[Insert series page:]

Cambridge Handbooks in Language and Linguistics
The Cambridge Handbook of Areal Linguistics

Preface

List of figures
List of tables
List of maps
List of contributors

Areas, areal features and areality
Raymond Hickey

Part I: Issues in areal linguistics

1 Why is it so hard to define a linguistic area? 31
   Lyle Campbell

2 Areas and universals 49
   Balthasar Bickel

3 Reassessing sprachbunds: A view from the Balkans 61
   Victor A. Friedman and Brian D. Joseph

4 Areal sound patterns: From perceptual magnets to stone soup 90
   Juliette Blevins

5 Convergence and divergence in the phonology of the languages of Europe 119
   Thomas Stolz and Nataliya Levkovych

6 Word prominence and areal linguistics 148
   Harry van der Hulst, Rob Goedemans, Keren Rice

7 Semantic patterns from an areal perspective 186
   Maria Koptjevskaja Tamm and Henrik Liljegren

Part II: Case studies for areal linguistics

8 The Germanic languages and areal linguistics 215
   Johan van der Auwera and Daniël Van Olmen

9 Britain and Ireland 242
   Raymond Hickey

10 Varieties of English 273
   Bernd Kortmann and Verena Schröter
<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Author/Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Slavic languages</td>
<td>Alan Timberlake</td>
</tr>
<tr>
<td>12</td>
<td>The Caucasus</td>
<td>Sven Grawunder</td>
</tr>
<tr>
<td>13</td>
<td>Western Asia</td>
<td>Geoffrey Haig</td>
</tr>
<tr>
<td>14</td>
<td>An areal view of Africa</td>
<td>Bernd Heine and Anne-Maria Fehn</td>
</tr>
<tr>
<td>15</td>
<td>Areal contact in Nilo-Saharan</td>
<td>Gerrit Dimmendaal</td>
</tr>
<tr>
<td>16</td>
<td>Niger-Congo languages</td>
<td>Jeff Good</td>
</tr>
<tr>
<td>17</td>
<td>The Kalahari Basin area as a ‘Sprachbund’ before the Bantu expansion</td>
<td>Tom Güldemann and Anne-Maria Fehn</td>
</tr>
<tr>
<td>18</td>
<td>South Africa and areal linguistics</td>
<td>Rajend Mesthrie</td>
</tr>
<tr>
<td>19</td>
<td>Jharkhand as a ‘linguistic area’</td>
<td>John Peterson</td>
</tr>
<tr>
<td>20</td>
<td>Sri Lanka and South India</td>
<td>Umberto Ansaldo</td>
</tr>
<tr>
<td>21</td>
<td>The Transeurasian languages</td>
<td>Martine Robbeets</td>
</tr>
<tr>
<td>22</td>
<td>The changing profile of case marking in the northeastern Siberia area</td>
<td>Gregory D. S. Anderson</td>
</tr>
<tr>
<td>23</td>
<td>Languages of China in their East and Southeast Asian context</td>
<td>Hilary Chappell</td>
</tr>
<tr>
<td>24</td>
<td>Language in the Mainland Southeast Asia area</td>
<td>N. J. Enfield</td>
</tr>
<tr>
<td>25</td>
<td>Southeast Asian tone in areal perspective</td>
<td>James Kirby and Marc Brunelle</td>
</tr>
<tr>
<td>26</td>
<td>The areal linguistics of Australia</td>
<td>Louisa Miceli and Alan Dench</td>
</tr>
</tbody>
</table>
27  *Languages of the New Guinea region*
    Malcolm Ross

28  *Languages of Eastern Melanesia*
    Paul Geraghty

29  *The Western Micronesian Sprachbund*
    Anthony P. Grant

30  *Native North American languages*
    Marianne Mithun

31  *The areal linguistics of Amazonia*
    Patience Epps and Lev Michael

32  *Linguistic areas, linguistic convergence, and river systems in South America*
    Pieter Muysken, Joshua Birchall, Rik van Gijn, Olga Krasnouhova, Neele Müller
INSERT LIST OF FIGURES
INSERT LIST OF TABLES
Contributors

GREGORY D. S. ANDERSON is Lead Scientist and President at Living Tongues Institute for Endangered Languages and Research Fellow at University of South Africa (UNISA). His main research interests are linguistic typology and historical linguistics, with particular focus on the languages of Siberia, Papua New Guinea, Africa, and the Munda and Tibeto-Burman languages of India. Recent publications include his monograph on Auxiliary Verb Constructions, as well as articles in Indian Linguistics, Language and Education, Studies in African Linguistics, Linguistic Typology and Oceanic Linguistics. Current research projects include descriptive grammars with text collections and dictionaries for the Munda language Gta’ of India and the Turkic language Xyzyl of Siberia, a monograph on Language Extinction, electronic online dictionaries and ethnobiological compendia for several languages of Papua New Guinea.

UMBERTO ANSALDO is Professor of Linguistics at The University of Hong Kong. His interests include contact linguistics, linguistic typology and language evolution. He specialises in languages of East, South and Southeast Asia as well as Pidgin and Creole studies. His most recent publication is Languages in Contact (with Lisa Lim, Cambridge University Press, 2015).

JOHAN VAN DER AUWERA is Professor of English and General Linguistics at the University of Antwerp in Belgium. His main research interests are the grammar and the semantics of modality, mood, negation and indefiniteness, both from language-specific, areal and typological as well as synchronic and diachronic perspectives. Publications relevant for this chapter include The Germanic Languages (with Ekkehard König, Routledge, 1994), ‘English do: on the convergence of languages and linguists’ (with Inge Genee, English Language and Linguistics 6, 283-307, 2002), The Languages and Linguistics of: Europe, A Comprehensive Guide (with Bernd Kortmann, de Gruyter Mouton, 2011) and ‘Modality and mood in Standard Average European’ in The Oxford Handbook of Mood and Modality, edited by Jan Nuyts and Johan van der Auwera (with Daniel Van Olmen, Oxford University Press, 2016).

BALTHASAR BICKEL holds the Chair of General Linguistics at the University of Zurich. Before this, he was a postdoc researcher at the University of California, Berkeley and then a professor of general linguistics at the University of Leipzig. His core interest are the regional and universal factors shaping the distribution of linguistic diversity over time. For this, Bickel applies methods ranging from the statistical analysis of typological databases and corpora to ethnolinguistic fieldwork and experimental methods. A special focus area is the Himalayas where Bickel has been engaged in interdisciplinary projects on endangered languages and developing and analyzing corpora of them. He has been editor of the journal Studies of Language and co-edited the volume Language Typology and Historical Contingency (2013, John Benjamins).

JULIETTE BLEVINS is Professor of Linguistics at The Graduate Center, City University of New York, and was previously a senior research scientist at the Max Planck Institute for Evolutionary Anthropology in Leipzig. Her main research interests are sound patterns and sound change, with a special focus on phonological typology, as detailed in her chapter ‘Evolutionary Phonology: A holistic approach to sound change typology’, in the
Oxford Handbook of Historical Phonology (Oxford University Press, 2014). She is currently working on a new reconstruction of Proto-Basque.

MARC BRUNELLE is Associate Professor in Linguistics at the University of Ottawa. His main research interests are the phonetics and phonology of Southeast Asian languages, with a special emphasis on tones and prosody. His research mostly focuses on Vietnamese and Eastern Cham. His work has been published in a variety of journals including Journal of Phonetics, Phonetica, Journal of the Acoustical Society of America, Diachronica and the Linguistic Review.

LYLE CAMPBELL (PhD, University of California, Los Angeles), Professor of Linguistics, University of Hawai‘i, Mānoa, has held appointments in Anthropology, Behavioral Research, Latin American Studies, Linguistics, and Spanish, and has taught in Australia, Brazil, Canada, Finland, Germany, Mexico, New Zealand, and Spain. His specializations are: languages documentation, historical linguistics, American Indian languages, and typology. He has published 20 books and about 200 articles, and won the Linguistic Society of America’s “Leonard Bloomfield Book Award” twice, for American Indian Languages (1997, Oxford University Press) and Historical Syntax in Cross-linguistics Perspective (with Alice Harris, 1995, Cambridge University Press). His current projects include the Catalogue of Endangered Languages (endangeredlanguages.com) and documentation of several indigenous languages of Latin America.

HILARY CHAPPELL holds a Research Chair as Professor in Linguistic Typology of East Asian Languages at the Ecole des Hautes Etudes en Sciences Sociales in Paris, an appointment she took up in 2005 after teaching in the Linguistics Department at La Trobe University in Melbourne for 18 years. She was originally awarded her doctoral degree in 1984 by the Australian National University in Canberra for her thesis, A semantic analysis of passive, causative and dative constructions in standard Chinese, and has over 60 publications on Chinese linguistics and typology, with a book, A Grammar and Lexicon of Hakka (with Christine Lamarre), and four edited volumes, including The Grammar of Inalienability (with William McGregor, Mouton de Gruyter, 1996), Sinitic Grammar: Synchronic and Diachronic Perspectives (Oxford University Press, 2001) and Diversity in Sinitic languages (Oxford University Press, 2015). Her main research interests are 1) rethinking the typological profile of Sinitic languages on the basis of in-depth explorations of their diversity and 2) studying of the diachronic grammar of the Southern Min or Hokkien dialect, using a corpus of late sixteenth and early seventeenth century materials.

ALAN DENCH is Professor of Linguistics and Dean of the Graduate Research School at The University of Western Australia. His principal area of expertise lies in the documentation and grammatical description of Australian Aboriginal languages, especially those of Western Australia. He has written grammars of three languages of the Pilbara – Panyjima, Martuthunira and Yingkarta – and is working towards a description of Nyamal. In addition to primary grammatical description he has made contributions to the historical and comparative analysis of Australian languages, and has written in the general area of ethno-linguistics. His work also includes contributions to studies of language contact.
GERRIT J. DIMMENDAAL is Professor of African Studies at the University of Cologne. His main research interests are the comparative study of African languages, language typology and anthropological linguistics. Recent publications include *Historical Linguistics and the Comparative Study of African Languages* (2011, John Benjamins), *Number: Constructions and Semantics. Case Studies from Africa, Amazonia, India and Oceania* (with Anne Storch, 2014, John Benjamins), and *The Leopard’s Spots: Essays on Language, Cognition and Culture* (2015, Brill).

N. J. ENFIELD is Professor and Chair of Linguistics at the University of Sydney, and a Research Associate at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. His research on language and cognition in social and cultural context is based on long-term field work in mainland Southeast Asia, especially Laos. His books include *A Grammar of Lao* (2007, Mouton de Gruyter), *Dynamics of Human Diversity: The Case of Mainland Southeast Asia* (2011, Pacific Linguistics), *The Cambridge Handbook of Linguistic Anthropology* (with Paul Kockelman and Jack Sidnell, 2014, Cambridge University Press) and *Mainland Southeast Asian Languages: The State of the Art* (with Bernard Comrie, 2015, de Gruyter Mouton).

PATIENCE EPPS is an Associate Professor of Linguistics at the University of Texas at Austin. Her research focuses on indigenous languages of Amazonia, particularly involving description/documentation and the study of language contact and change. Recent publications include *A Grammar of Hup* (2008, Mouton de Gruyter) and articles in *Linguistic Typology, International Journal of American Linguistics, Journal of Ethnobiology*, and *Journal of Language Contact*.

ANNE-MARIA FEHN is a research fellow at InBIO-CIBIO Porto and also affiliated to the Max Planck Institute for Evolutionary Anthropology Leipzig. She holds a Ph.D. in African Studies from the University of Cologne and has been a part of the ‘Kalahari Basin Area’ project of the EUROCORES programme EuroBABEL which was hosted at the Humboldt University of Berlin. Her research interests include language documentation and areal linguistics. She is currently working on ‘Khoisan’ languages of Botswana and Angola, and on varieties of Himba and Kuvale spoken in southwestern Angola.

VICTOR A. FRIEDMAN (PhD., University of Chicago, 1975) is Andrew W. Mellon Distinguished Service Professor in the Department of Linguistics at the University of Chicago, where he also holds an associate appointment in the Department of Anthropology. He is also Director of Chicago’s Center for East European and Russian/Eurasian Studies and president of the U.S. National Committee of the International Association for Southeast European Studies. He has held Guggenheim, Fulbright-Hays, ACLS, IREX, NEH, and other fellowships. His publications include *The Grammatical Categories of the Macedonian Indicative* (1977, second revised edition, 2014), *Turkish in Macedonia and Beyond* (2003), *Studies in Albanian and Other Balkan Languages* (2004), *Očerk i lakskogo jazyka* (2011), and *Makedonistički Studii* (2011) as well as more than 200 scholarly articles. His main research interests are grammatical categories and sociolinguistic issues related to contact, standardization, ideology, and identity in the languages of the Balkans and the Caucasus.
PAUL GERAGHTY graduated from Cambridge with an MA in Modern Languages (French and German), and earned his PhD from the University of Hawai’i with a dissertation on the history of the Fijian languages. He was Director of the Institute of Fijian Language and Culture in Suva from 1986 to 2001, and is currently Associate Professor in Linguistics at the University of the South Pacific and Adjunct Associate Professor at the University of New England in Australia. He is author and editor of several books, including The History of the Fijian Languages (University of Hawai’i Press, 1983), Fijian Phrasebook (Lonely Planet, 1994), Borrowing: a Pacific Perspective (with Jan Tent, Australian National University, 2004) and The Macquarie Dictionary of English for the Fiji Islands (Macquarie Dictionary Publishers, 2006) and numerous articles in professional journals and newspapers on Fijian and Pacific languages, culture, and history. He is also well known in Fiji as a newspaper columnist and radio and TV presenter.

RIK VAN GIJN is post-doctoral researcher at the University of Zürich. His main research interests are South American languages, areal typology, morphology, and complex sentences. Recent publications include co-edited volumes Subordination Strategies in Native South American Languages (John Benjamins, 2011) and Information Structure and Reference Tracking in Complex Sentences (John Benjamins, 2014). Current research projects look at areal distributions of subordination strategies and morphological patterns.

ROB GOEDEMANS is presently employed as Editor-in-Chief and ICT&Education coordinator at the Humanities Faculty of Leiden University. He is associated with the Leiden University Centre for Linguistics as a guest researcher. His main research interests are the phonology, phonetics and typology of stress and accent. He is currently involved in a research project, funded by the Dutch National Science Foundation, with Harry van der Hulst (University of Connecticut) and Jeff Heinz (University of Delaware). The goal of this project is to merge and enhance two large databases on stress to facilitate the advance of research in stress typology.

JEFF GOOD is Associate Professor of Linguistics at the University at Buffalo. His research interests center around comparative Benue-Congo linguistics, morphosyntactic typology, and the documentation of underdescribed Bantoid languages. His recent publications include The Linguistic Typology of Templates (Cambridge University Press, 2015) and articles in Language, Morphology, Diachronica, Studies in Language, the Journal of Pidgin and Creole Languages, and Language Documentation and Conservation. He is currently heading a research project investigating the relationship between multilingualism and language change in rural areas of the Cameroonian Grassfields.

ANTHONY P. GRANT is Professor of Historical Linguistics and Language Contact at Edge Hill University, having studied at York under Robert Le Page, defending his PhD at the University of Bradford in 1995 on Agglutinated Nominals in Creole French, and having previously worked at the Universities of Manchester, Sheffield, Southampton and St Andrews. A native Bradfordian and author of over four dozen books, articles and chapters, his special research interests are Native North American languages, Austronesian languages, Romani, creoles and pidgins and issues in language
documentation. He is editor of the forthcoming *Oxford Handbook of Language Contact* (Oxford University Press) and is completing a monograph on intimate language contact.

SVEN GRAWUNDER has been working since 2005 as a post-doctorate researcher at the Max Planck Institute for Evolutionary Anthropology in Leipzig. His main research topics are phonetic motivations of sound change with an areal perspective, and voice (production and perception) from an evolutionary perspective. Current projects involve the assessment of phonetic speaker variability in processes of neutralization as well as areal phonetic typology of glottalization, ejective stops, germination, palatalization and pharyngealization (mainly) in languages of the Caucasus.

TOM GÜLDEMANN is Professor of African Linguistics at the Department of African Studies of the Humboldt University Berlin. His main research interests are language description and typology as well as historical and areal linguistics. Major publications include *Quotative Indexes in African Languages. A Synchronic and Diachronic Survey* (Mouton de Gruyter, 2008) and *Beyond ‘Khoisan’: Historical Relations in the Kalahari Basin* (with Anne-Maria Fehn, John Benjamins, 2015). Current research projects deal with macro-areal linguistics in Africa in general and the linguistic and population history of southern Africa in particular.

GEOFFREY HAIG is Professor of General Linguistics at the Institute for Oriental Studies at the University of Bamberg. His main areas of research are the languages of the Middle East with a focus on Iranian languages, areal linguistics, documentary linguistics, and language typology, in particular corpus-based typology.

HARALD HAMMARSTRÖM studied Computer Science and Linguistics at the University of Uppsala (1997-2003). He then went on to do a PhD in Computational Linguistics at Chalmers University (2004-2009) focusing on computational models that cater to diverse kinds of languages. In his postdoctoral work he started documentation of the endangered language isolate Mor in Papua, Indonesia. At present he is research staff at the Max Planck Institute for Psycholinguistics in Nijmegen where he is engaged in empirical and computational approaches to linguistic diversity, genealogical/areal relationships and language universals.

grammaticalization theory.


HARRY VAN DER HULST is Professor in Linguistics at the University of Connecticut. His main area of research is phonology (of both spoken and sign languages). He is the author of Syllable Structure and Stress in Dutch (Foris, 1984) and The Phonological Structure of Words. An Introduction (with Colin Ewen, Cambridge University Press, 2001) as well as some 26 edited books and 150 articles/book chapters. He is also the editor-in-chief of The Linguistic Review (since 1990).

BRIAN D. JOSEPH is Distinguished University Professor of Linguistics and The Kenneth E. Naylor Professor of South Slavic Linguistics at The Ohio State University, where he has taught since 1979. He received his Ph.D. from Harvard University in 1978, writing his dissertation on syntactic change between Medieval and Modern Greek. Brian Joseph specializes in historical linguistics, Greek linguistics, and Balkan linguistics, and has published extensively in these areas, including the monograph The Synchrony and Diachrony of the Balkan Infinitive: A Study in Areal, General, and Historical Linguistics (Cambridge University Press, 1983). He served as editor of Diachronica from 1999-2001 and as editor of Language from 2002-2008.

SIMON VAN DE KERKE is a retired senior lecturer at Leiden University and a specialist in indigenous languages of Bolivia. He wrote his thesis on the morphological structure of the verb in Bolivian Quechua. Currently he is working on Pukina and Leko in particular. He wrote the grammatical sketches of these languages for the first volume of the Lenguas de Bolivia series (La Paz: Plural Editores, edited by Emily Irene Cirevels and Pieter Muysken; the sketch of Pukina with Willem Adelaar as co-author).

JAMES KIRBY has been a Lecturer in Phonetics at the University of Edinburgh since 2011. His research is on the phonetic and phonological underpinnings of sound change, with particular attention to tonogenesis and the phonetic mechanisms underpinning the realisation of tone and voice quality.

