LINGUISTIC AREAS

Convergence in Historical and Typological Perspective

Edited by
Yaron Matras, April McMahon and Nigel Vincent

Linguistic Areas

Convergence in Historical and Typological Perspective

Edited by

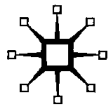
Yaron Matras
University of Manchester

April McMahon
University of Edinburgh

and

Nigel Vincent
University of Manchester

palgrave
macmillan



First published 2006 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N.Y. 10010
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN-13: 978–1–4039–9657–2
ISBN-10: 1–4039–9657–1

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

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

Library of Congress Cataloging-in-Publication Data
Linguistic areas : convergence in historical and typological perspective / edited by Yaron Matras, April McMahon, Nigel Vincent.
p. cm.
Includes bibliographical references and index.
ISBN 1–4039–9657–1 (cloth)
1. Areal linguistics. 2. Historical linguistics. 3. Typology (Linguistics) 4. Languages in contact. I. Matras, Yaron, 1963– II. McMahon, April M. S. III. Vincent, Nigel.
P130.L56 2006
417—dc22 2005056583

10 9 8 7 6 5 4 3 2 1
15 14 13 12 11 10 09 08 07 06

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham and Eastbourne

List of Abbreviations

1, 2, 3	first, second, third person
ABL	ablative
ABS	absolutive
ACC	accusative
ACTR	actor
AD	addressee
ADJ	adjective
AL	alienable
ALL	allative
ANA	anaphor
AOR	aorist
APPL	applicative
ARTCL	article
ASS	associative
AUGM	augmentative
AUX	auxiliary
BEN	benefactive
C	common
CAUS	causative
CLASS	classifier
CNC	concord
COMP	complementizer
COND	conditional
CONN	clause connector
CONV	converb
CONV:MAN	converb of manner
COP	copula
CSY	Central Siberian Yupik
CV	converb
D	dual
DAT	dative
DECL	declarative
DEI	deictic
DIM	diminutive
DIR	directional
DL	dual
DS	different subject
DSBJ	deposed subjectn

E	exclusive
EMPH	emphatic
EVID	evidential
F	feminine
FEM	feminine
FIN	finite
FUT	future
GEN	genitive
GIV	given
ILLAT	illative
IMP	imperative
IMPV	imperfective
INCH	inchoative
INF, INFIN	infinitive
INS, INSTR	instrumental
INTR	intransitive
IPFV	imperfective
IRR	irrealis
LOC	locative
M	masculine
MPO	multipurpose oblique
NEG	negative
NEGPOT	negative potential
NFUT	non-future
NM	noun marker
NOM	nominative
NP	noun phrase
OBJ	object
OBL	obligation
OBLQ	oblique
P, PL	plural
P/F	present-future
PAS, PASS	passive
PAST	past tense
PAT	patient
PERF	perfect
PL	plural
POSS	possessive
PP	postposition
PPL	participle
PQ	polar question
PRART	prepositional article
PRED	predicative
PRES	present

PRF	perfective
PRO	pronoun
PROB	probabilative
PROP	proper name
PROSEC	prosecutive
PRTCPL	participle
PST	past
PSYE	Proto-Siberian Yupik Eskimo
PURP	purposive
Q	question
QUOT	quotative
RECIPR	reciprocal
RED	reduplication
RED, REDUP	reduplication
REDPL	reduplication
REF	referential
REL	relative
REL	relativizer
RELV	relevance
REP	repetitive
RFLXV	reflexive
S, SG	singular
SBEN	self-benefactive
SBJ	subject
SF	stem-formant
SP	speaker
STAT	stative
SUBJ	subject
SUBJ	subjunctive
SUBORD	subordinator
TAG	tag question
TEMP	temporal
TR	transitive
TRANS	transitive
UNACCOMPL	unaccomplished
VOL	volition
XXVI	nominal class XXVI

Notes on the Contributors

Gregory D. S. Anderson is a Research Associate at MPI-EVA/Leipzig, Germany and the University of Oregon Department of Linguistics, Eugene, USA. He is the author of grammars on Xakas and Tuvan, two Tuvan–English dictionaries, and a volume on auxiliary verb constructions in Altai-Sayan Turkic languages. Forthcoming volumes are on language contact in South-Central Siberia, and a general–typological study of auxiliary verb constructions. Research interests include languages of Siberia, Munda and Austroasiatic languages, Burushaski and Eleme (Nigeria).

