a tone language (e.g. Vietnamese in contact with Chinese), but the development of tones is based on the loss of voicing of initials, the loss of final consonants, etc. The development of tones in some Tibetan dialects is of the latter type (see, for example, LaPolla 1989).

⁹ There is also Tibetan influence on the Qiang from the north and west (see e.g. Liu 1981, Lin Xiangrong 1990), to the extent that speakers of the north-western dialects have come to see themselves as Tibetans rather than Qiang (though still use the same appellation for themselves when speaking Qiang).

structure, the most extreme example of this being Vietnamese. It also seems to be the case that languages in the Sinosphere have simpler grammatical systems, but this brings us to the second part of this question of spheres: the subjective analysis of the linguists doing the recording of the languages. In India, linguists are trained in Sanskrit grammar, and so are familiar with paradigms and participles. They generally look for them in the Tibeto-Burman languages they describe, and often find them. They are not very familiar with tones, and do not consider them that important, and so even if the language has tones, they often will not be included in the description. On the other hand, the Chinese linguists are trained in Chinese linguistics, and so are often not familiar with paradigms and participles, but are very familiar with tones. They then generally do not describe the languages as having tight paradigms, etc., but very often find and describe Chinese-like tonal systems, even in languages (e.g. Burmese, rGyalrong) that could be argued to have register or pitch-accent systems.

As mentioned above, the Tibeto-Burman speakers followed two main lines of migration: west into Tibet and then down into Nepal, Bhutan, and northern India; and south-west down the river valleys along the eastern edge of the Tibetan plateau through what has been called the 'ethnic corridor' (Fei 1980, Sun 1983, Hoffman 1990). This split in the migration is responsible for the split between the Bodic languages and the rest of Tibeto-Burman. There is little information about the spread of Tibeto-Burman speakers into the Tibetan plateau other than that they spread from the north-east of Tibet (that is, the north-west of China; Stein 1961, Snellgrove and Richardson 1986, Ran and Zhou 1983, Hoffman 1990), but from the present wide geographic spread of the Tibetan dialects, from the closeness of the dialects, and from the fact that all dialects show some of the same uncharacteristically Tibeto-Burman features (such as non-Tibeto-Burman words for 'horse' and 'seven'), there must have been contact with non-Tibeto-Burman languages before the spread of Tibetan,¹⁰ and then the spread was relatively rapid, and into an area where there were no (or few) earlier inhabitants. There has also been quite a bit of contact with northern and central Asian languages since the original spread of the Tibetan dialects as well (see for example Laufer 1916).

There is also not much we can be sure of about the early history of Burma.¹¹ It is assumed that the original inhabitants were negritos. The migration of Tibeto-Burman speakers south into Burma must have started by at least the first century AD. Fourth-century Chinese records already talk of a barbaric tribe we might identify with the modern Jinghpaw in the far north of Burma and a civilized kingdom known as Pyu which controlled central and upper Burma. The Pyu were Tibeto-Burman speakers who had come down into Burma along the Irrawaddy valley.

¹⁰ See Hoffman (1990: ch. 4) on the prehistoric contact influence from the 'steppe peoples' (northern non-Sino-Tibetans) on the group that became the Tibetans.

¹¹ This section is a synthesis of information in Luce (1937, 1976), Luce and Pe Maung Tin (1939), Leach (1954), Hall (1981), FitzGerald (1972), Chen Xujing (1992), and Chen Ruixing (1995).

They adopted Theravadan Buddhism and their writing system (seventh century) from the Mon (Mon-Khmer), who controlled lower Burma and the Menam Chao Phya valley (now part of Thailand). The Chin (Zo), another Tibeto-Burman group, also came down into Burma some time before the ninth century and established a kingdom in the Chindwin valley.

In the eighth and ninth century, a kingdom called Nanzhao (Nan-Chao), in what is now western Yunnan Province in south-west China, came to dominate upper and most of lower Burma. Nanzhao was ruled by Yi (Lolo; Tibeto-Burman) speakers, but the people also included Bai (Tibeto-Burman), other Lolo-Burmese, and Tai speakers (Shiratori 1950, Backus 1981, Lin Zongcheng 1986). Some time before the eighth century, the migration of the Karen, another Tibeto-Burman group, down into Burma weakened the Pyu kingdom, and in 832 Nanzhao destroyed the Pyu kingdom. There are few mentions of the Pyu in records after 863. The Pyu and their language were simply absorbed into the succeeding political entities, with obvious effects on the culture of those entities.

The people we have come to think of as the Burmese had been in Yunnan, under the control of the Nanzhao kingdom, and moved down into Burma from the middle of the ninth century. They came down from the northern Shan states into the Kyaukse area south of Mandalay, splitting the Mon in the north and south, and pushed the Karens east of the Irrawaddy. About AD 1000 the Burmese conquered the Mon to the south, and the first Burmese kingdom, the Pagan kingdom, was founded in 1044. The court adopted much of Mon culture (it became the official court culture, and the Mon language (or Pali) was used for inscriptions; the Mon script also became the basis of the Burmese writing system). This was the early period of major contact and influence of the Mon on the Burmese, which lasted until the late twelfth century. Indian influence on the Burmese was mainly indirect through the Mon, or from Ceylon.

After the Nanzhao kingdom was conquered by the Chinese in the ninth century, the Dali (Tali) kingdom, which was ruled by a Tibeto-Burman people related to the modern Bai nationality, was established in the same area. This was then taken over by the Mongols in the thirteenth century. The Mongols then conquered northern Burma in 1283, but did not hold on to the territory. This gave the Shan, a Tai-speaking people who had been pushed by the Mongol invasion of Yunnan into the area between the Salween and the Irrawaddy, the chance to take over the upper and central areas of Burma. They covered both banks of the Irrawaddy and pushed the Chins out of the Chindwin valley into the hills to the west. Within about ten years the Shan controlled all of upper and central Burma. The Shan rulers adopted Burmese language and culture, and claimed to be descendants of the Pagan kings. Apart from this Shan state, there were several other Shan states in the north, and there was constant fighting among them. This fighting forced many Burmese south to Toungoo and Pagan, and this caused the Toungoo kingdom to become the more powerful state, and it eventually reconquered the Mon, who had become independent again after the Mongol invasion,

as well as the Arakanese (1784) and the rest of Burma. The Mon then became much more a part of Burma, and this began another period of Mon influence on the Burmese. Much of what we think of as Burma, such as the Irrawaddy Delta and Rangoon, was for most of its history part of a Mon kingdom.

Because of this legacy, there has been heavy influence of the Mon language on Burmese (Bradley 1980). Aside from the script and a large number of lexical loans, there has also been Mon influence on the suprasegmentals, in that Burmese 'tones' are unlike the usual Sino-Tibetan type of tones in being more like a register contrast (and in the Arakanese dialect of Burmese show vowel-height differences related to the registers), as is the case in Mon and other Mon-Khmer languages. There has been convergence in the vowel systems of Mon and Burmese, and to some extent the consonant system, where there has been a loss of contrast between alveolar fricatives and affricates versus palatal or alveopalatal fricatives and affricates, as in Mon. In Written Burmese there are also palatal finals (most finals have been lost from the spoken language), which do not usually occur in Tibeto-Burman languages, but do occur in Mon-Khmer languages. In terms of the word structure, Burmese has the typical sesquisyllabic structure of Mon-Khmer languages where the first 'half-syllable' or 'minor syllable' is unstressed and the second syllable is stressed (e.g. the Burmese pronunciation of the word Burma: [bə'ma]). This is a feature that characterizes a number of the languages in the area, as opposed to the Tibeto-Burman languages still in China, which generally do not show this pattern. Bradley (1980) attributes these influences to the fact that so many Burmese speakers were originally Mon speakers. Many of them are now monolingual in Burmese. In fact Burmese is spoken by many different ethnic groups, and so shows varieties in each area due to the influences of the local languages (Bradley (1996) has produced a map (with discussion) of the use of Burmese by different ethnic groups).

Another language which has had a major influence on Burmese is written Pali. Many Burmese texts are what is known as 'Nissaya Burmese', word for word translations of Pali texts which try to accommodate Burmese word order and grammar to the Pali original, and this led to influence on purely Burmese texts. 'Pali was regarded as the model of correctness in language, so that the closer to Pali one's Burmese was, the purer it seemed to be' (Okell 1965: 188). This written form eventually influenced the spoken form as well because of the influence of reading, education, and religion (Okell 1965).

The north of Burma continued to be populated by the Shan and the Tibeto-Burmans (principally Jinghpaw), and there has been much mutual cultural and linguistic influence, in some cases with subgroups of the Jinghpaw becoming Shan in language and culture and vice versa (Leach 1954: 293). In the eighteenth and nineteenth centuries these two groups extended into Assam, and the Jinghpaw brought thousands of Assamese slaves back into Burma. These formerly Indo-European-speaking slaves eventually assimilated to the Jinghpaw culture and language (Leach 1954: 294).

We see another type of contact situation in northern Burman, that is where two or more languages are in close contact, but no language is dominant, such as is the case with the Jinghpaw people (Dai 1987), which is similar to the situation that Dimmendaal (this volume) describes for the Suri group of Surmic. There are four subgroups within the Jinghpaw nationality, and each subgroup has its own language. These four groups often live together in the same villages and intermarry, and have very similar cultures, but keep the languages distinct in terms of exogamy, marrying other-language speakers, the children being considered speakers of the father's language even though they may speak one language to the father, one to the mother, and a third to the grandmother. Living in such a situation the people come to think in similar patterns and have similar cultures, and this leads to certain types of lexical and usage convergences among the languages. This is a clear case of adstratum influence. In the case of other Tibeto-Burman languages, contact has been not because they live within the same villages, but live relatively close to each other, and so become bilingual, and this can affect the languages. For example, in Lisu dialects in general, interrogatives are marked by a sentence-final particle, while in Yi dialects interrogatives are marked by reduplication of the verb. But in the Luquan dialect, the Lisu dialect closest to the Yi-speaking area, interrogatives can be, or always are, marked by reduplication of the verb (CASIML 1959: 3).

Tibeto-Burman migration into Nepal, Sikkim, and Bhutan was originally almost entirely from directly north, that is, Tibet (Poffenberger 1980), and so the earlier languages generally show a close relation to Tibetan. In Nepal (see Kansakar 1996) there are now over seventy different languages, possibly as many as a hundred (Grimes 1991). According to Kansakar (1996: 17), these include about fifty-six Tibeto-Burman languages, fourteen Indo-Aryan languages, one Austroasiatic language, one Dravidian language, and one isolate (Kusunda). Of the Tibeto-Burman languages, the Kiranti languages and what Bradley (1997) calls the Central Himalayan languages (Magar, Kham, Chepang, Newari) came into Nepal relatively early, and the Newars (now 3.7% of the population) had a kingdom in the Kathmandu Valley from at least the eleventh century until they were conquered by the Nepali-speaking Gurkhas in the eighteenth century. A large group of Tibetans moved into Nepal during the reign of the Tibetan leader Strong-Bstan-Sgampo in the seventh century and after, when the whole area down to the Bay of Bengal was part of the Tibetan kingdom; the Tamangs are said to be remnants of these people. Quite a few members of the Tamang-Gurung group have in the last one or two hundred years emigrated to north-east India or other areas (e.g. eastwards into Nepal) and now speak only Nepali. Among the Gurungs there is something of a cultural continuum of Buddhists in the north and Hindus in the south due to contact with Hindus in the south (Poffenberger 1980). Of those Gurungs still living in Nepal, 49.2% (221,271) no longer speak the Gurung language (Kansakar 1996: 23). The Sherpas came little by little into the eastern part of the country (Solu-Khumbu) from the Khams region of Tibet (the

eastern part of Tibet) starting in the sixteenth century (Oppitz 1974, cited in Nishi 1986).¹² There was also a relatively large migration of Tibetans from Krong to Langtang in the 1790s. These Tibeto-Burman speakers live mostly in the northern hills of the country, while the lowlands are now inhabited by Hindu Indo-Aryan speakers, many of whom migrated there between the eleventh and thirteenth centuries. A large number of Central Tibetan (Lhasa) speakers have come to Nepal and India since the failed 1959 uprising against Chinese rule in Tibet.

Nepali, an Indo-Aryan language, is the official language of Nepal, and so is used for official purposes and in education, law, and the media. Fifty per cent of the population are said to be native speakers of Nepali (Kansakar 1996). While all indigenous languages are recognized as national languages by the 1990 constitution, aside from Nepali, only two other languages (Maithili, Indo-Aryan; Newar, Tibeto-Burman) are offered in school (as electives) beyond the primary level. Nepali is clearly the dominant language, and 'non-Nepali speakers have been at a disadvantage in education, employment and other social benefits' (Kansakar 1996: 18). There is then great pressure to learn Nepali and this has caused an increase in bilingualism and language shift. Most of the people of the country now are bilingual in Nepali, and many languages show influence from Nepali, particularly the development of a dative/human patient ('anti-ergative', LaPolla 1992b) marker [lai], and in some cases convergence of grammatical categories and use, such as the convergence of the tense and ergative marking systems in Nepali and Newari (see Bendix 1974). Some of these convergences may be assisted somewhat by what Jakobson (1938) called 'linguistic affinity'; for example, the Tibetan dialects already had a locative marker [la] that could be used for dative and human patient marking. Quite a few of the languages, in fact almost all of the Kiranti (Rai) languages are endangered. In Bhutan, where there were in the past only Southern Tibetan (west) or Monpa (east) speakers, there are now a large number of Nepali speakers (though in recent years many have been expelled from Bhutan because of conflicts with the Bhutanese).

Quite a large number of Tibeto-Burman languages are found in the north-western and north-eastern parts of India and in Bangladesh, mainly languages that came from Burma in the east, but also some from Tibet in the north. They have been greatly affected by the cultures they have come into contact with. To give a few examples, in Kashmir two varieties of Tibetan have developed: Balti and Ladakhi. Balti is spoken in the (Pakistan-controlled) Moslem Baltistan area of northern Kashmir. The speakers of Balti are now also Moslems and write their language, which is a Western Tibetan dialect, with the Arabic script. Ladakhi is in the Indian-controlled area of Kashmir, and the speakers are still more culturally

¹² A competing theory, also mentioned by Nishi (1986), citing Qu (1985), is that the Sherpas migrated in the early to mid-thirteenth century. Qu (1985) also says that a part of this Sherpa population in Nepal moved back to Tibet (Shigatse in Central Tibet) about three hundred years ago, and that the speech of these migrants, due to influence from the surrounding central dialects, is now classified as a central dialect (rather than an eastern dialect), but still retains elements of the tone system of the eastern dialects.

Tibetan. In northern Himachal Pradesh and Uttar Pradesh, the speakers of Tibeto-Burman languages have all converted to Hinduism (Singh 1986, Tiwari 1986), and their languages are quite heavily influenced by the surrounding Indo-Aryan languages, such as in having non-native retroflex consonants and post-head relative clauses of the Indic type (see example (1) above). The borrowing of reflexive pronouns, aspect marking, postpositions, conjunctions, and certain other syntactic constructions is also common. Some languages, such as Raji (Jangali), a language of north-eastern Uttar Pradesh (ShreeKrishan, 2001b), are so mixed with features that it is hard to determine if it is a Tibeto-Burman language heavily influenced by Indo-Aryan and Munda, or a Munda language heavily influenced by Tibeto-Burman and Indo-Aryan. This led to it being classified in Grierson (1909, vol. 3, part 1: 177) as Tibeto-Burman, but in Sharma (1989) as Munda. As in Nepal, there have also been movements of non-Tibeto-Burmans into Tibeto-Burman areas. For example, the hill area of Darjeeling district of West Bengal was, before the twentieth century, inhabited mainly by speakers of Tibeto-Burman languages (Lepcha and Tibetan-related languages), but now, due to an influx of immigrants, the population is 80% Nepalese (Chaudhuri 1986).

In Manipur, there have been Meithei speakers for at least a thousand years, having moved there from Burma (Grierson 1909, vol. 3, part 3: 2). Meithei is written with a Bengali-based Indic orthography, and is heavily influenced by Indo-Aryan contact (see Chelliah 1997). Aside from being spoken by about one million Meitheis, it has become a lingua franca for many other ethnic groups in Manipur, and this has affected the form that it takes in each area where it is spoken, much as we saw for Mandarin and Burmese.

These languages that have some currency as a lingua franca or status language in an area (e.g. Meithei, Burmese, Tibetan, Mizo, Lahu, Jinghpaw, Mandarin) all show a sort of bidirectional influence: they are influenced by the native languages of the people who speak them, as we saw above, but at the same time they influence the native languages of those speakers. For example, as the dominant language of Burma, Burmese has had a major impact on many of the minority languages in the country (Bernot 1975); Stern (1962) discusses the influence of Arakanese Burmese on the lexicon and phonology of Plains Chin. In many cases there are wholesale shifts in language and culture to that of the Burmans (see Stern 1962 on the Chin).

There are a number of Tibeto-Burman speakers in northern Thailand, such as the Akha, Lahu, Gong, Mpi, and Karen. Aside from the Karen, most have moved down from China within the past few hundred years. For example, the Lahu migrated from Yunnan into Burma in the eighteenth and nineteenth centuries, and into Thailand and Laos only very recently (Matisoff 1986). Northern Thailand was originally populated by Tai speakers, and the recent arrivals (the Tibeto-Burman speakers) are now largely bilingual in Thai and their own languages, and their languages show quite a bit of Thai influence and even language shift (see for example the many Thai loanwords in Lahu given in Matisoff (1989) and the

discussion of areal features shared by the different languages in northern Thailand in Matisoff (1986), and see Bradley (1998a) on the language change and shift in progress of the Gong language; see Bradley (1986, 1998b) on the factors involved in the persistence or non-persistence of minority languages in Thailand).

3. Metatypy

I have argued elsewhere (LaPolla 2003) that language is not something separate from culture or cognition. How we represent some state of affairs represents how we conceive of that state of affairs, and how we conceive of it is related to cultural norms and experiences. When people learn some aspect of another language, if the influence of the culture associated with that language is not great, the borrowers will assimilate the borrowed form to their way of thinking. An example of this might be the distinction of animate and inanimate in relative pronouns in Chaudangsi, even though that distinction was not part of the borrowed structure. If there is heavy enough cultural contact, the contact may slowly change the way the borrowers conceptualize certain events, such that they develop what Bhattacharya (1974) has called 'new agreements in their outlook of life', thereby creating 'a common cultural core'; what Ross (this volume) gives as the reason for metatypy: speakers 'increasingly come to construe the world around them in the same way' as some other group. This common cultural core or construal of the world can then lead to the spread of certain constructions or linguistic patterns. For example, in the Wutun language (Chen Naixiong 1982), which is a heavily Tibetanized form of Chinese in Qinghai, rather than using two words for 'widow' and 'widower', as is standard in Chinese, the speakers of Wutun have come to agree with the Tibetans in not differentiating widows and widowers linguistically, and so use the Chinese form for 'widow' for both. The development of an inclusive/exclusive distinction in the first person plural pronoun in Northern Mandarin due to Altaic influence is another example, as making this distinction means having a clear cognitive category distinction that would lead to the use of different forms. This is true also of the example Ross (this volume) gives of the development of the formal distinction between alienable and inalienable possession in Proto-Oceanic because of Papuan contact.

When people are used to using a particular linguistic category in a language they use regularly, they will try to use it in any language they speak. In other words, if some category or lexical item they are used to using is not in one of the languages they are using, there is a perceived gap. Many Cantonese speakers in Hong Kong, when they speak English, will frequently use *then* (generally said with a rising tone) at the beginning of discourse segments or speech turns. They do this because there is a particle in Cantonese, [kəm³⁵], used in this way, and they feel the need for something with that function when they speak English. Substratum influence, such as the development of the aspect and complementizer patterns that have developed in Taiwanese Mandarin on the model of the Taiwanese dialect (Chappell, this volume) are of this nature. Heine (1994, see also 1997a, b, Heine

and Kuteva, this volume) has talked about the importance of event schemas for determining the type of grammaticalization you find in a language. These event schemas are ways of conceptualizing states of affairs. An example Heine discusses is comparatives. How speakers view a comparative situation, whether as a locational schema, an action schema, or whatever, will determine what sort of structure they use to express that situation. This way of thinking can change through contact with another culture, and lead to the development of what are commonly called calques, but are better seen as examples of metatypy. Matisoff (1991) discusses several types of grammaticalization common to the languages of South-East Asia that are based on particular types of schema, such as locative verbs becoming progressives, a verb meaning 'get' becoming an auxiliary meaning 'have to / must, able to' (see also Enfield, this volume), and a verb meaning 'give' becoming a causative or benefactive auxiliary. We can see how similar the ways of thinking and structure can become from the description of Rongpo (Chamoli District of Uttar Pradesh, India), a language that has been very heavily influenced by Hindi and Garhwali (Indo-Aryan), in Sharma (2001). In discussing a particular participial form, Sharma (p. 223–224) first gives the English translation in (3a), but says: 'In fact this translation is not very close in its meaning. The Hindi sentence is more appropriate', and then gives the sentence in (3b).

- (3) (a) *dī phəl gyi-tə jəpɪŋ yā*
 this fruit I-DAT eaten is
 'This fruit was eaten by me.'
- (b) *yəh phəl mɛrəː khaːyɑː huɑː hai*
 this fruit I + POSS eaten be + past is
 giving the sense—'I have the experience of eating this fruit in the past.'

One phenomenon in Tibeto-Burman that I think is a case of contact-induced metatypy is the parallel development of person-marking in a large number of Tibeto-Burman languages. The languages with person-marking are almost all spoken around the edge of the Tibetan plateau from north-west China down along the southern edge of the plateau, in an area of large-scale language contact, multilingualism, and mutual influence. I have given arguments elsewhere (LaPolla 1992a, 1994b) why person-marking should not be considered an archaic feature of Tibeto-Burman. Here I will just cite some examples of the person-marking forms in a number of languages to show how the same pattern of grammaticalization was followed in the different languages (similar to what happened in Australia—see Dixon 1980: 363, this volume).

The earliest example we have of person-marking in Tibeto-Burman is in Tangut, a dead language in which there are texts dating back to the eleventh century. In Tangut the optional verbal suffixes have the same phonetic form, including the tone, as the free pronouns (adapted from Kepping 1975, 1979, 1981, 1982, 1989; there is also a first and second person plural marker *ni*²; third person is not marked): see Table 2.

TABLE 2. Tangut person-markers and free pronouns

	Free pronouns	Verb suffixes
1sg	<i>ŋa</i> ²	- <i>ŋa</i> ²
2sg	<i>na</i> ²	- <i>na</i> ²

TABLE 3. Angami Naga person-markers and free pronouns

	Free pronouns	Verb prefixes	Noun prefixes
1sg	<i>ā</i>	<i>ā</i>	<i>ā-</i>
2sg	<i>nō</i>	<i>n̄-</i>	<i>n̄-</i>
3sg	<i>puō</i>	<i>puō-</i>	<i>puō-</i>

In the Kuki-Chin branch of Tibeto-Burman we find a person-marking system very similar to that in Tangut. In this system we find the Proto-Kuki-Chin pronouns **kai* ‘1sg’, **naŋ* ‘2sg’, and **a-ma* ‘3sg’ grammaticalized into the person-marking prefixes **ka-*, **na-*, and **a-* respectively (Thurgood 1985). Yet from the fact that the system is prefixal, and the fact that the pronouns that were the source of the prefixes are not the same as the Tangut forms (at least the 1sg and 3sg forms), and from the fact that the languages are not closely related, we can say that this system clearly developed independently of the Tangut system.

A middle case is the Kanauri-Almora branch, which has person-marking that is suffixal, like the Tangut system, but has a first person suffix derived from an innovative pronoun somewhat similar to that in Kuki-Chin. The forms are **-ga* (< **gai*) and **-na* (< **naŋ*) (there is no third person agreement suffix) (Thurgood 1985). We can still be confident of the independent origin of this system, though, because the source of the first person affix is different from that of Tangut, and though it may be similar to that of the Kuki-Chin system, it is a suffixal system.

A fourth case of clear independent development is the person-marking system of Angami Naga (Giridhar 1980), which involves prefixes clearly derived from the independent pronouns. The verbal prefixes are also isomorphic (except for the tone on the first person prefix) with the pronominal genitive noun prefixes (22 ff.): see Table 3. Again we see that not only is this a prefixing system, unlike the Tangut system, but it also derives from a set of free pronouns unique to Angami.

A fifth case is the person-marking prefixes of Mikir (Hills Karbi; Jeyapaul 1987). Again we have a prefixing system, but one quite different from those discussed above: see Table 4. That this system is a recent development can be seen not only from the fact that the free pronouns and the prefixes are so similar in form, but also from the fact that the verb prefixes retain the inclusive/exclusive distinction of the free pronouns.

TABLE 4. Mikir (Hills Karbi) person-markers and free pronouns

	Free pronouns	Verb prefixes
1sg	<i>ne</i>	<i>ne-</i>
1pl (exc)	<i>netum</i>	<i>ne-</i>
1pl (inc)	<i>itum ~ etum</i>	<i>i- ~ e-</i>
2sg	<i>naŋ</i>	<i>naŋ-</i>
3sg	<i>alaŋ</i>	<i>a-</i>

TABLE 5. Sgaw Karen person-markers and free pronouns

	Free pronouns	Verb prefixes
1sg	<i>ja³³</i>	<i>jǎ³³-</i>
1pl	<i>pu³³ we⁵⁵ θe³¹</i>	<i>pǔ³³ kǎ³¹-</i>
2sg	<i>na³³</i>	<i>nǎ³³</i>
2pl	<i>θu⁵⁵ we⁵⁵ θe³¹</i>	<i>θǔ⁵⁵ kǎ³¹-</i>

One last example is from the Delugong dialect of Sgaw Karen (Dai et al. 1991: 400); third person is unmarked: see Table 5. This system of verbal prefixes is very clearly of recent origin, being in the singular simply unstressed copies of the free pronouns, and unique to this dialect of Karen.

It is unlikely that so many languages developing person-marking in the same way is a coincidence, even given the fact that they are in most cases typologically similar. There must be some other factor, and I believe that factor is language contact, much as the Vietnamese development of tones in a way parallel to that of Chinese is at least partially due to contact with Chinese.

4. Conclusion

I have tried to show in this chapter that the history of the Sino-Tibetan-speaking peoples is one of frequent migration and contact with other languages and cultures, and each other, and that this contact has been a major influence on the development of the Sino-Tibetan language family. To understand why the languages of the family have the forms they do, and why there are difficulties in assigning a clear family-tree structure to the family, language contact must not only be taken into account, but must be considered a fundamental factor in the formation of the family.

But this then brings up a question. Those who do subgrouping (see note 3) often do not give the reasons for their groupings. In some cases there are clear isoglosses, but often subgrouping is affected by the author's subjective 'feel' of the language, shared features, or shared vocabulary, which are all often

influenced by its geographic location. Bradley (1997) is the most straightforward in this regard, as most of the names for his subgroups are geographic (e.g. 'Central Himalayan'). While some may argue that what is at issue is genetics, not location, there is value in grouping the languages geographically because contact has been so important in the development of the languages. This then brings us to a question raised in Dai (1997). Dai argued that the family tree model alone is not sufficient to account for the facts of Sino-Tibetan; we need to take into account language contact that has led to what he called 'language coalescence'. He asks, 'Is it not possible for two languages that were not originally related to become related through intense contact?' For example, could we not resolve the question of the relationship between Tai-Hmong-Mien and Chinese by saying they were not originally related but now are? If we accept geographic groupings that are most probably the result of areal contact, what does that mean for the concept of 'relatedness'?

References

- BACKUS, C. 1981. *The Nan-chao Kingdom and T'ang China's Southwestern Frontier*. Cambridge: Cambridge University Press.
- BARON, S. P. 1973. 'The classifier-alone-plus-noun construction: a study in areal diffusion'. Paper presented at the 6th International Conference on Sino-Tibetan Languages and Linguistics, San Diego.
- BAXTER, W. H. 1995. 'A stronger affinity . . . than could have been produced by accident': a probabilistic comparison of Old Chinese and Tibeto-Burman', pp. 1-39 of Wang (1995).
- BELLWOOD, P. 1992. 'Southeast Asia before history', pp. 55-136 of *The Cambridge History of Southeast Asia*, edited by N. Tarling. Cambridge: Cambridge University Press.
- BENDIX, E. H. 1974. 'Indo-Aryan and Tibeto-Burman contact as seen through Nepali and Newari verb tenses', *International Journal of Dravidian Linguistics* 3.42-59.
- BENEDICT, P. K. 1972. *Sino-Tibetan: A Conspectus*. Princeton-Cambridge Studies in Chinese Linguistics II. Cambridge: Cambridge University Press.
- BERNOT, D. 1975. 'Influence of Burmese language on some other languages of Burma', *Journal of the Siam Society* 63.96-109.
- BHATTACHARYA, S. 1974. 'Linguistic convergence in the Dravido-Munda culture area', *International Journal of Dravidian Linguistics* 4.199-213.
- BIELENSTEIN, H. 1959. 'The Chinese colonization of Fukien until the end of T'ang', pp. 98-122 of *Studia Serica Bernhard Karlgren Dedicata: Sinological Studies Dedicated to Bernhard Karlgren on his Seventieth Birthday*, edited by Søren Egerod. Copenhagen: Ejnar Munksgaard.
- BLUST, R. A. 1984-5. 'The Austronesian homeland: a linguistic perspective', *Asian Perspectives* 26.45-67.
- 1994. 'Beyond the Austronesian homeland: the Austric hypothesis and its implications for archaeology'. Paper presented at the Institute of History and Philology, Academia Sinica, 28 Feb. 1994.
- BRADLEY, D. 1980. 'Phonological convergence between languages in contact: Mon-Khmer

- structural borrowing in Burmese', *Proceedings of the Sixth Annual Meeting of the Berkeley Linguistics Society* 259–67.
- 1986. 'Identity: the persistence of minority groups', pp. 46–55 of *Highlanders of Thailand*, edited by J. Mckinnon and W. Bhruksasri. Singapore: Oxford University Press.
- 1996. 'Burmese as a lingua franca' (and associated map, no. 87), pp. 745–7 of *Atlas of Languages Used for Intercultural Communication in the Pacific, Asia, and the Americas*, edited by S. A. Wurm, P. Mühlhäusler, and D. T. Tryon, vol. 2: 1. Berlin: Mouton de Gruyter.
- 1997. 'Tibeto-Burman languages and classification', pp. 1–71 of *Tibeto-Burman Languages of the Himalayas* (Papers in Southeast Asian Linguistics 14, Pacific Linguistics Series A-86), edited by D. Bradley. Canberra: Australian National University.
- 1998a. 'Endangered languages and language change: Gong'. Paper presented to the 31st International Conference on Sino-Tibetan Languages and Linguistics, Lund, Oct. 1998.
- 1998b. 'Standardization of transnational minority languages in Asia: Lisu and Lahu'. Paper presented at the conference on Minority Languages in Context: Diversity and Standardization, Chur, Sept. 1998.
- BURLING, R. 1983. 'The *sal* languages', *LTBA* 7.1–32.
- CAO GUANGQU. 1997. 'Zhuang-Dongyu he Hanyu Min, Yue fangyan de gongtong dian' [Commonalities among Zhuang-Dong languages and the Min and Yue dialects of Chinese], *Minzu Yuwen* 2.54–60.
- CASIML (Chinese Academy of Sciences Institute of Minority Linguistics). 1959. *Lisuyu yufa gangyao* [Sketch of Lisu grammar]. Beijing: Kexue Chubanshe.
- CHANG, K. C. 1986. *The Archeology of Ancient China*. 4th edition. New Haven: Yale University Press.
- CHANG KUANG-YU. 1994. 'Wuyu zai lishi shang de kuosan yundong' [Linguistic diffusion of the Wu dialect], pp. 281–300 of *Current Issues in Sino-Tibetan Linguistics*, edited by H. Kitamura, T. Nishida, and Y. Nagano. Osaka: National Museum of Ethnology.
- CHAUDHURI, B. 1986. 'Economy, migration and development in the eastern Himalayas: a study with special reference to the hill areas of Darjeeling W. B.', pp. 72–85 of *Ecology, Economy and Religion of Himalays*, edited by L. P. Vidyarthi and Makhan Jha. New Delhi: Orient Publications.
- CHELLIAH, S. L. 1997. *A Grammar of Meithei*. Berlin and New York: Mouton de Gruyter.
- CHEN BAOKA. 1996. *Yuyan jiechu yu yuyan lianmeng* [Language contact and language coalescence]. Beijing: Yuwen Chubanshe.
- CHEN, M. Y., and WANG, W. S.-Y. 1975. 'Sound change: actuation and implementation', *Language* 51.255–81.
- CHEN NAIXIONG. 1982. 'Wutunhua chu tan' [A preliminary discussion of Wutun], *Minzu Yuwen* 1.10–18.
- CHEN RUXING. 1995. 'Lun Mianzu de xianmin zi Dian xi qianru Miandian de dayue luxian ji niandai' [On the probable path and date of the migration of the ancestors of the Burmese from Western Yunnan into Burma], *Dalu zazhi* 90.54–68.
- CHEN XUJING. 1992. *Chen Xujing Dongnanya gushi yanjiu heji* [Collection of studies on South-East Asia by Chen Xujing]. Hong Kong: Commercial Press.
- DAI QINGXIA. 1987. 'Lun Jingpozu zhixi yuyan' [On the languages of the Jingpo nationality], *Minzu Yanjiu* 3.69–77.
- 1997. 'Guanyu Han-Zangyu fenlei de sikao' [Thoughts on subgrouping in Sino-Tibetan], *Yuyan Jiaoxue yu Yanjiu* 4.4–10.

- DAI QINGXIA, HUANG BUFAN, FU AILAN, RENZENGWANGMU and LIU JUHUANG. 1991. *Zang-Mianyu shiwu zhong* [Fifteen Tibeto-Burman languages]. Beijing: Yanshan Publishing Co.
- LIU JUHUANG, and FU AILAN. 1987. 'Yunnan Mengguzu Gazhuoyu yanjiu' [On the Gazhuo language of the Mongolian people of Yunnan], *Yuyan Yanjiu* 1.151–75.
- ——— 1989. 'Guangyu woguo Zang-Mian yuzu xishu fenlei wenti' [On the problem of genetic subgrouping within the Tibeto-Burman languages of China], *Yunnan Minzu Xueyuan Xuebao* 3.82–92.
- DELANCEY, S. 1987. 'The Sino-Tibetan languages', pp. 799–810 of *The World's Major Languages*, ed. by Bernard Comrie. New York: Oxford University Press.
- 1991. 'Sino-Tibetan languages', pp. 445–9 of *The International Encyclopedia of Linguistics*, vol. 4, edited by William Bright. New York: Oxford University Press.
- DIXON, R. M. W. 1980. *The Languages of Australia*. Cambridge, London, and New York: Cambridge University Press.
- 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DRIEM, G. VAN. 1988. 'Reflexes of the Tibeto-burman *-t directive suffix in Dumi Rai', pp. 157–67 of *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg* (Pacific Linguistics C-104), edited by D. Bradley, E. J. A. Henderson, and M. Mazaudon. Canberra: Australian National University.
- DU RUOFU, YUAN YIDA, HUANG, J., MOUNTAIN, J., and CAVALLI-SFORZA, L. L. 1992. *Chinese Surnames and the Genetic Differences between North and South China*. JCL Monograph Series No. 5. Berkeley: Project on Linguistic Analysis.
- DWYER, A. M. 1992. 'Altaic elements in the Linxia dialect: contact-induced change on the Yellow River Plateau', *JCL* 20.160–78.
- EVANS, J. P. 1999. 'Contact-induced tonogenesis in Southern Qiang'. Paper presented at the Workshop on Qiangic Languages and Linguistics, Taipei, Nov. 1999.
- FAIRBANK, J. K., REISCHAUER, E. O., and CRAIG, A. M. 1989. *East Asia: Tradition and Transformation*. Revised edition. Boston: Houghton Mifflin Co.
- FEI XIAOTONG. 1980. 'Guanyu Woguo de minzu shibie wenti' [On the problem of distinguishing nationalities in China], *Zhongguo Shehui Kexue* 1.158–74.
- FITZGERALD, C. P. 1961. *China: A Short Cultural History*. London: The Cresset Press.
- 1972. *The Southern Expansion of the Chinese People: Southern Fields and Southern Ocean*. London: Barrie and Jenkins.
- FORREST, R. A. D. 1973. *The Chinese Language*. London: Faber and Faber.
- GE JIANXIONG, WU SONGDI, and CAO SHUJI. 1997. *Zhongguo yi min shi* [History of migrations in China]. Fuzhou: Fujian Renmin Chubanshe.
- GIRIDHAR, P. P. 1980. *Angami Grammar*. CIIL Grammar Series-6. Mysore: Central Institute of Indian Languages.
- GRIERSON, G. 1909. Editor of *Linguistic Survey of India*, vol. 3, parts 1–3: *Tibeto-Burman Family*. Calcutta: Superintendent of Government Printing.
- GRIMES, B. F. 1991. Editor of *Ethnologue: The Languages of the World*. 12th edition. Dallas: Summer Institute of Linguistics.
- HALL, D. G. E. 1981. *A History of South-East Asia*. London: Macmillan.
- HANSELL, M. D. 1989. 'Lexical borrowing in Taiwan'. Ph.D. dissertation, UC Berkeley.
- HASHIMOTO, M. J. 1976. 'Language diffusion on the Asian continent: problems of typological diversity in Sino-Tibetan', *CAAL* 3.49–63.
- 1980. 'Typography of phonotactics and suprasegmentals in languages of the East Asian continent', *CAAL* 13.153–64.

- 1986. 'The Altaicization of Northern Chinese', pp. 76–97 of *Contributions to Sino-Tibetan Studies*, edited by J. McCoy and T. Light. Leiden: E. J. Brill.
- 1992. 'Hakka in Wellentheorie perspective', *JCL* 20.1–48.
- HE JIREN. 1989. 'Yunnan Mengguzu yuyan ji qi xishu wenti' [The language of the Mongols of Yunnan Province and the problem of its geneological classification], *Minzu Yuwen* 1989.5.
- 1998. 'Guanyu Yunnan Mengguzu Gazhuoyu de xingcheng' [On the formation of the Gazhuo language of the Mongolians of Yunnan], *Minzu Yuwen* 4.51–4.
- HEATH, J. 1997. 'Lost wax: abrupt replacement of key morphemes in Australian agreement complexes', *Diachronica* 14.197–232.
- 1998. 'Hermit crabs: formal renewal of morphology by phonologically mediated affix substitution', *Language* 74.728–59.
- HEINE, B. 1994. 'Areal influence on grammaticalization', pp. 55–68 of *Language Contact Language Conflict*, edited by Martin Putz. Amsterdam and Philadelphia: Benjamins.
- 1997a. *Cognitive Foundations of Grammar*. New York and Oxford: Oxford University Press.
- 1997b. *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge: Cambridge University Press.
- HO DAH-AN. 1988. *Guilyu yu fangxiang: Bianqian zhong de yinyun jiegou* [Rule and direction: the changing structure of phonology]. Taipei: Institute of History and Philology.
- HOFFMAN, H. 1990. 'Early and medieval Tibet', pp. 371–99 of *The Cambridge History of Early Inner Asia*, edited by D. Sinor. Cambridge: Cambridge University Press.
- HUANG YUANWEI. 1990. 'Zhuangyu yu Yueyu, Wuming Guanhua de xianghu yingxiang' [The mutual influence of Zhuang, Cantonese, and Wuming Guanhua], pp. 173–8 of *Hanyu yu shaoshu minzuyu guanxi yanjiu* [Studies on the relationships between Chinese and the minority languages], edited by Zhongyang Minzu Xueyuan Xuebao Bianjibu. Beijing: Zhongyang Minzu Xueyuan.
- IWATA, R. 1995. 'Linguistic geography of Chinese dialects—Project on Han Dialects (PHD)', *Cahiers de Linguistique Asie Orientale* 24.195–227.
- JAKOBSON, R. 1938. 'Sur la théorie des affinités phonologiques entre les langues', pp. 45–59 of *Proceedings of the 4th International Congress of Linguists*, Copenhagen.
- JEYAPPAUL, V. Y. 1987. *Karbi Grammar*. Mysore: Central Institute of Indian Linguistics.
- JIN LIXIN. 1998. 'Han-Zangyu de mingci houzhui *-n' [The Sino-Tibetan noun suffix *-n], *Minzu Yuwen* 1.43–8.
- KANSAKAR, T. R. 1996. 'Multilingualism and the language situation in Nepal', *LTBA* 19.17–30.
- KEPPING, K. B. 1975. 'Subject and object agreement in the Tangut verb', *LTBA* 2.219–32.
- 1979. 'Elements of ergativity and nominativity in Tangut', pp. 263–77 of *Ergativity: Towards a Theory of Grammatical Relations*, edited by Frans Plank. London: Academic Press.
- 1981. 'Agreement of the verb in Tangut', *LTBA* 6.39–48.
- 1982. 'Once again on the agreement of the Tangut verb', *LTBA* 7.39–54.
- 1989. 'Xixiayu de jiegou' [The structure of the Tangut language], pp. 312–26 of *Zhongguo minzu shi yanjiu* [Studies on the history of the nationalities of China], vol. 2, edited by Bai Bin, Shi Jinbo, Lu Xun, and Gao Wende. Beijing: Zhongyang Minzu Xueyuan Chubanshe.
- KUBLER, C. C. 1985. *The Development of Mandarin in Taiwan: A Case Study of Language Contact*. Taipei: Student Book Co.

- LA POLLA, R. J. 1989. 'The phonetic development of Tibetan: a problem set', *LTBA* 11.91–5.
- 1992a. 'On the dating and nature of verb agreement in Tibeto Burman', *Bulletin of the School of Oriental and African Studies* 55.298–315.
- 1992b. 'Anti-ergative marking in Tibeto-Burman', *LTBA* 15.1–9.
- 1994a. 'Variable finals in Proto-Sino-Tibetan', *Bulletin of the Institute of History and Philology* 65.131–73.
- 1994b. 'Parallel grammaticalizations in Tibeto-Burman: Evidence of Sapir's "drift"', *LTBA* 17.61–80.
- 1995. 'Ergative marking in Tibeto-Burman', pp. 189–228 of *New Horizons in Tibeto-Burman Morpho-syntax* (Senri Ethnological Studies 41), edited by Yoshio Nishi, James A. Matisoff, and Yasuhiko Nagano. Osaka: National Museum of Ethnology.
- 2001. Editor of *A Linguistic Approach to Zhangzhung and Related Languages in the Indian Himalayas*, vol. 2: *The Tibeto-Burman Languages of Uttar Pradesh*. Osaka: National Museum of Ethnology.
- 2003. 'Why languages differ', pp. 113–144 of *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*, edited by D. Bradley, R. J. LaPolla, B. Michailovsky and G. Thurgood. Canberra: Pacific Linguistics.
- LAUFER, B. 1916 [1987]. 'Loan words in Tibetan', *T'oung Pao* 17.404–552. Reprinted in pp. 483–632 of *Sino-Tibetan Studies II*, edited by Hartmut Walravens and Lokesh Chandra. New Delhi: Rakesh Goel.
- LEACH, E. R. 1954 [1964]. *Political Systems of Highland Burma*. London and Atlantic Highlands, NJ: The Athlone Press.
- LEE, J. 1978. 'Migration and expansion in Chinese history', pp. 20–47 of *Human Migration: Patterns and Policies*, edited by William H. McNeill and Ruth S. Adams. Bloomington and London: Indiana University Press.
- 1982. 'The legacy of immigration in Southwest China, 1250–1850', *Annales de Démographie Historique* 1982: 279–304.
- and WONG BIN. 1991. 'Population movements in Qing China and their linguistic legacy', pp. 52–77 of Wang (1991a).
- LI FANG-KUEI. 1939. 'Languages and dialects', pp. 44–6 of *Chinese Yearbook*, 1938–9. Shanghai. (Republished in *JCL* 1.1–13.)
- LI JINGZHONG. 1994. *Yuyan yanbian lun* [On language change]. Guangzhou: Guangzhou Chubanshe.
- LI WEN-CHAO. 1995. 'The four grades: an interpretation from the perspective of Sino-Altaic language contact'. Paper presented at the 28th International Conference on Sino-Tibetan Languages and Linguistics, Charlottesville, VA, Oct. 1995.
- LIEN, C. 1987. 'Coexistent tone systems in Chinese dialects'. Ph.D. dissertation, UC Berkeley.
- 1993. 'Bidirectional diffusion in sound change revisited', *JCL* 21.255–76.
- 1997. 'The lexicon in competing changes', pp. 117–25 of *In Memory of Mantaro Hashimoto*, edited by A. O. Yue and Endo Mitsuaki. Tokyo: Uchiyama Shoten.
- LIN TAO. 1991. 'The scope of Pekinese', pp. 363–76 of Wang (1991a).
- LIN XIANGRONG. 1990. 'Sichuansheng Aba Zangzu Zizhizhou shuangyu diaocha baogao' [Report on the investigation of bilingualism in Aba Tibetan Autonomous Prefecture], *Zhongguo shaoshu minzu shuangyu jiaoxue yanjiu lunji* [Collected studies on bilingualism among China's ethnic minorities], edited by the Association for the Study of Bilingualism among China's Ethnic Minorities. Beijing: Minzu Chubanshe.

- LIN ZONGCHENG. 1986. 'Shi lun Tangdai xinan diqu Wuman Baiman ruogan wenhua tezhi' [Several characteristics of the culture of the Wuman and Baiman of the south-west during the Tang dynasty], *Zhongyang Minzuxueyuan Xuebao* 2.41–6.
- LIU GUANGKUN. 1981. 'Qiangyu zhong de Zangyu jieci' [Tibetan loanwords in Qiang], *Minzu Yuwen* 3.19–28.
- 1998. 'Lun Qiangyu shengdiao de chansheng he fazhan' [On the origin and development of tones in Qiang], *Minzu Yuwen* 2.1–8.
- Luce, G. H. 1937. 'The ancient Pyu', *Journal of the Burma Research Society* 27.239–53.
- 1976. 'Sources of early Burma history', pp. 31–42 of *Southeast Asian History and Historiography: Essays Presented to D. G. E. Hall*, edited by C. D. Cowan and O. W. Wolters. Ithaca and London: Cornell Press.
- and PE MAUNG TIN. 1939. 'Burma down to the fall of Pagan: an outline, part I', *Journal of the Burma Research Society* 29.264–73.
- MATISOFF, J. A. 1978. *Variational Semantics in Tibeto-Burman: The 'Organic' Approach to Linguistic Comparison*. Philadelphia: Institute for the Study of Human Issues.
- 1986. 'Linguistic diversity and language contact', pp. 56–86 of *Highlanders of Thailand*, edited by J. Mckinnon and W. Bhruksasri. Singapore: Oxford University Press.
- 1989. *The Dictionary of Lahu*. University of California Publications in Linguistics 111. Berkeley and Los Angeles: University of California Press.
- 1990. 'On megalocomparison', *Language* 66.106–20.
- 1991. 'Areal and universal dimensions of grammaticalization in Lahu', pp. 383–453 of *Approaches to Grammaticalization*, vol. 2, edited by E. C. Traugott and B. Heine. Amsterdam and Philadelphia: John Benjamins.
- MEI TSU-LIN. 1989. 'The causative and denominative functions of the *s- prefix in old Chinese', pp. 23–32 of *Proceedings of the Second International Conference on Sinology*. Taipei: Academia Sinica.
- MENG SIMU. 1998. 'Hanyu he Zhuangdongyu de miqie guanxi ji lishi wenhua beijing' [The close relationship between Chinese and Zhuang-Dong languages and its historical and cultural background], *Minzu Yuwen* 4.43–50.
- MICHAILOVSKY, B. 1985. 'Tibeto-Burman dental suffixes: evidence from Limbu (Nepal)', pp. 334–43 of *Linguistics of the Sino-Tibetan Area: The State of the Art. Papers Presented to Paul K. Benedict for his 71st Birthday* (Pacific Linguistics C-87), edited by G. Thurgood, J. A. Matisoff, and D. Bradley. Canberra: Australian National University.
- MOUNTAIN, J. L., WANG, W. S.-Y., DU RUOFU, YUAN YIDA, and CAVALLI-SFORZA, L. L. 1992. 'Congruence of genetic and linguistic evolution in China', *JCL* 20.315–30.
- NISHI YOSHIO. 1986. 'Gendai Chibetogo hoogen no bunrei' [A classification of Tibetan dialects], *Bulletin of the National Museum of Ethnology* 11.837–900.
- NORMAN, J. 1988. *Chinese*. Cambridge: Cambridge University Press.
- 1991. 'The Min dialects in historical perspective', pp. 325–60 of Wang (1991a).
- and MEI TSU-LIN. 1976. 'The Austroasiatics in ancient South China: some lexical evidence', *Monumenta Serica* 32.274–301.
- OKELL, J. 1965. 'Nissaya Burmese: a case of systematic adaptation to a foreign grammar and syntax', *Lingua* 15.186–227.
- OPPITZ, M. 1974. 'Myths and facts: reconsidering some data concerning the clan history of the Sherpas', *Kailash* 2.120–31.
- POFFENBERGER, M. 1980. *Patterns of Change in the Nepal Himalaya*. Delhi: The Macmillan Co. of India.

- PULLEYBLANK, E. G. 1983. 'The Chinese and their neighbors in prehistoric and early historic times', pp. 411–66 of *The Origins of Chinese Civilization*, edited by David N. Keightley. Berkeley: University of California Press.
- PULLEYBLANK, E. G. 1991. 'Chinese dialect studies', pp. 431–55 of Wang (1991a).
- QIAN NAIRONG. 1991. 'The changes in the Shanghai dialect', pp. 377–428 of Wang (1991a).
- 1997. *Shanghaihua yufa* [Grammar of the Shanghai dialect]. Shanghai: Renmin Chubanshe.
- QU AITANG. 1985. 'Han-Zang Yuyan diaozhi yanjiu de jiazhi he fangfa' [The value and methodology of studying the tonal contours of Sino-Tibetan languages], *Minzu Yuwen* 6.1–14.
- RAN GUANGRONG and ZHOU XIYIN. 1983. 'Lun Gan-Qing gu wenhua yu Qiangzu de guanxi' [On the relationship between the culture of the Gansu-Qinghai area and the Qiang nationality], pp. 215–34 of *Xinan minzu yanjiu* [Studies on the south-west nationalities], edited by Zhongguo Xinan Minzu Yanjiu Xuehui. Chengdu: Sichuan Minzu Chubanshe.
- LI SHAOming, and ZHOU XIYIN. 1984. *Qiangzu shi* [History of the Qiang nationality]. Chengdu: Sichuan Minzu Chubanshe.
- SAXENA, A. 1988a. 'On syntactic convergence: the case of the verb "say" in Tibeto-Burman', *Proceedings of the Fifteenth Annual Meeting of the Berkeley Linguistics Society* 375–88.
- 1988b. 'On the grammaticalization of the verb "say": a typological and diachronic study', in *Papers from the 3rd Pacific Linguistics Conference*, edited by S. DeLancey and R. Tomlin, Eugene: University of Oregon.
- SHAFFER, R. 1955. 'Classification of the Sino-Tibetan languages', *Word* 11.94–111.
- 1966. *Introduction to Sino-Tibetan*, part I. Wiesbaden: Otto Harrassowitz.
- SHARMA, D. D. 1989. *Tibeto-Himalayan Languages of Uttarakhand*, part I: Studies in Tibeto-Himalayan languages 3. Delhi: Mittal Publications.
- SHARMA, S. R. 2001. 'A sketch of Rongpo grammar', pp. 187–194 of LaPolla (2001).
- SHIRATORI, YOSHIRO. 1950. 'Nansyo kyuubi Dairi no minsoku to sono imin, Minka no gengo keitoo ni tsuite [On the tribes of Nan-chao and Ta-li and the descent of the languages of their descendents, the Min-chia], *Japanese Journal of Ethnology* 15.52–63.
- SHREE KRISHAN. 2001a. 'A Sketch of Chaudangsi Grammar', pp. 401–448 of LaPolla (2001).
- 2001b. 'A Sketch of Raji Grammar', pp. 449–501 of LaPolla (2001).
- SINGH, A. P. 1986. 'Beliefs and rituals among the Bhotias of Uttarakhand Himalaya—a study of persistence and change', pp. 168–71 of *Ecology, Economy and Religion of Himalays*, edited by L. P. Vidyarthi and Makhan Jha. New Delhi: Orient Publications.
- SNELGROVE, D., and RICHARDSON, H. 1986. *A Cultural history of Tibet*. Boston and London: Shambhala.
- STEIN, R. A. 1961. *Les tribus anciennes des marches sino-tibétaines; légendes, classifications et histoire*. Melanges publiés par l'Institut des Hautes Etudes chinoises. Paris: Presses Universitaires de France.
- STERN, T. 1962. 'Language contact between related languages: Burmese influences upon plains Chin', *Anthropological Linguistics* 4.1–28.
- SUN HONGKAI. 1981. *Qiangyu jianzhi* [A brief description of the Qiang language]. Beijing: Nationalities Press.
- 1983. 'Chuanxi "minzu zoulang" diqu de yuyan' [The languages of the 'ethnic corridor' region of western Sichuan], *Research on the Peoples of the Southwest* 1.429–54.
- 1988. 'Shilun Zhongguo jingnei Zang-Mian yu de puxi fenlei' [A classification of

- Tibeto-Burman languages in China], pp. 61–73 of *Languages and History in East Asia: A Festschrift for Tatsuo Nishida on the Occasion of his 60th Birthday*, vol. I, edited by P. K. Eguchi et al. Kyoto: Shokado.
- THURGOOD, G. 1985. 'Pronouns, verb agreement systems, and the subgrouping of Tibeto-Burman', pp. 376–400 of *Linguistics of the Sino-Tibetan Area: The State of the Art*, edited by G. Thurgood, J. A. Matisoff, and D. Bradley. Canberra: Pacific Linguistics.
- TIAN JIZHOU. 1989. 'Chuguo ji qi minzu' [The country of Chu and its nationalities], *Zhongguo Minzushi Yanjiu* 2.1–17.
- TIWARI, S. C. 1986. 'Some aspects of religious practices among the Rajis of Kumaon Himalaya', pp. 180–92 of *Ecology, Economy and Religion of Himalays*, edited by L. P. Vidyarthi and Makhan Jha. New Delhi: Orient Publications.
- TONG ENZHENG. 1998. *Gudai de Ba Shu* [Ancient Ba and Shu]. Chongqing: Chongqing Chubanshe.
- TREISTMAN, J. M. 1972. *The Prehistory of China*. Garden City, NY: Natural History Press.
- WANG HUIYIN. 1989. 'Chunqiu Zhanguo shiqi de minzu yuyan gaikuang he yuyan guanxi shuolue' [A brief discussion of the nationality languages and their relationships in the Spring and Autumn and Warring States periods], *Zhongyang Minzu Xueyuan Xuebao* 6.72–5, 79.
- WANG JUN. 1991. 'Language interaction in China', pp. 161–86 of Wang (1991a).
- WANG JUNG-KE. 1992. 'The Ch'iang of Ancient China through the Han Dynasty: Ecological Frontiers and Ethnic Boundaries'. Ph.D. dissertation, Harvard University.
- WANG, W. S.-Y. 1991a. Editor of *Languages and Dialects of China*. JCL Monograph Series 3. Berkeley: Project on Linguistic Analysis.
- 1991b. *Explorations in Language*. Taipei: Pyramid Press.
- 1995. Editor of *The Ancestry of the Chinese Language*. JCL Monograph Series 8. Berkeley: Project on Linguistic Analysis.
- and LIEN CHINEA. 1993. 'Bidirectional diffusion in sound change', pp. 345–400 of *Historical Linguistics: Problems and Prospectives*, edited by C. Jones. London: Longman Group.
- WIENS, H. J. 1967. *Han Chinese Expansion in South China*. Hamden, CT: Shoe String Press.
- XING GONGWAN. 1996. 'Han-Zangyu yanjiu he Zhongguo kaoguxue' [Sino-Tibetan studies and Chinese archeology], *Minzu Yuwen* 4.18–28.
- XU JIESHUN. 1989. 'Zhongguo gudai nanbei minzu guanxi shi bijiao yanjiu duanxiang' [A comparative study of the histories of the northern and southern nationalities in Ancient China], *Sixiang Zhanxian* 1.55–61.
- YOU RUJIE. 1982. 'Lun Taiyu liangci zai Hanyu nanfang fangyan zhong de diceng de yicun' [On the substrata and remnants of Tai classifiers in Chinese southern dialects], *Minzu Yuwen* 3.
- 1995. 'Zhongguo nanfang yuyan li de niao chong lei mingci citou ji xiangguan wenti' [The prefixes of bird and insect names in the languages of southern China and other questions], pp. 253–68 of Wang (1995).
- YUE-HASHIMOTO, A. O.-K. 1967. 'Southern Chinese dialects: the Tai connection', *CAAAL* 6.1–9.
- 1991. 'The Yue dialect', pp. 294–324 of Wang (1991a).
- 1993. 'The lexicon in syntactic change: lexical diffusion in Chinese syntax', *JCL* 21.213–53.

- ZHAO TONGMAO and LEE TSUNG DAO. 1989. 'Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation', *Human Genetics* 83,101–10.
- ZHOU ZHENHE. 1991. 'Migrations in Chinese history and their legacy on Chinese dialects', pp. 29–51 of Wang (1991a).
- and YOU RUJIE. 1986. *Fangyan yu Zhongguo wenhua* [Dialects and Chinese culture]. Shanghai: Shanghai Renmin Chubanshe.

On Genetic and Areal Linguistics in Mainland South-East Asia: Parallel Polyfunctionality of ‘acquire’

N. J. Enfield

This chapter raises questions concerning genetic and areal relatedness among languages of Mainland South-east Asia (hereafter MSEA),¹ mainly with reference to a widespread pattern of grammatical polyfunctionality involving a verb ACQUIRE. Although data are mostly from Sinitic and Tai, the issues of genetic versus areal relatedness arise across and throughout the five or more language families in the region.

I begin with introductory comments on the geographical, linguistic, and cultural situation of MSEA, including discussion of MSEA as a linguistic area. In §2, I present data from a synchronic case study of a polyfunctional verb ACQUIRE in MSEA languages, concentrating on two Tai languages (Lao and Northern

I would like to thank the editors for generously inviting me to contribute. I am indebted to the following people for helpful input: Sasha Aikhenvald, Umberto Ansaldi, Bob Bauer, Hilary Chappell, Gérard Diffloth, Tony Diller, Bob Dixon, Jerry Edmondson, Grant Evans, Nick Evans, Cliff Goddard, Randy La Polla, Jim Matisoff, Stephen Matthews, Andy Pawley, and Malcolm Ross. Unmarked Modern Standard Chinese examples are checked with native speakers. Lao examples are from my own corpus of texts (references are to L1, with page number), and fieldnotes (1996–9). Northern Zhuang examples are from Luo 1990, chapter 3, and Luo Yongxian, personal communication. South-Western Mandarin examples are from fieldnotes (Jing Hong, China, and Oudom Xay, Laos, September 1999) and consultation with Luo Yongxian in Brisbane, July 1998. (Transcription of South-Western Mandarin uses Pinyin, with tones unmarked.) Unmarked Kmhmu data are from fieldnotes (Vientiane, Laos, July 1998). Detailed supporting discussion of the data in Table 5 may be found in Enfield (2003). Pacoh data are from fieldnotes (Saravane, Laos, August–October 1999). Vietnamese data are from fieldnotes (Vietnam and Laos 1997–9), and Thompson (1987).

¹ Abbreviations for branches of language families are EMK (Eastern Mon-Khmer), NMK (Northern Mon-Khmer), SWT (South-Western Tai). Abbreviations for languages used in examples are as follows: AH (Ahom [SWT; India, Burma]), CA (Cantonese [Sinitic; China]), DG (Dong [Kam-Sui, China]), KH (Khmer [EMK; Cambodia, Laos, Thailand, Vietnam]), KM (Kmhmu [NMK; Laos, Vietnam, Thailand, China]), LAO (Lao [SWT; Laos, Thailand, Cambodia]), MSC (Modern Standard Chinese [Sinitic; China]), MU (Mulao [Kam-Sui; China]), NZH (Northern Zhuang [Northern Tai; China, Vietnam]), PA (Pacoh [EMK; Laos, Vietnam]), SWM (South-Western Mandarin [Sinitic; China, Laos]), TH (Thai [SWT; Thailand]), VN (Vietnamese [EMK; Vietnam]).

Zhuang) and two Sinitic languages (Cantonese and Modern Standard Chinese), with additional data from South-Western Mandarin. Section 3 brings in historical evidence from Tai(-Kadai) and Sinitic, as well as comparative evidence from Eastern Mon-Khmer. Issues arise concerning the distinction between areal and genetic relatedness, and it is noted, in addition, that while borrowing and common inheritance are two possible accounts for sharing of structures between languages, common language-internal mechanisms also need to be taken into account. The 'naturalness' of an innovation can result in a higher degree of common grammatical patterning due to independent innovation, and this naturalness may be defined with respect to human cognitive propensities, or to the semantic/grammatical developments made possible or likely by the language's given state of semantic and grammatical organization (or its *typological poise*).

1. Introductory discussion: the Mainland South-east Asian area

1.1. GEOGRAPHY

MSEA encompasses Vietnam, Cambodia, Laos, Thailand, Peninsular Malaysia, Burma, parts of north-east India, and extensive areas of southern and south-western China.

MSEA is a hilly monsoonal region with rivers descending into large basins, such as the Irrawaddy River valley in Burma, the broad Chao Phraya valley in central Thailand, and the long reaches of the Mekong from south-west China, through Laos, Thailand, and Cambodia, to the delta in southern Vietnam. Most of these flatter lowland areas have been well populated by paddy-rice farmers for centuries, and hillier regions show greater diversity, with wet-rice cultivation practised in some areas (where flat land can be found), and shifting 'dry-field' rice cultivation (on slopes) in others. Typically, those practising these different livelihoods also speak different languages.

Geography has naturally helped determine patterns of migration over the centuries, with large rivers and their tributaries hosting significant downstream southward migration, especially from south-west China into the lower hills and plains of Thailand, Laos, and Vietnam (Edmondson 1998a, b). Patterns of human movement have been complex and widespread, with ongoing juxtaposition of rather different peoples, whose relations are defined by important social and political factors (cf. Leach (1964) on north-east Burma; LaPolla, Chapter 9, on China).

1.2. LANGUAGE(S)

At least five widely accepted linguistic groups are found in MSEA, namely Hmong-Mien, Sino-Tibetan (including Sinitic and Tibeto-Burman), Tai-Kadai, Austroasiatic (including Mon-Khmer), and Austronesian (not discussed in this chapter). These include the national languages of Burma, Cambodia, China, Laos,

Thailand, and Vietnam. English and French have had a considerable recent presence, as Pali and Sanskrit have had in earlier times.

MSEA hosts great linguistic diversity, especially in the rugged mountain areas, yet these more isolated locations are unfortunately the places in which we know the least about grammar and semantics. Relatively little descriptive work has been done, and while fieldwork activity is increasing, not much of it is on grammar. Current work is mostly aimed at comparative reconstruction and especially language classification, and so is often restricted to small word lists. One reason for the general lack of fieldwork-based description has been inaccessibility of the relevant areas, due to political as well as geographical factors. With current political and economic development, research in the area is becoming easier.

1.3. CULTURE(S)

MSEA as a whole displays a range of cross-cutting and overlapping cultural commonalities, and cannot be considered a single distinct 'culture area'. While the religious, political, and economic influences of Sinitic and Indic cultures have been historically significant, and are obvious today, these have been predated by indigenous cultures which also have a modern presence, undercutting realms of Indic and Sinitic influence (cf. Steinberg 1987). Of relevance to the linguistic situation, many rather different groups have cohabited, and, perhaps more importantly, have been apt to fluidity in ethnic identity, for political and other reasons (cf. Evans 1999a, b, Keyes 1977, Leach 1964).

Movements of people, and associated social changes of linguistic consequence, have occurred at many levels of grain. More broadly, significant civilizations have dominated open river valleys (such as the Chao Phraya Basin), and in these circumstances, 'cohabitation' of different human groups has resulted in disappearance of cultural (and linguistic) distinctions. Consider, for example, those who are nowadays referred to as 'Tai', including significant minorities of southern China, north Vietnam, and north-east Burma, as well as dominant populations of Thailand and Laos. As an ethnic group, the Tai are defined by the 'genetic relatedness' of languages they speak (assumed by many to indicate speakers' common ancestry). This does not, however, correspond to a comparable level of genetic relatedness among the speakers themselves. There is evidence that today's Tai-speaking populations of lowland Thailand and Laos are mostly descendants of former Mon-Khmer-speaking inhabitants of the same area (Samerchai 1998). These people 'became Tai' linguistically and culturally, perhaps ultimately for economic reasons, related to superior agricultural technology that Tai-speaking populations brought with them from southern China (Hartmann 1998; cf. Leach 1964 on the same ethnolinguistic fluidity among neighbouring Kachin (Tibeto-Burman) and Shan (Tai) in north-east Burma).

There has been an epic history in MSEA of social movements and interactions,

with associated competition of fashions in both cultural and linguistic practice, between generations, and among neighbouring groups. These fluctuations, congregations, dispersals, cross-societal arrangements, and temporary 'equilibria', have resulted in today's complex cultural and linguistic situation.

1.4. MAINLAND SOUTH-EAST ASIA AS A LINGUISTIC AREA

Languages of MSEA share a great deal of grammatical structure, from broad typological traits to quite specific features, with varying degrees of overlap among languages (Bisang 1991, Clark 1989, Matisoff 1991, Chapter 11, Migliazza 1996). Languages of the region lack case-marking or cross-referencing in the usual sense of these terms. Disambiguation of 'who' and 'whom' relies on semantic and pragmatic context, and in the last resort is often achieved by constituent order. These languages are extremely open to leaving interpretation (e.g. of predicate–argument relations, tense, aspect–modality) to context, and both constituent order variation and ellipsis are common. Normal utterances are often impossible to interpret properly outside the contexts in which they actually occur.

All languages of MSEA use classifier constructions for enumeration, individuation, and other forms of nominal grounding. With respect to verb–phrase structure, the languages all display verb serialization. In phonology, lexical tone is an obvious areal feature, although not always found (e.g. it is mostly absent among Austroasiatic languages). Phonotactically, syllable-final consonants are highly restricted, with only a fraction of full consonant inventories permissible in syllable-final position.

Table 1 shows a few MSEA areal features across five language families (ignoring some exceptions).

Further to these broader generalizations, some more specific grammatical features enable subdistinctions. Tibeto-Burman languages are distinct from the rest in being mostly verb-final rather than verb-medial. This generalization is not absolute—note the presence of verb-final constructions in Tai and Sinitic, such as the so-called 'disposal construction' (of the form 'NP_{SUBJ} take NP_{OBJ} V_{TR}'), directional constructions with the likes of 'go' and 'come' in final position, and similar

TABLE 1. Some Mainland South-East Asian areal features

	Austroasiatic	Tai-Kadai	Hmong-Mien	Sinitic	Tibeto-Burman
Case-marking	–	–	–	–	–
Cross-referencing	–	–	–	–	–
Fusional affixing	–	–	–	–	–
Classifier constructions	+	+	+	+	+
Verb serialization	+	+	+	+	+
Lexical tone	±	+	+	+	±

TABLE 2. Some Mainland South-East Asian areal subdistinctions

	Austroasiatic	Tai-Kadai	Hmong-Mien	Sinitic	Tibeto-Burman
Verb-object	+	+	+	+	–
Prepositions	+	+	+	±	–
Adjective-standard of comparison	+	+	+	±	–
Head-modifier	+	+	+	–	±
Head-relative clause	+	±	+	–	–
Possessed–possessor	+	+	–	–	–

multiverb constructions. Some Austroasiatic languages are reportedly verb-final, and conversely, some Tibeto-Burman languages, such as those in the Karenic branch, are verb-medial, like their Tai and Sinitic neighbours. With respect to adpositions and comparative constructions, Tibeto-Burman languages have *post*-positions, and place the standard of comparison *before* the adjective predicating the quality of comparison. Sinitic languages are divided and/or mixed in these respects. Southern Sinitic languages group with the majority of MSEA languages in putting the standard of comparison after the element predicating the quality being compared. Sinitic languages in general use both postpositions (denominal) and prepositions (deverbal). Noun phrases are overwhelmingly head-initial in both Tai and Austroasiatic languages, while they are strongly head-final in Sinitic languages, and in Tibeto-Burman languages generally (different types of nominal attribution may display different head/attribute ordering; cf. Okell (1969) on Burmese). Hmong-Mien languages group with Mon-Khmer and Tai in having adjectives and relative clauses follow head nouns, but group with Sinitic and Tibeto-Burman in having possessors precede possesseds. These generalizations (again not exceptionless) are summarized in Table 2.

In some cases, specific features are common to certain languages only. For example, Vietnamese patterns with Khmer and Lao in that each have a possessive marker derived from a nominal meaning ‘stuff, things’ (cf. Clark 1989):

- (1)

h^uan *khɔ̌ɔŋ* *khɔ̌ɔj*

LAO house thing (‘of’) 1

‘my house’
- (2)

ptɛ̌ah *rəbɔ̌h* *kñom*

KH house thing (‘of’) 1

‘my house’
- (3)

nhà *của* *tôi*

VN house thing (‘of’) 1

‘my house’

Kmhmu, however, spoken literally among and between these languages, derives a possessive marker from a verb *dé'* 'take'/'get':²

- (4) *kmuul dé' ge*
 KM money POSS 3msg
 'his money'

In other respects, however, Vietnamese patterns grammatically like Sinitic languages, *not* like other Mon-Khmer languages and Tai. Consider, for example, pseudo-reflexive emphatic constructions and classifier constructions (using Modern Standard Chinese as a representative Sinitic language):

- (5) (a) LAO *láaw hían 'éeng*
 3 study self
 (b) MSC *tā zì(-jǐ) xué*
 (c) VN *nó tự học*
 3 self study
 'He learned/studied (it) by himself.'
- (6) (a) LAO *mǎa sǎam tǎo*
 dog three CL
 (b) MSC *sān zhī gǒu*
 (c) VN *ba con chó*
 three CL dog
 'three dogs'

Some languages feature competing options, with one construction 'genetically acquired', another 'contact acquired'. In Mulao, alternative orderings of nominal head and modifier are often possible, either in the Tai(-Kadai) head-initial style (7a), or Sinitic head-final style (7b):

- (7) (a) *a:n¹ ma⁴*
 MU saddle horse (cf. Lao *ʔān mǎa* [saddle horse] 'saddle')
 (b) *ma⁴ a:n¹*
 horse saddle (cf. MSC *mǎ ʔān* [horse saddle] 'saddle')

Another example of 'genetic' versus 'contact' acquired grammar in competition concerns causatives in Kmhmu. Mon-Khmer systems of productive derivational morphology (e.g. morphological causatives; Clark 1989: 200–2) are in decline, apparently due to areal pressure. Surrounding languages are isolating, displaying periphrastic and/or lexical causativization. In Kmhmu as spoken in northern Thailand, two types of causative construction are in competition.