MARIA KOPTJEVSKAJA-TAMM is Professor of Linguistics at Stockholm university. She has published extensively on various aspects of semantically oriented typology, combining synchronic and diachronic approaches across many languages. A large portion of her work focuses on the interplay between lexical and grammatical semantics. An
important direction in her work has been areal typology, with the main focus on the European and, particularly, the circum-Baltic languages, Östen Dahl and Maria Koptjevskaja-Tamm (eds) *The Circum-Baltic Languages: Typology and Contact*. 2 vols. (2001, John Benjamins). More recent publications include *New Directions in Lexical Typology* (special issue of *Linguistics*, 2013, co-edited with Martine Vanhove) and *The Linguistics of Temperature* (2015, John Benjamins).

BERND KORTMANN is Full Professor in English Language and Linguistics at the University of Freiburg, Germany, and Director of the Freiburg Institute for Advanced Studies. His main research interests include the areal typology of Europe, grammaticalization, language complexity, and grammatical variation across non-standard varieties of English. His publications include four monographs, nine edited volumes, two handbooks, one print and one open-access online atlas on grammatical variation in the anglophone world (2012/13), and about 90 research articles and reviews in journals and collective volumes. He is co-editor of the journal *English Language and Linguistics* and co-editor of two international book series.

OLGA KRASNOUKHOVA received her PhD at the Centre for Language Studies of Radboud University Nijmegen, the Netherlands. Her main research interests are linguistic typology, morphology and noun phrase. Her PhD thesis focuses on noun phrase structure in indigenous South American languages. Her other publications deal with such topics as attributive possession and the typology of demonstratives in South American languages.

NATALIYA LEVKOVYCH is Lecturer in General Linguistics at the University of Bremen. Her main research interests are language contact, areal linguistics and typology. Recent publications include *Po-russki in Deutschland. Russisch und Deutsch als Konkurrenten in der Kommunikation mehrsprachiger Gruppen von Personen mit postsowjetischem Hintergrund in Deutschland*. (2012, Brockmeyer).

HENRIK LILJEGREN is Researcher in Linguistics at Stockholm University. He is a co-founder of Forum for Language Initiatives, a resource and training centre for the many language communities in Pakistan’s mountainous North where he carried out linguistic fieldwork, primarily in the Indo-Aryan Palula community. His main research interests are areal-linguistic typology, Indo-Iranian languages (in particular in the north-west of the Indian subcontinent), case alignment, phonology and lexicography. He has also been engaged in revitalization efforts, mentoring language activists in local communities. He is presently leading a Swedish Research Council project, investigating language contact and relatedness in the Hindukush region.

RAJEND MESTHRIE is Professor of Linguistics at the University of Cape Town where he teaches sociolinguistics, including language contact and variation. He was head of the Linguistics Section (1998-2009), and currently holds an National Research Foundation research chair in migration, language, and social change. He was President of the Linguistics Society of Southern Africa (2002-9) and co-editor of *English Today* (2008-2012). Amongst his book publications are *Introducing Sociolinguistics* (with Joan Swann, Ana Deumert and William Leap, Edinburgh University Press, 2009), *Language in South Africa* (Cambridge University Press, 2002), *A Dictionary of South African Indian English* (Capetown: University of Cape Town Press 2010) and *The Cambridge Handbook of Sociolinguistics* (Cambridge University Press, 2011).
LUISA MICELI is an Associate Lecturer in Linguistics at the University of Western Australia. Her research interests include methodological issues in historical linguistics, the role of bilingualism in language change, and Australian languages (in particular Pama-Nyungan languages). She is currently collaborating on a research project investigating bilingual-led form differentiation in language contact.

LEV MICHAEL is Associate Professor of Linguistics at the University of California, Berkeley. With a methodological grounding in language documentation, his research focuses on the socio-cultural dimensions of grammar and language use, typology, language contact in South America, and the historical linguistics of Arawak, Tupi-Guarani, and Zaparoan languages. His recent publications include Evidentiality in Interaction (with Janis Nuckolls, John Benjamins, 2012) and Negation in Arawak Languages (with Tania Granadillo, Brill, 2014).

MARIANNE MITHUN is Professor of Linguistics at the University of California, Santa Barbara. Her main research interests are morphology, syntax, discourse, prosody, and their interrelations; language contact and language change; typology and universals: language documentation; North American Indian linguistics; and Austronesian linguistics. Among her major publications is The Languages of Native North America (Cambridge University Press, 1999).

PIETER MUYSKEN is Professor of Linguistics at Radboud University. His main research interests are language contact, Andean linguistics, and Creole studies. Recent publications include The Native Languages of South America: Origins, Development, Typology (with Loretta O’Connor, Cambridge University Press, 2014) and Surviving Middle Passage. The West Africa-Surinam Sprachbund (with Norval Smith, de Gruyter Mouton, 2014). Earlier books The Languages of the Andes. (Willem Adelaar with Pieter Muysken, Cambridge University Press, 2004).

JOHN PETERSON is Professor in General Linguistics at the University of Kiel in Germany. His areas of specialization include language description, especially with respect to the Munda and Indo-Aryan languages of eastern-central India, multilingualism and linguistic theory. He is editor of the series Brill’s Studies in South and Southwestern Asian Languages (BSSAL) and is a member of the editorial board of the Journal of South Asian Languages and Linguistics.

KEREN RICE is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. Her major research interests are in phonology (with a focus on markedness), morphology, indigenous languages of North America (in particular Dene [Athabaskan] languages), and fieldwork. Recent publications include papers on accent systems in North American indigenous languages, accent and language contact in North America, derivational morphology in Athabaskan languages, and sounds in grammar writing. She is one of the co-editors of the Blackwell Companion to Phonology.

MARTINE ROBBEETS is Associate Professor of Japanese Linguistics at Leiden University. Her main research interests are historical comparative linguistics, areal linguistics, diachronic typology, grammaticalization theory, morphology and
interdisciplinary research of linguistic prehistory. Recent publications include *Diachrony of Verb Morphology. Japanese and the Other Transeurasian Languages* (2015, de Gruyter Mouton) and various edited volumes such as *Copies versus Cognates in Bound Morphology* (with Lars Johanson, 2012, Brill), *Shared Grammaticalization* (with Hubert Cuyckens, 2013, John Benjamins) and *Paradigm Change* (with Walter Bisang, 2014, John Benjamins). Starting in September 2015, she will be conducting an interdisciplinary research project on the dispersal of the Transeurasian languages, funded by a Consolidator Grant from the European Research Council.

MALCOLM ROSS is Emeritus Professor in Linguistics at the Australian National University, where he held various positions from 1986 until his retirement in 2007. His main research interests are the histories of the languages of New Guinea and the Pacific, the study of language contact, and the methodologies of historical linguistics. Recent articles have appeared in *Oceanic Linguistics*, the *Journal of Historical Linguistics* and the *Journal of Language Contact*. He is co-editor with Andrew Pawley and Meredith Osmond of the volumes (four to date) of *The Lexicon of Proto Oceanic*. Current research projects include the Oceanic Lexicon Project (reconstructing the lexicon of the language ancestral to the Oceanic group of Austronesian languages) and reconstructing the history of the Trans New Guinea family of Papuan languages.

VERENA SCHRÖTER studied English language and literature and philosophy at the University of Freiburg, and completed her MA in 2010. She is a part-time lecturer at the English Department at the University of Freiburg, and is currently working on her PhD, exploring morphosyntactic variation in Southeast Asian varieties of English.

THOMAS STOLZ is Full Professor of General and Comparative Linguistics at the University of Bremen (Germany). His main research interests are areal linguistics, language contact, morphology, and language typology. Recent publications: *Competing Comparative Constructions in Europe* (with Sander Lestrade and Christel Stolz, Akademie-Verlag, 2013), *The Crosslinguistics of Zero-marking of Spatial Relations* (de Gruyter Mouton, 2014), articles in *Linguistics, Studies in Language, Sprachtypologie und Universalienforschung*. Current research projects on the areal linguistics and typology of spatial interrogatives as well as the morphosyntax of place names in cross-linguistic perspective.

ALAN TIMBERLAKE is Professor of Slavic linguistics and Director of the East Central European Center at Columbia University. A long-term research concern of his is the integration of sociolinguistic/communicative approaches and structural analysis. Recent publications include two articles on historical syntax for the volume on Slavic languages in the series *Handbücher zur Sprach- und Kommunikationswissenschaft* (de Gruyter) and three studies on the legends of Wenceslaus of Bohemia.

DANIEL VAN OLMEN is lecturer in historical linguistics and linguistic typology at Lancaster University. His main research interests are tense, mood, modality and pragmatic markers from a West Germanic, Standard Average European and typological perspective.
Preface

The current book aims at presenting a focussed and clearly structured volume on a topical field of linguistics, that of areal linguistics. This relates to many other fields such as language contact, typology and historical linguistics, to mention the three most directly involved. However, areal linguistics is more than each of these and unifies research into how languages come to share features diachronically and the manner in which this takes place. Areal linguistics is thus both an intersection between different subfields of linguistics and a domain of research in its own right.

For the current book a team of forty seven scholars came together to discuss issues surrounding areal linguistics in their particular fields of expertise. The editor is grateful to these colleagues for agreeing to contribute to this volume, helping to make it a comprehensive and linguistically insightful work on a topical subject in present-day linguistics.

In the preparation of this book Helen Barton, commissioning editor for linguistics at Cambridge University Press, was a great source of assistance and encouragement and ready to answer any questions which arose in the course of the project, so my thanks also go to her as well as to her colleagues at the press who work as a team to transform the manuscripts of authors into finished products in print.

Münster
October 2015
Areas, areal features and areality

Raymond Hickey

1 Introduction

The clustering of linguistic features in geographically delimited areas has long been recognised by researchers on a wide range of languages. In the course of the twentieth century this recognition gave rise to the notion of ‘linguistic area’ (Emeneau 1980; Matras, McMahon and Vincent, eds, 2006), a region in which shared features among a number of languages are found with more than chance probability. The reason for such sharing lies in contact between speakers whose own language comes under the influence of others in their environment. Admittedly, this view is simplistic, but it is useful as a first approximation because it focuses attention on speaker contact. Of course there are many contact scenarios and many situations of bi- or multilingualism (Field 2002) in which individuals speak different languages to varying extents. In such cases the contact is speaker-internal, so to speak.

1.1 Areal linguistics and linguistic areas

The term ‘linguistic area’ is a useful conceptual aid and in the early days of research helped to heighten scholars awareness of shared structural features among not necessarily related languages in circumscribed geographical areas. However, the term came to dominate research (Campbell 2006) so that scholars often felt that a binary decision had to be made as to whether a given geographical area could be classified as a linguistic area or not. This concern has not always proved to be fruitful. What can be more significant is research into the forces and mechanisms which lead to languages in a given area coming to share features. This approach would highlight the scholarly concern with areality, that is the areal concentration of linguistic features. How this concentration emerges and continues to develop is centre stage, not the attempt to attach the label ‘linguistic area’ to any given region.

1.2 Areal concentrations and geography

For areal concentrations to arise many centuries of prolonged contact and population interaction are usually required, especially as the common features of such areas usually belong to the closed classes of the languages involved, typically to the phonology and morphosyntax (Matras and Sakel, eds, 2007). Furthermore, the languages of a putative area show not only internal coherence but also recognisable external boundaries with languages immediately outside the area. So feature clustering is both positive within an area and negative vis-à-vis adjoining regions. The non-linguistic characteristics of an area involve its geography: regions bound by mountain ranges, large rivers, the sea on two or three sides (peninsulas) are all candidates for locations with areal concentrations of linguistic features.

The number of features which an area shares is a much-discussed matter in linguistic typology (Campbell, this volume; Croft 1990; Haspelmath, König, Oesterreicher and Raible, eds, 2001; Hickey 2001, 2003). However, the number does not need to be great and there are cases where single features are involved. In the main one can contend that the fewer the features shared in an area the more these must be typologically unusual (statistically rare across the world’s languages) for them to be areally significant.

1.3 Areality, contact and language change

An areal view of a region is a description of its language configuration at a certain point in time. Areal considerations become dynamic once processes of language change are considered. Generally in studies of change there is a tension between possible internal versus external factors (Hickey 2012). Some authors see internal causes as primary, unless there is no other possible account in which contact may then be appealed to, see Lass and Wright (1986) as a clear demonstration of this stance. Other authors see contact, i.e. external causes, as equally possible for change and by no means secondary to internal causes, e.g. Vennemann (2002). The weighting of internal and external factors in contact and change has consequences for the classification of languages, see McMahon (2013). It can also be significant in accounting for how, not just single features, but structural patterns arose in languages as in the discussion of possible contact-induced grammaticalisation (Heine and Kuteve 2003, 2005). However, not all discussions of contact-induced change involve an areal dimension, for instance, the literature on creolisation and contact, e.g. Holm (2004), Huber and Velupillai (eds, 2007), McWhorter (ed., 2000), Migge (2003), Siegel (1987), Thomason and Kaufman (1988), does not generally include considerations of areality.

1.4 The dynamics of areality

---

2 Change are a result of contact is treated in virtually all textbooks on language change. The following works are examples of studies in which contact-induced change occupies a central position: Jones and Esch (eds, 2002), Mufwene (2008), Nichols (1992), Sankoff (2002), Trudgill (1986).
The basis for areality is obviously language contact which leads to different degrees of feature transfer between languages. Certain developments in a language, and the community which speaks it, can be viewed as areality-enhancing and others as areality-diminishing. For instance, accommodation (Trudgill 1986) is areality-enhancing but dissociation (Hickey 2013) is areality-diminishing.

Dynamics of areality

*areality-enhancing*
- accommodation during contact (without shift);
- increase in bi-/multilingualism with sharing;
- feature transfer during language shift leading to sharing across at least two languages.

*areality-diminishing*
- dissociation between languages or varieties;
- decrease in bi-/multilingualism with loss of shared features;
- processes of standardisation or de-creolisation;
- importation of outside features to only some of the languages/varieties in an area.

1.5 Changes in areality

The degree of areality within a region is not constant. The processes outlined above can lead to changes in the level of areality. This means that investigations of putative feature-sharing areas at the present are just snapshots in a historical development which began in the past, in most cases in the deep past. And if a region has a high level of areality this does not mean that it will maintain this level, see the areality-diminishing processes listed above.

The following figures show some of the processes in more detail which can lead to a change in the areality level of a region containing different languages, whether genetically related or not.

Figure 1. Increase in areality due to close contact
Figure 2. Increase in areality due to language shift

Figure 3a. Feature development: decrease in areality
Figure 3b. Feature development: decrease in areality

Figure 4. Feature development: (coincidental) increase in areality
2 Issues in areal linguistics

Linguistic levels are affected to different degrees in regions sharing languages. It would seem that grammar is least affected in areal contact because it is the core of a language and consists of closed classes acquired in early childhood. However, as researchers on language contact have pointed out (Thomason 2001; Hickey 2010) there is no part of a language which cannot in principle be borrowed by speakers of another language. The details of this borrowing process, above all whether between speakers or internally for single bilingual speakers (Matras 2009), is a matter of discussion in the literature.

Levels of language and borrowing

Levels most affected

- Vocabulary (loanwords, phrases)
- Sounds (present in loan-words; otherwise through language shift)
- Speech habits (general pronunciation, suprasegmentals [stress, intonation])
- Sentence structure, word-order
- Grammar (morphology: inflections)

Levels least affected

2.1 Levels of language

Sound variation can be used to differentiate quite small areas as opposed to grammatical variation which tends to be typical of larger regions. The reason is probably that phonetic variation is immediately available for assessment in anyone’s speech whereas
grammatical features might not occur in any given stretch of discourse and so are not so suitable for fine differentiation, either spatially or socially.

The development of inflectional morphology is generally a community-internal phenomenon which takes several centuries to mature, consider the extensive morphology in many Indo-European languages. With adult language contact, morphological features are not usually transferred unless they are transparent, productive and easily separable from lexical bases. However, with child language learners, morphology can be borrowed with ease.

Syntactic transfer across speech communities can result via bilingualism and/or language shift but whether it can result from adult contact among speakers of different languages is disputed in the literature, see the discussion of the rise of the progressive in English (Hickey, this volume). The structures in question may be those used to express the same category or which represent the same organisational principle in two or more languages but where there is a difference in exponence, e.g. the change from one canonical word order to another, e.g. the rise of SOV in a language which previously had VSO as basic order. Probably the earliest example of contact-induced syntactic change shows just this: despite the posited verb-initial syntax of Proto-Semitic, Akkadian shows verb-final position which is assumed to be an areal influence from Sumerian (Zólyomi 2012: 402). Syntactic transfer is also common in cases of language shift where the original language of a speech community is abandoned and (nearly) all the speakers shift to the new language within a fairly well delimited period of time, a few centuries at most. In such situations speakers search for equivalents to structures in the target language which they are familiar with from their native language (Hickey 2001).

As an open class which speakers are consciously aware of, the lexicon is the first to experience transfer in adult language contact. All languages with a documented history exhibit lexical borrowings in their textual record, although the amount of borrowing can vary. For instance, Irish throughout its history has experienced many borrowings from Latin, Old Norse, Anglo-Norman and English whereas Icelandic has relatively few borrowings (Kvaran 2004: 145-149).

There is, however, an important caveat here: if the speakers of languages in contact have a negative attitude towards each other, then lexical borrowings become unlikely. But if the contact is very prolonged then structural transfer from closed classes, which show less conscious awareness among speakers, can occur, see the examples discussed in Epps and Michael (this volume).

### 3 Structure of the current volume

Part one of this volume is concerned with general issues in areal linguistics. The opening chapter by Lyle Campbell – ‘Why is it so hard to define a linguistic area?’ – considers the pitfalls in defining linguistic areas and discusses a large number of suggested areas while examining the features they contain and the languages spoken there.

In his contribution ‘Areas and universals’ Balthasar Bickel addresses the relationship of language universals to linguistic areas. The former has been researched intensively in recent years (Comrie and Östen, eds, 1984; Mairal and Gil, eds, 2006; Good, ed., 2008) and Bickel emphasises the necessity for this research to be coupled to
areal linguistics to increase the linguistic reliability of statements concerning areal features.3

In the field of areal linguistics, the Balkans is regarded as one of the classic cases of a linguistic area (Joseph 1983). In their chapter ‘Reassessing sprachbunds: A view from the Balkans’ Victor A. Friedman and Brian D. Joseph re-examine the evidence put forward for the Balkans and discuss the general relevance of these arguments. Also just what features can be regarded as defining for a linguistic area, and in what combinations, are considered.

The level of phonology, both segmental and suprasegmental, is catered for in three chapters. The first, by Juliette Blevins, ‘Areal sound patterns: From perceptual magnets to stone soup’ highlights the similarities between internal mechanisms for first language acquisition of phonology and external mechanisms by which sounds are borrowed and diffuse areally.

The chapter by Thomas Stolz and Nataliya Levkovych on ‘Convergence and divergence in the phonology of the languages of Europe’, reports on a large-scale typological study of sound systems across Europe and parts of Western Asia with a view to recognising and accounting for macro-patterns in sound systems, especially those which are probably triggered by language contact.

Suprasegmental phonology and contact (Clements and Gooden, eds, 2009) is addressed by Harry van der Hulst, Rob Goedemans, Keren Rice in their contribution ‘Word prominence and areal linguistics’. The authors present three cases studies and attempt to link the observations made in their attested data with those gleaned in second language phonology and loan phonology studies.