Peter Bakker is a Lecturer in the Department of Linguistics at Aarhus University, Denmark. He is the author of monographs on twins languages (*Autonomous Languages*, 1987) and the genesis of Michif (*A Language of Our Own*, 1997), co-editor of two books on mixed languages (*Mixed Languages*, with M. Mous, 1994, and *The Mixed Language Debate*, with Y. Matras, 2003). He also compiled a linguistic bibliography of Romani with Y. Matras (2003). His main research interests are language contact, language genesis and language change.

Walter Bisang is Professor of General and Comparative Linguistics at the University of Mainz, Germany. Since 1999 he has been the spokesman for a Collaborative Research Centre on 'Cultural and linguistic contacts: Processes of change in Northeast Africa and Western Asia' financed by the German Science Foundation. He is the author of *Hmong-Texte, eine Auswahl mit Interlinearübersetzung* (Zürich, 1988) and *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer* (Gunter Narr, 1992) and numerous articles. His main interests are language typology and language universals, formal versus functional linguistics, grammaticalization, and language contact–areal typology.

Claire Bowern is Assistant Professor in the Department of Linguistics at Rice University, Honston, Texas, USA. She is the editor (with Harold Koch) of *Australian Languages: Classification and the Comparative Method*. Her main research interests involve historical syntax and comparative reconstruction of Australian languages.

Lyle Campbell is Professor of Linguistics and Director of the Center for American Indian Languages at the University of Utah, Salt Lake City, Utah, USA. He is the author of fifteen books and 160 articles. He has twice won

the Linguistic Society of America's prestigious Leonard Bloomfield Book Award, first for jointly with Alice Harris *Historical Syntax in Cross-linguistics Perspective* (1995, Cambridge University Press), then for *American Indian Languages: The Historical Linguistics of Native America* (1997, Oxford University Press). His research specializations are Native American languages, endangered languages, historical linguistics, typology, sociolinguistics, linguistic anthropology, Uralic and Hispanic linguistics.

Tom Güldemann is a Research Associate in the Linguistics Section of the Max Planck Institute for Evolutionary Anthropology, Leipzig, and teaches at the Department of African Studies of the University of Leipzig, Germany. His main research interests are typology, historical linguistics, and the study of Khoisan and Bantu languages.

Lars Johanson is Professor of Turcology in the Seminar für Orientkunde at the Johannes Gutenberg-Universität in Mainz, Germany. He is the author of *Aspekt im Türkischen* (1971), *Alttürkisch als 'dissimilierende Sprache'* (1979), *Linguistische Beiträge zur Gesamtturkologie* (1991), *Strukturelle Faktoren in türkischen Sprachkontakten* (1992), *Discoveries on the Turkic Linguistic Map* (2001), *Structural Factors in Turkic Language Contacts* (2002), and other books. His main research interests are general and comparative Turkic studies, historical linguistics, code-copying, and aspectology.

Maria Koptjevskaja-Tamm is Professor in the Department of General Linguistics at Stockholm University, Sweden. She is the author of *Nominalizations* (1993, Routledge) and co-editor (with Östen Dahl) of *Circum-Baltic Languages* (2001, Amsterdam: Benjamins, 2 vols) and contributed to the *World Atlas of Language Structures* (2005, Oxford University Press). Her research interests are various morpho-syntactic phenomena (including nominalizations, possessive constructions, finiteness) in cross-linguistic and typological perspectives, and cross-linguistic research on lexical semantics, areal typology and traditional historical linguistics.

Yaron Matras is Professor of Linguistics in the School of Languages, Lingustics and Cultures at the University of Manchester, UK. His books include *Romani: A Linguistic Introduction* (2002), *The Mixed Language Debate* (co-edited with Peter Bakker (2003)), and *Markedness and Language Change* (with Viktor Elšík (2006)).

April McMahon is Forbes Professor of English Language at the University of Edinburgh, UK. She worked previously in the Department of Linguistics at the University of Cambridge and held a Chair in English Language and Linguistics at the University of Sheffield, both UK. Her research interests involve the interaction between phonological theory and historical evidence,

as well as issues of language comparison and classification. Her books include *Understanding Language Change* (Cambridge University Press, 1994), *Lexical Phonology and the History of English* (Cambridge University Press, 2000), *Change, Chance, and Optimality* (Oxford University Press, 2000), and *Language Classification by Numbers* (with Robert McMahon, Oxford University Press forthcoming 2005). She was formerly the President of the Linguistics Association of Great Britain.