² *Dé'* has other grammatical uses—e.g. as a dative marker—and may occasionally be used to refer to 'stuff'. Here, the 'stuff'/'possessive marker' polysemy has apparently been derived in the opposite direction to that assumed for the likes of Vietnamese, Lao, and Khmer (i.e. 'possessive marker' loses its initial 'possessed' complement, and refers to the possessed most generally, as 'stuff').

'Native' morphological causatives, in the following (b) examples, are formed from the simple verbs shown in the (a) examples (Suwilai 1987: 25 ff.):

- (8) (a) *nà: kà:j tà kì:*
 KM 3fsg come here
 'She came here.'
 (b) *rù: nà: p-kà:j*
 KM pull 3fsg CAUS-come
 'Pull her towards (me).'
- (9) (a) *só' pə' mäh*
 KM dog eat rice
 'Dogs eat rice.'
 (b) *nà: (mò:t mäh) pn-pə' só'*
 KM 3fsg take rice CAUS-eat dog
 'She took rice to feed to the dog.'

Periphrastic causativization provides a competing alternative. Two patterns are shown in (10), looking suspiciously like neighbouring Thai structures, shown in (11), following:³

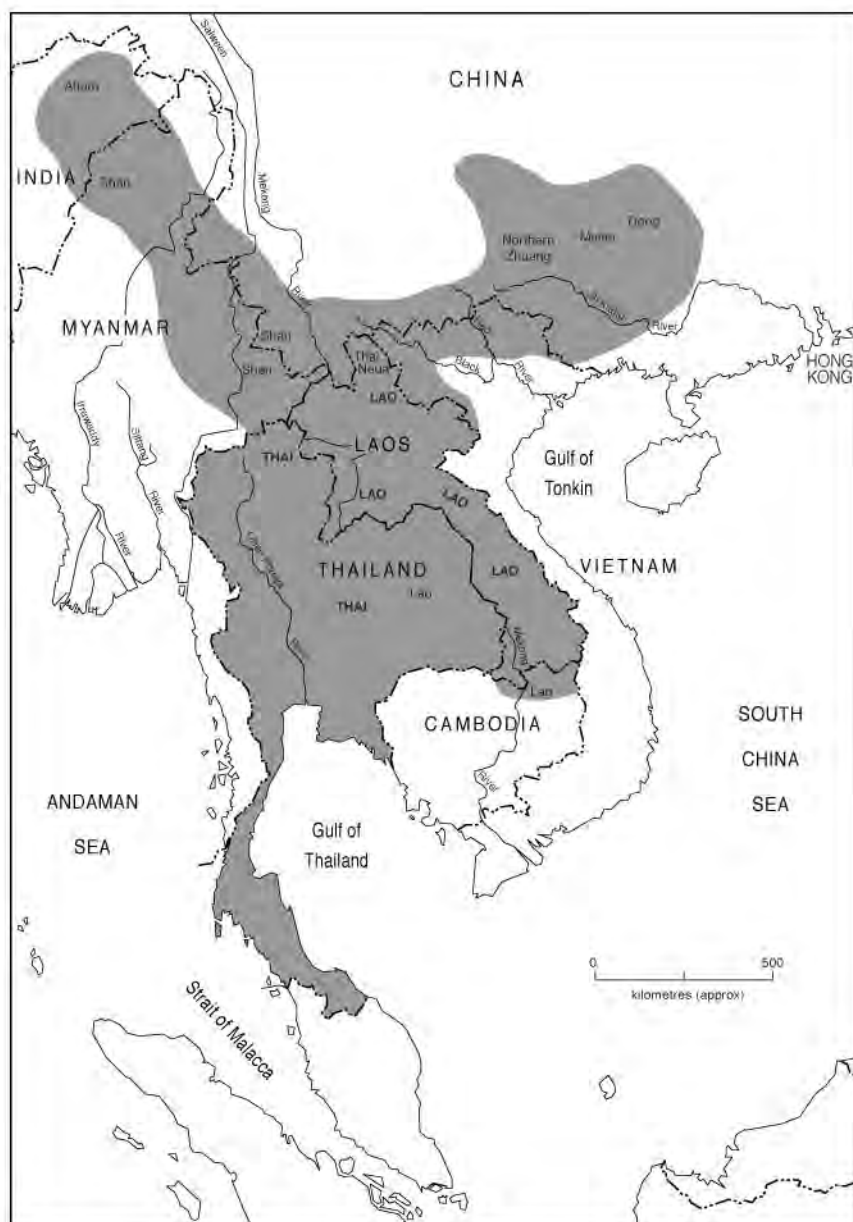
- (10) (a) *nà: 'úan só' pə' mäh*
 KM 3fsg give/make dog eat rice
 'She fed the dog.'
 (b) *nà: mò:t mäh 'úan só' pə'*
 KM 3fsg take rice give/make dog eat
 'She took rice to feed the dog.'
- (11) (a) *kháw hāj mǎa kin khâaw*
 TH 3 give/make dog eat rice
 'She let/made the dog eat rice.'
 (b) *kháw 'aw khâaw hāj mǎa kin*
 TH 3 take rice give/make dog eat
 'She took rice to give the dog to eat.'

1.5. VARIATION IN A LANGUAGE FAMILY DUE TO AREAL PRESSURES: THE CASE OF TAI(-KADAI)

The Tai family (a branch of 'Tai-Kadai') is standardly assumed to branch into Northern Tai, Central Tai, and South-Western Tai. Proto-Tai was probably spoken somewhere in the northern part of Guangxi Province, where there is greatest variety (Edmondson and Solnit 1997, Luo 1997, amongst others).

Application of the comparative method has led to reconstruction of a sizeable fraction of Proto-Tai, to a reasonable time depth (between one and three thousand

³ Translations of (10) and (11) differ here, not because they are not synonymous, but because translations of (10) are from the original source, Suwilai (1987). Translations of (11) are mine.



MAP 2. Approximate distribution of Tai languages

years). (For historical Tai, consult Benedict 1975, Edmondson and Solnit 1997, Gedney 1989, Li 1977, Luo 1997.) Li (1977) offered over 1,200 lexical items, in an orthodox and well-behaved reconstruction. While he did not speculate on a time or place for Proto-Tai, the idea that it was spoken in southern China at around the time of Christ or before had already been suggested by Chamberlain (1972), and is now widely assumed. Luo (1997) has reconstructed over 900 further Proto-Tai forms, many of which appear to have cognates in Sinitic languages (as do many of Li's original items). While both Li and Luo are directly concerned with Proto-Tai, their results are suggestive of either a Sino-Tai hypothesis (i.e. an earlier common origin to the Sinitic and Tai families) or an early period of contact. Many scholars have intuitions about this hypothesis, but available data at present are inconclusive (see, however, Bauer 1996, Dai 1991, and papers by Egerod, Gedney, Prapin, and Yue-Hashimoto in CAAAL 1976). There has also been an Austro-Tai hypothesis (Benedict 1975), that Tai languages belong in a sub-branch of Austronesian. This proposal has been less widely supported (cf. Gedney 1976).

Tai languages are now spoken across a large area, from south-east China to north-east India. Outside the two officially Tai-speaking states, Laos and Thailand, Tai speakers are surrounded by influential languages associated with nation-states, such as Assamese, Burmese, Cantonese, Khmer, Modern Standard Chinese, and Vietnamese, and are therefore often subject to strong pressure from language contact. In phonology, some unusual features are apparently due to contact. For example, where most Tai languages lack a voicing distinction in velar stops, some in north Vietnam do have a contrastive voiced velar stop (or fricative), as reported by Ross (1996). Vietnamese, with which these languages are in intensive contact, has—correspondingly—such a distinction (Thompson 1987: 25–8). In what follows, I am concerned not with phonology, but with contact-related differences in morphosyntactic behaviour among Tai-Kadai languages.

Most Tai-Kadai languages are almost exclusively head-initial, in nominal phrases particularly, while Sinitic languages display head-*final* nominal phrases. Some Tai-Kadai languages spoken in China follow Sinitic by allowing head-final nominal structures (cf. Gedney 1989: 122). Evidence from Mulao shows that left- and right-headed structures can be in competition, where right-headed tendencies are due to more recent convergence with Sinitic, with which speakers of these languages are now in intensive contact (cf. also Dong; Long and Zheng 1998).

The following Mulao example (12a) shows a head-initial noun phrase, a simple relative clause analogous in form to (12b) from Lao:

- (12) (a) *ngwa¹ lik⁸ na:n⁴*
 MU dog chase deer
 'a (deer-)hunting dog' (Wang and Zheng 1993: 29)
- (b) *mǎa lāj fáan*
 LAO dog chase deer
 'a deer-hunting dog' (also: 'A dog is chasing a deer.')

Compare this to the head-final noun phrase in (13a), a relative clause parallel in structure to Cantonese (and other Sinitic languages), illustrated in (13b) (with original glosses—‘PCL’ and ‘LP’ perform the same function):

- (13) (a) *ɲa:u³ twa³ jəu⁶ fong¹ jəu⁶ kuang¹ kə ga:n² ngwa⁴*
 MU one CL both tall both bright PCL house tile
 ‘a tiled house both tall and bright’ (Wang and Zheng 1993: 87)
- (b) *ngóh chéng ge gūngyàhn*
 CA 1 hire LP maid
 ‘the maid I hire’ (Matthews and Yip 1994: 88)

In the following example, the nominal head is ‘thread’, which appears both as a Tai form initially, and as a Sinitic form finally:

- (14) *pɹa:n⁶ tshə⁵ sjen⁴*
 MU thread machine thread
 ‘thread used for sewing machine’ (Wang and Zheng 1993: 31)

Here, *tshə⁵ sjen⁴* ‘machine thread’ is borrowed whole from Sinitic (cf. head-final Mandarin structure *jī xiàn* [machine thread] ‘machine thread’), and this whole expression is conceivably not analysed by speakers (i.e. is not headed one way or the other), becoming a simple modifier of the native Mulao nominal *pɹa:n⁶* ‘thread’, in the usual head-initial Tai order. However, the following example, a calque from Sinitic with noun-modifier order, includes one Tai and one Sinitic element, making it hard to imagine that the initial element (*fɪ¹* ‘fire’, a Tai word) is not recognized by speakers as a morphologically distinct modifier:

- (15) *fɪ¹ tshja¹*
 MU fire vehicle
 ‘train’ (Wang and Zheng 1993: 31)
 (cf. Mandarin *huǒ chē* [fire vehicle] ‘train’, Lao *lōt fáj* [vehicle fire] ‘train’)

A further case of syntactic variation in the noun phrase across Tai-Kadai concerns the classifier phrase, with a division between languages north, and south, of the Red River. In Northern Zhuang and Dong, spoken north of the Red River in south-west China, the normal order is [NUMERAL-CLASSIFIER-HEAD-(MODIFIER)]:

- (16) *sǎam ǎn lán hāa*
 NZH three CL house thatch.grass
 ‘three grass-thatched houses’

- (17) *ja²¹² tu²¹² jan²¹² sau³⁵ ta³³*
 DG two CL house 2 that
 ‘those two (animals) from your house’ (Long and Zheng 1998: 94)

These languages pattern like Sinitic (and Hmong-Mien) languages, which similarly place classifiers before head nouns (although the distinction is not

absolute; cf. e.g. Matthews and Yip (1994: 405)—in many languages, alternative orders are available, with a semantic/pragmatic distinction). Further south, in Lao and surrounding languages, the normal order is [HEAD-(MODIFIER)-NUMERAL-CLASSIFIER], with classifiers placed *after* head nouns:

- (18) *mǎa nāj sǒng tǔo*
 LAO dog big two CL
 'two big dogs'

While Tai languages display right-headed noun-phrase organization in situations of contact with Sinitic, *no* Tai language shows the full extent of nominal right-headedness found in Sinitic languages. No Tai language allows fully productive head-final ADJ-N order in simple attributive nominal phrases. Conversely, no Sinitic language displays N-ADJ as the basic, fully productive noun-attribute ordering. Even where head-modifier order is found to some degree, such as in Cantonese and other southern Sinitic languages, the pan-Sinitic head-final pattern remains dominant in simple noun-attribute expressions. However, the extent to which mixtures of types are allowed is an issue which deserves attention, of relevance to the question of whether Cantonese (and other southern varieties of Sinitic) are related at a deeper level to Tai languages, or as some suggest, have a Tai substrate (Bauer 1996). Nevertheless, it may be assumed that Proto-Tai and its Sinitic contemporarie(s) were head-initial and -final, respectively, in core noun-phrase organization. Areal influence, even when extreme, has not overridden this distinction.

A different example of contact-related grammatical variation in Tai concerns the verb phrase. Tai languages in Assam are surrounded by verb-final languages, and similarly are verb-final (cf. Diller 1992). In modern Ahom, like in Assamese (and unlike in, say, Thai), 'the usual order [of constituents in a transitive clause] is subject, direct object, verb' (Grierson 1903: 102; both examples from same, with original transcription, glosses, translation):

- (19) *luk ngi pun mǔng jau khau-ū-koi*
 AH son younger beyond country far entered-has
 'The younger son entered a foreign country.'

- (20) *mān-ko tāng khráng-ling tāk-lū tāk-pāng kin-jau-o*
 AH he all property diminished spent eaten-had
 'He had diminished, spent and eaten all the property.'

However, Ahom manuscripts from the fifteenth century display verb-*medial* clause organization, in the manner of other South-Western Tai languages, such as Lao and Thai (examples from Terwiel and Ranoo 1992: 80):

- (21) *sang khaw pak na la ka*
 AH if enter space front naga
 'If it enters the space in front of the naga.'

(22) *sang thuk neuw cik cong*

AH if reach star C.

'If it reaches the Cik Cong star.'

Over the last five hundred years, speakers of Tai languages in Assam have interacted with speakers of more dominant head-final languages of the Tibeto-Burman and Indo-Aryan families, changing the way they speak accordingly.

1.6. MODELLING 'LANGUAGE CONTACT' AND DIFFUSION IN MAINLAND SOUTH-EAST ASIA

The above examples of areal diffusion, along with the data presented in §2, are the product of a long and complex history of human relations, only some of which has been documented and/or inferred. For example, a fair amount is supposed about the spread of Tai speakers from southern China, west and south-west across mountains and along rivers in search of lowland river flats for paddy cultivation of rice. This often involved displacement of Tibeto-Burman and Mon-Khmer speakers, and was also often accompanied by the cultural/linguistic transformation of those non-Tai speakers (cf. Leach 1964, Condominas 1990). However, available coarse-grained descriptions of social history do not provide sufficient detail to account for the complex and context-dependent variables guiding speakers' choices about linguistic behaviour, ultimately determining the speech of their modern descendants.

The problem lies in the fact that linguistic change is necessarily and primarily a ground-level social process, the relevant mechanisms pivoting on identities, judgements, actions, and responses of individual speakers in real time. Speakers can detect when speech in their community begins to sound different (phonologically and grammatically), and these differences carry social significance, in the classical sociolinguistic sense (for example indicative of a speaker's age or background). Evidence of such details in the history of South-East Asia over the last two or three millennia is difficult, if possible at all, to find. One thing which must be ascertained in every case is the identifying value of particular linguistic choices in particular contexts, and this can apply to phonological choices, lexical choices, and grammatical choices (as for example in the case of Kmhmu, whose speakers may choose between a Thai-style periphrastic causative and a native morphological causative; cf. examples (9–10), above). Adoption of novel fashion in linguistic practice publicly advertises one's identification with others who adopt the same fashion (Le Page and Tabouret-Keller 1985), and one reason why great caution is needed in reconstruction of social conditions is that this identifying power of linguistic form goes beyond simplistic provision of absolute 'badges' or 'emblems' of imagined cultural, racial, or linguistic group membership.

A modern example from the context of this study is Ho, a South-Western Mandarin language spoken in the far north of Laos (Phongsaly and Oudom Xay provinces). Ho people are descendants of Chinese, associated with China by regular

contact with Chinese nationals, on both sides of the border. Even those who have never travelled to China are exposed to Modern Standard Chinese through electronic media, and through personal contact with travelling Chinese. The identifying value of Ho language is relative to context. In some contexts, it may signify and assert that one is 'Chinese' (in 'race') as opposed to 'Lao'. In China, it may signify and assert that one is 'south-western Chinese' (in terms of geographical affinity), as defined against other Chinese. Further, to speak Ho (a language spoken in Laos, as opposed to other Sinitic varieties spoken within the borders of China) may identify one as *non-Chinese* (in nationality). Thus, speaking Ho can signify 'being Chinese', 'being south-western Chinese', or 'not being Chinese', in different senses, and in different contexts, providing a Ho speaker with competing motivations for deciding when and if to use Ho at all. So-called linguistic 'emblems' must not be considered absolute, one-dimensional, and/or binary parameters in contact-induced change (cf. Milroy 1987). Different loyalties can be simultaneously maintained, different norms enforced.

While there is no space in this chapter to explore the relationship between social history and the contemporary linguistic situation in MSEA, any such endeavour will have to be undertaken with reference to explicit and plausible models of the ground-level social dimension of 'language contact' and change (cf. Thomason and Kaufman (1988) for some ideas; Ross (1997) and Enfield (2003) provide more explicit outlines), and will have to be genuinely informed by the findings of social anthropology, and especially sociolinguistics. Our most urgent requirement is empirically based and fine-grained multi-disciplinary research on grammar in *living cases* of speaker contact, since it is so difficult to *reconstruct* in sufficient detail the ethnography of inter-group communication.

2. Case study: polyfunctionality of ACQUIRE in Mainland South-East Asia

Most languages of MSEA have one verb-like morpheme which shows a strikingly similar and overlapping range of lexical and grammatical functions: a transitive verb 'come to have'; a preverbal modal/aspectual marker (typically 'get to', or 'have to'); a postverbal modal/aspectual marker (typically 'potential' or 'completive'); a marker of complex descriptive complement constructions such as resultative, adverbial, and potential expressions. I refer to this element as ACQUIRE.⁴

While in some languages, another verb may have the *basic* meaning 'come to have, acquire', the relevant item ACQUIRE both (a) has *some* meaning 'acquire', even if restricted, and (b) displays the basic range of secondary (both postverbal and

⁴ Unfortunately, English *acquire* does not reflect the basic, everyday nature of the verb in these languages, which is more like *get* with the non-agentive/non-controlled sense in *He got a parcel in the mail*. However, as a gloss 'get' is misleading in that it also has the agentive/controlled sense in *He carefully got a parcel out of the mailbox*. ACQUIRE in these languages *never* has this agentive/controlled sense. As a main verb of acquisition, it means 'come to have'.

preverbal) grammatical functions just described. Clark (1989) and Matisoff (1991) have briefly treated the issue in an areal perspective, Clark focusing on Hmong, Matisoff on Lahu (cf. also Bisang 1991). Elsewhere (Enfield 2003), I provide a detailed survey of the functions of ACQUIRE across a number of MSEA languages. In this section we look briefly at just four languages, two Sinitic and two Tai—Modern Standard Chinese, Cantonese, Northern Zhuang, and Lao.

2.1. MAIN TRANSITIVE VERB ‘ACQUIRE’

Modern Standard Chinese *dé* and Cantonese *dāk*, despite being usually glossed ‘get, obtain, gain, acquire’, do not normally appear as a main verb meaning ‘acquire’, but their historical source as ‘acquire’ is well established, and dictionaries invariably give these ‘acquisition’ glosses as primary meanings:

- (23) *sān sān dé jiǔ*
 MSC three three come.to.have nine
 ‘Three threes are nine.’

- (24) *tā dé bìng le*
 MSC 3 come.to.have illness CRS
 ‘S/he (has) got an illness.’

- (25) *nī douh hei dāk gwō chéuhng*
 CA this CL film come.to.have EXP prize
 ‘This film has won a prize.’

Northern Zhuang *dáy* and Lao *dāj* as main verbs are normal with the meaning ‘come to have’:

- (26) *kǔ dáy sǒng tuà, tē dáy sāam tuà*
 NZH 1 come.to.have two CL he come.to.have three CL
 ‘I got two and he got three.’

- (27) *phùu-nân pěn phùu dāj khāng dāj khǒng*
 LAO person-that be person come.to.have stuff come.to.have things
 ‘That person is the one who’ll get many things.’ (L1: 82)

2.2. ACQUIRE IN POSTVERBAL POSITION

Postverbal ACQUIRE has meanings associated with both ‘possibility’ and ‘achievement’. First, in the sense of ‘possibility’, the following examples from Lao, Cantonese, and Modern Standard Chinese show postverbal ACQUIRE as ‘can’:

- (28) *‘ān bō dāj*
 LAO read NEG can
 ‘He couldn’t read it.’ (L1: 49)

- (29) *hǎw bɔ̌ pāj nám khǎw ka bɔ̌ dāj*
 LAO 1 NEG go accompany 3 FP NEG can
 'I couldn't not go with them.' (L1: 658)
- (30) *m̄ go léuihjái hóu dá dāk ga*
 CA this CL girl very fight can PCL
 'This girl really knows how to fight.' (Matthews and Yip 1994: 242)
- (31) *jáu dāk ge lak*
 CA leave can PCL PCL
 '(We) can leave now.' (Matthews and Yip 1994: 242)
- (32) *yào bu dé*
 MSC want NEG can
 'cannot be wanted, undesirable' (Chao 1968: 453)
- (33) *shě dé*
 MSC abandon can
 'willing to give (something) up; can do without (something)'

The following Northern Zhuang and Lao examples illustrate an ambiguity of postverbal ACQUIRE, meaning either 'can', or signalling 'achievement' in a more finite context:

- (34) *náa thǎj dāj lɛ̃w*
 LAO paddy.field plough can PFV
 (i) '(This) field can be ploughed.'
 (ii) '(This) field has been ploughed.'
- (35) *nàa ɕwǎy dáy lo*
 NZH paddy.field plough can PCL
 (i) '(This) field can be ploughed.'
 (ii) '(This) field has been ploughed.'

The realis or 'achievement' readings in (34ii) and (35ii) are secondary, emerging pragmatically from literal assertion of 'possibility' in particular tense/aspect contexts. Observe the same alternation in these English examples:

- (36) *They were able to rescue only two of the children.* (Implies that they did.)
 (37) *I can smoke whole cigars without coughing.* (Implies that I do.)

Distinct from this 'achievement' interpretation, postverbal ACQUIRE may refer to a more complex notion of 'success' in the activity described in V_1 . Given 'acquire' as a simple meaning for ACQUIRE, examples like the following can be regarded as V_1 - V_2 resultatives ('V-and-acquire'), providing bridging contexts in which 'acquisition' and 'success' refer to the same sub-event (i.e. in which 'getting' something is what makes the said event successful):

- (38) *aũ t̄āai t̄ák dáy h̄au-l̄āai pȳă*
 NZH uncle elder catch come.to.have/succeed many fish
 ‘Great Uncle has caught a lot of fish.’

- (39) *mán h̄ăa p̄əm h̄ŭa n̄ân d̄āj* (l̄ēw)
 LAO 3 seek book CL that come.to.have/succeed PFV
 ‘He has found that book.’

In events described by verbs such as *nám* ‘pursue’, *h̄ăa* ‘seek’, and *cáp* ‘grab’, ‘acquisition’ and ‘success’ are contextually synonymous. With an acquisition verb in V_1 position, ACQUIRE as V_2 entails both ‘coming to have’ something and ‘succeeding’ in the V_1 task.

The semantics of postverbal ACQUIRE may then generalize in favour of this ‘succeed’ sense, becoming compatible with V_1 verbs which do not necessarily entail literal ‘acquisition’:

- (40) *s̄əp n̄āk-thám d̄āj*
 LAO be.examined AGT-dharma succeed
 ‘(I) passed my tests as a graduate in the dharma.’ (L1: 322) [‘possibility’ reading: ‘I am able to sit my tests as a graduate in the dharma.’]

- (41) *fáng b̄ə d̄āj*
 LAO listen NEG succeed
 ‘(It) can’t be understood/heard.’ (‘(One) can’t get a successful result from listening to (it).’) (L1: 52) [‘possibility’ reading: ‘... can’t listen to (it).’]

In this way, a postverbal ‘success’ function is established for ACQUIRE, derived originally from its main verb ‘acquire’ meaning, involving a resultative role. This two-step process is illustrated here:

- (42) 1. Simple resultative V_2 , ACQUIRE as ‘acquire’:
 $V_{ACQUISITION} + ACQUIRE$ ‘V-and-acquire something’ (entails ‘V-and-succeed’, given that the objective of V_1 is to acquire something)
 2. Meaning generalizes to ‘succeed’, ‘V’ slot opens to greater range of verbs:
 $\rightarrow V + ACQUIRE$ ‘V-and-succeed’

Now, a subsequent step, from ‘success’ to ‘possibility’, is enabled by a regular pragmatic property of V_1 - V_2 resultative constructions (associated with a high level of context-dependency in interpretation of interclausal relationships in these languages). Let us consider how it works.

The following V_1 - V_2 resultative constructions in Lao (same- and different-subject, respectively) have two interpretations, depending on whether the predicated V_2 ‘result’ is understood as a finite event (‘it is true that V_1 resulted in V_2 on a given occasion’), or less finitely, as habitual or potential (‘if/when/whenever V_1 is the case, V_2 results’):

(43) *nɛ̌m bɔ̌ thɔ̌ŋ*

LAO reach.for NEG reach

(i) '(I) didn't reach (it).' (on a given occasion, (I) reached for (it) and didn't reach (it))

(ii) '(I) can't reach it.' (if/when (I) reach for (it), (I) don't reach (it))

(44) *khɔ̌ɔj sāj pɔ̌n kabɔ̌ɔk n̄i n̄iŋ to-sāang tǎaj*

LAO 1 use gun CL this shoot CL-elephant die

(i) 'I shot an elephant dead with this gun.' (on a given occasion, I shot an elephant with this gun and it died)

(ii) 'I can shoot an elephant dead with this gun.' (if/when I were to shoot an elephant with this gun, it would die)

With the 'potential'/'possibility' interpretation in (43ii) and (44ii), postverbal ACQUIRE as resultative V_2 'succeed' thus expresses the most widely applicable sense of 'potential success', namely 'can' (i.e. 'if/when someone Vs, they succeed').⁵ This step may be added to the two steps described in (42):

(45) 1. Simple resultative V_2 , ACQUIRE as 'acquire': $V_{ACQUISITION}$ + ACQUIRE 'V-and-acquire something' (entails 'V-and-succeed', given that the objective of V_1 is to acquire something)

2. Meaning generalizes to 'succeed', 'V' slot opens to greater range of verbs:

→ V + ACQUIRE 'V-and-succeed'3. In non-finite contexts, resultative ' V_1 - V_2 ' is interpreted as 'can V_1 -and- V_2 '→ V + ACQUIRE 'can V-and-succeed', → 'can V'

Now, once this 'can' meaning for postverbal ACQUIRE is established as a distinct meaning, recall that by a different pragmatic inference (cf. (34–7), above), it can give an 'achievement' meaning (sometimes very close to the 'succeed' meaning).

The point of this more detailed discussion of semantic/pragmatic alternation for Lao postverbal ACQUIRE has been to show that (a) different pragmatic forces can encourage interpretations in more than one direction (i.e. from 'succeed' to 'can', and from 'can' to 'achievement'), and (b) typological features can encourage/account for such shifts. In this case, two areally widespread features—namely, scarce formal specification of dependency relationships among grammatically associated predicates (with corresponding high context-dependency in their interpretation), and V_1 - V_2 resultative constructions—combine to give 'potential result' readings for the V_1 - V_2 resultative strings (see further discussion, below).

⁵ This simplified proposal requires refinement, in particular to account for a distinction between two kinds of 'potential success' which arise with certain telic verbs (e.g. 'intentional object' verbs; Quine 1960: 219 ff.). For example, in most of these languages, 'seek' marked by postverbal ACQUIRE may mean 'can seek' or 'can find' (cf. similar examples (40–1), above). The emergence of a simple 'can' meaning out of 'potential success' would most likely have emerged through combination with verbs which entail their own result (e.g. 'kill'), and/or verbs of simpler semantic structure (although the path suggested here begins with semantically more complex verbs such as 'seek'). See Enfield (2003) for details.

In Cantonese, as in Lao and Northern Zhuang, we find similar interaction between the semantics of ‘possibility’/‘potential’ and ‘success’/‘achievement’, associated with postverbal ACQUIRE. The following means ‘did you *successfully* sit your exam?’ (i.e. ‘did you sit your exam and get a result?’; cf. (40–1), above), and not simply ‘were you able to sit your exam?’:

- (46) *léih háau-síh dāk-m-dāk a*
 CA 2 take-exam succeed-NEG-succeed PLC(Q)
 ‘Was your exam okay? [i.e. Did you pass?]' (Matthews and Yip 1994: 243)

There are traces in both Cantonese and MSC of a ‘success’ meaning of postverbal ACQUIRE in combination with certain other verbs. The following example shows postverbal *dé/dāk* in the usual idiom for ‘remember’:

- (47) *tā jì de (zhù)*
 MSC 3 remember succeed (be.placed)
 ‘S/he remembers (it).’
- (48) *léih m-gei-dāk-jó àh*
 CA 2 NEG-remember-succeed-PFV PCL(Q)
 ‘Have you forgotten?’ (Matthews and Yip 1994: 33)

There is less conclusive synchronic evidence in modern Sinitic of the semantic/pragmatic relationships described for postverbal ACQUIRE in Lao, especially since the main verb functions of ACQUIRE (e.g. as ‘acquire’ or ‘succeed’) are more restricted. However, in line with the pattern of development I have suggested for Lao here, Lamarre (2001) argues on the basis of synchronic comparative evidence that postverbal ACQUIRE in Sinitic became a marker of ‘success’ or ‘realization’ before its development into a marker of ‘potential’.

2.3. ACQUIRE IN POSTVERBAL DESCRIPTIVE COMPLEMENTS

A well-documented function of Modern Standard Chinese *dé* ACQUIRE and etymons in other Sinitic languages (such as Cantonese *dāk*, or Taiwanese Southern Min *tit*; Lien 1997) is its appearance in a class of complex postverbal descriptive complement constructions. Focusing only on MSC in this section (for details on similar patterns in Cantonese, see Matthews and Yip 1994), I make the distinctions set out in Table 3 for the purpose of this discussion.

Following are examples of the first type of construction, Manner (both examples from Li and Thompson 1981: 624):

- (49) *tā zǒu de hěn màn*
 MSC 3 walk MC very slow
 ‘S/he walks very slowly.’
- (50) *tā chuān de hěn piàoliang*
 MSC 3 dress MC very beautiful
 ‘S/he dressed very beautifully.’

TABLE 3. Types of 'V *de* COMP' construction in Modern Standard Chinese

Construction	Form	Meaning	Pattern of negation
1. Manner (MC)	$v_{1[\text{active}]}$ <i>de</i> $\dot{v}_{2[\text{stative}]}$	' v_1 in a v_2 manner'	v_1 (<i>de</i>) 'neg' $\dot{v}_2$ ' v_1 not in a v_2 manner'
2. Potential Manner (PMC)	$\dot{v}_{1[\text{active}]}$ <i>de</i> $v_{2[\text{bare stative}]}$	'can v_1 in a v_2 manner'	$\dot{v}_1$ 'neg' v_2 'cannot v_1 in a v_2 manner'
3. Potential Result (PRC)	$v_{1[\text{active}]}$ <i>de</i> $v_{2[\text{non-stative}]}$ (O) ($[v_1-v_2(O)]$ is <i>resultative</i>)	'can [v_1 and v_2 (O)]' ($[v_1-v_2(O)]$ is <i>resultative</i>)	v_1 'neg' v_2 (O) 'cannot [v_1 and v_2 (O)]'
4. Extent (EC)	$v_{1[\text{active}]}$ <i>de</i> v_2/S	' v_1 until v_2/S ; 'So v_1 that v_2/S '	negation internal to v_2/S 'so v_1 that not v_2/S '

The grave accent signifies stress.

The second type—Potential Manner—shows a familiar ambiguity (cf. §2.2) between 'potential' and 'realized':

- (51) *tā pǎo de kuài*
 MSC 3 run PMC fast
 (i) 'He can run fast.' [potential manner]
 (ii) 'He runs/is running fast.' [manner]

The third construction type shown in Table 3 requires that the verbs involved be in a resultative relationship, such as in the following, with V_1 *tiào* 'jump' and V_2 *guò-qu* 'go across':

- (52) *tā tiào guò-qu le*
 MSC 3 jump cross-go PFV/CRS
 'S/he (has) jumped across.' (Li and Thompson 1981: 55)

Insertion of *dé* between V_1 and V_2 here gives rise to the Potential Result construction, by which 'the action or process denoted by the first constituent of the compound *can* have the result denoted by the second constituent of the compound' (Li and Thompson 1981: 56):

- (53) *tā tiào de guò-qu*
 MSC 3 jump PRC cross-go
 'S/he can jump across.' (Li and Thompson 1981: 56)

- (54) *tā xǐ de gānjīng nèi ge xiāngzi*
 MSC 3 wash PRC clean that CL chest
 'S/he can wash that chest clean.' (Li and Thompson 1981: 477)

In the Extent complement construction (Type 4 in Table 3), 'the event in the first clause is done to such an extent that the result is the state expressed by the

stative clause or verb phrase' (Li and Thompson 1981: 626; both examples from same):

- (55) *tā xiào de zhàn bu qǐ lái*
 MSC 3 laugh EC stand NEG rise come
 'S/he laughed so much that she couldn't stand up.'

- (56) *tā jiāo de lèi le*
 MSC 3 teach EC tired CRS
 'S/he taught so much that s/he is tired.'

Tai languages do not display closely parallel patterning of these postverbal complement construction types, which are evidently more grammaticalized in Sinitic languages. Lao and Northern Zhuang do have adverbial constructions of Type 1 (in Table 3) and similar, in which postverbal ACQUIRE takes a stative verb complement. Here are some Lao examples:

- (57) *hāw hēt dāj nòj tam-tām*
 LAO 1 do/make MC small low-RDP
 'I made it small, quite low.' (L1: 90)

- (58) *cá' 'òk-mēē-phēē-lūuk dāj dīi nō'*
 LAO IRR give.birth.to-mother-propagate-child MC good PCL
 'They'll breed well, won't they.' (L1: 26)

Lao and Northern Zhuang also allow nominal complements in these adverbial constructions, such as the following temporal complement expressions marked by ACQUIRE:

- (59) *kǔ 'yū ki-nī dáy çip pǐ*
 NZH 1 live here TC ten year
 'I've lived here for ten years.'

- (60) *tɕon nân khòj pǎj nɕɔŋ-kháaj dāj sǎam dɕan*
 LAO time that 1 go N. K. TC three month
 'At that time, I'd been in Nong Khai for three months.' (L1: 596)

This type of temporal complement construction exists in MSC and Cantonese, but does not involve ACQUIRE (see §3.1 below, for discussion).

The Potential Manner construction (Type 2 in Table 3) is available in Lao and Northern Zhuang, resulting straightforwardly, as in the following Lao example, from the role of postverbal *dāj* ACQUIRE as 'can':

- (61) *mán lēēn dāj vāj*
 LAO 3 run can fast
 'S/he can run fast.'

Neither Potential Result nor Extent constructions (Types 3 and 4 in Table 3) marked by ACQUIRE in Sinitic are available in Lao and Northern Zhuang. However,

there are identical constructions to these involving markers other than ACQUIRE (see §3.1 below, for discussion).

2.4. ACQUIRE AS A PREVERBAL MARKER

Lao *dāj* ACQUIRE often directly precedes a main lexical verb, giving an aspectual/modal meaning translated in a range of ways—‘get to’, ‘have to’, ‘happen to’, ‘did’. The following example illustrates different context-dependent interpretations (assuming a past-tense context):

- (62) *kǔu dāj nāaj h^han*
 LAO 1 RPE move house
 (i) ‘I got to move house.’
 (ii) ‘I had to move house.’

The invariant meaning here is that ‘the main verb is the case because of something else that has happened before it’ (thus the gloss ‘(R)esult of (P)rior (E)vent’). Example (62) literally means ‘I moved house; this was because something else happened before this’, for which there is no direct translation equivalent in English. (A vaguely helpful rendition could be *It happened that I moved house.*) In specific contexts this meaning results in narrower interpretations. The ‘got to’ interpretation in (62i) would emerge if someone had been *given permission* to move, while ‘have to’ in (62ii) would emerge if someone had been *ordered* to move. The latter reading is normal for ACQUIRE as a modal in Tibeto-Burman languages like Burmese and Lahu; Okell (1969), Matisoff (1973) (in contrast to the ‘fortuitous’ or ‘benefactive passive’ usage of preverbal ACQUIRE in Vietnamese; Thompson 1987: 229).

Preverbally, Northern Zhuang *dáy* has a similar ‘resultant’ interpretation, typically expressed in translation by the likes of ‘manage to’, ‘get to’:

- (63) *pó-nāa dáy kǎn pyōng tōy hōu-sōu*
 NZH uncle RPE eat half bowl porridge
 ‘Uncle got the chance to have half a bowl of porridge.’
 (64) *tāai dáy hòt kyí çòn*
 NZH grandma RPE speak several words
 ‘Grandma managed to speak a few words.’

This more complex modal meaning is sometimes weakened, especially under negation, producing a kind of ‘assertive’ or realis expression (often translated into English using emphatic *do*):

- (65) *té bōu dáy pǎy*
 NZH 3 NEG RLS go
 ‘He didn’t go.’

- (66) *láaw bɔ̌ dâj bəŋg*
 LAO 3 NEG RLS look
 'He didn't (get to) look at them.' (L1: 41)

- (67) *láaw dâj sɔ̌n vɛ̌n-tǎa nám cék*
 LAO 3 RLS buy spectacles accompany Chinaman
 'He did buy the spectacles from the Chinaman.' (L1: 55)

Turning to Sinitic, MSC *dé* does not have this preverbal aspect/modality function. Note, however, the preverbal modal, *děi* 'should, must', which is written with the same character as *dé*, and which is widely presumed to be cognate:

- (68) *wǒ děi zǒu le*
 MSC 1 must/should walk CRS
 'I must/should go now.'

No such preverbal usage of Cantonese *dāk* is attested.

2.5. COMPARATIVE DATA FROM SOUTH-WESTERN MANDARIN

Let us now consider data from South-Western Mandarin, a dialect chain mostly spoken in Yunnan, China (in parts alongside Tai languages such as Shan/Tai-Lue and Lao). South-Western Mandarin is of interest due to its divided typological affiliation in the Sinitic family, between Modern Standard Chinese (with whom it also falls in terms of genetic grouping), and geographically more proximate languages (Sinitic and non-Sinitic) of peninsular MSEA.

The more common productive pattern for expressing 'can' in South-Western Mandarin uses postverbal *de* ACQUIRE (exactly as in Lao and Northern Zhuang), rather than preverbal *néng* as is usual in Modern Standard Chinese:

- (69) *nǐ bù néng zuò*
 MSC 2 NEG can do
 'You can't do it.'

- (70) *nǐ zuò bù de*
 SWM 2 do NEG can
 'You can't do it.'

Questions are normally formed in Northern Sinitic languages with a *p-not-p* construction:

- (71) *nǐ mài-bu-mài nǐ de zì-xíng-chē*
 MSC 2 sell-NEG-sell 2 POSS bicycle
 'Are you selling your bicycle (or not)?'

In Modern Standard Chinese, it is the preverbal modal *néng* 'can' which is targeted in question forms (and given as an affirmative answer), rather than the content verb, which follows:

(72) *nǐ néng-bu-néng qù*

MSC 2 can-NEG-can go

'Can you go (or not)?'

The affirmative answer is *néng (qù)* '(I) can (go)'. In South-Western Mandarin, the first verbal element may be repeated—as in Modern Standard Chinese—but this is the content verb rather than the modal:⁶

(73) *nǐ kě-bu-kě de*

SWM 2 go-NEG-go can

'Can you go or not?'

Postverbal *de* 'can' in South-Western Mandarin displays less of the 'full verb' or clausal head trappings than Modern Standard Chinese's preverbal *néng* 'can'. The affirmative answer to (73) is *kě de* '(Yes, I) can go', and *de* cannot appear alone as a yes-answer.

In Cantonese, as in Modern Standard Chinese, it is the modal which is more main-verb-like, since it is usually the target of *p-not-p* question formation, and may appear alone as an affirmative answer. However, in contrast to Modern Standard Chinese, the relative order of modal and content verb is reversed:

(74) Q: *léih hui dāk-m-dāk*

CA 2 go can-NEG-can

'Can you go?'

A: (*hui*) *dāk*

(go) can

'(Yes, I) can (go).'

While South-Western Mandarin falls within the same low-level branch of Sinitic (i.e. Mandarin) as Modern Standard Chinese, the grammatical differences correspond in part to geographical proximity with languages from outside this grouping. South-Western Mandarin is like closely neighbouring languages in some ways, and like languages which belong to its own 'family' in other ways. This is manifest in contrasting grammatical behaviour associated with postverbal ACQUIRE as a modal.

2.6. SUMMARY

Table 4 summarizes the extent of overlap of some functions of ACQUIRE in the sample languages.

We may now consider how these findings relate to the question of 'genetic' versus areal relatedness in languages of MSEA.

⁶ Other patterns of question formation, including use of a preverbal interrogative marker, are also found in SWM. There is evidently significant grammatical variation among SWM 'dialects'.

TABLE 4. Functions of ACQUIRE in five Mainland South-East Asian languages

Function	Lao (<i>dāj</i>)	Northern Zhuang (<i>dáy</i>)	Cantonese (<i>dāk</i>)	South- Western Mandarin (<i>de</i>)	Modern Standard Chinese (<i>dé</i>)
(i) main verb 'acquire'	+	+	+	+	+
(ii) preverbal modal 'get to'/'must'	+	+	—	+	+
(iii) postverbal modal marking	+	+	+	+	+
(iv) temporal adverbial complement (<i>'V_{ACQUIRE} t' = 'V has been the case for t'</i>)	+	+	—	—	—
(v) extent complement (<i>'V_{1 ACQUIRE} V₂' = 'So V₁ that V₂'</i>)	—	—	+	+	+
(vi) potential result complement (<i>'V_{1 ACQUIRE} V₂' = 'Can V₁-and-V₂'</i>)	—	—	+	+	+
(vii) manner complement (<i>'V_{1 ACQUIRE} V₂' = 'V₁ in a V₂ way'</i>)	+	+	+	+	+

3. Discussion

The complex pattern of grammatical behaviour surrounding ACQUIRE, sketched for just a few languages in §2, is uncannily replicated in dozens of other MSEA languages, also including languages of the Hmong-Mien and Mon-Khmer families (see Enfield (2003) for detailed treatment). What could explain the distribution of these complex grammatical patterns throughout MSEA? We may first consider historical evidence, where data is available. Second, we may expand our base for comparison, and look at further possible synchronic instantiations of the polyfunctionality of 'acquire'. Third, we may consider the possibility of common language-internal motivations for the emergence of patterns like the ones described.

3.1. HISTORICAL EVIDENCE FROM SINITIC AND TAI

Li (1977: 108, 285) reconstructs **ʔdai* (ACQUIRE) for Proto-Tai, probably spoken around the same time as Old Chinese (500 BC–AD 200 or after; Sun (1996: 3)). In Sinitic languages at this time, according to Sun (1996), most of the relevant functional extensions of *dé* ACQUIRE had not yet developed. Leading scholars in comparative and historical Tai linguistics (Tony Diller, Jerry Edmondson, and Luo Yongxian in personal communication, Bangkok, July 1998) agree that the Tai lexical item itself

(*ʔdai and descendants) could conceivably have been a borrowing from Sinitic, but none argue that it definitely is or is not. And Sinitic scholars do not say that the word originally came into Sinitic from Tai, but again this is conceivable. As it is, arguments for etymological relatedness of the ACQUIRE words in Tai and Sinitic are tenuous. Proto-Sinitic *tak has a voiceless unaspirated initial, whose counterpart in modern Tai is normally a voiceless aspirated stop (*th-*), not the voiced stop (*d-*) found in Lao *dāj*, Northern Zhuang *dáy*, and elsewhere (or the voiced lateral (*l-*), as in Shan *lài*; cf. Dong *lǐ*³²³). Further, while the Proto-Sinitic form has a final stop, the Tai forms do not.⁷ In the absence of further evidence, a fair conclusion is that *tak and *ʔdai are etymologically unrelated, and the close parallelism in function of the modern morphemes suggests long-term development of the *functional application* of the morpheme in each language family, either separately, or through borrowing of the semantic/grammatical ideas through contact. But even if *tak and *ʔdai were etymologically related, the parallel functional patterns are *just as likely* to be the result of this kind of separate and parallel development (possibly encouraged by diffusion), since at the time the lexical borrowing would have to have taken place (i.e. by the time of Proto-Sinitic and Proto-Tai), the various grammatical functions were hardly developed (at least for Sinitic; Sun 1996). In such a scenario, both genetic *and* areal factors would contribute to the widespread occurrence of a complex semantic and grammatical pattern.

The point is that *functions* may be duplicated closely without duplication of, or reference to, phonological form (i.e. by calquing). Sometimes particular functions are performed by similar structures in neighbouring languages, but the lexical material recruited to mark the structure is not the same. Recall the non-overlapping range of functions of ACQUIRE as head of a temporal adverbial complement, and extent adverbial complement, in Tai, and Sinitic, respectively, as shown in Table 4 (an extract of which is reproduced as Table 4a).

Sinitic languages use ACQUIRE as a complement head in expressions meaning 'VP₁ to such an extent that VP₂' (examples (55, 56) above), while Lao and Northern Zhuang do not. However, Lao and Northern Zhuang have structurally identical expressions which use 'until' where Sinitic languages use ACQUIRE:

- (75) *tě káng tâng pàak nài pǎy*
 NZH 3 speak until mouth tired go
 'He spoke so much his mouth got tired.'

- (76) *mán hēt sǎng dǎng cǒn phúak háw nǒn-bɔ-láp lǎj*
 LAO 3 make sound loud until group 1 lie-NEG-sleep at.all
 'S/he made such a racket we couldn't get to sleep at all.'

⁷ As Tony Diller points out in personal communication, if the original Sinitic form had a palatal stop final *-c, this would make an etymological relationship to Proto-Tai *-j more plausible. While there is no *phonemic* final stop in modern Tai reflexes of *ʔdai, the final -j is in various dialects consistently accompanied by a glottal stop; e.g. Stung Treng Lao (north-east Cambodia) [dajʔ⁵¹] 'ACQUIRE', Thai Neua (Laos) [lajʔ⁵¹] 'ACQUIRE'. (Data from fieldnotes.)

TABLE 4a. Extract from Table 4

Function	Lao (<i>dāj</i>)	Northern Zhuang (<i>dáy</i>)	Cantonese (<i>dāk</i>)	South- Western Mandarin (<i>de</i>)	Modern Standard Chinese (<i>dé</i>)
(iv) temporal adverbial complement (‘V ACQUIRE t’ = ‘V has been the case for t’)	+	+	–	–	–
(v) extent complement (‘V ₁ ACQUIRE V ₂ ’ = ‘So V ₁ that V ₂ ’)	–	–	+	+	+

Conversely, while Lao and Northern Zhuang use ACQUIRE in heading temporal complements (‘V for a period of t’; cf. examples (59, 60) above), Sinitic languages do not. However, the latter display an identical construction, as in the following Cantonese and Modern Standard Chinese examples, with a perfective particle in place of ACQUIRE (compare the Lao example (79)):

(77) *ngóh ga chē jā jó léuhng lihn géi*
CA 1 CL vehicle drive PFV two year some
‘I’ve been driving the car for over two years.’ (Matthews and Yip 1994: 205)

(78) *zhù le wú nián*
MSC live PFV five year
‘lived (there) for five years’

(79) *jūu dāj haa pī*
LAO live TC five year
‘lived (there) for five years’

3.2. COMPARATIVE EVIDENCE FROM EASTERN MON-KHMER

Table 5 shows forms and functions of ACQUIRE in nine Eastern Mon-Khmer languages, two of the Vietic branch (Vietnamese, Muong), three Katuic (Ngae, Pacoh, Katang), one Khmeric (Khmer), and three Bahnaric (Taliang, Alak, Brao).

The only forms known to be cognate among this set are Katuic [ɕʰeən] (Ngae), [ɕoon] (Pacoh), and [ɕəən] (Katang). Despite apparent formal similarity to these, Khmer [ɕaan] is unlikely to be etymologically related. Table 5 shows that in a relatively small geographical area (see the cluster of these languages in the south-eastern corner of Laos on Map 1), among languages of a single sub-branch of one among many neighbouring language groups, there are as many as seven separate etymons instantiating the complex pattern of ACQUIRE described above. Furthermore, none of these are likely to be related to Sinitic **tak* or Tai **ɬai*. With further forms realizing the pattern in MSEA (e.g. in Hmong-Mien) there are

TABLE 5. Functions of ACQUIRE in some Eastern Mon-Khmer languages

<i>Functions of ACQUIRE</i>	Vietnamese <i>dʉəŋk</i> ²¹	Muong <i>an</i> ²¹³	Ngae <i>ŋeən</i>	Pacoh <i>boon</i>	Katang <i>bəən</i>	Khmer <i>baan</i>	Taliang <i>bəəc</i>	Alak <i>dʉj</i>	Brao <i>dɔu</i>
(i) Main-verb ‘acquire’, non-agentive	1	1	1	1	1	1	1	1	1
(ii) Preverbal aspectual/modal	1	1	1	1	1	1	1	1	1
(iii) Postverb ‘with success’	1	1	1	%	1	1	1	1	1
(iv) Postverb ‘can’	1	1	1	0	1	1	1	1	1
(v) Marks postverbal descriptive complement	1	1	1	1	1	1	1	1	1

% signifies ‘indeterminate’.

thus over ten distinct etymons displaying the areally highly consistent semantic/grammatical pattern of ACQUIRE described in this study.

These data from Eastern Mon-Khmer strikingly demonstrate how languages which are both areally and genetically related can produce complex and near-identical patterns of grammatical polyfunctionality associated with lexical items which are from historically different sources (i.e. not 'the same word'). The borrowing of *ideas* for linguistic organization (without borrowing the attached phonological material) can include the whole polyfunctional potential of a particular semantic item, and not just one or another of its functional extensions. Thus, even if exponents of ACQUIRE in two languages may ultimately be cognate, their shared repertoire of grammatical and semantic functions (as opposed to their phonological form) is not necessarily due to this common 'genetic' origin.

In Tai languages, the pattern of ACQUIRE is associated with a single etymon, and it is known that Tai speakers moved relatively recently to the areas where Eastern Mon-Khmer languages are now spoken. The diversity in form of ACQUIRE in the latter languages suggests that at a stage when many of today's Eastern Mon-Khmer languages had already separated, speakers of these languages encountered and widely emulated a fashion of speech already long in vogue among Tai speakers, influential newcomers to the area.

3.3. SYSTEM-INTERNAL SOURCES FOR THE COMMON INNOVATIONS

Discussion so far has concerned lexical items and the sharedness of their functional behaviour among languages, due either to borrowing, or to common 'genetic' inheritance. Borrowing and inheritance can be regarded as external sources to a synchronic language system (i.e. an idiolect), since in both 'inheriting' and 'borrowing' an idea, a speaker relies on his social associates (be it his own kind or his neighbours) as sources. However, innovations can also be originated by creative individuals, and the creative imagination constitutes a synchronic system-internal source for innovation (which, if popular, becomes fashionable, takes hold, and eventually becomes common structural change; Durie and Ross 1996a: 15, Harris and Campbell 1995: 54, Ross 1997: 214–15). This is important in the present context because it reminds us that borrowing and inheritance do not provide the only common sources for separate linguistic systems.

Two issues which arise when considering the likelihood or possibility of system-internal innovations are, first, their conceptual naturalness, relating to constraints and propensities of human cognition and imagination, and second, the typological 'poise' of a linguistic system, i.e. how existing semantic/grammatical configurations may constrain or encourage stages of semantic and grammatical development. These concern degrees of naturalness both in a universal sense (relating to human cognition) and a relative sense (dependent on given typological configurations). I will argue in §3.3.2 that some functions of ACQUIRE (e.g. postverbal 'can') are distributed widely due to the corresponding distribution of a certain typological precondition (namely the area-wide availability of 'potential'

reading for resultative constructions), which enables common occurrence of the same semantic/functional development.

3.3.1. *Conceptual naturalness*

It is generally accepted that some functional extensions are cognitively more easily made than others, and indeed that some conceivable extensions are almost certainly made, while others almost certainly are not (cf. for example Hopper and Traugott (1993), Traugott and Heine (1991), Wilkins (1996), amongst others). The greater the naturalness of a semantic or structural extension, the more likely it is to occur in languages separately and spontaneously, with the effect that it may *appear* in retrospect as if borrowing/diffusion or common inheritance has occurred, when in fact none has—especially when the lexical items recruited for the extension in each case happen to be cognate (as may have been the case for ACQUIRE in Tai and Sinitic). Similarly, the greater the naturalness of a semantic or structural extension, the more readily the *idea* of making that extension may be *borrowed*. So, the presence in two languages of the same particularly ‘natural’ semantic extension does not help much in defining whether given words or grammatical elements are shared due to diffusion or inheritance, or indeed, coincidence.

It is intuitively easy to judge the conceptual naturalness of many lexical/idiomatic extensions, such that idiosyncratic expressions like ‘pig-crazy’ for ‘epileptic’ and ‘tooth-insect’ for ‘dental decay’ in many MSEA languages (Matisoff 1978: 70) seem less likely to be independently innovated than more globally attested expressions such as ‘foot’ for ‘tyre’ or ‘fire’ for ‘light’. So when we encounter relatively idiosyncratic semantic/grammatical extensions in languages of a single region, the likelihood that these have emerged coincidentally is low, and we may more readily suspect a non-coincidental relationship, such as borrowing/diffusion or common inheritance. But intuitions about semantic ‘naturalness’ are less forthcoming when it concerns the simpler or more abstract semantics of grammar, as for example with respect to various extensions from ‘acquire’, described above (i.e. to ‘possibility’, ‘success’, ‘necessity’, and so on).

In the present study, one pragmatic extension which seems conceptually ‘natural’, and likely to occur universally, is the inference from ‘possibility’ to ‘actuality’ in a finite context, by which *They were able to save two children*, implies, but does not entail, that they did (§2.2). The current study would benefit from further cross-linguistic work on the relevant semantic extensions of ‘acquire’, to gauge whether or not the functions examined here are typologically so ‘natural’ that their shared presence in the area is likely to be unremarkable, or even purely coincidental.

3.3.2. *Typological poise*

It has been claimed that structural diffusion is more likely to occur among languages which are already structurally similar (cf. references in Harris and

Campbell 1995: 123), or at least that syntactic borrowings should 'fit with innovation possibilities of the borrowing language' (Harris and Campbell 1995: 125). The result is a self-perpetuating process which gives rise to areal convergence, since structural borrowing or copying naturally increases the structural compatibility of the languages, thereby increasing the likelihood of further common structural borrowing or development, and so on.

The nature of a grammar sees a language 'poised' for particular semantic/grammatical developments, and less for others, determining the readiness or susceptibility of a language to realize a given extension. In at least this sense, speakers 'make do with what's historically presented to them' (Lass 1997: xviii). Similarly, if a language is *not* poised for a certain development, it may be less likely to occur. Note that the common poise of two neighbouring languages is logically independent of their areal or genetic relatedness, and it may be due to former contact, or to substratum interference. Also note that typological poise is more a measure of the likelihood of languages to independently make the same innovations, than of the likelihood of structural borrowing.

Let me illustrate with an example from this study. I have argued in §2.2 that the extension from a resultative V_2 'succeed' to 'can' is licensed by a combination of two typological features of MSEA languages: (a) a lack of overt marking of relations of subordination/dependency among grammatically associated predicates (e.g. verbs in series) with corresponding context-dependent openness in interpretation of those relations, and (b) resultative constructions of the form V_1 - V_2 .

V_1 - V_2 resultatives in many MSEA languages may be interpreted as either 'finite'/'realized', or 'non-finite'/'habitual'/'potential' (cf. (43-4), above):

- (80)
- | | | |
|--|--|--|
| $V_{1\text{cause/condition}}-V_{2\text{result}}$ |  | <p>(i) finite, on a given occasion
'V_1 and as a result V_2'</p> <p>(ii) non-finite, whenever
'can/would V_1 with the result that V_2'</p> |
|--|--|--|

As described for Lao in §2.2, when ACQUIRE as 'succeed' appears in resultative V_2 position, it similarly may be interpreted as non-finite/potential (80ii), just like any other resultative. Thus, V_1 -*'succeed'* may mean ' V_1 -and-succeed', or 'can/would V_1 -and-succeed'. 'Succeed' is the most general expression of 'result', and thus in this function occurs with the widest range of V_1 verbs. With this widest usage, under a non-finite/potential interpretation, postverbal resultative 'succeed' comes to have the simple meaning 'can', the most general expression of 'potential' (see note 5). A conceivable path is as follows (cf. (45), above):

- (81)
- (i) 'V-and-acquire' > (ii) 'V-and-succeed' > (iii) 'can V-and-succeed' > (iv) 'can V'

Assuming this order of development, we would expect that if a language had not made step (ii) (in which ACQUIRE generalizes, from 'success' as result of a verb of acquisition to 'success' more generally), then ACQUIRE in this language cannot

have made step (iv), to ‘can’. This is borne out by the only MSEA language I have found to lack the ‘can’ reading for postverbal ACQUIRE—namely, Pacoh (see the conspicuous anomaly on Table 5). Pacoh *boon* ACQUIRE as resultative V_2 expresses ‘success’ only by implication (as in (i)), as long as the aim of V_1 is to ‘acquire’ something. In these cases a non-finite/potential interpretation (available as a general property of unmarked V_1 - V_2 resultatives; cf. (80)) does allow a translation using ‘can’:

- (82) *kùù kòòp kuusên boon*
 PA 1 catch snake acquire
 (i) ‘I caught a snake/snakes.’
 (ii) ‘I can catch snakes.’

However, unlike its analogue in surrounding languages, *boon* ACQUIRE has apparently so far not taken the generalizing step (8iii), remaining inapplicable as a resultative V_2 whenever ‘success’ of V_1 does not involve ‘acquisition’. *Boon* has thus not appeared as resultative V_2 with a sufficiently broad range of V_1 verbs to have extended to simple ‘can’:

- (83) **pooq semuej (lêjq) boon*
 PA go S. (NEG) acquire
 ((You) can (not) go to Samoy.)

Pacoh speakers express ‘can’ using the postverbal modal *hooj*:

- (84) *pooq semuej (lêjq) hooj*
 PA go S. (NEG) can
 ‘(You) can (not) go to Samoy.’

In other MSEA languages examined, I assume step (8iii) has been taken, and the typological poise of these languages is what has allowed/encouraged the essential further steps (8iii–iv) in every case. Given the typological poise of Pacoh (specifically, its having V_1 - V_2 resultatives *combined with* its lack of an obligatory formal distinction between finite and non-finite readings of V_2 in a V_1 - V_2 string), there is no reason to think that it would not have done the same had it generalized a ‘success’ meaning for *boon* ACQUIRE (as in (8iii)).

It is thus implied that MSEA languages with rather different typological poise (such as verb-final Tibeto-Burman languages whose multiverb constructions are structured somewhat differently) may fail to realize certain functional/semantic developments such as those described here, or at least may fail to realize them in the same ways. In Burmese and Lahu, ACQUIRE does appear as both a main verb ‘acquire’, and as a verb-marking modal (see Okell (1969) on Burmese *yá* and Matisoff (1973) on Lahu *ǵa*), but it has been beyond the scope of this study to determine the nature and extent of the parallels in polyfunctionality of ACQUIRE with those languages. Further research is required to isolate the ‘poise effects’ I have suggested here, and test the hypothesis that similar typological poise can

provide an account for common structural development in languages. For the particular claim I have made concerning the grammatical development enabled by the combination of (a) non-marking and open interpretation of relations among verbs in series, and (b) V_1 - V_2 resultatives, one approach would be to isolate these two typological conditions in a sample of languages and look for evidence that it is indeed the combination of these two features which has the said effect.

In sum, it is important not to underestimate the significance of a language's typological poise, as a set of factors determining relative naturalness or likelihood of possible innovations. (Evolutionary biology may provide useful metaphors or even explanations here.) Conceptual 'naturalness' in linguistics is usually defined in terms of putative cognitive/conceptual universals (biologically based), and the parameter of typological poise introduces an idea that such 'naturalness' is often relative, or context-dependent.

4. Conclusion

The case study described in this work has demonstrated that the diachronic continuity of phonological forms and semantic/grammatical patterns associated with those forms can be separate matters altogether. The complex polyfunctionality of ACQUIRE is closely replicated among MSEA languages, yet is associated with a total of ten or more distinct etymons. We have compared Tai and Sinitic languages in particular, and have seen that close parallelism of function between Tai and Sinitic ACQUIRE is not accompanied by regular phonological correspondence. One might be tempted to overlook the imperfect correspondence between Sinitic **tak* and Tai **?dai*, and appeal to their uncannily similar functional behaviour as an indication of greater likelihood that they are 'genetically related'. But this would be unjustified, since, as the Mon-Khmer data presented in §3.2 show, even greater parallelism in semantic/grammatical polyfunctionality can be observed of words with no conceivable etymological relatedness at all. Even if the ACQUIRE words in Tai and Sinitic were commonly inherited, most if not all of their common complex semantic and grammatical behaviour has developed *since* the possible time of borrowing anyway. It is the *functional application*—not the form—that is shared, and most likely this has been in part borrowed, and in part independently innovated, given existing similarities in the semantic and typological profiles of these languages.

In addition to a distinction between 'genetic' and 'areal' relatedness of shared forms and/or functions, it is also important to investigate the possibility that similar or identical language-internal innovations have occurred, not due to mere coincidence, but encouraged by shared grammatical preconditions, by shared propensity for the said innovation, given already similar typological poise of the languages concerned. The more natural an innovation—and this 'naturalness' may be relative to the existing grammatical system—the more possible it is that

an areally shared feature is not directly due to diffusion but to parallel yet independent innovation.

References

- BAUER, R. S. 1996. 'Identifying the Tai substratum in Cantonese', pp. 1806–44 of *Proceedings of the Fourth International Symposium on Languages and Linguistics, Pan-Asiatic Linguistics*, V. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University at Salaya.
- BENEDICT, P. K. 1975. *Austro-Thai Language and Culture, with a Glossary of Roots*. New Haven: Human Relations Area Files Press.
- BISANG, W. 1991. 'Verb serialization, grammaticalization and attractor positions in Chinese, Hmong, Vietnamese, Thai and Khmer', pp. 509–62 of *Partizipation: das sprachliche Erfassen von Sachverhalten*, edited by Hansjakob Seiler and Waldfried Premper. Tübingen: Gunter Narr.
- CHAMBERLAIN, J. 1972. 'The origin of the Southwestern Tai', *Bulletin des Amis du Laos*, 7–8, 233–44.
- CHAO Y. R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- CLARK, M. 1989. 'Hmong and areal Southeast Asia', pp. 175–230 of *Southeast Asian Syntax*, edited by David Bradley. Canberra: Pacific Linguistics (A-77).
- CONDOMINAS, G. 1990. *From Lawa to Mon, from Saa' to Thai: Historical and Anthropological Aspects of Southeast Asian Social Spaces*. Canberra: RSPAS.
- DAI Q. 1991. 'On the affiliation of the Kadai (Zhuang-Dong) group: indications from the nature of the relationship between Tibeto-Burman and Chinese', pp. 51–63 of *Kadai: Discussions in Kadai and SE Asian Linguistics*, III, translated by Wu Hede.
- DILLER, A. V. N. 1992. 'Tai languages in Assam: daughters or ghosts?', pp. 5–43 of *Papers on Tai Languages, Linguistics, and Literatures, in Honor of William J. Gedney on his 77th Birthday*, edited by C. J. Compton and J. F. Hartmann. Northern Illinois University: Center for Southeast Asian Studies.
- DURIE, M., and ROSS, M. D. 1996a. 'Introduction', pp. 3–38 of Durie and Ross (1996b).
- 1996b. Editors of *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York: Oxford University Press.
- EDMONDSON, J. A. 1998a. 'The language corridor: new evidence from Vietnam'. Paper presented at the 1st International Conference on Tai Studies, Bangkok, 29–31 July 1998.
- 1998b. 'Tai migration patterns in interdisciplinary perspective: linguistics'. Paper presented at the 1st International Conference on Tai Studies, Bangkok, 29–31 July 1998.
- and SOLNIT, D. B. 1997. Editors of *Comparative Kadai: The Tai Branch*. Texas: SIL/UTA.
- ENFIELD, N. J. 2003. *Linguistic epidemiology: semantics and grammar of language contact in mainland Southeast Asia*. London: Routledge/Curzon.
- EVANS, G. 1999a. 'Introduction: what is Lao culture and society?', pp. 1–34 of Evans (1999c).
- 1999b. 'Ethnic change in highland Laos', pp. 125–47 of Evans (1999c).
- 1999c. Editor of *Laos: Culture and Society*. Chiang Mai: Silkworm Books.
- GEDNEY, W. J. 1976. 'On the Thai evidence for Austro-Thai', pp. 65–86 of *Ajia Afurikana-go no Keisu Keisan* 6. (Also pp. 117–66 of Gedney 1989.)
- 1989. *Selected Papers on Comparative Tai Studies*. Michigan: Ann Arbor.

- GRIERSON, G. A. 1903. *Linguistic Survey of India*, vol. 2: *Mon-Khmer and Siamese-Chinese Families (Including Khassi and Tai)*. Delhi: Motilal Banarsidass.
- HARRIS, A. C., and CAMPBELL, L. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- HARTMANN, J. 1998. 'A linguistic geography and history of Tai Meuang-Fai [Ditch-Dike] techno-culture', *Language and Linguistics* 16: 2.67–101.
- HOPPER, P. J., and TRAUGOTT, E. C. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- KEYES, C. 1977. *The Golden Peninsula: Culture and Adaptation in Mainland Southeast Asia*. New York: Macmillan.
- LAMARRE, C. 2001. 'Verb complement constructions in Chinese dialects: types and markers', pp. 85–120 of *Sinitic grammar: synchronic and diachronic perspectives*, edited by H. Chappell. Oxford: Oxford University Press.
- LASS, R. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- LEACH, E. R. 1964. *Political Systems of Highland Burma*. London: The Athlone Press.
- LE PAGE, R. B., and TABOURET-KELLER, A. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.
- LI, C. N., and THOMPSON, S. A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- LI, F. 1977. *A Handbook of Comparative Tai*. Oceanic Linguistics Special Publications 15. Honolulu: University of Hawaii Press.
- LIEN, C. 1997. 'Aspects of the evolution of *tit* in Taiwan Southern Min', *JCL*, 10.167–90.
- LONG Y. and ZHENG G. 1998. *The Dong Language in Guizhou Province, China*. Texas: SIL/UTA.
- LUO Y. 1990. 'Tense and Aspect in Zhuang'. MA dissertation, Australian National University.
- 1997. *The Subgroup Structure of the Tai Languages: A Historical-Comparative Study*. *Journal of Chinese Linguistics*, Monograph Series 12. Berkeley: University of California Press.
- MATTHEWS, S., and YIP, V. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- MATISOFF, J. A. 1973. *The Grammar of Lahu*. University of California Publications in Linguistics 75. Berkeley: University of California Press.
- 1978. *Variational Semantics in Tibeto-Burman: The 'Organic' Approach to Linguistic Comparison*. Philadelphia: Institute for the Study of Human Issues.
- 1991. 'Areal and universal dimensions of grammaticization in Lahu', pp. 383–453 of Traugott and Heine (1991).
- MIGLIAZZA, B. 1996. 'Mainland Southeast Asia: a unique linguistic area', *Notes on Linguistics* 75.17–25.
- MILROY, L. 1987. *Language and Social Networks*. 2nd edition. Oxford: Blackwell.
- OKELL, J. 1969. *A Reference Grammar of Colloquial Burmese*. London: Oxford University Press.
- QUINE, W. V. O. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- ROSS, M. D. 1997. 'Social networks and kinds of speech-community event', pp. 209–61 of *Archaeology and Language*, vol. 1: *Theoretical and Methodological Orientations*, edited by R. Blench and M. Spriggs. London: Routledge.
- ROSS, P. A. 1996. 'Archaic features in Baan Ha 'Tay', *Journal of Language and Linguistics* 14: 2.54–61.

- SAMERCHAI P. 1998. 'Tai migration in interdisciplinary perspective: physical anthropology'. Paper presented at the 1st International Conference on Tai Studies, Bangkok, 29–31 July, 1998.
- STEINBERG, D. J. 1987. Editor of *In Search of Southeast Asia*. Revised edition. Honolulu: University of Hawaii Press.
- SUN C. F. 1996. *Word-Order Change and Grammaticalization in the History of Chinese*. Stanford: Stanford University Press.
- SUWILAI P. 1987. *A Khmu Grammar*. Canberra: Pacific Linguistics (A-75).
- TERWIEL, B. J., and RANOO WICHASIN. 1992. Translators and editors of *Tai Ahoms and the Stars: Three Ritual Texts to Ward Off Danger*. Cornell, NY: Southeast Asia Program.
- THOMASON, S. G., and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- THOMPSON, L. C. 1987. *A Vietnamese Grammar*. 2nd edition. Honolulu: University of Hawaii Press.
- TRAUGOTT, E. C., and HEINE, B. 1991. Editors of *Approaches to Grammaticalization*. Amsterdam and Philadelphia: John Benjamins.
- WANG J. and ZHENG G. 1993. *An Outline Grammar of Mulao*, English translation by Luo Yongxian. Canberra: National Thai Studies Centre, Australian National University.
- WILKINS, D. 1996. 'Natural tendencies of semantic change and the search for cognates', pp. 264–304 of Durie and Ross (1996b).