The areal linguistics of semantics is covered by Maria Koptjevskaja Tamm and Henrik Liljegren in their chapter ‘Semantic patterns from an areal perspective’ in which they offer a typological classification of areal semantics determined on the basis of a number of lexico-semantic databases.

Contact studies in the European context are well established (Fisiak, ed., 1995; Dahl and Koptjevskaja-Tamm, eds, 2001; Kastovsky and Mettinger, eds, 2001; Ureland and Broderick, eds, 1991) and the chapters presented at the outset of the second part of the current volume examine and re-evaluate insights and conclusions reached about the areality of the Germanic and the Slavic languages. Johan van der Auwera and Daniël Van Olmen in ‘The Germanic languages and areal linguistics’ examine the Germanic languages as a whole, focussing on linguistic contacts with non-Germanic languages. The areality of English is the concern of both Raymond Hickey’s chapter on ‘British and Ireland’ as well as Bernd Kortmann and Verena Schröter’s contribution on ‘Varieties of English’. The approaches are somewhat different. Hickey looks at the diachrony of English in Britain and Ireland and considers historical contact with Celtic in some detail while Kortmann and Schröter cast a much wider net and classify the varieties of English world-wide from a typological perspective. The two chapters thus provide complementary views of developments within the English language.

The large family of Slavic languages are considered with regard to the various types and periods of contact they experienced over their long recorded history in the chapter ‘Slavic languages’ by Alan Timberlake. Here the role of external history and

3 There is also a body of literature on ‘vernacular universaals’ in the context of English studies, so-called Angloversals (Filppula, Klemola and Paulasto, eds, 2009).
social structure on language development is given particular attention.

The Caucasus and Western Asia (Haig and Khan, eds, 2015) are treated in two dedicated chapters: ‘The Caucasus’ by Sven Grawunder and ‘Western Asia’ by Geoffrey Haig. Both authors offer fine-grained presentations of language contact and contact-induced change in their respective areas in the modern era and present classifications of highly complex language regions based on the results of their research.

The areal linguistics of Africa (Güldemann, ed., 2016) begins with the overview chapter by Bernd Heine and Anne-Maria Fehn – ‘An areal view of Africa’ – in which the authors deal with various classification proposals and assess their merit in the light of the most recent research into Africa, especially from a typological perspective. This approach is continued by Gerrit Dimmendaal in ‘Areal contact in Nilo-Saharan’ where he looks at morphological structures, notably case systems, to attempt a classification of this large phylum in Eastern and Central Africa. The largest language group in Africa is scrutinised in Jeff Good’s chapter ‘Niger-Congo languages’. In particular he addresses questions concerning the internal composition of the family and the proposals for various subgroupings. He stresses the importance of understanding the relationship between Niger-Congo cultures and the Niger-Congo languages.

Southern Africa is the subject of two chapters. The first is ‘The Kalahari Basin area as a “Sprachbund” before the Bantu expansion’ by Tom Güldemann and Anne-Maria Fehn which examines the evidences for linguistic groupings in the Kalahari Basin, which stretches from southern Namibia through Botswana to south-west Zambia, before the Bantu migrations into southern Africa which affected the distribution of pre-Bantu languages in the region due to contact and mixture. The second chapter is ‘South Africa and areal linguistics’ by Rajend Mesthrie considers both native Bantu languages of South Africa and the two major European descendant languages, English and Afrikaans, and considers the possible cases of transfer across linguistic boundaries in South Africa among these languages in the centuries of the colonial and the more recent post-colonial period.

South Asia (Hock and Bashir, ed., 2016) is represented in this volume by two chapters, the first by John Peterson is ‘Jharkhand as a “linguistic area” in which he investigates contact across the Indo-Aryan Munda family border. The observable convergence is not just in the lexical area but encompasses many features of morphosyntax as well. In ‘Sri Lanka and South India’ Umberto Ansaldo investigates the predominance of a number of typological profiles for languages in different parts of the world and then applies the insights from this research to the linguistic situation in Sri Lanka.

The region of Northern Siberia, stretching to the Far East of Northern Asia, is the topic of Martine Robbeets chapter on ‘The Transeurasian languages’ in which she takes a fresh look at areality over this vast area by examining the realisation of a set of 27 features which show parallels among languages as far apart as Uralic in the west and Nivkh and Ainu in the east. A specific linguistic subsystem, that of case-marking, is examined by Gregory Anderson for languages of northeastern Siberia – ‘The changing profile of case marking in the northeastern Siberia area’. He notes that the system of case-marked clausal subordination is gradually being replaced by another case system under the influence of contact with Russian.

The main language families of China are outlined in Hilary Chappell’s ‘Languages of China in their East and Southeast Asian context’ in which she examines the areal linguistics of these large groupings. Then three small clusters are scrutinised in which
contact and transfer from non-Sinitic languages have occurred.

The convergence experienced by the languages of south-east Asia is examined in the chapter by N. J. Enfield, ‘Language in the Mainland Southeast Asia area’. He reports on the confirmation of standard wisdoms concerning these mainland languages, e.g. their structural similarities, as well as on newer research which challenges orthodox views such as how the noted convergence took place during historical social contact. A more specific investigation is offered by James Kirby and Marc Brunelle who look at phonological tone in the same group of languages in their chapter ‘Southeast Asian tone in areal perspective’. They describe the diverse kinds of tone systems found in Mainland Southeast Asia and question whether tone is indeed a strong indicator of convergence in this area.

Australia and the Pacific is dealt with in a series of four chapters. The first of these, ‘The areal linguistics of Australia’, by Louisa Miceli and Alan Dench looks at standard views on areal groupings in Australia against the background of their own detailed research in the Pilbara region of Western Australia. This work confirms the complex interaction of genetic relationship and language contact in the areal settings of Australia.

An in-depth examination of classification proposals for the languages of New Guinea (Palmer ed., 2015) and some neighbouring regions is given by Malcolm Ross in ‘Languages of the New Guinea region’. He specifically addresses the question of whether a typological profile can be established for New Guinea languages, or any subsection of these, which could account for structural similarities over and beyond those established by phylogenetic connections. By looking at a predetermined set of variables across a wide range of languages Ross was able to establish that while the occurrence of variable was not always in geographically contiguous areas, nonetheless constellations of variable values were recognised which were hardly the outcome of chance.

Languages from two of the major cultural zones of the island Pacific are examined in two subsequent chapters. In the ‘Languages of Eastern Melanesia’ Paul Geraghty examines the historical evidence for Melanesian languages and for Papuan languages in the area and considers the complex landscape of present-day Vanuatu and other locations such as the Solomon Islands and New Caledonia. The linguistic features he scrutinised cross-linguistically include serial verbs and numeral systems. In the ‘The Western Micronesian Sprachbund’ Anthony Grant examines the field of lexical borrowings and classifies the languages of the area as lexical donors and lexical recipients or indeed both.

The areal linguistics of the Americas is covered by the final three chapters. The first is ‘Native North American languages’, a tour de force by Marianne Mithun in which she brings her great expertise in this area to bear on issues of classification and cross-influence in a vast and complex language region. No less complex, but with different ecologies, are the areas of South America which are examined in two dedicated chapters. ‘The areal linguistics of Amazonia’ by Patience Epps and Lev Michael considers the role of languages functioning as linguae francae and examines various contact zones in South America, dealing with language shift and borrowing which has lead to grammatical diffusion (Aikhenvald and Dixon, eds, 2001), often of bound morphology. The final chapter is ‘Linguistic areas, linguistic convergence, and river systems in South America’ by a Dutch team – Pieter Muysken, Joshua Birchall, Rik van Gijn, Olga Krasnouhova, Neele Müller – who consider the role which the large rivers of South America have played in determining the linguistic landscape of this continent, especially where rivers have led to contact and convergence.
4 Outlook

The diversity of contributions in the present volume attests to the vitality of the field of areal linguistics. Furthermore, the range of perspectives on data is wide indeed and the interpretations of features found in these data are various, showing how actively issues of language contact and development in geographically defined regions are discussed in the scholarly community today. It is hoped that this volume will add to the knowledge and insights in areal linguistics and provide an impetus for further research in this field in the foreseeable future.

References


Winford, Donald 2013 ‘Social factors in contact languages’, in Bakker and Matras (eds), pp. 363-416.

Part I: Issues in Areal Linguistics
1 Why is it so hard to define a linguistic area?

Lyle Campbell

1 Introduction

This chapter examines the criteria that have been used to define linguistic areas with the goal of determining the adequacy of these criteria for the task. It evaluates general recommendations in the literature for how to establish linguistic areas, and it looks to see how they work with respect to several of the most widely acknowledged linguistic areas. It recommends new ways of looking at the overall enterprise of areal linguistics and illustrates both the recommended approach and the problems in defining linguistic areas through a case study of the Chaco as a potential linguistic area. It is generally acknowledged that linguistic areas are “notoriously messy”, “notoriously fuzzy” things (Thomason and Kaufman 1988: 95, Heine and Kuteva 2001: 396, Tosco 2000: 332), and that “what we understand about linguistic areas is depressingly meager” (Thomason 2001: 99). I argue that the reason for this is because the focus has misleadingly been on geography rather than on the diffused changes themselves. The goal of this paper is to re-examine areal linguistics and in so doing to arrive at a clearer understanding of the notion ‘linguistic area’. The conclusion reached is that it is individual historical events of diffusion that count, not the post-hoc attempts to impose geographical order on varied conglomerations of these borrowings.

2 Proposed linguistic areas

Scholars have struggled with both the definition of the concept ‘linguistic area’ and with determining what criteria might be valid or most useful for establishing linguistic areas – the two are closely interrelated, and we need to examine both. We could approach our task of examining these two vexed but key topics in one or both of two ways: (1) by examining the linguistic areas that have been proposed to see what they have in common (for the definition of the concept ‘linguistic area’) and what defines them (for perspectives on the criteria for establishing linguistic areas), or (2) by examining the shared traits upon which each linguistic area is based and then apply those criteria that have been proposed in the linguistic literature for determining the validity or the strength of each linguistic area, to see how effective the criteria may be in determining the various proposed linguistic areas.

The exercise of trying to approach areal linguistics in these two ways, however, runs into obstacles from the outset. Not everything that has been proposed as a linguistic area is necessarily accepted by knowledgeable linguists, and not all the criteria that have been proposed may be generally accepted, either. This makes the task of trying to understand the concept of linguistic area and the criteria for establishing linguistic areas seem circular. On the one hand, the linguistic areas – whether solid ones or not – depend on the criteria utilized to establish them. On the other hand, the criteria are arrived at by examining the linguistic areas – whether generally accepted or not. This makes it unclear that the areas can tell us what the criteria should be or that the criteria can tell us what the areas should be.
The difficulty can be understood more fully from a quick look at a list of some of the linguistic areas that have been proposed in the linguistic literature. These include but are not limited to the following:

<table>
<thead>
<tr>
<th>Proposed linguistic areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Africa as a linguistic area</td>
</tr>
<tr>
<td>2. Ethiopia, or the Ethiopian highlands</td>
</tr>
<tr>
<td>3. Southern Africa</td>
</tr>
<tr>
<td>4. Macro-Sudan Belt</td>
</tr>
<tr>
<td>5. West African Sahel Sprachbund</td>
</tr>
<tr>
<td>6. The Cape Linguistic Area</td>
</tr>
<tr>
<td>7. Yobe State (Nigeria) Linguistic Area</td>
</tr>
<tr>
<td>8. Benin-Surinam Transatlantic Sprachbund</td>
</tr>
<tr>
<td>9. Trans-Atlantic Sprachbund</td>
</tr>
<tr>
<td>10. European Sprachbund (Standard Average European Linguistic Area)</td>
</tr>
<tr>
<td>11. Ireland (see Hickey 1999, 2012)</td>
</tr>
<tr>
<td>12. Scotland</td>
</tr>
<tr>
<td>13. Basque and its neighbors</td>
</tr>
<tr>
<td>14. Balkans</td>
</tr>
<tr>
<td>15. Baltic</td>
</tr>
<tr>
<td>17. Uralic-Siberian Linguistic Area</td>
</tr>
<tr>
<td>18. Uralo-Altaic Sprachbund</td>
</tr>
<tr>
<td>19. Sumero-Akkadian (Sumero-Akkadian), Ancient Near East</td>
</tr>
<tr>
<td>20. Anatolian area</td>
</tr>
<tr>
<td>21. Carpathian (Danube) Linguistic Area</td>
</tr>
<tr>
<td>22. Onogur Sprachbund</td>
</tr>
<tr>
<td>23. Rossic Sprachbund</td>
</tr>
<tr>
<td>24. Siberian Linguistic Area</td>
</tr>
<tr>
<td>25. Upper Yenisei Sprachbund (Yenisei-Kirgiz area)</td>
</tr>
<tr>
<td>26. Volga-Kama Sprachbund</td>
</tr>
<tr>
<td>27. Ostyak (Ob-Yeniseic) Sprachbund</td>
</tr>
<tr>
<td>28. Volga-Oka Sprachbund</td>
</tr>
<tr>
<td>29. Yeniseic Sprachbund</td>
</tr>
<tr>
<td>30. Ob-Ugric Sprachbund</td>
</tr>
<tr>
<td>31. Core Uralic (Central Uralic) Sprachbund</td>
</tr>
<tr>
<td>32. Peripheral Uralic (Lateral Uralic) Sprachbund</td>
</tr>
<tr>
<td>33. Eastern Uralic Sprachbund</td>
</tr>
<tr>
<td>34. South Asian Linguistic Area (Indian subcontinent)</td>
</tr>
<tr>
<td>35. Sri Lanka Sprachbund</td>
</tr>
<tr>
<td>36. Caucasus</td>
</tr>
<tr>
<td>37. Northern Asia</td>
</tr>
<tr>
<td>38. Amdo Sprachbund (Qinghai-Gansu Sprachbund)</td>
</tr>
<tr>
<td>39. Northwest China sprachbund</td>
</tr>
<tr>
<td>40. Mainland Southeast Asia</td>
</tr>
<tr>
<td>41. Sepik River Basin</td>
</tr>
<tr>
<td>42. Bird’s Head Sprachbund</td>
</tr>
</tbody>
</table>
43. East Nusantara area (Eastern Indonesian Linguistic Area)
44. Timoric (Timor Leste)
45. North Vanuatu Sprachbund
46. The Pacific Linguistic Area
47. Australia as a linguistic area
48. Northwest Coast of North America
49. Northern Northwest Coast Linguistic Area
50. Northern California
51. Clear Lake Area (California)
52. South Coast Range Area (California)
53. California Linguistic Area
54. Southern California-Western Arizona Area
55. Plateau Linguistic Area
56. Great Basin Area
57. Pueblo Linguistic Area
58. Plains Area
59. Southeast Linguistic Area
60. Northeast Linguistic Area
61. Mesoamerica
62. Lowland Maya Diffusion Area
63. The Huehuetenango Sprachbund
64. Guaporé-Mamoré Linguistic Area
65. Vaupés(-Içana Basin) Linguistic Area
66. The Chaco
67. Amazonia, Amazonian Linguistic Area
68. Upper Xingu incipient linguistic area
69. Orinoco-Amazon Linguistic Area
70. Venezuelan-Antillean Linguistic Area
71. Andes Linguistic Area (Andean Area)
72. Colombian-Central American Linguistic Area
73. Lower Central America Linguistic Area
74. Lowland South American Linguistic Area
75. Southern Cone Linguistic Area

It might be thought that a good number of these proposed linguistic areas do not have general acceptance and that therefore perhaps we could gain traction on our task by considering only the best established, most widely accepted linguistic areas. However, even this tactic is unclear, since it would reduce the task to examining the very small number of areas that enjoy a reasonable measure of general acceptance: the Balkans, South Asia (Indian subcontinent), Baltic, Mesoamerica, and perhaps the Northwest Coast of North America, plus or minus one or two. (For the traits shared in each of these areas, see, for example, Campbell 2013: 298-305). However, the task gets even bleaker: even in the case of these most widely acknowledged linguistic areas, there is no general agreement upon exactly which languages belong to each linguistic area, not on what the traits of the areas are, nor on what the geographical extent of each is. Nevertheless, let’s look further.
3 Different kinds of linguistic areas

The tasks to be tackled are made all the more complicated because of the differing sorts of linguistic areas talked about in the literature – not all areas are equal, or even similar, in their histories or composition. Things called linguistic areas have included entities of widely divergent character and historical backgrounds, differing in social, cultural, political, geographical, attitudinal, historical, and other factors (cf. Aikhenvald and Dixon 2001: 11, 13, Dahl 2001: 1458, Kuteva 1998: 308-9). As Thomason (2001: 104) explains, “[linguistic areas] arise in any of several ways – through social networks established by such interactions as trade and exogamy, through the shift by indigenous peoples in a region to the language(s) of invaders, through repeated instances of movement by small groups to different places within the area.” The different sorts of linguistic areas found in the literature include: incipient ones, only beginning to form and with as yet few shared traits; moribund and decaying ones, where due to many changes after the area was actively formed, fewer traits are currently recognizable among the languages than would formerly have been the case; layered ones (with new layers and old layers of diffusion from different sources); overlapping ones, where different areas formed on top of one another or overlapping one another at different times for different reasons; multilateral vs. unilateral areas; areas with some nuclear (core) languages and other peripheral ones; areas resulting from rapid conquest, from population spread and migration (traits moving with movement of speakers), and others through home-grown, stay-in-place contact (movement of traits but not of peoples); disrupted areas with “latecomers, earlier drop-outs, and temporary passers-by” (Stolz 2002: 265); etc.

Clearly then, “linguistic area’ is not a uniform phenomenon, either socially or linguistically” (Thomason 2001: 115). These different kinds of linguistic areas raise doubts about the utility of whole concept ‘linguistic area’. Are all these different kinds of things legitimate ‘linguistic areas’, given their very different make up and the large differences in the circumstances of their origins and histories? It appears that there is little in the notion ‘linguistic area’ to unite these different sorts of “linguistic areas” other than the fact that they all involve borrowing in some way, but borrowings of different sorts, for different reasons, in different settings, and at different times. This leads me to agree with Dahl’s (2001) conclusion:

In the end, we are led to the following more far-going question about the notion of area: to what extent do areas … have a reality of their own and to what extent are they just convenient ways of summarizing certain phenomena? At the most basic level, linguistic contact relationships are binary: one language influences another. An area is then simply the sum of many such binary relationships. (Dahl 2001: 1458.)

I argue here that a linguistic area, to the extent that the concept may be of any value, is merely the sum of borrowings among individual languages in contact situations. If we abandon the search for an adequate geographically-oriented definition of the concept and for criteria for establishing linguistic areas and instead focus on understanding the borrowings, those contingent historical facts and the difficulty of determining what qualifies as a legitimate linguistic area cease to be problems.

With this discussion in mind, let us turn now to look at the vexed problem of attempts at defining the notion ‘linguistic area’.
4  The concept ‘linguistic area’

In broad terms, we might think that a general definition of the term ‘linguistic area’ would contain reference to a geographical area in which, due to language contact, languages of a region come to share certain structural features. Linguistic areas are also referred to at times by the terms Sprachbund, diffusion area, adstratum relationship and convergence area. The central feature of a linguistic area is generally held to be the existence of structural similarities shared among languages of the geographical area in question, usually coupled with the assumption that the reason the languages of the area share these traits is because at least some of them are borrowed. This definition or something similar to it is characteristic of the notion employed in much of the work in areal linguistics. However, it is not that simple, and there are many disagreements, which make even this notional definition, however appealing, unacceptable to some scholars who work in areal linguistics, and it does not fit all the entities called ‘linguistic areas’ in the literature.