Robert McMahon is a clinical molecular geneticist at the Western General Hospital in Edinburgh, UK. His work involves tracing inherited conditions through families, and in particular has researched and provided genetic services for cystic fibrosis, muscular dystrophy, inherited cancer and osteogenesis imperfecta (brittle bone disease). He has published a range of articles in professional and scientific journals, and maintains a research interest in issues of human genetics and evolution, and their relationship with language. He is co-author of *Language Classification by Numbers* (forthcoming, OUP).

Simon Musgrave is a post-doctoral Fellow in the Department of Cultures and Linguistics at Monash University, Victoria, Australia. His research interests include languages of Maluku, Indonesia; the syntax and typology of Austronesian languages; non-derivational theories of syntax; and computational tools for linguists. He is co-editor (with Peter Austin) of the volume *Voice and Grammatical Relations in Austronesian Languages* (forthcoming).

Thomas Stolz is Professor of Linguistics at the University of Bremen, Germany. He is the author of *Gibt es das kreolische Sprachwandelmodell?* (1986), *Sekundäre Flexionsbildung* (1992) and *Die Numeralklassifikation im Klassischen Aztekischen* (2001). His research interests are language contact, areal linguistics, typology, language change, morphology, grammaticalization, and the history of linguistic thought.

Nigel Vincent is Associate Dean and Mont Follick Professor of Comparative Philology in the Department of Linguistics at the University of Manchester, UK. He has published Widely a Romance languages and has recently translated into English Tore Janson's book *A National History of Latin* (2004). He is a former president of the Philological Society and for ten years was editor of the *Journal of Linguistics*.

Introduction

April McMahon, Yaron Matras and Nigel Vincent

Editing a volume on linguistic areas is both a fascinating and a worrisome undertaking, for essentially the same reason: 'linguistic area' is a rather ill-defined and amorphous notion, for which 'The number of definitions is almost coextensive with the number of linguists working in the field' (Stolz, this volume, p. 33). Working with a notion which is yet to be defined conclusively can be exhilarating, since by definition there must be work to be done; but it can also be frustrating, since it can sometimes feel as though metatheoretical discussions are impeding progress on the all too necessary description and analysis of particular languages and linguistic areas. It is precisely because of the potential collision between the need for clear and detailed description on the one hand, and theoretical and definitional considerations on the other, that we feel this volume is so timely. For exactly the same reason, the volume does not seek only to describe particular situations, or only to discuss the theoretical issues in a more general and discursive way, but to consider both perspectives, and indeed also the influence of each on the other.

Textbook accounts of linguistic areas, perhaps typically, begin by giving apparently straightforward and uncontentious definitions. Campbell (1998: 299) tells us that '*Areal linguistics*, related to borrowing ... is concerned with the diffusion of structural features across language boundaries within a geographic area', while Trask (1996: 315) suggests that:

> centuries of contact between languages can lead to a particularly striking result: several neighbouring but unrelated languages can come to share a number of structural properties with one another, properties which they do not share with their closest genetic relatives elsewhere. A group of languages in which this situation obtains is called a LINGUISTIC AREA, or, using the German term, a SPRACHBUND.

Both Trask and Campbell also introduce the term 'convergence' for the process of progressive structural assimilation that is characteristic of linguistic areas.

However, this appearance of definitional harmony is short-lived, and Campbell in particular goes on to raise a number of salient and persistent difficulties with the notion of linguistic area. Although we have seen that both he and Trask characterize linguistic areas as involving structural borrowing, Campbell (1998: 300) proceeds almost immediately to invoke

'not only borrowed words, but also shared elements of phonological, morphological or syntactic structure'. So, is lexical as well as structural convergence necessary in a linguistic area? How is that 'or' to be interpreted – would we characterize a contact relationship as a linguistic area if lexical and phonological borrowing were involved, but not morphology or syntax, or do we need evidence of interpenetration of features across the board? Does bidirectionality of borrowing relationships count more than the number of features borrowed? Where do we place the cut-off point between a 'normal' case of borrowing, and something severe and profound enough to merit the label of a linguistic area? Is the number of languages involved one of the more salient features, as implied by Thomason's (2001) characterization of linguistic areas as necessarily involving three or more languages? Or is the definition of linguistic areas more a negative one? Thomason (2001: 99) further suggests that 'perhaps the most prominent similarity... is the lack of an explanation for many or most of the areal features'. It is notable also that Trask, Campbell and Thomason all work mainly by illustration rather than by definition in their outlines of linguistic areas: in other words, all work towards what a linguistic area is by giving examples of what a linguistic area has been said to be, rather than by giving a fairly watertight definition and some examples of situations the definition excludes. Even Heine and Kuteva (2005: 174), who do provide a set of characteristics they see as indicative of linguistic areas, note that 'This characterization is fairly general, it is not meant to be a definition; rather, we use it as a convenient heuristic for identifying possible instances of sprachbunds.' In turn, the illustrations given tend to show how the concept is elastic and problematic, since it is debatable how far the candidate examples share unifying characteristics. In short, we move rapidly from an apparently clear and helpful definition to the admission that 'what we understand about linguistic areas is depressingly meager, compared to what we don't understand about them' (Thomason 2001: 99).