Genetic versus Contact Relationship: Prosodic Diffusibility in South-East Asian Languages

James A. Matisoff

Perhaps the most striking phonological feature of the South-East Asian linguistic area (which is here defined broadly to include north-east India, the Himalayan region, and China south of the Yangtze) is the proliferation of systems of contrastive prosodic laryngeal effects—tones and phonation types—that have spread through all the language families of the region, and that have developed more elaborately here than anywhere else on earth. Before the universal phonetic mechanisms of ‘tonogenesis’ were well understood, cross-linguistic similarities in tonal systems were naively taken as *prima facie* evidence of genetic relationship; and before contact or diffusional phenomena had been studied from a sophisticated sociolinguistic point of view, the hypothesis of genetic relationship was felt to be especially convincing if there were ‘regular’ tonal correspondences between similar-looking vocabulary items in different languages. Thus Vietnamese was once thought to be related to Tai because of the overall similarity in their tone systems. More plausibly (but still erroneously), Tai and Hmong-Mien have often been included in the Sino-Tibetan family because of their large number of shared vocabulary items with regularly corresponding tones.

We now realize that language contact, if intense enough, can affect absolutely all areas of linguistic structure, and that words can easily be borrowed into an unrelated language along with their tones.

To have any hope of unravelling the strands of genetic and contact relationship in a complex linguistic area, we need to distinguish well-established genetic groupings from shaky or fanciful ones. Such an enterprise is never easy, but is especially challenging in the South-East Asian context, where ancient written records are relatively few, and where the languages tend to have monosyllabic morphemes and minimal inflectional apparatus.

This chapter begins with a discussion of some theoretical issues involved in establishing genetic relationship (§1), followed by a summary of the genetic grouping of South-East Asian languages (§2). In §3, we present some South-East Asian areal features which have resulted from widespread multilingualism. The

remainder of the paper focuses on tone as an areal feature. Section 4 briefly examines the general relationship between syllable structure and tone. In §5 and §6, we narrow the discussion down to Sino-Tibetan and Tibeto-Burman, first presenting a typology of Tibeto-Burman tone systems (§5), then a discussion of the still controversial question of the mono- vs. polygenesis of tone in this family (§6). Section 7 describes how intense language contact has led to cases of drastic changes in syllable structure, and the homogenization of prosodic systems throughout the region. Finally, §8 raises some theoretical questions for further investigation.

1. Theoretical issues in establishing genetic relationship

1.1. SCALE OF COMPARISON: MICRO-, MACRO-, MEGALO-

At relatively shallow time depths (e.g. 2,000 years BP), *microlinguistic* comparative reconstruction is possible, even in the absence of extensive written records, as long as one is dealing with a well-ramified family with surviving members in several branches. Regularity of sound correspondences can be insisted upon (even for vowels!), and exceptions to phonological rules or semantic discrepancies can be explained to everyone's satisfaction. This happy state is familiar to specialists in Tai, Loloish, or Bantu—and a fortiori to Romance philologists.

Extensive written records and morphological complexity (as in Indo-European or Semitic) and/or a large number of highly diversified daughter languages (as in Austronesian, Tibeto-Burman, or Austroasiatic) permit *macrolinguistic* work, enabling us to push back the clock to perhaps 6,000–8,000 years BP. At this level there are many unsolved and perhaps insoluble problems, though the basic validity of the family grouping is not in serious question.

At remoter time depths, the classic distinction between genetic and other types of relationship breaks down. Sound correspondences are not regular, semantics are questionable, cognates are few. Too many alternative explanations for perceived similarities are possible: chance, borrowing, areal typological convergence, universal tendencies, faulty analysis, wishful thinking. I have dubbed this sort of speculative endeavor *megalocomparison* (Matisoff 1990a).

All scales of linguistic comparison are legitimate, as long as one realizes that the rules of the game are quite different at each level. The broader the scale, the more acutely problematical the following theoretical issues become.

1.2. THE FAMILY-TREE MODEL

The classical 'Stammbaum' or 'family tree' metaphor for characterizing degrees of linguistic genetic relationship has for a century been recognized as a vast oversimplification. Languages rarely split off cleanly from their relatives. A much more appropriate image for what one finds in linguistic areas like South-East Asia might be the 'thicket', an impenetrable maze of intertwined branches. Instead of clear-cut

migrations of population groups, one finds slow ‘percolations’ or ‘filtrations’ of small groups of people.

1.3. ‘CORE VOCABULARY’ AND THE RATE OF LEXICAL REPLACEMENT

Idiosyncratic morphological features (e.g. parallel exceptional forms in inflectional paradigms) have long been appreciated as especially valuable indicators of genetic relationship. Unfortunately, in languages with minimal morphologies, like most of those in the South-East Asian linguistic area, this criterion is of little use, and one is forced to rely on lexical similarity. It has been suggested that genetic relationships can be inferred from similarities in ‘core vocabulary’. As the editors state in their Introduction, this has been shown to be without foundation.

Furthermore, it has been persuasively argued that the rate of linguistic change of all kinds is highly sensitive to extra-linguistic events, with long eras of relative stasis giving way to periods of rapid change prompted by military, political, or demographic upheavals at irregular intervals (Dixon 1997; see §8 below).

1.4. REGULARITY OF CORRESPONDENCE

Since every natural language is rife with irregularities, and since every modern language is (as Mary Haas put it) ‘a proto-language with respect to the future’, it is unreasonable to expect that all etymologically related forms in daughter languages will exhibit perfect regularity of phonological correspondence. Still, if we are to do historical reconstruction at all, we must never abandon the *ideal* of regularity. While there are stunning examples of perfectly cognate forms that have little or no surface phonetic similarity,¹ the stranger the correspondences, the more independent evidence is required to back them up. (In fact if forms from not particularly closely related languages are *too* similar, it should arouse one’s suspicions that perhaps borrowing or pure chance is involved.)

Still, it is all too easy to abuse notational devices and ad hoc explanations to make just about any correspondence achieve a specious air of regularity. It is not enough to set up ‘tables of correspondences’ without presenting all the data *that either confirm or disconfirm* the fillers of the cells in the table.

The trick is to steer a middle course between etymological promiscuity and a stodgy insensitivity to the mechanisms of linguistic variation. (See Matisoff 1978, 1994a.)

1.5. SEMANTIC LATITUDE AND AREAL SEMANTICS

It is even more of an art to decide how much semantic divergence may be tolerated among reflexes of the same etymon. Roots may indeed undergo spectacular semantic changes through time, and the glottochronological dogma against

¹ For example, ‘two’, Latin *duo*: Armenian *erku*; ‘eye’, Latin *oculus*: Modern Greek *mati* < (**op-*)*ma-ti-on* < **ok^W-mŋ-ti-on*; ‘what’, German *was*: Russian *cto*; ‘eight’, written Tibetan *brgyad*: Lahu *hi*.

accepting semantically shifted cognates in determining degrees of genetic relationship goes much too far. (See the critique of this dogma—which disregards cognates like German *Hund*: English *hound*—in Matisoff 1978: 99–106.) However, the bigger the semantic leap the better the phonological correspondence must be between the putative cognates. Otherwise the phonological and semantic arguments are like two drunks supporting each other.

Crucially, it should not automatically be assumed that semantic associations attested in one linguistic area are universally valid. Among the supposed cognates offered by Sagart (1990) to demonstrate a genetic link between Chinese and Austronesian is Proto-Austronesian **pusuq* ‘heart, central leaf’ and Old Chinese **swia* (re-reconstructed **s-j-wa?*) ‘marrow’, since marrow is supposedly ‘the heart of a bone’. Yet, aside from the dubious phonological correspondence, there is no evidence at all that marrow has ever been conceived in a ‘heartlike’ way by East Asian peoples. (What ‘marrow’ is related to conceptually—both within and without the South-East Asian linguistic area—is ‘brain’.) Similarly, after admitting that ‘the abundance of comparisons of the type *water/sap* over the type of *water/water* seriously diminishes the credibility of any hypothesis of genetic relationship’, Vovin (1993: 1) attempts to prove the Altaic affiliations of Japanese by such comparisons as Proto-Japanese **momo* ‘peach’ to Proto-Manchu-Tungus **ňang-ta* ‘nut’ (perhaps because such an association exists in North Caucasian languages). Sometimes a semantically dubious etymology is presented as if the meaning association were obvious, even though it may never have been clearly attested in any language family. As support for his Austro-Japanese theory, Benedict (1990: 193) compares Indonesian *ikan* ‘fish’ (< PAn **Sikan*) to Japanese *ika* ‘squid’ (< Proto-Japanese **yika*), since ‘squid, like fish, have long been a staple food source for the Japanese’.

As we shall see (§3.2), the notion of ‘areal semantics’ is just as valid as that of ‘areal phonology’. However, once a semantic association has already been established on independent grounds within a linguistic area, similar associations found elsewhere may well have confirmatory force. Just as ‘brain’ ↔ ‘marrow’ is unmistakably attested both in Tibeto-Burman and Indo-European, so I have hypothesized that two supposedly distinct but homophonous Proto-Tibeto-Burman roots **dyam* ‘full’ and **dyam* ‘straight, flat’ are really one and the same, offering as additional evidence the phonological similarity and intercontamination between two semantically similar Indo-European roots represented by Latin *plānus* ‘flat’ and *plēnus* ‘full’ (Matisoff 1988).

1.6. THE PHONOLOGICAL SLIGHTNESS OF SINOSPHERIC LANGUAGES

Not only is inflectional morphology at a minimum in Chinese-type (‘Sinospheric’) languages, but morphemes are monosyllabic, immensely complicating the task of reconstruction. As Dixon puts it (1997: 41): ‘A cognate set between polysyllabic forms provides much better evidence than one involving monosyllables, or single-segment forms. If the verb “go” is *gimlar-* in two

TABLE 1. Lahu homophonous monosyllables

Proto-Tibeto-Burman	Proto-Lolo-Burmese	Lahu monosyllables	Lahu disyllables	
* <i>b-r-gya</i>	* <i>ʔra</i> ¹	<i>ha</i>	<i>tê ha</i>	'hundred'
* <i>s-gla</i>	* <i>s-la</i> ³	<i>ha</i>	<i>ha-pa</i>	'moon'
* <i>s-lyā</i>	* <i>s-l(y)a</i>	<i>ha</i>	<i>ha-tē</i>	'tongue'
* <i>s-hla</i>	* <i>sla</i>	<i>ha</i>	<i>ǝ-ha</i>	'spirit'
* <i>g-ya:(p)</i>	* <i>ʔ-ya</i> ¹	<i>ha</i>	<i>ha ve</i>	'winnow'

languages, this is a stronger evidence of relationship than if it were *-a-*.³ Maybe so, but with due care it is still possible. I have reconstructed a very solid Proto-Tibeto-Burman numeral **a* 'one' on the testimony of Qiang (Sichuan) and Hruso of Arunachal Pradesh (two obscure languages that could not have been in contact), a root which has yet to be uncovered elsewhere in Tibeto-Burman (Matisoff 1997: 23).

While it is true that the homophony problem is severe in phonologically depleted languages (e.g. those of the Loloish branch of Tibeto-Burman), this is compensated for by the pervasive strategy of compounding. See Table 1. In fact there is a good deal of derivational morphology in South-East Asian languages, often taking the form in Tibeto-Burman and Mon-Khmer languages of semantically obscure prefixal elements, pronounced with schwa vocalism (cf. the Proto-Tibeto-Burman and Proto-Lolo-Burmese reconstructed forms in Table 1). These 'bulging monosyllables' or 'sesquisyllables' play a key role in the evolution of prosodic systems (see §4), and the additional phonetic material they contain comes in very handy for historical reconstruction.

Putative cognate identifications between monosyllabic and sesqui- or disyllabic languages vary greatly in their persuasiveness, ranging from the obvious (e.g. Vietnamese and Written Khmer—see Table 9 below), to the much less convincing (e.g. Tai and Hmong-Mien comparisons with Austronesian; see Benedict 1975), down to the far-out (Chinese/Austronesian; see Sagart 1990, 1993a, 1993b).

2. Recognized language families of South-East Asia

Mainland South-East Asia is home to five to six hundred languages, belonging to five great language families. (For more detailed statistics on the number and distribution of the languages in each family, see Matisoff 1991c.) Everyone now recognizes the validity of these basic macro-groupings:

- Austroasiatic (about 150 languages), comprising the 11 branches of Mon-Khmer plus the Munda languages of India. See Figure 1.
- Sino-Tibetan, comprising Chinese on the one hand and the 250–300 languages of the Tibeto-Burman family on the other. See Figure 2.

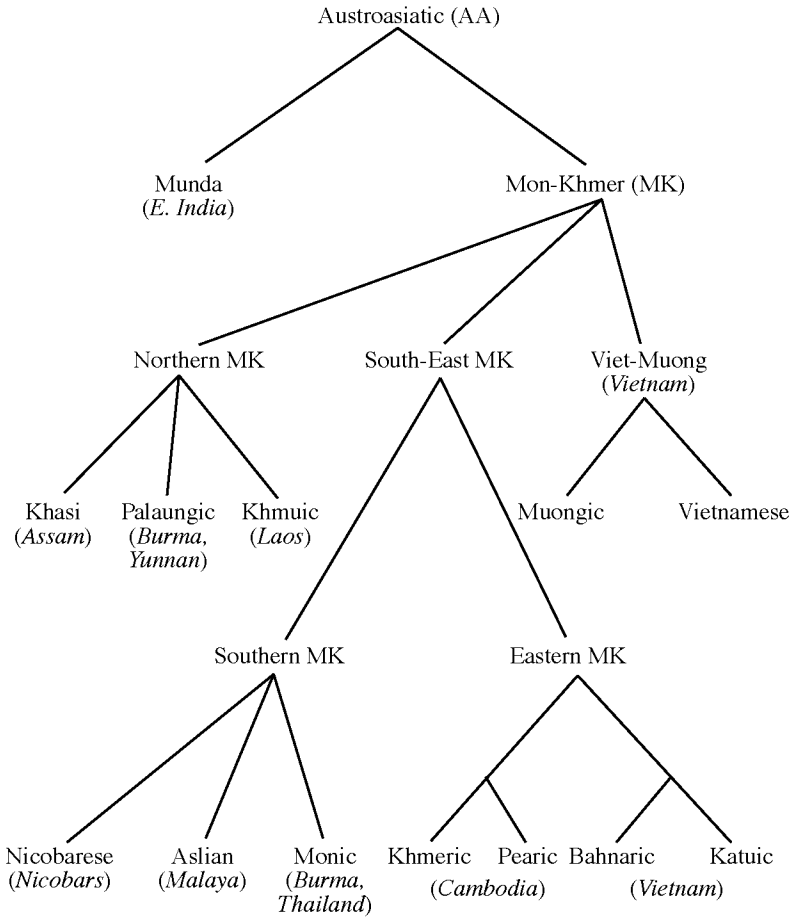


FIGURE 1. Subgroups of Austroasiatic

- (c) *Tai-Kadai* (sometimes called simply *Kadai*), consisting of about 20 languages in Tai proper (subdivided into Northern, Central, and South-Western), plus the more distantly related ‘outlier’ languages scattered through Southern China, including the *Kam-Sui* group, Lakkia, Be, Li (Hlai), Gelao, and several others. See Figure 3.
- (d) *Hmong-Mien* (= Miao-Yao), including 30–40 languages divided into two major groups, Hmongic and Mienic, and several intermediate languages like Ho Nte (She) and Pateng (Na-e). See Figure 4. (For more recent classifications of the Hmong-Mien family, see Strecker 1987: 2–3 and Niederer 1998: 49–56.)
- (e) Austronesian is an enormous family with close to 1,000 members, spoken mostly in Oceania, but represented on the mainland of South-East Asia by

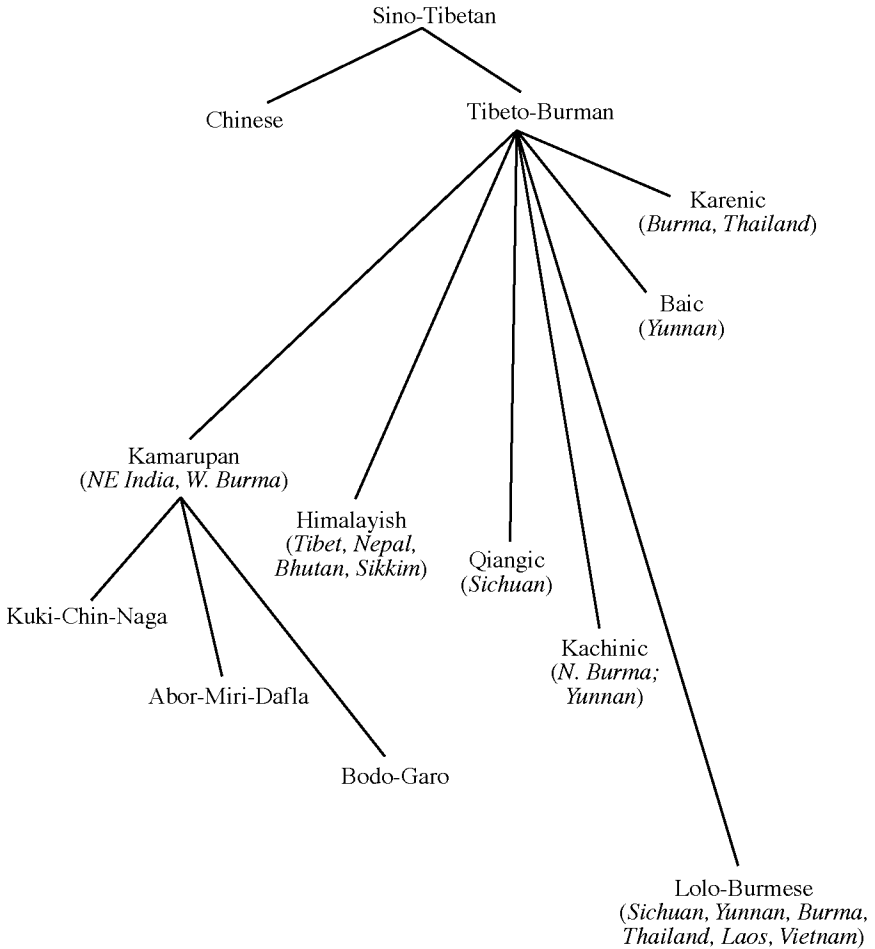


FIGURE 2. Sino-Tibetan and Tibeto-Burman

Malay and Cham. The Chamic languages, spoken on the island of Hainan and in South Vietnam and Cambodia, are of particular interest because of the extreme typological changes they have undergone under the influence of Sinospheric and Mon-Khmer languages (see below §7).

At the level of megalocomparison, virtually all possible higher-order groupings of these five basic language families have been proposed. The fact that there is no consensus in opinion here reinforces the conclusion that all these megalogroupings are speculative in the extreme.

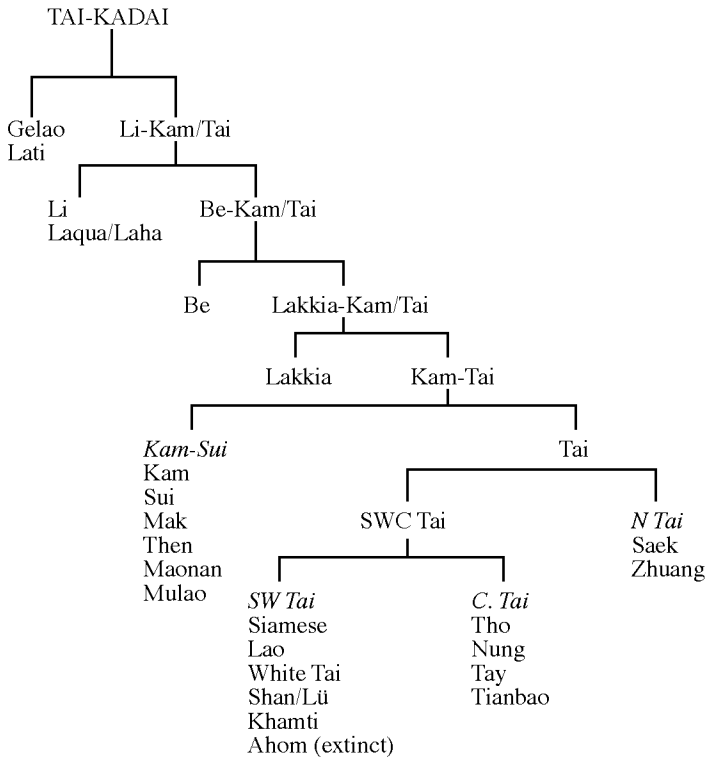


FIGURE 3. The Tai-Kadai family

3. Areal features in South-East Asia

3.1. SOUTH-EAST ASIA AS A LINGUISTIC AREA

In their splendid study of language-contact phenomena, Thomason and Kaufman (1988: 74–6) set up a five-point scale of intensity of contact influence, ranging from [1] casual contact (lexical borrowing only), through [2]–[4] (with slight, more, and moderate structural borrowing), to [5] (with heavy structural borrowing). Under Type-[5] conditions (best described sociolinguistically, in terms of the dynamics of the intense contact between the linguistic communities), no aspect of structure is immune to influence or replacement, and a language may even undergo radical changes in its basic phonological and/or grammatical typology (see §7). Thomason and Kaufman characterize these changes as involving ‘major structural features that cause significant typological disruption . . . changes in word structure rules . . . categorial as well as more extensive ordering changes in

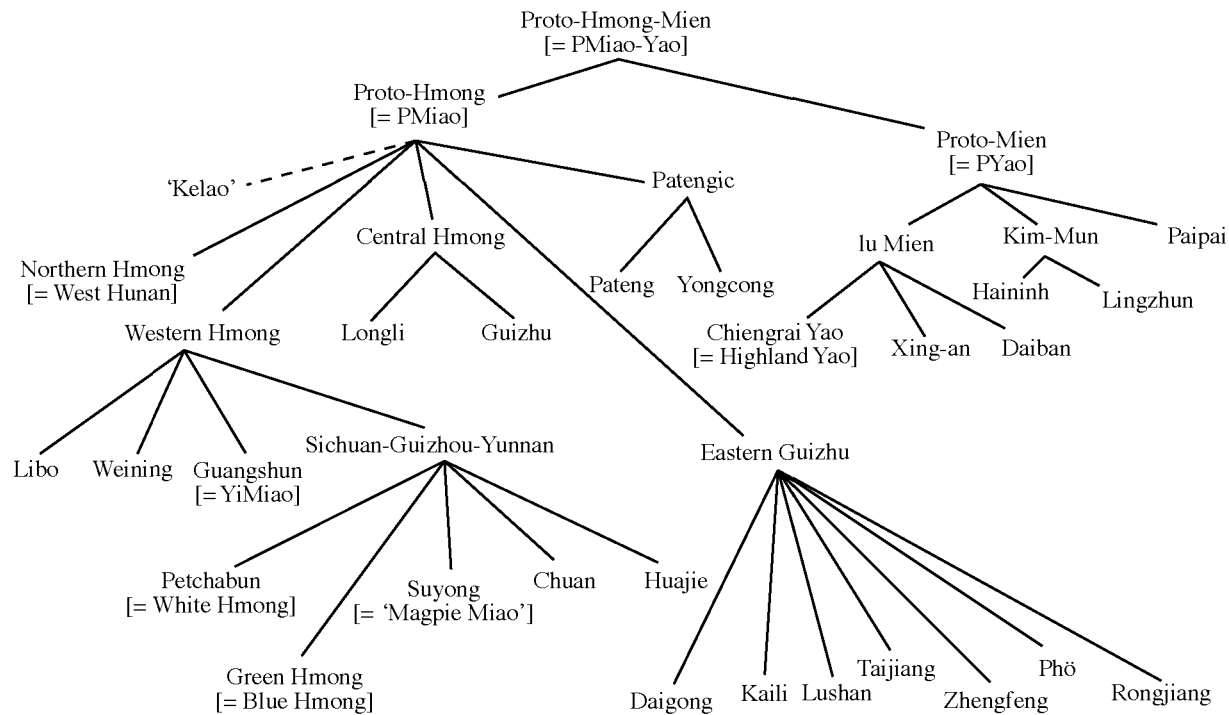


FIGURE 4. The Hmong-Mien family

morphosyntax (e.g. development of ergativity; added concord rules, including bound pronominal elements).²

Thomason and Kaufman make a distinction between a multilateral linguistic area, with only a few area-wide features but with many instances of localized diffusion and where the directionality of influence is unclear, and a non-multilateral area where one can 'establish the source of interference features and the direction and mechanism of diffusion'.³ 'What a long-term multilateral [area] seems to promote ... is the gradual development of isomorphism in all areas of structure^[4] except the phonological shapes of morphemes' (Thomason and Kaufman 1988: 96).

Cogent as these remarks are, this dichotomy between multilateral and non-multilateral linguistic areas seems quite artificial when applied to South-East Asia. Does not every 'linguistic area' arise from an accumulation of individual cases of 'localized diffusion'? On the one hand there is widespread structural isomorphism throughout the region, and the directionality of influence has often not been clear, especially in prehistoric times, and has often reversed itself according to the vicissitudes of cultural history. Benedict (1975) surmises that the Tai peoples might have influenced the early Chinese more than vice versa in the dim past, although this certainly changed later. The Mons and the Khmers once exerted great cultural and linguistic influence on the Burmese and Thai, respectively, though the balance has decisively tipped in the other direction in modern times.

Yet there certainly are numerous widespread East and South-East Asian areal features that go well beyond 'localized diffusion', and in modern times the directionality is often obvious, with the two greatest centres of influence being the civilizations of China and India. I would even claim that South-East Asia comprises two linguistic areas at once: one 'vertical', distinguishing the languages of the hard-scrabble minority populations of the hills from those of the major languages of the plains (one important difference is the lack of elaborate honorific language or status-based pronominal systems in the languages of the humble hill-dwellers); and one 'horizontal', cutting across the entire region.

3.2. SOUTH-EAST ASIAN AREAL FEATURES AND ASPECTS OF STRUCTURE

One could write a whole book on this topic, but here we can only list a few of the area-wide features that give South-East Asia languages their special flavour:

² Cf. the ongoing controversy about the Tibeto-Burman 'verb pronominalization' (subject and/or object marking on the verb), which is found in several branches of the family. Some scholars consider this 'head-marking' to be a feature inherited from Proto-Tibeto-Burman; others (including myself) feel that it is secondary, helped along by contact influence from Indo-Aryan languages.

³ Thomason and Kaufman employ the term 'Sprachbund' for a multilateral situation and 'linguistic area' for a non-multilateral one. Other scholars consider 'Sprachbund' and 'linguistic area' to be synonyms.

⁴ Including semantics, as we shall see (§3.2), leading to the phenomenon of *intertranslatability*.

(a) GRAMMATICAL

- (i) topic-prominence (not subject-prominence; see Li and Thompson 1981)
- (ii) aspect (not tense) as the most important verbal category
- (iii) verb serialization and verb concatenation (see Matisoff 1991a; Enfield, this volume)
- (iv) sentential nominalizations: treating whole sentences as noun-like, without embedding them into any larger unit (see Matisoff 1973b)
- (v) complex systems of particles (usually several dozen per language), often demonstrably grammaticalized from root nouns or verbs
- (vi) lack of grammatical gender; no case systems for common nouns
- (vii) classifier systems (not plural markers on common nouns)
- (viii) compounding as a key morphological process
- (ix) elaboration: a special type of quadrisyllabic reduplication, often with the first and third, or second and fourth, syllables the same.

(b) LEXICOSEMANTIC

- (i) highly specific verbs—Diffloth (1993) reports an Aslian (Mon-Khmer of Malaysia) monosyllabic verb that means ‘to stack up flat round objects (like pancakes)’; a profusion of lexical distinctions in verbs of manipulation like ‘cut’, ‘carry’, ‘wash’
- (ii) psycho-collocations: expressions for intellectual activities, qualities of personality, or emotions, containing a morpheme which explicitly mentions the receptacle or arena where the psycho-phenomenon unfolds (heart, mind, spirit, liver, etc.; see Matisoff 1986)
- (iii) sentence-final particles with the exclusive function of expressing emotional tone
- (iv) receptivity to new lexical items; mixing of native and foreign items in compounds and collocations
- (v) parallel lexicalizations, calques, intertranslatability, e.g. ‘pig’ + ‘crazy’ = ‘epilepsy’, ‘fly’ + ‘shit’ = ‘freckle’, ‘eye’ + ‘foot’ = ‘anklebone’ (see Matisoff 1978: 70)
- (vi) parallel situational formulas, greetings: *Have you eaten yet?*; *Where are you going?* (but not *Good morning* or *God be with you.*)

(c) PHONOLOGICAL

- (i) prime importance of the syllable as the unit of phonological structure
- (ii) apparent phonetic slightness of the syllable (see §1.6), usually compensated for by rich systems of syllable-onsets or tonal contrasts
- (iii) unstressed prefixal syllables combined with fully stressed root syllables, constituting ‘bulging monosyllables’ or ‘sesquisyllables’ (cf. the ‘compounding/prefixation cycle’, §4)
- (iv) prenasalized obstruents; voiceless and glottalized sonorants

- (v) imploded voiced stops, but only labial and dental, not velar
- (vi) apical vowels after sibilant initials
- (vii) no manner contrasts in syllable-final stops; restriction of final consonants to (at most): /-p -t -k -m -ŋ -r -l -s / (also -c and -ñ for Mon-Khmer).

Most importantly in the context of the present chapter:

- (viii) tone-proneness
- (ix) changes in manner of initial consonants⁵ (especially the devoicing of *voiced obstruents and the voicing of *voiceless sonorants), with concomitant tonogenetic or registrogenic effect
- (x) key tonogenetic role of prefixes, which often underlie consonantal mutations; similar tonogenetic effects of s- and ʔ-
- (xi) vowel length interacting with tone in stopped syllables (Tai, Mien, Cantonese)

3.3. GRADING SOUTH-EAST ASIAN CONTACT SITUATIONS IN TERMS OF THE THOMASON AND KAUFMAN SCALE OF INTENSITY OF CONTACT

It is interesting to try roughly to categorize contact situations in South-East Asia according to the Thomason and Kaufman scale, though any attempt to situate them at precise points on the continuum is necessarily impressionistic.

Almost any geographically contiguous languages in South-East Asia, regardless of their individual genetic affiliations, are sure to exercise at least a [1]–[2]-level influence on each other. Among the innumerable examples of slight to moderate influence that could be cited are the following:

- (a) WHERE THE PRESTIGE OF THE DONOR AND RECEIVER LANGUAGES IS ABOUT EQUAL (ADSTRATAL)
 - (i) (in Yunnan) (Tibeto-Burman) → Palaung-Wa (MK)
 - (ii) (in NE India) Khasi (MK) ↔ Barish (Tibeto-Burman)
 - (iii) (in the Himalayas) Mon-Khmer → Lepcha (Tibeto-Burman)⁶
 - (iv) (in peninsular South-East Asia) Thai → Vietnamese
Southern Thai ↔ Malay (AN)
Southern Thai → Kelantan Chinese
(northern Malaysia)
- (b) WHERE THE DONOR LANGUAGE HAS HIGHER PRESTIGE THAN THE RECEIVER
 - (i) (in Laos) Lao (Tai) → Khmuic (MK)
 - (ii) (in Burma) Shan (Tai) → Lahu (Tibeto-Burman)
 - (iii) (in Yunnan) Yunnanese Mandarin → Lahu
 - (iv) (in Malaysia) Malay (AN) → Aslian (MK)

⁵ What Haudricourt called *mutations consonantiques*. Such mutations in the history of Indo-European (e.g. Grimm's Law or the Second Germanic Sound Shift) have never led to tonogenesis.

⁶ This is supposedly a 'substratal' influence. See Forrest (1962).

It is also not hard to find cases of contact influence that would receive a rating of [3]–[4]:

- (i) (in Burma) Mon (MK) → Burmese (Tibeto-Burman)
Features of Old (Written) Burmese ascribable to Mon influence include final palatal consonants /-c ñ/ (see (c) in §3.2) and a phonation-prominent prosodic system (§5.3).
- (ii) (in Nepal) Nepali/Kashmiri/Hindi (Indo-Aryan) →
Newari (Tibeto-Burman)
The complicated periphrastic verbal forms in Newari, as well as a major proportion of its lexicon, show heavy Indo-Aryan influence, to the point where it is difficult to determine what the closest Tibeto-Burman relatives of Newari might be.
- (iii) (in peninsular South-East Asia) Thai ↔ Khmer (Varasarin 1975)
The Khmer, inhabitants of peninsular South-East Asia long before the Tai peoples, transmitted Indic writing to the Siamese and Lao, as well as a major lexical component to the Siamese (but not so much to the Lao) language (see Varasarin 1975).
Malay → Kelantan Chinese (see Teo Kok Seong 1993)

Most interesting for our purposes are cases of extremely intense contact (meriting a [5] rating), resulting in typological change, or *metatypy*. All too often the relative power of the languages in such close contact is so disparate that one of them dies, no matter how radically it changes its original typological profile. Such a thanatoglossic fate seems to be in store for the Hayu (Tibeto-Burman) language of east Nepal, for example, which has been steadily losing ground to Nepali (Indo-Aryan) during the past century (compare the descriptions of Hodgson (1880) and Michailovsky (1988)), and which no longer has monolingual child speakers. When the number of speakers on the receiving end is large, however, intense contact leads merely to profound structural influence. Several such cases involving prosodic phenomena will be discussed (§7), including: Chinese phonotactic and prosodic influence on Vietnamese, Tai, and Hmong-Mien; Mon phonational influence on Burmese, with subsequent Burmese tonal influence on Karenic; and the tonal and registral adventures of the Chamic branch of Austronesian, originally disyllabic and non-tonal, under Mon-Khmer and Sinospheric influence.

4. Syllable structure and tone

The Tibeto-Burman family is remarkable for its typological diversity: phonological, morphological, and grammatical. Much of this diversity is explicable in terms of the interinfluence of the two great linguistic areas dominated by China and

TABLE 2. Proto-Tibeto-Burman, Written Tibetan, and Lahu syllable canons

					[T]		
Proto-Tibeto-Burman	(P ₂)	(P ₁)	Ci	(G)	V(:)	(Cf)	(s)
Written Tibetan	(P ₂)	(P ₁)	Ci	(G)	V	(Cf)	(s)
		T					
Lahu	(Ci)	V					

India: the Sinosphere and the Indosphere. Although Tibeto-Burman morphemes are basically monosyllabic, the Tibeto-Burman monosyllable varies in complexity from that of Written Tibetan (which closely approximates what is set up for Proto-Tibeto-Burman) and that of Lahu. See Table 2.

The abundant presence of prefixes (or pre-initials) means that Proto-Tibeto-Burman was really more *sesquisyllabic*—i.e. a syllable-and-a-half in length—than strictly monosyllabic. Many Tibeto-Burman languages are sesquisyllabic to the present day. (Mazaudon (1974: 84–90) divides up the Tibeto-Burman family into ‘schwa-languages’ vs. ‘non-schwa languages’, according to whether they are sesquisyllabic or strictly monosyllabic. The term ‘sesquisyllable’ was coined in Matisoff (1973a). Haudricourt refers to words of this type as ‘quasi-monosyllabiques.’)

Sinospheric Tibeto-Burman languages tend to be more strictly monosyllabic than others. Since they also preserve final consonants and prefixes less well than many Indospheric languages, they are usually more tonally complex than less uncompromisingly monosyllabic languages. Strictly monosyllabic languages seem especially ‘tone-prone’:

There is something about the tightly structured nature of the syllable in monosyllabic languages which favors the shift in contrastive function from one phonological feature of the syllable to another. . . . So tightly interdependent are these neighboring vowels and consonants, that certain phonetic features seem to have bounced back and forth from vowel to consonant and back again through the history of the Tibeto-Burman languages. (Matisoff 1973a: 78–9)

In my view prosodic contrasts are constantly arising and being lost in the languages of this area, concomitantly with changes in syllable- and word-structure. These changes naturally include alterations in the manner of articulation of initial consonants (what Haudricourt called *mutations consonantiques*). The loss of a manner distinction in initial consonants has different consequences according to whether the language was already tonal or not: if the language was tonal, a loss of contrast can cause a tonal split; if the language was not tonal, a loss of manner contrast can cause a phonational difference, as in Austronesian (Cham of Cambodia and Vietnam; §7.3) or Austroasiatic (Lamet, Riang (Palaungic group)).

Very schematically we can envision the complementary cycles of tonality and syllable-type more or less as in Figure 5. (I have discussed this ‘compounding/prefixation cycle’ in several publications and talks, including Matisoff 1973a: 82–4, 1978: 58–72, 1990b.)

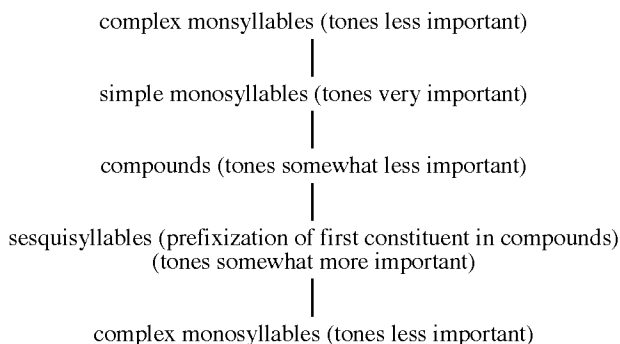


FIGURE 5. The compounding/prefixation cycle

In favourable cases it can be demonstrated that the sources for the ‘minor’, unstressed prefixal portions of sesquisyllables were independent morphemes to which meanings can be assigned (e.g. WB *pərwak* ‘ant’ < Proto-Tibeto-Burman **bəw-rwak* (**bəw* ‘insect’); WB *səmak* ‘son-in-law’ < Proto-Tibeto-Burman **za-mak* (**za* ‘son; child’)). Some languages find themselves caught up in different stages of the cycle at the same time, so that they include both tonal and atonal dialects (Tibetan, Qiang, Khmu (Mon-Khmer)). Many Tibeto-Burman languages with complex monosyllables (i.e. relatively good consonantal preservation) are only marginally tonal, or have no phonemic tonal contrasts at all.

Tone is by no means a simple matter of relative pitch, but rather a complex bundle of features, including phonation-type, tongue position, pharyngeal tension, vowel length, and contour. Whatever the exact interrelationships of these phonetic mechanisms may be, the fundamental opposition seems to be between what we could call the *tense* vs. *lax laryngeal syndromes*. See Table 3. (Matisoff 1973a: 76.)

In view of their diversity in terms of syllable structure, it is not surprising that the tonal Tibeto-Burman languages differ in the size of their *tone-bearing unit*,

TABLE 3. Laryngeal attitudes

Tense-larynx syndrome	Lax-larynx syndrome
higher pitch/rising contour	lower pitch/falling contour
association with -ʔ	association with -h
voicelessness	voicedness, breathiness
retracted tongue root (see Gregerson 1973)	advanced tongue root
‘creaky’ laryngeal turbulence	‘rasping’ laryngeal turbulence
larynx tense and/or raised	larynx lax and/or lowered
reduced supraglottal cavity	distended supraglottal cavity

varying from the single syllable to 'phonological words' which may contain two, three, four or more syllables. Tone systems may also vary in the role played by phonational (register) differences, as opposed to mere pitch and contour contrasts. It is in fact impossible to draw a strict dividing line between 'tone' and 'phonation'.

5. Typology of Tibeto-Burman tone systems

The rough typological distinctions in the following sections are not mutually exclusive. Burmese is simultaneously a (mildly) sesquisyllabic language and a phonation-prominent one. Jingpho is highly sesquisyllabic but not particularly phonational. Tamang Risiangku is phonation-prominent, but also has a word-tone system.

5.1. OMNISYLLABIC TONE LANGUAGES: THE CASE OF LAHU

Lahu is a Sinospheric, strictly monosyllabic language; like Chinese (especially southern dialects like Cantonese), but unlike Mandarin, Lahu has no unstressed or tonally 'neutral' word- or phrase-final syllables. If postpositional particles are in danger of losing their stress, they just become fused with other particles, and the particle combination as a whole has stress. Furthermore Lahu totally lacks unstressed prefixal syllables with schwa vocalism; prefixes (including the ubiquitous δ - < Proto-Lolo-Burmese $*aŋ$ -) are fully stressed and tonal. (This δ -prefix, which sometimes serves to nominalize verbs, usually disappears in compounds: *u* 'lay an egg', δ -*u* 'an egg', *gãʔ-u* 'hen's egg'.) Even polysyllabic loanwords receive a tone on each syllable (*kə̌mɪtɪ* 'committee').

As indicated in Table 2, Lahu is a language with very simple monosyllables, with no initial consonant clusters or final consonants, and no contrast in vowel length; many syllables lack an initial consonant entirely. In compensation there is a rich system of seven tones, five open and two checked. The two checked tones (as well as one of the open ones⁷) descend from earlier syllables with final $*st$ ops. It is easy to find minimal septuplets illustrating all seven Lahu tones. See Table 4.

The phonological simplicity of the Lahu syllable has led to massive homophony. The language has resorted to two strategies to preserve contrastivity, one phonological (the proliferation of tones) and one morphological (compounding). Even with tonal contrasts, there remain many homophonous monosyllables; this is handled, in Lahu (as for example in Mandarin) by compounding or collocation. See the Lahu syllables pronounced *ha* (all under mid-tone, unmarked in the transcription), in Table 1. In sum, Lahu has monosyllables and innumerable di- and trisyllabic compounds, but no sesquisyllables.

⁷ This is the well-known Lahu high-rising tone, which descends by 'glottal dissimilation' from syllables with Proto-Lolo-Burmese $*voiced$ glottalized or $*voiceless$ sibilant initials and a $*final$ stop. See Matisoff (1970), where the word *tonogenesis* was first used.

TABLE 4. A Lahu minimal tonal septuplet: *ca* on all seven tones

Transcription	Tonetics	Description	Glosses
<i>ca</i>	<i>ca</i> ³³	mid level	'look for; seek'
<i>cá</i>	<i>ca</i> ³⁵	high rising	(1) 'boil'; (2) 'join'
<i>câ</i>	<i>ca</i> ⁵³	high falling	'to eat'
<i>cà</i>	<i>ca</i> ²¹	low falling	'be ferocious'
<i>cā</i>	<i>ca</i> ¹¹	low level	'to feed'
<i>cà?</i>	<i>ca</i> ⁴	high checked	'string; rope'
<i>cà?</i>	<i>ca</i> ²	low checked	'to push'

5.2. SESQUISYLLABIC TONE LANGUAGES: THE CASE OF JINGPHO

Jingpho (also known as Kachin), one of the most important Tibeto-Burman languages, spoken in north Burma and adjacent areas of Yunnan and India, has well-preserved final stops and nasals, a robust system of three principal tones in syllables ending with a vowel or nasal, and a two-way tone contrast in stopped syllables. It is also a language with a high percentage of sesquisyllabic words, but relatively few disyllabic compounds. We may divide Jingpho words into four structural types.

(a) MONOSYLLABIC

T
C_i (G) V (C_f)

Purely monosyllabic words are relatively rare in Jingpho, though they certainly exist. If they end in a vowel or nasal they may appear either under the high-tone /¹/ (55), mid-tone /²/ (33), or low-tone /³/ (31), e.g. *khú* 'be smoky'; *khron* 'spread quickly'; *sai* 'blood' (there is also a rare and secondary falling tone /⁴/ (51), usually in a sandhi relationship with the low tone. Verbs in the low tone acquire this falling tone when negated). Stopped syllables may end in /-p -t -k -ʔ/. Since Proto-Tibeto-Burman *-k has developed into Jingpho final -ʔ (e.g. 'pig' Proto-Tibeto-Burman *p^wak > Jingpho wàʔ), modern Jingpho -k occurs only in loanwords (especially from Shan and Burmese). There is a high vs. low tonal contrast in these 'dead' syllables, e.g. *yáʔ* 'night', *ʔúp* 'bank a fire' vs. *myiʔ* 'eye', *láp* 'leaf'.⁸

(b) PRENASALIZED

T
N-C_i (G) V (C_f)

A frequent syllable onset is the syllabic nasal *ŋ*, which assimilates in position of articulation to the following root-initial, and which may take a full tone.

⁸ This tonal contrast in stopped syllables was implausibly imputed by Maran (1971) to a voicing contrast in the *final* consonant. This analysis involves positing a voiced equivalent to glottal stop, a phonetic impossibility. I have tried to correlate this Jingpho tonal split in stopped syllables to the one that occurred in Loloish, with equivocal but suggestive results. See Matisoff (1974, 1991a), and §6 below.

Before noun-roots, it has been shown to derive frequently from Proto-Tibeto-Burman **r-* (Benedict 1972: 109).⁹ This syllabic nasal under the high tone /ń-/ fills an important grammatical role: this is the negative morpheme (< Proto-Tibeto-Burman **ma*). Examples: *m-būŋ* 'wind (n.)' *ń-lūŋ* 'stone' (< Proto-Tibeto-Burman **r-luŋ*; cf. Mikir *arlon*), *ŋ-khyūn* 'kidney'; *ń-lû* 'not have' (< *lû* 'have').

(c) SESQUISYLLABIC

T
C₃-C₁ (G) V (C_f)

The typical Jingpho word is sesquisyllabic (for example, all the numerals from 'one' to 'ten' are sesquisyllables, except for *krúʔ* 'six' and *śī* 'ten'). The vowel of the minor syllable is always unstressed schwa. No fewer than twenty-one consonants (including ʔ-, sometimes regarded as zero-initial) may begin the minor syllable, though only five of them are common, and twelve are marginal or dialectal. (No clusters like *prə-* or *krə-* may occur in these minor syllables—unlike the situation in many Mon-Khmer languages, or in Khmer words borrowed into Siamese.)

A rough count of the entries beginning with each prefix in Hanson (1906) gives some idea of their relative frequency (approximate number of pages in parentheses):

<i>Very frequent:</i>	<i>mə-</i> (41.5); <i>ʔə-</i> (37); <i>kə-</i> (35.5); <i>lə-</i> (27.5); <i>śə-</i> (24.3)
<i>Fairly frequent:</i>	<i>gə-</i> (9.3); <i>ǰə-</i> (6.8); <i>sə-</i> (6.7)
<i>Rare:</i>	<i>tśə-</i> (4.5); <i>pə-</i> (4); <i>khə-</i> (3); <i>də-</i> (3); <i>phə-</i> (1.5); <i>tsə-</i> (1)
<i>Less than one page:</i>	<i>tə-, thə-, bə-, nə-</i> (Hkauri dialect), <i>rə-</i> (Hkauri), <i>ŋə-</i> (Hkauri)

Total of all sesquisyllables: 232.6 pages, or about a third of the dictionary.

In some of their occurrences, several of these stressless prefixes have relatively clear meanings, and sometimes it is clear which full morpheme they derive from, e.g. *śə-* and *ǰə-* 'causative' < Proto-Tibeto-Burman **s-*; *mə-* 'stative' < Proto-Tibeto-Burman **m-*; *lə-* 'action with the hands or feet' < Proto-Tibeto-Burman **lak* 'hand'. In most cases, however, their meaning, if any, is quite obscure.

It has been claimed (e.g. by Maran (1971), a native speaker) that there is a two-way tonal contrast in minor syllables; Dai *et al.* (1983) distinguish all three tones in these syllables (though the high and low tones are much more frequent here than the mid tone). I confess I have never perceived any such contrast in Maran's speech (he was my consultant in 1963). Even if it exists, it is certainly secondary, undoubtedly reflecting the influence of the tone in the following major syllable.

⁹ There are twenty-four pages of words with ń- in Hanson's 739-page dictionary (1906). The reflex of Proto-Tibeto-Burman prefixal **m-* is the very frequent Jingpho prefix *mə-* (below).

(d) DISYLLABIC

Jingpho has relatively few compounds composed of two monosyllabic morphemes. Such compounds tend to have their first syllables reduced (cf. *lə* 'action with the hands or feet' < Proto-Tibeto-Burman **lak* 'hand'). When a sesquisyllabic free word becomes a constituent in a compound it may lose its prefix: *məsìn* 'liver' > *sin-wóp* 'lungs' ('spongy liver'). Many disyllables have a meaningless but syllabic prefix as their first element, e.g. *gùm-phrò* 'silver'. Many others, however, do consist of two root-morphemes: *wōi-bō* 'monkey-fern'; *tsùn-lōŋ* 'island'; *phùŋ-tāŋ* 'echo'; *phùt-kái* 'sit cross-legged'.

Jingpho is thus a language which in spite of its high degree of sesquisyllabicity is still fully tonal in major syllables, and is perhaps becoming so in minor syllables as well.

5.3. PHONATION-PROMINENT TONE SYSTEMS:¹⁰ THE CASE OF BURMESE

We shall take Burmese as an example of a phonation-prominent Tibeto-Burman tone language, though many other Tibeto-Burman languages, notably Lhasa Tibetan (see Shefts 1968; Chang and Shefts Chang 1968; Mazaudon 1974: 49–54), are similar.¹¹ Spoken Rangoon Burmese has three tones that descend from *live (open or nasal-final) syllables. Those syllables which descend from Proto-Lolo-Burmese *dead syllables (with final stops */-p -t -k/) are uniformly pronounced with a short high tone and a clear glottal stop; following syllables, even in close juncture, do not undergo voicing of their initial consonant, though syllables after the three 'live' tones do undergo such voicing. Note that there is no tonal contrast in Burmese stopped syllables, in sharp contrast to the Loloish languages (and even other languages of the Burmish group), all of which have at least a two-way contrast in such syllables (Matisoff 1972, 1991b). A four-way minimal contrast among these tones is shown in Table 5.

The three open tones do have pitch differences, but also concomitant vowel length and (crucially) phonational differences. Vowels under Tones 1 and 2 are relatively long (especially in open as opposed to nasal-finalled syllables); those under Tone 3 are 'half-long', ending in a lax glottal catch; while those under Tone 4 are quite short and end with a sharp glottal stop. Pitch *per se* does not seem to be a very reliable or salient feature for distinguishing Tones 1 and 2. Tone 1 is relatively low, but sometimes is realized higher as it reacts with certain phrasal intonations. Tone

¹⁰ I am using this term by analogy to the concepts 'topic-prominent' vs. 'subject-prominent' (Li and Thompson 1981: 15ff.). Two of the first three Tibeto-Burman languages I studied, Jingpho and Lahu, lack significant phonational features, which made me slow to recognize their fundamental importance, both in Tibeto-Burman and areally.

¹¹ Another well-known example among Himalayan languages is Chepang (see Caughley 1972) with a three-way contrast between clear, breathy, and creaky voice that e.g. Weidert (1987) takes as a direct inheritance from the Proto-Tibeto-Burman system of phonational contrasts. See also Ostapirat (1997), who compares the phonations of Chepang to the tones of Chin languages. The most detailed and reliable account of phonation in a Himalayan language remains Mazaudon (1973: 61–107).

TABLE 5. Modern Burmese tones

Tone 1 (clear)	<i>la</i>	'come' (< PLB * <i>la</i> ¹)
Tone 2 (breathy)	<i>lā</i>	'mule; yes-no question particle' (< PLB * <i>la</i> ²)
Tone 3 (creaky)	<i>la'</i>	'moon' (< PLB * <i>la</i> ³)
Tone 4 (stopped)	<i>laʔ</i>	'fresh; new' < WB <i>lat</i> ; 'be uncovered; empty' < WB <i>lap</i>

2 has two variants: one allotone is high level (in non-phrase-final position); the other, phrase-final variant, has a decided fall at the end. There is a region of pitches in the mid-range of the voice where it would be hard to distinguish monosyllables in isolation, were it not for the considerable phonational difference between the tones—Tone 1 is clear, while Tone 2 is decidedly breathy.

This sort of phonational contrast is of course most typical of the Mon-Khmer languages, and the Burmese system has long been suspected of having undergone Mon influence. Yet we also find such systems elsewhere in Tibeto-Burman, e.g. in Himalayish (see the discussion of Tamang, §5.4), as well as throughout the South-East Asian linguistic area.

Mazaudon (1974: 60–2) has suggested that certain Loloish tonal and manner developments are more comprehensible if one assumes that the prosodic system of Proto-Lolo-Burmese was basically phonational rather than 'melodic'. In Matisoff (1979: 27–9), I showed that the Proto-Lolo-Burmese *voiced series of obstruents developed differently in Sani according to the tone of their syllable: they remained voiced under Tone 2, but became voiceless under Tone 1, furnishing an unusual example of the tone determining the initial rather than vice versa. Mazaudon observes that this might be due to some phonational feature associated with Tone 2 that retarded the loss of voicing (or that favoured its retention), undoubtedly breathiness. This is roughly analogous to a situation in Akha, where the Proto-Lolo-Burmese *voiceless series of stops becomes aspirated in non-checked syllables, but unaspirated in laryngealized ones (from */-p -t -k/)—a sort of dissimilatory tension between aspiration and laryngealization. Even closer to the Sani developments is Mandarin, where the Middle Chinese *voiced series devoiced under all tones, but became aspirated only under the *píngshēng*—evidently the marked phonational features of the oblique tones were incompatible with aspiration. All these phenomena are reminiscent both of Grassman's Law (concerning the loss of the first of two aspirated consonants in successive syllables in Greek and Sanskrit), and the phenomenon of 'glottal dissimilation' noted for Lahu, whereby checked syllables that also have glottalized initials lose all of their marked phonational features and acquire a clearly phonated high-rising tone (Matisoff 1970, 1972). (The principle of glottal dissimilation actually holds also for Lisu and Sani, as well as for Ahi and Nasu—Mazaudon 1974: 23, 43.)

Within Sino-Tibetan it is always taken for granted that Chinese *píngshēng*, and the Tibeto-Burman tones that supposedly correspond to it (e.g. Proto-Lolo-Burmese Tone *1) are phonationally neutral or unmarked, while the other open

tones have some kind of special phonation, breathy or creaky. This is largely because *pingsheng* words are roughly twice as frequent as the words under *B and *C put together. Yet in Lolo-Burmese, Tones *1 and *2 are of roughly equal frequency. By the way, Burmese and Chinese do not agree here, with the Burmese creaky tone corresponding to Chinese Tone *C if anything, while Burmese breathy tone corresponds to Chinese *B (*shǎngshēng*).

5.4. PHONATION-PROMINENT WORD-TONE SYSTEMS WITH 'TONE SPREADING':

THE CASE OF TAMANG RISIANGKU

A paradigm example of a non-omnisyllabic tone language is the Risiangku dialect of Tamang (Nepal), described definitively by Mazaudon (1973). This dialect has four tones, but the 'tone-bearing unit' is not the syllable but the phonological word. (Many Tibetan dialects, and many other Himalayish languages, have similar systems; see Sprigg 1966.)

This is a word-tone language, or *langue à ton de mot*. Each of the four tones has a distinctive manifestation in words of all syllable-types: monosyllables, sesquisyllables, disyllables, trisyllables and quadrisyllables. (The tonetics are different, e.g. for disyllabic words and for sequences of two monosyllables.) Particles are toneless, and never occur in isolation; they combine with the previous root-morpheme to form phonological words. This type of system resembles, but is different in crucial respects from 'pitch-accent' systems like that of Japanese. In Japanese there are only two pitch possibilities (not four as in Tamang), and the most prominent (i.e. high-pitched) syllable is not necessarily the first in the word (whereas in Tamang tone always inheres in the first syllable of the phonological word). More crucially, all four tones in Tamang can manifest themselves on a single syllable, unlike the case in pitch-accent languages. (See the discussion concerning 'Langues à plusieurs types d'accent' in Mazaudon 1973: 91–2.)

The tonetic features of the four tones include in Tamang pitch, length, and phonation type. There is a complicated bundle of features associated with each tone:

- Tone 1: high, short, constricted, tense
- Tone 2: mid-high long; unmarked phonationally
- Tone 3: rising, lax, breathy (lower than Tone 2)
- Tone 4: very low, falling (in initial position), breathy

The phonation type determines the manner of the initial consonant. Syllables under one of the *clear* tones (1 and 2) may have aspirated initials, but not voiced ones; syllables under one of the *breathy* tones (3 and 4) may not have aspirated initials, but may have voiced ones.¹² Mazaudon (p. 82) refers to the clear/breathy distinction as one of *registre*. Within each register, one tone is higher in pitch than

¹² Interestingly, men's and women's speech differ in the effects breathiness has on the initial consonant: men have a *voiced* breathiness in Tones 3 and 4, while women have a *voiceless* breathiness.

TABLE 6. The tones of Tamang Risiangku

	Clear	Breathy
High	1	3
Low	2	4

the other: 1 and 3 are higher than 2 and 4, yielding a four-way system of oppositions. See Table 6.

Historically these four tones can be demonstrated to have resulted from the splitting of the two Proto-Tamang tones *A and *B (*A > 1 and 2; *B > 3 and 4). This split occurred when the *voiceless nasals became voiced and the *voiced stops and fricatives became voiceless unaspirates. Very similar mutations have caused the splitting of the Proto-Tai tones in Siamese (except that Proto-Tai *voiced stops became Siamese voiceless aspirates).

Vowel length is contrastive only in initial stressed open syllables, never in the second syllable of a disyllabic word, or in closed syllables. Length is considered to be a distinctive feature of the vowels, not of the tones. As in Mon-Khmer, vowels in the breathy register are more centralized than in the clear register.

If this were an omnisyllabic tone language, one would expect $4 \times 4 = 16$ possible tonal patterns in disyllables, and $4 \times 4 \times 4 = 64$ patterns in trisyllables. Instead one finds only four patterns in Risiangku words, no matter how many syllables that word may have. Atonic syllables do not constitute a 'neutral tone'; their contours are part of the distinctive bundle of features of the particular word-tone they belong to.

Mazaudon's account of the behaviour of non-tonic syllables may serve as an excellent definition of the phenomenon of 'tone-spreading':

Phonétiquement . . . les syllabes non initiales ne présentent ni une répétition du ton précédent, ni une réalisation spéciale constante, ni une variation libre, mais varient en fonction du ton du lexème et de leur propre position par rapport au début et à la fin du mot, de manière à supporter une partie de la courbe caractéristique du ton du lexème. [Phonetically, non-initial syllables involve neither a repetition of the tone of the preceding syllable nor a special fixed realization nor free variation; but they vary depending on the tone of the lexeme and on their own position with respect to the beginning or the end of the word, so as to maintain their role in the characteristic curve of the tone of the lexeme.]

The suprasegmentals of this language constitute a system intermediate between omnisyllabic tone and pitch-accent—i.e. between languages which have a tonal contrast on all syllables and those where a single syllable is not enough for the development of a tonal contrast.

5.5. MARGINALLY TONAL AND TONELESS LANGUAGES

There are a host of ways in which a language may be marginally tonal. In Lotha Naga (based on my work in a one-semester Field Methods class in 1981–2), for

example, there are only about a half dozen pairs of utterances distinguished by tone (comparable in a sense to the few minimal tonal pairs one can find in a language like Swedish). Often there are phonetic pitch differences that a language can afford to ignore, since the conditioning factors for the differences (typically a contrast between voiced and voiceless initials) has not been lost. This is the case in certain dialects of Bwe Karen (Blimaw, Geba; see Henderson 1979), as well as in Naxi (which, unlike the Loloish languages proper, has not undergone a clear tonal split in checked syllables).

Finally, of course, there are many Tibeto-Burman languages that are not tonal at all, belonging to the Himalayish, Qiangic, and Kamarupan branches of the family. Yet all branches of the family have at least some tonal members—and all known Lolo-Burmese, Kachin-Nung, Karenic, and Baic languages are fully tonal.

6. Mono- versus polygenesis of tone in Sino-Tibetan and Tibeto-Burman

Benedict doubted at first that the Chinese tones could be related to those of Tibeto-Burman languages (Benedict 1948). By the time the *Conspectus* was published (1972: 197), he had changed his mind, claiming a basic correspondence between Middle Chinese *A tone (*píngshēng*) and Tone *1 of Proto-Lolo-Burmese on the one hand; and Chinese *B (*shǎngshēng*) and Tone *2 of Proto-Lolo-Burmese on the other, with additional data from Karenic and Nungish offering support. See Table 7, reproduced from Benedict (1972: 196).

Benedict considered Chinese *qùshēng* (Tone *C) as a ‘sandhi tone’, while others (Haudricourt, Pulleyblank) derive it from an *-s suffix. Tone *3 of Lolo-Burmese (marked by creaky phonation) has a different, probably prefixal origin. It should be noted that the ‘sandhi vs. suffix’ theories are not necessarily mutually exclusive. The source of the sandhi could well have been a suffixal morpheme of the shape *-s.

Benedict’s sweeping conclusions were arrived at by what he called the method of ‘teleoreconstruction’ (1976): using key bits of data to cut through the forest of complex and often contradictory information from individual languages. To his credit, however, he emphasized the fragmentary nature of the tonal information available on most Tibeto-Burman languages in the 1960s and 1970s, and the

TABLE 7. Suggested Sino-Tibetan tonal correspondences

	Karen	Burmese	Trung	Chinese
Tone *A	I (high) II (low)	level (‘tone 1’)	mid-falling	<i>píngshēng</i>
Tone *B	III (high) IV (low)	falling (‘tone 2’)	high-level	<i>shǎngshēng</i>

complexity of it all. Much subsequent work has been done at the level of Tibeto-Burman subgroups, trying to relate the proto-tone categories of one subgroup to another. So far these efforts have met with mixed success, and are susceptible of a variety of interpretations:

- (a) **KARENIC AND LOLO-BURMESE** The basic work on the tones of Proto-Karen was done by Haudricourt (1942–5, 1975), who ultimately reconstructed three proto-tones in non-checked syllables plus a single checked syllable-type.¹³ It is relatively straightforward to find fairly regular tonal correspondences between Proto-Karenic and Proto-Lolo-Burmese, though there are many problematic cases (see Benedict 1972: 150–2, 196, where Haudricourt's prior work is not mentioned). The apparent contradiction between Benedict's views on the genetic position of Karenic as being outside Tibeto-Burman proper, and the relative ease with which correspondences may be found between the tonal systems of Karenic and Proto-Lolo-Burmese, lead one to wonder whether this suspicious similarity is due to diffusion rather than descent from a common inherited system (Matisoff 1973a: 81).
- (b) **JINGPHO AND LOLO-BURMESE** My first attempt to relate the tone systems of Jingpho and Lolo-Burmese led to inconclusive results (Matisoff 1974). Although there does seem to be some correlation between the relatively rare Jingpho high tone /' / and Proto-Lolo-Burmese Tone *2, Lolo-Burmese correspondences to the other two open Jingpho tones are not regular. In stopped syllables, there is a fairly strong match between Lolo-Burmese *high-stopped and Jingpho low-stopped, though with many exceptions. My conclusion is that we are not justified in setting up a higher-order subgroup (facetiously called 'Jiburish') on the basis of tonal correspondences. In a later study using new data on the tone systems of Burmish languages other than Burmese (Achang, Atsi/Zaiwa, Maru/Langsu, Bola), I concluded that roughly the same tonogenetic mechanisms were at work in the checked syllables of all these languages, but that the details of the process were quite different from language to language, especially as concerns the tonogenetic effects of particular combinations of *prefix and *root-initial (Matisoff 1991b: 106–11). The tonal splits in Burmish, Loloish, Naxi, and Jingpho checked syllables were thus seen to be parallel independent developments.

¹³ Haudricourt's pioneering 1942–5 article was unfortunately ignored by Jones (1961) (who reconstructed two proto-tones along with three syllable-final laryngeal features) and Burling (1969) (who improved on Jones's system in many respects, but reconstructed a six-tone system all the way back to Proto-Karen without taking account of the secondary nature of the tonal split from the original *three-tone system caused by the 'mutations consonantiques' in syllable-initial position). See also Haudricourt (1961). Note that checked syllables behave exactly like syllables ending in a nasal or a vowel in tonal splits conditioned by initial consonants: i.e. if a language splits its three open tones */A B C/ into six /A₁ A₂ B₁ B₂ C₁ C₂/ because of e.g. a loss of voicing contrast in initial consonants, its checked tone *D should also split into D₁ D₂.

- (c) **KARENIC AND TAMANGIC** Attempts to correlate the two tones of Proto-Tamangic (§5.4) with the two primary unstopped tones of Karenic, and by implication with Benedict's putative Proto-Tibeto-Burman distinction between proto-tones *A and *B, have not met with success, since the correspondences appear to be random (Mazaudon 1974: 55, 1985).
- (d) **RECENT WORK FAVOURING MONOGENESIS** Based largely on data collected through original fieldwork in north-east India and west Burma, Weidert (1987) worked out the tonal correspondences among a number of Kuki-Chin-Naga languages to his satisfaction. He went on to compare this proto-system to the three phonation types of Chepang (Nepal), and felt he had discovered the proto-prosodic system for all of Tibeto-Burman: a three-way proto-contrast in phonation type (clear, breathy, creaky). Very recently, Ostapirat (1997) independently demonstrated by an internal reconstruction of the tonal system of Tiddim Chin that it could be related to the phonation types of Chepang. I have been surprised to find from Ostapirat's examples that there might even be a correlation between these tone classes and those of Lolo-Burmese.

It will be a long time before we will be able to resolve the arguments about the common origin vs. independent development of the infinitely various tone systems of Tibeto-Burman languages. A key complicating factor is the undeniable ease with which tone systems or phonational habits may be diffused across languages or language families in a 'tone-prone' linguistic area.

7. Tonogenetic parallels in South-East Asian languages: the Sinospheric *Tonbund*

In view of the extreme difficulties in establishing a unitary tone system for Proto-Tibeto-Burman or Proto-Sino-Tibetan, it is all the more striking that regular tone correspondences can be established for large chunks of the vocabulary which Chinese, Vietnamese, Tai, and Hmong-Mien have in common. This fact alone—that the tonal categories so closely correspond across these several different language families—constitutes the 'suspicious similarity' that leads one to invoke borrowing/contact rather than genetic relationship.

Already in 1924, Jean Przyluski assigned Vietnamese to the Mon-Khmer family in spite of its tonal correspondences to Chinese, and expressed strong doubt in general on the criteriality of tone for genetic relationship. This sentiment is echoed by Haudricourt (1954a: 207) in connection with Hmong-Mien:

'Le système tonal primitif [des langues Miao-Yao] comportait seulement 3 tons pour les mots terminés par une voyelle ou une consonne sonore, et un seul pour les mots à consonne sourde finale. Donc le même système que l'ancien chinois, le thai commun ou l'ancien vietnamien. *Ces similitudes phonologiques ne préjugent aucunement des parentés généalogiques des langues miao-yao, celles-ci ne peuvent être fondées que sur le vocabulaire.*

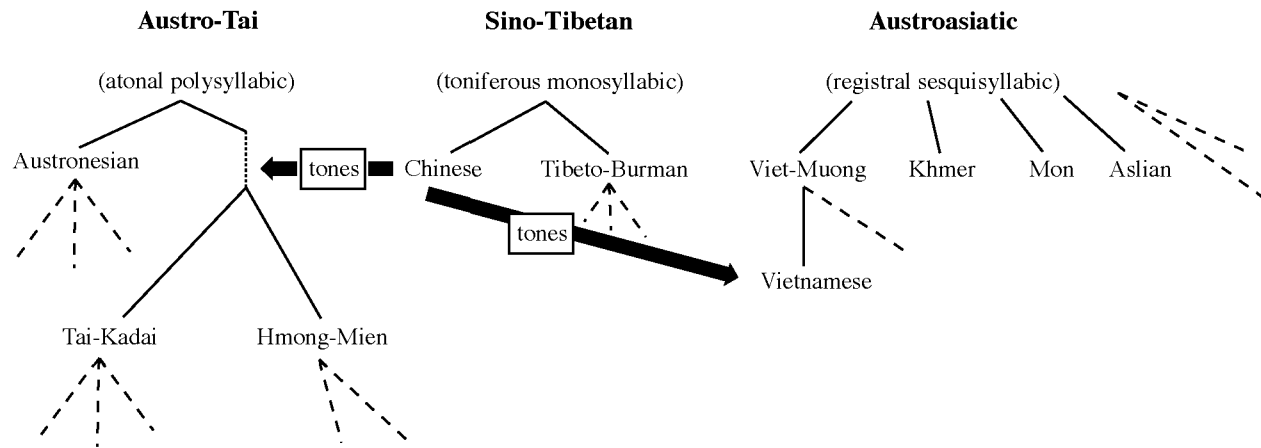


FIGURE 6. Chinese tonal influence on Tai-Kadai, Hmong-Mien, and Vietnamese

TABLE 8. Sino-Xenic tone correspondences (oldest stratum)

Chinese	<i>píngshēng</i>	<i>qùshēng</i>	<i>shǎngshēng</i>
Vietnamese	<i>ngang/huyền</i>	<i>hỏi/ngã</i>	<i>sắc/ngặng</i>
Tai	A (unmarked)	B (<i>máj-?èk</i>)	C (<i>máj-thoo</i>)

[The original tonal system [of the Miao-Yao languages] had only three tones for words ending in a vowel or in a voiced consonant, and just one tone for words ending in a voiceless consonant. It was the same system as Ancient Chinese, Common Tai and Ancient Vietnamese. *These phonological similarities cannot be used to establish genetic links for Miao-Yao languages, since these connections can only be established on the basis of shared vocabulary.*]

and has been noted repeatedly ever since (e.g. Downer 1963).

At least in the case of Vietnamese, everyone is now agreed that this is due to a relatively late diffusion of Chinese tonal categories into this Mon-Khmer language. If one adopts Benedict's 'Austro-Tai hypothesis' (an unsubstantiated example of megalocomparison), the diffusional explanation holds for Tai and Hmong-Mien as well: these branches of the originally atonal and disyllabic Austro-Tai stock became monosyllabic and tonal under Chinese influence, diverging from Austronesian, which remained atonal and disyllabic. See Figure 6 and Table 8.

Note that these tonal relationships are considered to hold only among borrowed or areal vocabulary items. Two different strata of Vietnamese–Chinese tonal correspondences may be distinguished, with a curious reversal of phonation types in different periods: (a) in the oldest period of loans from Chinese to Vietnamese (third to sixth centuries AD, as in Table 8), *qùshēng* words were borrowed as Vietnamese *hỏi/ngã*, while *shǎngshēng* words were borrowed as *sắc/ngặng*; (b) in the Sino-Vietnamese of around the tenth century, it was the opposite: *shǎngshēng* words were borrowed as *hỏi/ngã*, while *qùshēng* words were borrowed as *sắc/ngặng* (Haudricourt 1954a, Mazaudon 1974: 60). This seems to show that the basic phonational opposition was between clear (or unmarked) and marked. Sagart and Lee (1998) have recently demonstrated that two strata of Chinese loans into Bai can be distinguished on tonal grounds. For similar phonational oscillation between the 'marked' tones, cf. Burmese and Chinese (see §5.3).