Different approaches have been taken in attempts to establish linguistic areas, with different implications for the definition of the concept. The circumstantialist approach (see Campbell 1985) mostly just lists similarities found in the languages of a geographical area and allows the list of shared traits to suggest diffusion and to define the linguistic area accordingly. Here, concrete evidence that the shared traits are diffused is not required. Circumstantialist areal linguistics has been criticized for not eliminating chance, universals, and possibly undetected genetic relationships among the languages as other possible explanations for shared traits. The historicist approach (Campbell 1985) seeks concrete evidence that the shared traits of an area are diffused (borrowed). This approach is more rigorous, although the lack of clear historical information in many cases has often led scholars to rely on the less trustworthy circumstantialist approach. Linguistic areas are often defined, surprisingly, on the basis of a rather small number of shared linguistic traits.

5  Criteria for determining linguistic areas

The following criteria are the major ones that have been utilized to define linguistic areas: the number of traits shared by languages in the area, bundling of the shared traits (clustering at geographical boundaries), and different weight attributed to different areal traits (under the assumption that some kinds of diffused changes constitute stronger evidence than others of areal affiliation). Let’s look at these a bit more closely.

5.1  Number of shared traits

merrier: some areas are more securely established because they contain many shared diffused traits and other areas with fewer shared traits are weaker (Campbell 1985, 2006a). In the linguistic areas considered above, we see considerable variation in the number and kind of traits they share and which define them.

5.2 Number of language families

The number of language families involved have played a role in some scholars’ ideas about the criteria for linguistic areas. Some have required that two or more language families be involved to define a linguistic area (see Emeneau 1980 [1956]: 124, 1980 [1965]: 127, 1980 [1978]: 1, Aikhenvald and Dixon 2001: 11, Schaller 1975: 58, Tosco 2000, van der Auwera 1998: 260). However, this appears not a strong requirement, since the Balkan Linguistic Area, probably the best-known and most widely accepted of all linguistic areas, involves only Indo-European languages for most scholars (though some include also Turkish in the area) (cf. Masica 1992:110). Some scholars temper the multiple language family requirement, suggesting instead that at least some of the languages of an area should not be closely related if only languages from one family are involved – although details of what ‘not closely related’ means in this context means left unaddressed (cf. Campbell 1985: 25, 1994: 1471, Matthews 1997: 351, Stolz 2002: 261). Nevertheless, the idea of some minimum level of required linguistic diversity among the languages of a linguistic area has not really been a focus of attention.

5.3 Trait weight

It is easy to see the value of the relatively greater weight or importance attributed to some traits than to others for defining linguistic areas, but how to assign weight is not clear. For instance, the borrowed word-order patterns in the Ethiopian linguistic area provide an instructive example. Ethiopian Semitic languages exhibit a number of areal traits diffused from neighboring Cushitic languages. Several of these individual traits, however, are interconnected due to the borrowing of the SOV (Subject–Object–Verb) basic word order patterns of Cushitic languages into the formerly VSO Ethiopian Semitic languages (Campbell et al. 1988). Typologically, the orders Noun-Postposition, Verb-Auxiliary, Relative Clause-Head Noun, and Adjective-Noun all tend to correlate with one another and to co-occur with SOV order cross-linguistically. If the expected correlations among these constructions are not taken into account, we might be tempted to count each one of these as a separate shared areal trait. Their presence in Ethiopian Semitic languages might seem to reflect several different diffused traits (SOV counted as one, Noun-Postposition as another, and so on), and they could be taken as several independent pieces of evidence defining a linguistic area. However, from the perspective of expected word order pattern co-occurrences, these word order arrangements are not independent traits, but may be viewed as the diffusion of a single complex feature, the overall SOV word-order type with its various expected coordinated orderings of typologically interrelated constructions. However, even though the borrowing of the SOV basic word order type may count only as a single diffused areal trait, many scholars would rank it as counting
for far more than many other less complex individual traits based on the knowledge of
how difficult it is for a language to change so much of its basic word order.¹

5.4 Trait bundling

With respect to the criterion of the bundling of areal traits, some scholars had thought that
such clustering at the borders of a linguistic area might be necessary for defining
linguistic areas correctly. However, linguistic areas are not typically viewed this way.
Rather, they are similar to traditional dialects in this regard. Often, one trait may spread
out and extend across a greater territory than another trait, whose territory may be more
limited, so that traits’ territories do not coincide, they do not ‘bundle’ together at some
border. Some would distinguish different kinds of traits based on their belief that some
kinds of linguistic features, such as phonological ones, may be able to diffuse more easily
and thus cover more terrain than others, such as grammatical features. The most typical
pattern is where languages within the core of an area may share a number of features, but
the geographical expanses of the individual traits may vary considerably one from
another. However, in a situation where the traits do coincide at a clear boundary, rare
though this may be, the definition of a linguistic area which matches their boundary is
considered relatively secure. An additional complication (pointed out to me by Ray
Hickey) is that even where a few features might coincide, especially ones that diffuse
more easily – the seeming bundling could be due just to coincidence. Unfortunately,
however, in most linguistic areas, the areal traits do not have the same boundaries,
offering no clearly identifiable outer border of the linguistic area in question. As
Emeneau (1980 [1965]: 1366) said, “in linguistic area studies it is doubtful if there will
every emerge isogloss-bundles.”

6 Defining ‘linguistic area’

A common perception is that the term ‘linguistic area’ is difficult to define (cf. Heine and
define the concept [linguistic area] ever since [Trubetzkoy 1928], mainly because it isn’t
always easy to decide whether a particular region constitutes a linguistic area or not.” The
principal definitions that have been given of the notion ‘linguistic area’ (or of related and
more or less synonymous terms, Sprachbund, diffusion area, convergence area, etc.) are
surveyed in Campbell (2006a), and are not repeated here in the interest of space. In spite
of prolonged efforts to define ‘linguistic area’, there is no general agreement on its
definition, and even for the most widely accepted linguistic areas, such as the Balkans,
scholars do not agree wholly on which languages belong to the area, what linguistic traits
characterize the area, and what its precise geographical extent is. In Stolz’ (2002: 260)
words, “these terms [Sprachbund, linguistic area, and areal type] seem to invite as many
meanings and readings as there are linguistic minds to contemplate them.” Stolz (2002:

¹As Raymond Hickey points out (p.c., 2015), the apparent difficulty of borrowing basic word order is supported by Irish. Despite the enormous influence from English on Irish (seen in many syntactic features), the English SVO word order has not made a single dent in Irish, which staunchly maintains VSO basic word order (with ccorrelated N+Adj, Nom+Gen, etc.), see Hickey (2010: 13).
259) is correct that “the search for clearcut definitions [of “Sprachbund, linguistics area and areal type”] has been largely futile and will probably never come to a really satisfying conclusion.”

I argue that too much effort has been wasted on trying to define the concept, that little progress has been made, and that it would be more productive to investigate the facts of linguistic diffusion without the concern for defining linguistic areas. Every ‘linguistic area’, to the extent that the notion has any meaning at all, arises from an accumulation of individual cases of ‘localized diffusion’; it is the investigation of these specific instances of diffusion, and not the pursuit of defining properties for linguistic areas, that will increase our understanding and will explain the historical facts. With the focus rather on specific instances of borrowing, many of the unresolved issues and indeterminacies which have dogged areal linguistics from the outset cease to be relevant questions.

7 New concept

Areal linguistics might appear to have been somewhat schizophrenic, sometimes concentrating on situations like set intersection and at other times on situations like set union. We have sought to define linguistic areas in a set-intersection-like manner, based mainly on traits shared by all the members of the thing being designated a linguistic area, but we have also set up linguistic areas in a more set-union-like way, seeing anything whatsoever found in the thing being designated a linguistic area as evidence defining that area so long as that trait is found in more than one language in the region. Some might say we prefer the former, with set-intersection-like traits as the main protagonists in our cast of characters, but that at the same time we also call upon the latter, with the set-union-like traits as supporting cast members. However, because we have not sorted out the consequences from inclusion of these difference kinds of traits, the definitional muddle has persisted. It could be argued that our interests are better served by making the differences explicit, with appropriate terminology to help avoid confusion. We might reserve the term linguistic area sensu stricto (LASS) for a geographical region defined by shared diffused traits mostly contained within and shared across the languages of a clearly delimited geographical space (set-intersection-like). We need a different term for the entity that focuses more on the shared traits themselves regardless of whether they show up in all the languages thereabouts or coincide in their distribution within some bounded geographical space – the set-union-like situations. I suggest for this the term trait-sprawl area (TSA). The word ‘sprawl’ here reflects the fact that the individual traits can pattern in disordered ways, with some crisscrossing some languages while others crisscross other languages, with some extending in one direction, others in another direction, with some partially overlapping others in part of their distribution but also not coinciding in other parts of their geographical distribution. The focus of the TSA is the actual diffused traits themselves, with the emphasis on answering the question which is the prime directive for historical linguists, “what happened?” – that is, with focus on what changes took place, rather than on defining some geographical area itself. To the extent that any clear LASSes with geographical focus may be established, we will still want to

2 Note here also the belief of some linguists, mentioned earlier, that phonological features might sprawl differently from grammatical or lexical traits.
answer the same question about the linguistic traits shared by the languages of the LASS of what happened.

For TSA, what I have in mind are the cases where clearly there is considerable sharing of linguistic traits in some region even though it is difficult by the application of the usual criteria for setting up linguistic areas to determine whether the case in question can legitimately be defined as a linguistic area. I have in mind cases where different traits are shared by some adjacent languages but that the set of individual traits can vary greatly in terms of which languages may share which ones of them and how different they can be in their territorial expanse. Let me illustrate the notion of TSA and what it means for the problems of defining linguistic areas with a brief case study – the Grand Chaco of South America.

Others have observed similar difficulties for defining linguistic areas due to different behavior with respect to which languages share which traits, though not conceptualized precisely as proposed here. Daumé (2009) has developed a Bayesian approach to attempt to quantify traits and thus kinds of linguistic areas. Muysken et al. (2015: 209), following Lindstedt (2000), note that “a Sprachbund may exhibit gradience in language membership, from ‘core’ languages in the Sprachbund to ‘non-core’ languages ... thus the Balkans can be characterized as a Sprachbund with ‘soft’ boundaries: there is no sharp transition for Sprachbund to Sprachbund”. They explore possible quantitative definitions of linguistic areas, though their areal traits are not weighted.

8 Is the Gran Chaco a linguistic area?

The Gran Chaco is the extensive dry lowland plain of northern Argentina, Paraguay, southeastern Bolivia, and southern Brazil. It has been considered a culture area characterized by cultural traits shared across ethnic boundaries (cf. Murdock 1951, Miller 1999), and it has sometimes been considered a linguistic area, too, although the evidence has not been clear (see Tovar 1961, Rona 1969-72, Kirtchuk 1996, Grondona 2003, Viegas Barros 2002: 140, Adelaar and Muysken 2004: 386, 499, Aikhenvald 2011, Messineo 2011, Campbell and Grondona 2012). More than twenty languages from six language families are found in the Chaco region (see Grondona and Campbell 2012 for details). The language families are: Guaicuruan (Waykuruan), Matacoan (Mataco-Mataguayan), Mascyan (Mascoian, Mascoyan, Lengua-Mascoy, Enlhet-Enenlhet), Lule-Vilelan, Zamucoan, and Tupí-Guaranían (a branch of Tupían). There were also a number of poorly known extinct languages in the region that are not dealt with here (see Campbell 2012a for details). A look at what makes determining whether the Chaco is a linguistic area difficult will help to clarify central issues addressed in this chapter and the concept of TSA. Details of the Chaco linguistic traits are not repeated here (for that see Campbell and Grondona 2012), but some of the main ones are discussed sufficiently to help clarify what is at stake here.

8.1 SVO Word order

Shared basic word order has been cited as a trait of Chaco languages (see Tovar 1961: 195, Tovar and Tovar 1984:202, Adelaar and Muysken 2004: 499, Messineo 2011,
Campbell 2012b. SVO basic word order is found in Matacoan and Guaicuruan languages, and Ayoreo (Zamucoan). SVO basic word order may be characteristic of several Chaco languages, but not of all of them; Lule and Vilela have SOV (with Noun-Adjective, Modifier-Head) (Comrie et al. 2010: 91-92). SVO also does not set Chaco languages clearly apart from languages of neighboring parts of South America.

8.2 Gender

Grammatical gender has been suggested as a Chaco areal trait (Tovar 1961, Aikhenvald 2000: 80). Matacoan, Guaicuruan, Zamucoan, and Mascoyan languages have a masculine-feminine gender distinction. However, gender as a grammatical category is also found widely elsewhere in South America, but, again, not in all languages of the Chaco.

8.3 Genitive classifiers

Matacoan, Guaicuruan, Maskoyan, and Zamucoan languages have a genitive (or possessive) noun classifier construction for possessed domestic animals (cf. Fabre 2007, Messineo 2011, Campbell 2012b). In these languages it is not possible to say directly, for example, 'my cow'; rather, a 'possessive domestic animal classifier' that bears a possessive pronominal prefix is necessary. Genitive classifiers are rare though several other South American Indian languages also have them, but in these languages it is usually a member of a larger classifier system with several other noun classifiers. These Chaco languages typically have only the genitive classifiers, not other kinds of noun classifiers, and this may distinguish them from languages of other areas.

8.4 Rich set of demonstratives

Chaco languages typically have a rich system of demonstratives, with forms distinguished on a number of different semantic parameters which include visible vs. not visible, often also proximate, distal, extended horizontally, extended vertically, three-dimensional, known from first hand experience, known from hearsay, no longer existent, etc. (see Campbell 2012b, Campbell and Grondona 2012, Messineo et al. 2011). However, this is far from unique to Chaco languages; numerous other South American languages also have complex demonstrative systems.

8.5 Active-stative verb alignment

Active-stative alignment characterizes many Chaco languages and may be an areadefining feature (Grondona 2003, Adelaar and Muysken 2004: 499). Matacoan and Guaicuruan languages are clearly active-stative; Enlhet (Maskoyan) appears to be (Grubb 1914: 319). Guaraní is well known for having active-stative alignment. Active-inactive languages which operate on the semantic criterion of event (for active) and non-event or stative (for inactive) seem to be limited mostly to the Chaco and adjacent regions –
though not exclusively; for example Tunica, an extinct language isolate of Louisiana, also
has active-stative alignment (Heaton 2012). This feature, however, also extends beyond
the Chaco, found in several Tupí-Guaraní languages. Thus is an important feature of
the Chaco, but not limited to the Chaco region alone.

8.6 Lack of verbal tense

Matacoan, Guaicuruan, Ayoreo, and Guarani for the most part do not mark tense on verbs
– tense in these languages is either determined from context or signalled by adverbials,
demonstratives, or directionals. Several have nominal tense (instances where a nominal
or part of a noun phrase, not the verb, carries the tense information for the proposition)
(cf. Vidal and Gutiérrez 2010; Tonhauser 2007, 2008; Campbell 2012b). Nevertheless,
nominal tense is also found in several other South American languages, not just in the
Chaco (Campbell 2012a).

8.7 Directional verbal affixes

Chaco languages typically have a complex set of directional verbal affixes, found for
example in Matacoan, Guaicuruan, Maskoyan, Zamucoan, and Chiriguano (Tupí-
Guaraní) (cf. Campbell and Grondona 2012, Messineo 2011). This trait may
characterize most Chaco languages. Again it is not limited to the Chaco, however. Many
other languages in southern South America have directional affixes on verbs (Campbell
2012b).

8.9 Other postulated areal traits in the Chaco

A number of other traits have been mentioned as possibly diffused among Chaco
languages. Most of these, however, are not compelling as evidence for a Chaco linguistic
area (assessed in Campbell and Grondona 2012). Several are commonplace and can be
found easily in languages around the world. Others are found not just in the Chaco but
also widely in other areas of South America. This does not mean these traits are not
diffused, but only that they provide no strong evidence for designating the Chaco as a
clearly delineated linguistic area, since they do not distinguish Chaco languages from
languages in neighboring areas that also bear these traits. In other cases, the trait in
question is limited to only a few languages of the Chaco and so does not provide
evidence of the area as a whole.

All the phonological traits that have been mentioned as possibly diffused in the
Chaco are either widely spread beyond the Chaco or are limited to only a few of the
Chaco languages. These include lack of voiced stops; simple vowel systems; voiceless
bilabial fricative \( \phi \) (Gerzenstein 2004); vowel nasalization (Adelaar and Muysken
2004: 499); vowel harmony (Adelaar and Muysken 2004: 499; cf. Gerzenstein and
Gauldier 2003); palatalization (Messino 2003: 36; Messino 2011 limits it to
palatalization of coronal consonants); glottalized (ejective) consonants (only in a few
Chaco languages, Matacoan and Vilela-Lule) (Campbell 2006b, Campbell and Grondona
2012, Comrie et al. 2010: 93-94); uvular (postvelar) consonants (Campbell 2006b,
Other suggested areal traits in the Chaco that are either widely spread in languages beyond the Chaco or found only in some of the Chaco languages, or both, include: prefixing (vs. only suffixing); alienable/inalienable possessive marking (see Comrie et al. 2010; dismissed as not very significant by Campbell and Grondona 2021); unspecified possessor marker (denoting unpossessed forms, Grondona 2003; Campbell 2012b); plural object suffixes on verbs (Grondona 2003); inclusive-exclusive contrast in first person plural pronominal forms; negated adjectives for antonyms (for example, Nivaclé nipitexa ‘short’ [ni- NEG + pitex ‘tall, long’ + -a NEG ]). (See Campbell 2012b.)

8.10 Is the Chaco a linguistic area?

Most of the diffused traits involving languages of the Chaco do not provide strong support for a Chaco linguistic area, particularly not as a LASS. Most are found extending also into languages beyond the Chaco, and others are characteristic of only a few of the languages within the region. Only a few of the traits seem true of a majority of Chaco languages, but none of these is unique to the area and some are quite commonplace in the world, for example SVO word order. This raises the question about whether or how a Chaco Linguistic Area might be defined. Tupí-Guaranían illustrates the problem. Opinion has diverged about whether Tupí-Guaranían languages should be considered members of a Chaco linguistic area, and no grounds for excluding Chiriguano and Tapiete have as yet been found (see Comrie 2010: 86 for discussion). Tupí-Guaranían shares most of the Chaco traits just listed. Since Tupí-Guaranían extends far beyond the Chaco region, though it has representatives also in the Chaco, its inclusion in a Chaco linguistic area would extend the linguistic area way beyond the geographical extent of the Chaco region – hardly a “Chaco” area if defined in that way. If Tupí-Guaranían is included, since these languages also share many traits with languages of the Amazonian area (see Campbell 2006b, Campbell and Grondona 2012 for details), how could we establish what belongs to the Chaco linguistic area and what to the Amazonian linguistic area and how the two areas are distinguished? If Tupí-Guaranían is not included, the areal definition of the Chaco as defined on the basis of shared traits is compromised, since many of the traits seemingly reflective of a Chaco area are also found in neighboring languages well beyond the Chaco region. This sprawling, overlapping, or intertwining of shared traits among Chaco languages and languages of the Andes, of Amazonia, and elsewhere complicates any attempt to define a Chaco linguistic area with recognizable boundaries. In short, the Chaco is a good example of a TSA, but does not qualify as a LASS – in fact hardly any proposed linguistic area qualifies well as a LASS.