We do not only have a confusion over definitions, or perhaps a proliferation of partially overlapping definitions, but also a proliferation of partially overlapping terms. Campbell (1998) talks not only about linguistic areas, but also about *Sprachbunds*, diffusion areas, adstratum relationships and convergence areas. While these are seen to a large extent as being intertranslatable terms, there are attempts elsewhere to subdivide linguistic areas, so that Heine and Kuteva (2005: 172), for example, distinguish *Sprachbunds*, which are 'defined by the presence of a limited set of linguistic properties'; metatypy, where 'the languages concerned exhibit a high degree of mutual intertranslatability'; and grammaticalization areas, which 'are the result of one and the same historical process, more specifically, of the same process of grammaticalization, even if there may be other properties in addition' (though Heine and Kuteva also concede that these types may not be mutually exclusive). This foregrounding of grammaticalization echoes Campbell's

(1998: 306) suggestion that different features may be weighted differently in the establishment of a linguistic area, though he does not single out a particular feature or feature type. Again, however, we return to the paradoxes that seem inherent in ideas of convergence, since Heine and Kuteva (2005: 182) suggest that 'The contribution that the study of grammaticalization can make to defining areal relationship is a modest one' (whereas, on the other hand, they see cases of metatypy, like Kupwar in India or the East Tucanoan-Tariana contact region in north-west Amazonia, as almost prototypical cases of convergence).

These definitional questions, of course, are not peculiar to discussions of linguistic areas. In the literature on many linguistic topics, there have been lengthy discussions of what particular constellation of features we need to identify in order to 'approve' a given label. On the other hand, we can ask whether the very existence of those contributory features means there is no need to recognize the superordinate category of grammaticalization or convergence in their own right. Both these rather mechanical, feature-counting and checking approaches can be challenged. Accepting that components exist does not necessarily rule out the recognition of the higher-order category they compose: hydrogen and oxygen are real, but so is water. On the other hand, checking off the number and type of features in any specific case before we agree to identify, say, a linguistic area can be equally side-tracking. The conclusion might be that, since we already know that so many linguistic categories and phenomena are graded, with prototypical and less clear cases, we should not be so reluctant to recognize that there might be more central and more peripheral linguistic areas.

Of course, if there were a clear and absolute ruling on how we define linguistic areas, there would be far less need for this volume. The papers included here fall into two categories, although it is perhaps fitting that there is no absolute dividing line between the two: while some focus on how, and indeed whether, linguistic areas can be defined, others deal with the characteristics of individual cases, more or less clear, of convergence.

Of the chapters that focus on individual convergence areas, it is in keeping with the theme of the volume that few deal with completely uncontentious or generally accepted examples: either new examples are proposed (as in Anderson); or there is a concentration on individual, less clear aspects of better-known areas (Bakker); or the focus is on a much-discussed area, but convergence is not accepted at face value (as in Bowern's discussion of Australia). Anderson's outline of Siberia as a linguistic area provides a sound descriptive basis for a discussion of a relatively unfamiliar situation, setting the scene for the assessment of a range of features that appear not to be typologically, logically or genetically related, and which are therefore plausibly indicative of contact. Similarly, Güldemann's account of possible influence of Tuu on Khoekhoe develops arguments for a Cape linguistic area, and provides a salutary counterpoint to the Greenbergian

assumption of relatedness among all Khoisan languages. Johanson and Bakker focus on the contribution of individual languages to larger, putative linguistic areas, the former on Turkic in the Caucasus, and the latter on Portuguese and Malay, two newly 'Indianized' immigrant languages introduced relatively recently to the well-known Sri Lankan *Sprachbund*. This in turn is connected to the larger South Asian linguistic area; Heine and Kuteva (2005: 177) suggest that such 'macro-areas', also including the Balkans, and Meso-America, 'are notoriously messy', and can better be characterized by regarding them as being composed of several smaller 'micro-areas', which may be more amenable to investigation. Bisang takes a different approach to a further macro-area, namely Ethiopia, arguing that we should not seek to be too stringent in our application of absolute criteria for the identification of linguistic areas, but that equally we should not reject the category altogether (as Campbell suggests); rather, we need to relax the concept, recognizing a more clinal, graduated concept of zones of contact-induced structural convergence.