7.1. VIETNAMESE AND THE OTHER MON-KHMER LANGUAGES:

FROM SESQUI- TO MONO-

Ironically, although Vietnamese is the Mon-Khmer language with the most speakers (over 60 million; Khmer only has 7 million, Mon 700,000), it is the least typical typologically. The tonogenetic process in Vietnamese went hand in hand with a change in its syllable structure from the typically Mon-Khmer sesquisyllabic to the Sinospheric monosyllabic. See Table 9 (Gage 1987).

TABLE 9. Vietnamese in Mon-Khmer perspective: from sesqui- to monosyllables

Vietnamese	Other Mon-Khmer	
<i>măng</i>	Old Mon <i>tbang</i>	'bamboo shoot'
<i>gấu</i>	Sedang (Bahnaric) <i>rokóu</i>	'bear' (n.)
<i>ngày</i>	Written Khmer <i>thngai</i>	'day'
<i>ngái</i>	Written Khmer <i>chngaai</i>	'far'
<i>kháp</i>	Bahnar <i>hokop</i>	'join'
<i>vuốt</i>	Old Mon <i>sumpot</i>	'rub'
<i>rắn</i>	Written Khmer <i>msa</i>	'snake'
<i>phủi</i>	Rongao (Bahnaric) <i>həpuih</i>	'sweep'
<i>rú</i>	Bru (Katuic) <i>brôu</i>	'wooded mountain'
<i>năm</i>	Written Khmer <i>chnam</i>	'year'

7.2. THE MON–KAREN–BURMESE PROSODIC COMPLEX

The Karenic peoples (now concentrated along the Burmese–Thai border) were among the first Tibeto–Burman groups to penetrate into what is now Burma. There is evidence to suggest that they stood in a servile relationship to the Mon, with whom they came down from the north at the same time (around the middle of the first millennium AD). The Mon attained a higher level of culture than other groups, absorbing elements of Indic civilization, including Buddhism, the concept of kingship, and a *devanāgarī* writing system. Later the ethnic Burmans made their way southwards, conquering another Tibeto–Burman people called the Pyu, and eventually establishing domination over all the varied ethnic groups of the region, including the Mon, in the process absorbing many Mon cultural traits.

The Karenic peoples seem to have been on the receiving end of linguistic influence throughout history. Almost alone among the overwhelmingly SOV (better: verb-final) Tibeto–Burman languages, Karenic has SVO word order. (The only other Tibeto–Burman group with SVO order is Baic (north-west Yunnan), which has been under enormous Chinese influence for millennia, with some dialects reported to have as high as 75% of their lexicon consisting of Chinese loanwords.) This was among the main reasons that led Benedict (1972: 3–6, 127 ff.) to consider Karenic to have split off from the other Tibeto–Burman languages at a very early date, and in fact to elevate Karenic to the status of a coordinate branch with 'Tibeto–Burman proper' in a larger 'Tibeto–Karen' family. Few people go along with this nowadays, attributing the anomalous Karenic word order to influence from Mon (and also perhaps from contiguous Tai languages).

We have also noted that the Proto–Karen tonal system correlates surprisingly well with that of Proto–Lolo–Burmese, despite the wide genetic distance between

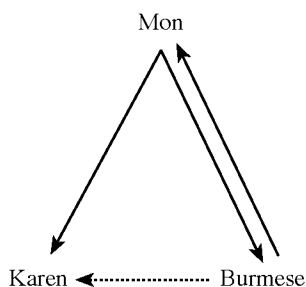


FIGURE 7. Interinfluence of Mon, Karen, and Burmese

these two Tibeto-Burman subgroups, as measured for example by percentage of shared inherited vocabulary (see §6).

To complete the triangle, it has been plausibly suggested (see Bradley 1982) that the phonation-prominent nature of the Burmese tone system (§5.3) is due to influence from Mon, which itself, in typical Mon-Khmer fashion, has a thorough-going phonational ('register') contrast between clear and breathy voice. I would add that the relatively high degree of sesquisyllabicity in Burmese, compared to the other Lolo-Burmese languages, is also due to influence from Mon's typically sesquisyllabic structure.

If all these assumptions are correct, Mon influenced Karenic (word order) and Burmese (phonational contrasts), while Burmese later influenced Karenic (tone system), as schematically shown in Figure 7.

In recent times the interethnic pecking order between Burman and Mon has changed drastically, as the Burmans have increasingly overwhelmed the Mons culturally and demographically, so that we must now add an arrowhead in the opposite direction. This is in keeping with what we observed in §3.1 about the multilateral vicissitudes of a true 'linguistic area'.

7.3. CHAMIC: POLY- TO SESQUI- AND POLY- TO MONO-

Perhaps the most amazing example of prosodic mutability in response to outside influences is Chamic, originally a typical polysyllabic Austronesian language group closely related to Acehnese. In the course of their migrations the Chams came into contact with monosyllabic languages on the island of Hainan, and developed a strictly monosyllabic, highly tonal dialect now known by their autonym, Utsat or Tsat. (These Muslim Chams are known as Huikui in Chinese.) The Cham dialects of those who settled in southern Vietnam and Cambodia, on the other hand, became sesquisyllabic and acquired phonational contrasts under influence from Khmer and other Mon-Khmer language-groups of Vietnam like Bahnaric. (See Haudricourt 1984, Edmondson and Gregerson 1993, Thurgood

TABLE 10. Typologically changed Chamic

Malay [polysyllabic]	Rade (Chamic of S. Vietnam) [sesquisyllabic]	Tsat (Cham of Hainan) [monosyllabic]	
<i>pinang</i>	<i>mnang</i>	<i>na:ŋ</i> ³³	'areca palm'
<i>tulang</i>	<i>klang</i>	<i>la:ŋ</i> ³³	'bone'
<i>lembu</i>	<i>êmô</i>	<i>mo</i> ³³	'cow'
<i>telinga</i>	<i>knga</i>	<i>ŋa</i> ³³	'ear'
<i>bulan</i>	<i>mlan</i>	<i>luŋ</i> ¹¹ (<i>phian</i> ¹¹)	'moon'
<i>jalan</i>	<i>êlan</i>	<i>la:n</i> ¹¹	'road'
<i>tasik</i>	<i>ksi</i> [?]	—	'sea'
<i>kulit</i>	<i>klit</i>	<i>liʔ</i> ²⁴	'skin'
<i>langit</i>	<i>êngit</i>	<i>ŋiʔ</i> ²⁴	'sky'
<i>semangat</i>	<i>mngat</i>	—	'soul'
<i>tali</i>	<i>klei</i>	<i>laβ</i> ³³	'string'
<i>ribu</i>	<i>êbau</i>	<i>phæ</i> ¹¹	'thousand'

1996.) The Chamic languages have thus undergone two degrees of reduction from their original disyllabic structure (poly- > sesqui- > mono-), as shown in Table 10.¹⁴

This vacillation in syllable-type, along with an under-appreciation of the role of diffusion in typological change, led earlier scholars like Schmidt (1906) to consider Cham to be a 'mixed language' (*Mischsprache*, or *langue mixte* in the French translation), intermediate between Malay and Austroasiatic.¹⁵

7.4. TONAL DIFFUSIONAL SCENARIOS, PAST AND PRESENT

The strange fact that tones and other prosodic features are eminently diffusible has manifested itself in all sorts of contact situations: from as far back as we can reconstruct to situations which are observable synchronically, before our very eyes.

¹⁴ Tsat data from Thurgood (1996); Rade/Malay comparisons from Ferlus (1996). Ferlus's interesting article gives several other sets of forms showing the passage from sesqui- to monosyllabism in other Mon-Khmer and Tai-Kadai languages, including Muong (sesqui-) vs. Vietnamese (mono-); Laven (sesqui-) vs. Nyaheun (mono-) (both Vietic branch of MK); Old Mon (sesqui-) vs. Modern Mon (still sesqui- but approaching mono-); Proto-Kam-Sui (Tai-Kadai) (sesqui-) vs. Sui (mono-).

¹⁵ If anything in East or South-East Asia is a good candidate for *Mischsprache* status it is Japanese, which may well have both an Austronesian and a Korean-type (some would say Altaic) component. For conflicting views on this thorny subject, see Hinloopen-Labberton (1924); Miller (1971); Solomon (1974); Murayama Shichirō (1976, 1978), Kawamoto (1977–8), Benedict (1985a), Martin (1966, 1996), Serafim (1993).

(a) PREVIOUSLY NON-TONAL LANGUAGE BORROWS TONES ALONG WITH LEXICAL ITEMS

(i) Chinese > Vietnamese, Tai, Hmong-Mien

According to the classical account of Vietnamese tonogenesis (Haudricourt 1954b), Vietnamese acquired its tones initially through normal compensatory processes (making up for the loss of laryngeal finals and the loss of the voicing distinction in initial consonants), but this development must have been stimulated by the huge number of lexical loans from Chinese, whose original tonal categories were faithfully preserved in the newly tonal borrowing language.

If one believes in Benedict's 'Austro-Tai hypothesis', a similar explanation must be invoked for the originally non-tonal Tai-Kadai and Hmong-Mien families. See Figure 6.

(ii) Lao (Tai family) > Khmuic (Mon-Khmer)

Much more recently, it has been reported that the Khmuic language known as 'U' has acquired a simple tone system through the transphonologization of a vowel-length contrast, stimulated by the tonal ambience created by Lao and other coterritorial tonal languages (see Svantesson 1988).

(b) TONAL LANGUAGE B BORROWS FROM ANOTHER TONAL LANGUAGE A THAT HAS HIGHER PRESTIGE A 'culturally recessive' tonal language may add new tones to its system to accommodate loanwords from a more prestigious tonal language. A spectacular example is furnished by the tonal virtuosi who speak the Punu (= Bunu) dialect of Mien, which already had eight tones of its own in native syllables, but has added three more new tones exclusively for loans from Zhuang (Tai) and more recently from Chinese (Mao and Chou 1962: 243).

Similarly, a lower prestige language may violate one of its own prosodic/phonotactic constraints in order to accommodate borrowed material. Among many examples which could be cited:

- (i) The Samsao dialect of Mien permits -VV? (glottal stop after a long vowel) only in loans from Chinese: *tòo?* [low tone] 'read'; *hòo?* [low tone] 'study'.
- (ii) Thai high tone occurs on long checked syllables only in loans from English (Gandour 1979: 96).

(c) TONAL LANGUAGE ASSIGNS TONES TO BORROWINGS FROM TONELESS LANGUAGES This is a situation analogous to the assignment of gender to loanwords from a language that lacks the category (e.g. the treatment of English nouns as masculine or feminine by German/Norwegian immigrants to the United States). At least two distinct strategies may be employed by a tonal language in order to accommodate loans from a non-tonal language.

(i) The most common, most 'unmarked' native tone may be used:

Khmer, Indic, and Austronesian loans with vocalic or nasal finals were almost all borrowed into Proto-Tai under Proto-Tai tone *A (Benedict

1942: 598, Gedney 1946, Gandour 1979). This led Gedney to conclude that 'tone *A was the normal level tone, with tones *B and *C so markedly different from it as never to be used in pronouncing syllables of words borrowed from a toneless language'. (For studies of the way Thai assigns tones to loanwords from Malay, see Court (1975); for the tonal treatment of English loanwords into Thai, see Bickner 1980; Gandour 1979.)

- (ii) The rarest native tone may be used, in order to avoid homophony with indigenous lexical items:

The rarest of the Lahu non-stopped tones is low level (11), marked with a macron, which derives only from Proto-Lolo-Burmese Tone *2 words with *glottalized or *voiceless sibilant initials. This is the tone of choice for the relatively few recent loans from English: *lālī* 'truck; lorry'; *kāmīti* 'committee'.

The rare Cantonese high-rising tone (35) is used mostly for loans from English (Kiu 1977).

8. Theoretical implications and desiderata for the future

Is syllable-type really predictive of tonogenetic possibilities, or is there nothing more than a rough correlation between, for example, monosyllabicity and tone-proneness? Is there a necessary connection between sesquisyllabicity and the birth of phonational systems ('registrogenesis')? Even though phonation has reached its fullest development in the sesquisyllabic Mon-Khmer family, not all Tibeto-Burman sesquisyllabic languages are phonational (e.g. Jingpho, §5.2), and some are phonational but predominantly monosyllabic (e.g. Burmese, §5.3).

It is high time to attempt a world-wide typology of tone systems, broad enough to encompass African and Mesoamerican prosodic systems as well as those of East and South-East Asia. Which typological traits are independent, and which are interrelated? Is it universally true that the functional load of tone contrasts is in inverse proportion to consonantal degeneration? Can we find languages with rich inventories of both initial and final consonants that also have complex tonal systems?

Can we ever reconstruct the phonetics of proto-tone systems? How stable are phonation types through time? Are tone and phonation really different aspects of one and the same phenomenon? Does one have logical primacy over the other, or is that a chicken-and-egg question? Are the principles of tono- and registro-genesis everywhere the same?

To what can we ascribe the surprising diffusibility of prosodic features? It seems to me that part of the answer lies in the *perceptual salience* of the rise and fall of the human voice, as well as of that mysterious entity known as the *basis of articulation* (French *base d'articulation*), i.e. the habitual setting of the articulators in the pronunciation of a particular language. Ontogenetically and impressionistically speaking, the first linguistic feature babies seem to acquire is the intonational

pattern of the language they hear around them. Even before they are capable of articulating consonants, babies pass through an adorable phase of vocalizing nonsense syllables with perfectly native intonation, amusing their adult friends who are likely to say, 'Just listen to him—it sounds like he's making a speech!' Contrariwise, a native-like intonation is usually the very last feature to be mastered by adult learners of a foreign language.

Still, one might well seek some more precise explanation for any especially rapid diffusion of prosodic traits throughout a linguistic area. It is a remarkable fact that a tremendous spate of tonogenetic and registrogenic activity occurred all over the South-East Asian linguistic area in the twelfth and thirteenth centuries, triggered by the devoicing of the previously *voiced series of obstruents in many Middle Chinese and Hmong-Mien dialects, in Siamese and other Tai languages, in Karenic, in Burmese and many Loloish languages, and in Vietnamese, Khmer, and other Mon-Khmer languages. It is interesting to note that this period was roughly contemporaneous with the Mongol invasions that convulsed Eurasia in those centuries. Is it going too far to regard these extralinguistic events as a sort of *punctuation* in the sense of Dixon (1997), a period of upheaval that shook up a previously stable prosodic constellation in South-East Asia? Could the peoples of the region have been so terrified by the Golden Hordes that they hardly dared to vibrate their vocal cords, dooming the *voiced obstruents to transphonologize into mere breathy voice or lower tone?

While we need not take the details of this terrible explanation too seriously, it is not at all implausible to invoke some sort of indirect correlation between linguistic evolution and events in the extralinguistic world. One need only cite the incalculable effects the Norman Conquest had on English; or the fact that the great Xixia (Tangut) civilization of the Gansu-Tibet borderlands was utterly wiped out by the Mongols, leaving its complex logographic writing system an eternal puzzle.

Yet explanations for historical events can never be simple. Who is to say which linguistic changes are due to purely internal pressures, as opposed to those which are due to language contact, or even to extralinguistic events?

References

- BENEDICT, P. K. 1942. 'Thai, Kadai, and Indonesian: a new alignment in Southeastern Asia', *American Anthropologist*, n.s. 44.576–601.
- 1948. 'Tonal systems in Southeast Asia', *Journal of the American Oriental Society* 68.184–91.
- 1972. *Sino-Tibetan: A Conspectus*. Cambridge: Cambridge University Press.
- 1975. *Austro-Thai Language and Culture, with a Glossary of Roots*. New Haven: Human Relations Area Files Press.
- 1976. 'Austro-Thai and Austroasiatic: the Austro-linkage', pp. 1–36 of *Austroasiatic Studies*, part I, edited by P. N. Jenner, L. C. Thompson, and S. Starosta. Honolulu: University of Hawaii Press.

- BENEDICT, P. K. 1985a. *Toppakô: Tônan Azia no gengo kara Nihongo e* [A breakthrough: from the languages of East Asia to Japanese.] Tokyo: CAAAL, Monograph Series 10.
- 1990. *Japanese/Austro-Tai*. Ann Arbor: Karoma.
- BICKNER, R. J. 1980. 'Thai tones and English loanwords: a proposed explanation'. Paper presented at the 13th International Conference on Sino-Tibetan Languages and Linguistics at the University of Virginia.
- BRADLEY, D. 1982. 'Register in Burmese', pp. 117–32 of *Papers in Southeast Asian linguistics* 8. Canberra: Pacific Linguistics.
- BURLING, R. 1969. 'Proto-Kâren: a reanalysis'. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics* 1.1–116.
- CAUGHLEY, R. 1972. *A Vocabulary of the Chepang Language*. Kirtipur, Nepal: Summer Institute of Linguistics and Tribhuvan University.
- CHANG, K. and SHEFTS CHANG, B. 1968. 'Vowel harmony in spoken Lhasa Tibetan', *Bulletin of the Institute of History and Philology, Academia Sinica (Taipei)* 40.53–124.
- COURT, C. 1975. 'The segmental and suprasegmental representation of Malay loanwords in Satun Thai: a description with historical remarks', pp. 67–88 of *Studies in Tai linguistics in Honor of William J. Gedney*, edited by J. G. Harris and J. R. Chamberlain. Bangkok: Central Institute of English Language.
- DAI QINGXIA, *et al.* 1983. *Jing-Han cidian* [Jingpho-Chinese Dictionary]. Kunming: Yunnan People's Publishing Co.
- DIFFLOTH, G. 1993. 'The lexical evidence for Austric, so far'. Paper presented at Conference on Austronesian-Mainland Asian Connections, Honolulu.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DOWNER, G. B. 1963. 'Chinese, Thai, and Miao-Yao', pp. 133–9 of *Linguistic Comparison in South-East Asia and the Pacific*, edited by H. L. Shorto. London: School of Oriental and African Studies.
- EDMONDSON, J. A., and GREGERSON, K. J. 1993. 'Western Cham as a register language', pp. 61–73 of *Tonality in Austronesian languages*, edited by J. A. Edmondson and K. J. Gregerson. Honolulu: University of Hawaii Press.
- FERLUS, M. 1996. 'Evolution vers le monosyllabisme dans quelques langues de l'Asie du Sud-Est'. Talk presented at the Société de Linguistique de Paris, 23 Nov.
- FORREST, R. A. D. 1962. 'The linguistic position of Rong (Lepcha)'. *Journal of the American Oriental Society* 82.331–5.
- GAGE, W. 1987. 'Vietnamese in Mon-Khmer perspective', pp. 493–524 of *Southeast Asian Studies Presented to André-G. Haudricourt*, edited by S. Ratanakul, D. Thomas, and S. Premrirat. Bangkok: Mahidol University.
- GANDOUR, J. 1979. 'Tonal rules for English loanwords in Thai', pp. 94–105 of *Studies in Tai and Mon-Khmer Phonetics and Phonology in Honour of E. J. A. Henderson*, edited by T. L. Thongkum *et al.* Bangkok: Chulalongkorn University Press.
- GEDNEY, W. J. 1946. 'Indic Loanwords in Spoken Thai'. Ph.D. dissertation, Yale University.
- GREGERSON, K. J. 1973. 'Tongue-root and register in Mon-Khmer', pp. 323–69 of *Austroasiatic Studies*, vol. 1, edited by P. N. Jenner, L. C. Thompson, and S. Starosta. Honolulu: University of Hawaii Press.
- HANSON, O. 1906. *A Dictionary of the Kachin Language*. Rangoon: Baptist Board of Publications.
- HAUDRICOURT, A.-G. 1942–5. 'Restitution du karen commun', *Bulletin de la Société de Linguistique de Paris* 42.103–11.

- 1954a. 'Introduction à la phonologie historique des langues miao-yao', *Bulletin de l'École Française d'Extrême-Orient* 44.555–74.
- 1954b. 'De l'origine des tons en vietnamien', *Journal Asiatique* 242.68–82.
- 1961. 'Bipartition et tripartition des systèmes de tons dans quelques langues d'Extrême-Orient', *Bulletin de la Société de Linguistique de Paris* 56.163–80.
- 1975. 'Le système de tons du karen commun', *Bulletin de la Société de Linguistique de Paris* 70.339–43.
- 1984. 'Hainan-dao jizhong yuyan de shengdiao' [The tonality of the languages of Hainan], *Minzu Yuwen* 4.17–25.
- HENDERSON, E. J. A. 1979. 'Bwe Karen as a two-tone language? An enquiry into the interrelations of pitch, tone and initial consonant', pp. 301–26 of *Southeast Asian Linguistic Studies*, edited by N. D. Liem. Canberra: Pacific Linguistics.
- HINLOOPEN-LABBERTON, D. V. 1924. 'Preliminary results of researches into the original relationship between the Nipponese and the Malayo-Polynesian languages', *Journal of the Polynesian Society* 33.244–80.
- HODGSON, B. H. 1880. *Miscellaneous Essays Relating to Indian Subjects*. 2 vols. London: Trübner.
- JONES, R. B. 1961. *Karen Linguistic Studies: Description, Comparison, and Texts*. Berkeley: University of California Press.
- KAWAMOTO, T. 1977–8. 'Towards a comparative Japanese-Austronesian', *Bulletin of the Nara University of Education* 26.23–49 (part I), 27.1–24 (part II).
- KIU, K. L. 1977. 'Tonal rules for English loanwords in Cantonese', *Journal of the International Phonetic Association* 7.17–22.
- LI, C. N. and THOMPSON, S. A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.
- MAO TSUNG-WU, and CHOU TSU-YAO. 1962. 'A brief description of the Yao language' (in Chinese), *Zhongguo Yuwen* 113.141–8. (Reprinted in English translation as pp. 239–55 of *Miao and Yao Linguistic Studies: Selected Articles in Chinese*, translated by Chang Yü-hung and Chu Kwo-ray. Ithaca, NY: Cornell University Southeast Asia Program, 1972.)
- MARAN, L. 1971. *Burmese and Jingpho: A Study of Tonal Linguistic Processes*. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics, vol. 4. Urbana, IL.
- MARTIN, S. 1966. 'Lexical evidence linking Korean with Japanese', *Language* 43.185–251.
- 1996. *Consonant Lenition in Korean and the Macro-Altaic Question*. Honolulu: University of Hawaii Press.
- MATISOFF, J. A. 1970. 'Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case study', *Journal of the American Oriental Society* 90. 3–44.
- 1972. *The Loloish Tonal Split Revisited*. Research Monograph 7. Berkeley: University of California, Center for South and Southeast Asia Studies.
- 1973a. 'Tonogenesis in Southeast Asia', pp. 71–95 of *Consonant Types and Tone*, edited by L. M. Hyman. Los Angeles: UCLA.
- 1973b. *The Grammar of Lahu*. University of California Publications in Linguistics 75. Berkeley: University of California Press.
- 1974. 'The tones of Jinghpaw and Lolo-Burmese: common origin vs. independent development', *Acta Linguistica Hafniensia* 15.153–212.
- 1976. 'Austro-Thai and Sino-Tibetan: an examination of body-part contact relationships', pp. 256–89 of *Genetic Relationship, Diffusion and Typological Similarities of East*

- and *Southeast Asian Languages*, edited by M. J. Hashimoto. Tokyo: Japan Society for the Promotion of Science.
- MATISOFF, J. A. 1978. *Variational Semantics in Tibeto-Burman: The 'Organic' Approach to Linguistic Comparison*. Philadelphia: Institute for the Study of Human Issues.
- 1979. 'Problems and progress in Lolo-Burmese: quo vadimus?', *LTBA* 4: 2.11–43.
- 1986. 'Hearts and minds in Southeast Asian languages and English: an essay in the comparative lexical semantics of psycho-collocations', *Cahiers de Linguistique Asie Orientale* 15.5–57.
- 1988. 'Universal semantics and allofamic identification—two Sino-Tibetan case studies: STRAIGHT / FLAT / FULL and PROPERTY / LIVESTOCK / TALENT', pp. 3–14 of *Languages and History in East Asia: Festschrift for Tatsuo Nishida on the Occasion of his 60th Birthday*, edited by A. Sato. Kyoto: Shokado.
- 1990a. 'On megalocomparison', *Language* 66.106–20.
- 1990b. 'Etymologizing monosyllables: headaches in the compilation of the *Sino-Tibetan Etymological Dictionary and Thesaurus*'. Paper presented at Summer Meeting of the Society for the Study of the Indigenous Languages of the Americas, Vancouver, 15 Aug.
- 1991a. 'Areal and universal dimensions of grammaticization in Lahu', pp. 383–453 of *Approaches to Grammaticalization*, vol 2, edited by E. C. Traugott and B. Heine. Amsterdam and Philadelphia: John Benjamins.
- 1991b. 'Jiburish revisited: tonal splits and heterogenesis in Burmo-Naxi-Lolo checked syllables', *Acta Orientalia* 52.91–114.
- 1991c. 'Endangered languages of mainland Southeast Asia', pp. 189–228 of *Endangered Languages*, edited by R. H. Robins and E. M. Uhlenbeck. Oxford: Berg.
- 1994a. 'Regularity and variation in Sino-Tibetan', pp. 36–58 of *Current Issues in Sino-Tibetan Linguistics*, edited by H. Kitamura, T. Nishida, and Y. Nagano. Osaka: 22nd International Congress on Sino-Tibetan Language and Linguistics.
- 1997. *Sino-Tibetan Numeral Systems: Prefixes, Protoforms, and Problems*. Canberra: Pacific Linguistics.
- MAZAUDON, M. 1973. *Phonologie tamang*. Paris: Société des Etudes Linguistiques et Anthropologiques de France.
- 1974. 'Tibeto-Burman tonogenetics', *LTBA* 3:2.1–123.
- 1985. 'Proto-Tibeto-Burman as a two-tone language? Some evidence from Proto-Tamang and Proto-Karen', pp. 201–29 of *Linguistics of the Sino-Tibetan Area*, edited by G. Thurgood, J. Matisoff, and D. Bradley. Canberra: Pacific Linguistics.
- MICHAILOVSKY, B. 1988. *La langue Hayu*. Paris: Editions du Centre National de la Recherche Scientifique.
- MILLER, R. A. 1971. *Japanese and the Other Altaic Languages*. Chicago: University of Chicago Press.
- MURAYAMA, S. 1976. 'The Malayo-Polynesian component in the Japanese language', *Journal of Japanese Studies* 2.413–36.
- 1978. *Nihongo keitō no tankyū* [An investigation into the lineage of Japanese]. Tokyo: Taishūkan Shoten.
- NIEDERER, B. 1998. *Les langues Hmong-Mjen (Miáo-Yáo): phonologie historique*. Munich: Lincom Europa.
- OSTAPIRAT, W. 1997. 'Tiddim Chin tones in historical perspective', *LTBA* 21:1.235–48.
- PRZYLUCKI, J. 1924. 'Les langues austroasiatiques', pp. 385–403 of *Les langues du monde*, edited by A. Meillet and M. Cohen. Paris: E. Champion.

- SAGART, L. 1990. 'Chinese and Austronesian are genetically related!' Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, at Arlington, Texas.
- 1993a. 'Chinese and Austronesian: evidence for a genetic relationship', *JCL* 21.1–62.
- 1993b. 'Old Chinese and Proto-Austronesian'. Paper presented at the Conference on Austronesian–Mainland Asian Connections, Honolulu.
- and LEE Y.-J. 1998. 'The strata of Bai'. Paper presented at the 31st International Conference on Sino-Tibetan Languages and Linguistics, at the University of Lund, Sweden.
- SCHMIDT, W. 1906. 'Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens', *Archiv für Anthropologie (Braunschweig)*, n.s. 5.59–109.
- SERAFIM, L. A. 1993. 'Comments on Austronesian–Japanese hypothesis'. Paper presented at the Conference on Austronesian–Mainland Asian Connections, University of Hawaii.
- SHEETS, B. 1968. 'Glottalization and the falling tone in Lhasa Tibetan'. Monthly Internal Memorandum, Phonology Laboratory, University of California, Berkeley.
- SOLOMON, J. R. 1974. 'Japanese and the other Austronesian languages: a survey of the literature'. Paper presented at Austronesian Conference, Honolulu.
- SPRIGG, R. K. 1966. 'Lepcha and Balti Tibetan: tonal or non-tonal languages?', *Asia Major*, n.s. 12.185–201.
- STRECKER, D. 1987. 'The Hmong-Mien languages', *LTBA* 10:2.1–11.
- SVANTESSON, J.-O. 1988. 'U'. *LTBA* 11:1.64–133.
- TEO KOK SEONG. 1993. 'A Sociolinguistic Description of the Peranakan Chinese of Kelantan, Malaysia'. Ph.D. dissertation, University of California, Berkeley.
- THOMASON, S. G., and KAUFMAN, T. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- THURGOOD, G. 1996. 'Language contact and the directionality of internal drift: the development of tones and registers in Chamic', *Language* 72.1–31.
- VARASARIN, U. 1975. *Les éléments khmers dans la formation de la langue siamoise*. Paris: Société d'Etudes Linguistiques et Anthropologiques de France.
- VOVIN, A. 1993. 'Is Japanese related to Austronesian?' Paper presented at Conference on Austronesian–Mainland Asian Connections, University of Hawaii.
- WEIDERT, A. 1987. *Tibeto-Burman Tonology*. Amsterdam: John Benjamins.

Language Contact and Areal Diffusion in Sinitic Languages

Hilary Chappell

This analysis includes a description of language-contact phenomena such as stratification, hybridization, and convergence for Sinitic languages. It also presents typologically unusual grammatical features for Sinitic such as double-patient constructions, negative existential constructions and agentive adversative passives, while tracing the development of complementizers and diminutives and demarcating the extent of their use across Sinitic and the Sinospheric zone. Both these kinds of data are then used to explore the issue of the adequacy of the comparative method to model linguistic relationships inside and outside the Sinitic family. It is argued that any adequate explanation of language family formation and development needs to take into account these different kinds of evidence (or counter-evidence) in modelling genetic relationships.

In §1 the application of the comparative method to Chinese is reviewed, closely followed by a brief description of the typological features of Sinitic languages in §2. The main body of this chapter is contained in two final sections: §3 discusses three main outcomes of language contact, while §4 investigates morphosyntactic features that evoke either the north–south divide in Sinitic or areal diffusion of certain features in South-East and East Asia as opposed to grammaticalization pathways that are cross-linguistically common.

1. The comparative method and reconstruction of Sinitic

In Chinese historical phonology, various methods have been applied with relative success to the Sinitic family in the reconstruction of both stages of Middle and Old Chinese. In *Études sur la phonologie chinoise* (1915–26), Karlgren published his

I would like to thank to the following colleagues for comments and critique: Sasha Aikhenvald, Tim Curnow, Bob Dixon, Nick Enfield, Geoffrey Haig, Bernd Heine, Tania Kuteva, Randy LaPolla, Jim Matisoff, Alain Peyraube, Laurent Sagart and participants at the workshop on ‘The connection between areal diffusion and the genetic model of language relationship’ held at the Research Centre for Linguistic Typology at the Australian National University in August 1998.

This research forms part of an Australian Research Council Large Grant project ‘A semantic typology of complex syntactic constructions in Sinitic languages’ (1997–9).

ground-breaking reconstruction of Middle Chinese according to three main sources: an analysis of rhyme tables based on the early seventh century dictionary *Qiyùn* (CE 610), Sinoxenic readings from Japanese and Vietnamese, and data from nineteen dialects which he collected while carrying out fieldwork in China from 1910 to 1912. Strictly speaking, he did not apply the comparative method to these dialect data but determined the phonological system of Middle Chinese on the basis of the *Qiyùn*, interpreting and assigning phonetic values to the rhyme categories.¹ Note that the *Qiyùn* dictionary was compiled as a guide to the correct pronunciation for the recitation of the classics. Hence, its precise relation to the spoken language of its time is not transparent. Many scholars believe that it is based on several different spoken dialects of the time and not just that of the capital, Chang'an (present-day Xi'an), while others believe it reflects educated speech from the sixth century CE, that is, the end of the Nanbeichao dynasty (Northern and Southern dynasties, 420–589 CE).

Karlgren later worked on the reconstruction of Old Chinese based on his Middle Chinese reconstruction in conjunction with an analysis of the rhyme categories of the *Shījīng* [Book of Odes] and the information which could be deduced from the phonetic components present in most Chinese written characters. Old Chinese hypothetically reflects the elevated speech of the late Zhou period of fifth to third centuries BCE, in the view of some scholars, or the even earlier period of the Western Zhou in the view of others (roughly the first half of the first millennium BCE). These are not, however, uncontroversial issues, for which a fuller discussion may be found in Sagart (1999) or for a contrary view, in Baxter (1992).

The *Shījīng* is an anthology of poems from 1000–500 BCE, compiled in the sixth century BCE. An early observation made by scholars in China was that characters which rhymed in it generally contained the same phonetic element. Karlgren's contribution was similarly to interpret and assign values to the categories of initials and finals in the *Shījīng* (Book of Odes) which would obey regular phonetic laws for development into those he had earlier posited for Middle Chinese. Karlgren's second reconstruction was published in 1940 as *Grammata Serica* with a revised version appearing in 1957. Given the lack of records of real dialect materials from the late Nanbeichao and Sui periods to which Middle Chinese roughly corresponds, the reconstruction of Old Chinese could not avoid being the more hypothetical of the two. Karlgren's postulation of these two earlier stages of the Chinese language inspired further work by sinologists resulting in revisions and new breakthroughs, and provided indisputable evidence for the genetic relationship of Sinitic languages, albeit mainly on the basis of phonology and the lexicon.² Nonetheless, the focus on phonetic laws and the use of the

¹ I am indebted to Laurent Sagart for this clarification of Karlgren's approach.

² For more discussion of the reconstructions for either Old or Middle Chinese, see Norman (1988), also Baxter (1992) and Sagart (1999, 2001).

Neogrammarian approach with its assumption of homogeneous data in Chinese linguistic reconstruction was early criticized by Grootaers (1943) and Serruys (1943) as the sole means of relating dialects to Old and Middle Chinese. In particular, they both objected to Karlgren's use of character lists for elicitation and dialect dictionaries based on the reading of standard Chinese characters. The reading lists not only required literate language informants but could also hardly avoid producing the literary pronunciations which by definition hold a close relationship to the standard language, Mandarin, and thus neatly supported his reconstruction (see also §3 on stratification). In many cases, these pronunciations represented morphemes not used at all in the local *patois* which belong to the purely colloquial level.

In the same study, Grootaers (1943) shows how methods in geographical linguistics can be successfully applied to capturing dialect isoglosses in Northern Chinese for both the innovation and extent of use of phonetic and lexical features, based on 'real' colloquial items. Similarly, Hashimoto (1992) pioneered the use of *Wellentheorie* (wave theory) in Chinese linguistics to account for the spread of tonal categories and phonetic features such as retention or loss of voicing in Chinese dialects. The use of lexical and morphological data has also been incorporated in various handbooks produced by Beijing University in the 1960s such as *Hànyǔ fāngyán cíhuì* [A lexical list for Chinese dialects] and *Hànyǔ fāngyán gàiyào* [An outline of Chinese dialects] compiled by Yuan (1989) which includes syntactic data. More recently the inutility of the family-tree model to explain how languages develop in a relatively stable environment is raised by Hashimoto (1992: 32) for Hakka and by Dixon (1997) for the general case.

In sections 3 and 4 which follow, it is argued that the family-tree model, used alone, is inadequate to capture the complexities of linguistic phenomena created during the course of evolution and geographical distribution of a language family: the comparative method and the family-tree model simply cannot account for all the facets associated with language change and development, and to be fair were never intended to do so. They need to be used in conjunction with other methods to account for the effects of language contact such as stratification, hybridization, and convergence, not to mention other possible outcomes such as mixed languages and language obsolescence.

2. Typological features of Sinitic

Sinitic languages form a sister group with the Tibeto-Burman languages of the Sino-Tibetan language family located in East and South-east Asia. As a language family, Sinitic languages are as diverse as the Romance or Germanic languages within the Indo-European family. The spoken forms of Chinese languages are not mutually intelligible: a speaker of Suzhouese, a Wu dialect, will not understand a compatriot from Quanzhou, who speaks a Southern Min dialect. Even within dialect groups such as Min or Yue there is a high degree of mutual unintelligibil-

ity between subdivisions such as Coastal versus Inland Min, or one of the Guangxi Yue dialects versus Hong Kong Cantonese Yue.

Typologically, Sinitic languages are tonal languages which show analytic or isolating features, though in some Min languages, for example, the development of case markers and complementizers from lexical verbs, and the use of a range of nominal suffixes, has moved further along the path of grammaticalization than in Mandarin. Complex allomorphy is also widespread in Min dialects, exemplified by the many variant forms for each negative marker in Fuzhouese (North-Eastern Min) and for the diminutive suffix in Southern Min.

Tone sandhi (or tone change) can be used to code morphological functions in Chinese languages. For example, in Toishan Cantonese, aspectual distinctions such as the perfective and the plural form of pronouns can be signalled in this way. Tone sandhi phenomena are, however, most conspicuous in the Min and Wu dialect groups where citation- or juncture-forms for each syllable differ from contextualized forms. Although Sinitic languages have SVO basic word order, object preposing is a common contrastive device and postverbal intransitive subjects are common in presentative constructions. The modifier generally precedes the modified element. This means that subordinate or backgrounding clauses typically precede main clauses while attributives precede head nouns and adverbs precede verbs. Well-known exceptions to this rule are presented by the case of gender affixes on animal terms and certain semantic classes of nominal compounds and adverbs in many Southern Sinitic languages.

The ten major Sinitic languages (or Chinese dialect groups) that are generally recognized are listed below:

I.	Northern Chinese (Mandarin)	北方話
II.	Xiang	湘
III.	Gan	贛
IV.	Wu	吳
V.	Min	閩
VI.	Kejia or Hakka	客家
VII.	Yue dialects	粵
VIII.	Jin dialects	晉
IX.	Hui dialects	徽
X.	Pinghua	

Mandarin covers the largest expanse of territory from Manchuria in the north-east of China to Yunnan and Sichuan provinces in the south-west. Apart from the Jin dialects, the eight other dialect groups fall neatly into almost complementary geographical distribution with Mandarin, covering the east and south-east of China: Xiang dialects are largely concentrated in Hunan province, Gan in Jiangxi, Wu in southern Jiangsu and Zhejiang provinces, Min in Fujian, Kejia in north-eastern Guangdong, south-western Fujian and parts of Jiangxi and Sichuan provinces, Yue in both Guangdong and Guangxi provinces, Hui dialects in



MAP. Chinese dialects



southern Anhui and adjacent areas of Jiangxi and western Zhejiang provinces, and the Pinghua dialects in Guangxi. The Jin dialects in Shanxi province and Inner Mongolia represent the only non-migrant dialect group to be found in northern China, apart from Mandarin. The reader is referred to the map for the exact locations.

2.1. A NOTE ON CHINESE DIALECT HISTORY

According to Bellwood (this volume), archaeological evidence points to Neolithic settlements in two areas of modern China—the middle and lower Huang He (Yellow River) and the Yangzi River valleys. These can be dated to around 7000 BCE. However, reconstruction of Proto-Chinese, based on the diversity found in modern dialects, cannot hope to reach much further back than the first millennium BCE (see §1).

Overall, the development of Sinitic languages over the last two and a half millennia can be aptly modelled in terms of its history of imperialist unification and expansion accompanied by ensuing periods of relative equilibrium. These were in turn regularly punctuated by periods of disunity and temporary fragmentation of the Chinese empire. During the formation time of the Sinitic group, the major migrations of the Han Chinese took place from northern China to various regions in the south, for which a detailed coverage of population movements in China over the last several millennia is provided in LaPolla (this volume) while a brief history of Chinese dialects is given in Chappell (2001c) and thus is not recapitulated here.

The general consensus regarding the approximate time of diversification of Chinese into the present-day dialect groups is around the time of Medieval Chinese during the Sui (581–618) and early Tang dynasties (618–907) for Yue, Xiang, and Gan but earlier, during the transitional period for the Han dynasty (206 BCE–220 CE) for the ancestral language(s) of Wu and Min. Sagart (1988, 2001) and You (1992: 97) claim that Wu, Xiang, Yue, and Gan developed directly from earlier stages of Northern Chinese whereas Min was probably a secondary development from a Southern Sinitic language such as Wu (or Proto-Wu-Min), and Hakka, similarly, a secondary development from Southern Gan during the Tang period. Ting (1983) and Norman (1988: 189) do not entirely concur with this view regarding Min, holding that there is a strong demarcation line between Wu and Min linguistic territory, but agree on the early split. The larger dialect picture for Sinitic languages was thus essentially in place by the end of the Southern Song (1127–1279), apart from the later formation of the Hui dialects by the early Ming dynasty (1368–1644).

Sagart aptly describes dialect groups as ‘fuzzy entities that owe (as) much (of) their make-up to contact as opposed to vertical inheritance’ (1997: 298–9). He further argues for the difficulty of using isoglosses to determine dialect boundaries given that innovations may be obliterated or reversed through contact, with

the result that the family-tree model is only strictly applicable to rarer situations where diversification and loss of contact co-occur, as for Austronesian, concurring with Dixon (1997). The history of Sinitic languages certainly presents a case in point, exemplifying the difficulties that could arise if the family-tree model and comparative method were exclusively used to represent genetic relationships. The implication is that a fuller description of the evolution of Sinitic languages necessarily involves modelling genetic relatedness as well as the characteristics of *Mischsprachen*, 'mixed languages', (see Heine and Kuteva, this volume) combining substratum or superstratum features of 'step-parent' contact languages (Dixon 1997: 71). These, in their turn, can be either genetically related or unrelated which has further typological ramifications. Next I consider some aspects of areal diffusion in the South-east and East Asian region before beginning on the main discussion.

2.2. AREAL DIFFUSION

Mantaro Hashimoto has convincingly argued for a north-south divide for Chinese languages on the basis of phonological, lexical, and syntactic evidence (see Hashimoto 1974, 1976a, 1976b, 1986). His thesis essentially has the following argumentation: Chinese languages are sandwiched between Altaic languages in the north and Tai languages in the south, with the typogeographical consequence of Altaicization of northern Chinese varieties and Taiization of Southern Sinitic. Furthermore, he observes that the north-south opposition can be clearly perceived in features such as the increasing number of classifiers, tones, and consonantal endings to syllables, not to mention the monosyllabic nature of morphemes as one moves southwards. He notes that some varieties of Northern Chinese show agglutinative tendencies, witnessed in the existence of a postposition for accusative/dative case in Qinghai Mandarin, stress-accent dominance over tone, and adoption of O-V structures as in North-Western Mandarin dialects spoken in Qinghai and Gansu provinces. Other broad divisions are the typically MODIFIER-MODIFIED word order in the north versus MODIFIED-MODIFIER order for some structures in the south; different comparative strategies; different word orders for the 'double object' or ditransitive construction; and aspect and tense distinctions maintained in the south while merged in the north.

To this could be added the more limited use of patient-marking or disposal constructions where the direct object is positioned before the main verb and preceded by a special marker, for example, the extensively researched *bǎ* 把 construction in Mandarin: S – *bǎ* – O – V. In its canonical form, it codes a highly transitive event that affects a referential object with a specifiable effect or result state. Cheung (1992) has shown that Cantonese, which uses the Medieval Chinese exponent *jeung*¹ 將 [*jiāng*], is restricted to transitive verbs, whereas Mandarin also allows its use with intransitive verbs provided there is a causative interpretation (see Chappell 1992a). Furthermore, the use of *jeung*¹ is more a feature of formal discourse than colloquial Cantonese, evidence of Mandarin influence. Similarly,

Hakka also reportedly uses this construction much less frequently than Mandarin (Yuan 1989).

Bisang (1996) presents a typology of classifiers according to their functions in South-East and East Asian languages, showing a similar set of geographical correlations with respect to enumeration, referentialization, and other parameters. In Cantonese, for example, classifiers may also be used as possessive and relative clause markers, thus showing a greater alliance with Tai languages as opposed to Northern Chinese which does not permit this function.

With regard to Northern Chinese, Hashimoto (1986: 95) suggests that a pidgin Chinese developed when Altaic peoples became Sinicized, and that while they adopted Chinese lexicon and morphology they retained the syntax of Altaic, and possibly its phonetic system as well. This must be a two-step process, however: presumably what is meant by Altaicization follows on as the next step after cultural Sinicization, whereby the superstrate Altaic syntactic structures slowly diffuse into the different varieties of Northern Chinese and then gradually southwards into other Sinitic languages by virtue of the prestige of Mandarin. He observes that this is not unique to northern Chinese: the Ong-Bê language of south-western China, a Tai language, has undergone the same process of Sinicization (1986: 95), as too pre-war Korea with respect to the effect of Japanese on Korean.

Matisoff (1991: 386, this volume) refines Hashimoto's basic classification by dividing the larger South-East Asian zone into two main areas: the Sinospheric and the non-Sinospheric. The Sinospheric area includes Southern Sinitic (basically Sinitic languages south of the Yangzi) and the language families which have been in close cultural contact with China such as Hmong-Mien, Tai-Kadai, Vietnamese in the Mon-Khmer branch of Austroasiatic, and certain branches of Tibeto-Burman such as Lolo-Burmese. The non-Sinospheric languages include Austronesian languages, many Mon-Khmer languages, and Tibeto-Burman languages, for example those found in north-eastern India and Nepal.

According to Matisoff (1991) some of the broad grammatical features which unify the South-East Asian area into a linguistic zone are the following:

- (a) development of modal verbs > desiderative markers, 'be likely to'
- (b) development of verbs meaning 'to dwell' > progressive aspect markers
- (c) development of verbs meaning 'to finish' > perfective aspect markers
- (d) development of verbs meaning 'to get, obtain' > 'manage', 'able to', 'have to'
- (e) development of verbs of giving > causative and benefactive markers
- (f) development of verbs of saying > complementizers, topic and conditional markers
- (g) formation of resultative and directional compound verbs through verb concatenation.

With respect to Sinitic, all of these pathways of grammaticalization apply to Northern Chinese as well, with the exception of a 'say' verb developing into a complementizer and the limited use of 'give' with a causative meaning. Both these

paths of grammaticalization are treated in §4 for Southern Sinitic languages while other pathways, such as for *get* verbs, are analysed in depth in Enfield (Chapter 10). Next, I discuss some linguistic phenomena that are the result of language contact, illustrating some of the potential difficulties for modelling the outcomes of language contact including stratification, metatypy, hybridization, and convergence.

3. Language contact: stratification, hybridization, and convergence

Synchronically, there are three main outcomes of language contact situations for Sinitic languages: stratification, hybridization, and convergence. Examples of all three outcomes are discussed in this main section. Stratification and hybridization of syntactic and morphosyntactic forms are a widespread phenomenon in Sinitic languages.

3.1. STRATIFICATION

Stratification has resulted from the systematic introduction of certain features of the prestige language in China for the purposes of reciting classical texts; or as forms borrowed from this standard language (different varieties of Mandarin). Moreover, this has occurred more than once in the historical development of several of the major Chinese dialect groups such as Min which has three such layerings from Northern Chinese: the Han dynasty stratum (206 BCE–220 CE); the Nanbeichao stratum (420–581 CE) and the late Tang stratum (eighth to tenth centuries). The degree of stratification varies along a continuum from minor phonological differences, as in Hakka, to major stratification of the lexicon and a marked contrast between the literary and colloquial pronunciations as in Southern Min. The differences in pronunciation are known as *wén-bái yì-dú* 文白異讀 in Chinese linguistics. The *bái* or vernacular pronunciation for each syllable in a given dialect represents the native morpheme which may or may not have a *wén* or reading doublet whose pronunciation has been adopted from Northern Chinese.

For example, in the Xiamen or Amoy dialect of Southern Min, words in the reading pronunciation which end in a velar nasal often have a nasalized vowel in the cognate colloquial form: the character for ‘name’, 名, has the literary form *bêng* versus colloquial *miáⁿ*. In other cases the relationship is not so straightforward: the preposition ‘to, with’ written as 共 has *kā* as its colloquial pronunciation but *kiōng* as its reading pronunciation, with the latter closer to the modern standard Mandarin /kuŋ/¹ in form. Similarly, the possessive morpheme 其 has *ê* for its colloquial pronunciation but *kí* for its literary one, closer to Mandarin /tɕʰi/². In many cases, it first needs to be established whether there is any cognacy at all. There clearly is none for the suppletive relationship between these possessive morphemes, nor for the two readings of the diminutive suffix 仔 which has *á* for the colloquial as opposed to *tsú* for the literary. Again, the reading form resembles modern standard Mandarin very closely, which is /tsɿ/³. As argued below, the diminutive suffix has evolved from another morpheme for ‘son’ in Min: *kiáⁿ*.

Most non-Mandarin Sinitic languages show this kind of phonological and lexical stratification as a result of different periods of intense contact with Mandarin, particularly with the emergence of an official court language in the mid- to late-Tang period (eighth to tenth centuries CE), a koine based on the language of the capital, Chang'an, where a north-western dialect of Northern Chinese was spoken. This was brought to southern regions during the migrations of the later Tang dynasty and is the basis of the reading or literary pronunciation in most Southern Sinitic languages, as noted above. In some dialect groups, a second overlay of a more eastern variety of Northern Chinese occurred after the establishment of the Liao (937–1125 CE), Jin (1115–1234 CE) and Yuan dynasties (1271–1368 CE) in northern China, whose capitals were located in the region of Beijing. It is significant that both koines are associated with flourishing vernacular literatures (Norman 1988) and the strong tendency to standardize language use that accompanies the consolidation of an imperial system of government. More traditional research has mainly concentrated on describing the phonological correspondences between the reading and colloquial pronunciations of characters. Recent pioneering work on syntax by Zhu Dexi (1990) and Anne Yue-Hashimoto (1991) has uncovered several different strata for the syntax of interrogative forms in Southern Sinitic (see §3.3). For the purposes of any kind of comparative work, the native stratum must first be clearly separated from the imported stratum.

3.2. LEXICAL AND MORPHOLOGICAL STRATIFICATION

Lien's study of morphological change in Taiwanese Southern Min (2001) shows that this historical process of layering has resulted in different kinds of stratificational distinctions in the lexicon for the native colloquial morphemes versus the 'alien' literary forms. Taiwanese Southern Min belongs to the subdivision of Coastal Min and is closely related to the Xiamen (Amoy), Quanzhou, and Zhangzhou dialects spoken on the south coast of Fujian province. It is the first language of over 73% of the population in Taiwan, despite the fact that Mandarin is the official language.

As Lien observes, since this variation is present in everyday colloquial language, it cannot simply be explained as the existence of separate registers resulting from the impact of Mandarin on Southern Min during the Tang period. He discusses cases of morphological competition which have been synchronically resolved in favour of either the colloquial or literary stratum and concludes on the basis of his data that the diffusion is clearly bidirectional.

For example, the morpheme *lâng* 儂 'person' represents the first type where this colloquial form is in the ascendant over the literary and unproductive bound morpheme *jîn* 人 which also means 'person' but was borrowed from the Tang Northern Chinese koine. It is not cognate with *lâng*. Couplets thus exist, such as *toā lâng* 大儂 'adult' versus *tāi jîn* 大人 'police officer' (a polite vocative akin to 'Sir'), where both are formed with morphemes for 'big' + 'person'. This is

indicative, Lien argues, of *jîn* developing a special idiomatic meaning in many of its compounds. The literary morpheme *jîn* generally occurs with less frequency as a suffix than *lâng*, according to a statistical count made by Lien. It is much less likely to occur affixed to disyllabic stems, and never with those from the colloquial stratum. Furthermore, in coining new words, he notes, the younger generation prefers the native morpheme *lâng*.

Similarly, for numerals, the colloquial forms are used for cardinal numbers while the literary forms are used for giving telephone numbers and for calendar years in the Gregorian or western calendar. Lien observes, however, that in the case of ordinal numbers, the colloquial forms are winning out from the lexeme 'third' upwards. He attributes this outcome to the lack of literacy in the native language, Taiwanese Southern Min, as opposed to high literacy in the official language, Mandarin: it is nowadays rare for younger generation first-language speakers of Taiwanese to be instructed in the reading pronunciations and forms of Southern Min.

The second type, where the literary form is more productive than the colloquial form, is represented by suffixes which are in complementary distribution such as colloquial *ke* versus literary *ka* (which share the etymon for 'family' 家). These are used as agentive suffixes or nominalizers but, significantly, in different semantic fields: the first, colloquial form *ke* shows a broader application as it is used not only for family relationships but also for those pertaining to the old agrarian society such as 'head-servant' and 'master' and names for relatives-in-law while the second, literary form *ka* applies to higher status professions of the new industrialized society such as 'writer', 'connoisseur', 'diplomat', 'statesperson'. Nonetheless, colloquial *ke* has become 'inert' and unproductive.

Similarly, colloquial *sai-hū* versus literary *su* act as agentive suffixes, the first referring to trades and crafts that require manual labour, while the second refers to professions that require intellectual skills. This is shown in Tables 1 and 2, reproduced from Lien (2001).

TABLE 1. Derivatives with colloquial suffix *sai-hū* 師傅 in Southern Min

Agent noun	Gloss	Translation
<i>thô'-chúi sai-hū</i> 塗水師傅	mud-water-master	'bricklayer'
<i>chúi-tiēn sai-hū</i> 水電師傅	water-electricity-master	'electrician/plumber'
<i>iū-chhat sai-hū</i> 油漆師傅	oil-paint-master	'painter'
<i>bak-chhiūⁿ sai-hū</i> 木匠師傅	wood-wright-master	'carpenter'

TABLE 2. Derivatives with literary suffix *su* 師

Agent noun	Gloss	Translation
<i>i-su</i> 醫師	treat medically-master	'doctor'
<i>kâu-su</i> 教師	teach-master	'teacher'
<i>ūiūi-su</i> 畫師	draw-master	'artist'
<i>káng-su</i> 講師	talk-master	'instructor'

Both these cases contrast with the outcome for the competition between morphemes for 'person' in that the literary form is very productive, and a clear semantic division of labour is apparent. Lien characterizes the colloquial stratum as typified by basic and popular vocabulary, versus the technical and cultural vocabulary representative of the literary stratum. Despite this mixing and integration of the literary stratum into everyday language, convergence of the two strata is not likely, particularly where the semantic specialization of the two sets has occurred, as for *ke* and *ka* and *sai-hu* and *su*. Lien concludes that only a bi-directional diffusion of features can explain the continuing coexistence of these strata.

3.3. SYNTACTIC STRATIFICATION: PREVERBAL INTERROGATIVE MARKERS

Zhu (1990) and Yue-Hashimoto (1991) discuss the complementary distribution in Sinitic languages of neutral interrogative constructions using the Northern Sinitic strategy of VP-NEG-VP as opposed to Southern Sinitic constructions using either a preverbal interrogative adverb (ADV-VP) or a VP-NEG-(PRT) form for this type of yes/no question. These interrogatives are described as neutral in terms of any presupposition concerning the response. The type which uses the ADV-VP form is found in some Southern Min and Wu dialects but also in certain South-Western and Lower Yangzi Mandarin dialects of Anhui province, while the VP-NEG-(PRT) form is characteristic of Hakka and Yue dialects.

Yue-Hashimoto is able to pinpoint different strata for these interrogative structures by comparing several colloquial Southern Min texts from the Ming and Qing dynasties (dating from the sixteenth century onwards) written in the Chaozhou and Quanzhou dialects. Her analysis of these texts enables her to resolve apparent counter-examples where certain Min dialects possess all three strategies described above and thus seem to belie this basic Northern versus Southern distinction. She argues that the ADV-VP form using the adverbial interrogatives *kě* 可 or *qí* 豈 belongs to a residual premodern colloquial stratum found in certain Southern Min dialects such as Yilan in Taiwan and Shantou (Swatow) in north-eastern Guangdong province, China. This contrasts with the form of VP-NEG-(PRT) which has been in use over many centuries and represents a standard

and native Southern Min stratum, while VP-NEG-VP represents the non-native stratum which has been borrowed from Northern Chinese. Further comparisons with non-Sinitic languages are made: the ADV-VP form is commonly found in Tibeto-Burman while the VP-NEG form is typical of Kam-Tai, though languages in both families show use of the VP-NEG-VP strategy which overall appears to have the widest distribution in Sino-Tibetan, presumably through diffusion.

3.4. SYNTACTIC HYBRIDS AND METATYPY

Another consequence of language contact is the mixing or hybridization of syntactic forms. There are many clear-cut cases of this in Sinitic languages where native and borrowed syntactic strategies are eclectically combined into the one new form. This is quite distinct from the situation known as metatypy (Ross 1996) where the syntactic configuration for a construction is borrowed from the prestige language entailing the calquing of its grammatical exponents by the appropriate morphemes. When metatypy occurs, it may replace the native strategy (if there is one—see §4.3 on complementizers) or it may be used side by side with this native form, possibly in different speech levels or registers. Hong Kong Cantonese shows an unusual case of retention of the native form, in combination with metatypy and hybridization for the relative-clause construction which I next examine.

Matthews and Yip (2001) have coined the useful term of *ditaxia* which refers to the parallel use of two syntactic structures in different registers. This lays the basis for analysing a third peculiar construction for the relative clause which has made a recent appearance in Hong Kong Cantonese. The two main relative-clause structures can be thus described: colloquial Cantonese employs classifiers as relative markers as in (1) while formal Cantonese employs a structure using the possessive *ge*³ as in (2) which mirrors the use of Mandarin *de* as a relativizer. Compare the following two examples:

(1) Colloquial Cantonese: Relative Clause + DET + CL + HEAD NP

佢	唱	嗰	首	歌	好	好	聽
<i>Koei</i> ⁵	<i>coeng</i> ³	<i>go</i> ²	<i>sau</i> ²	<i>go</i> ¹	<i>hou</i> ²	<i>hou</i> ²	<i>teng</i> ¹
3sg	sing	that	CL	song	very	good	listen

‘the song she sings is very nice’

(2) Formal Cantonese: Relative Clause + GEN + HEAD NP

佢	唱	嘅	歌
<i>Koei</i> ⁵	<i>coeng</i> ³	<i>ge</i> ³	<i>go</i> ¹
3sg	sing	PRT	song

‘the song(s) she sings’

Typologically, the relational, including possessive, use of the classifier in colloquial Cantonese given in (1) is characteristic of other southern Chinese dialect

groups such as Southern Min but also of Tai and Hmong-Mien languages, showing further evidence of the affinity among the Sinospheric languages (see Bisang 1992). The construction in (2) is an example of metatypy based on the prestige language, Mandarin. A third and innovative construction represents a hybridization of these two, where both the classifier and *ge*³ are present with the form [DET + CL + GEN (= *ge*³) + N]:

(3) Hybridization: Relative Clause + DET + CL + GEN + HEAD NP

佢	唱	嗰	首	嘅	歌
<i>Koei</i> ⁵	<i>coeng</i> ³	<i>go</i> ²	<i>sau</i> ²	<i>ge</i> ³	<i>go</i> ¹
3sg	sing	that	CL	PRT	song
'the song she sings'					

At this point, a reasonable surmise might be that such examples of Cantonese show a lack of mastery over the newer Mandarinized form of the relative-clause structure. It is interesting to learn, however, that the hybrid relative clause construction tends to be used in more formal and public registers such as broadcasting and sermons, and is therefore classified as pseudo-High in register by Matthews and Yip. Possibly it serves a double purpose: on the one hand it has an emblematic status for Cantonese speakers—it can be used to show linguistic solidarity and Cantonese identity by retaining the classifier as a marker of the relative clause—yet on the other hand speakers retain the use of ‘posh’ Cantonese by means of the counterpart of the Mandarin relative clause, with use of the genitive marker *ge*³ (see Aikhenvald, this volume, on the topic of emblematicity). An explanation involving syntactic hypercorrection does not appear to be relevant in this case.

A similar phenomenon can be observed in both Taiwanese Southern Min and Hakka for the comparative construction where the native strategy using an adverb ‘more’ is combined with the cognate for Mandarin *bǐ* 比 ‘compare’ (see Ansaldo 1999 on this kind of double marking). Zhu (1990) also examines a hybrid structure for neutral yes/no questions where an adverbial interrogative marker is used together with a VP-NEG-VP form. This is found in some Lower Yangzi Mandarin dialects, in the Suzhou dialect (Wu) and in the Shantou dialect (Southern Min) (see also §3.3). Similarly, Chappell (1992b and 2001b) notes hybridization for the evidential (or experiential aspect) marker in Taiwanese Southern Min, where the native strategy of a preverbal marker *bat* 別 from the verb ‘know’ is combined with the verb enclitic *koè*, calqued on Mandarin *guò* 過 ‘cross, pass through’.

3.5. THE PLASTIC COMMON LANGUAGE

Wu (1992) describes a variety of Changsha Mandarin called *sùliào* or ‘plastic’ *pǔtōnghuà* in which convergence is taking place between the local Xiang dialect and the official language, *pǔtōnghuà*. *Pǔtōnghuà*, literally ‘the common language’, is based on the pronunciation of educated speakers of the prestige dialect of

Beijing Mandarin in combination with the vocabulary and grammar of model works of vernacular literature written in Northern Chinese dialects. This definition was promulgated for the official language of China in 1955 (Chen 1999: 25). Speakers prefer to use Changsha Xiang but in official and formal situations they are encouraged to use *pǔtōnghuà*. Although the convergence is unidirectional — in the direction of Mandarin — it is far from complete.

When speakers accommodate to *pǔtōnghuà*, a language over which they may not have full command, a special tone correspondence is set up which neither belongs to the Changsha Xiang dialect nor to *pǔtōnghuà*, yet symbolizes that speakers have adopted an official speech level which is as close as they can possibly come to *pǔtōnghuà*. Even when non-standard lexical items are used, specific to the Xiang dialect, or speakers are unable to distinguish velar from alveolar nasal endings, let alone retroflexes from dental sibilants (as they should in standard Mandarin), the mere fact that they are using this special tone correspondence suffices for their speech to be considered ‘official’, that is, as plastic *pǔtōnghuà*.

By way of contrast, if speakers use the right lexicon and grammar for *pǔtōnghuà* but retain their own Changsha Xiang tone pattern, their speech remains irredeemably Changsha Xiang. The reason is as follows: first, it needs to be noted that Changsha Xiang has seven tones, whereas both plastic *pǔtōnghuà* and ‘real’ *pǔtōnghuà* have only four. Wu (1992: 137–8) explains how the correspondences between the Middle Chinese sources for the modern tones in standard Mandarin and colloquial Changsha Xiang differ. Changsha speakers base their rules for conversion of Xiang tones into plastic *pǔtōnghuà* on the historical relationships for their *own* dialect with Middle Chinese. It is this local interpretation which has created the special tone correspondences that act as a marker of plastic *pǔtōnghuà*.

In the final section, I examine the outcomes of language change: are pathways of grammaticalization triggered by a certain set of typological preconditions in the given language; is it due to areal diffusion of a morphosyntactic feature or, more broadly, merely attributable to common language universals of grammatical change?

4. Shared grammaticalization pathways in Sinitic, areal diffusion, and language universals

In this section, I examine five sets of data in Sinitic: the source of the diminutive suffix, the feature of negative existential verbs ‘there is not / there are not’, the development of complementizers from verbs of saying, adversative passives, and some constructions which express inalienable possession. Some of these phenomena unify Sinitic as a family while others bear witness to the grouping of languages in the South-East and East Asian zone as a *Sprachbund* or linguistic area. In this section, the attempt is made to distinguish which features represent a pathway of grammaticalization that is cross-linguistically unremarkable, which are the result

of areal diffusion, and which could be seen as special typological features of Sinitic languages.

4.1. EARLY SOUTHERN MIN DIALECT GRAMMAR AND EVIDENCE FOR GRAMMATICALIZATION: THE DIMINUTIVE

Early seventeenth-century texts on Southern Min dialects provide an invaluable source for the diachronic study of the grammar of their modern counterparts in that they are largely written in the special dialect characters for vernacular Hokkien. Below, I compare the diminutive of modern Southern Min dialects such as Taiwanese and Amoy (Xiamen) with those found in the *Arte de la lengua Chiō Chiu* (1620), a grammar on the same type of dialect written in Spanish.³

In Sinitic languages, the diminutive has its source in various morphemes for 'son' which may have 'child' as a secondary meaning. A morpheme for 'child' is the common source crosslinguistically for diminutives (see Heine, Claudi, and Hünemeyer 1991: 79–88, 1993: 38). For example, Mandarin uses the suffix 兒 *ér* 'son' while Cantonese employs tone sandhi, changing the citation tone to high rising tone, the cheshirization of an earlier segmental morpheme meaning 'son'. Cheshirization refers to the attrition of segmental phonemes, which leave a mere trace of their former phonetic substance, such as the tone.⁴ In Taiwanese Southern Min, the diminutive is formed with the suffix *-á*. It can be related to the lexeme for 'son', 子 *kiaⁿ*, used in the *Arte* (1620: 2b, 11a, 12b) and to *kiáⁿ* 'son' in contemporary Taiwanese and Amoy, for which the character 囡 is used as well.⁵ Note that the stem of the word used for 'child' in the *Arte*—簡仔 *kín nia* (1620: 15) or 囡仔 *gín-á* ~ *gín-ná* in contemporary Taiwanese—cannot be the source for this diminutive on phonological grounds (see Lien 1998).

In the early seventeenth-century grammar of Southern Min, the following description is given for the diminutive (1620: 10):

³ This work was most likely a collaborative effort of Spanish Dominican missionaries and Chinese interpreters living in a Chinese Sangleys community near Manila in the late sixteenth and early seventeenth centuries. On phonological grounds, van der Loon identifies the dialect used in these manuscripts as the vernacular of Hai-cheng as spoken around the turn of the seventeenth century (1967: 132). He shows conclusively that it differed in certain phonological features from the dialect of Zhangzhou city, to which prefecture this harbour town belonged. It appears that the Sangleys or Chinese traders had migrated from this port in southern Fujian province during the late sixteenth century, with many eventually settling in and around Manila.

⁴ See Matisoff (1991) for more on 'cheshirization', to whom we owe the coining of this evocative term.

⁵ This morpheme *kiáⁿ* 'child, son' is in fact used to exemplify the tone category which is accompanied by nasalization, according to the missionaries' classification. Note that in the Spanish romanization *k-* is used interchangeably with *gu-* and *qu-* for the unaspirated voiceless velar plosive initial /k/, as seen in the diminutive forms given in (4). Furthermore, nasalization has not been marked for these diminutive forms, suggesting that it had already been lost at this stage, in contrast to its lexical use as 'child'.

(4) *Arte de la lengua Chiō Chiu* (1620)

'The diminutive is formed with the final particle *ia* or *nia* or *guia*:

kéiguià 圭仔 "little chicken" [*pollito*]

bôguia 帽仔 "little hat" [*sonbrerillo*]

tôguia 刀仔 "little knife" [*guedillito*]

In contemporary Taiwanese, the three corresponding words are *ke-á* 'chicken, little chicken'; *bô-á* 'hat' and *to-á* 'knife, small knife' respectively, indicating partial bleaching of the diminutive feature.⁶

I suggest that in this early grammar of Southern Min, the *Arte*, an incipient stage of development for the diminutive can be viewed, where its form can still be clearly related to the morpheme for 'son', unlike contemporary Southern Min where the form has atrophied to *-á* and can be used not only as a diminutive but also as a marker for the noun category:

(5) Taiwanese Southern Min:

一 張 桌仔 合 兩 張 椅仔

chit tè toh-á *kap n̄ng tè í-á*

one CL table-NOM and two CL chair-NOM

'a table and two chairs' (not: 'a small table and two small chairs')

It is interesting to find that the lexeme *kiáⁿ* can nonetheless still be used as a kind of suffix to mark the young of animal species, postposed after the reduced diminutive form used as a noun marker:

(6) 牛仔囡 狗仔囡

gû-á-kiáⁿ *káu-á-kiáⁿ*

ox-NOM-offspring dog-NOM-offspring

'calf' 'puppy'

Further support for the proposed grammaticalization pathway of 'son' > DIMINUTIVE comes from Yang (1991: 166) who points out that the diminutive suffix in the Chaozhou dialect of Southern Min retains the full form of *kiáⁿ*.

(7)

Chaozhou: *tiaⁿ kiáⁿ* contrasting with Xiamen, Zhangzhou, Taiwanese:

tiaⁿ-á.

'a small cooking pot'

Yang also quotes the Tang poet Gu Kuang 顧況 who annotates the character 囡, pronounced with an alveopalatal initial /tɕiǎn/ in modern Mandarin, as having the meaning 'son' in colloquial Min in §13 of his poem *Shàngǔ zhī shé*

⁶ Note that only one of the variants listed by the *Arte* is illustrated by the examples in (4). This is discussed further in Chappell (2000).

- (8) 囙 音 蹇， 閩 俗 呼 子 為 囙。
Jiǎn yīn jiǎn mǐn sù hū zǐ wéi jiǎn
(word) sound jian Min custom call son as jian
‘The sound of this character is *jiǎn*, the Min usually call “son” *jiǎn*.’

The more general case of semantic change from ‘child’ to diminutive morpheme is well attested in other languages of the world, for example, in Jurafsky (1996) and Heine *et al.* (1993: 38) while the use of diminutives with probable source morphemes in sex-specific ‘son’ is characteristic of Sinitic (for more data, see Huang 1996). The *Arte* provides the hard evidence for this semantic change into a diminutive suffix, affecting the morpheme ‘son’ in Southern Min (see also Chappell 2000). Given the widespread occurrence of the first type of conceptual shift cross-linguistically, I conclude that while this more semantically specific case may be a shared development in Sinitic languages, it only partially characterizes it typologically.

4.2. NEGATIVE EXISTENTIAL CONSTRUCTIONS: ‘THERE IS NO / THERE ARE NO’

Southern Sinitic languages display a large number of negative morphemes which can be used to negate propositions at clause level. Furthermore, the semantic space for negation is carved up by subtle modal and aspectual nuances. In particular, Southern Min languages show a highly differentiated set of negative adverbs, most being fused forms combining one of the first two negatives listed in Table 3 with various modal verbs and showing different degrees of bondedness.