Interestingly, Comrie et al. (2010: 89) declare their faith that the Chaco is a linguistic area by acknowledging but dismissing the complications just discussed. They say:

From our theoretical perspective, the existence of phonological, grammatical or lexical features in the languages of the Chaco that are repeated in other areas of South America does not invalidate the characterization of the Chaco region as a linguistic area. The identification of linguistic areas in other geographical regions of South America has shown that often the traits listed are very general and are found widespread throughout South America, see for example the list of traits of Amazonian languages in Dixon and
Aikhenvald (1999). Therefore, not only is the identification of common traits important, but also their overall manifestation in a given area, in non-related languages (Aikhenvald and Dixon 2006).³

I take this statement not as a defense of the Chaco as a true linguistic area, but as a declaration of interest in shared traits found in and around Chaco languages, no matter what the ultimate geographical distribution of the various traits might be – that is, I take this to be an endorsement of the argument put forward here that what is important are the diffused changes themselves, not attempts to set up linguistic areas defined by the overall reach of these traits. That is, I take the statement as a tacit acknowledgement of the Chaco as a TSA, only. More precisely, Comrie et al. (2010: 125) conclude that the Chaco is a core-periphery (núcleo-periferia) type linguistics area (see also Campbell 2006), with Guaicuruan and Matacoan languages at the core and Vilela (Lule-Vilelar) and Tapiete (Tupí-Guaranían) on the periphery. This view, unfortunately, suffers from the fact that even the so-called core languages do not have several of the traits talked about as areal ones, as well as the acknowledged inconvenience that most of these traits extend far beyond the Chaco to languages of other areas. The notion of a core-periphery type language area is an interesting one, but does not seem to have overcome the problems for defining linguistic areas. Since it is not clear that the “core” hangs together in this case, we appear to be left with a lot of trait sprawling, a TSA at best.

There does appear to be considerable diffusion of structural traits involving Chaco languages, but these do not come together in such a way as to suggest a cohesive geographical area. Rather, they show varying linkages with languages and regions outside the Chaco on all sides, while at the same time often not linking all Chaco languages. This is not a surprising finding, since the Chaco as a cultural area is also not distinguished clearly from surrounding regions, and Chaco groups underwent cultural influences from many directions (cf. Métraux 1946: 210).⁴ The linguistic traits shared

³ ?Desde nuestra perspectiva teórica, la existencia de características fonológicas, gramaticales o léxicas en las lenguas del Chaco que se repiten en otras áreas de América del Sur no invalida la caracterización de la región chaqueña como área lingüística. La identificación de áreas lingüísticas en otras regiones geográficas de América del Sur ha mostrado que muchas veces los rasgos señalados son muy generales y se encuentran extendidos a lo largo de América del Sur – véase, por ejemplo, la enumeración de rasgos de las lenguas amazónicas en Dixon y Aikhenvald (1999). Por lo tanto, no importa sólo la identificación de los rasgos comunes, sino su manifestación en conjunto en un área determinada, en lenguas no emparentadas (Aikhenvald y Dixon, 2006).” [From our theoretical perspective, the existence of phonological, grammatical or lexical characteristics in the languages of the Chaco, which are also to be found in other areas of South America, does not invalidate the characterisation of the Chaco region as a linguistic area. The identification of linguistic areas in other geographical regions of South America has shown how often significant features are very general and can be encountered throughout South America – consider, for instance, the list of features for the Amazonian languages in Dixon and Aikhenvald (1999). Therefore, it is not the identification of common features which is important, but their occurrence in a specific area in languages which are not genetically related (Aikhenvald and Dixon, 2006), translation, RH].

⁴ As Raymond Hickey pointed out to me, the topography in the case of the Chaco region also favors the spread of cultural traits and linguistic traits in and from various directions. Since there are no sharp geographical features that set the Chaco off from neighboring regions, presumably movement on dry lowland plains of peoples with contact among their cultures and their languages
among and beyond the Chaco languages appear to parallel the distribution of the cultural traits. This evidence is too weak to declare a Chaco linguistic area *sensu stricto*, a LASS. However, the sprawling of trait among Chaco languages and beyond is a good example of a *trait-sprawling area* (TSA), the concept advocated here to avoid definitional problems with the concept of linguistic area, focussing on actual diffused traits and on answering the question, “what happened?”, rather than on fixing a geographical area to be defined by its shared linguistic traits.

9 Conclusions

So, what does all this mean? The goal should be to understand the changes themselves, in particular in this instance the changes diffused across languages. It is not important nor practical to try to force the various sprawling, overlapping, and intertwining diffused traits to fit together into coherent geographical spaces with clearly defined borders. If we succeed in finding out what happened, we will know which changes are due to borrowing and which to other factors, and we will know how the changes are distributed among the languages involved. Whatever geographical patterning there may be will fall out naturally from this fuller historical account. There is no longer any need to continue to struggle over the intractability of defining the concept ‘linguistic area’ and the specific ‘linguistic areas’.

References


is easier, not as impeded as would be the case if mountains or other geographical barriers were involved.


Messineo, Cristina, Javier Carlol y Harriet Manelis Klein. 2011. Los determinantes demostrativos en las lenguas guaycurúes y mataguayas (Región del Gran Chaco. Memorias del V Congreso de Idiomas Indígenas de Latinoamérica, 6-8 de octubre de 2011, University of Texas Austin.


2 Areas and universals

Balthasar Bickel

1 Introduction

In explanations of how linguistic structures are distributed in the world, the pendulum has swung back from an emphasis on universals, that has dominated the second half of the twentieth century, to a renewed emphasis on local developments and areal diffusion. This shift in emphasis started over twenty years ago, with Dryer (1989), who drew attention to large-scale diffusion as an important possible confounding factor in the statistics of universals and with Nichols (1992), who set out to test universals but instead discovered an intriguing set of large-scale areal patterns. The general shift has gained further impetus with Maslova (2000), who raised the possibility that the current distribution of linguistic structures does not so much reflect universal tendencies but rather accidental skewings during the early developments of language, including skewings due to areal diffusion. Dediu and Levinson (2013) even suggest that the current distribution might reveal signals from ancient contacts with now-extinct sister lineages (Neanderthals, Denisovans and perhaps other unknown lineages) in Eurasia and the Sahul area.

The shift in emphasis has considerably challenged prospects for the discovery of universals. It has become clear that the sheer frequency of some patterns in the world does not in itself suggest effects from a universal principle of language: a pattern can be dominant in frequency worldwide just because it spread around the world through historically contingent cases of areal diffusion. For example, the fact that non-verb-initial languages are far more frequent than verb-initial languages does not necessarily suggest that the former are preferred as a matter of principle (e.g. because of a processing principle). The observed frequency distribution could just as well reflect a multitude of far-reaching but accidental spreads in the distant path (e.g. expansion of verb-final structure in the Eurasian steppe, spread of verb-medial order in Southeast Asia in the wake of Tai migrations, spread of verb-medial order together with the Bantu expansion etc.).

But how can we assess and demonstrate such kinds of areal diffusion? The problems here are by no means smaller than the challenges facing research on universals. First, the repeated spread of some pattern might itself be driven by universal forces: if a structure is universally preferred by some principle (e.g. by processing ease), we would in fact expect it to diffuse in contact situations more easily than universally dispreferred structures. For example, there is growing evidence that simple morphology, e.g. in case marking, is favored by adult learners (Bentz and Winter, 2013, summarizing earlier work), and there is tentative evidence from Artificial Language Learning experiments that certain word order patterns are more easily acquired by adults than others (Culbertson et al., 2012). Findings like these imply that some patterns spread more easily in the wake of second language learning, i.e. in contact situations, than other patterns, possibly resulting in large-scale areal spreads.

1 My thanks go to Johanna Nichols and Raymond Hickey for helpful comments on an earlier draft of this chapter. All shortcomings are my own.
A second problem is similar in kind to the possible fallacies when interpreting worldwide frequencies as universal preferences: if a pattern is more frequent in an area than outside it, this can be attributed to diffusion in contact only to the extent that we can be sure that the pattern did not arise many times independently because of some universal principle. For example, dependent marking on arguments (by case or adpositions) is known to occur with increased frequencies all-over Eurasia, with only few exceptions in Southeast Asia (Bickel and Nichols, 2009). This does not by itself establish that dependent marking spread in this area: it is equally possible that dependent marking is favored in the area independently in each family. This could be caused by the fact that dependent marking here is mostly embedded in verb-final sentence structures. These structures are expected to favor dependent marking universally (Greenberg, 1963; Nichols, 1992; Siewierska, 1996) because this kind of marking makes sentence processing (Hawkins, 2004) and/or meaning transmission more efficient (Gibson et al. 2013; Hall et al. 2013).

Obviously, none of these problems can be solved by keeping the pendulum of emphasis active, sometimes attributing frequency patterns to areal diffusion, sometimes to universal trends. What is needed instead is an approach that focuses on the interaction between area formation and universals, seeking to identify the extents to which the two explain how linguistic structures are distributed in the world. This chapter discusses some possible ways of making progress in this question, drawing mostly on the theoretical framework of Distributional Typology (Bickel, 2015).

In the following, I begin the discussion by exploring the kinds of processes that lead to area formation and universal patterns (Section 2). This leads me to suggest ways of distinguishing the statistical footprint of the relevant processes (Section 3). Section 4 illustrates the methods via recent case studies, and Section 5 concludes the chapter.

2 Functional and event-based triggers of language change

As already noted, one of the key challenges in understanding areal patterns is that they need not result from historical contingencies but can just as well reflect universally preferred patterns of diffusion. The traditional opposition between areas and universals fails to capture the underlying processes and causes here.

As an alternative conceptual opposition, genetics offers the contrast between ‘horizontal’ and ‘vertical’ transfer, but these notions do not help much either: linguistic diffusion is always a vertical, diachronic process (Croft et al., 2011). When we say that a structure (say, verb-final order) has spread in an area, what is meant is that the languages in this area changed their structure so as to mirror the structure of their neighbors, or that they selectively kept structures that mirror those of their neighbors. Such processes can easily take several generations until their results stabilize in a community, and as Johanson (1992) has emphasized some time ago, these processes are fundamentally based on copying and imitating (calquing). Horizontal transfer in genetics, by contrast, involves the direct transfer of concrete genetic material across synchronically co-existing individuals (as is common in bacteria).2

2 Lexical borrowing would seem more similar to horizontal transfer than structural copying, but even lexical borrowing does not involve the direct transfer of Saussurian signs. Rather, the
Given these conceptual problems, it is advisable to replace the traditional opposition between areas and universals by one that is grounded in the causes and conditions of distributional patterns. Approach this perspective, the main contrast turns out to be between functional and event-based triggers (inducers) of language change.

**Functional triggers** are grounded in the biological/cognitive or social/communicative conditions of language, such as specific processing preferences (e.g. Hawkins, 2004; Christiansen and Chater, 2008) or specific sociolinguistic constellations (e.g. Trudgill, 2011; Lupyan and Dale, 2010) that systematically bias the way linguistic structures evolve. The defining property of functional triggers is that they affect transition probabilities universally, independent of concrete historical events. For example, if it is true that processing principles cause verb-final word order to associate with dependent marking, we expect this to cause a higher probability of languages changing towards than away from this association, and this transition probability is the same in any language, at any time.

But note that such kinds of universally fixed transition probabilities do not always result in universally wide-spread patterns: a functional trigger may be tied to a biological or social condition that has itself a limited distribution in the world. Possible such cases are discussed by Dediu and Ladd (2007) for genetic and Everett et al. (2015) for environmental bases of phonological tone, and by Evans (2003) for specific social structures as favoring specific grammars and lexicalizations in kinship expressions (known as ‘kintax’). To the extent that the underlying functional trigger is indeed real, the probabilities for developing and maintaining the relevant structures (tone, kintax) are universally fixed even if they cannot yield the same results all over the world because the conditions do not obtain everywhere. When the conditions are met, however, functional triggers are expected to yield systematic, replicated patterns. Therefore, such triggers can also be called *principles*.

The expected systematicity of effects also entails that a functionally preferred structure can easily spread when languages are in contact, leading to area formation: if a certain structure, or a certain association between structures, is preferred by processing mechanisms or a certain sociolinguistic setting, there is every reason to expect that when speakers can select between variants in multilingual settings, they select the variant that complies with the preferred structure or the preferred association of structures.

**Event-based triggers** are tied to single historical events, leading to idiosyncratic, once-off changes. A classical example is the RUKI rule of Indo-European which retracted *s after *r, *u, *k, and *i at some point in the history of the family. Event-based triggers are especially prominent in language contact situations, when a structure is copied not because it has a universally high probability of developing, but out of mere fashion. For example, relative pronouns and ‘have’-based perfect tenses appear to have a very low probability of developing, as suggested by their extremely rare occurrence (cf. Comrie and Kuteva, 2013 and Dahl and Velupillai, 2013, respectively). As far as we know, the structures do not seem to be particularly preferred by either processing or social conditions of language use. However, where they occur, they appear to have spread in the wake of intensive language contact events, viz. in Europe during the transition period between antiquity and the middle ages (Hastelmath, 1998; Heine and Kuteva,
Another example is the finding by Bickel et al. (2014) that differential object and differential subject marking are strongly dispreferred worldwide, but spread widely (to different extents) in two hotbeds, once in Southwestern Eurasia and once in Australia. In these cases, event-based area formation rests mostly on the preferred development (grammaticalization) of a structure. But like in other diachronic processes, the key effect can also be the preferred maintenance, rather than innovation, of a structure. Examples of this are ergative case alignment or pronominal gender. These structures have been observed to rarely emerge, but when they do emerge, they tend to persist in areal clusters (cf. Nichols, 1993 and Bornkessel-Schlesewsky et al., 2008 on ergativity; Nichols, 1992 and Bickel, 2013 on gender).

What these examples suggest is that both functional and event-based triggers can lead to area formation. However, because functional triggers come with universally fixed transition probabilities, they can be expected to leave a different statistical footprint in the typological record than event-based triggers. Specifically, functional triggers can be expected to leave the same footprint across geographical regions. Event-based triggers, by contrast, are tied to specific historical contingencies, and so they are expected to leave signals only in single regions.

Quantification of this contrast is challenging because the relevant regions are often unknown. Since Dryer (1989) and Nichols (1992), typologists usually operate with a predefined set of regions (e.g. Africa, Eurasia, New-Guinea/Australia and the Americas, or more fine-grained distinctions). Functional triggers are then expected to leave the same statistical signal across all these regions while event-based triggers are expected to leave statistically different signals in each region. This is the basic state-of-the-art model in which we can test the effects of functional vs. event-based triggers, and I will discuss the relevant methods for this model and case study applications below. Before getting there, however, I wish to elaborate a bit further the challenges posed by the definition of regions.

Humans, linguists among them, are good at detecting spatial patterns. Browsing the World Atlas of Language Structures (WALS, Dryer and Haspelmath, 2013), for example, one can easily discern dozens of potentially interesting clusterings in space. What is not so easy is to tease apart genuine from spurious clusters. Statistical methods are only of limited help here because a statistical correlation can of course itself be spurious (as was recently highlighted for typological data by Roberts and Winters, 2013). What is needed is robust causal theories that motivate specific scenarios of event-based triggers and the distributional patterns that can be predicted from these scenarios, i.e. what are called Predictive Areality Theories by Bickel and Nichols (2006). The key idea of Predictive Areality Theories is that areal patterns are predicted by what is known from social/cultural history and archeology, from language spreads and contact events, and from migration patterns as revealed for example through population genetics. An example is the Predictive Areality Theory of the Eurasian area which has good support in both the historical record of ancient migrations and language spreads and the population genetic record. Together, these records suggest multiple waves of intensive language contact events over at least the past 14,000 – 19,000 years (for a summary of the historical record, see Nichols, 1998, and for the genetic record, Rootsi et al., 2007).

3 Methods
The discussion so far suggests that the relative impact of functional and event-based triggers can be best evaluated through statistical modeling, following the well-established framework of Generalized Linear Models (cf. Baayen, 2008 and Johnson, 2008 for linguistically-oriented introductions): the distribution of some structure (say a specific word order) is predicted by the interaction of the conditions that are caused by the assumed triggers, for example, conditions of area formation that are caused by specific historical contact events (e.g. Eurasia) and conditions of universal preferences that are caused by functional triggers (e.g. verb-final clause structures favoring certain argument-marking patterns, or social organization types favoring certain ‘kintax’ patterns).

There are various ways of implementing such models, and there is no consensus yet which implementation is most suitable. The literature includes logistic (Sinnemäki, 2010) as well as log-linear (Poisson) models (Cysouw, 2010a; Bickel, 2011), and some authors model areas as random factors (Cysouw, 2010a; Jaeger et al., 2011), others as fixed factors (Bickel et al., 2009; Sinnemäki, 2010). In practice, these choices often have relatively little impact on results. Still, we clearly need more ‘meta-typological’ research of the kind presented in Maslova (2008) and Cysouw (2010b) to resolve the underlying issues.

What is far more pressing, however, is the question of how one can in fact assess whether a hypothesized diachronic process – functional or event-based, alone or through interaction – leaves a signal in the typological record. Synchronic frequency counts can be deceptive because we cannot assume that the distributions of our samples have reached what is known as stationarity (Maslova, 2000; Cysouw, 2011; Dediu and Cysouw, 2013). Stationarity means that when individual languages change types, this is compensated for by other changes elsewhere in the population of languages so that the total frequencies of types remain the same. If this is the case, synchronic frequencies provide a direct and reliable estimate of transition probabilities. But such a situation cannot be assumed a priori for typological data, especially since event-based triggers do not lead to universally fixed and constant transition probabilities (Maslova, 2000). Further, empirical studies on large families have shown that synchronic frequency distributions can indeed sometimes differ markedly from the frequencies that one would expect under stationarity (see for example Cysouw, 2011 for a demonstration of this with regard to word order and adposition patterns in Austronesian).

In response to these challenges, current approaches have moved away from interpreting raw synchronic frequency patterns and now aim at estimating trends in transition probabilities within individual families (from which in turn one can calculate the expected stationary distribution for each family). There is a number of proposals that are currently being explored, but they all share the basic insight that differences and trends in transition probabilities can be estimated from languages that belong to the same family. If we know that some languages belong to the same family (because the family has been demonstrated using the Comparative Method), we know that the properties of these languages have developed through a chain of innovation and retention from a single proto-language (corresponding to what is known as a Markov chain in mathematics). The challenge then is to estimate the development of this chain. Some methods rely on pairs of languages that belong to the same family each, sampled at the same time depth (Maslova, 2004; Maslova and Nikitina, 2007; Dediu and Cysouw, 2013), others rely on exhaustive samples of languages in each family, at variable time depths (Bickel, 2011,
and still others on estimated or reconstructed tree topologies, of variable size and time depth (Dunn et al., 2011; Dediu and Levinson, 2012). Each method has its advantages and disadvantages. Pair-wise methods make the problematic assumption that within an assumed time depth, there was at most one change and no reticulation (Dediu and Cysouw, 2013), but they have the advantage that one needs to sample only two languages per family. Exhaustive-sample methods require more data. This is a seeming disadvantage, but it has the advantage that the method can pick up more signals. Tree-based methods have the highest resolution, but they limit historical models to trees and exclude the possibility of wave and linkage models although such models are often plausible, especially from the point of view of areal diffusion processes (François 2014). Another requirement of tree-based methods is estimates of branch lengths, as these provide the relevant time intervals within which rates of change are calculated. This can be problematic when, as is usually done, branch length estimates are gained from lexical data (cognate replacements), for lexical change is not necessarily a good calibration stick for structural change. After all, lexical stems can be conservative while syntax may at the same time change rapidly, as one would precisely expect in fact for substrate effects (Thomason and Kaufman, 1988).