Stolz, also on the basis of general and theoretical discussion rather than concentration on a specific example, tends to agree with Campbell that historical linguists can become too preoccupied with exactly how we define linguistic areas, concluding (p. 46) that we would do better to 'either strip the term of its unwelcome and much too suggestive connotations or abolish it for good (but it should be kept in the virtual museum of linguistic thought as an example of how difficulties and misunderstandings can be created by terminology)'. Other contributions, however, problematize the individual examples, but without necessarily seeking to do away with the general concept. Thus, Koptjevskaja-Tamm assesses the contribution of two completely independent sets of features, one syntactic and the other prosodic, to the identification of the Circum-Baltic linguistic area; but although finding a whole series of indubitably contact-induced similarities, she is unable to identify any similarity that unites all the languages spoken around the Baltic. The case for the existence of a Circum-Baltic convergence area thus remains unproved: but this is an excellent example of how we must check and interrogate our evidence if we are to tell convincing cases of convergence from less convincing ones, or properly assess the extent and type of contact that must be at issue for a linguistic area to be identified.

In Chapter 10, Bowern also challenges a claimed linguistic area, though on somewhat different grounds. She does not take issue with the geographical distribution of the linguistic features themselves, but rather with the historical mechanisms by which they have come to populate their existing territory. Bowern assesses Dixon's punctuated equilibrium approach to Australian languages, and proposes an alternative account involving divergence in situ of neighbouring dialectal speech communities, which none the less remained in contact. Genetically related languages would in this case be difficult to subgroup and place on a family tree, since in some cases

isoglosses would overlap, leading to conflicting signals of subgrouping; without seeking to argue against the concept of linguistic areas in a global sense, Bowern none the less challenges it as the only plausible historical account of the contemporary Australian linguistic situation.

Bowern also raises the question of how we might in practical terms determine whether specific shared features do reflect contact, or whether they may reflect some other historical explanation, either genetic inheritance or parallel development perhaps. This is absolutely crucial to our development of general theories of convergence, and to our testing of specific cases, and two final chapters address this question in particular. In Chapter 3, McMahon and McMahon consider how quantitative methods might allow us to identify contact-induced features, and how these might behave in simulated and real language data; Musgrave, in Chapter 9 though also raising these general issues in outline, focuses specifically on a particular test case, involving shared features between Austronesian and non-Austronesian languages of Eastern Indonesia. Although these might be seen as contact-induced, and might therefore encourage the proposal of Eastern Indonesia as a linguistic area, Musgrave argues that the data do not support this interpretation. Future discussions might certainly benefit from both the application of quantitative or statistical methods, and the careful analysis of the particular dataset at issue.

The chapters collected here, then, represent a whole range of approaches to the idea of linguistic areas: some are positive about the general approach, though they may not find evidence for convergence in a specific situation; others argue that the concept is unhelpful; and yet others argue for its modification, or the application to new data. Together, they provide an appropriately multi-faceted picture of a complex and fascinating topic, which while not easily characterized, is none the less interesting, not least because of its applicability to many language contact situations, regardless of geography and typology. Bringing this range of viewpoints together depends above all on the authors, and we thank them for their prompt delivery of their manuscripts, and their full engagement with readers' comments. The idea for this volume originated in a North-West Centre for Linguistics conference in 2002 (although not all the papers from the conference are included in the volume, and some papers here were not presented there), and we thank Francesco Goglia, Diane Blakemore and Kevin Watson for their assistance with the organization of that meeting, and the Mont Follick Fund of the University of Manchester for its contribution to costs. We also thank an anonymous reader who, along with the three editors, read and commented on all papers. Finally, we are grateful to the School of Languages, Linguistics and Cultures of the University of Manchester for a contribution to the costs of preparing this manuscript. We hope that this volume will contribute to further debate and discussion of the notion of linguistic areas.

References

Campbell, L. (1998) *Historical Linguistics: An Introduction* (Edinburgh: Edinburgh University Press). (1999 American rights edition, Cambridge, Mass.: MIT Press.)

Heine, B. and T. Kuteva (2005) *Language Contact and Grammatical Change*, Cambridge Approaches to Language Contact, 3 (Cambridge: Cambridge University).

Thomason, S. G. (2001) *Language Contact: An Introduction* (Edinburgh: Edinburgh University Press).

Trask, L. (1996) *Historical Linguistics* (London: Arnold).

Contents