In Sinitic, it is typically the marker used to negate perfective clauses which also has a fully verbal use meaning ‘there is no Y / there are no Y’ with one nominal argument. This set of verbs in Southern Sinitic can also occur in a transitive

TABLE 3. Taiwanese Southern Min negative markers

<i>bó+</i> V	無	Negation of perfective contexts, attributive predicates
<i>m̄+</i> V	唔	Negative marker for property verbs, imperfective contexts, and unwillingness to V
(<i>íá</i>) <i>bē+</i> V	(猶)未	Negation of expectation: ‘have not (yet) V-ed’
<i>boē+</i> V	勿會	Negation of ability/possibility to V: ‘unable to V’
<i>boài+</i> V	勿愛	Negation of perfective desiderative: ‘didn’t want to V’
<i>m̄mài+</i> V	唔愛	Negation of imperfective desiderative: ‘don’t want to V’
<i>mài+</i> V	莫	Negative imperative: ‘Don’t V!’
<i>m̄mó+</i> V	唔好	Negative hortative: ‘You shouldn’t V!’
<i>m̄bién+</i> V	唔免	Negation of necessity: ‘You don’t need to V’

syntactic frame as the negative possessive verb: 'X has no Y'. I describe the semantic and syntactic features of negative existential verbs in more detail in Chappell (1994) and observe that their prior lexical meaning is often 'lose' as exemplified by (9) for Cantonese 冇 *mo⁵* where the meaning is ambiguous between the two uses:

(9) Cantonese:

已經	冇	哩	個	嘅	權勢	啊
<i>yi⁵ging¹</i>	<i>mo⁵</i>	<i>lei⁵</i>	<i>goh³</i>	<i>ge³</i>	<i>kuen⁴sai³</i>	<i>a¹</i>
already	NEGV	this	CL	PRT	power	PRT

'(This prime minister) had already lost his power.' *or*:
'The prime minister no longer had any power.'

Standard Mandarin does not possess such a negative existential or negative possessive verb. It must use the negative perfective marker *méi* preposed before the verb *yǒu* 'there is', shown in (10).

(10) Mandarin:

沒	(有)	人	了
<i>méi</i>	(<i>yǒu</i>)	<i>rén</i>	<i>le</i>
NEG	(there:be)	person	CRS

'There's nobody here.'

Omission of *yǒu* 'there is' is possible but should not be confused with an analysis of *méi* as a monomorphemic negative existential verb (which it is not), since *yǒu* can always be added back in. It appears that the same situation applies in many Tibeto-Burman languages where a negative adverb or prefix beginning with *m-* is used (see Matisoff 1991: 388, 393–4), and also in Thai. In other words, these languages similarly do not have a special negative existential verb. Hence this is a Southern Sinitic feature, not attested in either Northern Chinese or evidently in the other half of the Sino-Tibetan language family. It is neither a Sinospheric typological feature nor a pan-Sinitic one. Nor is it well documented cross-linguistically, given that Payne (1985) discusses this type of negation for only a few Austronesian languages but does not include it as a negation type.

4.3. COMPLEMENTIZERS

In Taiwanese Southern Min, a complementizer similar in function to English *that* has grammaticalized out of the verb 'to say' *kóng* 講. Matisoff (1991: 398–400) describes this path of grammaticalization as an example of the general category of verbs developing into verb particles in South-East Asian languages, represented by Thai, Khmer, and Lahu. Like these three languages, the Southern Min verb 'say' is also used at the end of a non-final clause and before the intonation break to introduce the complement clause. It is not fully grammaticalized since it may be omitted. Moreover, it forms a kind of verb complex with the preceding matrix

verb which must belong to one of the following verb classes: speech act, cognition, or perception, and it directly introduces the embedded clause, as in (11):

(11) Taiwanese Southern Min:

遐	個	敵對	的	武將	共	笑	講
<i>Hia</i>	<i>ê</i>	<MC: <i>díduì</i> >	<i>ê</i>	<i>búchiòng</i>	<i>kā</i>	<i>chhiò</i>	<i>kóng</i> ,
that	CL	opposing	L	general	PRETR	laugh	SAY _{that}

這	是	號作	<J: 猴面	冠者>.
<i>che</i>	<i>sì</i>	<i>hō-tsò</i>	<J: <i>Sarumen</i>	<i>Kanja</i> >.
this	be	name.as	monkey.face	youngster

‘Those generals who opposed him mocked him (General Toyotomi) as the one who should be called “monkey-face boy”?’ (*Japanese tales* 629–30) (Note: MC = Mandarin Chinese insert; J = Japanese insert)

In this first stage of grammaticalization, when ‘say’ verbs are used as quotative markers, the lexical meaning is not completely bleached. Examples such as *chhiò kóng* could still be rendered as ‘laughed (at him) saying’ while in the second stage where *kóng* is used with cognitive verbs such as *siūⁿ* ‘think’, its literal meaning is less plausible: ‘think saying’. The putative path of development is outlined in Chappell (forthcoming e) in addition to other grammaticalized or partially grammaticalized uses of *kóng* as a metalinguistic marker of explanation; an evidential marker of hearsay; a component of a compound conditional marker; a topic introducer and as a clause-final marker of assertions and warnings. It has not yet developed a purposive function, which may indicate that certain of its several grammaticalization pathways are relatively ‘young’ (Bernd Heine, p.c.).

There has been only very little study of this phenomenon in typological work on Sinitic languages to date. In Chappell (forthcoming e), I show that this development has proceeded as far as the quotative stage in some Yue and Wu dialects and less far in standard Mandarin. For the Yue dialect of Cantonese, ample evidence can be found of the use of *wa⁶* ‘to speak’ in conversational and narrative texts where it functions as such a quotative marker with speech-act verbs. Note, however, that *wa⁶* does not form a verb complex with the preceding speech-act verb: this is clear in that it can be separated from the verb by a noun denoting the direct object:

(12) Cantonese:

贊	哩	個	男仔	話...
<i>jaan³</i>	<i>lei⁵</i>	<i>goh³</i>	<i>laam⁴jai²</i>	<i>wa⁶...</i>
praise	this	CL	young.man	say...

‘(she) praised this young man saying...’

Although a verb complex with ‘say’ as V_2 is not a possible strategy for introducing complement clauses in standard Beijing Mandarin, or *pǔtōnghuà* (as opposed to such a use for quotations), it is in the regional variety known as

Taiwanese Mandarin. It is striking that Taiwanese Mandarin does not choose the cognate verb for *kóng*, which is *jiǎng* in Mandarin, to create the new syntactic calque but instead makes use of its functional equivalent, the high frequency verb *shuō* 說, in the configuration SUBJECT-VERB₁-*shuō* + CLAUSE:

(13) Taiwanese Mandarin:

那	我	希望	說	這	個	願望
<i>nà</i>	<i>wǒ</i>	<i>xīwàng</i>	<i>shuō</i>	<i>zhèi</i>	<i>ge</i>	<i>yuànwàng</i>
CONJ	1sg	hope	SAY _{comp}	this	CL	wish

很	快	就	到	了
<i>hěn</i>	<i>kuài</i>	<i>jiù</i>	<i>dào</i>	<i>le</i>
very	quickly	then	arrive	PFV

‘So I hope that this wish will be realized very soon.’

(14) Beijing Mandarin:

*我	希望	說
* <i>wǒ</i>	<i>xīwàng</i>	<i>shuō</i>
1sg	hope	say

However, this does not provide supporting evidence just for the north-south divide for Sinitic languages: it appears that Sinitic is encircled by language families and language isolates (such as Japanese and Korean) that all possess complementizers which have developed from verbs of saying. This feature has been described in the relevant literature for individual languages belonging to Tibeto-Burman, Tai-Kadai, Hmong-Mien, Indic, Dravidian, and Altaic (see Matisoff 1991, Saxena 1988).

Since this semantic change is also cross-linguistically well attested (it occurs widely in various language families of Africa—see Frajzyngier 1996 for Chadic, Amberber 1995 for Amharic, Heine *et al.* 1991: 216–17, 246–7, Heine *et al.* 1993: 190–8 for a larger sample of languages), it seems that the grammaticalization of *kóng* into a complementizer in Taiwanese Southern Min is most likely a language-internal development. It has simply drawn on its own resources (Dixon 1997) to recreate a syntactic device which was in fact available in Classical and Middle Chinese, as attested in the written register.

Indeed, earlier periods of written Chinese made use of verbs of saying such as *yuē* 曰 (Classical Chinese) and *dào* 道 (Medieval Chinese) as quotative markers, although not as fully fledged complementizers (described in Chappell forthcoming e). This means that not only does Sinitic have its own inherited language-internal devices upon which to analogize but it also has access to patterns and processes which can be imitated from surrounding unrelated language families.

It seems that this has taken place in recent times for sister languages within Sinitic, the case in point being the calquing of the Taiwanese Southern Min complementizer into Taiwanese Mandarin. This is an unusual development in

terms of the direction of metatypy from a less prestigious to a more prestigious language, and note that there are many other examples of Taiwanese Southern Min constructions which have been borrowed into the Taiwanese variety of Mandarin (see Kubler 1985). This probably reflects linguistic creativity in transferring favoured syntactic forms and devices into Mandarin where gaps exist, rather than a negative description in terms of interference from the first language.

Further research on dialect materials would be in order to show irrefutable evidence for the view that the development of a complementizer in Taiwanese Southern Min is a purely independent innovation, triggered however by a combination of factors: a conducive environment in terms of areal typological features and the existence of appropriate language-internal characteristics.

Unlike the case for negative existential verbs, the existence of a complementizer in Southern Min and some Wu and Yue dialects tallies well with Matisoff's inclusion of Southern Sinitic in the South-East Asian linguistic area. The theoretical problem remains, however, of distinguishing between areal diffusion and a putative language universal for the development of complementizers from verbs of saying, given the right typological preconditions.

4.4. ADVERSATIVE PASSIVES

Matisoff (1991) points out that verbs of giving typically develop into causatives and benefactives in South-East Asian languages. In Southern Sinitic languages, verbs of giving are also used to form the passive construction. For example, most Hakka dialects use the high frequency verb *pun*⁴⁴ 'to give' as both the passive and the benefactive marker, while Cantonese does the same with *bei*² < 'give'.

A further characteristic feature of passives which unites Sinitic is that the colloquial forms are both adversative and agentful. This appears to be an unusual development for 'give' (compare this with data in Heine *et al.* 1993: 97–103). Such a description applies to standard Mandarin as well where only the *bèi* passive has an agentless form although it has lost its adversative feature in some contexts. Note that the *bèi* passive belongs to more formal discourse, in contrast to the agentive colloquial passives formed by *jiào* 'make' and *ràng* 'let' (see Chappell 1986).

Norman (1982: 245) observes that these two Northern Chinese passives formed with the causative verbs *jiào* 'make' and *ràng* 'let' are unique amongst Sinitic languages, as opposed to the use of verbs of giving. He argues that this is not an independent development in Mandarin but rather is due to Manchu superstrate influence on Chinese. In Manchu and other Altaic languages the same structure can be used for both passive and causative meanings. In support of this view, an earlier study by Hashimoto (1987: 46) contrasts standard Mandarin with Mandarin dialects on the periphery of the Northern Chinese zone which continue to use verbs of giving as passive markers. This suggests that 'give' verbs as passive markers are an older feature.

The adversative feature appears to be an areal feature as not only do South-East

Asian languages such as Thai and Vietnamese show this restriction, but also Japanese (see Shibatani 1994). Hence there are different allegiances for each of these features: some evince the north–south divide in Sinitic (verbs of giving versus causative verbs used as passive exponents), some are relevant to the South-East and East Asian area (the adversative feature), while this particular development for ‘give’ is possibly specific to Southern Sinitic within the Asian zone, and is quite rare cross-linguistically (Bernd Heine, p.c.).

4.5. POSSESSION

4.5.1. Pronominal systems and inalienable possession

In general there are no separate morphological classes for alienable and inalienable possession in Sinitic languages; nonetheless, there is a weaker reflection of this distinction in the fact that genitive marking is facultative for kin relationships as well as other important social relationships, body parts, and spatial orientation, particularly when the possessor is pronominal (see Chappell and Thompson (1992) on Mandarin genitives):

PRONOUN_{POSSESSOR} (genitive NOUN_{POSSESSED}
marker)

(15) Mandarin:

你	(的)	母親	先生	(的)	耳朵	裹
<i>nǐ</i>	(<i>de</i>)	<i>mǔqīn</i>	<i>xiānsheng</i>	(<i>de</i>)	<i>ěrdǔo</i>	<i>lǐ</i>
2sg	(GEN)	mother	teacher	(GEN)	ear	in
		‘your mother’			‘in the teacher’s ears’	

Hakka is unusual within Sinitic in having a special portmanteau genitive form for pronominal possessors which can be considered as a kind of case marker (Table 4). These special genitive forms are not generally used, however, with inanimate nouns such as ‘fountain pen’ in (17) but, again typically, with kin as in (16):

TABLE 4. Meixian Hakka pronouns

	Nom/Acc	Gen
1sg	<i>ŋa¹¹</i> 崖	<i>ŋa⁴⁴</i>
2sg	<i>ŋ¹¹</i> 你	<i>ŋia⁴⁴</i>
3sg	<i>ki¹¹</i> 佢	<i>kia⁴⁴</i>

(16) 我 老弟

ηa^{44} $lau^{31}-t'ai^{44}$
 1sg younger.brother
 'my younger brother'

With inanimate nouns, as in example (17), the genitive marker *ke* is used with the nominative/accusative form of the pronominal possessor:

(17) 我 嘅 鋼筆

ηai^{11} ke^{53} $kong^{53}pit^{11}$
 1sg GEN pen
 'my fountain pen' (* ηa^{44} $\emptyset$ $kong^{53}pit^{11}$)

This semi-covert distinction is reflected more clearly in syntax in the form of the double patient construction, discussed next.

4.5.2. Double-patient constructions

The double-patient construction is shared by all Sinitic languages. It is syntactically unusual in that its intransitive process verb appears to take two arguments, one more than the verb valency should allow, recalling the 'one-too-many-argument' problem described in Shibatani (1994). The two arguments of the intransitive verb designate possessor and possessum. Furthermore, the nouns in this possessive relationship occur non-contiguously and belong to different constituents. Specifically, the possessor appears in the canonical position for grammatical subject (S) clause-initially, while the possessum appears postverbally in the canonical object position (O). The verb must be a so-called 'unaccusative' non-volitional one such as 'go red', 'go white', 'limp', 'increase' (literally: 'become more'), 'fall out', or 'rot', which takes a semantic undergoer as its subject. An example of this construction from Cantonese is given with its structural formula:

Double patient construction:

NOUN_{POSSessor} VERB_{INTRANSITIVE} NOUN_{PART/KIN TERM}

(18) Cantonese Yue:

棵 樹 落 咗 好 多 葉
 Poh^1 sue^6 lok^6 joh^2 ho^2 doh^1 yip^6
 CL_{REF} tree fall PFV very many leaf
 'That tree has lost many leaves [more literally: The tree fell very many leaves].'

In Chappell (1999), I argue that the relationship of inalienable possession licenses the use of two arguments with an intransitive verb. It can only be used for part-whole relations and, in a more restricted fashion, for kin. While this construction is a shared feature of Sinitic, as with the study of complementizers,

it has not been extensively researched. The same situation applies for South-East Asian languages: it is not possible in Lahu (Matisoff, p.c.) but a similar construction appears to exist in Lao (Nicholas Enfield, p.c.). At this stage, it is difficult to determine if such a construction is typologically defining for Sinitic.

5. Conclusion

The family-tree model appears to work reasonably well for Sinitic as far as phonology and some aspects of morphology are concerned; nonetheless, this only accounts for a small part of a much more complex linguistic picture: the family-tree model is unable to capture the effect of successive waves of Mandarinization of Southern Sinitic languages, stratifying lexical and syntactic components as shown in §3 for nominal affixes in Southern Min and interrogative constructions in Southern Sinitic languages. Nor can it handle the cases where convergence is well under way with the Mandarinization of Changsha Xiang, albeit by means of an intermediate language known as *sùliào* or 'plastic' *pǔtōnghuà*. The initial stages of this process of convergence include widespread occurrence of metatypy and hybridization of syntactic forms in Sinitic, as illustrated by the example of Hong Kong Cantonese relative-clause constructions. Hence, a more delicate and subtle treatment of the question of genetic affiliation is needed.

Note that the processes of metatypy and convergence may not always be in the direction of the official language of prestige: in Taiwan, massive calquing and metatypy from Southern Min into Taiwanese Mandarin is taking place, as briefly described for the use of complementizers. It can be conjectured that this is because Southern Min, and not Mandarin, is emblematic of current loyalties and serves as a 'badge' of being Taiwanese. Such developments involving language contact cannot be easily captured in terms of genetic affiliation while they would skew the data in any study using the comparative method.

Section 4 investigated the problems of determining whether certain syntactic and morphological features could be the outcome of shared developments in a language family, while others are simply the result of areal diffusion or are common cross-linguistically, requiring no particular typological preconditions. Five areas of morphosyntax were thus examined: similarities and differences with cross-linguistically attested pathways of language change were described for the five areas of diminutives, negatives, complementizers, passives, and inalienable possession with additional language-specific features being noted in some of these cases: first, diminutive suffixes in Sinitic were shown to have their source not in a morpheme for 'child' but in the more sex-specific 'son' (which nonetheless may have the secondary meaning of 'child' or 'offspring' in some, but not all, of these languages). Second, the large inventory of negative markers in Sinitic languages was also briefly described. The fact that these grammaticalize out of a fusion of basic negative markers and modal verbs appears to be typologically unusual in the light of cross-linguistic studies such as Payne (1985). Third, it was observed

that complementizers with a source in a verb of saying are common cross-linguistically although the Southern Min development is relatively young, while that for Cantonese Yue is only in an incipient stage. Fourth, passive exponents in Southern Sinitic languages were described as typically having their source in verbs of giving, yet it is unusual cross-linguistically for this type of passive also to express adversity and to require an agent. Fifth, for the expression of inalienable possession at the level of nominal syntax, the Meixian Hakka dialect presents an interesting and typologically uncharacteristic case for Sinitic since it uses a portmanteau morpheme in precisely this function. This distinction is typically covert in most Sinitic languages, and can at best be only detected for syntactic constructions such as the double patient with intransitive verbs and two patient nouns. Yet different pronouns and nominal constructions to code alienable versus inalienable possession are very common cross-linguistically (see Chappell and McGregor 1995).

To reconstruct the history of a language family adequately, a model is needed which is significantly more sophisticated than the family tree based on the use of the comparative method. It needs to incorporate the diffusion and layering process as well as other language-contact phenomena such as convergence, metatypy and hybridization. The desideratum is a synthesis of all the processes that affect language formation and development.

References

- AMBERBER, M. 1995. 'Functional verbs in predicate formation. Event type hierarchy and grammaticization', pp. 1–11 of *Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society*.
- ANSALDO, U. 1999. 'Comparative Constructions in Sinitic: Areal Typology and Grammaticalization in Chinese Languages'. Ph.D. dissertation. University of Stockholm, Sweden.
- Arte de la lengua Chiō Chiu. 1620. Manuscript from Biblioteca Universitaria Provincial Barcelona entitled 'Gramatica China'. Anonymous author(s). Manila.
- BAXTER, W. H., III. 1992. *A Handbook of Old Chinese Phonology*. Berlin: Mouton de Gruyter.
- BEIJING DAXUE ZHONGGUO YUYAN WENXUE XI YUYANXUE JIAOYANSHI [Linguistics section of Beijing University Chinese Languages and Literatures Department]. 1964. Editors of *Hanyu fangyan cihui* [A lexical list for Chinese dialects]. Beijing: Beijing University.
- BISANG, W. 1992. *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer*. Tübingen: Gunter Narr.
- 1996. 'Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages', *Studies in Language* 20:3.519–97.
- CHAPPELL, H. 1986. 'Formal and colloquial adversity passives in standard Chinese', *Linguistics* 24:6.1025–52.
- 1992a. 'Causativity and the *ba* construction in Chinese', pp. 509–30 of *Partizipation: das sprachliche Erfassen von Sachverhalten*, edited by Hansjakob Seiler and Walfried Prempner. Tübingen: Gunter Narr.

- 1992b. 'Towards a typology of aspect in Sinitic languages', *Zhongguo Jingnei Yuyan ji Yuyanxue: Hanyu Fangyan* [Chinese languages and linguistics: Chinese dialects] 1.67–106.
- 1994. 'A typology of negative verbs of existence and possession in Sinitic languages', pp. 303–11 of *Current Issues in Sino-Tibetan Linguistics*, edited by Hajime Kitamura, Tatsuo Nishida, and Yasuhiko Nagano. Osaka: National Museum of Ethnology.
- 1999. 'The double unaccusative in Sinitic languages', pp. 197–232 of *External Possession and Related Noun Incorporation Constructions* (Typological Studies in Language Series 39), edited by Doris L. Payne and Immanuel Barshi. Amsterdam: John Benjamins.
- 2000. 'Dialect grammar in two early modern Southern Min texts: a comparative study of dative *kít*, comitative *càng* and diminutive *-guià*', *JCL* 28.248–302.
- 2001a. Editor of *Sinitic Grammar: Synchronic and Diachronic Perspectives*. Oxford: Oxford University Press.
- 2001b. 'A typology of evidential markers in Sinitic languages', pp. 56–84 of Chappell 2001a.
- 2001c. 'Synchrony and diachrony of Sinitic languages: A brief history', pp. 3–27 of Chappell 2001a.
- Forthcoming. 'Grammaticalisation of the verb *kóng* "say" in Taiwanese Southern Min'.
- and MCGREGOR, W. 1995. Editors of *The Grammar of Inalienability. A Typological Perspective on Body Part Terms and the Part-Whole Relation*. Empirical Approaches to Language Typology 14. Berlin: Mouton de Gruyter.
- and THOMPSON, S. A. 1992. 'The semantics and pragmatics of associative *de* in Mandarin discourse', *Cahiers de Linguistique Asie Orientale* 21.199–229.
- CHEN, P. 1999. *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press.
- CHEUNG, S. H.-n. 1992. 'The pretransitive in Cantonese', *Zhongguo Jingnei Yuyan ji Yuyanxue: Hanyu Fangyan* [Chinese languages and linguistics: Chinese dialects] 1.241–303.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- FRAJZYNGIER, Z. 1996. *Grammaticalization of the Complex Sentence: A Case Study in Chadic*. Amsterdam: John Benjamins.
- GROOTAERS, W. 1943. 'La géographie linguistique en Chine—nécessité d'une nouvelle méthode pour l'étude linguistique du chinois', *Monumenta Serica* 8.103–66.
- HASHIMOTO, M. J. 1973. *The Hakka Dialect: A Linguistic Study of its Phonology, Syntax and Lexicon*. Cambridge: Cambridge University Press.
- 1974. 'Aspect and tense in Asian languages', *CAAAL* 1.15–25.
- 1976a. 'Language diffusion on the Asian continent: problems of typological diversity in Sino-Tibetan', *CAAAL* 3.49–63.
- 1976b. 'The double object construction in Chinese', *CAAAL* 6.33–42.
- 1986. 'The Altaicization of Northern Chinese', pp. 76–97 of *Contributions to Sino-Tibetan Studies*, edited by J. McCoy and T. Light. Leiden: E. J. Brill.
- 1987. 'Hanyu beidongshi de lishi, quyu fazhan' [The history and areal development of the Chinese passive], *Zhongguo Yuwen* 1.36–49.
- 1992. 'Hakka in Wellentheorie perspective', *JCL* 20.1–49.

- HEINE, B., CLAUDI, U., and HÜNNEMEYER, F. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
- , GÜLDEMANN, T., KILIAN-HATZ, C., LESSAU, D., ROBERG, H., SCHLADT, M. and STOLZ, T. 1993. *Conceptual Shift: A Lexicon of Grammaticalization Processes in African Languages*. Afrikanistische Arbeitspapiere 34–5. Cologne: Institut für Afrikanistik.
- HUANG B. 1996. Editor of *Hànyǔ fāngyán yǔfǎ lèibiǎn* [Concordance of Chinese dialect grammar]. Qingdao: Qingdao Publishers.
- JURAFSKY, D. 1996. 'Universal tendencies in the semantics of the diminutive', *Language* 72:3–533–78.
- KARLIGREN, B. 1915–1926. *Études sur la phonologie chinoise*. *Archives d'études orientales*, v.15 (in 4 parts). Leiden: E. J. Brill.
- 1940. 'Grammata Serica, script and phonetics in Chinese and Sino-Japanese', *Bulletin of the Museum of Far Eastern Antiquities* 12.1–471.
- 1957. 'Grammata serica recensa', *Bulletin of the Museum of Far Eastern Antiquities* 32.1–332.
- KUBLER, C. C. 1985. *The Development of Mandarin in Taiwan: A Case Study of Language Contact*. Taipei: Student Book Co.
- LIEN, C. 1998. 'Taiwan Minnanyu cizhui 'a' de yanjiu [A study of the affix 'a' in Taiwanese Southern Min]', pp. 465–83 of *Dì'ér jiè Taiwan Yuyan Guoji Yantaohui Lunwen Xuanji*. Taipei: Crane Publishing Co.
- 2001. 'Competing morphological changes in Taiwanese Southern Min', pp. 309–39 of Chappell 2001a.
- LOON, P. VAN DER. 1966. 'The Manila incunabula and early Hokkien studies', part I, *Asia Major*, n.s. 12.1–43.
- 1967. 'The Manila incunabula and early Hokkien studies', part II, *Asia Major*, n.s. 13.95–186.
- MATISOFF, J. 1991. 'Areal and universal dimensions of grammatization in Lahu', pp. 383–453 of *Approaches to Grammaticalization*, vol. 2, edited by E. C. Traugott and B. Heine. Amsterdam and Philadelphia: John Benjamins.
- MATTHEWS, S., and YIP, V. 2001. 'Aspects of contemporary Cantonese grammar: the structure and stratification of relative clauses', pp. 266–82 of Chappell 2001a.
- NORMAN, J. 1982. 'Four notes on Chinese–Altaic linguistic contacts', *Tsinghua Journal of Chinese Studies* 14:1–2.243–7.
- 1988. *Chinese*. Cambridge: Cambridge University Press.
- PAYNE, J. 1985. 'Negation', pp. 197–242 of *Language Typology and Syntactic Description*, vol. 1, *Clause Structure*, edited by Timothy Shopen. Cambridge: Cambridge University Press.
- ROSS, M. D. 1996. 'Contact-induced change and the comparative method: cases from Papua New Guinea', pp. 180–217 of *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*, edited by M. Durie and M. D. Ross. New York: Oxford University Press.
- SAGART, L. 1988. 'On Gan-Hakka', *Tsing Hua Journal of Chinese Studies*, n.s.18:1.141–60.
- 1997. 'On distinguishing Hakka and non-Hakka dialects', *JCL* 26:2.281–302.
- 1999. *The Roots of Old Chinese*. Amsterdam: John Benjamins.
- 2001. 'Vestiges of Archaic Chinese derivational affixes in modern Chinese dialects', pp. 123–42 of Chappell 2001a.
- SAXENA, A. 1988. 'The case of the verb "say" in Tibeto-Burman', *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society* 375–88.

- SERRUYS, P. L.-M. 1943. 'Philologie et linguistique dans les études sinologues', *Monumenta Serica* 8.167–221.
- SHIBATANI, M. 1994. 'An integrational approach to possessor raising, ethical datives, and adversative passives', *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society* 461–86.
- TING, P.-H. 1983. 'Derivation time of colloquial Min from Archaic Chinese', *Bulletin of the Institute of History and Philology* 54:4.1–14.
- WU, Y. 1992. 'Studies in Chinese Lexicology: Investigations into the Xiang Dialect'. Ph.D. dissertation, University of Melbourne.
- YANG, H.-F. 1991. *Taiwan Minnanyu yufagao* [A grammar of the Minnan language of Taiwan]. Taipei: Ta An Press.
- YUAN, J. 1989 [1960]. Compiler of *Hanyu fangyan gaiyao* [An outline of Chinese dialects]. Beijing: Wenzi Gaige.
- YUE-HASHIMOTO, A. 1991. 'Stratification in comparative dialectal grammar: a case in Southern Min', *JCL* 19:2.139–71.
- YOU, R. 1992. *Hanyu fangyanxue daolun* [Chinese dialectology]. Shanghai: Shanghai Jiaoyu Press.
- ZHU, D. 1990. 'Dialectal distribution of V-neg-VO and VO-neg-V interrogative sentence patterns', *JCL* 18:2.209–30.

Areal Diffusion versus Genetic Inheritance: An African Perspective

Gerrit J. Dimmendaal

1. Introduction

The genetic classification of African languages has a long and partially turbulent history. Whereas our understanding of specific linguistic areas on the continent has improved considerably over the past decades, the increased knowledge in most cases has resulted in confirmation of hypotheses on their genetic links as formulated by Greenberg (1963). The four major language families according to this classification are Afroasiatic, Khoisan, Niger-Congo, and Nilo-Saharan. Only a few groups have been subject to genetic reclassification over time, as a result of improved documentation.¹ The internal classification, on the other hand, and the integrity of larger subgroups as proposed by Greenberg and others, have been subject to extensive debates. In at least one case (discussed below), that of the Kwa and Benue-Congo languages, investigators came to realize that at an earlier stage of their scientific investigation areal diffusion had come to be mixed up with genetic inheritance.

This contribution was written when the author was a visiting scholar at the Australian National University, Canberra. I am deeply indebted to Sasha Aikhenvald and Bob Dixon, directors of the Research Centre for Linguistic Typology, for making this visit possible; I would also like to thank them as well as the referees for their thorough comments on an earlier version of this chapter.

¹ The Kadu languages (Sudan), for example, are now generally considered to be part of Nilo-Saharan, rather than the Kordofanian branch of Niger-Congo, as in Greenberg's classification; the Mao languages of south-western Ethiopia, classified as part of Koma (Nilo-Saharan) by Greenberg, are in fact part of the Omotic branch of Afroasiatic. Also in south-western Ethiopia, probably an ecological refugium area for thousands of years, there are—what appear to be—two linguistic isolates, Ongota (Biraile) and Shabo. These two highly endangered languages, which were not listed by Greenberg (1963), may constitute the last representatives of independent African stocks. Finally, a language called Laal (Chad), whose existence was not known among Africanists at the time of Greenberg's classification, has defied genetic classification so far, although some Africanists claim it is an Adamawa-Ubangian (Niger-Congo) language. Least well established, in the present author's view, is the genetic status of Khoisan. Whereas there is solid grammatical evidence for a Central Khoisan group, with distant genetic links with the Sandawe language in Tanzania, the genetic affiliation of Northern and Southern Khoisan as well as the Hadza language in Tanzania with the former or with each other is far from clear at present. Khoisan may therefore be an areal grouping.

This study sets out to describe two case studies of areal diffusion (§2), first between Swahili and other coastal Bantu languages, next between Baale and other Surmic languages. These would seem to represent instances of convergence of a type frequently encountered in Africa between genetically related languages. A comparison of these two case studies, the first involving a Niger-Congo subgroup, the second a Nilo-Saharan subgroup, shows that their respective historical outcomes are slightly different for reasons to be discussed below.

The findings for these two case studies are used as a methodological basis for the second part of the chapter (§3), where diffusion within the Niger-Congo family at large is discussed. The present study focuses on this latter family rather than on one of the other major families on the African continent for a number of reasons: first, a considerable amount of comparative work has been carried out for this genetic group; also, most of the languages involved are spoken in a geographically contiguous zone; finally, there are a number of well-attested areal phenomena, some of which also cut across the genetic boundaries of Niger-Congo. The results of these case studies are compared to those for Australia and other parts of the world (§4).

2. Two cases of areal diffusion

2.1. THE INFLUENCE OF SWAHILI ON OTHER COASTAL BANTU LANGUAGES

Swahili ranges amongst the major languages of Africa in terms of number of speakers. As suggested by the etymology of its name (derived from the Arabic word for 'coastal belt'), it used to be the language of the coast. Its geographical expansion is generally assumed to have taken place via the coast of East Africa, from the mouth of the Tana river (Kenya) down to northern Mozambique, over a period of more than one thousand years. Its expansion into the African interior dates back to a relatively recent period in history, mainly the nineteenth century.

Swahili belongs to an extensively studied language group, itself one of the earliest established genetic units on the continent, called Bantu. In view of massive grammatical and lexical affinities of Bantu to languages of West Africa, as pointed out by Westermann (1927) amongst others, a gradual awareness grew that these latter languages must be genetically related to Bantu. According to Greenberg (1963), Bantu must be a relatively late split-off of Niger-Congo (or Kongo-Kordofanian, as it was called at the time), a position which now has acquired universal acceptance amongst specialists in the area. The internal subclassification of Bantu on the other hand has turned out to be notoriously difficult, above all because various potentially diagnostic features involving phonological and morphological innovations, have an areal distribution, and so shared innovations are sometimes hard to distinguish from areally diffused innovations.

In their monumental work on Swahili, Nurse and Hinnebusch (1993) classified Swahili as one of the six members of the Sabaki group (after a main river in the

area). The latter is part of a larger grouping called the North-East Coast Bantu languages. As pointed out by the same authors, the most visible sign of outside encroachment is in the Swahili lexicon, and with it borrowed phonemes, as well as lexical items from Arabic, Persian, Turkish, and various Indian and European languages. As further stated by Nurse and Hinnebusch (1993: 310, 320 and *passim*) there is almost no sign of direct morphological or syntactic borrowing from these languages in the inflectional or derivational system of Swahili. At the language-internal level, however, the situation is different. Areal diffusion or borrowing between dialects is at times hard to distinguish from inheritance and shared innovations, as Nurse and Hinnebusch (1993: 463) have pointed out.

During its southern expansion, Swahili influenced other more distantly related coastal Bantu languages, amongst them a language of Mozambique, Khoti (Ekhoti). The Khoti speech community is relatively small compared to Swahili or the neighbouring Makhuwa language; according to Grimes (1996: 316), Khoti has 41,287 speakers, and Makhuwa 5,208,000 or more.² According to Schadeberg (1997), Khoti is closely related to the Southern Bantu language Makhuwa, but it has converged towards Swahili as a result of heavy influence from the latter. Swahili and Makhuwa differ considerably phonologically. Moreover, a lexico-statistic count reveals that Swahili and Makhuwa share 49% basic vocabulary; Khoti and Makhuwa share 67%. On the other hand, Swahili and Khoti also share 67% basic vocabulary, thus suggesting a kind of 'double lexical alliance' for Khoti. Schadeberg conjectures that massive relexification in Khoti towards Swahili lies at the heart of this remarkable fact. The extensive borrowing, also of basic vocabulary, has led to the blurring of sound correspondences, or rather to double correspondence sets, as the following examples help to illustrate.

(1) Swahili	Khoti	Makhuwa	
-pata	-patha	-vara	'get'
-pita	-vira	-vira	'pass'
-tukana	-tukhana	-ruwana	'insult'
-tuma	-ruma	-ruma	'send'
-tfeza	-feza	-teya	'play'
-tŋimba	-thipa	-tipa	'dig'

Whereas the correspondence sets between Swahili and Makhuwa are regular (for example: $p \sim v$; $t \sim r$, $mb \sim p$), the correspondences between Khoti and Makhuwa are not; as a result of the relexification, Khoti has lexical forms virtually identical to those in (Standard) Swahili in certain examples (e.g. 'get', 'insult', 'play'), whereas in others it is virtually identical to Makhuwa. The Khoti noun-class system is still closer structurally to Makhuwa. Also, its productive type of derivational morphology in the verb is formally and semantically identical to Makhuwa,

² Such numbers of speakers would be considered huge in comparison with figures common among Australian or Amazonian groups (Sasha Aikhenvald and Bob Dixon, p.c.).

as stated by Schadeberg (1997: 17–18). Moreover, the inflectional morphology (e.g. pronominal subject and object marking, tense) is shared with Makhuwa rather than with Swahili.

These days, the Khoti have little knowledge of Swahili, according to Grimes (1996: 316). However, in order to explain the deep borrowing (of basic vocabulary) into what is now known as Khoti, intimate knowledge of Swahili in former times must be assumed. The Khoti case is best understood as one where Makhuwa or a variety of southern Bantu extremely close to Makhuwa, constituted the matrix language. The relexification occurred either through intensive borrowing of lexical items, or alternatively, through ‘correspondence mimicry’; where speakers regularly use two or more related languages, they may develop an intuitive grasp of some of these correspondences and use them to convert the phonological shapes of words from one lect to another. (Compare also (Ross 1997) for such processes in Papua New Guinean languages, or Evans (1998) for Australian languages.)

2.2. THE SURI OF THE ETHIOPIA–SUDAN BORDERLAND

Whereas there is considerable disagreement on the internal classification of the Nilo-Saharan family at the higher levels, lower-level units such as Nilotic, Saharan or Central Sudanic have been recognized for some time now. (Greenberg’s 1963 subclassification of this phylum is given in the Appendix below.) One of these lower-level genetic units includes a group of languages known today as Surmic (and listed as group 2 within the list of the ten Eastern Sudanic subgroups by Greenberg 1963: 85). There is now a fairly well-established subclassification for Surmic, supported by shared phonological, lexical, and grammatical innovations; see Dimmendaal (1998a) for a summary of arguments, and Dimmendaal and Last (1998) for a general survey of this group of languages. On the other hand, there is also evidence for areal diffusion between its members. One such clearcut case is that between Baale and Tirma-Chai.

Genetically, Baale clearly is to be grouped with the Didinga-Murle languages, with which it forms the South-Western branch of Surmic (Moges Yigezu and Dimmendaal 1998). The social grouping of its speakers, however, is rather different. The Baale, who live in the border area between Ethiopia and Sudan and who number around 9,000 (or possibly less), are in close contact with the Tirma and Chai, speakers of South-Eastern Surmic languages. The Tirma and Chai are more numerous, estimates ranging between 20,000 and 40,000 people. The Tirma, Chai, and Baale frequently intermarry and hold common ceremonies. They all refer to themselves as Suri (or Surma) people. Tirma and Chai are mutually intelligible, and are closely related members of a dialect cluster to which Mursi also belongs. Speakers of Mursi, however, do not refer to themselves as Suri people; they see themselves as belonging to a different ethnic group, a classic case, then, where language and ethnicity are not isomorphic.

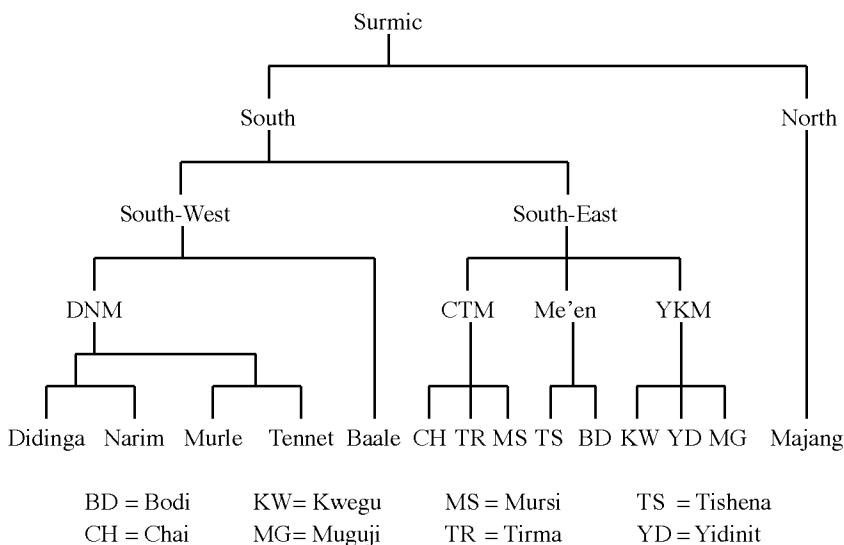


FIGURE 1. Surmic subclassification (Dimmendaal 1998a)

As a result of the intensive interaction and networking between speakers of the South-Western Surmic language Baale with the South-Eastern Tirma and Chai speakers, Baale now shows various typological properties which are absent from its closest linguistic relatives (Didinga-Murle), but which are common in Tirma and Chai. Phonologically, for example, Baale clearly has converged towards Tirma and Chai. Tirma and Chai do not have word-final stops. Baale has lost word-final stops, at least at the phonetic level, as a comparison with cognates in the Didinga-Murle languages shows.

- (2) Baale: Didinga: Murle:
mèèlé *meelek* *melek* 'axe'
wé *uwec* *weec* 'four'

Stops were protected from loss in Baale whenever another suffix followed (e.g. as a result of number suffixation or case-marking for nouns); the result now is a rather complex system of morphophonemic alternations in the language, where the stops still have to be posited underlyingly (e.g. *mèèlé* < *k* >).

- (3) singular plural
mèèlé < *k* > *mèèlé-k-ká* 'axe'
 [mèèlé] [mèèlékká]

Whereas Baale has nine vowels and a classic African system of A[dvanced] T[ongue] R[oot] harmony (see below in §3.1), the [+ATR] mid-vowels *e* and *o*

are extremely close acoustically to [+ATR] i and u respectively. Baale's closest relatives, the Didinga-Murle languages, also have ATR-vowel harmony systems. South-eastern Surmic languages like Tirma and Chai, on the other hand, have seven-vowel systems without vowel harmony; possibly, then, Baale is on its way to developing a seven-vowel system as well. Also, the tonal structure of Baale appears to have been influenced by that of Tirma-Chai, although more research is required on these related languages.

As in Khoti, there has been extensive borrowing of basic vocabulary from Tirma and Chai into Baale, in particular with respect to kinship terminology. Sporadic sound changes in items which are cognate—and which presumably are recognized as such by speakers—are rare. The loss of the final nasal in the Baale word for 'water', for example, is not part of a regular historical rule; this irregularity is best explained as a case of 'correspondence mimicry'.

(4)	Didinga	Baale	Chai	
	<i>maam</i>	<i>máá</i>	<i>maa</i>	'water'

The final bilabial nasal is still found in Baale when a case suffix follows, as in the instrumental form *máammé*. These sporadic changes, involving phonological modification in Baale of inherited common Surmic vocabulary, are similar to the phonological modifications in Khoti items such as 'get', 'insult', and 'play' in example (1) above. This process of correspondence mimicry may be distinguished from lexical borrowing involving transfer of lexical material as a corollary of cultural influence. For example, Baale has borrowed various words from Tirma-Chai relating to social structure (e.g. 'wedding' *wóllollo*; Tirma *wololo*); these items are not attested in Baale's closest relatives, the Didinga-Murle languages.

Baale probably also borrowed an aspectual particle *wa* 'just now, recently' from Tirma-Chai. This marker is found in several South-Eastern Surmic languages (cf. Last and Lucassen 1998: 384), but no such marker is attested, as far as is known, in Baale's closest relatives, Didinga-Murle.

There has been extensive grammatical convergence in Baale towards Tirma and Chai; evidence for some dramatic innovations under the influence of Tirma and Chai is emerging in this respect. Because our findings are part of research in progress, only the more obvious cases are illustrated below.

Whereas nominal compounding is rare in Didinga-Murle, it is found rather frequently in Baale; the order in Baale endocentric as well as exocentric compounds is modifier-head, as in Tirma or Chai. Because of the rather idiomatic nature of several of these compounds in Baale, their origin is best explained as a result of calquing from Tirma and/or Chai, where identical patterns occur. Compare:

(5)	Tirma	Baale	
	<i>way tugo</i>	<i>ǎ́ǎ ǎ́ǎ<k></i>	'nipple [lit. breast mouth]'
	<i>dori jagare</i>	<i>kis-sǎ</i>	'wall [lit. house leg]'

In addition, the categorization of adjectival concepts has been affected in Baale. Whereas in the closely related Didinga-Murle languages concepts such as ‘big’ or ‘small’ are expressed as prototypical verbs, Baale conjugates them in a way parallel to nominal predicates, i.e. in combination with a copula ‘be’, a strategy also found in Tirma and Chai. Compare Tennen (data from Randal 1998) as a typical representative of Didinga-Murle:

(6) **k-eéni anná demézzóht**

1sg-be I.NOM teacher

‘I am a teacher.’

(7) **maán-ê tíí na íí kó**

tan-PL cows.NOM DEM

‘These cows are tan-coloured.’

Baale consistently treats such attributive concepts in the same way as nominal predicates, in that both require an auxiliary or copula (data from author):

(8) **ándá kééni baàlèjini**

I.ABS 1sg.be Baale.ABS

‘I am a Baale.’

(9) **ándá kééni càllé**

I.ABS 1sg.be well

‘I am well/fine.’

Compare Chai (data from Last and Lucassen 1998):

(10) **ṣàgénù á tirmàgà**

those.ABS 3pl.be Tirma.ABS

‘They are Tirma (people).’

(11) **yòg á ràrà**

they.ABS 3pl.be tall

‘They are tall.’

(12) **àpè k-àni àrà**

I:ABS 1sg-be tall

‘I am tall.’

The Didinga-Murle languages are head-marking, verb-initial languages; Baale allows for verb-initial constituent order, but like Tirma and Chai, it has a rather free constituent order (allowing for VSO, as well as SVO, OVS, SOV; Dimmendaal 1998b). The typological shift in the conjugational behaviour of property concepts in Baale confirms an observation made by Dixon (1997: 125) on the link between head-marking versus dependent-marking grammar, and the corresponding grammatical status of adjectives.

In spite of the rather dramatic morphosyntactic changes in Baale, native speakers still think of their language as being highly similar to Didinga-Murle. As my

main informant for this language, Alemu Olekibo, once pointed out, Baale and Murle are similar to each other, in the same way that Tirma and Chai are. (For further details see Moges Yigezu and Dimmendaal 1998.)

2.3. A COMPARISON OF KOTI AND BAALE

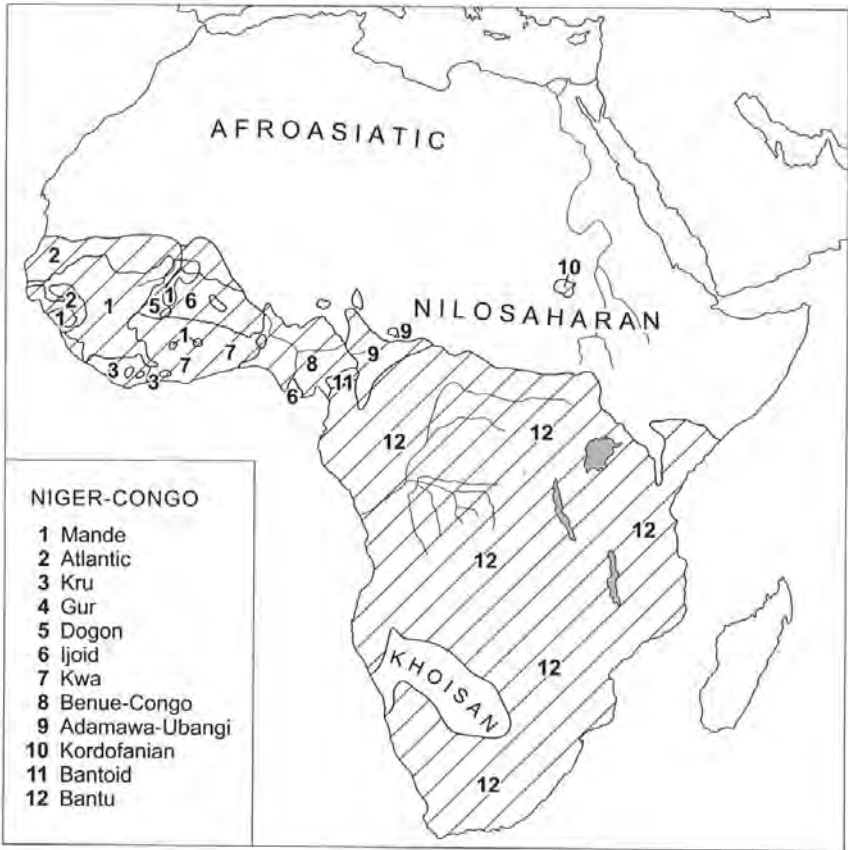
When comparing the two case studies above, one may observe a number of common properties. The relatively small population size—compared to their dominant neighbours—presumably has facilitated the diffusion of borrowings amongst speakers of Khoti and Baale. The transfer of linguistic features into their respective languages follows naturally from a situation of bilingualism with diglossia amongst speakers of Khoti and Baale, which allows for a deep influence from the speech of prestigious neighbouring communities. But at the same time the Baale and Khoti communities maintained a partly separate identity, in that they upheld their own family ties, cultures and partly distinct ethnic history. Neither the Baale nor the Khoti gave up their first language, presumably because they wanted to maintain a double identity. Why they decided to do so can only be explained through more in-depth sociological and historical research.

Both the Khoti and the Baale cases represent instances of linguistic change as a result of rapid horizontal assimilation towards the structure of neighbouring languages to which they are genetically related. The Khoti came into contact with speakers of a prestigious seafaring nation with a powerful religion, Islam; the Baale, who are agricultural specialists, came into contact with the widely admired pastoral cultures of the Tirma and Chai people. However, unlike the situation in Khoti, there is little evidence for ‘correspondence mimicry’ between Baale and Tirma-Chai, presumably because the genetic distance (which is also reflected in their typological distance) is too big. On the other hand, there is little evidence, either in Baale or in Khoti, for extensive grammatical borrowing, for example of function morphemes. What speakers appear to have copied in the case of Baale is a pattern (e.g. with compounding and constituent order) rather than the actual morphosyntactic elements themselves. The typological gap to be bridged in the case of Baale must have been considerably wider than in the case of Khoti, morphologically as well as in terms of constituent order and clause structure, thus resulting in rather dramatic changes in this Surmic language.

With these two case studies at the ‘micro-level’ in mind, let us have a closer look at the macro-level, where we are dealing with diffusion over a much larger area and with a considerably greater time depth. What evidence is there for areal diffusion, and to what extent has this complicated the reconstruction, or obliterated the subclassification of Africa’s largest language family?

3. Diffusion versus genetic inheritance in Niger-Congo

Niger-Congo (also known as Kongo-Kordofanian, or Niger-Kordofanian) is the largest family on the African continent in terms of number of languages as well as



MAP 1. The language families of Africa

geographical spreading (see also Map 1.) There is relatively little disagreement on lower subgroupings within this stock. Several of these smaller units were already established *de jure* by nineteenth-century scholars such as Koelle in his pioneering study *Polyglotta Africana* (1854). Greenberg's (1963) classification contains important innovations at the macro-level, however, for example in that the Adamawa-Ubangi (Adamawa Eastern) languages as well as the Kordofanian languages of Sudan were also included in the Niger-Congo by him. (See also Dimmendaal (1993), and Newman (1995) for an assessment of Greenberg's contribution to the genetic classification of African languages.)

A more recent attempt at a subclassification of this family is that by Williamson (1989a). Genetically, the most diversified subgroup within Niger-Congo is Benue-Congo (Williamson 1989b: 261). Its 'rake-like' structure probably should not be

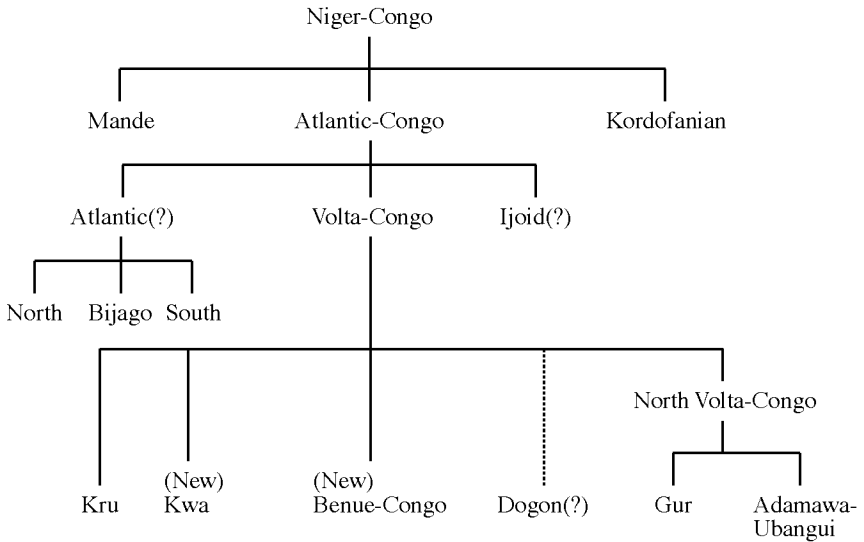


FIGURE 2. The subclassification of Niger-Congo (Williamson 1989a)

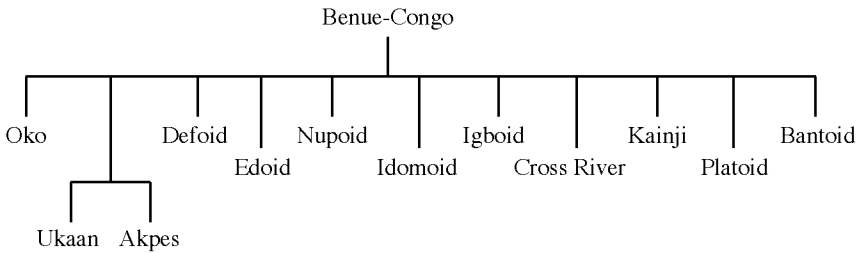


FIGURE 3. The subclassification of Benue-Congo (Williamson 1989b)

interpreted to represent a sudden and massive dispersion of various groups. Rather, the ‘flat’ structure represents absence of solid and convincing criteria for internal subclassification at this point in time. (Note that the term Bantoid refers to Bantu and its closest relatives.)

Ijoid, as well as several of the groups now classified as part of Benue-Congo (Yoruboid, Edoid, Nupoid, Idomoid, Igbooid) were classified as (Eastern) Kwa in Greenberg (1963). It is probably fair to state that their inclusion within Kwa was motivated to some extent by the observed typological similarity with languages still classified under Kwa today. For example, these various languages share such features as ATR-vowel harmony, nasalized vowels, reduced noun-class systems, and serial-verb constructions. (These properties are further discussed below.)

Apart from Greenberg's (1963) list, an extensive list of likely lexical cognates with a widespread distribution across Niger-Congo has been presented by Mukarovsky (1976–7), who also makes some preliminary attempts at reconstruction of proto-forms. By the criteria of regular sound correspondences among these languages and of the reconstruction of proto-forms, Niger-Congo is not a proven genetic unit. Nevertheless, considerable historical-comparative work, using classical Neogrammarian methods of regular sound correspondences as well as grammatical comparison, has been carried out over the past decades, most prominently in the scholarly work of John Stewart (e.g. 1970, 1971, 1976, 1983, 1994). In his historical-comparative work Stewart has concentrated on Volta-Congo. Stewart (1976: 7) makes the following observation with respect to his comparative findings on this major sub-branch within Niger-Congo:

We find that closely similar ancestral sound systems can be reconstructed independently for the Kwa and Gur but possibly not for the Benue-Congo languages; the proto-Bantu sound system, however, can be plausibly regarded as a modified form of the proto-Volta-Congo system as it emerges from the study of the Kwa and Gur languages.

Amongst these phonological phenomena essentially absent in Bantu, but widespread in other subgroups within Volta-Congo or Niger-Congo as a whole, are vowel harmony and nasalized vowels; these are discussed next.

3.1. VOWEL HARMONY OF THE CROSS-HEIGHT TYPE

In a now classic article, Stewart (1967) has shown that vowel alternation in Akan (Kwa; Ghana) is governed by a distinctive feature involving tongue root advancement (ATR) harmony. In such an ATR vowel-harmony system there are ten alternating vowels: A harmony set of five [–ATR] vowels, **ɪ, ɛ, ɔ, ʊ**, and a harmony set of five [+ATR] vowels, **i, e, ä, o, u**. 'The relation of the first set to the second is one of unmarked to marked, so that one would expect to find a constant articulatory feature extending throughout harmony spans with **i, e, ä, o, u**', according to Stewart (1967: 202). This type of alternation between [–ATR] and [+ATR] vowels in morphemes (resulting in alternations of the following type: **ɪ/i, ɛ/e, ɔ/ä, ʊ/o, ʊ/u**) is a central property of such systems. Two Akan examples:

- (13) **ɔɔbetú** 'it (sc. the hen) is going to lay'
oobetú 'he is going to pull it out'

Stewart (1976: 9) further observes that '[i]t would appear that this type of vowel harmony affords the vowel system an extraordinary stability, the scope for vowel shifting which would not seriously interfere with the harmony being very limited'. A vowel reduction, for example with ***ɪ** and ***ʊ** shifting to and merging with ***ɛ** and ***o** indeed would seriously affect the operation of the harmony system; it would result in morphophonemic alternations between **ɛ/e**, next to **ɛ/i**, and **ɔ/o** next to **o/u**. Such historical mergers accordingly lead to the breakdown and loss of cross-height harmony. (See also Archangeli and Pulleyblank (1994) on this topic.)

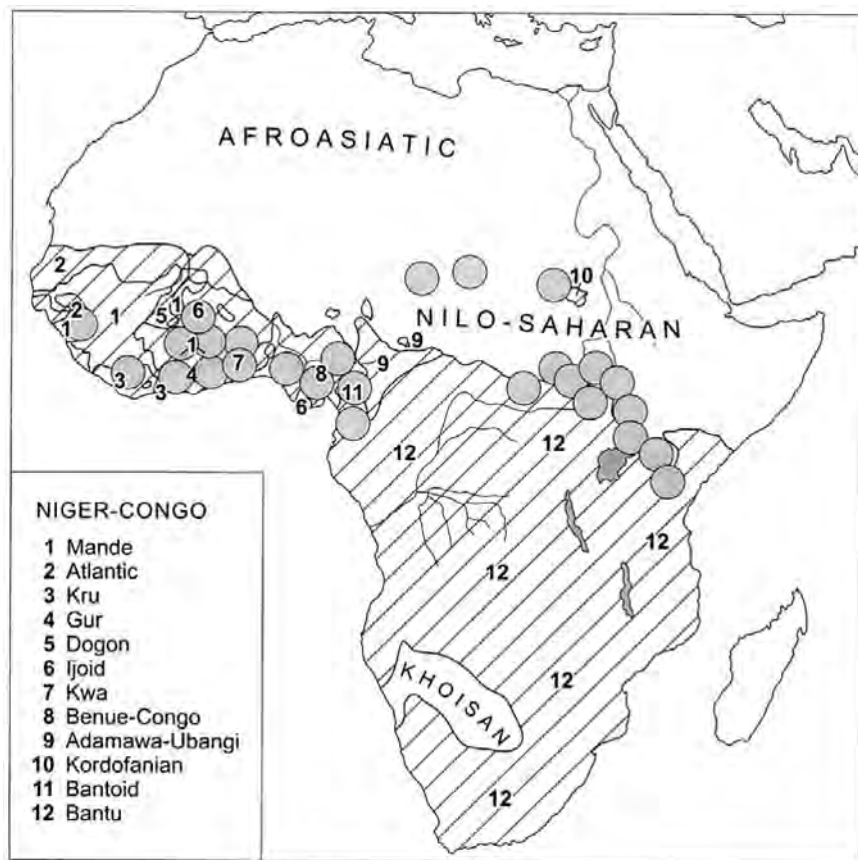
TABLE 1. Vowel systems in Niger-Congo

Atlantic	between 5 and 10	Benue-Congo	
Mande	between 5 and 9 vowels	Edoid	between 7 and 10
Kru	between 7 and 9	Nupoid	between 5 and 10
Gur	between 5 and 10	Idomoid	10
Kwa	between 7 and 10	Defoid	between 7 and 9
Ijoid	9	Igboid	between 8 and 10
Adamawa-		Platoid	between 5 and 9
Ubangi	between 5 and 10	Kainji	between 5 and 9
Kordofanian	between 5 and 7 (or 9?)	Bantoid	between 6 and 10

Classic vowel-harmony systems of the Akan-type, with ten vowels (or 'slightly less than classic' systems with nine vowels, in which the low vowel *a* lacks a [+ATR] counterpart, are widespread across Niger-Congo, as Table 1 helps to illustrate. However, as the same table shows, there is also considerable variation within each of the sub-branches. Atlantic languages such as Dyola manifest classic properties of cross-height vowel harmony (cf. Sapir 1965), whereas in other Atlantic languages (e.g. in Limba, which has seven oral and seven nasal vowels) there is no trace of such a system.

A widespread assumption among Niger-Congo specialists appears to be that languages lacking this type of vowel harmony may have lost it. For example, Bendor-Samuel (1992: 98) notes that given vowel systems found today, it may eventually be possible to reconstruct Proto-Niger-Congo with ten contrasting vowels. But as Creissels (1994: 102–3) has pointed out with respect to historical-comparative work in Niger-Congo, 'il y a une tendance à interpréter toute alternance vocalique faisant intervenir des distinctions d'aperture comme le vestige d'un ancien système à harmonie d'avancement qui se serait dégradé' [there is a tendency to interpret every vowel alternation involving a distinction of openness as a trace of an archaic system of advancement harmony, which has been partially lost]. And there are additional, complicating factors. What is striking, is the apparent *instability* of such systems, as the variation within the genetic subgroups presented in Table 1 shows. From the variation within these reasonably well-defined subgroups, one could equally well conclude that languages may easily develop ATR-harmony through areal diffusion, in particular, as we shall see below, if one takes into account the geographical distribution of such harmony systems.

The core area for vowel harmony is formed by Kwa, the western representatives of Benue-Congo, and Ijoid. For a number of subgroups within Benue-Congo and Kwa, ten-vowel systems have in fact been reconstructed: compare Williamson (1983–4) for a survey of Cross River and Igboid, and Elugbe (1989) for Edoid; see also Williamson (1983–4) for Ijoid. On the other hand, there are well-established Benue-Congo representatives such as Proto-Bantu, reconstructed with a seven



MAP 2. [ATR] harmony in Niger-Congo and Nilo-Saharan

vowel system (*j, *i, *e, *a, *o, *u, *ʏ) and no vowel harmony. (Compare the reconstructions in Guthrie (1967–71), or Meeussen (1967).) Southern Bantu languages like Tswana have nine vowels (without vowel harmony) going back to a seven-vowel system (Creissels 1994: 7). In the northern Bantu borderland, a number of languages are in the process of developing a nine-vowel inventory with vowel harmony. Bila, for example, has characteristics of both a (classic Bantu) seven-vowel system and a nine-vowel system with vowel harmony. The former is still apparent in noun roots in which $V_1 = V_2$. But the language is in the process of acquiring a nine-vowel system, for example in its verbal system, and with nouns in which V_1 is different from V_2 (Conny Kutsch Lojenga, p.c.).

Some of the north-western Bantu languages, such as Tunen (south-eastern Cameroon), have ATR-harmony as well. But this system again must be a later

development. Stewart and Van Leynseele (1979: 51) have argued that Proto-Bantu had a classic system of cross-height vowel harmony with nine (or possibly ten) vowels, and that it inherited this system largely unchanged from Proto-Volta-Congo; consequently, the cross-height vowel harmony in the Bantu language Tunen was inherited from Proto-Bantu, according to them. However, as shown by de Blois (1981), the Tunen ten-vowel system can be derived from the seven-vowel system of Proto-Bantu (*i, *i *e, *a *o *u *ʊ), if one assumes that the vowels (represented by the symbols) i, e, a, and o had a distinctive feature [–ATR] in pre-Nen, and that these [–ATR] vowels developed [+ATR] counterparts by way of assimilation rules involving the development of ([+ATR]) i, e, ə, o, u in words containing ([+ATR]) i and u. (See also Stewart (1983: 33–5) on the same subject.) This type of assimilation rule appears to be a common historical source for the innovation of cross-height harmony in Niger-Congo languages. Compare, for example the Ijesha dialect and other eastern varieties of Yoruba (Nigeria) which developed a nine-vowel system with cross-height harmony out of a Proto-Defoid seven-vowel system, according to Capo (1985), who further observes (p. 117) that the prolonged contact with Edoid and Igbirra languages operated as a catalyst. Unadapted borrowing from such classic vowel-harmony languages usually results in further expansion of the ATR contrast.

There is a clear-cut areal dimension to these varying vowel systems, as the cases discussed above already suggest. For the Edoid branch within Benue-Congo, for example, Elugbe (1989) reconstructs an original ten-vowel system. This system has been retained in Degema, which is surrounded by Ijoid languages, which also have classic vowel-harmony systems. On the other hand, Williamson (1989a: 110) points out that in the Ijoid language Nkɔrɔɔ 'vowel harmony has almost completely disintegrated ... due to the influence of neighboring languages, particularly Obolo [Cross River; Gerrit Dimmendaal]'.

In his discussion of Gur, Naden (1989: 154) observes that the Central Gur languages have seven vowels. 'These facts do not exclude the possibility that an ancestral language may have had vowel harmony and contrastively nasalized vowels, but only show that Central Gur languages do not furnish any reason for postulating such systems.' When taking into account not only such internal variation within subgroups but also their areal distribution, the logical conclusion to be drawn is that vowel harmony can be acquired and lost easily.

The ultimate answer as to the direction of change in early Niger-Congo vowel systems obviously comes from the application of the historical-comparative method. But as Stewart (1994: 176) points out:

It has proved extremely difficult to find regular sound correspondences across Ewe and Akan ([both Kwa; Gerrit Dimmendaal])... It has in fact proved much less difficult to find sound correspondences across Akan and Proto-Bantu, even though the latter is classified not as Kwa but as Benue-Congo... The explanation appears to be that there has been very extensive soundshifting in both Kwa-to-Ewe and Kwa-to-Akan but relatively little in Volta-Congo-to-Kwa and Volta-Congo-to-Bantu.

There may be a number of reasons for these complications. As we saw above for Khoti, ‘correspondence mimicry’ between relatively closely related languages may blur sound correspondences. Also, prosodic phenomena such as vowel harmony apparently spread relatively easily (through unadapted borrowing and rephonologization, next to language-internal vowel assimilation). These severely complicate the establishment of sound correspondences between closely related languages.

On the basis of a systematic comparison between Akan (Kwa), Proto-Tano-Congo (the latest common ancestor of Benue-Congo plus Kwa) and Common Bantu as reconstructed by Guthrie (1967–71), Stewart (1983) posits nine oral vowels for their common ancestor, Proto-Tano-Congo, ‘and seven of these with considerable confidence’ (Stewart 1983: 26). Tano-Congo differs from Volta-Congo, in that Kru is not included in the subgrouping. The Proto-Bantu forms in Table 2 represent a transcription of the Common Bantu forms in terms of their presumed ATR values. Note also that several of these reconstructed Proto-Tano-

TABLE 2. Comparative Tano-Congo

Akan	Proto-Tano-Congo	Proto-Bantu	Common Bantu	
<i>bɪŋ</i>	* <i>ˈbidɪ</i>	* <i>-bid</i>	* <i>-bɪdò</i>	‘dirt’
<i>ciri</i>	* <i>gidi</i>	* <i>-gid</i>	* <i>-gɪd-</i>	‘abstain’
<i>tiŋ</i>	* <i>ˈtina</i>	* <i>tina</i>	* <i>-tɪnà</i>	‘root’
<i>wu</i>	* <i>ku</i>	* <i>ku</i>	* <i>-kú-</i>	‘die’
<i>huru</i>	* <i>pudu</i>	* <i>pud</i>	* <i>-pɹ́do</i>	‘foam’
<i>huru</i>	* <i>pudu</i>	* <i>pud</i>	* <i>-pɹ́d-</i>	‘froth over’
<i>bɪŋ</i>	* <i>ˈbidɪ</i>	* <i>bid</i>	* <i>-bɪd-</i>	‘be cooked’
<i>sɪ</i>	* <i>cɪ</i>	* <i>cɪ</i>	* <i>-cɪ</i>	‘underneath’
<i>cɪ</i>	* <i>kɪ</i>	* <i>kɪ</i>	* <i>-kɪ-</i>	‘dawn’
<i>boro</i>	* <i>ˈbodo</i>	* <i>bod</i>	* <i>-búd-</i>	‘hit’
<i>soro</i>	* <i>jodo</i>	* <i>jodo</i>	* <i>-jùdú</i>	‘top, sky’
<i>mɔ</i>	* <i>mœ</i>	* <i>mœ</i>	* <i>-múé</i>	‘you (pl)’
<i>seŋ</i>	* <i>dedɪ</i>	* <i>ded</i>	* <i>-déd-</i>	‘be suspended’
<i>ceŋ</i>	* <i>pedɪ</i>	* <i>edi</i>	* <i>-yédi</i>	‘moon’
<i>ceŋ</i>	* <i>pedɪ</i>	* <i>edo</i>	* <i>-yédu</i>	‘white’
<i>Fɔw</i>	* <i>bɔbu</i>	* <i>bɔmb</i>	* <i>-bòmb-</i>	‘be wet’
<i>pɔŋ</i>	* <i>ˈkwɔɔŋɔ</i>	* <i>kɔɔn</i>	* <i>-kóónud-</i>	‘break off’
<i>wɔ</i>	* <i>wɔka</i>	* <i>ɔka</i>	* <i>-yókà</i>	‘snake’
<i>da</i>	* <i>ˈdaádɪ</i>	* <i>daad</i>	* <i>-dáád-</i>	‘lie down’
<i>saw</i>	* <i>tapɪ</i>	* <i>tap</i>	* <i>-táp-</i>	‘draw (water)’
<i>sə</i>	* <i>təto</i>	* <i>tato</i>	* <i>-tátù</i>	‘three’

Congo roots (e.g. 'die', 'snake', 'lie down', 'three') are widespread across the Niger-Congo family; compare also Mukarovsky (1976–7.)

According to Stewart (1983), the mid vowels **e* and **o* are least well established: 'Unfortunately, Akan does not have many roots with *e* or *o* as the only vowel, and it is consequently not possible to establish Proto-Bantu correspondents for these two vowels with anything like as much confidence as for the remaining seven.'