A key problem faced by all methods is that they require samples from within families. Yet about half of the known families on our planet are represented in available databases by only a single member (either because no other members are known at all or because no relevant data are available). For testing theories with universal scope or large-scale area effects, data from these single-member families can be absolutely critical. The only proposal so far that attempts to solve this problem is based on an extrapolation algorithm (Bickel, 2013).

The basic strategy is to first perform estimates on diachronic biases (differences in transition probabilities) within larger families, using one of the methods mentioned above. This results in an overall estimate on the probability for a family to show a diachronic bias (i.e. a significant difference in estimated transition probabilities) vs. being diverse (with no significant difference in the direction of estimated transition probabilities). For the typological variable of Agent-Before-Patient vs. Patient-Before-Agent order, for example, there is a .69 probability for families to show a bias (in one of the two possible directions) and a .31 probability to be diverse (Bickel, 2013). Extrapolation then proceeds by randomly selecting 69% of the small families (with between one and, say, four languages) and declaring them as stemming from an unknown larger family with a bias. The actual value of the bias is read off the data (e.g. if Basque happens to be picked up as belonging to an unknown biased family, it would be taken to reflect a bias towards Agent-Before-Patient order, since this is the pattern shown by Basque). Allowing for statistical deviates (the sole known survivor of a family may be the odd one out, having undergone atypical development) and re-sampling the extrapolations in order to assess error margins, the algorithm then yields estimates on how many families are likely to have been biased vs. not biased diachronically, and when

---

3 Software implementation is available and documented for all methods. Maslova’s pair-wise method is available via Dediu and Cysouw (2013); my own exhaustive-sample-based familybias method via Bickel (2013). Tree-based methods rely on software developed for genetics, such as BayesTraits (Pagel and Meade, 2013), MrBayes (Ronquist and Huelsenbeck, 2003) or geiger (Harmon et al., 2008).
they are estimated to have been biased, how many families have been biased in which direction.

Another important challenge for all methods is that the statistical signals of diachronic biases is often very weak: functional principles and historical contact events can only ever trigger possible diachronies, but they do not themselves guarantee the emergence of actual developments: the actualization of a possible diachrony is entirely a matter of social propagation within a speech community. This process is fraught with multiple confounds of competing factors, ranging from competing functional principles and competing contact events to matters of social prestige and language ideology in a given community. As a result of this, it often takes many functionally-driven or event-based processes in individuals before the result of this can be detected in the diachronic biases of a language family.

4 Areal signals of Eurasia, hidden behind universals

A number of recent studies have re-assessed the evidence for functional principles that have been hypothesized to underlie statistical universals in the literature. While controlling for areal patterns, some of these studies have revealed evidence for event-based triggers of change that have affected distributions in addition to, but independently of, the relevant functional principles. These studies are interesting for the purposes of the present chapter because they suggest that evidence for event-based area formation can come from universals research. I focus here on two recent case studies that concern the Eurasian area.

Performing log-linear analyses on diachronic biases, Bickel (2015) reports that families within Eurasia are significantly more likely to be biased towards developing and maintaining dependent marking on arguments (specifically, towards formal distinctions between agent and patient noun-phrases of two-argument clauses) than families outside Eurasia. This difference is independent of the equally significant effect that leads verb-final families or sub-families to biases towards dependent marking. In other words, the event-based diachrony that favors the emergence and maintenance of dependent marking in Eurasia is statistically independent of the functional principles that makes verb-final structures easier to process and/or transmit in communication if there is dependent marking than if there is no such marking (Hawkins, 2004; Hall et al., 2013; Gibson et al., 2013).

An earlier study (Bickel, 2011) looked at the Greenbergian correlation between the order of dependent noun phrases with respect to the head noun (left-branching [[NP] N] vs. right-branching [N [NP]]) and the order of patient noun phrases with respect to the verb (left-branching [[NP] V] vs. right-branching [V [NP]]), relying on Dryer’s data in WALS (Dryer, 2005a, 2005b). The study controlled for various continent-sized areal patterns. However, since left-branching NPs are particularly frequent in the core of

---

4 See Sinnemäki (2010) for a case study showing evidence for Southeast Asia and for what Güldemann (2008) calls the ‘Macro-Sudan Belt’ in Africa, while testing a functional theory relating zero-marking and SVO order.

5 When families are split with regard to word order, as is the case for Indo-European or Sino-Tibetan, the study estimated biases within homogenous subgroups of the family. See the original study for details, and Bickel (2013) for theoretical justification.
Eurasia (Masica’s (2001) ‘Indo-Turanian’ area), it is possible that any areal effect on NP order is specifically caused by the difference between diachronic biases within vs. outside Eurasia. The Eurasian peripheries in Europe and Southeast Asia tend to show deviating NP orders (e.g. right-branching dependent NPs in French and Thai, respectively), but as argued in Bickel (2015) for other parts of grammar, deviating patterns like these may represent younger developments than those that shaped the Eurasian area as a whole. Incipient evidence for this possibility comes from the fact that left-branching NPs are attested even in the peripheries, sometimes as the dominant structure (Basque in Europe, Karen in Southeast Asia) and sometimes as an alternative pattern (English left-branching John’s house vs. right-branching the house of John).

Based on these considerations, I estimated family biases in NP orders and performed a log-linear analysis of the association of these biases with VP order and the location within vs. outside Eurasia. The results are reported in Table 1.

<table>
<thead>
<tr>
<th>Area</th>
<th>VP order</th>
<th>Biased families</th>
<th>Ratio of biases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No bias ([NP]</td>
<td>[N [NP]])</td>
</tr>
<tr>
<td>Inside Eurasia</td>
<td>[V [NP]]</td>
<td>3.67</td>
<td>7.93</td>
</tr>
<tr>
<td></td>
<td>[[NP] V]</td>
<td>6.86</td>
<td>45.30</td>
</tr>
<tr>
<td>Outside Eurasia</td>
<td>[V [NP]]</td>
<td>10.61</td>
<td>36.92</td>
</tr>
<tr>
<td></td>
<td>[[NP] V]</td>
<td>37.76</td>
<td>97.42</td>
</tr>
</tbody>
</table>

Table 1: Estimated diachronic biases of NP order in families with right-branching vs. left-branching VPs inside vs. outside Eurasia (data from Dryer 2005a, 2005b; estimation method from Bickel 2013 with extrapolation to isolates and small families).

The statistical analysis suggests that both area and VP order are significant and at the same time independent of each other. This means that the family biases are best modeled by effects both from functional principles preferring harmony in branching direction and an event-based process that increases the number of families that are diachronically biased towards left-branching NPs in Eurasia. This can also be seen directly in the differences between the ratio of left-branching vs. right-branching families in the last column of Table 1. Inside Eurasia, families biased towards left-branching NPs outrank families biased towards right-branching NPs by a factor of almost 54. Outside Eurasia this factor is only about 11. No comparable difference can be observed with biases in right-branching NPs. This confirms the hypothesis that the Eurasian spreads specifically targeted left-branching NPs – not as a structure with a universal preference but as something that happened to be locally attractive for copying, a mere fashion.

---

6 This can be tested using a likelihood ratio test to compare models with vs. without the relevant interaction between variables: AREA × NP ORDER BIAS χ²=4.96, p=.03; VP ORDER × NP ORDER BIAS χ²=112.2, p<.001. The three-way interaction is not significant (χ²=1.79, p=.18), suggesting independence of the area and VP order effects. Family biases in NP order were estimated using the familybias function that is made available at https://github.com/IVS-UZH/familybias for use in R (R Development Core Team, 2013), with default parameter settings. When families are split in VP order, biases are estimated within non-split subgroups, just like in the other case study (see the previous note).
The two case studies reported here show that critical evidence for event-based area formation can sometimes come from research on universal correlations. In both case studies, a hypothesized functional principle is confirmed by looking at estimates of diachronic biases in families. But the functional principles alone fail to explain the distribution sufficiently. In addition, there is evidence that event-based triggers of language change also significantly contributed to the observed spatial distribution.

5 Conclusions

Research on areas and research on universals have long been considered as irreconcilable opposites. The studies surveyed in this chapter have shown that both strands of research require each other. Research on areas is needed in order to control for alternative explanations when studying universals. Conversely, research on universals sometimes provides critical evidence for areal developments in particular regions. Progress in each domain requires a deeper and better understanding of the relevant trigger that causes the observed distributions: in-depth research on the cognitive/biological and social/communicative principles that favor the development and maintenance of specific structures, and at the same time, in-depth research on the historical and population-genetic events that intensified language contact during specific periods in specific regions.

Good causal theories in these domains make clear predictions on likely or less likely pathways of diachronic development. We have now a range of estimation techniques and databases at our disposal that make it possible to test these predictions. Good theories bring with them higher resolution in their predictions and modern estimation techniques require dense samples. This means that for all the methodological and theoretical progress that has been made on the question of areas and universals, the most urgent task remains the empirical groundwork of analyzing ever-larger arrays of languages across the world and making these analyses available in databases.

References


Dediu, Dan and Michael Cysouw. 2013. Some structural aspects of language are more stable than others: a comparison of seven methods. PLOS One 8. e55009.

Dediu, Dan and Robert D. Ladd. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. Proceedings of the National Academy of Sciences of the USA 104. 10944–10949.


Lupyan, Gary and Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS One* 5. e8559.


3 Reassessing sprachbunds: A view from the Balkans

Victor A. Friedman and Brian D. Joseph

1 Basic facts about sprachbunds, in the Balkans and elsewhere

It is almost impossible to talk about the Balkans from a linguistic standpoint and not utter the term “sprachbund” or one of the variously used English counterparts, such as “linguistic area”, “linguistic union”, “convergence area”, or “linguistic league”.1 Indeed, among linguists, one of the things that the Balkans are best known for is being a sprachbund, that is to say, a zone -- a geographic grouping -- of languages with similarities, especially of a structural nature, that are the result of language contact rather than descent from a common ancestor or typological universals.

The Balkan sprachbund, taking in Albanian, Greek, the South Slavic languages Bulgarian, Macedonian, and some of dialects of the Bosnian-Croatian-Serbian-Montenegrin (BCSM) complex, the Eastern Romance languages Aromanian, Romanian, and Meglenoromanian, the co-territorial dialects of the Indic language Romani, and to some extent the co-territorial dialects of Judezmo (brought to the Ottoman Balkans by Jews expelled from the Iberian peninsula), and Turkic (especially West Rumelian and Gagauz), is noted for a large number of “areal features” — first called “Balkanisms” by Seliščev (1925) — covering aspects of phonology, morphology, syntax, semantics, and lexicon. For concise overviews of these Balkan features, one can consult Friedman 2006a, 2006b, 2011a, Joseph 2003, or Joseph 2010, with more details to be found in handbook-like presentations such as Sandfeld 1930, Schaller 1975, Feuillet 2012, Asenova 2002, Demiraj 2004, and in the compendious Friedman and Joseph 2016.

The result of this linguistic convergence is that, in many instances, one can map between Balkan languages simply by taking note of relevant vocabulary differences.2 For instance, many Balkan languages converge in using impersonal (third person singular) nonactive verb forms (involving either mediopassive (MP) morphology or reflexive marking (RX)) with a dative experiencer in the sense ‘X feels like...’ (literally “to-X DAT VERB_NONACT.3SG ...”), as illustrated in (1) for the meaning ‘I feel like eating burek’ (cf. Friedman and Joseph 2014a):

1 The German Sprachbund means literally “language-union”; none of the English terms proposed in its place has really caught on, so we use here the German word as a borrowing into English, and we nativize it. We therefore write it with a lower-case initial letter and form the plural as sprachbunds, not Sprachbünde; in this way, it is like pretzel or other German loanwords in English (plural pretzels, not *Pretzeln). As is seen below, just as English pretzel has similar but not identical connotations to its German source, so, too, our understanding of sprachbund is not the literal translation from German that has disturbed scholars such as Stolz (2006).

2 We adhere to a distinction between the purely geographic designation “language of the Balkans” and the more specifically contact-affected and structurally and lexically convergent “Balkan language”; thus Croatian is a language of the Balkans without being a Balkan language, whereas Macedonian is both.
Similarly, several of the languages converge in the order of elements marking negation, future tense, mood, and argument structure in the verbal complex, as in (2):

(2) Macedonian ne ḱe (da) mu go davam⁵
    Albanian s’ do tē j+ a jep
    Romanian nu o să i+ l dau
    Dialectal Greek: ñē ñē na tu to ñōso⁶
    'I will not give it to him.'

This word-by-word or even morpheme-by-morpheme “translatability” between languages is what led Jernej Kopitar (1829: 86) to famously characterize the Balkans as an area where “nur eine Sprachform herrscht, aber mit dreyerley Sprachmaterie.”⁷ Similarly, Miklosich (1861: 6-8) remarked on the convergence by noting such features as:

(3) a. future with ‘want’+ infinitive⁸

---

³ This construction does not occur in standard Modern Greek, hence the example from a northern dialect, one in close contact with Macedonian. The experiencer object is accusative, reflecting the northern use of the accusative for indirect objects, parallel to the genitive in southern dialects of Greek and datives in other languages.

⁴ Albanian and Daco-Romanian here have merged dative and accusative weak pronouns into a single portmanteau word indicated by +; the Greek case is generally labeled “genitive” but it fills typical dative functions.

⁵ In Macedonian da here is non-standard or has a meaning of attenuation.

⁶ We give here dialectal Greek, instead of the standard language, so as to be able to cite an invariant third person form of ‘want’, ñē, as the future marker, offering a more direct parallel with the Macedonian, Albanian, and Romanian markers. Standard Greek has ña (from ñē na) and while ña na (via vowel assimilation) occurs dialectally, ña is not the third person form of ‘want’ making the parallel less direct.

⁷ “Only one grammar dominates but with three lexicons (literally: ‘language material’”).

⁸ Most of these features were also noted by Kopitar. Infinitive here is to be understood only in historical terms. Synchronously, the formulation is ‘future marked by a particle descended from ‘want’ plus finite form (with or without the so-called subjunctive marker’). Miklosich (1861)
b. lack of infinitive, with replacement by a finite verb plus a conjunction

c. merger of genitive and dative

d. the un-Romance postposing of the definite article

It was Trubetzkoy, writing in 1923, who took such observations and coined the notion of _jazykovoj sojuz_ ‘linguistic union’ in Russian and the German _Sprachbund_, in his more famous 1928 formulation (Trubetzkoy 1930).

In subsequent years, the features in Miklosich’s (1861) account were significantly expanded and elaborated, especially in Sandfeld’s 1930 classic study, Asenova 2002, and other works, and further features were taken as characteristic of the Balkan sprachbund, as given in (3, continued):

(3) e. replacement of conditional by anterior future

f. object reduplication (proleptic use of clitic pronouns)

g. simplification of the declensional system

h. replacement of synthetic comparatives by analytic ones

i. development of a perfect using the auxiliary ‘have’

j. the so-called narrative imperative

k. evidential forms or usages

l. certain types of dative subject constructions

m. shared lexicon from Turkish, Romance, Slavic, Greek, and a presumed ancestor of Albanian.

The isoglosses for these and other important features are complex, e.g., there are various remnants of person marking for some future markers; remnants of the infinitive survive to varying degrees; the conditional meaning of the anterior future extends into the BCSM of Montenegro and Bosnia; the postposed definite article is absent from some regions that have the other features; evidentials sometimes come from perfects and sometimes from futures, and so on (see Joseph 1983, Hamp 1989, Friedman 2006b, Greenberg 2000, Belyavski-Frank 2003).

---

9 In modern terms, a subjunctive particle. The item in question can also mark independent clauses, generally with a modal sense, as in wishes and polite commands.

10 This is a construction wherein an imperative, a form ostensibly co-indexing a second person subject, can be used with a first or a third person subject, to render narration more vivid.

11 Phonological features do not figure significantly in this list. While there are some shared phonological innovations among co-territorial dialects of different languages, they are too localized, diverse and diffuse to be generalized in the same way as morpho-syntactic features and the lexicon (Friedman 2008, cf. also Hamp 1977 on schwa as a [non-] Balkanism). This difference may be because getting phonology right is less relevant to second-language users, whereas syntactic and morphosyntactic issues are more likely to hinder communicative acts. Moreover, phonology tends to function as an emblematic identity marker. The sort of deeper familiarity with another language that would foster phonological convergence generally occurs at the local level.
The notion of a geographic area characterized by languages that are similar in various ways owing to contact rather than genealogical heritage has been extended to groupings in other parts of the world, including South Asia (Emeneau 1956, Masica 1976), the Pacific Northwest (Beck 2000), Meso-America (Campbell, Kaufman and Smith-Stark 1986), the Vaupés River region in the Amazon (Aikhenvald 2006b, Epps 2006), the Caucasus (Grawunder, this volume; but see also Tuite 1999), Ethiopia (Bisang 2006, but see Masica 2001), and Mainland Southeast Asia (Enfield 2005).

In each case studied, there are languages occupying the same space for a reasonably long time, there are various convergent features, and there is the clear evidence of lexical borrowings showing contact among speakers of the different languages. Putting all that together — especially if there is evidence of divergence in at least some of the languages from earlier stages without the convergent features — one is led to the conclusion that the speaker interactions in the region in question are the reason for the convergence. Thus, one indeed has a geographic grouping of languages showing structural convergences as a result of contact: a sprachbund.

But, there are reasons to think that identifying a sprachbund is not as simple as might appear at first glance. In this respect, Tuite (1999) on the Caucasus is highly instructive. He argues that while glottalization and certain phraseological calques do indeed appear to be areal in the Caucasus, on careful examination the oft-cited feature of ergativity is in fact realized in such different ways that it cannot be taken as a commonality, and that the Caucasus region does not, therefore, represent a sprachbund in Trubetzkoy’s original sense. In section 2, therefore, we outline some key questions concerning the identification of a sprachbund and then turn to a consideration of how one might answer those questions in general in sections 3 through 6, looking to the Balkans in particular throughout.

2 Interrogating the “sprachbund”

Although a well-established concept in contact linguistics and historical linguistics, the notion of “sprachbund” is not without problems. In particular, the following questions are among those that need to be addressed when considering the utility and validity of this construct. These questions pertain to the nature of the languages involved, in (4a), to the nature of the features involved, in (4b), and to the nature of the causes, in (4c). Moreover, there are others involved in the problematization of the notion of “sprachbund” that concern the delineation of the region and the groups, as in (4d), and with the assessment of the construct, as in (4e).