If ATR vowel harmony in Niger-Congo is indeed the result of diffusion, a hypothesis which at this point remains inconclusive, we must be dealing with an ancient convergence phenomenon, given its wide spreading, not only across Niger-Congo, but also into neighbouring Afroasiatic and Nilo-Saharan languages. The Tangale group within the Chadic branch of Afroasiatic probably developed harmony systems as a result of long-term contact with neighbouring Benue-Congo languages (see Jungraithmayr 1992–3). Vowel harmony systems have also been found in East African representatives of Afroasiatic such as the Omotic language Hamar (Lydall 1976), presumably as a result of diffusion from some Nilo-Saharan group in the area. With respect to Nilo-Saharan vowel systems, there is usually considerable variation between and within genetically well-defined subgroups. (See the Appendix for Greenberg's subclassification of Nilo-Saharan.) For example, within Central Sudanic, there are classic ten-vowel or nine-vowel systems, but also seven-vowel systems. Similarly, Nile Nubian has five, but Hill Nubian probably has nine vowels. There are Nilotic and Surmic languages with classic ATR systems, and those without. As in the case of Niger-Congo, there is a clear-cut areal dimension to the dichotomy between languages with and those without ATR systems. (Compare also Map 2.) Again, this could be interpreted in either way. Languages or language groups in the centre of the area or bordering on the area with vowel-harmony systems are more prone to retain this feature. Continued contact, also with unrelated languages sharing the same phonetic properties, may contribute to the conservation of features, as with laryngealized consonants in some Indo-European languages of ancient Anatolia. (See also Watkins, this volume; further evidence for areal norms as important factors in the retention or loss of phonological features in eastern Africa is given in Dimmendaal (1995).)

If ATR-harmony is old in Niger-Congo, the seven-vowel systems of Mande or Bantu, spoken in lateral or relic zones—at least from a geographical point of view—would represent innovations. Since, however, languages or language groups can easily acquire such features, early split-offs like Mande may have retained a more archaic seven-vowel system, with only a few southern Mande languages bordering on the core area where vowel harmony is found, e.g. languages such as Bissa, having copied this typological feature through areal diffusion at a much later point. Given the generally conservative nature of Bantu (with a well-established original seven-vowel system), one would predict that vowel harmony is an instance of (early) diffusion. Only further historical-comparative studies will tell us what the direction of change was.

3.2. THE CONTRAST BETWEEN ORAL AND NASALIZED VOWELS

When plotting another phonological feature, that of nasalized vowels, on the map of Africa, the emerging picture differs from that for vowel harmony. While a contrast between oral and nasalized vowels is common in Niger-Congo, it is relatively rare in Nilo-Saharan or Afroasiatic languages.

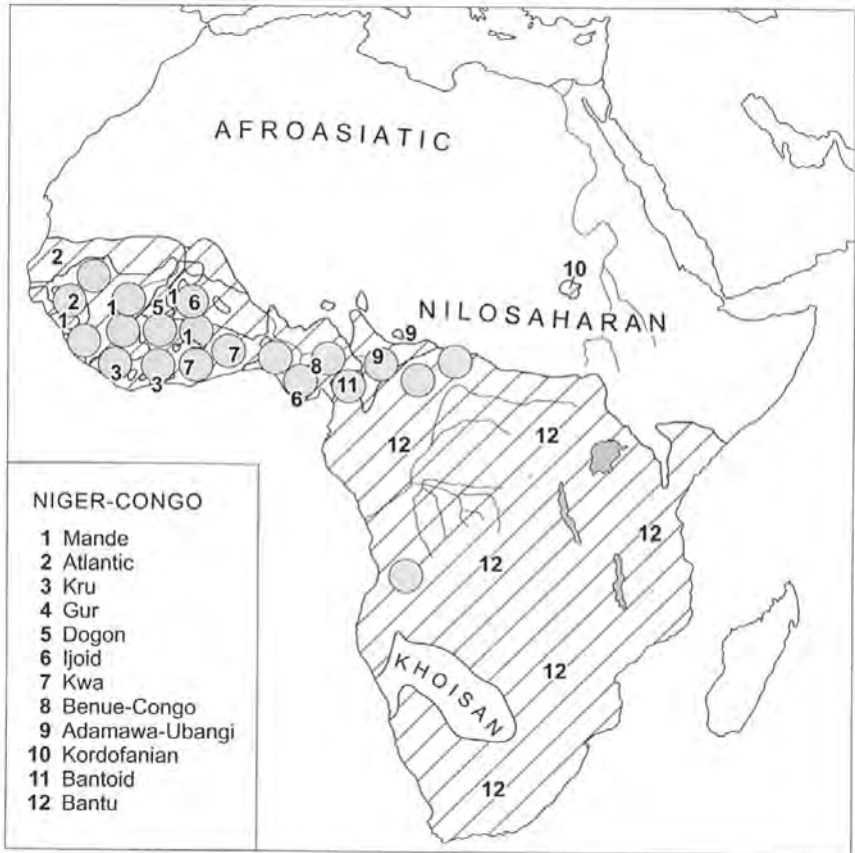
Whereas cross-linguistically nasalized vowels are fairly common (cf. Ladefoged and Maddieson 1996), the synchronic, typological interest in the case of Niger-Congo languages with this feature lies in the fact that several of them have been claimed to have nasalized vowels but no nasal consonants (cf. Bole-Richard (1985) on the Mande languages Dan and Tour, the Kru languages Grebo and Nyabwa, the Kwa languages Ebie and Mbatto, the Gur language Bwamu, the Ubangi language Yakoma, and Edoid (Benue-Congo)).

A survey of the various branches of Niger-Congo shows that nasalized vowels as such are common across the family, even in subgroups such as Mande, which many Africanists would consider to be an early split-off from Niger-Congo. Within subgroups, however, the distribution is at times uneven. For example, whereas ten out of eleven groups in Kwa have nasalized vowels, in Cross-River this phenomenon appears to be restricted to one low-level subgroup, the Ogoni (Kegboid) languages; see Faraclas (1989).

The picture becomes even more intricate when considering Bantoid. Nasalized vowels are common in the north-western Bantu zone, but relatively rare elsewhere in this subgroup. Neither Guthrie (1967–71) nor Meeussen (1967, 1980) reconstruct nasalized vowels for Proto-Bantu. Stewart (1973) assumes, on the basis of historical-comparative evidence, that nasalized vowels were lost in Proto-Bantu by a merger with their oral counterparts. Such mergers of course are not uncommon cross-linguistically. But the actual historical process of loss may have been more intricate and complex. In at least one Bantu language, UMBundu (Angola), there are nasalized vowels, sometimes occurring in forms which must be reflexes of Proto-Bantu roots as reconstructed by Guthrie (1967–71) or Meeussen (1980); compare the latter as a source for Proto-Bantu reconstructions and their reflexes

TABLE 3. Nasalized vowels in Niger-Congo

Atlantic	yes	Benue-Congo	
Mande	yes	Edoid	yes
Kru	yes	Nupoid	yes
Gur	yes	Idomoid	no
Kwa	yes	Defoid	yes
Ijoid	yes	Cross-River	yes
Adamawa-		Kainji	yes
Ubangi	yes	Platoid	yes
Kordofanian	no (?)	Bantoid	yes



MAP 3. Nasal vowels in Niger-Congo

in UMBundu (with nasalization marked below the vowel; data from Schadeberg 1981).

Proto-Bantu	UMBundu	
(14) * -tú _í	é-tw _í	‘ear’
* -dà	óva-là	‘intestines’
* -cù	óva-sù	‘urine’

Spontaneous nasalization does occur in languages, but there is no evidence in terms of a conditioning factor for such a process historically in UMBundu. The historical picture is ‘aggravated’ by the fact that these forms have cognates without any trace of nasality in other Bantu languages, as far as is known at present. What is more, the Proto-Bantu forms have not been reconstructed with nasalized

vowels, but one finds cognates with a nasalized vowel elsewhere in Volta-Congo, for example in cognate forms in the Kwa language Akan (Stewart 1998); compare also the cognate forms for 'blow', which lack a nasalized vowel:

	Proto-Bantu	UMbundu	Akan	
(15)	*-túí	é-twí	a-sú	'ear'
	*-túd-	-téja	-tonú	'forge'
	*-púd-	-fél-a	-huw	'blow'

Consequently, the UMbundu nasalized vowels cannot simply be explained as innovations. Instead, lexical diffusion and gradual loss of nasalized vowels in Bantu languages (independently or through areal diffusion, or a combination of these two processes) appears to be the only plausible alternative explanation. (Prosodic features are known to be particularly prone to diffusion, as shown by Matisoff, this volume.) If the latter hypothesis is correct, the gradual historical loss did indeed result in phonological development towards a common prototype in Bantu. Such areal spreading of phonological as well as grammatical features is common in Bantu, as shown in Guthrie (1967–71).

Nasalized vowels appear to be absent from the Kordofan group (within Niger-Congo), many of which are surrounded by Nilo-Saharan languages. Nasalization of vowels is not common in Nilo-Saharan languages, except in the Ngambay-Mundu and Mbay dialect of Sara, i.e. in one Central Sudanic language bordering on Adamawa-Ubangi (Niger-Congo) languages, where this feature is prominent (as Table 3 shows); areal diffusion from the latter group would therefore appear to be the most plausible explanation.

There are other phonological features of Niger-Congo languages with an areal distribution extending beyond their genetic boundaries, for example labial-velar stops *kp* and *gb*. As pointed out by Greenberg (1983), these universally rare consonants are essentially restricted to Niger-Congo within Africa. Such consonants are also common in the Central Sudanic branch of Nilo-Saharan. In fact, these languages share several typological features with neighbouring Niger-Congo languages (e.g. vowel harmony, as well as the universally rare constituent order type which Heine (1976) called the type B languages; the latter are characterized by S AUX OV word order combined with postpositions, and adverbial clauses which may precede or follow the main clause). Consequently, areal diffusion from Niger-Congo into Nilo-Saharan is the most plausible explanation. Elsewhere (Dimmendaal 1995), it has been shown how such labial-velar consonants entered the eastern representatives of Nilo-Saharan, Nilotic languages such as Alur and Kuku. There is extensive bilingualism amongst speakers of these Nilotic languages and neighbouring Central Sudanic languages, where labial-velar consonants abound, as noted above. The latter sounds entered (Western Nilotic) Alur and (Eastern Nilotic) Kuku in the first place through unadapted lexical borrowing, in particular in ideophonic words, from these neighbouring Central Sudanic languages. But there is a second, language-internal source as well, a shift in

phonetic norm. This becomes evident from a comparison with their closely related relatives. Kuku, for example, is a dialect of the Bari cluster within Eastern Nilotic. Bari proper (and other Eastern Nilotic languages) have labialized velars, where Kuku has labial velar stops as an alternative.

- (16) Bari proper Kuku
 logwake? 'flea' *logwake, logbake* 'tick'

The same historical drift, hardening for labialized velars, can be observed in Western Nilotic Alur. Compare alternative pronunciations such as *kwâyá* or *kpâyá* 'jest, joke' (Dimmendaal 1995).

Although labial-velar stops are widespread in Niger-Congo, their historical status is still problematic. Bendor-Samuel (1971: 155) reconstructs a contrast between simple and labialized velars for Gur. But, as pointed out by the author, the latter frequently have labial-velar reflexes in the present-day languages, as a result of partly independent shifts in phonetic norm. If we recognize the importance of diffusion through areal contact between languages (whether closely related or distantly related) as an important trigger for phonetic change, such potential complications in the establishment of sound correspondences receive a natural explanation.

3.3. NOUN CLASSES

Noun class systems are found in different areas across the world: the east Caucasus (Rieks Smeets, p.c.), the Amazon and some Papuan languages (Sasha Aikhenvald, p.c.), and northern Australia (Bob Dixon, p.c.). On the African continent, this classificatory system is essentially restricted to Niger-Congo. Here we find a system of up to twenty or so individual classes paired in genders, with agreement marking by way of concordial noun-class markers, as well as cross-reference marking on verbs for subjects and objects whose shape depends on the noun-class with which they are co-indexed. Williamson (1989a) provides a survey of the various ways in which such noun class systems are expressed across Niger-Congo. This internal variation within the family is summarized in Table 4.

It is important, from a genetic point of view, to observe that several of the noun classes are cognate, obviously going back to a common ancestral form. Compare for example the gender-bearing human referents among its prototypical members, class 1 (sg) in comparative Bantu terminology, which has a cognate *gu-* in Kordofanian groups such as Heiban, *gu-* in Proto-Atlantic, *o-* in Proto-Togo Remnant, *ù-* in Proto-Benue-Congo, and a suffix *-o* in Gur. (This variation between prefixal and suffixal forms, as in Gur, is further discussed below.)³ Similarly, class 3 (containing 'tree' or tree names amongst its prototypical members) has *gu-* in Heiban (Kordofanian), *gu-* in Proto-Atlantic, *o-* in Proto-Togo Remnant, *ú-* in Proto-Benue-Congo. The noun class 5 is reconstructed as

³ Reconstructions are based on Doneux (1975) for Atlantic, Heine (1968) for Togo Remnant, de Wolf (1971) for Benue-Congo, Manessy (1975) for Gur; see also Williamson (1989a).



MAP 4. Noun classes in Niger-Congo

**li-* for Proto-Heiban (Kordofanian), **de-* in Proto-Atlantic, **li-* in Proto-Togo Remnant, **li-* in Proto-Benue-Congo, and again a suffix *-li* in Gur. The noun class containing terms for liquids, class 6 (in Bantu terminology) has **N-* in Proto-Heiban (Kordofanian), **ma-* in Proto-Atlantic, *N-* in Togo Remnant, **ma-* in Proto-Benue-Congo, and *-ma* in Oti-Volta (Gur). This short summary does not constitute an exhaustive listing.

Williamson (1989a: 31–40) assumes an original system of prefixation. This hypothesis is highly plausible for a number of reasons. Firstly, the geographical distribution of these alternative systems across Niger-Congo as a whole suggests an original prefixation, rather than a suffixation, system. Extensive noun-class prefixation systems are common in geographically distant subgroups such as Kordofanian, the (Northern) Atlantic branch, the Togo Remnant languages within Kwa, and major Benue-Congo groups such as Bantoid or Cross-River (see also

TABLE 4. Noun-class affixation in Niger-Congo

	Noun	Agreement
Atlantic	prefixes reconstructed; also suffixation	yes
Mande	no evidence	no
Kru	suffixation	yes
Gur	prefixes (innovating suffixes and proclitics)	yes
Kwa	(remnant) prefixes	yes
Ijoid	(reduced) prefixation, suffixation (and gender)	remnant form
Benue-Congo		
Defoid	petrified prefixes	no
Edoid	reduced prefixation	yes
Idomoid	(petrified) prefixation	yes
Nupoid	reduced (petrified; some suffixation)	no
Igboid	petrified prefixes	no
Cross River	prefixes (partly reduced)	yes
Kainji	prefixation (partly reduced)	yes
Platoid	towards suffixation	yes
Bantoid	prefixation (reduction and loss in north)	yes
Adamawa-Ubangi	(petrified suffixes; no evidence in Ubangi?)	no
Kordofanian	prefixation	yes

Map 3). And, as explained above, several of these markers are cognate. Secondly, reconstruction work at lower level units with internal variation as to the position of noun-class affixes relative to the root, such as Atlantic, unambiguously points towards an earlier prefixation system. Thirdly, plausible mechanisms of diachronic change can be invoked, explaining how one moves from a prefixation to a suffixation system (whereas assuming an inverse process would not work). Let us have a closer look at the second and third argument.

Reduced prefixation systems are easily explained as natural historical outcomes of more elaborate systems. Reduction, as well as total loss, is found in various Kwa groups (Comoe, Gbe), and Benue-Congo groups such as Central Jukunoid, Platoid, or the Lower Cross branch of Cross-River. As pointed out by Faraclas (1989: 389), 'Cross River nominal class-concord systems . . . typify almost every possible stage of simplification of the proto-Benue-Congo system, from full retention in some conservative Upper Cross and Bendi languages to near complete elimination in the Ogoni group'. Whereas generally speaking the degree of productivity of these systems indeed seems to coincide with that of neighbouring Niger-Congo groups, there are also striking exceptions; the Cross-River language Kohumono, for example, has a highly elaborate noun-class system whose complexity contrasts with that of the (petrified) system of the neighbouring Igboid group (which also belongs to Benue-Congo).

In order to explain the positional variation in noun class affixation across

Niger-Congo, Welmers (1973: 209) took recourse to the principle of archaic heterogeneity. But the question whether Proto-Niger Congo had prefixes, suffixes, or both, can now be answered less ambiguously. There is a well-known mechanism described by Greenberg (1978), which helps to account for this variation in an elegant and plausible manner. It involves the grammaticalization of (one set of) demonstratives into simple nominal markers along a number of stages:

- Stage I: reduction into a definite article
 Stage II: widening of distribution and development into a non-generic article, used in contexts in which a specific but unidentified item is referred to, i.e. there is a presupposition of reference
 Stage III: obligatory (grammaticalized) marker of nouns

The Atlantic branch, for which an original system of noun-class prefixation has been reconstructed by Doneux (1975) and De Wolf (1971), may help to illustrate how these historical reinterpretations affect the position of noun classes relative to nominal roots. A number of authors have shown how nominal specification or definiteness-marking may affect the noun-class system synchronically in Atlantic languages. Sapir (1965: 68) has shown for Diola-Fogny (Senegal) that definiteness-marking in this language involves the suffixation of a 'binder' (mostly *a-*) plus the concordial class marker. It would be erroneous therefore to think that prefixes and suffixes are the same markers in these Atlantic languages, they obviously are not, although they are grammatically related.

- (17) *ji-sek* *mu-sek* 'small woman' 'small women'
 e-yen *si-yen* 'dog' 'dogs'
- (18) *ji-sek-aj(u)* 'the small woman'
 mu-sek-am(u) 'the small women'
 e-yen-ey 'the dog'
 si-yen-as(u) 'the dogs'

Once suffixal markers (going back to nominal modifiers) start losing their sense of definiteness and become the regular noun-class markers (a complex reinterpretation process which itself needs to be explained), one may end up with a system of circumfixes, as in the Atlantic language Serer. Prefixes are rare in Serer plurals; instead, consonant alternation occurs (Mukarovsky 1983). This remnant of erstwhile prefixation has become the general pattern in Fulfulde, where suffixes are the regular markers of singular and plural noun classes, without such suffixes adding a sense of definiteness (Breedveld 1995, Mukarovsky 1983).

- (19) Serer Fulfulde
 o-kor-oxa *gor-ko* 'man'
 gor-va *wor-fe* 'men'
 o-tew-oxa *debb-o* 'woman'
 rew-va *rew-fe* 'women'

Noun-class suffixation may also arise from definiteness marking with concomitant deletion of the nominal prefix. Such a synchronic system has been described by Hoffmann (1967) for the Platoid (Benue-Congo) language Dakarkari:

- (20) *d-gyàŋ* 'egg' *c-gyàŋ* 'eggs'
gyàn dónè 'the egg' *gyàn cónè* 'the eggs'

Within Benue-Congo, there are indeed languages (e.g. Tiv) or groups (within Kainji as well as Platoid) which have developed suffixal systems with simultaneous loss of prefixation. Such suffixation systems are also attested elsewhere in Niger-Congo, e.g. in Kru, Adamawa, and Gur. In Adamawa, these have become petrified nominal elements (Stefan Elders, p.c.). Occasionally, derivational suffixes have been misinterpreted as noun-class suffixes by investigators, as shown by Storch (1997) for Jukun; whereas some of the more central varieties of this Benue-Congo branch have innovated, other varieties have retained classical Niger-Congo noun-class prefixes.

Concord systems generally speaking are more conservative than markers on the noun across Niger-Congo. There are, however, areas in particular in Nigeria where concord prefixes have decayed more than noun prefixes; this applies, for example, to Edoid, (south-west) Plateau, and Ijoid. A similar system of noun-class prefixation without agreement has been reported for northern Bantu borderland languages such as Amba, Bera, Bhele, Bila, Kaiku, and Komo, which are bordering on Ubangian (Niger-Congo) and Central Sudanic (Nilo-Saharan) groups; these latter do not have noun-classes systems. (As pointed out by Conny Kutsch Lojenga (p.c.), singular/plural distinctions are only functional with animate nouns in these Bantu languages.)

Maintenance of agreement with reduction of nominal class-marking on the head noun is a natural historical drift across Niger-Congo, as argued by Demuth, Faraclas, and Marchese (1986). Alternatively, loss of agreement and subsequent petrification and loss of marking on the head noun typically arises in contact situations with languages lacking noun classes altogether, it seems. In the case of the Nigerian Niger-Congo languages, contact with neighbouring Chadic languages may have triggered this early loss of agreement. The case of Ijoid, a language group spoken in the Niger-Delta and thus geographically far apart from Chadic (at least today), remains enigmatic, also because this group has developed gender-marking on nouns and a verb-final syntax, typological features which are absent elsewhere in the area.

An obvious, related question to ask would be how definiteness-marking on the noun was expressed in the earliest stages of Niger-Congo. There appear to be two plausible options: (a) by way of independent markers following the head noun (as already illustrated for Dakarkari above); (b) by way of pre-prefixes or augments preceding the nominal class prefix. This latter strategy is common, for example, in Bantoid. Augments have been reconstructed for Proto-Bantu by Meeussen (1967). Their original function has been retained in Bantu languages like Dzamba; in others, Herero for example, the augmented form has become the normal form of the noun

(a process which again seems to follow the historical drift discussed by Greenberg (1978)). Bantu languages like Swahili have abandoned the use of augments altogether, and have taken recourse to the use of demonstratives (following the head noun) to enhance the interpretation of definiteness in nominal reference.

There is now growing evidence that the origin of the augment predates Bantu; as Williamson (1993) has argued, for example, the Igboïd group within Benue-Congo has traces of such an earlier augment or pre-prefix. It is possible therefore that this strategy was old in Niger-Congo, and in use before recourse was taken to alternative (postnominal) strategies, also because the shape of noun-class prefixes in a number of Niger-Congo subgroups outside Bantu appears to be better accounted for historically, if the prefixes are taken to be reflexes of erstwhile (CV-) prefixes preceded by (V-) augments. (Compare also the description of the Atlantic language Temne by Creissels (1991: 93).)

Noun-class systems may emerge as a result of areal contact. This we notice, for example, in the Nilotic (Nilo-Saharan) language Luo, which has undergone extensive lexical borrowing from neighbouring Bantu languages, e.g. of nouns which have retained their singular and plural noun-class prefix in Luo. In addition, there is a language-internal process whereby heads of endo-centric compounds are developing into noun-class prefixes, as with the word for 'mouth' *dhək* (Dimmendaal, forthcoming).

(21) *dhó-lúo* 'the Luo language'

The Janus-like properties of *dhó-*, between that of lexical head and prefix, is manifested in the morphophonemic alternation, e. g. the shift of [–ATR] *ɔ* to [+ATR] *o*, and the deletion of the velar stop (compare *dhək*), a process which lexical heads in Luo compounds do not undergo. The emerging noun-class system is still different, however, from that of neighbouring Bantu languages, in that there is no agreement-marking on nominal modifiers in Luo. The typological *Umwandlung* itself is the outcome of massive language shift from Bantu (Niger-Congo) towards a Nilo-Saharan language (Dimmendaal, 2001).

3.4. SERIAL VERBS

A stereotypical view of African languages sometimes encountered in the general literature is the presence of serial verb constructions. In actual fact, this phenomenon has a rather restricted distribution both genetically and areally. It is found in a largely contiguous zone stretching from the Ivory Coast to Nigeria, in languages belonging to different subgroups of Niger-Congo: Kwa, Western Benue-Congo, and Ijoid; in addition, it is found in neighbouring Gur languages such as Dagbani, Kasem, Mampruli, Supyire, or Vagala, or Mande languages such as Ligbi.⁴ (See

⁴ See also Carlson (1994) for an interesting discussion of structural differences between serial verb constructions and consecutivization in Supyire, a Gur language typologically akin to Kwa languages to its south and south-east.



MAP 5. Serial verbs in Niger-Congo

also Map 5.) The same phenomenon is attested in Atlantic creoles, for which a West African substratum is commonly assumed.

It is important, from a typological and historical-comparative perspective, to distinguish serial verb constructions from the more common phenomenon of verb consecutivization. The latter is distinct from verb serialization in several respects: consecutivization allows for separate negation-marking on the two verbs; also, the sequential (temporal) order for the two verbs is crucial; moreover, there is no need for shared objects; and usually, no lexicalization is involved in such consecutive verb constructions.

Consecutives are common in African languages belonging to different genetic groupings, and with different constituent order types. Moreover, they are found in languages with opulent verb morphologies as well as in languages with more restricted morphological systems. What these languages share in common is a

strategy whereby 'and' is avoided as a clausal connector. Compare Turkana, a Nilotic language of Kenya (data from author):

- (22) à-pàm-ɪ akimuj daaŋ(ɪ) to-lot nàwuy
 3.PAST-eat-ASP food all 3.CONSGO home
 '(S)he ate all the food, and went home.'

Serial verbs on the other hand function as single predicates, they share either subject or object arguments, and refer to a single event on a par with monoverbal clauses; usually, no independent tense/aspect/modality/illocutionary force-marking is found on the second verb, negation has scope over the entire clause, and restrictions occur on the expression of pronominal subjects and/or objects in such complex predicates; if the main verb occurs clause-finally (as in Ijoid; Williamson 1965), full inflection tends to be found on the final rather than the preceding verb.

Out of context, a clause may have an ambiguous reading between a complex predicate and consecutivization, as pointed out by Ikoro (1996) for the Cross River language Kana. Apart from the features mentioned above, important clues for their distinct syntactic status come from the optional use of connectives in consecutivization; other tests for constituency usually are relativization, 'extraction' in order to mark focus, topicalization, prosodic (phrasal) phonology, and nominalization; see also Déchaine (1993) for a lucid discussion.

Cross-linguistically, there are different types of serial verb constructions (compare Déchaine 1993, Durie 1997). Periphrastic causatives ('make/give/let x do y'), for example, are widespread cross-linguistically. The latter, however, are quite compatible with verbal valency-changing markers in a particular language, or with case-marking strategies. In this sense, they are not a predictor of a language type. A central feature of West African languages with serial verbs, however, is the lack of three-place predicates. Instead, a second verb (prepositional verb, coverb, verbid) is required to host a third argument. With such serial verb constructions, one can usually distinguish an asymmetrical type, with one verb being derived from a large open class, the other verb being selected from a small, closed set, e.g. 'give' in dative constructions ('do x give y'), or 'send' in locative constructions. Compare again the Nigerian Cross River language Kana as described by Ikoro (1996):

- (23) mǎǎ dàǎǎ kpé má dògò
 I.PROG lift bicycle send.to Dogo
 'I am carrying a bicycle to Dogo.'

Whereas in constructions involving semantic roles such as benefactive or location it is the second verb which is derived from a small set, it is generally the first verb in constructions expressing instrument, manner, or comitative that is drawn from such a restricted set of 'coverbs'.

Basing himself upon extensive experience with verb serialization in Papuan languages, Andy Pawley (p.c.) has suggested that the absence of three-place

TABLE 5. Serial verbs in Niger-Congo

	Serial verbs	Derivational suffixes
Atlantic	no	yes
Mande	no	yes
Kru	no	yes
Gur	yes	yes
Kwa	yes	no
Ijoid	yes	no
Benue-Congo		
Defoid	yes	no
Edoid	yes	no
Nupoid	yes	no
Idomoid	yes	no
Igboïd	yes	no
Cross River	yes	yes

predicates in these West African languages may be epiphenomenal. Whereas this observation no doubt is justified with respect to Papuan languages, where a continuum of clause chaining strategies occurs, the situation in the case of West African Niger-Congo languages must be different. Outside this spread zone for serial verbs, there is a widespread tendency to use verb morphology (head-marking) in order to express valency modification, for example in such distantly related members as Atlantic, Kru, and Bantoid (see also Voeltz 1977). In Mande, there is a rather restricted degree of verbal valency marking; here, adpositions play an important role, a strategy which is found as a concomitant feature of serial-verb languages and languages using verbal valency-marking.

Presumably, the most extensive system of head-marking (on the verb) is found in Atlantic and Bantu. A typical example from the latter group is the applicative (expressing benefactive, malefactive, motion towards, purpose), as in Swahili:

- (24) **ni-ku-pik-i-e** **chakula**
 1sg.SUBJ-2.O-cook-APPL-SUBJ food
 'Shall I cook some food for you?' (compare *-pika* 'cook')

Such language types may be further divided into the so-called asymmetrical type, where only one of the postverbal NPs exhibits primary object syntactic properties (e.g. passivizability, object-marking on the verb, adjacency to the verb), or the symmetrical type language; in the latter more than one NP can display 'primary object' syntactic properties.

Such verbal strategies are notably absent from the area where serial verbs are used. Interestingly, serial verbs are not common in the Togo Remnant languages of Ghana and Togo. These Kwa languages, situated in the heart of the area where serial verbs are common, (still) have verbal extensions. They combine three-place

argument structures (including the use of prepositions) with serial-verb constructions, as shown by Ford (1988). Note, *inter alia*, that the same languages are also rather conservative with respect to their (extensive) noun-class systems.

The historical shift away from head-marking on the verb, and towards a complex predicate type with two (serial) verbs, does not merely involve a shift towards an alternative morphosyntactic strategy; languages in the transitional zone between these typological zones make this clear. Compare again Kana, a Cross River language spoken in the eastern border area where serial verb languages meet those using verbal valency-marking. Kana has serial verbs as well as verb-marking strategies, often used as alternative strategies dictated by information structure (Ikoro 1996), as the following examples help to illustrate; by using a serial verb 'give', the verbal act of feeding is focused upon.

- (25) **bàrilè āā nè zǐá ɲwíí**
 Barile PROG give food child
 'Barile is feeding a child.'

- (26) **bàrilè āā sú zǐá nè ɲwíí**
 Barile PROG take food give child
 'Barile is *feeding* a child.'

Such a restructuring (e.g. through a 'reworking' of existing constructions, for example by using 'give' as a serial verb) accordingly has major consequences for the organization of the syntax–semantics interface.

The interesting question is: when and how did this innovation, the use of serial verb constructions, spread as an areal feature? Whereas the problem of its actualization probably will remain unsolved, the current distribution of this phenomenon (as well as other morphosyntactic and phonological features) calls for a link with social developments in the area as one relevant factor. There are at least three major languages in the area, Akan, Yoruba, and Igbo, whose dominant status is associated historically with the establishment of centralized states, and with the founding of major cities in the area. This innovation begs for inferences about changing (expanding) network structures in urban settings, all the more since Akan, Yoruba and Igbo are also intercommunity or contact languages. Presumably, important social significance was attached to such morphosyntactic innovations; they became emblematic features, and copying them may have served as an act of identity.

Given the vast number of speakers for languages such as Akan, Yoruba, and Igbo, language shift must have occurred in favour of these intercommunity languages, next to language maintenance with copying of this feature (and other typological properties). The fact that a number of smaller, isolated communities such as those speaking (conservative) Togo Remnant languages, were not affected by the diffusion of this typological feature (or less so), suggests that they were not incorporated into these larger networks.

From a historical-comparative point of view, the spreading of verb serialization represents a further instantiation of areal convergence towards a common prototype, next to such features as vowel harmony, nasalized vowels, labial-velar stops, and reduced noun-class systems. There appears to be no way of determining whether these changes occurred rapidly or, alternatively, over a period of a few thousand years.

In spite of these diffusion processes, the genetic classification did not become obliterated, for example with respect to Eastern Kwa and Western Benue-Congo or Ijoid, because diffusion involved copying of typological features, without extensive grammatical borrowing.

4. Some answers and some further questions

When plotting areal features of African languages on a map, for example isoglosses representing phonological characteristics, the result is not always a bunching of exclusively shared isoglosses. Tone, for example, constitutes an ancient diffusional trait, covering major parts of the continent. Within this area there are languages sharing ATR-harmony, and within as well as outside the vowel-harmony zone, there are languages with nasalized vowels. Consequently, the emerging synoptic chart is reminiscent of dialect maps. (Compare also Hock (1988) for similar statements regarding areal diffusion and convergence in the Eurasian area.) Nevertheless, it is still possible to define one or several *Sprachbünde*, with the proviso that some features extend beyond their respective territories, or did so in the past.

What does this mean in historical terms? Or, phrased alternatively, how do we account for this either in terms of diffusion or inheritance and loss? Relative chronology appears to be a key factor. Tone is found in Niger-Congo, Nilo-Saharan as well as Khoisan. Whether this situation came about as a result of areal diffusion (resulting in a common prototype), we will probably never know. There is evidence on the other hand that neighbouring Afroasiatic languages developed tone through areal diffusion from Niger-Congo and Nilo-Saharan languages.

Apart from suprasegmental and segmental phenomena, there are other widespread, and therefore ancient, Africanisms, for example lexical idiosyncrasies such as the frequent non-distinctness of a term for 'smell' and 'hear' (compare Swahili: *nasikia samaki* 'I (can) smell fish'; *-sikia* also means 'hear'). These may have originated in one of the major phyla still found today, but our historical methods do not allow us to reconstruct the ultimate source, except in transitional areas.

The position defended in this contribution is that there is little evidence for extensive morphological borrowing in African languages; noun classes, for example, are the result of genetic inheritance, not diffusion. Diffusion of morphological properties does occur, also between dialects or closely related languages (compare also the contribution by Heine and Kuteva, this volume), and this may complicate historical-comparative work. Nevertheless, it is possible to reconstruct

earlier systems, as long as distantly related members have not been affected by such areal diffusion, and retain essential properties of their common ancestor. This methodological principle also applies to syntactic phenomena such as serial-verb constructions, which are restricted to West African representatives of Niger-Congo (many of which also share reduced noun-class systems and phonological properties such as nasalized vowels and vowel harmony). Here the picture is somewhat reminiscent of a 'Big Bang', with languages surrounding the zone where serial verbs are common representing more archaic systems (as shown through the comparative method). Moreover, and this is important from a methodological point of view, areal diffusion did not obscure the original genetic relationship (although it did confuse researchers at one point).

In a sociolinguistic situation where multilingualism is the norm—as is true for large parts of Africa—areal diffusion is what one expects. The observed changes, however, would appear to be rather different from those described by Dixon (this volume) for Australia, or Aikhenvald (this volume) for the Amazon. So how come? Maybe Africanists have not looked hard enough for areal diffusion of morphological material, an option which cannot be entirely excluded as a potential explanation. But Africanists like to think they do know a fair amount about areal types and genetic subgrouping, and the considerable degree of consensus amongst scholars seems to confirm this. Neither the type of lexical diffusion nor the type of grammatical borrowing found apparently in Australian languages or languages of the Amazon are common in Africa. There are no obvious linguistic reasons for this, e.g. in terms of typological distance, since these would appear to be as large in Africa as anywhere else in the world. Moreover, to invoke Goddard's Law (Watkins, this volume) 'a language can do whatever it wants to with whatever material it has to hand, if it wants to'. And so any typological gap can in principle be bridged. The small size of speech communities appears to be one factor which sets the situation in Australia and the Amazon apart from that in Africa. In these Australian and Amazonian communities, intensive social networking with other groups speaking distinct languages usually was or still is needed, in order to create sizeable production units. This in turn may create a basis for heavy borrowing also at the grammatical level. Such small-scale societies are also found traditionally in the region where Khoisan languages are spoken. Unfortunately, our understanding of the historical divergence (and convergence) between many of these languages is still incomplete. However, Vossen (1997), in his pioneering historical-comparative work on Central Khoisan languages, does not report any evidence for the type of deep morphological borrowing observed for Australia or the Amazon. This would leave a deeply rooted difference in the role played by language in contact situations in these African communities as against Australia and the Amazon as the only plausible, alternative explanation for the observed distinct outcomes. Only intensive research in the Khoisan area can help us to create a deeper understanding of the social background to these diverging attitudes. Hopefully, we will

succeed in completing this research endeavour before most of the Khoisan languages have disappeared.

Appendix—The subclassification of Nilo-Saharan according to Greenberg (1963)

1. Songhai
2. Saharan (group)
3. Maban (group)
4. Fur
5. Chari-Nile
 - Eastern Sudanic
 1. Nubian (group)
 2. Murle, Longarim, Didinga, Suri, Mekan, Murzu, Surma, Masongo
 3. Barea
 4. Ingassana
 5. Nyima, Afitti
 6. Temein, Teis-um-Danab
 7. Merarit, Tama, Sungor
 8. Dagu (group)
 9. Nilotic (group)
 10. Nyangiya, Teuso
 - Central Sudanic (group)
 - Kunama
 - Berta
6. Coman (group)

References

- ARCHANGELI, D. B., and PULLEYBLANK, D. 1994. *Grounded Phonology*. Cambridge, MA: MIT Press.
- BENDOR-SAMUEL, J. 1971. 'Niger-Congo: Gur', pp. 141–78 of Sebeok (1971).
- 1989. Editor of *The Niger-Congo Languages*. Lanham: University Press of America.
- 1992. 'Niger-Congo', pp. 93–100 of *International Encyclopedia of Linguistics*, edited by W. Bright, Oxford : Oxford University Press.
- BOLE-RICHARD, R. 1985. 'Hypothèse sur la genèse de la nasalité en Niger-Congo', *Journal of West African Languages* 15:2.3–28.
- BREEDVELD, J. O. 1995. *Form and Meaning in Fulfulde: A Morphophonological Study of Maasinangkoore*. Leiden: Centre for Non-Western Studies.
- CAPO, H. B. C. 1985. 'On the high non-expanded vowels in Yoruboid', *Studies in African Linguistics* 16:1.103–21.
- CARLSON, R. 1994. *A Grammar of Supyire*. Berlin: Mouton de Gruyter.
- CREISSELS, D. 1991. *Elements de syntaxe generale*. Paris : Presses Universitaires de France.
- 1994. *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: ELLUG.
- DE BLOIS, R. 1981. 'On the Origin of Underlying Cross-Height Vowel Harmony in Nen'. MA dissertation, Leiden University.

- DÉCHÂINE, R.-M. 1993. 'Serial verb constructions', pp. 799–825 of *Syntax: ein internationales Handbuch zeitgenössischer Forschung*, vol. 1, edited by J. Jacobs *et al.* Berlin: Walter de Gruyter.
- DEMUTH, K., FARACLAS, N., and MARCHESE, L. 1986. 'Niger-Congo noun class and agreement systems in language acquisition and historical change', pp. 453–71 of *Noun Classes and Categorization*, edited by C. Craig. Amsterdam: John Benjamins.
- DE WOLF, P. P. 1971. *The Noun-Class System of Proto-Benue-Congo*. The Hague: Mouton.
- DIMMENDAAL, G. J. 1993. 'Review of Keith Denning and Suzanne Kemmer (eds.), *On Language: Selected Writings of Joseph H. Greenberg* (1990)', *Language* 69.796–807.
- 1995. 'The role of bilingualism in Nilotic sound change', pp. 85–109 of *Sound Change*, edited by M. Dominicy and D. Demolin. Amsterdam: John Benjamins.
- 1998a. 'Surmic languages and cultures: an introduction', pp. 1–31 of Dimmendaal and Last (1998).
- 1998b. 'A syntactic typology of Surmic from an areal and historical-comparative point of view', pp. 32–85 of Dimmendaal and Last (1998).
- 2001. 'Language shift and morphological convergence in the Nilotic'. *Sprache und Geschichte in Afrika* 16/17: 83–124.
- and LAST, M. 1998. Editors of *Surmic Languages and Cultures*. Nilo-Saharan 13. Cologne: Rüdiger Köppe.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DONEUX, J. L. 1975. 'Hypothèses pour la comparative des langues atlantiques', *Africana Linguistica* 6.41–129.
- DURIE, M. 1997. 'Grammatical structures in verb serialization', pp. 289–354 of *Complex Predicates*, edited by A. Alsina, J. Bresnan, and P. Sells. Stanford: CSLI Publications.
- ELUGBE, B. O. 1989. *Comparative Edoid: Phonology and Lexicon*. Port Harcourt: University of Port Harcourt.
- EVANS, N. 1998. 'Iwaidja mutation and its origins', pp. 115–49 of *Case, Typology and Grammar: In Honor of Barry J. Blake*, edited by A. Siewierska and Jae Jung Song. Amsterdam: John Benjamins.
- FARACLAS, N. G. 1989. 'Cross-River', pp. 377–99 of Bendor-Samuel (1989).
- FORD, K. C. 1988. 'Structural features of the Central-Togo languages', pp. 126–54 of *The Languages of Ghana*, edited by M. E. Kropp Dakubu. London: Kegan Paul.
- GREENBERG, J. H. 1955. *Studies in African Linguistic Classification*. New Haven: Compass Publishing Company.
- 1963. *The Languages of Africa*. Special issue of the *International Journal of American Linguistics* 29:1.
- 1978. 'How does a language acquire gender markers?', pp. 47–82 of *Universals of Human Language*, vol. 3: *Word Structure*, edited by J. H. Greenberg *et al.* Stanford: Stanford University Press.
- 1983. 'Some areal characteristics of African languages', pp. 3–21 of *Current Approaches to African Linguistics*, vol. 1, edited by I. Dihoff. Dordrecht: Foris.
- GRIMES, B. F. 1996. *Ethnologue*. Dallas: Summer Institute of Linguistics.
- GUTHRIE, M. 1967–71. *Comparative Bantu: An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. 4 vols. Farnborough: Gregg Press.
- HEINE, B. 1968. *Die Verbreitung und Gliederung der Togoorestsprachen*. Berlin: Dietrich Reimer.

- 1976. *A Typology of African Languages Based on the Order of Meaningful Elements*. Kölner Beiträge zur Afrikanistik 3. Berlin: Dietrich Reimer.
- HOCK, H. H. 1988. 'Historical implications of a dialectological approach to convergence', pp. 283–328 of *Historical Dialectology: Regional and Social*, edited by J. Fisiak. Berlin: Mouton de Gruyter.
- HOFFMANN, C. 1967. 'An outline of the Dakarkari noun class system and the relation between prefix and suffix noun class systems', pp. 237–59 of *La classification nominale dans les langues négro-africaines*. Paris: CNRS.
- IKORO, S. M. 1996. *The Kana Language*. Leiden: Research School CNWS.
- JUNGRAITHMAYR, H. 1992–3. 'On vowel systems in Chadic', *Folia Orientalia* 29.119–29.
- KOELLE, S. W. 1854. *Polyglotta Africana*. London: Church Missionary House.
- LADEFOGED, P. and MADDIESON, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
- LAST, M., and LUCASSEN, L. 1998. 'A grammatical sketch of Chai, a Southeastern Surmic language', pp. 375–436 of Dimmendaal and Last (1998).
- LYDALL, J. 1976. 'Hamer', pp. 393–438 of *The Non-Semitic Languages of Ethiopia*, edited by M. L. Bender. East Lansing: African Studies Center, Michigan State University.
- MANESSY, G. 1975. *Les langues oti-volta*. Paris: SELAF.
- MARCHESE, L. 1988. 'Noun classes and agreement systems in Kru: a historical approach', pp. 323–41 of *Agreement in Natural Language: Approaches, Theories, Descriptions*, edited by M. Barlow and C. A. Ferguson. Stanford: Center for the Study of Language and Information.
- MEEUssen, A. E. 1967. 'Bantu grammatical reconstructions', *Africana Linguistica* 3.79–121.
- 1980. *Bantu Lexical Reconstructions*. Tervuren: Musée royal de l'Afrique Centrale.
- MOGES YIGEZU and DIMMENDAAL, G. J. 1998. 'Notes on Baale', pp. 233–80 of Dimmendaal and Last (1998).
- MUKAROVSKY, H. 1976–1977. *A Study of Western Nigritic*, vols. 1 and 2. Beiträge zur Afrikanistik 1–2. Vienna: Afro-Pub.
- 1983. 'Die Nominalklassen im Serer und im Ful', *Afrika und Übersee* 66.175–90.
- NADEN, T. 1989. 'Gur', pp. 141–68 of Bender-Samuel (1989).
- NEWMAN, P. 1995. *On Being Right: Greenberg's African Linguistic Classification and the Methodological Principles which Underlie it*. Bloomington, IN: Institute for the Study of Nigerian Languages and Cultures.
- NURSE, D., and HINNEBUSCH, T. 1993. *Swahili and Sabaki: A Linguistic History*. Berkeley: University of California Press.
- RANDAL, S. 1998. 'A grammatical sketch of Tennenet', pp. 219–72 of Dimmendaal and Last (1998).
- ROSS, M. D. 1997. 'Social networks and kinds of speech-community event', pp. 209–61 of *Archaeology and Language*, vol. 1: *Theoretical and Methodological Orientations*, edited by R. Blench and M. Spriggs. London: Routledge.
- SAPIR, J. D. 1965. *A Grammar of Diola-Fogny: A Language Spoken in the Basse-Casamance Region of Senegal*. Cambridge: Cambridge University Press.
- SCHADEBERG, T. C. 1981. 'Nasalization in UMBundu', *JALL* 4.109–32.
- 1997. *De Swahili-talen van Moçambique* [The Swahili languages of Mozambique]. Amsterdam: Koninklijke Academie van Wetenschappen.
- SEBEOK, T. E. 1971. Editor of *Current Trends in Linguistics*, vol. 7: *Linguistics in Sub-Saharan Africa*. The Hague: Mouton.

- STEWART, J. M. 1967. 'Tongue root position in Akan vowel harmony', *Phonetica* 16.185–204.
- 1970. 'Tongue root position in the Volta-Comoe languages and its significance for the reconstruction of the original Bantu vowel sounds', *African Language Studies* 11.340–50.
- 1971. 'Niger-Congo, Kwa', pp. 179–212 of Sebeok (1971).
- 1973. 'The lenis stops of the Potou Lagoon languages and their significance for pre-Bantu reconstruction', pp. 1–49 of *Papers in Ghanaian Linguistics*, edited by M. E. Kropp Dakubu. University of Ghana Institute of African Studies Research Review, Supplement 4.
- 1976. *Towards Volta-Congo Reconstruction*. Inaugural lecture. Leiden: Leiden University Press.
- 1983. 'The high unadvanced vowels of Proto-Tano-Congo', *Journal of West African Languages* 13:1.19–36.
- 1994. 'The comparative phonology of Gbe and its significance for that of Kwa and Volta-Congo', *JALL* 15.175–93.
- 1998. 'An explanation of Bantu vowel height harmony in terms of a pre-Bantu nasalized vowel lowering'. Paper presented at the 28th Colloquium on African Languages and Linguistics, Leiden.
- and LEYNSEELE, H. VAN 1979. 'Underlying cross-height vowel harmony in Nen (Bantu A.44)', *JALL* 1.31–54.
- STORCH, A. 1997. 'Where have all the noun classes gone? A case study of Jukun', *JALL* 18.157–70.
- VOELTZ, E. 1977. 'Proto-Niger-Congo Verb Extensions'. Ph.D. dissertation, UCLA.
- VOSSEN, R. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Cologne: Rüdiger Köppe.
- WELMERS, W. 1973. *African Language Structures*. Berkeley: The University of California Press.
- WESTERMANN, D. 1927. *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Beiheft zu den Mitteilungen des Seminars für Orientalische Sprachen, Jahrgang 29. Berlin: Walter de Gruyter.
- WILLIAMSON, K. 1965. *A Grammar of the Kolokuma Dialect of Ijò*. West African Language Monographs 2. Cambridge: Cambridge University Press.
- 1983–4. 'Vowel merger in harmony languages', *Journal of the Linguistic Association of Nigeria* 2.61–82.
- 1989a. 'Niger-Congo overview', pp. 3–45 of Bendor-Samuel (1989).
- 1989b. 'Benue-Congo overview', pp. 247–74 of Bendor-Samuel (1989).
- 1993. 'The noun prefixes of New Benue-Congo', *JALL* 14.45.

Convergence and Divergence in the Development of African Languages

Bernd Heine and Tania Kuteva

In the present chapter a bird's-eye view of some problems associated with contributions to the linguistic history (or prehistory, as some would say) of Africa is offered. It has been stimulated by recent contributions on linguistic methodology, especially by attempts to relate linguistic findings to more general observations on the evolution of the human species. The conclusion reached is that contact-induced language change and the implications it has for language classification in Africa are still largely a *terra incognita*.

1. Introduction

For roughly half a century, work on the reconstruction of African languages and their relationship has been based on the work of Joseph Greenberg (1949, 1955, 1963). What this work has established in particular are findings such as the following:

- (a) The most easily accessible way of describing the historical relationship of these languages is by reconstructing their genetic relationship patterns.
- (b) The multitude of African languages can be reduced to four genetically defined units, called families by Greenberg and phyla by others. These units are Niger-Congo (or Niger-Kordofanian), Nilo-Saharan, Afroasiatic, and Khoisan.
- (c) There are various methods available to the linguist for historical reconstruction. The task of the linguist is to choose that method that appears to be best suited to solve a particular problem. Which method is most suitable in a given

The present paper is based on research that has been generously funded by the Deutsche Forschungsgemeinschaft (German Research Society) to which we want to express our deeply felt gratitude. The paper has benefited greatly from discussions with and critical comments by colleagues at the Institut für Afrikanistik of the University of Cologne. We wish to express our gratitude most of all to Erhard Voeltz, Ulrike Claudi, Marianne Bechhaus-Gerst, and Friederike Lüpke for many suggestions. We are also grateful to Sasha Aikhenvald, Lionel Bender, Hilary Chappell, Gerrit Dimmendaal, Bob Dixon, Geoffrey Haig, Malcolm Ross, and other contributors to the present volume for valuable comments.

situation depends primarily on the time factor involved. Historical processes that happened within the last century require different methods of analysis from processes dating back a thousand or two thousand years.

- (d) Most of the work to establish language phyla in Africa, including that of Greenberg (1963), has used the method of resemblances, which is based on the assumption that in order to establish that two or more languages are genetically related, or to determine the degree to which they are related, one simply needs to demonstrate that these languages share a sufficient number of lexical (and/or grammatical) items that are similar in form and meaning. The main problems associated with this method concern the question of how the notions 'sufficient number' and 'similarity in form and meaning' can be defined.
- (e) On account of such problems, many students of African linguistics consider this method to be of doubtful value, and some would reject it altogether, arguing that reliable reconstructions of genetic-relationship patterns can only be achieved by means of the comparative method. According to Nichols (1992: 2), this method works reliably only up to a time depth of roughly 8,000 years. So far, it has not been possible to apply the comparative method appropriately to any of the four African language phyla.

Greenberg's genetic classification of African languages is by now widely accepted, but it also leaves many questions on the prehistory of Africa unanswered. Reconstructing family trees is helpful to define one kind of historical process, but it contributes little to our understanding of what has happened in Africa for example in terms of linguistic interaction across languages. With the present chapter we wish to draw attention to the need that exists to study language contact and the ways it may be relevant to linguistic classification in Africa.

2. Language contact

While there are a number of studies on how African languages influence one another, we know little about how this affects linguistic relationship. Still, there are a few studies that suggest that areal forces and linguistic relationship based on contact between languages may cut across genetic boundaries, and a number of convergence areas (or areal groups) have been identified.

2.1. AREAL LINGUISTICS

To start with, there is some evidence to suggest that the African continent forms a convergence area of its own. There are very few linguistic properties that are found almost only in Africa, like click consonants, which occur in all southern African and East African Khoisan languages, in many Bantu languages of southern Africa, and in Dahalo, a Cushitic language of eastern Kenya. But there are a number of features that are widespread in Africa but less common, or uncommon, outside

Africa. Phonetic features are in particular the presence of labiovelar stops (*kp* and *gb*), of implosive stops (*b*, *d*, *g*), of prenasalized stops (*mb*, *nd*, *ng*), or of vowel harmony based on the tongue root (+/- advanced tongue root position). Among morphological characteristics one finds the widespread occurrence of a set of verbal derivative extensions expressing grammatical functions such as passive, causative, applicative/benefactive, and reciprocal. One might also mention the presence of noun-class systems, based on the distinction human vs. non-human or animate vs. inanimate (rather than masculine vs. feminine) and distinguishing a larger number of classes, but the occurrence of such systems appears to be genetically determined: they are common in the Niger-Congo family but essentially absent elsewhere.¹ A negative areal feature can be seen in the nearly complete absence of ergative languages in Africa. Semantic features characterizing the African continent are certain polysemies of nouns and verbs. For example, the noun for 'wild animal' also denotes 'meat', and the verb for 'eat' has 'conquer' and 'have sexual intercourse with' as additional meanings in many African languages; see Greenberg (1959: 23), Gilman (1986) for details.²

But also within Africa, some convergence areas have been identified (see e.g. Greenberg 1959: 24–5). The most frequently mentioned example is north-eastern Africa. Within roughly the last two millennia, the highlands of Ethiopia appear to have favoured cultural and linguistic exchange on a massive scale, with the effect that the languages of this region now share a number of linguistic properties (Ferguson 1976).

The Kalahari basin of southern Africa appears to form another convergence area; it provides an instance of a refuge area where people have been living over centuries and probably millennia without much interference from outside. It is the homeland of the Khoisan-speaking Bushmen or San peoples. Lewis-Williams (1984) suggests that there has been ideological continuity in San culture for at least two millennia and possibly for as long as 26,000 years, where ideological continuity implies some degree of continuity in social relations.³ As Güldemann (1997) argues, the Kalahari basin convergence area is not confined to languages conventionally classified as belonging to the Khoisan phylum; rather, it also includes a Bantu language, Tswana (Güldemann 1997).

One linguistic domain that appears to be particularly prone to contact-induced change is word order, more precisely the arrangement of main clause

¹ An example of such a noun-class system outside Niger-Kordofanian can be found in !Xun (Ju|'hoasi), a North Khoisan language of Namibia, Botswana, and southern Angola. What distinguishes this system from canonical Niger-Congo systems is in particular that in !Xun, nouns are not overtly marked for gender (cf. Heine 1981: 210).

² Concerning more areal properties of African languages, see Gilman (1972, 1986), Creissels (2000).

³ Lewis-Williams (1984) suggests in particular that the San trance dance strengthens kinship relationships, which in turn structure the aggregation and dispersal necessary to distribute people over available resources. Whatever the significance of such suggestions may be, we have to be aware that they are not based on any historically relevant data and therefore have to be approached with care.

constituents. Based on a survey of the order of meaningful elements in African languages, Heine (1976) concludes that there are a number of linguistically defined areas cutting across boundaries of language families. One such area consists of a large part of West Africa where Mande, Gur (Voltaic), and western Kwa languages are spoken. In addition to these languages, which are traditionally classified as Niger-Congo, this area also includes Songhai, a language usually classified as belonging to the Nilo-Saharan phylum. What characterizes this area most of all is the presence of a possessor-possessee word-order syntax which is not confined to the noun phrase but has also affected the structure of the clause (see Claudi 1993). Another area, called the Rift Valley Convergence Area, is defined by the presence of verb-initial (VSO) syntax, very rarely encountered elsewhere in Africa.⁴ The languages of this East African area belong to Greenberg's (1963) Nilo-Saharan (Surma, Kuliak, Eastern Nilotic, and Southern Nilotic) and Khoisan families (Hadza).

What these studies suggest is, first, that previous research, like that summarized in §1, has relied too heavily on discovering genetic-relationship patterns, ignoring the fact that investigating areal relationship provides a complementary—and equally rewarding—approach to reconstructing Africa's linguistic history. Second, they also suggest that what constitutes areal relationship is still largely unclear. Terms such as 'linguistic area', 'areal group', or 'convergence area' are notoriously fuzzy; as a rule, using them is tantamount to claiming that there is a set of linguistic properties exhibiting an areal distribution that cannot be reconciled with what we know about the genetic relationship of the languages concerned and that the most reasonable explanation therefore is contact-induced relationship. Third, these studies also suggest that we still know very little about the overall situation of areal relationship in Africa and, perhaps more importantly, that we still lack adequate methods and models for describing this kind of relationship. Still, whether, or to what extent, existing models capture salient characteristics of convergence areas remains unclear considering our as yet largely inadequate empirical knowledge of language contact and its implications for language classification. Fourth, and consequently, what we need most urgently is a more detailed account of what happens when languages, or more exactly, when speakers of different languages, are in contact.

The most likely consequence of such situations is lexical borrowing. There are quite a number of studies that describe how one African language has borrowed part of its vocabulary from another language. As a rule, nouns account for by far the largest part of borrowed material, followed by verbs, interjections, and conjunctions, with affixal morphology being much less likely to be affected by language contact. While this is the expected case, there are nevertheless examples

⁴ The only other VSO-languages reported so far are the Berber languages of north-western Africa, a few Chadic languages, and Krongo, a Kordofan language nowadays considered to belong to the Kadu branch of Nilo-Saharan.

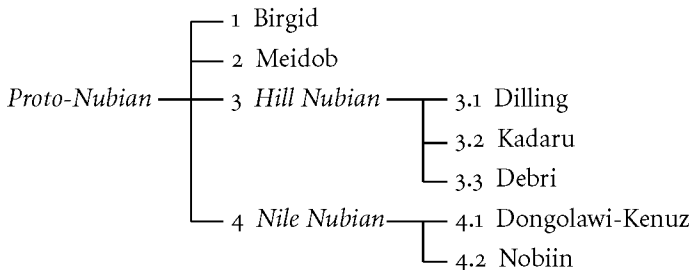
to suggest that the lexicon need not be the primary domain of contact-induced change. The following case study involving the Nile Nubian languages of Egypt and Sudan provides such an example.

2.2. NILE NUBIAN

Prior to the construction of the Aswan dam, the Nile Nubian languages were spoken by some 200,000 to 400,000 people along the Nile River between Aswan in the north and Old Dongola in the south in southern Egypt and the northern end of the Republic of Sudan (see Werner 1987: 29–30). Four Nile Nubian dialects tend to be distinguished in the relevant literature: Kenuz (Kenzi, Kunuzi), Fadijja (Fadidja, Fadičča), Mahasi (Mahas), and Dongolawi (Dongola).⁵ Following Bechhaus-Gerst (1984), on which the present account is based, Kenuz and Dongolawi are treated as one group,⁶ referred to as Dongolawi-Kenuz, and Mahasi and Fadijja are also grouped together under the term Nobiin. Both groups show a close relationship: their phonological systems are said to be identical, and their morphological and lexical inventories are very similar; a lexicostatistic count yielded 70% of cognates between the two groups.

Nile Nubian shows genetic relationship with the following language groups: the Hill Nubian languages and dialects of Kordofan, such as Debri, Kadaru, and Dilling, and the Birgid and Meidob languages spoken in Darfur, over five hundred kilometres away from Nile Nubian (see Map). On the basis of lexicostatistic counts, Nubian has been classified as described in (1).

- (1) A lexicostatistic classification of Nubian (Bechhaus-Gerst 1984: 17; groups are italicized)



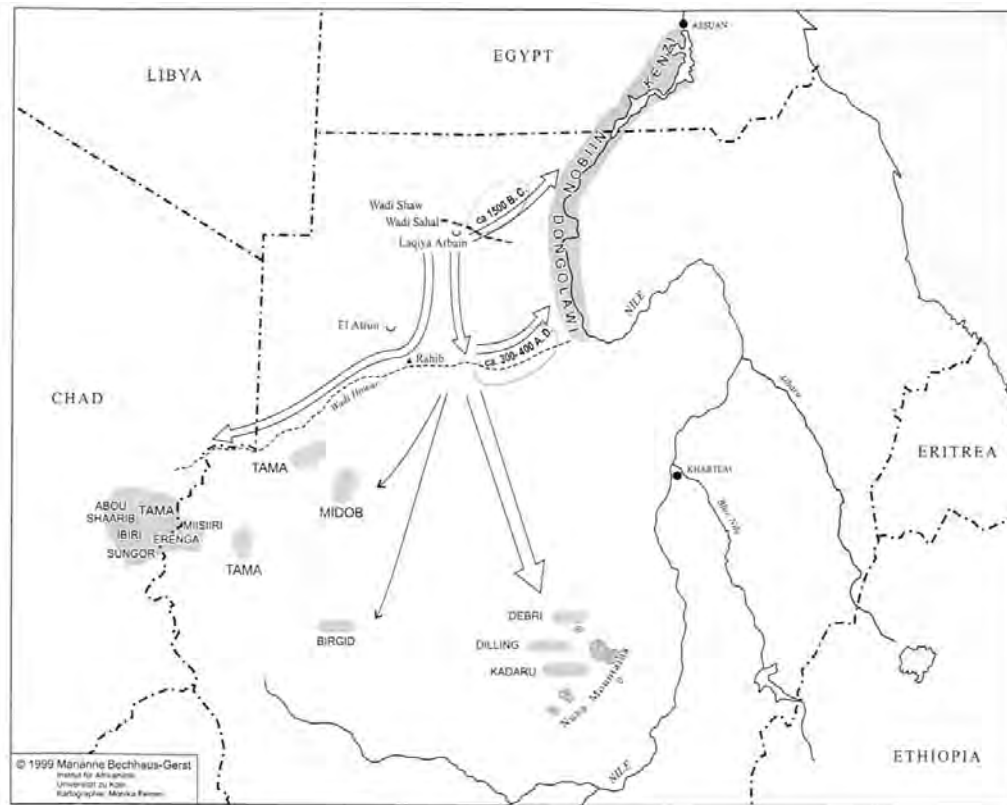
On the basis of such lexicostatistic data, Thelwall (1982) concludes that Dongolawi and Nobiin⁷ represent a most recent genetic split within Nubian.⁸ More detailed

⁵ This four-fold distinction is not shared in every detail by all authors who have written on the subject; virtually every author has come up with a different description of Nile Nubian dialects or languages.

⁶ The split of Kenuz and Dongolawi into different dialects appears to be a very recent one (cf. Werner 1987: 28).

⁷ Thelwall (1982) does not consider Kenuz in his calculations.

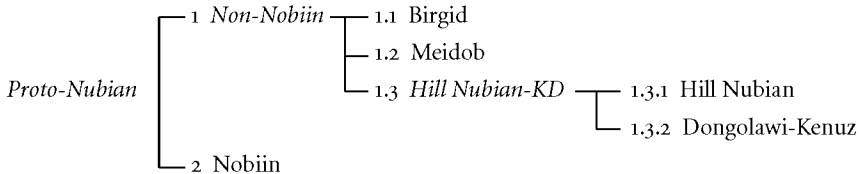
⁸ Thelwall's (1982) classification differs also in few other details from that of Bechhaus-Gerst, e.g. in his claim that the first split of Proto-Nubian led to a separation of Meidob from the rest of Nubian.



MAP. Reconstructed migrations of the Nubian and Tama peoples from their presumed homeland in the Wadi Shaw/ Laqiya region

work based on linguistic, archaeological, and Egyptological evidence suggests, however, that (1) does not reflect the actual genetic relationship pattern holding between these languages; rather (2) is a much more appropriate tree diagram:

(2) The genetic classification of Nubian (Bechhaus-Gerst 1984: 121; groups are italicized)



The lexicostatistic tree is at variance with the 'genetic' tree in two ways in particular:

- (a) It suggests that Nile Nubian is a genetic unit, while a more comprehensive analysis shows that the two subgroups of Nile Nubian (Kenuz-Dongolawi and Nobiin) belong to different primary branchings of Proto-Nubian.
- (b) It does not represent Nobiin as splitting off from Proto-Nubian before all other languages did.

In general, lexicostatistics has turned out to be a fairly reliable tool for establishing first hypotheses on genetic relationship in Africa: in most cases where the comparative method and lexicostatistics have been employed, they yielded similar results (see below). The question is: how is it possible that in the case of Nubian there is such a divergence between the two tree diagrams?

On the basis of combined linguistic, archaeological, and other evidence, Bechhaus-Gerst (1984) volunteers the following answer (see also Thelwall 1982: 32). The Nubian languages were 'originally' spoken in the dry regions of Kordofan and Darfur nearly five hundred kilometres west of the Nile. Roughly three millennia ago, speakers of what is nowadays referred to as Nobiin migrated to the Nile and settled there, adopting a riverine economy and culture.

About one thousand years later, another group of Nubians, now represented by speakers of Kenuz-Dongolawi, also migrated to the Nile via the Bayuda Steppe where they met the Nobiin. With the collapse of the Meroitic empire in the fourth century AD at the latest, Nobiin speakers became the dominant power along that part of the Nile. Their language became a written language used in church and in trade, commonly known as Old Nobiin (Bechhaus-Gerst 1996: 298, 304). The result was heavy borrowing, which was largely though not entirely unilateral: the high-prestige language Nobiin was the main donor, but there were also Kenuz-Dongolawi loans that entered Nobiin. While having stayed separated for no less than a thousand years, Kenuz-Dongolawi and Nobiin became more and more similar, to the extent that they have almost become dialects.⁹

⁹ Werner (1987: 24) insists that Kenuz-Dongolawi and Nobiin are not mutually intelligible and hence proposes to consider them as different languages.

The separation of the Nobiin-speaking and the Dongolawi-Kenuz-speaking people was so long that present-day descendants of both groups do not share a common Nubian identity; traditionally, the Nubians (Nobiin) call Dongolawi-Kenuz *ofkiriin bannid* 'language of the slaves' or *bideriin bannid* 'language of the poor'. Both groups also differ in their traditions about their origin. The Nobiin claim to be the only genuine Nubians of African origin, while the Dongolawi-Kenuz believe they are descendants of immigrants from the Arabian peninsula (Bechhaus-Gerst 1996: 298).

The case of Nile Nubian is also of interest with regard to principles of language classification. Genetic relationship is widely assumed to be based on the family-tree model, even if there are a few examples that seem to challenge such an assumption (Thomason and Kaufman 1988). Nile Nubian offers another case of a challenge. There is some evidence to suggest that Dongolawi-Kenuz is a 'hybrid' language between Old Nobiin and pre-contact Dongolawi. This evidence is of the following kind (see Bechhaus-Gerst 1996: 305 ff.):

- (a) PHONOLOGY. Dongolawi-Kenuz has borrowed almost its entire phonological system from Nobiin, even if there was also some borrowing in the reverse direction.¹⁰
- (b) MORPHOLOGY. Dongolawi-Kenuz has borrowed much of its morphology from Nobiin, in particular the following items:¹¹
 - (i) the postpositions *bokon* 'until' and *takki* 'when'
 - (ii) demonstrative pronouns
 - (iii) interrogative pronouns
 - (iv) the plural suffix *-gu* with pronouns
 - (v) the plural suffix *-ri* (in loanwords only)
 - (vi) the suffix *-ke(n)* marking habitual aspect (Kenuz only)
 - (vii) the plural object suffixes with the verb *den-/tir-* 'give'
 - (viii) verbal suffixes for the resultative/perfective
 - (ix) verbal suffixes for the stative
 - (x) verbal suffixes for the durative
 - (xi) verbal prefixes for the durative/habitual
 - (xii) verbal prefixes for the intentional/ingressive (future).

Conversely, the influence of Dongolawi-Kenuz on Nobiin was rather restricted. Bechhaus-Gerst (1996: 306) finds the following elements of

¹⁰ Nobiin has borrowed the phoneme /b/ through late loanwords (Bechhaus-Gerst 1996: 306). Concerning the techniques used to determine directions of borrowing, see Bechhaus-Gerst (1996).

¹¹ Bechhaus-Gerst (1996) gives several types of arguments why these morphological forms are borrowed rather than retained. That morphemes might be adopted into one language from another language, provided that the two languages are significantly similar typologically, has already been observed in other language-contact situations in other geographical areas (cf. the borrowing of Bulgarian inflectional verb endings into Meglenite Romanian; Sandfeld 1938: 59). It remains unclear, however, whether, or to what extent, the borrowed morphology has replaced previously existing grammatical categories.

Dongolawi or Kenuz origin in Nobiin: the formative suffix for ordinal numerals, a second set of personal pronouns based on plural pronouns, the suffix *-ndi* (> *-ni*) with possessive pronouns, and the nominal plural suffixes *-ii* and *-nci* in loanwords.

- (c) LEXICON. Even though loanwords and loan translations from Nobiin into Dongolawi-Kenuz can be found, the latter appears to have retained much of its own vocabulary.¹² There is no evidence of massive lexical borrowing from Nobiin in Dongolawi-Kenuz.¹³

Modern Dongolawi-Kenuz is not just a later historical state of one language, rather it is the 'daughter' of both pre-contact Dongolawi-Kenuz and Nobiin. Whether the process underlying this situation requires specific circumstances to happen, like the presence of a close, or even a common, genetic link between the languages concerned (Jeffrey Heath, p.c.), requires further investigation. What is obvious from the data available is that we are dealing with an instance of convergence, not towards a common prototype (see Dixon 1997), but rather of one language towards another, i.e. Dongolawi-Kenuz to Nobiin. To conclude, we are faced

- (a) with the emergence of a new language, modern Dongolawi-Kenuz, whose genetic position can no longer be described unambiguously in terms of a tree-diagram model, and
- (b) with the continuation of another language, Nobiin, which is slightly 'changed' due to borrowing.

3. Grammaticalizing metatypy

Cases like Nile Nubian do not seem to be very common in Africa. What this case suggests however is, first, that until now such situations of intensive language contact have not received the kind of attention they deserve. Second, that we still know very little about the various kinds of sociolinguistic settings that may be present in situations of intensive language contact, and how each setting affects language structure. And third, that previous research on contact-induced language change has focused primarily on lexical, phonological, and morphological interference. What we lack most of all is more information about how language contact affects meaning and the arrangement of meaningful elements in

¹² That areal influence can strongly affect grammar but spare the lexicon is noteworthy but not entirely uncommon. Aikhenvald (this volume) observes that the Vaupés linguistic area of north-west Amazonia is characterized by the presence of a number of common grammatical features while lexical borrowing is absent.

¹³ The discovery of lexical items turns out to be a difficult exercise since Dongolawi-Kenuz has been assimilated phonologically to Nobiin, which means that there are hardly any phonological clues which would be of help when deciding whether a given lexical correspondence is due to borrowing or to common inheritance (cf. Bechhaus-Gerst 1996: 186).

discourse. Apart from a few, largely impressionistic observations, not much is known about how meaning and meaningful structures behave in language-contact situations.

Examples of meaning-transfer from language to language, irrespective of the form this meaning may take in a given language, are more common in Africa than is commonly believed; they are instances of what Ross (1996, 1997, Chapter 6) calls *metatypy*. Note that metatypy is not confined to lexical semantics. It may involve entire constructions or predications and, since it has to do with the combination of meaningful elements, it also has a syntactic component. Underlying metatypy there appears to be a strategy whereby speakers aim to adapt their ways of saying things to those of the target language by reorganizing their expressions of meaning (semantics) and the way meaningful elements are arranged (syntax). Adaptation affects in particular the following domains of language structure:

- (a) the range of meanings expressed by a given word or phrase,
- (b) the patterns of syntactic encoding, and
- (c) the nature of idiomatic expression.

Metatypy has been treated traditionally as calquing. The difference between calquing and metatypy is actually one of degree rather than kind. While the term *calquing* tends to be used for the 'translating' of lexical items, referring to what happens to individual words or groups of words, *metatypy* captures 'loan translation' on a larger scale, relating to more general patterns of linguistic expression; it leads essentially to a change in structural type. No attempt is made here to trace a boundary between the two (for a discussion of the differences between calquing and metatypy, see Ross, this volume). Following Ross (1997: 241) we will assume that underlying metatypy there is a strategy employed by speakers to reduce their cognitive and linguistic-processing burden by bringing their construal of reality into line with that of speakers of another language, what we may term the target language. The result is that the languages concerned become more readily inter-translatable. Note that in metatypy the form, i.e. the phonological substance used to encode meaning, remains unaffected.

Metatypy can be held responsible for a variety of new structures of language use, for example new conventionalized expressions, phrases, idioms, proverbs, and patterns of syntactic encoding. In addition, there appears to be one type of metatypy that leads to the emergence of new grammatical categories: we will refer to this type as grammaticalizing metatypy.

Grammaticalization has been described as a process leading from lexical to grammatical and from grammatical to even more grammatical forms. While such a description has turned out to be useful, it tends to ignore that, more often than not, the relevant process is not confined to individual units such as morphemes or words; rather it involves the reinterpretation of more complex semantic structures as structures serving the expression of grammatical functions. The following is a sketchy treatment of grammaticalizing metatypy, meant to illustrate the

potential this notion may have for understanding certain patterns of contact-induced language change.

There is a kind of semantic structure that tends to be used cross-linguistically for expressing grammatical functions. The term used for this structure is *event schema* (see Heine 1993, 1997a, 1997b for details). Event schemas present stereotyped situations with which we are constantly confronted; they are propositional in structure and take the form of simple predications describing what one does (Action), where one is (Location), who one is accompanied by (Companion), what exists (Existence), etc. There is only a limited pool of such schemas recruited for the expression of grammatical functions. The way event schemas can affect grammatical encoding may be illustrated with two examples. In §3.1 we will look at comparative constructions, while §3.2 will be devoted to reflexive markers.

3.1. COMPARATIVES

Our concern here is more narrowly with the way the standard of comparison is encoded in comparative constructions of inequality (also called superior comparatives). These are constructions having the form *X is Y-er than Z*, where *X* is the comparee (or item compared), *is Y-er* is the predicate, and *than Z* is the standard of comparison. There is only a handful of event schemas that tend to be recruited time and again in the languages of the world to express and grammaticalize this notion. Perhaps the most widespread schema is one in which the standard of comparison is presented by means of an ablative or locative source morphology, as in the following example:

(3) Yaaku (Eastern Cushitic; Afroasiatic)

kédén ké cén òù àì
 tree COP big from house
 'The tree is bigger than the house.'

Quite a different way of expressing the notion of a comparison of inequality is to establish a polar contrast between the comparee (*X*) and the standard of comparison (*Z*). Polarity may involve either antonymy (= presence of property *p* vs. presence of property *q*), as illustrated in (4a), or a negative-positive contrast (= presence vs. absence of property *p*), as in (4b).

(4) (a) Sika (Moluccan; Austronesian; Stassen 1985: 44)

dzarang tica gahar, dzarang rei kesik
 horse that big horse this small
 'That horse is bigger than this horse.'

(b) Hixkaryana (Carib; Stassen 1985: 44)

kaw-ohra naha Waraka, kaw naha Kaywerye
 tall-not he.is Waraka tall he.is Kaywere
 'Kaywerye is taller than Waraka.'

TABLE 1. The main event schemas used for encoding comparative constructions (see Heine 1997b: 112)

Form of schema	Label of schema
X is Y surpasses Z	Action
X is Y at Z	Location
X is Y from Z	Source
X is Y to X	Goal
X is Y, Z is not Y	Polarity ¹⁴

TABLE 2. Event schemas serving as sources for the grammaticalization of comparatives of inequality (Sample: 109 languages of world-wide distribution; Stassen 1985, Heine 1997b: 128)¹⁵

Source Schema	Europe	Asia	Africa	The Americas	Indian/ Pacific Ocean	Total
Action	0	4	13	1	2	20
Location	0	4	3	4	1	12
Source	0	18	4	9	1	32
Goal	1	0	3	3	3	7
Polarity	0	0	0	10	10	20
TOTAL	14	26	23	28	18	109

These are two schemas commonly recruited to express the notion of a comparative of inequality; we will refer to them as the Source and the Polarity Schemas, respectively. But these are not the only schemas; the whole range of schemas most commonly employed cross-linguistically is summarized in Table 1.

In principle, speakers of a given language may select any of these schemas to develop a new comparative construction, and in many languages, more than one schema has been grammaticalized. It would seem, however, that there is one important factor that influences the choice of schemas, and this factor has to do with geography: neighbouring peoples are more likely to draw on the same schema for a specific purpose than peoples living at some distance from one another. The result is that there are geographically defined regions where a preference for a specific kind of grammaticalizing metatypy can be observed. Table 2

¹⁴ There is reason to assume that the Polarity Schema differs from the other schemas in that it tends to be only weakly grammaticalized, if at all (Geoffrey Haig, p.c.)—to the extent that in some languages where comparative notions are expressed by means of polarity it remains unclear whether there is any justification for assuming that such expressions really have the status of a grammatical category. More research is needed on this issue.

¹⁵ This table differs slightly from that presented in Heine (1997b: 128) in that Classical Arabic and Biblical Hebrew are treated here as ‘Asian languages’, which means that, instead of a category ‘Africa and Middle East’ we now have a category ‘Africa’, which includes only languages that have been spoken natively in Africa for at least one millennium.

summarizes the results of a cross-linguistic survey of these constructions. Note that the sample of 109 languages has been established on what Stassen (1985) argues is a genetically and areally balanced selection of the world's languages (see Stassen 1985 for details).

What the figures in Table 2 suggest is that the macro-areas distinguished are each characteristic of a particular choice of event schemas. In the European languages of the sample, the vast majority are characterized by what Stassen (1985) calls 'particle comparatives', that is, by constructions whose etymological source is opaque: 92% of all European languages of the sample, including English, have grammaticalized their major comparative construction to the extent that it is no longer possible to determine unambiguously the schema from which it is historically derived.¹⁶ In Asian languages there is a clear preference for the Source Schema: more than two thirds (69%) of all sample languages spoken on the Asian continent make use of the Source Schema. What unites the Americas and the region of the Indian and Pacific Oceans again is the widespread grammaticalization of the Polarity Schema to a comparative construction; no sample language outside this general area has been found to have drawn on this schema.

Africa as a macro-area also exhibits a clear preference pattern: more than half of all African sample languages (57%) have grammaticalized the Action Schema to comparative constructions. But perhaps more significantly, almost two thirds (65%) of all languages of our world-wide sample having made use of this schema are spoken in Africa. There is some variation in the exact shape this schema may take, the main ones being either [*X is Y surpasses Z*], as in (5a), or [*X surpasses Z (at) Y-ness*], illustrated in (5b). What is common to all of them is that the standard of comparison is presented by means of a verb meaning 'surpass, defeat, exceed', and the like, that is, the comparee (X) surpasses the standard (Z) with reference to the quality in question (= the predicate Y).

- (5) (a) Swahili (Bantu; Niger-Congo)
 Nyumba yako ni kubwa kushinda yangu
 house your be big to.defeat mine
 'Your house is bigger than mine.'
- (b) Hausa (Chadic; Afroasiatic; Wolff 1993: 221)
 naa fi Muusaa wàayoo
 I surpass Moses cleverness
 'I am cleverer than Moses.'

This areal distribution sets the African continent apart from the rest of the world: one can predict with a certain degree of probability that if one finds a language that expresses the notion of a comparative of inequality by means of the Action Schema then that is likely to be an African language.¹⁷ Note that the Action

¹⁶ But see Heine (1997b: 117–18) for possible etymologies.

¹⁷ Geoffrey Haig (p.c.) points out, however, that the Action Schema is also very common in Papua New Guinea.

Schema is not confined to some particular language phylum or phyla in Africa; rather, its distribution cuts across genetic and regional boundaries.¹⁸

But there is also an areal patterning within Africa. While one can expect the Action Schema to have given rise to comparative constructions in any part of the continent, there is also a regional patterning. In Table 2 we find four instances of the Source Schema in Africa, and three of them involve Ethiopian languages: Amharic, an Ethio-Semitic language, and Beja and Bilin, both Cushitic languages.¹⁹ A more detailed analysis suggests in fact that it is the Source Schema, rather than the Action Schema, which is the most common source for grammatical categories of a comparative of inequality in these language groups. This distribution might suggest that we are dealing with a genetic rather than an areal feature since Ethio-Semitic and Cushitic languages are both branches of Afroasiatic. But the Source Schema is also found in non-Afroasiatic languages of Ethiopia, e.g. in Kunama (where the ablative postposition or suffix *-kin* encodes the standard of comparison):

- (6) Kunama (Nilo-Saharan; Böhm 1984: 94)
 Marda- kin Kunama maida
 Nera²⁰- from Kunama be.noble
 'A Kunama is more noble than a Nera'

These observations suggest that grammaticalizing metatypy provides yet another feature defining the Ethiopian highland region as a linguistic area: instead of the otherwise prevailing pattern of forming comparatives of inequality in Africa by means of the Action Schema, it is the Source Schema which is favoured in this region.²¹

3.2. REFLEXIVES

That the use of the Action Schema for expressing a comparative of inequality belongs to those features that characterize Africa as a linguistic area has already been mentioned by Greenberg (1959). Greenberg's examples also include that of reflexive marking: he observes that 'he himself' translates in African languages as 'he with his head' (1959: 23). This example relates to expressions for what tend to be referred to as emphatic reflexives, but one can generalize by saying that, in fact,

¹⁸ As we noted above, many languages have grammaticalized more than one schema. With reference to example (5a), for example, one should mention that Swahili uses not only the Action Schema but also the Location Schema (having grammaticalized the form *kuliko* 'where there is' to a standard marker: see Heine 2000).

¹⁹ The only African language outside the Ethiopian area is Nama, a Central Khoisan language (Stassen 1985: 40).

²⁰ Böhm uses the term 'Barea' instead of Nera; the former term is no longer considered by the speakers of this language to be appropriate, hence we have replaced it by 'Nera'.

²¹ This observation might suggest that the Ethiopian area is an extension of the Asian 'macro-area' where the Source Schema is the clearly predominant one (see Table 2).

TABLE 3. Nominal sources for reflexive/reciprocal markers in African languages (Sample: 62 languages; for 25 of these, no nominal source could be found; see Heine 2000).

Nominal meaning	Number of occurrences	Percentages
body	20	51.2
head	9	23.1
owner	3	7.7
comrade	2	5.1
life	2	5.1
relative	1	2.6
soul	1	2.6
person	1	2.6
TOTAL	39	100.0

a number of African languages have grammaticalized reflexive pronouns (including markers for emphatic reflexives) which are etymologically derived from terms for the body-part 'head', as illustrated in the following example:

- (7) Hausa (Chadic; Afro-Asiatic; Kraft and Kirk-Greene 1973: 231)

Sun kashè kân-sù

they kill head-their

'They have killed themselves.' (i.e. 'they have committed suicide')

In examples such as (7) we are dealing with a propositional schema where an object noun phrase which is co-referential with the subject noun phrase is grammaticalized to a reflexive marker. Such a schema can in fact be said to constitute an areal feature of Africa. However, Greenberg's example is in one respect not entirely satisfactory: in the majority of cases it is not a noun for 'head' that is employed in African languages as the head of the object noun phrase but rather the noun for 'body', as in the following example:

- (8) Yoruba (Kwa, Niger-Congo; Awolaye 1986: 4)

Nwosu rí ara rē

Nwosu saw body his

'Nwosu saw himself.'

That this example illustrates the clearly predominant type found in Africa is suggested by the figures in Table 3, based on a sample of sixty-two African languages from all major genetic groupings and regions of the continent.²²

Table 3 suggests that by far the most common nominal meaning to encode reflexive (and reciprocal²³) concepts in African languages is to use the noun 'body'

²² The reader is referred to Heine (2000) for more details.

²³ Reciprocal categories are not considered here; they differ in some ways from reflexive ones; see Heine (2000).

as the object noun in a propositional schema of the kind *X sees/hits/kills X's body*, and this schema has been grammaticalized to a reflexive category.²⁴ More than half of the sample languages for which a nominal source could be established use this schema, and we can in fact say that it may well constitute an areal property of the African continent: given any unknown African language, one may predict with more than chance probability that that language will have a grammaticalized form of the above schema, let us call it the 'body'-schema, to express reflexivity.

Compared to this, the number of African languages using a schema of the form *X sees/hits/kills X's head*, that is, which have grammaticalized the noun 'head' to a reflexive pronoun, is fairly small: less than one fourth of our sample languages appear to have done so. Furthermore, these languages are spoken in one specific area, the sub-Saharan belt of West Africa, roughly between Senegal and Cameroon, and they belong to two different language families: Niger-Congo (West Atlantic) and Afroasiatic (Chadic).²⁵ There is only one exception in our sample, which is Kemantney (Kimant), a Central Cushitic language of Ethiopia, which also has the 'head'-schema.

This suggests that, in addition to the pan-African distribution of the 'body'-schema, there is also an areal pattern based on the 'head'-schema, also cutting across genetic boundaries. Note that two languages of the sub-Saharan belt area, Margi and Mina (both belonging to the Chadic branch of Afroasiatic), have two reflexive categories, derived respectively from the 'body'-schema and the 'head'-schema.

3.3. SUMMARY

These observations suggest first, that among those linguistic features that are indicative of an areal rather than a genetically defined distribution there are patterns involving neither phonetic, nor phonological nor morphosyntactic forms; rather they involve meaning and the way meaning is encoded. Our concern was with meaning relating not to lexical semantics but to grammatical categories. Thus we considered not merely event schemas, that is, meaningful propositions, but rather the way these schemas are employed for the expression of grammatical functions. In other words, we were dealing with metatypy of a specific kind, a process involving a two-stage strategy, whereby speakers not only adopt²⁶ a certain semantic configuration or schema, but also the idea that this configuration be used for encoding grammatical meaning.

Second, both examples presented involve a pan-African areal patterning on the one hand, and a more restricted regional one on the other. Both patternings cut

²⁴ Lionel Bender (p.c.) observes that 'neck' and 'foot' are additional sources for reflexive markers in some African languages.

²⁵ These languages are Fulani, Diola, Hausa, Margi, Mina, Pero, Kwami, and Lele. The first two are West Atlantic languages, while the rest are Chadic. Note that the languages spoken further south, along the West African coast, are excluded from this belt.

²⁶ Whether this is done consciously or unconsciously is an issue we cannot dwell on here.

across boundaries of genetic units. The most plausible explanation, therefore, would seem to be one in terms of language contact. What we now need most urgently is, first, sociolinguistic micro-studies of speech communities in contact that would allow us to describe in more detail why and how exactly people adopt propositional schemas of the kind discussed in this section from other speech communities.

4. Conclusions

Our observations of some problems of historical linguistics in Africa may be summarized in the following way. First, we observed that quite a bit of progress has been made in the genetic classification of African languages; still, our knowledge of more remote relationship patterns is severely limited. Second, while the family-tree model, based on the one-parent assumption, has turned out to be the only one to describe genetic relationship appropriately, there are cases such as Nile Nubian which may be viewed as an additional challenge for this model. We should be aware, however, that models used in comparative linguistics, in the same way as models used elsewhere in the humanities, are based on probabilities rather than on exceptionless laws. Examples like Nile Nubian may suggest that searching for a family tree no longer makes much sense; still, looking at the overall situation of language history in Africa, such cases are statistically hardly significant. Thus, it seems advisable not to select, or to develop, models on the basis of such spectacular and unusual cases but rather on what is the expected case, that is, on what is most likely to have happened in the history of the language concerned, or of the people speaking that language.

Third, previous work has been overly concerned with a search for the origin of Africa's present linguistic diversity, and to this end new genetic classifications were proposed time and again to reduce present-day variety to earlier unity. Such work could rely on a set of readily applicable methods, and on an attractive, logically coherent model for describing linguistic relationship: the family-tree model. Compared to this work, research on contact-induced linguistic relationship is still in its infancy. What makes areal language classification particularly difficult are problems such as the following: (a) there are no reasonable findings to guide the student of areal linguistics as to how many features would be required to define an areal group, or how to determine its boundaries; from the little we know, boundaries of areal groups are notoriously fuzzy; (b) there are also no ready-made methods and models to classify languages according to contact-induced relationship.

Fourth, genetic linguistics rests primarily on the comparison of form–meaning units, that is, on observations on correspondences between morphemes and words of different languages. Similar approaches have been used to reconstruct areal diffusion processes, perhaps most successfully for the reconstruction of lexical borrowing. One goal of the present chapter is to suggest that language contact manifests

itself in the same way in semantic transfers, more specifically in grammaticalizing metatypy and, most likely, also in other kinds of metatypy. It may happen that people borrow a comparative or a reflexive morpheme from another language but, as we argue in this chapter, they are more likely to borrow conceptual templates, like event schemas, to develop a new comparative or reflexive category.

References

- AWOLAYE, Y. 1986. 'Reflexivization in Kwa languages', pp. 1–14 of Dimmendaal (1986).
- BECHHAUS-GERST, M. 1984. 'Sprachliche und historische Rekonstruktionen im Bereich des Nubischen unter besonderer Berücksichtigung des Nilnubischen', *SUGIA (Sprache und Geschichte in Afrika*, Cologne) 6.7–134.
- 1996. *Sprachwandel durch Sprachkontakt am Beispiel des Nubischen im Niltal: Möglichkeiten und Grenzen einer diachronen Soziolinguistik*. Language Contact in Africa 3. Cologne: Köppe.
- BLENCH, R., and SPRIGGS, M. 1997. Editors of *Archaeology and Language*, vol. 1: *Theoretical and Methodological Orientations*. London: Routledge.
- BÖHM, G. 1984. *Grammatik der Kunama-Sprache*. Beiträge zur Afrikanistik 22. Vienna: Afro-Pub.
- CLAUDI, U. 1993. *Die Stellung von Verb und Objekt in Niger-Kongo-Sprachen. Ein Beitrag zur Rekonstruktion historischer Syntax*. Afrikanistische Monographien 1. Institut für Afrikanistik, Universität zu Köln.
- CREISELS, D. 2000. 'Typology', in Heine and Nurse, forthcoming.
- DIMMENDAAL, G. J. 1986. Editor of *Current Approaches to African Linguistics*, vol. 3. Dordrecht, Cinnaminson: Foris Publications.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DURIE, M., and ROSS, M. D. 1996. Editors of *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York: Oxford University Press.
- EHRET, C., and POSNANSKY, M. 1982. Editors of *The Archeological and Linguistic Reconstruction of African History*. Berkeley and Los Angeles: University of California Press.
- FERGUSON, C. A. 1976. 'The Ethiopian language area', pp. 63–76 of *Language in Ethiopia*, edited by M. L. Bender, J. D. Bowen, R. L. Cooper, and C.A. Ferguson, London: Oxford University Press.
- GILMAN, C. 1972. 'The Comparative Structure in French, English, and Cameroonian: Pidgin English: An Exercise in Linguistic Comparison'. Dissertation, Northwestern University, Evanston, IL.
- 1986. 'African areal characteristics: sprachbund, not substrate?', *Journal of Pidgin and Creole Languages* 1.33–50.
- GREENBERG, J. H. 1949. 'Studies in African linguistic classification: I. The Niger Congo family', *Southwestern Journal of Anthropology* 5.79–100.
- 1955. *Studies in African Linguistic Classification*. New Haven: Compass Publishing Company.
- 1959. 'Africa as a linguistic area', pp. 15–27 of *Continuity and Change in African Cultures*, edited by W. R. Bascom, and M. J. Herskovits. Chicago: The University of Chicago Press.

- 1963. *The Languages of Africa*. The Hague: Mouton.
- and BERRY, J. 1971. Editors of *Linguistics in Sub-Saharan Africa*. Current Trends in Linguistics 7. The Hague, Paris: Mouton.
- GÜLDEMANN, T. 1997. 'The Kalahari Basin as an object of areal typology: a first approach', *Khoisan Forum* (Cologne), Working Paper 3.
- HEINE, B. 1976. *A Typology of African Languages Based on the Order of Meaningful Elements*. Kölner Beiträge zur Afrikanistik 3. Berlin: Dietrich Reimer.
- 1981. 'African noun class systems', pp. 189–216 of Seiler and Lehmann (1981).
- 1993. *Auxiliaries. Cognitive Forces and Grammaticalization*. New York: Oxford University Press.
- 1997a. *Possession: Cognitive Sources, Forces, and Grammaticalization*. Cambridge: Cambridge University Press.
- 1997b. *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
- 2000. 'Polysemy involving reflexive and reciprocal markers in African languages', pp. 1–29 of *Reciprocals: Form and Function*, edited by Z. Frajzyngier and T. S. Curl. Amsterdam: John Benjamins.
- and NURSE, D. 2000. Editors of *African Languages: An Introduction*. Cambridge: Cambridge University Press.
- KRAFT, C. H., and KIRK-GREENE, A. H. M. 1973. *Hausa*. Teach Yourself Books. London: St Paul's House.
- LEWIS-WILLIAMS, J. D. 1984. 'Ideological continuities in prehistoric Southern Africa: the evidence of rock art', pp. 225–52 of Schrire (1984).
- NICHOLS, J. 1992. *Linguistic Diversity in Time and Space*. Chicago: University of Chicago Press.
- ROSS, M. D. 1996. 'Contact-induced change and the comparative method: cases from Papua New Guinea', pp. 180–217 of Durie and Ross (1996).
- 1997. 'Social networks and kinds of speech-community event', pp. 209–61 of Blench and Spriggs (1997).
- SANDELFELD, K. 1938. 'Problèmes d'interférences linguistiques', pp. 59–61 of *Actes du Quatrième Congrès Internationale de Linguistes*. Copenhagen: Einar Munksgaard.
- SCHRIRE, C. 1984. Editor of *Past and Present in Hunter Gatherer Studies*. New York: Academic Press.
- SEILER, H., and LEHMANN, C. 1981. Editors of *Apprehension: das sprachliche Erfassen von Gegenständen*, part I: *Bereich und Ordnung der Phänomene*. Tübingen: Gunter Narr.
- STASSEN, L. 1985. *Comparison and Universal Grammar: An Essay in Universal Grammar*. Oxford: Blackwell.
- THELWALL, R. 1982. 'Linguistic aspects of greater Nubian history', pp. 39–56 of Ehret and Posnansky (1982).
- THOMASON, S. G., and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- WERNER, R. 1987. *Grammatik des Nobiin (Nilnubisch): Phonologie, Tonologie und Morphologie*. Nilo-Saharan 1. Hamburg: Rüdiger Köppe.
- WESTERMANN, D. 1927. *Die westlichen Sudansprachen und ihre Beziehungen zum Bantu*. Beiheft zu den Mitteilungen des Seminars für Orientalische Sprachen, Jahrgang 29. Berlin: Walter de Gruyter.
- WOLFF, E. 1993. *Referenzgrammatik des Hausa*. Hamburger Beiträge zur Afrikanistik 2. Münster, Hamburg: LIT.

What Language Features Can Be ‘Borrowed’?

Timothy Jowan Curnow

1. Introduction

One of the issues which always arises in discussions of language contact is the question of which features can be transferred (or ‘borrowed’) from one language to another. The one definite conclusion which almost every examination of language contact has inevitably come to is that expressed by Thomason and Kaufman (1988: 14): ‘as far as the strictly linguistic possibilities go, any linguistic feature can be transferred from any language to any other language’. Indeed, it is obvious without the need for examination of any data that, if no constraints are placed on the languages in question, any linguistic change which can occur must be transferable from one language to another. A community which speaks ‘a language’ does not wake up one morning having changed a feature of that language. Instead, the change spreads from an original point of innovation throughout the language community—that is, a linguistic feature is transferred from one language, that of speaker A, to ‘another’ language, that of speaker B (cf. Milroy 1997, especially 315–17). While the ‘languages’ of different speakers within a single language community are clearly very closely related, there is no obvious boundary beyond which ‘borrowing’ should not be considered to take place: is it borrowing if it occurs between two dialects of a language, is it borrowing if it occurs between two closely related languages, is it borrowing if it occurs between two unrelated languages, and so on.

The only recent dissent from the position that anything can be transferred appears to be Myers-Scotton (1998: 291), who claims that ‘*not* anything can happen in language contact. Rather, one can make principled, rather specific, predictions about expected effects.’ However, while Myers-Scotton’s arguments predict that there are only particular *paths of development* which can occur in

I would like to thank all the contributors to this volume, and especially Bob Dixon and Sasha Aikhenvald, who read an earlier draft of this chapter. Clearly, being something of a summary chapter, I owe an enormous debt to all contributors for their data and discussion; the points discussed in this chapter were almost all mentioned at various times by different people in their chapters or at the initial workshop for this volume.

language-contact situations, just about any *resulting situation* is possible by combinations of convergence and code-switching, the switching of the 'matrix language', and the possibility of stopping the transference at any point. This is similar in some respects to Ross's distinction (§2.7 of Chapter 6) between four types of contact-induced change, with different outcomes for each type and, at least for some types, a hierarchy of changes within that type.

Given that, in general, any feature of a language can be transferred to another language, the focus is then often shifted to the question of whether there is any sort of universal order in which elements are transferred, whether there are any universal constraints on which elements are transferred, and what sorts of features can interfere with these universals of borrowing.

This chapter examines the different sorts of 'borrowing' which the various contributors to this volume have exemplified in their chapters, and positions these in terms of the wider literature. After a brief look at what the term 'borrowing' may mean in §2, various possible types of hierarchy and constraints are considered in §3, then impediments to the development of constraints are examined in §4: social factors, availability of data, and the possibilities of multiple causation. Following this, §5 is an examination of what may be borrowed, a summary of those features which are discussed within the chapters of this volume as being transferred from one language to another. Finally §6 contains the conclusions which can be drawn from this.

2. 'Borrowing'

2.1. WHAT IS 'BORROWING'?

'Borrowing' is the term most commonly used in any discussion of language-contact phenomena, but it is used in a variety of different ways. While some authors attempt to maintain a distinction between terms such as 'borrowing' and 'contact-induced change' (see, for example, Harris and Campbell 1995: 122), many do not, apparently using the term 'borrowing' to cover all contact-induced changes, although sometimes with reservations about the use of the term (so that, for example, Ross (§2.1 of Chapter 6 and 1988: 413) suggests that the term 'borrowing' is infelicitous when applied to syntactic change). Thus 'borrowing' may sometimes include the addition, loss or retention of features under contact.

While the use of the term 'borrowing' to cover more than its prototypical meaning is not, in itself, a problem, it often leads to the obscuring of issues in the development of a 'borrowing hierarchy'. Thus, to use an early example, Lehmann (1962: 212–13) discusses borrowing, and considers that the borrowing of vocabulary is more common than that of syntax, morphology, and phonology. However, the borrowing of vocabulary includes, according to Lehmann, 'loanwords', 'calques', and 'extensions', with this last covering the situation where a word changes meaning under the influence of a similar form in another language, as

Australian Italian *fattoria* shifted meaning from 'farm' to the current 'factory' under the influence of the similar-sounding English word *factory* (Clyne 1997). While loanwords are one of the most common results of language contact, this particular sort of extension of meaning is presumably not. But the differences in their positioning on any potential scale of adoptability is simply hidden, as the use of the term 'borrowing' with regard to lexicon makes one think automatically of loanwords.

There is also the additional complication that contact may induce a language to lose a category or distinction which is not found in the language with which it is in contact. One important issue to consider is whether a hierarchy for borrowing is the same hierarchy as that for loss—if, as has often been suggested, tone or an inclusive–exclusive distinction in pronouns is easily 'borrowed', does this mean only that these features are easily transferred from a language which has them into a language which lacks them, does it also mean that they are easily lost from a language which has them when it is in contact with a language which lacks them, or does it mean precisely the opposite, that they are extremely difficult to lose from a language which has them when in contact with a language which does not? In the development of constraints on borrowing, we need to establish whether loss of features is to be included or excluded, or whether a separate 'loss hierarchy' should be developed in parallel with the 'borrowing hierarchy'. One practical problem related to this issue is the fact that, as pointed out by Dimmendaal (§3.1 of Chapter 13), we often only have access to synchronic information, and consequently may not even know whether particular features (such as vowel harmony in Niger-Congo languages) were lost from one set of languages and retained in another, or 'added' to the first set and not to the second.

Even more complex is the possibility that contact with a language with a particular feature may cause a language to retain a similar feature which it may otherwise have lost. Thus Watkins (Chapter 3) suggests that laryngeal consonants may have been retained in Anatolian Indo-European languages under contact with Semitic languages containing laryngeal consonants. Is there a 'retention hierarchy', similar to the 'borrowing hierarchy' and the 'loss hierarchy', or is retention an interaction of one of these hierarchies together with a principle of multiple causation (see §4.2)?

Similar issues to those found with the term 'borrowing' arise with other terms such as 'diffusion'. When Dixon (1997: 19) states that 'prosodic and secondary contrasts such as tone, glottalisation, nasalisation . . . typically diffuse', he presumably means that the existence of contrasts based on these features in one language can easily give rise to the existence of contrasts in a neighbouring language. But are these contrasts also lost easily when a language is in contact with a neighbouring language lacking these features? Or are the easily borrowed features retained as long as possible?

One of the issues in discussing borrowing is thus to establish precisely how broad a range of feature-transference is intended to be covered. In the remainder

of this chapter, the term is to be understood as broadly as possible, including addition, loss, and retention of features under contact.

2.2. WHAT IS 'BORROWED'?

In the various sorts of hierarchy of borrowability and constraints on borrowing which have been proposed (see §3), different units of language are considered to be easier or more difficult to transfer from one language to another. Thus, for example, it is often claimed that 'nouns' are more easily borrowed than 'verbs', or that 'free grammatical morphemes' are more easily borrowed than 'bound grammatical morphemes'. However, it is not entirely clear what these sorts of statements are intended to refer to, because for any particular instance of transfer of an item from one language to another, there are *two* morphosyntactic systems which have to be taken into account, the system in the original language and the system in the borrowing language, as well as some sort of potentially universal semantic system.

For example, consider the apparently simple claim that verbs are relatively difficult to borrow. It is often claimed that to the extent that this is true, it is understandable, since verbs tend to be highly inflected (e.g. Campbell 1993: 104, Dixon 1997: 20). Presumably, this means highly inflected in the original language—speakers have trouble deciding exactly what the root is, and this lowers the possibility of borrowing; this would accord with Heath's (1978: 105) statement that 'haziness of boundaries' impedes diffusion. In this case, what is important for lack of borrowability is that the potential loanword is a verb in the original system.

On the other hand, referring to exactly the same claim about the difficulty of borrowing verbs, Meillet (cited in Thomason and Kaufman 1988: 348) considers that French does not borrow verbs because it is difficult to incorporate foreign elements into the complex inflectional system of French. Here what is being claimed as important is that the potential loanword is a verb in the borrowing system. (In fact, many languages with a great deal of verbal inflection 'cheat'; for example the Iranian languages Kurmanji and Zazaki (see §4.2 of Chapter 8), many Mayan languages (Thomason and Kaufman 1988: 349), the Colombian/Ecuadorian language Awa Pit (Curnow 1997) and many other languages all have uninflected-word-plus-auxiliary-verb structures in the language, and these strategies are co-opted for borrowing verbs, which are borrowed in some relatively uninflected form, then inflected within the language on the auxiliary verb, thereby avoiding the incorporation of the foreign word directly into the inflectional system.)

Weinreich (1953: 36–7), in yet another analysis of the statement that verbs are hard to borrow, considers that the reason for the relative difficulty of borrowing verbs is lexical-semantic rather than grammatical. That is, languages are more likely to borrow a word which refers to a concrete object rather than a word which refers to an action. Under this analysis, what is important is that the potential loanword is a word referring to an action.

These different interpretations of the apparently simple statement that verbs are not easily borrowed have entirely different results. In fact, if the borrowing constraints are to be in any sense universal, none of the versions of the constraints in question are about verbs at all—one claims that it is difficult to borrow a lexical item which is highly inflected in the original language, one claims that it is difficult to borrow a lexical item if it will end up in a class which is highly inflected in the borrowing language, and one claims that it is relatively difficult to borrow concepts relating to actions.

Similar problems arise with grammatical rather than lexical borrowing. It is often claimed that derivational morphemes are more easily borrowed than inflectional morphemes (cf. Lass 1997: 190, Moravcsik 1978: 112, Thomason and Kaufman 1988: 74–5). To begin with, as will be seen in §5.10 below, it is possible for a language to borrow affixal forms, or else simply to borrow the idea of a particular affixal category (e.g. number) or exponents of that category (e.g. dual), but to develop the forms for their expression by language-internal means. Is a statement of ‘derivation > inflection’ intended to cover both borrowing of form and language-internal restructuring?

On the other hand, it is sometimes claimed (e.g. Harris and Campbell 1995: 135–6) that rather than a statement based on the formal distinction between derivation and inflection, the data which this generalization is intended to capture is better expressed in semantic terms—affixes with clear semantic content (most derivational affixes, but some inflections such as number) are more easily borrowed than semantically weak or redundant affixes (such as verb agreement).

Equally, derivational affixes are often ‘one-off’ (non-paradigmatic) affixes, while inflectional affixes are usually found only in tightly constrained paradigms. Perhaps the constraint should be phrased in terms of the possibility of transfer of individual affixes (or their meanings) versus the possibility of transfer of entire paradigms.

The main issue here, then, is that it is not a straightforward matter to decide how to categorize any particular instance of a borrowing, nor which categories should be used. A lack of borrowed ‘verbs’ may simply reflect the tendency of verbs to be more often inflected in languages than nouns, or it may reflect the relative conceptual difficulty of borrowing a word for an action rather than a concrete object. A greater preponderance of borrowed derivation over inflection may be relevant as such, or else it may simply reflect the greater semantic content of derivational morphemes over inflectional morphemes.

In order to establish constraints on what may be borrowed, we need to know not just that few verbs compared to nouns were borrowed from language X to language Y, but many other facts: are verbs inflected in language X and language Y compared with nouns, does language Y have a strategy for borrowing verbs, and so on. Rather than knowing that more derivational than inflectional affixes are borrowed, we need to know which affixes were and were not borrowed,

whether form and meaning or simply meaning was borrowed, whether the affix in question is part of a paradigm or not, and so on; and this information is often not available.

3. Scales of adoptability, hierarchies, and constraints

Attempts to define limits on which features can be transferred from one language to another have been made for over a hundred years. Many of these attempts have been phrased in terms of a 'hierarchy of borrowability' (Lass 1997: 189), 'borrowing hierarchy' (Wilkins 1996) or 'scale of adoptability' (Haugen 1950), usually phrased in terms of the order of transfer of particular grammatical, lexical, or semantic categories, or the likelihood of transfer of categories. Other attempts have been more generally phrased in terms of constraints, where it is considered that particular features cannot be transferred until certain conditions are satisfied; of course, hierarchies can always be explicitly restated in terms of constraints.

There are three main types of hierarchy which have been developed. The earliest hierarchies only consider lexical items of different types, so that for example Haugen (1950) ranks nouns as easiest to borrow followed by verbs; adjectives are harder still to borrow, then adverbs, prepositions, and so on. These sorts of hierarchy can either be discussed purely in terms of description (like Haugen's), or else given some form of functional explanation, whereby, for example, 'content words' are considered to be more easily borrowed than 'function words' since 'the former have a clear link to cultural content and the latter do not' (Appel and Muysken 1987: 171).

A slightly different type of hierarchy is that where units of different grammatical levels are considered to be ranked according to their ease of borrowing; for example, Ross (1988: 12) suggests that lexical items belonging to open sets are easiest to borrow, followed by lexical items belonging to closed sets, syntax, non-bound function words, bound morphemes, and finally that phonemes are most difficult to borrow. As with the lexical hierarchies, this type of hierarchy can be descriptive or seek an explanation, as Ross's notion of 'metatypy' (see §2.1 of Chapter 6) seeks to do through the idea of the reorganization of the semantic patterns of a language to fit another language's ways of saying things. These two types of hierarchy are not, of course, mutually exclusive, so that it is possible to integrate a level-internal hierarchy (nouns > verbs > adjectives) and a multiple-level hierarchy (lexical items > syntax > bound morphemes).

While the two previous types of hierarchy are contact-induced but in some senses context-free, the third type of hierarchy is exemplified by Thomason and Kaufman's (1988: 74–6) 'borrowing scale', where different features are expected to be borrowed depending on the type and strength of contact between two languages, with casual contact between languages leading only to borrowing of non-basic vocabulary; slightly more intense contact leading to the borrowing of

some function words and minor phonological, syntactic, and lexical semantic features; and leading through to very strong cultural contact, with heavy structural borrowing causing major typological disruption.

There are various issues which are relevant to all hierarchies. Hierarchies usually assume some sort of underlying categorization of linguistic features (nouns versus verbs, bound morphology versus free morphology), but in fact we have no *a priori* reason for believing that all the different changes within each section are equally easily transferable—loss, retention, and acquisition of a phonological feature such as tone from a contact-language may all be equally easy, or one may be far more difficult to achieve than the others. Even given a particular type of change (say, acquisition of a feature) within one section (say, the lexicon), there is no reason to assume that acquisition of all new lexical items will be equally easy, and all more difficult than, say, acquisition of any free grammatical morpheme. That is, it is not enough simply to examine one lexical feature (say, the transference of a form and a meaning for a traditional cultural item) and one morphological feature (say, the restructuring of the tense system), to find that cross-linguistically words for cultural items are borrowed more frequently than tense systems are restructured, and then claim that this shows that lexicon is more open to being adopted than morphology—it may be that lexical items (form plus meaning) from some other field, say the field of colour terms, is more difficult to transfer than tense systems.

While hierarchies all involve some sort of ordering, this may be of different kinds. Many hierarchies, such as Haugen's (1950) hierarchy which claims that nouns are more easily borrowed than verbs, order their elements in terms of ease of borrowing, or likelihood of borrowing, all other things being equal. As noted above, Thomason and Kaufman's (1988) hierarchy is not so much an ordering of transfer of language features, but rather a claim that different sorts of transference of features are related to different levels of language contact—the features are ordered in terms of level of contact between languages. On the other hand, Ross's definition of metatypy (see §2.1 of Chapter 6) states that semantic reorganization occurs first, followed by the restructuring of syntax, which begins at the level of sentences and clauses, then reorganizes phrase-level features, before finally restructuring word-internal features; lexical borrowing and phonological assimilation are explicitly excluded from this hierarchy. A similar ordering is found for Haig's linear alignment (see §5 of Chapter 8), although he argues for syntactic weight rather than level of organization as the factor determining the order of restructuring, on the basis of relative clauses. In these two cases the ordering does not seem to be related to ease of borrowing or intensity of contact, but rather gives a necessary order of borrowing: sentences and clauses must be restructured before phrases which must be restructured before words. Of course, each of these levels involves many different features, and it is not clear how far the restructuring at any level must go before any other level can begin to be restructured, although the presentations of metatypy and linear alignment in this volume suggest that

restructuring of sentences and clauses must be complete before phrasal restructuring can begin, and so on.

Moravcsik (1978) deals with the same sorts of issues as contact hierarchies, but introduces a different approach, using constraints on borrowing rather than a hierarchy of borrowing. The seven constraints she introduces are: non-lexical properties of a language cannot be borrowed unless lexical items have been borrowed first; no member of an unaccentable class (e.g. bound morphemes) can be borrowed unless a member of an accentable class which contains the unaccentable member (e.g. an inflected word) is borrowed first; a noun must be borrowed before any non-nominal lexemes can be borrowed; a lexical item whose meaning is verbal can never be borrowed; inflectional affixes cannot be borrowed before some derivational affix is borrowed; grammatical morphemes must be borrowed with their linear order with respect to their head; and if a class contains (some) uninflected words, at least some of the words borrowed into that class must be uninflected. (For discussion of and counter-examples to Moravcsik's hypotheses, see Campbell (1993) and Trask (1996: 314–15).)

The sorts of hierarchies and constraints presented here have often been proposed, but in almost all cases counter-examples have been found; consequently they are more usually talked about as tendencies rather than rigid universals. However, even assuming that there *are* tendencies for some things to be transferred between languages more easily than others, there is a wide variety of impediments to the development of such a hierarchy.

4. Impediments to the development of constraints on borrowing

4.1. SOCIAL, POLITICAL, AND HISTORICAL CONTEXT

Perhaps one of the most important factors which needs to be taken into account in developing and using any constraints on borrowing are language-external influences—the social context in which the language contact took place, and the attitudes of the speakers involved towards their language or languages.

The most obvious social or historical distinction is the distinction between 'borrowing' and 'interference through language shift', to use the terms of Thomason and Kaufman (1988). Borrowing is 'the incorporation of foreign features into a group's native language by speakers of that language' (p. 37), while language-shift interference or substratum interference occurs when 'a group of speakers shifting to a target language fails to learn the target language perfectly. The errors made . . . then spread to the target language as a whole' (p. 39). The importance of the distinction here is that the features which are transferred from one language to another appear to differ depending on the type of contact: when borrowing (in Thomason and Kaufman's terms) occurs from a foreign into the

native language,¹ it is likely to be in the areas of lexicon and perhaps morphosyntax, but not phonology (but see Trask's (1998) discussion of the influence of Spanish phonology on Basque, although this may have occurred under conditions of long-term bilingualism); in contrast, when substratum influences transfer features from the native into a foreign language, these are likely to be strongest in phonology and morphosyntax, but not in lexicon. Clearly, which of these situations has occurred may affect the relative ease of transfer of features from different levels of language; that is, affect the constraints on borrowing.

Most people interested in borrowing would state that they are only interested in borrowing from a foreign language into a native language, and that a borrowing hierarchy should thus be restricted to situations which involve Thomason and Kaufman's (1988) borrowing rather than substratum influence. Unfortunately for the development of borrowing-only constraints, given a language which has obviously undergone some form of contact-induced change, we often simply do not have the historical data which would assure us that borrowing, rather than substratum influence, has happened. We cannot simply rely on the fact that speakers tell us that the language they speak is their traditional language—Vakhtin (1998: 327) notes that Copper Island Aleut speakers claim that their language is one hundred per cent Aleut and has no resemblance to Russian, despite the fact that the analysis which Vakhtin has been led to is that Copper Island Aleut arose when native speakers of Russian (descendants of Aleut speakers) incorporated Aleut lexical items into their native Russian.

Particular social situations can also confuse the issue of 'foreign' versus 'native' languages. In the Vaupes multilingual region, speakers are expected to practise exogamy, marrying someone who has a different language as their traditional language (Aikhenvald 1996, Sorensen 1972 [1967]). In cases such as this, a child tends to be exposed at least to the language of his or her ethnic identity (the language of the father) as well as to another language (that of the mother) from childhood, with neither language being 'foreign'. Interestingly, the resulting changes found are that Tariana has converged structurally with Tucano languages but remained lexically distinct (because of cultural attitudes; see §4.1 of Chapter 7), and these results are those which Thomason and Kaufman (1988) would expect to find in cases of language shift (i.e. when a large population of non-Tariana speakers attempted to speak Tariana as a second language). The social situation of Tariana is thus neither strictly one which would imply borrowing, nor one which would imply substratum influence—should data from Tariana then be allowed in borrowing-only constraints?

¹ While Thomason and Kaufman (1988) use the terms 'native language' and 'foreign language', these are perhaps not the best terms to use to describe language-contact situations, at least prolonged ones, although they are better than Myers-Scotton's (1998) L1 and L2. In many language-contact situations, speakers are often balanced bilinguals from birth, with native-speaker competence in 'their' traditional language and the other language or languages spoken in the region, and, of course, may identify different languages as 'their' language in different social contexts (cf. §1.6 of Chapter 10).

More generally, of course, there are all the other social issues of languages in contact which will not be discussed here (see §4 of the Introduction for mention of similar factors). Is there official or *de facto* multilingualism, symmetrical or asymmetrical multilingualism with languages of equal or unequal status, or society-wide diglossia (see Clyne (1997) for discussion of many of these concepts)? Is the contact between the languages and between speakers relatively peaceful, or is there a high level of 'language conflict' (cf. Nelde 1997)? Is the language group a closed or an open group, loose-knit or tight-knit (see §2.7 of Chapter 6)? It is highly probable that different forms of social and political contact between languages will impede certain features of language transference, thus disturbing any purely language-internal scale of adoptability. (Of course, we often do not have the data available to evaluate the historical relations between two languages and language groups; see §4.2.)

Equally, social effects in language-contact situations may be more particularly language-based. In a variety of multilingual situations, code-switching (at the level of lexical items) is considered inappropriate. In the Vaupes region, this is a general societal norm, and consequently while Tariana is structurally similar to other languages of the region, having acquired many features of surrounding Tucano languages, it is lexically almost entirely distinct (see §4.1 of Chapter 7). This 'emblematicity' of certain features of language, and issues of identity, are not only powerful, but can change very quickly—thus Geoff Haig (p.c.) notes that Turkish frequently borrowed words from Arabic and Persian, but with the founding of the modern Turkish state and the related rise of nationalism and national identity, this borrowing ceased more or less instantly. (For more discussion of the effects of such social features on retarding or assisting contact-induced change, see Enfield's discussion of identity in §1.6 of Chapter 10, and Ross's discussion of emblematicity and norm enforcement in §2.1 of Chapter 6.)

A final social issue which will be mentioned here is that of language obsolescence or death. In many cases, the data available for studies of language contact are carried out in cases where the language of lesser prestige is 'endangered', being 'lost', 'displaced' (Brenzinger 1997) or 'dying'. While there are many issues surrounding language endangerment (see for example the essays in Dorian (1989) and Grenoble and Whaley (1998)), the issue relevant to borrowing is that 'many obsolescent languages undergo structural changes' (Craig 1997: 256), and it is not clear whether many of the structural changes which have been observed in language-contact cases are changes which would happen in stable bilingual situations (where speakers are and continue to be fully fluent speakers of both languages), or are structural changes particular to language death. Thus while Haig discusses the extreme convergence of the Ardeşen dialect of Laz with Turkish (see §4.3 of Chapter 8), in particular with regard to the restructuring of the nominal case system, the most complete convergence is found only with young urban semi-speakers of Laz. It is not clear that this more complete convergence is a 'natural' language-contact change, or is caused by language attrition;

and consequently this sort of change should perhaps not be included as data for borrowing constraints, at least initially.

Of course, the problem is that there is no such thing as context-free borrowing—every language-contact situation comes with a particular social context. As with all social phenomena, each context is slightly different, and it is thus difficult to generalize across cases, when we do not know which features are and are not relevant.

4.2. RELIABLE DATA AND MULTIPLE CAUSATION

In order to develop any constraints on borrowing, data on language contact is required, and perhaps the biggest impediment is simply the lack of fully reliable data, both on the current lexicon and grammar of specific languages of the world and those languages they have been in contact with, but also on the historical developments and changes within languages. Given the existence of languages A and B in contact, where A and B share certain features, it may appear that these features have been transferred from one language to the other. If there is another dialect of B, called V, which is not in contact with A and which does not have the shared features of A and B, it would seem fairly definite that these features have been transferred from A to B. Unfortunately, this is not necessarily true, and without extensive historical records there is no way of knowing whether it is true or not. There are several cautionary tales available in the literature to show that, when history is examined, often features which are ‘well known’ to have been transferred from one language to another turn out not to have been. Two of these hypothesized contexts of contact-induced change will be reviewed here to give an indication of the sorts of errors which can be made without careful historical work.

Harris (1991 [1987]) gives clear examples of similarities between Irish English and Irish Gaelic (and differences from standard English) which appear to be and are often considered to be contact-induced. Both Irish English and Irish Gaelic make extensive use of clefting, in a way which is not possible in standard English (*it's looking for more land a lot of them are*); and the tense-aspect systems of Irish English and Irish Gaelic show many similarities, distinct from standard English. In particular, corresponding to an English perfect construction with *have*, Irish English uses four different structures depending on the meaning: an ‘extended-now’ construction (*I know his family all me life*), a ‘hot-news’ construction (*a young man's only after getting shot out there*), a ‘resultative’ construction (*I've it pronounced wrong*), and an ‘indefinite anterior’ construction (*I never saw a gun in my life*). The similarities in clefting and perfect equivalents between Irish English and Irish Gaelic have led many to the obvious conclusion that these parallels are found because Irish English adopted these features from Irish Gaelic. Harris shows that, while it is likely that the use of clefting has indeed been transferred into Irish English from Irish Gaelic, the tense-aspect system of Irish English has not been remodelled on the basis of Irish Gaelic (with the exception of one structure), but

is in fact largely a retention of the Early Modern English tense-aspect system, which has been modified with the increase in use of the perfect construction in the development of standard Modern English.

The second set of examples come from Lass (1997: 197–207), and relate to South Africa, with English influence on Afrikaans, and Afrikaans influence on English. Standard Afrikaans has the standard West Germanic verb-final subordinate clauses in most cases: 'I have said, that I sick was'. In modern spoken Afrikaans, however, there is an alternative structure, with the deletion of the complementizer and second-position verb: 'I have said, I was sick'. This appears, quite obviously, to be related to the heavy contact of Afrikaans with English. In fact, however, exactly the same feature is found widely in spoken Dutch, Frisian, and German, which have not had the same sort of contact with English, and thus would appear to be an internal development or a historical retention.

Afrikaans is likewise often believed to have had an influence on varieties of South African English, in the pronunciation of /r/. At least some varieties of English in South Africa have a non-approximant /r/, realized usually as a tap, and this has been related to contact with the non-approximant /r/ of Afrikaans, with second-language speakers of English using their own /r/ rather than the English /r/. However, the development of an approximant /r/ is relatively recent in English, and at the time of the colonization of South Africa many varieties of English had a tap /r/, similar to modern Afrikaans—and these varieties of English include most of those which were spoken in areas from which there was heavy migration to South Africa.

While these examples may seem obvious, and simple cases of lack of careful thought on the part of those who have suggested contact origins for these features, these suggestions have been made, and without an examination of the historical positions of English and Afrikaans seem intuitively reasonable. The *only* reason that they have been rejected is our knowledge about the history and dialect variations of the languages in question. Without that historical record, they would certainly be accepted—and in many cases where contact-induced change is suggested, we simply do not have the historical data to ensure that there are not internal reasons for the change, as noted by Dixon (§1 of Chapter 4). In other cases, as can be seen from Dench's careful and cautious discussion of languages of the Pilbara region of Australia (see Chapter 5), it is clear that there has been retention of features, internal development of new features, and diffusion of features from language to language—unfortunately, there is often no way to tell for any particular feature whether it has been retained, is an internal development in a group of languages, or has been diffused (or from which language to which). Thus in the majority of cases we are simply not aware whether proposed contact-induced changes have alternative analyses or not; and the potentially suspect data which is obtained from many of these analyses of phenomena as contact-induced is the only data which we have for building our constraints on borrowing.

Of course, a further problem in these cases is the possibility of multiple causation. Both Harris (1991 [1987]) discussing Irish English and Lass (1997) discussing Afrikaans and South African English accept that language contact may have helped reinforce the internally developed systems (cf. also the discussion of 'acquire' in some Mainland South-East Asian languages by Enfield in §3.3 of Chapter 10, and the shared grammaticalization pathways of Sinitic discussed by Chappell in §4 of Chapter 12). One of the relatively recent realizations in historical work is that a historical change is not necessarily either just internal to the language or just caused by contact—these two can interact, and multiple causation of a change is possible, with a related structure in a contact language influencing the development or expansion of a new structure in a language (Appel and Muysken 1987: 162, Harris and Campbell 1995: 407, Harris 1991 [1987]: 209, Thomason and Kaufman 1988: 57–61).

However, multiple causation affects our ability to devise universal constraints. If a particular feature is caused partially by contact-induced change, but partially by language-internal change, we simply cannot assign such a change to a place in a hierarchy of borrowability, since the likelihood of transfer of a feature depends on the particular pre-change state of the language.

The pre-change state of the language is also relevant to the concept of 'structural compatibility', the idea that languages have to be somehow similar in order to facilitate borrowing. As an absolute hypothesis—'we would expect syntactic influence only when the two languages had a good deal of syntactic similarity to begin with' (Allen 1980: 380)—the hypothesis is clearly false (see counterexamples in Harris and Campbell 1995: 123–7). However, the weaker form of the hypothesis, that borrowing is more easily achieved if there is something similar between the two languages, is obviously relevant to the establishment of a hierarchy of borrowability. Thus Haig cautiously suggests that Laz may have been more strongly affected by Turkish than Kurmanji and Zazaki because the pre-change structure of Laz was more similar to Turkish than the pre-change structure of the other two languages (see §4.4 of Chapter 8). However, if a borrowing hierarchy is designed to indicate the order in which features are universally borrowable, then including cases where features are only easily borrowed because they are similar to features already existing in the borrowing language will lead to an erroneous ranking in the universal hierarchy.

The problem of the pre-change state of the language influencing the borrowability of particular items is not relevant only to the transfer of structural features, of course. At the level of morphology, languages sometimes co-opt an already existing functional morpheme into a new role, because of its phonological similarities to a functional morpheme in a contact language. In Chapter 3, Watkins notes that Eastern Ionic Greek altered the use of a morpheme *-ske-* to become an iterative imperfective marker, presumably under the influence of Anatolian languages such as Hittite, which had a semantically marked imperfective suffix *-ske-* or similar. The same phenomenon is found in lexical borrowing, so that

Australian Italian has altered the meaning of the standard Italian *fattoria* 'small farm' to 'factory' under the influence of Australian English (Clyne 1997); and in §2.2 of Chapter 13, Dimmendaal discusses what he calls correspondence mimicry, where a word such as *máá* 'water' in Baale has lost the original final nasal (compare Didinga *maam*) on the basis of the similar Tirma-Chai word *maa* 'water'. (See also Dench's discussion of explicit correspondence mimicry by a Martuthunira speaker in §2.2 of Chapter 5.) In all of these cases, the transfer from one language to another was presumably influenced by the existence of a similar morpheme in the pre-change state of the language. These cases consequently indicate multiple causation, and cannot be included in the development of a universal, context-free borrowability hierarchy.

5. What language units are borrowed? A summary

To establish any sort of constraints on borrowing or hierarchy, or even to talk about transference or borrowing in any general terms, particular features are required as the points on the hierarchy. This section discusses the different sorts of possible items which can be transferred from one language to another in a summary fashion, using examples found in the chapters in this volume.

While for convenience the examples of contact-induced change discussed here are divided into various categories, this is simply to make the discussion more accessible. The categories are not definitive, and some examples of change straddle more than one category.

5.1. PHONETICS

While none of the chapters in this volume specifically addresses phonetic change, it is clear that language-contact can affect phonetic features of languages—for example, the change in many European languages from an alveolar to a uvular /r/ is presumably contact-induced (Trudgill 1974). A partial case of this sort, without complete phonetic change, is noted by Dench (see §2.1 of Chapter 5): the lamino-dental stop /th/ in Martuthunira has been lenited in some environments, making it phonetically more similar to nearby languages; this has occurred with no change to the phonology of the language.

5.2. PHONOLOGY

A variety of phonological changes are possible under language contact. The simplest change is presumably the addition of a phoneme; Resígaro appears to have added the glottal stop phoneme to its inventory under contact with nearby languages (see §4.2.2 of Chapter 7).

Watkins discusses a variety of phonological changes which occurred in ancient Anatolia, such as the convergence in the inventory and distribution of stops (see Chapter 3). Dimmendaal (see §3.1 of Chapter 13) discusses the areal spread of

[+ATR] vowel-harmony systems in Niger-Congo languages. In §2.2, Dimmendaal also discusses a number of phonological changes in Baale, apparently under influence from Tirma-Chai, including an interesting phonotactic patterning, whereby word-final stops are lost in Baale, paralleling the phonotactics of Tirma-Chai. Similar phonotactic convergence is described by Aikhenvald (in §4.2.2 of Chapter 7), who notes for example that Resígaro has the same syllable structure as its neighbours, different from languages genetically related to it.

One area of phonology which appears particularly susceptible to change is suprasegmental features. Tone has been introduced into languages which previously did not have it: for example, Resígaro has acquired tonal contrasts (see §4.2.2 in Chapter 7), as have Tibeto-Burman languages in contact with Chinese (§2.2 in Chapter 9), and Matisoff (§7 in Chapter 11) discusses tone as an areal feature in South-East Asia. While diffusion of tonal contrasts is only discussed as a positive feature, one language acquiring tone from another, Dimmendaal discusses the opposite for the feature of nasality on vowels—it appears that what has diffused through most Bantu languages is the loss of a nasal-oral distinction, rather than its acquisition (see §3.2 of Chapter 13).

Phonological change is often linked to the borrowing of lexical items, so that for example Swahili is considered to have borrowed phonemes together with lexical items (cf. §2.1 of Chapter 13). However, the expected link between lexical borrowing and phonological change is not always there—Tariana has borrowed scarcely any lexical or grammatical forms from Tucano (see §4.1.1 of Chapter 7), and yet Tariana phonology is very similar to Tucano phonology. In contrast, Takia has borrowed some lexical items from Waskia (§2.2 of Chapter 6), and yet the phonologies of the two languages have diverged rather than converged (§2.3 of Chapter 6).

While there are sometimes claims that phonology is the first point of convergence between languages (cf. Watkins's citation of Trubetzkoy in Chapter 3), and it is in many of the examples given in this volume, this is not necessarily the case—as noted above, Takia and Waskia have converged grammatically, and some lexical items have been borrowed, but the phonologies have diverged.

5.3. LEXICAL FORM-AND-MEANING

The most traditional of all borrowed items is the loanword, and examples will not be discussed here, as the literature abounds in examples (see also §4.2.1 in Chapter 7). One interesting case, however, involves what might be called accommodation of lexical form, with the form of a word being altered under contact with a similar form with similar meaning in another language: for example, Dimmendaal shows that Khoti has undergone relexification, altering some of its forms to make them more similar to Swahili (see §2.1 of Chapter 13).

5.4. LEXICAL FORM ONLY

One of the most interesting examples of contact-induced change which was raised by a few contributors to this volume at the initial workshop (although not in any

individual chapter) was the possibility of a language 'borrowing' a word which did not actually exist in the source language—as though a form had been borrowed with no meaning. The particular example discussed was the German word *Handy* 'mobile phone', which many Germans 'know' was borrowed from English. Unfortunately there does not seem to be an English word for it to be borrowed from. A similar example is available from French, with the existence of a noun *footing* 'jogging' (Tony Liddicoat, p.c.). While both of these words could conceivably be English words, and have the meaning in question, neither of them actually is. (For more examples of this phenomenon, see Stefanowitsch (1999).)

5.5. STRUCTURE OF THE LEXICON

Similar to the accommodation of lexical form discussed in §5.3, there is also what might be called accommodation of lexical meaning, of two types: loan homonyms and loan synonyms (to use the terminology of Haugen (1950)). There are no clear loan homonyms, where the meaning of a lexical item is changed because of its formal similarity to a word in another language, in the chapters in this volume (but see the examples of Australian Italian *fattoria* 'factory' in §4.2; and compare the extension of use of Tariana *-ri* on the basis of Tucano *-ri* discussed in §4.1.2 of Chapter 7). Examples of loan synonyms, where the meaning of a word is extended to fit the pattern of lexical extensions of a word in another language with a similar basic meaning, are also scarcely mentioned explicitly, although LaPolla notes in §3 of Chapter 9 that the Wutun word which was historically only 'widow', with a separate word for 'widower', has been expanded to cover both meanings, presumably through its intimate contact with Tibetan languages, which only use one word undifferentiated for sex. While the lexical extension of words is not explicitly discussed in many chapters, it seems clear that convergence of lexical extensions is considered implicitly to form a part of the general reorganization of semantic patterns discussed by Ross (§2.1 of Chapter 6), LaPolla (§3 of Chapter 9) and Heine and Kuteva (§3 of Chapter 14). In addition, Ross's discussion of the reorganization of semantic patterns shows that this process is not limited to individual lexical items, but extends to compound words and metaphors (so that Takia and Waskia both express 'person' as 'man-woman', and both say literally 'I am putting out my eye' for 'I am waiting'; see §2.1 of Chapter 6); and Enfield shows that the idea of lexical extension continues on from purely lexical to grammatical uses, with the lexical morpheme for 'acquire' being used for similar grammatical purposes (modal marker, aspect marker, in resultative and potential constructions) in four South-East Asian languages (see §2 of Chapter 10); see also many of the examples of 'shared grammaticalization pathways' in Sinitic in §4 of Chapter 12.

5.6. EXPRESSIVE WORD FORMS

Perhaps more closely related to lexical influence than any other level of linguistic analysis is the diffusion of expressive word forms from one language to another,

so that the same process (such as reduplication, or reduplication with phonetic alterations) is used for similar semantic and pragmatic effects in neighbouring languages. Thus Haig notes that many Anatolian languages have an expressive construction where a morpheme is reduplicated with the initial segment of the reduplication being replaced by *m*, to give an idea of ‘and so on’ (Turkish *dergi mergi* ‘magazines, journals and so on’ from *dergi* ‘magazine’; Laz *toli moli* ‘eyes and stuff, the face’ from *toli* ‘eye’; see §3.7.2 of Chapter 8).

5.7. INTERJECTIONS AND DISCOURSE MARKERS

Two other areas of the lexicon perhaps require special treatment: interjections and various sorts of discourse markers. While these are different features, they are linked, in that both interjections and discourse markers are in some senses separate from syntax in a way in which the remainder of the lexicon is not.

Few examples of borrowed interjections are given in the chapters in this volume, although Haig seems to suggest that they are frequently borrowed (either in form or construction) in the languages of Anatolia (§3 of Chapter 8). Perhaps most interestingly, Aikhenvald (p.c.) notes that interjections have been borrowed in form from Tucanoan languages into Tariana, despite the otherwise complete absence of lexical loans.

Discourse markers appear to be easily transferred from language to language (cf. §2.1 of Chapter 6). In some cases both form and use are transferred (although as the study of the meaning of discourse markers is not well advanced within linguistics, it is not clear that the use of such markers is precisely carried over); however, in these cases their status is not always clear. Thus Dimmendaal (p.c.) suggests that the frequent use of Swahili discourse markers in various varieties of African English may more strictly speaking be code-switching rather than borrowing.

Other cases of transference are clearly not code-switching as there is no borrowing of forms, but simply strong similarities of use and position of discourse markers. Examples include the sentence connectors found in Hittite and Hattic (see Chapter 3), the topic-switch markers of eastern Anatolia (see §3.7.1 of Chapter 8), and the use of *then* at the beginning of discourse segments in Hong Kong English to coincide with the use of a Cantonese particle (see §2 of Chapter 9).

5.8. FREE GRAMMATICAL FORM-AND-MEANING

Unbound grammatical forms and meanings are usually considered to be borrowed with more frequency than their bound equivalents. There are a few examples in this volume, although interestingly almost all examples are in languages which have also borrowed bound morphemes. Resígaro has borrowed two pronouns from Bora, although with some changes (see §4.2.2 of Chapter 7); Dongolawi-Kenuz has borrowed some postpositions and demonstrative and

interrogative pronouns (see §2.2 of Chapter 14); and Laz has borrowed the complementizer *ki* from Turkish, although the morpheme itself was borrowed into Turkish from Iranian languages (see §3.1.1 of Chapter 8).

5.9. BOUND GRAMMATICAL FORM-AND-MEANING

There are relatively few examples (in this volume or elsewhere) of bound grammatical forms and meanings being transferred from one language to another. In all cases only particular exponents of grammatical categories appear to have been borrowed, rather than an entire paradigm of forms.² Thus Resígaro has borrowed dual number markers, some classifiers, and some oblique case affixes (see §4.2.2 of Chapter 7); and Dongolawi-Kenuz has apparently borrowed a wide variety of bound morphology, including plural suffixes and verbal aspect markers (see §2.2 of Chapter 14). Kurmanjî and Zazaki have borrowed the form and meaning of the Turkish protasis enclitic *-sE* (see §3.2 of Chapter 8), although it is not clear whether enclitics should be treated as bound or free morphemes.

An interesting example of borrowing of bound morphology is found in Taiwanese Southern Min (see §3.2 of Chapter 12). Taiwanese Southern Min has borrowed the form *su* from Mandarin as an agentive suffix, but with a slightly different use; while it is the general agentive suffix in Mandarin, it is only used in Taiwanese Southern Min to form terms for intellectual professions, with the native *sai-hū* being used as an agentive suffix for lower-status professions.

5.10. GRAMMATICAL CATEGORIES

Grammatical distinctions of various types may be transferred from one language to another, with or without borrowing of forms.

In some cases, an additional distinction is added to an already existing grammatical category. Thus, for example, Arawak languages in general have a category of number, but with only singular and plural. Under contact with Bora, Resígaro has borrowed forms and established a third term in the system, dual, while retaining its inherited singular and plural as they were, although presumably the plural is now 'more than two' rather than 'more than one' (see §4.2.2 of Chapter 7). A particularly common additional distinction which languages appear to acquire in contact situations is that between inclusive and exclusive first person, either by borrowing a form, as in Resígaro (see §4.2.2 of Chapter 7), or by reanalysis of existing material, as in Northern Mandarin (see §3 of Chapter 9); note that Takia has not lost this distinction under contact with Waskia (see §2.1 of Chapter 6). Distinctions in a system can be lost under contact, with Takia having lost the

² The only examples from general language-contact literature which involve the 'borrowing' of entire paradigms appear to be Ma'a and Copper Island Aleut, and these are somewhat contested, in that some scholars believe they are originally Bantu and Aleut languages respectively, with extensive borrowing from Cushitic and Russian, while others consider that they are originally Cushitic and Russian, with borrowing from Bantu and Aleut.

distinction between consumable and neutral alienable possession (see §2.1 of Chapter 6).

In other cases, rather than simply add or lose an additional term in the system, the entire structure of a grammatical category is reorganized. Arawak languages and Bora both have a tense system expressed through suffixes; however, the systems are quite distinct. The Arawak language Resígaro, in contact with Bora, has completely reorganized its tense system to match that of Bora, although it has done so without borrowing any forms, but by reanalysing already existing suffixes, so that for example the new future tense suffix *-vá* is in origin an incomplete aspect suffix (see §4.2.2 of Chapter 7). A similar restructuring occurred in the case system of Hittite to develop an ergative, probably from an ablative-instrumental with resegmentation (see Chapter 3).

In addition to adding or losing a term in a system of grammatical categories, or reorganizing the system, languages can create or lose an entirely new grammatical category under contact. Thus in §3.3 of Chapter 13, Dimmendaal discusses the areal nature of noun class and concord systems in Niger-Congo and nearby languages, and notes that these systems can develop under contact (with both borrowing of forms together with lexical items, and internal development), as well as being able to be lost under contact with languages without noun-class systems. In some Australian languages such as Kugu-Muminh (see §3.2 of Chapter 4), bound person-and-number marking has grammaticalized under areal influence from earlier pronouns; and similar paths of grammaticalization of person-marking under contact are suggested by LaPolla for some Tibeto-Burman languages (see §3 of Chapter 9).

As well as addition of items to a grammatical category, reorganization of the semantics of the category, and addition and loss of grammatical categories, it is possible for a language to reorganize its exponents of grammatical categories under contact. While there are no clear cases in this volume, an example is discussed by Trask (1996: 310–11), who notes that Old Armenian had a standard Indo-European system of fusional case-and-number-marking. While Modern Armenian expresses the same categories, and uses native material to do so, the system of expression has been remodelled on the basis of Turkish, with case and number being expressed agglutinatively, so that following a root comes the plural marker (identical for all cases) followed by the case marker (identical regardless of number).

5.11. POSITION OF MORPHOLOGY

One of the interesting features of the influence of one language on another with regard to free or bound grammatical morphology appears to be the importance of the position of these elements with respect to some other element or elements. One of the universals suggested by Moravcsik (1978) is that a grammatical word cannot be borrowed unless the linear order with respect to its head is also

'borrowed', and this is probably the universal which has stood up best to the test of time and counter-examples.³

There are relatively few clear examples of the importance of position with borrowed bound morphology in this volume. As mentioned in §5.9 above, there are few examples of bound morphological forms being borrowed, and while in all cases the morpheme in question then occurs in the same position as in the original language, the examples involve languages of the same typological profile in any case. Thus, for example, Resígaro has borrowed the dual markers as suffixes from Bora suffixes, but Resígaro only has suffixes (see §4.2.2 of Chapter 7).

It is important to note that Moravcsik's universal simply specifies linear order, not any hierarchical relationship. Campbell (1993: 103) suggests that Moravcsik's universal is somewhat counter-intuitive, since it suggests that when a language is in contact with languages of different constituent order types, it can only borrow a grammatical word by opposing its native constituent order. But this confuses the levels of word form and word class (cf. §2.2 above). For example, Turkish has borrowed the complementizer *ki* from Iranian languages, retaining its linear position (see §3.1.1 of Chapter 8). However in Iranian languages such as Persian, the morpheme *ke* is the first element in the complement clause, whereas in Turkish and Laz the corresponding element (*ki*) is the last element of the matrix clause—clearly an entirely different hierarchical relationship between *ki* and other elements of the sentence exists in the two languages, but with precisely the same strict linear positioning (cf. Gerritsen and Stein's (1992: 6) comment that 'syntactic change caused by borrowing does not have to result in the same construction in the receiving as the giving language').

While the discussion of the position of borrowed bound and unbound morphology has usually only related to borrowed forms, it is interesting to examine the case of reanalysis of existing native material. It appears that in some cases having an appropriate grammatical category is not sufficient for a language in contact—it requires the morpheme expressing the category to be in the same position as in the contact language.

Sometimes this is achieved by simply reordering existing equivalent morphemes (although this appears to be surprisingly uncommon); in other cases new morphemes are reanalysed from a different system, with the old morphemes expressing the category in question being lost. Thus in §2.1 of Chapter 6, Ross

³ Trask (1996: 314–15) gives what appears to be a counter-example from Basque, which has borrowed the Spanish preposition *contra* 'against' as a postposition. While this is the best example I have seen, it is not completely convincing. First, it is clear that some adpositions are more lexical than others (cf. Myers-Scotton 1998: 293); and second, while in some uses *contra* is a simple preposition in Spanish, in other cases it forms the noun-like portion of a complex preposition, *en contra de N*, literally 'in against of N, in N's against'. Both of these constructions are found in Basque (with different cases on N), and potentially at least (pending historical investigation) *contra* was borrowed first as a type of noun for use in the more complex construction; and because of the structure of Basque NPs, this would place *kontra* after its possessor N, precisely the same position as it now has as a simple post-position.

discusses the development of a system of determiners in Takia from earlier demonstratives under contact from Waskia. In fact Takia already had a system of determiners, but these were linearly in the 'wrong' place, while the demonstratives were in the 'right' place; and rather than move its existing determiners to match the order patterns of Waskia, Takia developed a new series of determiners.