(4) Questions to ask concerning “sprachbund” as a viable notion

a. Language-based issues
i. Is there a minimum number of languages needed to identify a sprachbund?
ii. Must the languages be unrelated to one another? If relatedness is allowed, how closely related can they be?

b. Feature-based issues
iii. Is there a minimum number of features needed to identify a sprachbund?
iv. How should the features be distributed across the languages? Must all features be found in all the languages? Do some features characterize some languages as forming
the “core” of the sprachbund? If so, how does one assess the contribution of the noncore -- peripheral or marginal -- features or languages?

c. Cause-based issues
v. If contact is the basis for the convergence at issue in an area (and not some other causal factor or mere chance), is there a type of contact that is needed to create a sprachbund?
vi. What else might play a role in the formation of a sprachbund?

d. Delineational issues
vii. How do we identify the boundaries of a sprachbund, if any? Are there different degrees of membership, as suggested by the core/periphery distinction in (iv)?

e. Assessment issues
viii. Is the evidence giving a basis for identifying a sprachbund the effects of past sprachbund construction or is the sprachbund an on-going phenomenon?

In what follows, these questions are elaborated upon and some answers given.\textsuperscript{12}

3 Answering the language-based issues

First, to address (4a.i), since language contact is involved in the basic definition of a sprachbund, clearly the minimum number of languages necessarily involved is greater than one, but that provokes another question, namely how much greater? Two languages would necessarily constitute the logical minimum for a sprachbund just as the minimum needed in genetic linguistics for a language family — i.e., a grouping for which contact is irrelevant — is one, as in the case of so-called isolates, languages not demonstrably related to any other.\textsuperscript{13} While for a sprachbund, one has to have at least two to tango, i.e. at least some contact between speakers of historically distinct systems, Thomason (2001:99) is among those who insist that a sprachbund must be at least a ménage à trois, a point to which we return below.

Trubetzkoy (1930), in his formulation of the difference between the Sprachfamilie and the Sprachbund, made no mention of boundaries or numbers. He was attempting both to account for and to distinguish the two diachronic ways languages come to resemble one another, what Labov (2007) has termed transmission and diffusion (cf. also Hamp 1977). As Trubetzkoy recognized, the Sprachfamilie is distinguished by the existence of regular sound correspondences across the member languages, deriving from regular sound changes each underwent, that can be determined using the Comparative Method and found in both grammatical morphemes and core vocabulary. The Sprachbund, on the other hand, as Trubetzkoy defined it, was characterized by shared syntax and

\textsuperscript{12} Much of the material in sections 3 and 4 is adapted from Chapter 3 of Friedman and Joseph (2016), and much of sections 5 through 7 is from Chapter 8.

\textsuperscript{13} For example, Sumerian, Zuñi and Basque, to name a few. Such isolates of course can have internal dialect diversity — quite rich in the case of Basque, cf. Trask (1996) — a situation that stretches the notion of “language isolate” through its intersection with the vexing language-versus-dialect question. Also, known isolates may in fact form a stock or phylum with some other existing language, but such connections are not demonstrable given our current state of knowledge and methodology.
morphosyntax, non-systematic phonological correspondences, and common “culture words” (*Kulturwörter*).

While the regularity of sound correspondences has a predictability that neatly parallels the scientific method, the distinction between *core vocabulary* and *culture words* is not uniform and is susceptible to social manipulation. Thus, for example, in the Pomak dialects of Greece, numerals and basic kinship terms are Turkish loanwords, even though the dialects are clearly Slavic in origin, arguably because the speakers of these dialects view Turkish as having importance to their identity as Muslims.¹⁴ Thus, any given body part or basic verb of motion, feeling, bodily function, etc., representing words that would certainly be part of any lexical “core”, has the potential to be replaced by a loan.¹⁵ In this respect, Romani is illustrative, perhaps precisely because of its massive multilateral contact. Romani has a pre-Byzantine, mostly Indic, core that accords remarkably well with the notional concept in its basics.

The usefulness of the concept of language family is considered to be self-evident since it provides an historical basis for accounting for language resemblances and relations. Still, as noted above, absent evidence demonstrating a relationship of a given language to any other, the existence of a family with only a single member poses no problem to the concept of language family. Similarly, for defining a given language family, it is precisely the shared history of regular sound change combined with notions of core vocabulary and basic grammar that enable us to speak of boundedness, although Thomason and Kaufman (1988) cogently question the rigidity of such conceptions. It can be argued that nineteenth and early twentieth century ideas connected with the need to establish purities of lineage in “races” were carried over to languages as well, whence Schleicher’s (1850:143) characterization of the Balkan languages as “misbegotten sons” [*missrathenen Söhne*] that are “the most corrupt [*die verdorbensten*] in their families” and Whitney’s (1868: 199) characterization of structural borrowing as a “monstrosity”. In a world suffering anxieties about “purity” of race and origin, and one in which political (national) boundaries were in the process of being drawn and redrawn, it is not surprising that such concerns would also permeate academic discourse.

Moreover, the difference between a language and a dialect or the definition of a dialect boundary remains, to some extent, a social or political artifact. A particularly telling example from the Balkans is the way in which conflicting Serbian and Bulgarian territorial claims to Macedonia in the nineteenth and first half of the twentieth centuries were bolstered by the selection of different isoglosses in the South Slavic dialect continuum. Serbian linguists chose the monophonemic (versus diphonemic) reflex of Common Slavic *tj/dj*, while Bulgarian linguists chose the isogloss for the presence of a postposed definite article (see Friedman 2003). Even after the recognition of Macedonian as a separate language within Yugoslavia, Yugoslav linguists continued to treat all

¹⁴ Likewise both Albanian and Bosnian, with their significant numbers of Muslim speakers, use more Turkish kinship terms than co-territorial or neighboring Macedonian, Serbian, or Croatian. On the other hand, Romani dialects in these regions, whose speakers are predominantly Muslim, have native (or Slavic) kinship terms for these relations, arguably as a boundary-marking device. On the other hand, Bulgarian-speaking Christians from Thrace used Turkish numerals like their Pomak neighbors (Kodov 1935). Thus numerals can be highly conservative, but they are also subject to lexical borrowing.

¹⁵ This includes closed word classes that are generally felt to be resistant to borrowing, such as pronouns, complementizers, or conjunctions; see §4 on “ERIC” (conversationally based) loanwords, which defy common assumptions about such resistance.
dialects with /u/ from the Common Slavic back nasal as at least “transitional” to Serbian.

In the case of the sprachbund, however, Trubetzkoy’s intent is sometimes forgotten or misunderstood. He was not talking about any situation of bilingual contact but rather those in which there was a range of similarities in syntax, lexicon, morphosyntax, and even phonology, but precisely without regular sound correspondences and shared core vocabulary. An underlying assumption was areal contiguity, but it is the very nature of areality that raises the question of defining the “area”. Masica (1976:11) writes:

Some [instances of convergence]... involve only two or three contiguous languages. These may merely be instances of what is possibly a tendency for contiguous languages anywhere in the world—or at least contiguous dialects of contiguous languages—to resemble each other in some way or another. Even if every Indian language turns out to be linked to its neighbors by special two-by-two relationships, forming a continuous network covering the subcontinent, this in itself would not establish India as a special area, especially if similar arbitrary [our emphasis VAF/BDJ] linkages continue beyond India....

And Thomason (2001: 99) is explicit about numbers:

The general idea is clear enough: a linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor. The reason for requiring three or more languages is that calling two-language contact situations linguistic areas would trivialize the notion of a linguistic area, which would then include all of the world’s contact situations except long-distance contacts (via religious language..., etc.)... the linguistic results of contact [among more than two languages] may differ in certain respects.

But, as Hamp (1989) had already pointed out with respect to former Yugoslavia, even the Balkans can be understood as part of a “crossroads of sprachbunds”, with “a spectrum of differential bindings, a spectrum that extends in different densities across the whole of Europe and beyond”. A crucial characteristic of the Balkan sprachbund, and by extension, of any sprachbund, is that it is, in the words of Thomason and Kaufman (1988: 95), “messy,” i.e. directionality of the sharing can be difficult to determine. In fact, it is precisely the “messiness” of multiple causation and mutual reinforcement that is characteristic of a sprachbund such as the Balkans. We have enough historical data to know that certain features are innovations in all of the languages for which there are attestations. While hints may occur in this or that earlier stage of this or that language, and typological parallels may exist elsewhere, in the end we have a situation in which languages underwent the same innovation in the same place at the same time, under known conditions of mutual multilingualism. Under such circumstances, attempting to locate a single cause may not be merely futile, it may in fact smack of the same nineteenth century anxieties about purity alluded to above. Thus, for example, Leake

16 A problematic aspect of Masica’s formulation is the assumption that contiguity entails communication. This is not necessarily always the case.
Friedman & Joseph -- page 74 of 393

(1814) attributed the commonalities of the Balkan languages to Slavic, Kopitar (1829), Miklosich (1861) and others to a substratum of one or more of the unattested or poorly attested ancient languages of the Balkans (by definition unprovable), Sandfeld (1930) to Byzantine Greek, and Solta (1980) to Balkan Latin (chiefly). Nonetheless, given the complexity of the evidence, the attempt to identify a single source for all the Balkan commonalities is downright wrong, and the attempt to prove that each innovation arose completely independently in each language strains credulity (see Joseph 1983 on multiple causation, Friedman 2007 on the irrelevance of identifying a single source for each and every phenomenon, and below on the Janus face of genealogical and areal linguistics). A sprachbund is thus a stable situation of mutual multilingualism (cf. also Aikhenvald 2006a), as opposed to, for example, unidirectional multilingualism such as the vertical multilingualism of the Caucasus (Nichols 1997). A bilingual situation in which the two languages were more or less socially equal and stable would indeed be a type of sprachbund situation.

For example, the Geg (northern) Albanian variety spoken in the town of Debar in Macedonia matches the local Macedonian dialect in having no vowel nasalization, a phonological detail which occurs in no other Geg dialect. However, other Albanian dialects northeast of Debar have only one nasal vowel, which is a severely reduced inventory for Geg. To this can be added the fact that the peculiar rounded reflex of the Common Slavic back nasal in the Debar region matches exactly the reflex of the equivalent vowel in Albanian. To exclude Debar Macedonian and Albanian from consideration in the overall convergence of Macedonian and Albanian, and to treat their localized phonological convergence as irrelevant to convergences found across other Balkan languages, would simply be an arbitrary decision rooted in an aprioristic minimum number of languages needed for a sprachbund.

We are faced then really with two problems: the problem of ‘boundaries’ (which subsumes both the territorial implication of ‘area’ and the membership implications of ‘union’ and ‘league’) and the problem of ‘number’. These problems are reminiscent of the difficulties in defining concepts such as ‘nation’, ‘empire’, ‘state’, ‘ethnicity’ or ‘culture’ as well as ‘language’, ‘dialect’, ‘pidgin’, and ‘creole’. What level of control constitutes a ‘state’? How big must a ‘state’ be to be an ‘empire’? How many ‘nations’ must it comprise? When is the speech of a community a ‘dialect’ of another ‘language’ and when is it a separate ‘language’? What is ‘separate’ and how much intelligibility is required before it is ‘mutual’? From a general theoretical point of view, it does not actually seem that the kind of diffusion that takes place among three or more languages is in any way qualitatively different from diffusion that is possible between two languages. For a speaker of a given language -- and the role of speakers in language contact must never be ignored -- diffusion involves the entry of an extra-systemic feature into their existing system: whether a feature of system X that comes into system Y was original to Y or entered Y from another system Z is irrelevant to the reality of the presence of that feature in X and Y (and possibly Z); in either case, contact and diffusion between X and Y were responsible for the shared presence of that feature. Thus, it is important not to

17 In vertical multilingualism, people in higher villages know the languages of those down the mountain, but those in the lowlands do not bother to learn highland languages.

18 Haugen (1966) remains a classic account of this issue and should be read by those who write about the Balkans regardless of their field.
confuse a methodological issue for linguists with the realities of a contact situation for speakers.

Moreover, contact phenomena are never arbitrary. They are embedded in social relations as well as the structures of the languages that manifest them. In a sense, the village of Kupwar in Maharashtra state in India (Gumperz and Wilson 1971, also Masica 1976: 11), with its convergence among Marathi, Urdu, and Kannada, is a linguistic area, albeit one that is part of a larger area, just as the dialects spoken in it are parts of larger languages. If Marathi, Urdu, and Kannada can be shown to converge over a broader region at least to some extent, then we can say that in a sense, Kupwar is the Debar of India (or vice versa).

Trubetzkoy’s original motivation for the terminological distinction between Sprachfamilie and Sprachbund means that two languages related by diffusion can constitute a Sprachbund just as two languages related by transmission from a common source can constitute a Sprachfamilie. The crucial difference is that the Sprachfamilie can involve an isolate, while the Sprachbund by definition requires more than one member for diffusion to take place. There is, however, another issue in the definition of a Sprachbund as understood by Thomason and Kaufman (1988: 95) — who, unlike Thomason 2001 (see above), do not explicitly impose a tripartite requirement on the concept — namely that of directionality. It is generally agreed that in bilateral language contact situations, there is usually asymmetry in the direction of transfer. As argued above, if two languages were demonstrably genetically different enough that similarities resulting from diffusion could be identified but not the directionality (or if the directionality is symmetrical), such a unit could arguably be described as a sprachbund. On the other hand, a situation such as that which we find in the Balkans clearly involves diffusion, but directionality can be variable. For our purposes here, determining directionality is desirable but not requisite, and it can even be argued that it is ultimately irrelevant (cf. Iliievski 1988 [1973] on internal versus external factors). Moreover, in the end the size question does not affect the Balkans — regardless of any minimum considered necessary for a sprachbund, the Balkans, with up to six distinct contributing language groups (Albanic, Indic, Hellenic, Romance, Slavic, and Turkic), would qualify.

As to question (4a.ii) regarding relatedness, these six contributing language groups constitute only two different language families, Indo-European for the first five and also Turkic.19 Does the relatedness of five of the six groups vitiate the sprachbund in this case? Most assuredly not, if the evident convergence is there and is contact-related. The task of identifying the source of features in a given language potentially involved in a convergence is complicated by the possibility of (genealogical) inheritance being responsible for the feature occurring in two related languages rather than diffusion (contact), but this methodological issue for linguists — ease of analysis — should not be equated with realities for speakers.

4 Answering the feature-based issues

Points similar to those in section 3 can be made in a response to question (4b.iii), regarding the number of features for determining a sprachbund. All researchers here are

19 Leaving aside here the vexed question of whether Turkic fits into a larger language grouping such as Altaic.
noncommittal: note Thomason’s reference to “some [our emphasis, VAF/BDJ] structural features”. Clearly, the more convergent features one can find, the more compelling the case becomes for a convergence area, but could a single feature constitute the basis for a sprachbund? Here again we can refer to Trubetzkoy’s original idea, which specified convergences at various levels. There is a parallel to be drawn here with a key aspect of genetic (genealogical) linguistics, namely dialect subgrouping. A single highly significant feature, i.e. nontrivial and unlikely to arise independently, can be taken as diagnostic or emblematic for distinguishing one dialect from another and at the same time for establishing subgrouping. The rhoticization of Proto-Albanian *n intervocically to r is perhaps the single most salient feature distinguishing Tosk (southern) dialects of Albanian from Geg (northern), although the isoglosses for the development of stressed schwa from the low nasal vowel, the general loss of nasality, va from original vo in initial position, and ua from ue all coincide with the r/vn isogloss or deviate from it by no more than 20 kilometers.20 Similarly, Joseph and Wallace (1987) argue that the parallelism in the first person singular of the verb ‘to be’ in Latin and Oscan, with both showing a reflex of an enclitic allomorph of a strong form *esom, even though a single innovation, is diagnostic of the existence of an Italic branch within Indo-European subsuming these two languages of ancient Italy.

Moreover, simply toting up the presence of some set of features and, based on that, scoring the languages as to their degree of membership, as in Campbell, Kaufman and Smith-Stark (1986), Reiter (1994), Haspelmath (1998), van der Auwera (1998a), and Lindstedt (2000), fails to provide an accurate picture of a linguistic area. When applied to the Balkans, complex phenomena are treated as unitary, so that the facts ‘on the ground’ disappear from view. And what it means to “count” a feature as present in some language is far from trivial. For instance, the ‘feels-like’ construction in (1) is absent from Standard Modern Greek and occurs only in some dialects, specifically those in areas where the majority of the population spoke Macedonian, Albanian, and Aromanian into the mid-twentieth century. Does Greek count as a language with this construction? Yes and no. Yes, in that it is found in some variety of Greek, but no, in that it is not widespread across all of Greek. How would that fact, and the fact of the specific dialect distribution of that feature within Greek, be reflected in the scoring? A binary assignment of 1 (presence) or 0 (absence) would not reflect the facts well, nor would a percentage-based score, e.g. 0.1, to signal presence in a small percentage of dialects. And in any case, does the absence across most of Greek render the Kastoria Greek use of that construction any less significant for the speakers of that dialect? We think not.

Even if in principle one should pay attention to a single feature, it is certainly true that the more features one can conclusively identify as convergent due to contact, the stronger the case is for a sprachbund. Nonetheless, when speaking of features we cannot really give a quantifiable threshold, a magic number metric for how much shared vocabulary or how many shared features determine a sprachbund. In this regard, areal linguistics is like its Janus-twin of genealogical linguistics: the criterion for relatedness is simply that the systematic and correspondent similarities are too many to be a

20 On the level of morphosyntax, there are Tosk dialects with infinities marked with me, usually thought of as a Geg feature. In this case, the me-marked infinitives are best viewed as reflecting the proto-Albanian infinitive (so Altimari 2011). This helps support an argument that while regular sound change is genealogically diagnostic, morphosyntax can be more problematic.
coincidence, but attempts to quantify what would constitute coincidence, e.g. Ringe (1992), have been problematic and are not widely recognized as valid. 21

In the case of the Balkans, for reaching any conclusions about the sprachbund, we also have two advantages. First, there is historical documentation for long stretches of time for most of the languages. 22 Second, there is comparative evidence in the form of related languages, both within the most immediate genealogical groups (e.g. for comparisons with Balkan Romance, not only Italian within Eastern Romance, but beyond that, within Romance more generally, also Spanish and French as well as the non-Balkan dialects of Judezmo) and across more distant relatives (e.g. Celtic and Germanic for comparisons involving the Indo-European branches in the Balkans). There is also a qualitative side that cannot be ignored in the assessment of any given feature.

Related to the issue of the number of features is the question (see 4b.iv) of what distribution of features is needed in the group to permit classification as a sprachbund. In particular, must the features identified as diagnostic be present in all of the languages? Belić (1936) and Mladenov (1939) adduce the piecemeal geographic distribution of some Balkan features as a problem for the sprachbund construct (so also Birnbaum 1968). The discussion above in section 3 addresses that criticism, since a cluster-based approach means that features need not be widespread to be relevant. 23 Smaller convergence areas that nonetheless overlap, when taken together, determine more extensive geographic zones in which convergence is to be found. The convergent phonological features of Debar Albanian and Macedonian, for instance, when joined with other features the languages share, obviously form a wider area of convergence involving these two languages. So too, the Kastoria Greek convergence with Albanian, Macedonian, and Aromanian seen in (1) overlaps with the distribution of the ‘want’-based future, found across all of Balkan Greek, and thus adds to the strength of the sprachbund as far as these languages are concerned.