This position-matching of reanalysed material between languages does not seem to be an absolute rule, however. La Polla suggests that a variety of Tibeto-Burman languages have developed person-marking on verbs under contact, but in some of these languages the new category is expressed through verb prefixes, while others use verb suffixes (see §2 of Chapter 9). Potentially, of course, this is a distinction between bound and free morphology, with free morphological items matching linear order, while bound items do not.

5.12. SYNTACTIC FRAMES

Contact-induced changes in syntactic frames are those changes where, with no alteration in the paradigms of morphological material or what would normally be considered extensions in the meanings of the categories, particular categories are used in circumstances where they were not previously used. This possibility has not received much attention in the literature, but some cases have been presented. Thus Harris and Campbell (1995: 142) discuss the findings of Timberlake (1974) that the use of nominative case for objects in impersonal constructions in Russian has been borrowed from Finnish. Haig's discussion of the reorganization of the case system of Ardeşen Laz in §4.3 of Chapter 8 makes the situation there appear very similar; comparable too is the relative-clause construction in formal Cantonese discussed by Chappell in §3.4 of Chapter 12, and some of the event schema convergences discussed by Heine and Kuteva in §3 of Chapter 14. This sort of change is not necessarily associated with any constituent-order change; it is purely the use of particular morphological patterns in particular contexts.

5.13. CLAUSE-INTERNAL SYNTAX

The transfer of order of constituents at various levels from one language to another has been extensively discussed in language-contact studies (see, for example, Harris and Campbell 1995: 136–41), and many examples of this are found in this volume—for example, Dimmendaal shows that clause-level constituent order in Baale has been 'freed up' in line with neighbouring Tirma-Chai (see §2.2 of Chapter 13), Aikhenvald notes that Tariana has verb-final clauses in line with Tucano languages (see §4.1.2 of Chapter 7), and Heine and Kuteva note the verb-initial convergence in the Rift Valley of Africa (see §2.1 of Chapter 14).

Examples of change of order within major constituents is less frequent, although Takia has developed close similarities to Waskia, developing for example postpositions rather than prepositions and postnominal determiners rather than prenominal determiners (see §2.1 of Chapter 6). It is not entirely clear, however,

whether features such as these indicate transference of within-constituent order from one language to another, or are at least partially attributable to languages regularizing their word-order typology (cf. Harris and Campbell 1995: 136–41): once Takia had changed its clause-level constituent order to verb-final, there would presumably be typological pressure towards consistent head-final constituents as well as language-contact influences.

5.14. BETWEEN-CLAUSE SYNTAX

The transfer of between-clause syntax has not been as extensively studied as within-clause syntax. However there are two sets of good examples in this volume. In §3 of Chapter 8, Haig discusses a wide variety of similarities of clause-linkage strategies in Anatolian languages, including complement clauses, relative clauses, and various adverbial clauses, such as 'after' clauses, 'so that' clauses, 'nevertheless' clauses, and 'either-or' clauses. In many cases the structuring of these clauses is identical across languages, regardless of whether borrowed forms or native material is involved. Similarly, Ross describes the development in Takia of clause-chains of a type not usually found in Oceanic languages, but commonly used in Waskia, in §2.1 of Chapter 6. Takia has developed this method of clause-linking under contact with Waskia, but has reanalysed native conjunctions to form the non-final clause enclitics.

5.15. DISCOURSE TYPES AND DISCOURSE ORGANIZATION

One final issue which is not explicitly addressed in this volume is the transfer of discourse-level features such as genre types and the organization of presentation of discourse. These features are implicit in discussions of metatypy (see §2.1 of Chapter 6), linear alignment (see §5 of Chapter 8), and common ways of construing of the world (see §3 of Chapter 9). In fact, however, it is not clear whether discourse genres and discourse organization will necessarily transfer from language to language. In particular, these sorts of features are probably most prone to being affected by the type of contact. It seems intuitively reasonable that in a situation of stable multilingualism where the same sorts of tasks can be carried out in all languages, with the language chosen depending on location and participants, convergence of features such as discourse genres and discourse organization might be expected. On the other hand, language contact often occurs in situations of diglossia (Ferguson 1959), in which case a convergence of discourse genres may not be expected, as there are certain tasks which can only be carried out in one language or another. Whether there are clear cases of convergence of discourse organization in conditions of diglossia remains to be seen.

6. Conclusions

This chapter has examined the transfer of language features from one language to another, concentrating on issues which have arisen in the other chapters in this

volume. It discussed first what sorts of 'borrowing' need to be taken into account—acquisition of features, loss of features, retention of features, or some combination of these. The issue of the formal or functional use of terms (such as 'verb' or 'derivation') in constraints was examined, as was the issue of borrowing of forms versus borrowing of concepts expressed by restructuring native material. Then, after examining the sorts of hierarchies and constraints which can and have been developed, it looked at some of the factors discussed in this volume and elsewhere which impede the development of hierarchies and constraints on borrowing—the social, political, and historical context of the languages in contact, borrowing versus substratum influence, emblematicity constraints, the problem of language death, the issue of the reliability of data, and the problem of multiple causation for language change. Finally there was an overview of the sorts of features which can be affected when two (or more) languages are in contact, giving a summary of features which the various contributors to this volume consider have been transferred from one language to another.

What conclusions can we draw about the development of universal constraints on borrowability on the basis of the data presented here? Unfortunately, the probable conclusion is that we may never be able to develop such constraints. We would need to take into account far more information than is usually available, and factor out all possible influences, whether of a sociopolitical or historical nature, or to do with the pre-existing structure of the languages before contact. At the same time, we would need to examine an astonishingly broad range of potentially transferable features of language.

This is not to say, of course, that there is not a wide variety of issues in contact-induced change which can be effectively examined and discussed without a series of universal constraints or a general hierarchy of borrowability. A great deal of interesting and useful data can be collected, and with care some potentially fruitful hypotheses can be developed on the basis of this data. It is possible that a variety of constraints on borrowing in particular contexts can be developed. But the attempt to develop any universal hierarchy of borrowing should perhaps be abandoned.

References

- AIKHENVALD, A. Y. 1996. 'Areal diffusion in north-west Amazonia: the case of Tariana', *Anthropological Linguistics* 38.73–116.
- ALLEN, C. L. 1980. *Topics in Diachronic English Syntax*. New York: Garland.
- APPEL, R., and MUYSKEN, P. 1987. *Language Contact and Bilingualism*. London: Edward Arnold.
- BRENNINGER, M. 1997. 'Language contact and language displacement', pp. 273–84 of Coulmas (1997).
- CAMPBELL, L. 1993. 'On proposed universals of grammatical borrowing', pp. 91–109 of *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, Rutgers University, 14–18 August 1989*, edited by H. Aertsen and R. J. Jeffers. Amsterdam: John Benjamins.

- CLYNE, M. 1997. 'Multilingualism', pp. 300–14 of Coulmas (1997).
- COULMAS, F. 1997. Editor of *The Handbook of Sociolinguistics*. Oxford: Blackwell.
- CRAIG, C. G. 1997. 'Language contact and language degeneration', pp. 257–70 of Coulmas (1997).
- CURNOW, T. J. 1997. 'A Grammar of Awa Pit (Cuaiquer): An Indigenous Language of Southwestern Colombia'. Ph.D. dissertation, Australian National University.
- DIXON, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- DORIAN, N. C. 1989. Editor of *Investigating Obsolescence: Studies in Language Contraction and Death*. Cambridge: Cambridge University Press.
- FERGUSON, C. A. 1959. 'Diglossia', *Word* 15, 325–40.
- GERRITSEN, M., and STEIN, D. 1992. 'Introduction: On "internal" and "external" in syntactic change', pp. 1–15 of *Internal and External Factors in Syntactic Change*, edited by M. Gerritsen and D. Stein. Berlin: Mouton de Gruyter.
- GRENOBLE, L. A., and WHALEY, L. J. 1998. Editors of *Endangered Languages: Language Loss and Community Response*. Cambridge: Cambridge University Press.
- HARRIS, A. C., and CAMPBELL, L. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.
- HARRIS, J. 1991 [1987]. 'Conservatism versus substratal transfer in Irish English', pp. 191–212 of *Dialects in English: Studies in Grammatical Variation*, revised edition, edited by P. Trudgill and J. K. Chambers. London: Longman.
- HAUGEN, E. 1950. 'The analysis of linguistic borrowing', *Language* 26, 210–31.
- HEATH, J. 1978. *Linguistic Diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- LASS, R. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- LEHMANN, W. P. 1962. *Historical Linguistics: An Introduction*. New York: Holt, Rinehart, and Winston.
- MILROY, J. 1997. 'Internal vs. external motivations for linguistic change', *Multilingua* 16, 311–23.
- MORAVCSIK, E. 1978. 'Language contact', pp. 3–123 of *Universals of Human Language*, vol. 1: *Method and Theory*, edited by J. H. Greenberg, C. Ferguson, and E. Moravcsik. Stanford: Stanford University Press.
- MYERS-SCOTTON, C. 1998. 'A way to dusty death: the Matrix Language turnover hypothesis', pp. 289–316 of Grenoble and Whaley (1998).
- NELDE, P. H. 1997. 'Language conflict', pp. 285–300 of Coulmas (1997).
- ROSS, M. D. 1988. *Proto Oceanic and the Austronesian Languages of Western Melanesia*. Canberra: Pacific Linguistics, Australian National University.
- SORENSEN, A. P., JR. 1972 [1967]. 'Multilingualism in the northwest Amazon', pp. 78–93 of *Sociolinguistics: Selected Readings*, edited by J. B. Pride and J. Holmes. Harmondsworth: Penguin Books.
- STEFANOWITSCH, A. 1999. 'Sum: Pseudo-loanwords', *Linguist List* 10.1388 (22 September 1999). <http://linguistlist.org/issues/10/10-1388.html>
- THOMASON, S. G., and KAUFMAN, T. S. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- TIMBERLAKE, A. 1974. *The Nominative Object in Slavic, Baltic, and West Finnic*. Munich: Otto Sagner.

- TRASK, R. L. 1996. *Historical Linguistics*. London: Arnold.
- 1998. 'The typological position of Basque: then and now', *Language Sciences* 20.313–24.
- TRUDGILL, P. 1974. 'Linguistic change and diffusion: description and explanation in socio-linguistic dialect geography', *Language in Society* 3.215–46.
- VAKHTIN, N. 1998. 'Copper Island Aleut: a case of language "resurrection"', pp. 317–27 of Grenoble and Whaley (1998).
- WEINREICH, U. 1953. *Languages in Contact*. The Hague: Mouton.
- WILKINS, D. P. 1996. 'Morphology', pp. 109–17 of *Kontaktlinguistik: ein internationales Handbuch zeitgenössischer Forschung* = *Contact linguistics: An international handbook of contemporary research*, vol. 1, edited by H. Goebel, P. H. Nelde, Z. Sary, and W. Wölck. Berlin: Walter de Gruyter.

Index of Authors

- Aikhenvald, A. Y. 1–26, 64, 146, 152, 167–95, 214,
222, 342, 360 n., 377, 388, 401 n.12, 420,
426, 428, 432
- Alavi, B. 201
- Allen, C. L. 424
- Allin, T. R. 175, 182, 185–9
- Alpher, B. 84
- Amberber, M. 349
- Andersen, H. 14, 155, 158
- Andrews, P. 198
- Ansaldi, U. 342
- Appel, R. 222, 417, 424
- Archangeli, D. B. 368
- Aronson, H. 211
- Aschmann, R. P. 175, 182, 184, 186–8
- Atatürk, K. 16
- Austin, P. 90, 107–8, 110, 112, 114–16, 120
- Auwers, J. van der 12
- Awolaye, Y. 407
- Backus, C. 237
- Bakker, P. 19, 159, 214
- Barker, F. 139 n.
- Barnas, U. 196 n., 201
- Baron, S. P. 233
- Baudouin de Courtenay, J. 5
- Bauer, R. S. 264, 266
- Baumhoff, M. 39
- Baxter, H. 225, 329, 344 n.3
- Beard, J. 106
- Bechhaus-Gerst, M. 397, 399–400, 401 n.3
- Bedir Khan, E. D. 202–3, 205
- Behzad, F. 201
- Bellwood, P. 27–43, 232, 334
- Bender, L. 408 n.24
- Bendix, E. H. 239
- Bendor-Samuel, J. 369, 377
- Benedict, P. K. 225–6, 264, 294, 306, 308, 313–14,
318, 320 n., 321
- Benveniste, E. 48–9
- Bernot, D. 214
- Berry, J. 39
- Berry, M. S. 39
- Bettinger, R. 39
- Bhattacharya, S. 242
- Bickner, R. J. 322
- Bielenstein, H. 231
- Birdsell, J. B. 65
- Bisang, W. 220, 259, 269, 336, 342
- Blake, B. J. 86, 92, 95, 97
- Bläsing, U. 195 n.
- Blevins, J. 117
- Blust, R. A. 31–2, 134, 233
- Böhm, G. 406
- Bole-Richard, R. 374
- Bolt, J. E. 82
- Bradley, D. 226 n., 238–9, 242, 319
- Braun, F. 195 n.
- Breedveld, J. O. 380
- Brendemoen, B. 195 n.
- Brenzinger, M. 421
- Bright, W. 13, 14
- Brinton, D. G. 170
- Brody, J. 18
- Broeder, W. 195 n.
- Bulut, C. 195 n., 196 n.
- Burgess, E. 173
- Burling, R. 226 n., 314 n.
- Campbell, L. 1, 11, 13, 14, 18, 38, 64, 98, 139, 145,
154, 180, 185, 200, 206, 211, 217, 219–21,
283, 285, 415–16, 419, 424, 431–3
- Cao Guangqu 233
- Cao Shuji 228
- Capell, A. 89, 136
- Capo, H. B. C. 371
- Carlson, R. 382 n.
- Caughley, R. 309 n.11
- Cavalli-Sforza, L. 40
- Chamberlain, J. 264
- Chang, K. C. 227
- Chang Kuang-yu 233
- Chao, Y. R. 270
- Chappell, H. 1, 5, 10, 12, 17, 18, 22, 159, 233, 235
n.7, 242, 328–58, 424
- Chaudhuri, B. 241
- Chelliah, S. L. 241
- Chen, M. Y. 232
- Chen, P. 343
- Chen Baoya 232
- Chen Naixiong 234, 242
- Chen Ruxing 236 n.11
- Chen Xujing 236 n.11
- Cheung, S. H.-N. 335
- Chou Tsu-yao 321
- Clark, M. 259–61, 269
- Claudi, U. 344, 349, 396 n.
- Clyne, M. 414, 421, 425
- Coelho, E. 167 n.
- Coelho, L. 167 n.

- Comrie, B. 175, 220
Condominas, G. 267
Court, C. 322
Craig, A. M. 227
Craig, C. G. 421
Creissels, D. 369–70, 382, 395 n.2
Curnow, T. J. 16, 19, 22, 113 n., 195 n., 217, 412–36

Dai Qingxia 226 n., 228, 239, 245, 264, 308
Dawkins, R. M. 206
De Blois, R. 371
De Wolf, P. P. 377 n., 380
Déchaine, R.-M. 384
Dehghani, Y. 209, 221
DeLancey, S. 226 n.
Dempwolff, O. 134
Demuth, K. 381
Dench, A. 1, 4, 6, 7, 10, 20, 64, 105–33, 157, 423, 425
Denison, N. 146
Diamond, J. 38
Diffloth, G. 301
Diller, A. 19, 220, 266, 279
Dimmendaal, G. J. 3, 8, 13, 15, 16, 22, 239, 358–92, 414, 425–6, 428, 430
Divshali, S. 201
Dixon, R. M. W. 1–26, 27–30, 36, 48–9, 59–62, 64–104, 105, 110, 114, 118, 131, 138–9, 153–4, 157, 168, 175, 191, 212, 227, 232, 243, 293–4, 323, 330, 335, 349, 360 n., 364, 377, 388, 401, 414–15, 423
Donaldson, T. 74, 110, 111
Doneux, J. L. 377 n., 380
Dorian, N. C. 421
Dorleijn, M. 195 n., 196 n., 203, 209, 213
Douglas, W. 120
Downer, G. B. 317
Drewes, A. J. 151
Driem, G. van 225 n.1
Du Ruofu 231 n.4
Dumézil, G. 196 n., 211, 214, 216
Durie, M. 283, 384
Dutton, T. E. 138
Dwyer, A. M. 234

Edmondson, J. A. 256, 262, 264, 279, 319
Elders, Stefan 381
Elfenbein, J. 206
Elugbe, B. O. 369, 371
Emeneau, M. B. 5, 11, 206, 209
Enfield, N. J. 1, 12, 14, 17, 18, 20, 195 n., 243, 255–90, 301, 337, 353, 421, 424, 427
Esenç, T. 196 n., 211, 214, 216
Evans, G. 29 n., 92–3, 127, 258, 361
Evans, J. P. 92–3, 235
Evans, N. 361

Fairbank, J. K. 227
Faraclas, N. 374, 379, 381
Fei Xiaotong 236
Ferguson, C. A. 395, 433
Ferlus, M. 320 n.14
Feuerstein, W. 198
FitzGerald, C. P. 227, 236 n.11
Foley, W. A. 144
Ford, K. C. 386
Forrest, R. A. D. 225 n.2, 302 n.6
Frajzyngier, Z. 349
Friedman, V. A. 17, 146
Fu Ailan 226 n., 228, 245

Gage, W. 317
Gandour, J. 321–2
Ge Jianxiong 228
Gedney, W. J. 264, 322
Gerritsen, M. 431
Gil, D. 160
Gilij, Father 170
Gilman, C. 395
Giridhar, P. P. 244
Goddard, I. 60, 98
Golovko, E. V. 159
Grace, G. W. 134, 139, 146–8, 156–7
Grannes, A. 209
Green, D. 167 n., 172
Green, H. 167 n., 172
Green, I. 82
Greenberg, J. H. 3, 8, 66, 89, 98, 145, 167–8, 220, 358 n., 359, 361, 366–8, 373, 376, 380, 382, 393–6, 406
Gregerson, K. J. 305, 319
Grenoble, L. A. 421
Grierson, G. 226 n., 241, 266
Grimes, B. F. 239, 360–1
Grootaers, W. 329–30
Gruiter, M. de 155
Gudschinsky, S. 7
Güldemann, T. 346, 349–50, 395
Gumperz, J. J. 10, 13, 17, 147
Guthrie, M. 370, 374, 376

Haase, M. 146, 152–3
Haig, G. 11, 12, 17, 18, 20, 195–224, 404 n.14, 405 n.17, 418, 421, 424, 428, 432–3
Haiman, J. 146
Hale, K. L. 80, 89, 93, 118
Hall, D. G. E. 236 n.11
Hansell, M. D. 234
Hanson, O. 308
Harris, A. 18, 139, 145, 211–12, 217, 219, 220–1, 283–5, 413, 416, 424, 432–3
Harris, J. 422, 424
Hartmann, J. 258
Hashimoto, M. J. 230, 330, 335–6, 338, 350

- Haspelmath, M. 13, 153, 206
 Haudricourt, A.-G. 302 n.5, 304, 313–15, 319, 321
 Haugen, E. 417–18, 427
 He Jen 228
 Headland, T. 41
 Heath, J. 13, 16, 19, 44, 60–1, 82, 91, 180, 222, 227, 401, 415
 Heine, B. 2, 7 n., 8, 10, 22, 195 n., 206, 242–3, 284, 335, 344, 346, 348–51, 376, 377 n.1, 387, 393–411, 427, 432
 Henderson, E. J. A. 313
 Hercus, L. A. 78
 Hill, J. 38
 Hillman, G. 34, 35
 Hinloopen-Labberton, D. V. 320 n.15
 Hinnebusch, T. 359–60
 Hjelmlev, L. 148 n.
 Ho, M.-L. 160
 Ho Dah-an 232
 Hock, H. H. 191, 202, 387
 Hoddinott, W. G. 82
 Hodgson, B. H. 303
 Hoffman, H. 236
 Hoffmann, C. 381
 Höhlig, M. 16, 18
 Holisky, D. A. 196 n., 198, 203–5, 210–11, 214–15
 Hopper, P. J. 284
 Huang, B. 245, 346
 Huang Yuan Wei 233
 Huber, R. Q. 175
 Hübschmann, H. 59
 Hünemeyer, F. 344, 349
 Hunt, T. 41

 Ikehara-Quebral, R. 41
 Ikoro, S. M. 384, 386
 Iwata, R. 233

 Jackson, J. 177
 Jakobson, R. 158, 239
 Jamison, S. 59
 Jasanoff, J. 59
 Jeyapaul, V. Y. 244
 Jin Lixin 225 n.1
 Johanson, L. 11, 17, 195 n., 209, 217–18, 221
 Johnson, S. 80, 157
 Jones, R. B. 314 n.
 Joseph, B. D. 1
 Junggraithmayr, H. 373
 Jurafsky, D. 346

 Kahn, M. 199
 Kansakar, T. R. 239–40
 Karlgren, B. 328
 Kaufman, T. 1, 12, 13, 132, 145–6, 152, 154, 158–9, 180, 200, 206, 219, 268, 298–300, 302, 400, 412, 415–16, 418–20, 424

 Kawamoto, T. 320 n.15
 Kepping, K. B. 243
 Keyes, C. 258
 Kilham, C. 80
 Kilian-Hatz, C. 346, 349–50
 Kirk-Greene, A. H. M. 407
 Kiu, K. L. 322
 Koelle, S. W. 366
 Kofod, F. M. 82
 Köhler, O. R. A. 8
 Kraft, C. H. 407
 Kubler, C. C. 159, 234, 350
 Kühne, C. 51
 Kurylowicz, J. 63
 Kuteva, T. 2, 7 n., 8, 10, 22, 206, 243, 335, 387, 393–411, 432
 Kutscher, S. 196 n., 205

 Labov, W. 147
 Ladefoged, P. 374
 Lamarre, C. 281
 LaPolla, R. J. 3, 6, 10, 21, 225–54, 256, 334, 427, 432
 Lass, R. 60, 285, 416–17, 423–4
 Last, M. 361, 363
 Laufer, B. 236
 Laughren, M. 111
 Laycock, D. C. 147, 156
 Le Page, R. B. 267
 Leach, E. R. 236 n.11, 238, 256, 258, 267
 Lee, J. 139, 228, 230
 Lee, Y.-J. 317
 Lee Tsung Dao 231 n.4
 Lehmann, W. P. 413
 Lescot, R. 196 n., 202–3, 205
 Lessau, D. 346, 349–50
 Levelt, W. J. M. 148–9
 Lewis, G. 208–9
 Lewis-Williams, J. D. 395
 Li, C. N. 18, 273–5, 301
 Li, F. 264
 Li Fang-Kuei 226 n.
 Li Jingzhong 231 n., 232
 Li Shaoming 227
 Li Wen-Chao 230
 Liddicoat, T. 427
 Lien, C. 231, 232, 339–40, 344
 Lin Tao 230
 Lin Xiangrong 235 n.9
 Lin Zongcheng 237
 Liu Guangkun 235
 Liu Juhuang 226 n., 228, 245
 Lojenga, C. K. 370, 381
 Long Y. 264–5
 Longacre, B. 39 n.
 Loos, E. E. 182
 Lorenz, M. 201

- Loukotka, C. 182, 184
Lucassen, L. 363
Luce, G. H. 225 n.2, 236 n.11
Luo Y. 262, 264, 279
Lydall, J. 373
Lynch, J. 138, 158
- Mabry, J. 39, 40
McCarthy, J. 198
McConvell, P. 65
McGregor, W. 354
MacKenzie, D. 196 n.
McSwain, R. 138, 147
Maddieson, I. 374
Manessy, G. 377 n.
Mao Tsung-wu 321
Maran, L. 307 n., 308
Marchese, L. 381
Marmion, D. 117, 120
Martin, S. 320 n.15
Martins, S. A. 167 n., 191
Martins, V. 167 n., 191
Masica, C. P. 1, 13, 180
Matisoff, J. A. 5–6, 11, 12, 13, 16, 17, 22, 225–6, 234, 235 n.7, 241–3, 259, 269, 276, 284, 286, 291–327, 336, 344 n.5, 347, 349–50, 353, 376, 426
Matras, Y. 195 n.
Matteson, E. 171, 173
Matthews, S. 265–6, 270, 273, 281, 341–2
Mattissen, J. 196 n., 205
Mazaudon, M. 304, 309–11, 315
Meeussen, A. E. 370, 374, 381
Mei Tsu-lin 225 n.1, 233
Melchert, H. C. 52, 54
Meng Simu 233
Merlan, F. C. 71, 73
Michailovsky, B. 225 n.1, 303
Migliazza, B. 259
Militarev, A. Y. 10
Miller, R. A. 320 n.15
Milroy, J. 412
Milroy, L. 268
Minor, E. E. 182
Mithun, M. 98
Moges Yigezu 361, 365
Moravcsik, E. 416, 419, 430–1
Mountain, J. L. 231 n.4
Mukarovsky, H. 368, 373, 380
Murayama, S. 320 n.15
Murdock, G. P. 33
Muro, M. 39
Muysken, P. 156, 222, 417, 424
Myers-Scotton, C. 412, 420 n., 431 n.
- Nash, D. 84
Nedjalkov, I. 220
Nelde, P. H. 421
Nesin, A. 207
Nestmann, L. 196
Newman, P. 366
Nichols, J. 31, 63
Niederer, B. 296
Nimuendajú, C. 180
Nishi Yoshio 239
Noble, G. K. 171
Nordlinger, R. 82–3
Norman, J. 231, 233, 334, 338, 344 n.3, 350
Nurse, D. 146, 359–60
- O'Grady, G. N. 89–90, 98, 108, 110–12, 114–15, 118
Okell, J. 238, 260, 276, 286
Oliver, J. R. 171
Oppitz, M. 239–40
Ostapirat, W. 309 n.11, 315
- Pamulkan, M. 80
Paol, J. N. 139
Papen, R. 214
Parks, D. 14
Paterson, A. 117
Paul, L. 195 n., 196 n., 198, 201–5, 208–9, 213
Pawley, A. K. 31, 136, 149, 384
Payne, David L. 168–70, 173–4, 187
Payne, Doris L. 168–9
Payne, J. 347, 353
Pe Maung Tin 236 n.11
Pensalfini, R. T. 82
Pietrusewsky, M. 41
Platt, J. T. 159–60
Plog, S. 40
Poffenberger, M. 239
Pontius, J. 16
Pootchemunka, J. 80
Possehl, G. 38
Prince, E. F. 150
Przyluski, J. 315
Pulleyblank, E. G. 226–7, 232, 313, 368
- Qian Naong 233
Qu Aitang 239 n., 240 n.
Quine, W. V. O. 272 n.
- Radcliffe-Brown, A. 108
Ran Guangrong 227, 236
Randal, S. 364
Ranoo Wichasin 266
Ray, S. H. 136
Reed, R. B. 175
Rehburg, J. 139
Reid, L. 41, 175
Reischauer, E. O. 227

- Renfrew, C. 32, 34, 36, 42
 Renzengwangmu 245
 Richardson, H. 236
 Rivet, P. 182, 184
 Roberg, H. 346, 349–50
 Roberge, P. T. 159
 Romaine, S. 222
 Romney, A. K. 39
 Rosenberg, M. 34
 Ross, M. D. 10, 12, 14, 17, 20, 109–10, 131–2,
 134–66, 195 n., 200, 217–18, 242, 268, 283,
 341, 360, 402, 413, 417–18, 421, 427, 431,
 433
 Ross, P. A. 264
 Rowan, O. 173
- Sagart, L. 294–5, 317, 329, 334, 344 n.3
 Salmons, J. 18
 Salzer, J. 196 n., 201
 Samerchai, P. 258
 Sandfield, K. 400 n.11
 Sands, K. 93, 110, 118–19
 Sapir, E. 3, 49, 169, 369, 380
 Sasse, H.-J. 11, 146, 149–50, 152, 222
 Saxena, A. 235, 349
 Schadeberg, T. 360–1, 375
 Scheffler, H. 108
 Schladt, M. 346, 349–50
 Schleicher, A. 4
 Schmidt, J. 4
 Schmidt, W. 89, 320
 Schroeder, C. 195 n.
 Seki, L. 14
 Serafim, L. A. 320 n.15
 Serruys, P. L.-M. 329
 Shafer, R. 226 n.
 Sharma, D. D. 241
 Sharma, S. R. 243
 Sharp, J. 123
 Shefts, B. 309
 Sherzer, J. 1, 13, 14, 180
 Shevelov, G. Y. 5
 Shibatani, M. 351–2
 Shiratori, Yoshiro 237
 ShreeKrishan 234, 241
 Silverstein, M. 120
 Singh, A. P. 241
 Smeets, R. 377
 Smith, I. 80
 Smith-Stark, T. 1, 12, 13, 180, 200, 219
 Snellgrove, D. 236
 Solnit, D. B. 262, 264
 Solomon, J. R. 320 n.15
 Soper, J. 202 n., 209, 212, 220
 Sorensen, A. P. Jr. 15, 177, 420
 Sprigg, R. K. 311
 Stassen, L. 206, 403–5, 406 n.18
- Stefanowitsch, A. 427
 Stein, D. 431
 Stein, R. A. 236
 Steinberg, D. J. 258
 Stern, T. 241
 Stewart, J. 368, 371–4, 376
 Stolz, T. 346, 349–50
 Storch, A. 381
 Strecker, D. 296
 Sun C. F. 279–80
 Sun Hongkai 226 n., 235–6
 Sutton, P. 80
 Suwilai, P. 262
 Svantesson, J.-O. 321
 Swadesh, M. 175
 Syder, F. 149
- Tabouret-Keller, A. 267
 Tandilava, A. 198, 214
 Taracha, P. 51, 54
 Teo Kok Seong 303
 Tépahae, P. 158
 Terwiel, B. J. 266
 Thelwall, R. 397, 399
 Thiesen, W. 182, 185–9
 Thomason, S. 132, 145–6, 152, 154, 158–9, 206,
 268, 298–300, 302, 400, 412, 415–16,
 418–20, 424
 Thompson, L. C. 264, 276
 Thompson, S. 273–5, 301, 351
 Thurgood, G. 146, 152, 244, 320–1
 Thurston, W. R. 138, 146, 149–50, 155
 Tian Jizhou 231 n.5
 Timberlake, A. 432
 Tindale, N. 108
 Ting, P.-H. 334
 Tiwari, S. C. 241
 Tong Enzheng 232
 Tosco, M. 2
 Tovar, A. 171
 Tovar, L. de 171
 Trask, L. 17, 221, 419–20, 430, 431 n.
 Traugott, E. C. 284
 Treistman, J. M. 227
 Trubetzkoy, N. S. 17, 52, 59, 426
 Trudgill, P. 425
 Tschenkéli, K. 205
 Tuite, K. 195 n., 209, 211
 Tuominen, S. 139
 Turnbull, W. 87
- Vakhtin, N. B. 159, 420
 Van Bruinessen, M. 198
 Van der Loon, P. 344 n.4
 Van Leynseele, H. 371
 Van Valin, R. D. Jr. 144
 Vanilişi, M. 198–214

- Varasarin, U. 303
Voegelin, C. 89–90, 98, 107, 110
Voegelin, F. 89–90, 98, 107, 110
Voeltz, E. 385
Vossen, R. 388
Vovin, A. 294
- Wang, J. 264–5
Wang, W. S.-Y. 231–2
Wang Huiyin 227
Wang Jun 230–1
Wang Ming-ke 228
Waters, B. 139
Watkins, C. 5, 9, 13, 16, 17, 19, 20, 23, 44–63, 227, 388, 414, 424–6
Wavrin, R. de 182, 184
Weidert, A. 309 n.11, 315
Weinreich, U. 149–50, 152, 415
Welmrs, W. 380
Werner, R. 397, 399 n.
Westermann, D. 359
Westphal, E. O. J. 8
Whaley, L. J. 421
Whiffen, T. W. 182
Wiens, H. J. 228
Wilkins, D. P. 284, 417
Williamson, K. 366–7, 369, 371, 377–8, 382, 384
- Wilson, R. 10, 13, 17, 147
Windfuhr, G. 217
Wise, M. 182
Wodarg, A. 196 n., 201–2, 204–5, 208–9, 214
Wolff, E. 405
Wolmby, T. 80
Wong Bin 228–30
Wordick, F. 116
Wu, Y. 342–3
Wu Songdi 228
Wurm, S. A. 89–90
- Xing Gongwan 227
Xu Jieshun 228
- Yang, H.-F. 345
Yip, V. 265–6, 270, 273, 281, 341–2
You Rujie 228, 230–1, 233, 334
Yuan, J. 330, 335
Yue-Hashimoto, A. O.-K. 233, 264, 340
- Zhao Tongmao 231 n.4
Zheng G. 264–5
Zhou Xiyin 227, 236
Zhou Zhenhe 228, 230–1, 233
Zhu, D. 338, 340

Index of Languages and Language Families

- Acehnese 319
 Achagua 174–5, 184, 187, 190–1
 Achang 314
 Adamawa 381
 Adamawa-Ubangi 358 n., 366, 376
 Adyghe 16, 18
 Aeolic 58
 African languages 2, 3, 7 n., 8–9, 22, 358–89,
 393–436
 Afrikaans 159–60, 423–4
 Afroasiatic 8, 11, 31, 36, 196, 358, 373–4, 387, 393,
 403, 405–8
 Agaw 3
 Ahi 310
 Ahom 266
 Akan 368–9, 371–3, 376, 386
 Akha 310
 Akkadian 51
 Alak 281
 Albanian 11, 45, 56, 146–50
 Aleut 159, 420, 429 n.
 Algonquian 29, 33, 214
 Altaic 220, 231, 234, 242, 294, 335–6, 349–50
 Alur 376–7
 Amba 381
 Amharic 219, 349, 406
 Amoy 337–8, 344
 Amuesha 176
 Anatolia(n) 44–63, 195–224, 424, 428
 Anejom 158
 Anêm 138, 146, 149, 150
 Angami Naga 244
 Añun 172
 Aputina 174
 Arabic 16, 33, 151, 196, 198–9, 206, 213, 359–60,
 421
 Classical 15, 404 n.15
 Arakanese 238
 Arakanese Burmese 241
 Aramaic 196
 Aranda 108
 Arandic 86
 Arawá 168
 Arawak 2, 6, 10, 11, 21, 29, 167–94, 429–30
 Ardeşen 211–12, 214–15, 217, 222, 421, 432
 Armenian 23, 45, 49, 56, 57, 59, 146, 195 n., 196,
 209, 221, 293 n., 430
 Aromanian 150
 Arop-Sissano 156
 Arvanitic 146, 151–2
 Asianic 51
 Aslian 301–2
 Assamese 264, 266
 Assyrian, Old 45, 51
 Athapaskan 15, 28, 29, 33, 98
 Atlantic languages 369, 377–80, 382, 385, 408
 Atsi/Zaiwa 314
 Australian languages 4, 20, 29, 64–133, 361, 430
 Austroasiatic 37, 40, 231–3, 239, 256, 260, 292,
 295, 320, 336
 Austronesian 9, 28, 30, 31, 32, 40, 41, 42, 134, 136,
 146, 151, 233, 256, 264, 292, 294–6, 303–4,
 317, 319, 321, 334, 336, 347, 403
 Awa Pit 415
 Azeri 221
 Baagandji 77–9, 85
 Baale 15, 16, 22, 359, 361–5, 425–6, 432
 Bahnaric 319
 Bahwana 174–6
 Bai 237, 317
 Baic 313, 318
 Balkanic 56
 Balti 240
 Baltic 45
 Baluchi 217
 Baniwa 169–70, 172–9, 181, 184, 186–7, 189–90
 Bantoid 367, 369, 374, 378, 385
 Bantu 22, 30, 37, 42, 146, 292, 358–411, 426, 429 n.
 Bare 173–5, 181
 Bari 377
 Barish 302
 Basque 10, 17, 87, 146, 152, 153, 221, 420, 431 n.
 Baure 176
 Be 296
 Beja 406
 Bendi 379
 Bengali 28
 Benue-Congo 3, 358, 366, 368–9, 371–4, 377–9,
 381–2, 387
 Bera 381
 Berber 11, 396 n.
 Bhele 381
 Bhutanese 240
 Bila 370, 381
 Bilin 406

- Birgid 397
Bissa 373
Blimaw 313
Bodic 236
Bodish 225
Bola 314
Bora 182, 184–8, 190, 428–30
Bora-Muinane 182, 184, 186–8, 191
Bora-Witoto 21, 168–9, 176, 182, 184, 186–90
Brao 281
Bulgarian 56, 150, 152, 400 n.11
Bunaban 90
Burmese 236–8, 241, 260, 264, 276, 286, 303,
306–7, 309–11, 314, 317, 319, 322–3
Bwamu 374
Bwe Karen 313

Campa 174
Cantonese 233, 242, 256, 264–6, 269, 273, 275,
278, 281, 302, 306, 322, 331, 335–6, 341–2,
344, 347–8, 350, 352, 353, 428, 432
Cape Dutch Creole 159
Carian 50
Carib 168, 171, 403
Caucasian languages 196
 Northwest 16
Celtic 45, 53
Chadic 349, 373, 381, 396 n., 405, 407–8
Chai 361–5
Cham 297, 304, 319–20
Chamic 297, 303, 319–20
Chamicuro 174
Chamorro 151
Chaozhou 231–2, 340, 345
Chaudangsi 234, 242
Chepang 239, 309 n.11, 315
Chibchan 98
Chibchan-Paezan 98
Chin 241, 309
Chinese 16, 18, 30–2, 227, 231–2, 234–5, 242, 245,
294–5, 303, 306, 310–11, 313, 315, 317–18,
321, 328, 330, 334, 336, 350, 426
 Ancient 317
 Classical 349
 dialects 225–6, 228, 330–4, 335
 Medieval 335, 349
 Middle 310, 313, 323, 328–30, 343, 344 n.3, 349
 Modern Standard 256, 261, 264, 268–9, 273–4,
277–8, 281
 Northern 330–4, 335–8, 341, 343, 347, 350
 Old 279, 294, 328–30, 344 n.3
Chu 231
Circassian 196
Comoe 379
Copper Island Aleut 159, 420, 429 n.
Cree 159
Cross-River 369, 371, 374, 378–9, 384, 386

Cushitic 3, 146, 394, 403, 406, 408, 429 n.
Czech 16

Dagbani 382
Dahalo 394
Dakarkari 381
Dan 374
Debri 397
Degema 371
Delugong 245
Didinga 425
Didinga-Murle 361–4
Didinga-Narim 362
Dilling 397
Diola 408 n.
Diola-Fogny 380
Djeragan 90
Djingulu 82–3
Dong 264–5, 280
Dongolawi 397, 400–1
Dongolawi-Kenuz 397, 399–401, 428–9
Dravidian 5, 10, 11, 29, 36, 37, 147, 235, 239, 349
Dutch 155, 159, 423
Dyirbal 71
Dyola 369
Dzamba 381

Ebrie 374
Edoid 367, 369, 371, 374, 381
English 3, 28, 32, 49, 66, 149–50, 152, 159–61, 207,
216, 234, 242–3, 258, 268 n., 276, 294,
321–2, 422–3, 427
 African 428
 Australian 425
 Basilectal Singapore 160
 Early Modern 423
 Hong Kong 428
 Irish 422, 424
 Modern 2, 423
 Old 2
 Singapore Colloquial 159–60
 South African 423–4
Eskimo-Aleut 98
Ethio-Semitic 406
Ethiopian languages 406
Etruscan 45, 51, 56
Ewe 371

Fadijja 397
Finnish 432
Flinders Island Language 95
French 159, 214, 258, 320, 415, 427
Frisian 423
Fulani 408 n.25
Fulfulde 380
Fuzhouese 331

- Gaelic, Irish 422
 Gan 231, 331–4
 Garhwali 243
 Gascon 146, 153
 Gaulish 45
 Gbe 379
 Geba 313
 Gelao 296
 Georgian 205, 209, 211
 German 3, 16, 65, 146, 293 n., 294, 423, 427
 Germanic 45, 49, 330
 West 423
 Gong 242
 Grebo 374
 Greek 3, 11, 15, 17, 23, 44, 45, 49, 51, 55–9, 62, 146,
 150–2, 196, 206, 209, 221–2, 310
 Classical 59
 Modern 59, 293 n.
 Pontus 195 n.
 Guahibo 168
 Guajiro 172, 174, 182
 Guanche 11
 Gumbaynggirr 69, 95
 Gur 368, 371, 374, 377–8, 381–2, 396
 Central 371
 Gurnu 78–9
 Gurung 239
 Gyalrong 236

 Hadza 358 n., 396
 Haida 98
 Hakka 231, 330, 335, 337, 340, 342, 350–1, 354
 Hamar 373
 Hangzhou 228
 Harakmbet 168
 Hattic 51, 54–5, 428
 Hausa 405, 407, 408 n.25
 Hayu 303
 Hebrew, Biblical 404 n.15
 Heiban 377–8
 Herero 381
 Himalayan, Central 239
 Himalayish 310–11
 Hindi 202, 243, 303
 Hittite 17, 45, 49, 51–5, 57–8, 180, 424, 428,
 430
 Middle 54
 Old 55
 Hixkaryana 403
 Hkauri 308
 Hmong 269, 295
 Hmong-Mien 37, 226–7, 230–3, 256, 259–60, 265,
 279, 281, 291, 296, 303, 315, 317, 321, 323,
 336, 342, 349
 Ho 267–8
 Ho Nte 296
 Hohôdene 174

 Hokkien 344
 Hopi 38
 Hruso 295
 Hui dialects 331–4
 Hungarian 10
 Hurrian 51–5

 Idomoid 367
 Igbirra 371
 Igbo 386
 Igboid 367, 369, 379, 382
 Ignaciano 174–6, 182
 Ijesha 371
 Ijoid 3, 367, 369, 371, 381–2, 384, 387
 Ilwana 146
 Inapari 174–5, 182
 Indic languages 234–5, 241, 321, 349
 Middle 56
 Indo-Aryan 10, 56, 147, 196, 235, 239–41, 267, 300
 n.2, 303
 Indo-European 4–7, 9, 13, 17, 19, 28–32, 36–8,
 41–2, 44–63, 170, 195–6, 235, 238, 292,
 294, 373, 430
 Indo-Iranian 45, 56–7
 Indonesian 284
 Ionic 58
 Ionic Greek 180
 Eastern 424
 Iranian languages 11, 21, 196, 199, 202, 209–10,
 212–13, 217, 415, 429, 431
 Irish, Old 45
 Italian 10, 146, 151, 425
 Australian 414, 425, 427
 Italic 45, 49, 51

 Jangali 241
 Japanese 3, 37, 152, 159, 234, 294, 311, 328, 336,
 349
 Javanese 155
 Javindo 155
 Jê 168
 Jin dialects 331
 Jinghpaw 238, 241
 Jingpho 306–9, 314, 322
 Jiwari 108, 116, 123, 125
 Jukun 381
 Jukunoid, Central 379

 Kabardian 196
 Kachin, *see* Jingpho
 Kachin-Nung 313
 Kadai, *see* Tai-Kadai
 Kadaru 397
 Kadu 358 n., 396 n.
 Kaiku 381
 Kainji 381
 Kalkatungu 65, 86–7

- Kam-Sui 296, 320 n.14
Kam-Tai 341
Kana 384, 386
Kanauri-Almora 244
Kannada 13, 147
Kanyara 107–8, 112
Karbi 244–5
Kardu 89, 98
Karen 225, 245, 314, 318, 319
Karenic 260, 303, 313–15, 318, 323
Kariera 108
Kartvelian 21, 195–6, 203, 211, 214
Karwan 90–1
Kasem 382
Kashmiri 303
Katang 281
Katuic 281
Kegboid 374
Kejia 331
Kelantan Chinese 302–3
Kemantney 408
Kenuz 397, 400, 401
Kenuz-Dongolawi, *see* Dongolawi-Kenuz
Keresan 39
Kham 239
Khamti 220
Khasi 302
Khmer 260, 281, 295, 303, 308, 317, 321, 323, 347
Khmeric 281
Khmu 261, 267, 305
Khmuic 302, 321
Khoisan 8, 358, 387–9, 393–6, 406 n. 19
Khoti 15, 360, 363, 365, 372, 426
Kimant 408
Kiranti 239–40
Kmhmu, *see* Khmu
Kohumono 379
Koma 358 n.
Komo 381
Kongo-Kordofanian, *see* Niger-Congo
Kordofan 396 n., 397
Kordofanian 358 n., 366, 376–8
Korean 336, 349
Krongo 396 n.
Kru 372, 374, 381, 385
Kugu-Muminh 80–1, 83, 430
Kugu-Nganhcara, *see* Kugu-Muminh
Kuki-Chin 244
Kuki-Chin-Naga 315
Kuku 376–7
Kuliak 396
Kunama 406
Kuot 156
Kurdish 199, 212, 218
Kurmanji Kurdish 21, 196, 198–9, 201–10, 212–14, 217–18, 415, 424, 429
Kurrama 106 n., 108, 109, 114–18, 129
Kusunda 239
Kwa 3, 358, 367–9, 371–2, 374, 376, 378–9, 382, 385, 387, 396, 407
Kwami 408 n.
Laal 358 n.
Ladakhi 240
Lahu 241, 269, 276, 286, 302, 304, 306–7, 309 n.10, 310, 322, 347, 353
Lakkia 296
Lamet 304
Lao 255, 260–1, 264–6, 269–73, 275–7, 280–1, 302–3, 321, 353
Latin 49, 52, 56, 293 n., 294
Latino-Faliscan 49
Laven 320 n.14
Laz 21, 196, 198–217, 222, 421, 424, 428–9, 431–2
Lele 408 n.25
Lepcha 241, 302
Lhasa 240
Lhasa Tibetan 309
Li 296
Ligbi 382
Limba 369
Lisu 239, 310
Lokono 170, 172, 174
Lolo 237
Lolo-Burmese 225, 237, 295, 306, 309–11, 313–15, 318, 322
Loloish 292, 295, 307 n., 309–10, 313–14, 323
Lotha Naga 312
Luo 8, 382
Luquan 239
Lusi 146, 149, 150
Luvian 49–58, 180
Lycian 50–3, 59
Lydian 50, 52
Ma'a 429 n.
Macedonian 56, 146
Madak 156, 160
Magar 239
Magori 138
Mahasi 397
Maipure 170
Maisin 138, 146, 154
Maithili 240
Makedonski, *see* Macedonian
Makhuwa 360–1
Makú 10, 168, 170, 191
Malay 28, 297, 302–3, 320, 322
Malayo-Polynesian 31
Maltese 151
Mampruli 382

- Manchu 230, 350
 Mandarin 17, 159, 228, 230, 232–4, 241, 265,
 306, 310, 330–1, 335–45, 347–8, 350–1,
 429
 Beijing 343, 349
 Changsha 342
 Modern 231, 337
 Northern 242, 429
 Northwestern 335
 Southwestern 267, 277–8
 Taiwanese 17, 159, 235, 242, 349–50, 353
 Yunnanese 302
 Mande 373–4, 382, 385, 396
 Mantharta 107–8, 112
 Mao 358 n.
 Marathi 13
 Margi 408
 Mar(r)ngu 90, 107
 Martuthunira 108–9, 114–18, 129–30, 425
 Maru/Langsu 314
 Mayali 74
 Mayan languages 415
 Mbabaram 1
 Mbatto 374
 Mbay 376
 Media Lengua 156
 Meidob 397
 Meithei 241
 Melanesian 158
 Meso-American 12
 Miao-Yao, *see* Hmong-Mien
 Michif 19, 159
 Mien 302, 321
 Mikir 244–5
 Milyan 50
 Min 231, 330–4, 337, 345
 Northeastern 331
 Southern 17, 159, 233, 235 n.7, 330–1, 337,
 340–2, 344–7, 350, 353–4
 Taiwanese Southern 273, 338–9, 342, 344–50,
 429
 Min-Yue 231
 Mina 408
 Mindi 81–2
 Mingrelian 211, 212
 Minkinan 91
 Mixe-Zoquean 219
 Mizo 241
 Moluccan 403
 Mon 225, 237–8, 303, 317–19
 Mon-Khmer 225, 227, 237–8, 256, 258, 261, 267,
 279, 287, 295–6, 301–3, 305, 308, 310, 312,
 315, 317–19, 321–3, 336
 Eastern 256, 281, 283
 Northern 260
 Monpa 240
 Moxo 170, 174
 Muinane 187
 Muinane Witoto 184
 Muk-Thang 85
 Mulao 261, 264–5
 Munda 37, 241, 295
 Mundaic 29
 Muong 281, 320 n.14
 Murle 365
 Mursi 361
 Nadëb 191
 Nakh-Daghestanian 206
 Nama 406
 Nan-Chao 237
 Nanda, *see* Nhanta
 Narrinyeric 90
 Nasu 310
 Navajo 28
 Naxi 313–14
 Nepali 235, 303
 Newari 235, 239–40, 303
 Ngae 281
 Ngaliwuru 82
 Ngambay-Mundu 376
 Ngarla 109, 119, 123–5, 127
 Ngarluma 108, 129
 Ngayarda, *see* Ngayarta
 Ngayarta 107, 111, 112
 Ngiyambaa 74
 Ngumbin 90
 Nhanta 89, 108, 114, 117
 Niger-Congo 8, 22, 358–9, 365–6, 368–9, 371,
 373–4, 376–82, 385, 387–8, 393, 395–6,
 405, 407–8, 414, 426, 430
 Niger-Kordofanian, *see* Niger-Congo
 Nilo-Saharan 8, 22, 358–9, 361, 373–4, 376, 381–2,
 387, 393, 396, 406
 Nilotic 361, 373, 376–7, 382, 384, 396
 Nobiin 397, 399–401
 Nubian 7 n., 373, 397–400, 409
 Numic 38, 39
 Numidian, East- 11
 Nungish 313
 Nupoid 367
 Nyabwa 374
 Nyaeun 320 n.14
 Nyamal 109, 123–5, 127
 Nyangumarta 107, 123–4
 Nyiyaparli 116, 123–4
 Nyulnyulan 90
 Obolo 371
 Ocaina 170, 182, 185
 Oceanic 20, 31, 41, 134–66, 242, 433
 Ogoni 374, 379
 Omotic 358 n., 373
 Ong-Bê 336

- Ongota 358 n.
Orma 146
Osco-Umbrian 49
Oti-Volta 378
- Pacoh 286
Palaic 49–50, 52, 54
Palaung-Wa 302
Palaungic 304
Pali 237–8, 258
Palikur 172–5, 182, 191
Paman, Northern 65
Pamphylian 59
Pano 168
Panyjiima 116–18, 129
Papuan 20, 40, 92, 136, 138, 143–4, 146, 149,
142–54, 156, 175, 242, 361, 377, 384–5
Pareci 173–6, 182
Pateng 296
Payungu 108, 119, 123–4
Pazar 214
Pero 408 n.25
Persian 198, 200–2, 205–6, 209, 221, 360, 421,
431
 Middle 202
Phan Rang Cham 146, 152
Phrygian 45, 50, 55–7
Piapoco 173–5, 181, 184, 187, 190–1
Picene, South 49
Pilbara languages 105–32
Pinghua 331
Pipil 220
Piro 173
Pisidian 50
Platoid 379, 381
Polynesian 31
Punu 321
Purduna 108, 114–16
Pütönghuà 233 n., 342–3, 348
- Qiang 235, 295, 305
Qiangic 225
Quanzhou 338, 340
Quechua 155–6
- Rade 320
Rai 240
Raji 241
Resígaro 2, 10, 17, 19, 21, 169–70, 172–6, 182–91,
214, 425–6, 428–31
Rhaeto-Romance 146
Riang 304
Romance languages 53, 221, 292, 330
Romanian 56
 Meglenite 152, 400 n.11
Rongpo 243
Russian 18, 159, 293 n., 420, 429 n., 432
- Sabellic 49
Saharan 361
Samsao 321
Sandawe 358 n.
Sani 310
Sanskrit 3, 258, 310
Sara 376
Sauris German 146
Semitic 2, 30, 45, 51, 52, 195, 199, 209, 292, 414
Serbo-Croatian 150
Serer 380
Sgaw Karen 245
Shabo 358 n.
Shan 280, 302, 307
Shanghainese 233
Shantou 340, 342
Shiriwe 191
Siamese 303, 308, 312, 323
Sichuan 295
Sidetic 50
Sika 403
Sinitic 5, 12, 18, 22, 225–6, 228–34, 256, 259–61,
264–6, 268, 273, 275, 277, 279–80, 284,
287, 328–54, 424, 427
Sino-Tai 264
Sino-Tibetan 21, 22, 31, 37, 225–46, 256, 291–2,
295, 310, 313–15, 347
Sinospheric languages 21, 235, 294, 296, 303–4,
306, 317, 336, 342, 347
Slavic 5, 45, 56
Songhai 396
Spanish 32, 151–3, 156, 220–1, 345 n.6, 420, 431 n.
Standard Average European 10, 12, 13, 153, 214
Sudanic 361, 373, 376, 381
Sui 320 n.14
Supyire 382
Surma 396
Surmic 22, 359, 361–3, 365, 373
Suzhou 342
Suzhouese 330
Svan 211
Swahili 15, 22, 359–61, 382, 385, 387, 405, 406 n.18,
426, 428
Swedish 313
- Tagalog 151–2
Tahitian 28
Tai 31, 37, 225–7, 230–3, 237, 241, 255, 258–62,
264–7, 279–81, 284, 287, 291–2, 295–6,
302–3, 312, 315, 317–18, 321, 323, 335–6, 342
Tai-Kadai 256, 259, 262, 264–5, 296, 321, 336, 349
Taino 172
Taiwanese 233–73, 344–5
Tajik 212, 213, 217
Takia 20, 138–54, 160–1, 426–7, 429, 432–3
Takis 39, 60
Taliang 281

- Tamang 310, 312, 315
 Tamang Risiangku 306, 311–12
 Tamangic 315
 Tang koine 231
 Tangale 373
 Tangut 243–4
 Tano-Congo 372
 Tanoan 39
 Tarascan 98
 Targari 90
 Tariana 11, 13, 15, 17, 19, 146, 152, 167–94, 222,
 420–1, 426–8, 432
 Temne 382
 Tennet 364
 Teréna 174, 176
 Thai 31, 33, 220, 241, 262, 266, 302–3, 321–2, 347,
 351
 Thalanyji 108, 110, 121, 124
 Tharrkari 108, 114–18, 123–4
 Tibetan 236, 239–41, 293 n., 304–5, 309, 311,
 427
 Tibeto-Burman 3, 21–2, 30, 225–6, 234–44,
 258–60, 267, 276, 286, 292, 294–5, 300 n.
 2, 302–5, 307–10, 313–15, 318–19, 322, 330,
 336, 341, 347, 349, 426, 430–1
 Tibeto-Karen 225, 318
 Tiddim Chin 315
 Tirma 361–5
 Tirma-Chai 15, 16, 361, 363, 365, 425–6, 432
 Tiv 381
 Tiwi 85
 Tocharian 45, 49
 Togo Remnant 377 n., 377–8, 385–6
 Tok Pisin 147, 150–1, 160–1
 Tour 374
 Tsat 319–20
 Tswana 370, 395
 Tubatulabal 39
 Tucano 10, 13, 15, 17, 19, 29, 146, 168–9, 176–81,
 186, 188–91, 222, 420–1, 426–8, 432
 Tunceli 208
 Tunen 370–1
 Tupí 168, 171
 Turkana 384
 Turkic 195–6, 208, 209, 212–13, 217, 221, 234
 Turkic-Iranian 11
 Turkish 15, 16, 21, 146, 195–224, 360, 421, 424,
 428–31
 Ubangi 374, 381
 Ugaritic 51
 Ukrainian 158
 Umbindhamu 85
 UMBundu 374–6
 Umpila 65
 Uralic 30
 Uralo-Altaic 3
 Urartean 51
 Urdu 13, 147
 Uto-Aztecan 29, 30, 32, 33, 38–40, 42, 60, 220
 Utsat, *see* Tsat
 Uzbek 205, 208, 212
 Vagala 382
 Vietic 281
 Vietnamese 146, 227, 235, 245, 260, 261 n., 264,
 276, 281, 291, 295, 303, 315, 317–18, 321,
 323, 328, 336, 351
 Volta-Congo 368, 371–2, 376
 Voltaic 396
 Wa 260
 Wambaya 82–3, 90, 91
 Wapishana 174–5
 Wardaman 71, 73
 Warekena 173, 175, 181
 Warienga 90
 Warlmanpa 74
 Warlpiri 74
 Warumungu 83, 85
 Waskia 20, 138–40, 142, 138–54, 426–7, 429, 431,
 433
 Watjarri 89, 98, 107–8, 114, 120, 123–4
 Waurá 172–4, 176
 Welsh 150
 Western Desert language(s) 74, 85, 107, 118, 120,
 123–4, 129
 Wik 65, 79
 Wik-Me'nh 65
 Wik-Muminh, *see* Kugu-Muminh
 Wik-Mungknh 80–1
 Wik-Ngathrr 65, 80–1, 83
 Witoto 170, 182, 185–6, 190–1
 Witoto-Ocaina 182
 Wororan 90
 Wu 228, 231, 233, 330–4, 340, 342, 348, 350
 Wutun 242, 427
 Xiamen 337–8, 344–5
 Xiang 231–2, 331–4, 342–3
 Changsha 343, 353
 !Xun 395 n.1
 Yaaku 403
 Yakoma 374
 Yalarnnga 65, 86
 Yavitero 174
 Yawalapiti 174, 176
 Yi 239
 Yilan 340
 Yindjibarndi 106 n., 108, 114–18, 129
 Yingkarta 107–8, 114, 116–17, 120–1, 123–4
 Yinhawangka 129
 Yitha-Yitha 95

- Yolngu 97
Yongxiang 232
Yoruba 371, 386, 407
Yoruboid 367
Yucuna 174–6, 187, 191
Yue 233, 252, 254, 330–4, 340, 348, 350
Zazaki 21, 196, 199, 201–14, 217, 415, 424, 429
Zhangzhou 338, 345
Zhuang 255–6, 265, 269–70, 273, 275–7, 280–1,
321
Zo 237
Zuñi 39

Subject Index

- ablative 55, 68, 126, 129, 403, 406, 430
- absolutive 51, 68, 120, 123, 125
- accusative 20, 55, 83, 92, 110, 121, 123–30, 136, 335, 355
- active 129, 171
- affricate 238
- agriculturalism 9, 10, 14, 19, 32–42, 136, 138, 191
- anaphora 86
- aspect 68, 140–2, 171, 214, 259, 301, 335, 336, 384, 422–3, 427–9, 430
- aspiration 225 n.1, 310
- auxiliary 82–3, 167, 364, 415
- benefactive 243, 276, 336, 350, 384–5, 395
- bilingualism 14, 15, 41, 50, 61, 146–7, 149, 152, 169, 210, 215, 239, 240, 241, 365, 376, 420
- borrowability, hierarchy of 14, 19, 23, 412–34
- borrowing:
 - attitudes to 15–16, 169, 176–7, 180
 - constraints on 412–34
 - lexical 2, 29, 139, 145, 150, 152, 169, 177, 182–4, 212–14, 280, 283, 291, 298, 317, 321, 360–1, 363, 376, 396, 401, 409, 413–21, 424–8
 - morphological 18, 21, 130, 185–8, 190, 200, 215, 338, 360, 365, 388, 400, 413, 415–20, 424
 - phonological 2, 184–5, 283, 360, 371–2, 400, 413, 417, 420, 425–6
 - semantic 139
 - syntactic 2, 17, 18, 139, 145, 182, 184, 285, 341, 360, 413
- calquing 17, 18, 109, 129, 145, 169, 180–1, 185, 200, 204, 205, 243, 265, 280, 301, 342, 349, 353, 363, 402, 413
- case-marking system 20, 21, 68, 106, 120–30, 214–16, 222, 421
 - absolutive/ergative 68
 - nominative/accusative 68, 105, 112, 123, 168, 169
 - split ergative 54, 105, 112, 120–1, 124, 172, 175
 - tripartite 105, 110, 121, 123–5
- causative 18, 56, 199, 225 n.1, 261–2, 267, 335, 336, 350–1, 395
 - periphrastic 17, 199, 262, 267, 384
- change, system-altering 16, 60
- change, system-preserving 16, 60
- classifier 17, 69, 83, 168, 172–3, 175–6, 178, 182, 184, 185–9, 227, 230, 233, 265–6, 269, 301, 335, 336, 341–2, 429
- clause linkage 21, 143, 196, 204–5, 219–20, 433
- clitic 66, 73, 75, 77, 83, 116
- code-switching 199, 221–2, 413, 421, 428
 - see also* language mixing
- comitative 68, 384
- community, externally open/closed 14, 110, 131, 155–9, 421
- community, internally tightly-knit/loosely knit 14, 110, 131, 155–9, 421
- comparative 2, 205–6, 260, 342, 403–6, 410
- comparative method 7 n., 20, 44–63, 113, 262, 328–30, 335, 353–4, 388, 394, 399
- complementizers 22, 127–8, 200–2, 235, 328, 336, 341, 343, 347–50, 352, 353–4, 423, 429, 431
- connective 18, 55, 384
- constituent order 18, 136, 179, 199, 218, 220–1, 259, 364–5, 376, 431–3
 - see also* word order
- core vocabulary 7, 83, 89–90, 111, 169, 175, 176, 182, 293
 - see also* lexicostatistics
- cosubordination 144, 154
- coverb 71, 73–4, 82
- creolization 30, 152, 158–61, 234
- dativ 68, 115, 116, 119, 120, 121, 125–7, 130, 188, 240, 261 n., 335, 384
- deictic 55, 69
- demonstrative 69, 93, 123–4, 172, 175, 178, 186, 189, 380, 382, 400, 428, 432
- dependent-marking 19, 65, 75, 364
- development, convergent/parallel 3–4, 9, 66, 75, 98, 243, 314
- devoicing 52, 323
 - see also* voicing
- diglossia 15, 365, 421, 433
- diminutive 328, 337, 343, 344–6, 353
- ditaxia 18, 341
- dominance 11, 13, 15
- double patient construction 22, 352–3
- emblematicity 17, 18, 23, 110, 146, 152, 157–8, 267–8, 342, 353, 386, 421
- emphatic 180, 261
- enclitic 54–5, 75, 77, 80–3, 91, 140–4, 154, 171–2, 201, 203, 204, 205, 207–8, 342, 429, 433
- equilibrium 9–10, 11, 13, 19, 20, 23, 48, 49, 61–2, 84, 131, 138–9, 153–4, 157, 191, 259, 334
 - homeostatic 48, 52, 55, 62–3, 64–5
 - see also* punctuated equilibrium model
- ergative 20, 51, 53, 54, 68, 69, 70, 83, 92, 93, 95 n., 97 n., 116–21, 123–30, 136, 214, 217, 240, 395

- family tree model 4–7, 9, 20, 21, 28, 57, 59, 62, 65,
153–4, 226, 332–5, 353–4, 394, 409
inadequacy of 1, 4–5, 22, 28, 59, 90, 246, 292,
330, 353, 409
fortition 53, 105, 114, 115
fricative 67, 238, 264
- geminate 52–3
gender 2, 54–5, 171–2, 178, 186, 217, 301, 321, 331, 381
genitive 58, 59, 68, 115, 125–6, 204, 351
glide 115
glottal stop 69, 117, 185, 280 n., 307 n., 309, 425
glottochronology 7, 29, 31, 293
grammaticalization 88, 171, 176, 186, 235, 243,
244, 275, 328, 331, 336, 343, 345, 347–9,
380, 402–4, 406 n.18, 407–8, 427, 430
Grassmann's Law 3, 310
Grimm's Law 13, 302 n.
- head-marking 19, 65, 68, 86, 171, 214, 300 n.2,
364, 385–6
hybridization 5, 330, 337, 341–2, 353–4
- imperative 69, 128
imperfective 17, 58, 180, 424
inchoative 128, 130
inner dynamic 3–4, 66, 74–5, 83
instrumental 55, 68, 118, 188, 189, 430
interrogative 69, 93, 124, 189, 233, 239, 278 n.,
340–2, 353, 400, 429
intransitive subject 120, 124
intransitive verb 54, 171, 335, 352, 354
isogloss 4, 6, 10, 54, 64, 87, 92, 93, 112, 209, 330,
334, 387
isomorphism 10, 13, 17, 22, 156, 179, 217–18, 244,
300, 361
- language:
attitudes 13, 15–16
see also borrowing, attitudes to
change, rate of 7, 13, 15, 23, 29, 55, 60–2, 84, 293
death/extinction 11, 62, 65, 152, 421
mixing 22, 177, 180
see also code-switching
shift 28, 31, 40–1, 138, 140, 157–61, 240–2, 382,
419–20
lenition 53, 70, 105, 114–15
lexicostatistics 7, 89–98, 110–12, 360, 397–9
lingua franca 14, 147, 158–9, 241
loan word 15–16, 45, 49, 51, 54, 180, 217, 227, 232,
241, 321–2, 401, 413–15, 426
loan-translations 18, 145
locative 68, 88, 116, 119, 130, 177, 189, 210–11, 240,
384, 403
- marriage, language and 14, 15, 29, 40, 109, 152,
159, 177, 239, 420
- metatypy 5, 16, 18, 20, 139, 145–61, 218, 242–5,
303, 337, 341–2, 350, 353–4, 402, 404, 408,
410, 417–18
migration 22, 49, 51, 56, 153, 168–70, 171 n., 191,
225–46
modality 259, 384
monolingualism 19, 61, 181, 303
monomorphemic verb 71, 73–4, 347
morphology:
derivational 134, 225, 261, 295, 360, 416
inflectional 134, 294, 360–1, 416
morphonemic alternations 106, 116, 118–20, 130,
362, 368
multilingualism 14, 15, 19, 20, 29, 109, 146,
169–70, 176, 177, 181, 191, 217, 222, 243,
291, 388, 420–1, 433
- nasalization 2, 337, 345 n.6, 375, 414
nominalization 125–6, 172–3, 210–12, 384
nominative 55, 121, 123–7, 352
noun class 2, 3, 8, 22, 65, 69, 83, 87, 92, 367,
377–82, 387, 395, 430
- oblique case 169, 179, 185, 188, 189, 428
official language 198, 240, 338, 342, 343, 353
- Pama-Nyungan 'idea' 20, 29, 64, 89–98
particle 54–5, 186, 189, 205–6, 212, 235, 239, 301,
306, 347, 363
passive 128–9, 199, 353–4, 395
adversative 18, 328, 343, 350–1, 354
derivational 128–30
inflectional 128–9
perfective 129, 213, 331, 347
phonological reduction 75, 80
polylectalism 146–7, 152, 155–6
possession:
alienable 136, 173, 242, 430
inalienable 136, 143, 173, 185, 242, 343, 351–2,
353–4
postposition 142–4, 199, 203–4, 212, 219–21, 241,
260, 335, 376, 400, 406, 428, 432
preposition 142–3, 199, 220–1, 260, 386, 417,
432
prestige, language and 6, 15, 337, 342–3, 350, 353,
365, 399, 421
pronominal system 83, 95, 110, 123–4, 300, 351–2
pronoun 68, 69, 95, 96, 116, 121, 123, 144, 169,
171–2, 176, 185–6, 190, 230, 242, 244, 401,
408, 428
anaphoric 54
bound 4, 20, 66, 68, 69, 71, 73, 75–83, 86, 87,
97 n., 123
free 20, 68, 75, 77–8, 80–1, 180, 243–5
number-segmentable 92, 95, 98 n.
relative 211, 234–5, 242
prosodic system 292, 295, 303, 310, 315, 321

- proto-language 4, 7, 8, 9, 28, 30, 31–2, 38, 45, 49,
 53, 57, 84, 85, 86, 96, 110, 171, 293
 prototype 49, 62, 233, 387, 401
 punctuated equilibrium model 1, 19, 44, 60–1, 65
 punctuation 9–10, 11, 13, 19, 20, 23, 27, 30, 36, 42,
 48, 49, 61–3, 65, 84, 85–6, 131, 139, 153, 191
 purposive 68, 70, 125, 188

 register 18, 238, 311, 338, 341–2
 relational noun 12, 143–4
 relative clause 21, 129, 211–12, 221, 234, 241, 260,
 264–5, 336, 341–2, 353, 432–3
 relic language 10, 124
 relic linguistic area 20, 64, 86
 religion, language and 14, 15, 51, 59, 399

 similarity, typological 113, 245, 367
 spirantization 52–3, 59
 stratification 330, 337–41
 stress 54, 56, 65–6, 69
 subgrouping 21, 83–99, 110, 115, 127, 131, 161,
 167–91, 226 n., 234, 245–6, 372, 388
 subordinate clause 17, 69, 86, 109, 125–7, 202,
 210–12, 217, 219, 331, 423
 switch-reference 12, 65, 69, 87, 177, 179–80
 syllable structure 65, 67, 230, 292, 303–6, 317,
 319–20, 426

 tense 68, 73, 77–80, 82–3, 128, 140–2, 167, 179, 214,
 240, 259, 335, 361, 384, 418, 422–3, 430

 tone 2, 22, 168, 185, 227, 230–3, 235–6, 238, 242,
 243, 245, 291–323, 330–1, 335, 343–4, 426
 tone-proneness 12–13, 302, 304
 tonogenesis 13, 291, 306 n., 314, 317, 321–3
 typological poise 256, 283, 285–7

 universal properties/tendencies 1, 5, 168, 200,
 292

 verb:
 existential 3, 347, 350
 inflection 80, 125, 127, 128, 415
 morphology 19, 110, 152, 188
 serial 3, 8, 22, 144, 217, 259, 301, 367, 382–8
 simple 71, 73–4, 82–3, 262
 system 57, 81
 valency 68, 150, 169, 173, 214, 352, 384–5
 vernacular 15, 31, 337, 343, 344
 vowel:
 assimilation 71
 harmony 3, 8, 22, 363, 367–74, 387–8, 395, 426
 height 238
 length 53, 54, 302, 305, 309, 312, 321
 nasalized 3, 8, 56, 367–8, 371, 374–7, 387–8
 system 22, 54, 67, 158, 230, 238, 362–3, 368–73

 wave metaphor/theory 4–5, 59, 330
 word order 12, 220, 225, 233, 238, 318, 331, 335,
 395–6
 see also constituent order