As with other issues, here too a comparison with genealogical linguistics is instructive. According to Bird’s (1982) compilation of the distribution of roots reconstructed for Proto-Indo-European in the various branches of that family – Bird operates with 14 such branches – only one root, *tēu- ‘swell’, is found in all 14 branches. Moreover, there are only eight roots that occur in 13 branches. The number of nonisolated roots increases as the threshold for distributions decreases, so there are 28 roots attested in 12 of the branches, and so on. Taken in this light, the absence of postpositioning for the definite article in Greek, for example, is much less important than its absence in the non-Torlak dialects of former Serbo-Croatian. Similarly, the distributions of ‘have’ and ‘want’ futures take on different meanings in different geographic contexts. The point is that it is not the absolute totality of features that all

21 Chang, Cathcart, Hall and Garrett (2015) show that certain more sophisticated and refined statistical methods, while still not replacing what traditional methods show, offer promise of enhancing historical understanding in situations where ancient records are lacking.

22 The historical record for Greek begins in ancient (BCE) times, and so also for Indic (in the form of Sanskrit) and Romance (in the form of Latin); Turkish is attested from the eighth century and Slavic from the ninth-tenth centuries, and Albanian shows up in extended texts in the fifteenth century CE.

23 Cf. also articles such as Steinke (1999) and Reiter (1999) and the discussion of defining features mentioned in §1 (and in footnote 8). Note also Masica (1976), whose approach to mapping South Asian features involves selecting a few morphosyntactic features and mapping them in all directions as far as possible.
languages share but rather the cumulative effect of smaller convergence zones that justifies the concept of sprachbund.

Having determined that distribution need not be uniform and that the quest for an absolute minimum of determining features is unrealistic without a qualitative assessment of each feature, the next question is whether certain types of features are more relevant than others. The methodological issue of the sprachbund as a consistently definable unit is that despite the parallel first drawn by Trubetzkoy (1923) between the genetic linguistic family and the areal linguistic league, the manner of selecting the correspondences used to define the latter has not been systematized. Contact phenomena, however, do not have the type of systemic invariance found in regular sound change or shared morphology, the bedrock of demonstrable genealogical origin. Contact-induced change, by its very nature, involves a complex ecology of choices among competing systems (cf. Mufwene 2001). Trubetzkoy recognized all types of features, and earlier works, e.g., Miklosich (1861), gave prominence to the lexicon while more recent works, e.g., Thomason (2001: 100), give primacy to structural commonalities. Within the group of non-lexical features taken to be more important, calques are especially significant.24 Campbell et al. (1986) use evidence from calques and shared metaphors in Meso-American languages to argue for a Meso-American sprachbund, since some degree of bilingualism is needed for calquing to occur and to spread (cf. also Ross 2001 on metatypy).25 Likewise, especially crucial are intimate borrowings and in particular conversationally based loans — “ERIC” loans (those that are “Essentially Rooted In Conversation”) as defined and discussed in Friedman and Joseph (2014b, 2016: Chap. 4). ERIC loans include closed classes and generally borrowing-resistant items including kinship terms, numerals, and pronouns, conversationally based elements such as greetings, idioms, and phraseology, and discourse elements such as connectives and interjections. These are important as they are precisely the lexical items that depend on – and thus demonstrate — close, intimate, and sustained everyday interactions among speakers. They give direct evidence of communication between speakers that is not “object” oriented, not purely aimed at satisfying the needs one speaker may have.

Since contact is involved, except for cases of shared retentions, defining a sprachbund presupposes an innovation and thus a drift away from a prior state and toward a state resembling that occurring in another language; in this way, sprachbund phenomena typically involve both convergence on a new type by two or more languages but concomitant divergence from earlier types (possibly preserved in genetically related languages outside the sprachbund) as well. In numerous oral presentations, Andrei N. Sobolev (University of St Petersburg, Marburg University) has claimed that the definition of Balkanisms is circular: Balkan languages have certain features and those features constitute the Balkan sprachbund. This is an unfair, even inaccurate characterization. One must begin with the fact that a variety of languages has been spoken in a multilingual

---

24 We call calques “nonlexical” here since they involve the transfer/copying not of specific lexical material but rather of conceptual structures; there is often a lexical dimension too, but it is not requisite, cf. Enfield (2005, 2008) on linguistic epidemiology.

25 Other cases where calques have been used in arguing for a sprachbund include Nuckols (2000) with regard to a Central European area taking in Hungarian, Czech, Slovak, and German, and Gil (2011) with regard to a Mekong-Mamberambo linguistic area (from the Mekong river area in Southeast Asia, and southeast along the Malay archipelago to the Mamberambo river in western New Guinea), where part of the evidence was the shared phraseology of ‘EYE-DAY’ meaning SUN.
environment for a long time. For most of these languages the previous stages are well attested. If we omit this starting point, then it is indeed possible to accuse Balkan linguistics of circularity. However, it is precisely the diachrony at our disposal that enables us to identify convergent features among the Balkan languages. This is Ilievski’s (1988 [1973]) point, alluded to above, when he argues that what is important is not the source of convergence but the fact of convergence.

If not all features one is considering must be found in all the languages being assessed, then the next relevant question is to query, as in (4b.iv), whether some set of features, and thus some set of languages, constitute the core of the sprachbund, i.e. the most prototypical members as revealed by their incorporation of features x, y, z etc. This is a reasonable question, and would seem to lend itself to the quantitative approaches discussed above, assessing strength of membership by occurrence of features. Many scholars who have addressed the Balkan sprachbund have in fact written in terms of different gradations of “membership” in the sprachbund. Schaller (1975), for instance, classifies Balkan languages as “ersten Grades” (‘of the first order’) and “zweiten Grades” (‘of the second order’), where some features are taken as more telling than others. Schaller’s approach has been seriously criticized (e.g. by Joseph 1987). Not all features are found to the same extent in all languages. Thus, for example, Schaller classifies Greek as a “second degree Balkan language, but Joseph (1983) demonstrates that the loss of the infinitive is not realized uniformly across the various Balkan languages, with Macedonian and Greek both showing the total loss of the infinitive while the other languages show infinitives to varying, often limited, degrees.

However, the cluster approach as envisioned by Hamp (1989) and outlined in section 3 above provides a basis for understanding the variable realization of given features across the Balkans: each feature has its own spatial trajectory of diffusion, and thus it is unrealistic to expect full “compliance”, so to speak, by all languages on all features. The diffusion takes place between speakers, not between languages in some abstract sense, and moreover, it takes place in socially and geographically defined space: this is thus the “speaker-plus-dialect approach” advocated in Friedman and Joseph (2016: Chapter 3.3). Furthermore, given that each feature could have a different point of origin, it is evident that some features will have different distributions from others. Thus the cluster approach in a sense means that no particular feature or set of features is privileged as diagnostic for the sprachbund but rather all convergent features contribute to the sprachbund, each in its own locale. The more locales a feature occurs in, the more salient it becomes to the linguist, though not necessarily to the speaker. Thus, as in other cases discussed already, the needs and interests of the linguist and those of the speaker do not necessarily coincide.

5 Answering the cause-based questions

Turning now to the questions pertaining to cause, the response must start with an acknowledgment that speaker-to-speaker contact is responsible for the diffusion of features in the Balkans, as in any sprachbund, and it is therefore responsible for the

26 “Dialect” is meant in the sense of regional or social dialect, distinguished from the standard varieties of the languages in question, which, as noted below, are too often the basis for comparisons and judgments concerning a sprachbund, at least for the Balkans.
convergence observed therein. But, as (4c.v) asks, it must be considered whether there is a particular type of contact that is needed. Based on what is seen in the Balkans, the relevant contact is not casual contact but intense contact, specifically, in the typical case, multi-lateral, multi-directional, mutual multilingualism. This “four-M” model means that several languages are involved (minimally two, cf. §3 but typically more) and that speakers are multilingual, each speaking some version of the language of others, and overall it is mutually so, in that speakers of language X know language Y and speakers of Y know X, with the result that features can flow in either direction from one language to another. The qualifier “some version of” is important because the speakers are not necessarily perfect bilinguals, but rather have sufficient knowledge of the other language(s) to communicate, and their interlocutors presumably alter aspects of their own usage in the direction of these imperfect speakers. The multilingualism could be stable, or could be a first step towards wholesale language shift, but in either case, an imperfect knowledge of the other language(s) could be involved, knowledge to the extent needed for the relevant communicative acts.

Moreover, as noted in §4 regarding the sorts of features that are significant, what we have defined as ERIC loans give evidence of the intense, mutual, and sustained contact needed for a sprachbund. The ability of speakers of different languages to interact on a regular basis in non-need-based ways was fostered by a particular socio-historic milieu, and we turn to that to answer question (4c.vi), regarding other factors that might figure in the development of a sprachbund.

Based on our available documentation, processes that may have been set in motion, or at the very least begun to be reinforced during the middle ages, and even features that may have appeared in the written record at earlier dates, achieved their current state during the five-century period of control of the Balkans by the Ottoman Turks, referred to as the Pax Ottomanica. Ottoman rule created a political and socioeconomic stability in the central Balkans that allowed for the stable multilingualism necessary for the convergence effects that constitute the “sprachbund”. Some comparisons with other contact situations are especially helpful for pinning down what it was about the Balkans that led to the massive observed convergence.

Heath (1984: 378) presents an interesting view: “It now seems that the extent of borrowing in the Balkans is not especially spectacular; ongoing mixing involving superimposed European languages vs. native vernaculars in (former) colonies such as [the] Philippines and Morocco is, overall, at least as extensive as in the Balkan case even when (as in Morocco) the diffusion only began in earnest in the present century”. Heath’s observations are important for several reasons. On the one hand, the point that significant change can take place rapidly concurs with the view that it was precisely during the Ottoman empire that the Balkan sprachbund was formed. The examples from Morocco and the Philippines, however, all involve lexical items or reinterpreted morphemes rather than morphosyntactic patterns. Moreover, the relationship of colonial languages to indigenous ones is roughly equivalent to that of Turkish to the Balkan Indo-European

---

27 The response that follows draws heavily on Friedman (2014).
28 In Bulgarian this period is referred to as turskoto igo ‘the Turkish yoke’. In a similarly telling example of differences of point of view, the Byzantine Empire is sometimes referred to in Bulgarian as vizantijskoto igo ‘the Byzantine yoke’.
29 On rapidity of change see Mufwene (2004: 203) and Dixon (1997), regarding “punctuated equilibrium”, but on this latter, cf. also Joseph (2001a).
languages at the beginning of the Ottoman conquest. While Turkish did maintain a certain social prestige owing to unequal power relations, there is nonetheless a significant difference between recent European colonial settings lasting a century or so and the five centuries of Turkish settlement in the Balkans during which the language became indigenized and members of all social classes were Turkish speakers. To this we can add that the complexity of indigenous power relations prior to conquest is another part of the picture that is easier to tease out in the Balkans than in European colonies owing to longer histories of documentation of the languages prior to conquest. It is precisely this background of long-term, stable language contact with significant documented history that makes the Balkans an interesting model for comparing and contrasting with other contact situations.

It has been noted above in section 4 that a strictly quantitative approach to the Balkans, or really any sprachbund, is fraught with problems. Further consideration of numbers leads to another telling comparison with the Balkans. Hamp (1977) includes a critique of the conflation of areal and typological linguistics seen in Sherzer (1976) in describing indigenous languages of North America. Among Hamp’s (1977: 282) points is that which he refers to as “gross inventorizing” of what he characterizes as “a Procrustean bed of parameters” (Hamp 1977: 283) which cannot capture the historical depth and specificity that gives meaning to areal developments. Such number games played with a small set of features, characterized by Donohue (2012) as “cherry picking”, can produce maps in which languages seem to mimic modern politics, e.g. Haspelmath (1998: 273), which shows a French-German-Dutch-North Italian “nucleus” to a presumed “Standard Average European”, with the Indo-European Balkan languages at the next level once removed, and with Turkish entirely outside of “Europe”. A subsequent representation (Haspelmath 2001: 107) has only French and German at its core, with Albanian and Romanian as part of the next closest level, Bulgarian and the former Serbo-Croatian entity beyond that, and Turkish still totally outside. Van der Auwera (1998b: 825-827) has dubbed such constructs the “Charlemagne Sprachbund” on the assumption that Charlemagne’s short-lived (800-814) empire, or its successor, the Holy Roman Empire, was the nucleus for a linguistically unified Europe whose influence is detectable today in the mapping out of synchronic feature points. This is, in essence, an extension of Sherzer’s (1976) methodology in Europe (cf. also König 1998: v-vi), and is best seen in the work of the EUROTYP project. To be sure, as with Sherzer (1976), the assembled data are welcome. The over-arching quasi-historical conclusion, however, is misleading and the lack of attention to historical and dialectological detail of the type called for by Hamp (1977) is problematic.

Van der Auwera’s (1998b: 827) formulation that on the basis of EUROTYP’s investigations “the Balkans do indeed get their Sprachbund status confirmed” still gives the impression of the Balkan languages being defined vis-à-vis (the rest of) Europe. But the linguistic realities of the Balkan sprachbund (as identified by Trubetzkoy) took their modern shape internal to the Ottoman Empire — not Obolensky’s (1971) Byzantine Commonwealth — in the regions that were part of the Ottoman Empire from the fourteenth to the early twentieth centuries. This area, during this period, was where, as Olivera Jašar-Nasteva said, with one teskere (travel document) you could travel the whole (Balkan) peninsula.

Hamp (1977: 280) recognizes that areal features “may be crudely labeled Post-Roman European”, but, for example, the spread of the perfect in ‘have’ into the Balkans has nothing to do with Charlemagne. The construction was a Late Latin innovation,
whose origins are already apparent in Cicero and Julius Caesar (Allen 1916: 313), and it made its way into the Balkans with the Roman armies, settlers, and Romanized indigenous populations. It became the preterite of choice — independently — in French and Romanian (except in the south; see Pană Dindelegan 2013: 33), and continues to displace the aorist in other parts of both Western and Eastern Romance. In Balkan Slavic it was precisely those populations in most intensive contact with Balkan Romance that became Aromanian which developed independent ‘have’-perfect paradigms, namely those in what is today the southwest of the Republic of Macedonia and adjacent areas in Greece and Albania (see Goląb 1976, 1984: 134-136 for details). Moreover, it is hardly coincidental that in Bulgarian dialects, it is precisely those that were spoken along the route of the Via Egnatia where similar perfect paradigms developed. As for Greek, as Joseph (2000a and references therein) makes abundantly clear, the use of ‘have’ as a perfect auxiliary is in fact of very different, albeit also Roman, origin. The use of ‘have’ plus a petrified infinitival form as a future in post-Classical Greek — itself a Latin-influenced innovation — gave rise to an anterior future with the imperfect of ‘have’ that became a conditional and then a pluperfect. This pluperfect then provided the model for the formation of the perfect using a present of ‘have’. This stands in stark contrast to the Romance perfect, which began as ‘have’ plus past passive participle, the participle then abandoning agreement, which is exactly the construction that was calqued into Macedonian (and some Thracian Bulgarian). On the other hand, the perfect in the Romani dialect of Parakalamos in Epirus (Matras 2004) is clearly calqued on Greek, as is the very innovation of a verb meaning ‘have’. Albanian also has a perfect in ‘have’ plus participle, and the participle itself is historically of the past passive type found in Romance and Slavic. The directionality is difficult to judge. The Albanian perfect was securely in place by the time of the first significant texts in the sixteenth century — a time when it was still not well established in Greek — but the relationship to Latin or Romance influence is difficult to tease out. Such perfects are not found in the Torlak dialects of former Serbo-Croatian, a region where early contact was likely with populations whose languages are presumed to have been ancestral to Albanian and modern Balkan Romance, and where there were significant Albanian-speaking populations until 1878 (Vermeer 1992: 107-108). The Slavic dialects of Kosovo and southern Montenegro — where contact with Romance lasted into the twentieth century30 and where it is on-going with Albanian, albeit strained — do not show such developments. This fact itself may be due to the importance of social factors in language change. Living in close proximity does not necessarily produce shared linguistic structures. A certain level of coexistential communication must also involve social acceptance. On the western end of the old Roman Empire, Breton is the only Celtic language with a ‘have’ perfect, and the directionality is clear. Still to describe all these perfects as part of a “Charlemagne Sprachbund” is to do violence to historical facts.

The spread of ‘have’ perfects exemplifies linguistic epidemiology in Enfield’s (2008) sense. And thanks to the depth and detail of our historical records, we can tease out the facts. Thus, for example, Donohue (2012) showed that WALS (2005) features for the main territorial languages of Europe, when “decoded into binary format, then pushed

30 According to Rexhep Ismajli (p.c.), when Pavle Ivić was conducting field work on the old town former Serbo-Croatian dialect of Prizren (southern Kosovo) in the mid-twentieth century, he asked a group of old women to count in the old-fashioned way (po-starinski) and they began: unà, dao, trei, patru,... ‘one, two three, four’ (counting in Aromanian).
through computational algorithms (Splitstree) that cluster languages on the basis of "best shared similarity" produce groupings for Germanic, Slavic, Balkan, Romance, and Celtic. These clusters, Donohoe is careful to point out, are explicitly synchronic and not diachronic, and details within groupings are interesting only because we already know the history. Thus, Icelandic and Faroese come out closer to German than to Scandinavian, while Afrikaans is closer to Scandinavian than to Dutch; Polish comes between Belarusian and Ukrainian, on the one hand, and Russian, on the other, while Portuguese is much closer to French than to Spanish. Moreover, the ability to differentiate areal from genealogical causality — what prompted Trubetzkoy’s sprachbund in the first place — is missing. These results demonstrate clearly Hamp’s (1977) point: typological, areal, and genealogical linguistics are independent disciplines, the former achronic, the latter two “twin faces of diachronic linguistics” (Hamp 1977: 279). Nonetheless, despite its many sins of omission and commission (underrepresentation of so-called non-territorial languages, itself a problematic, almost bureaucratic notion, absence of crucial dialect facts, misanalyses, misleading generalizations, etc.), WALS (2005) is a blunt instrument that, if wielded with care and sensitivity, can at least spur us to consider other approaches, as Donohue (2012) has productively done in his discussion of Australia.

In the context of the putative Charlemagne sprachbund, it is instructive to cite here Jakobson’s (1931/1971) concept of the Eurasian sprachbund. Jakobson deviated significantly from Trubetzkoy’s (1930) emphasis on morphosyntactic structure by positing phonologically based sprachbunds, specifically a Eurasian one, concentrating on consonantal timbre (meaning here palatalization) and prosody (meaning here pitch accent or tone), with nominal declension mentioned in a footnote. Eurasia was the center. For nominal declension, Germano-Romance Europe and South and Southeast Asia were the peripheries. For phonological tone, the Baltic and Pacific areas were the peripheries (with West South Slavic, most of Serbo-Croatian and Slovenian, as a relic island), while for palatalization the core was roughly the boundaries of the Russian Empire, with the inclusion of eastern Bulgaria (imagined as Russia’s potential zadunajskaja gubernaja ‘trans-Danubian province’ during the nineteenth and into the twentieth century). He even suggested that palatalization finds its most complete expression in Great Russian [sic], and it is thus no coincidence that it is the basis of the Russian literary language, i.e. the language with a pan-Eurasian cultural mission (Jakobson 1931/1971: 191).

It is also important to remember that, while Masica (2001: 239) warns against confusing “recent political configurations” with “linguistic areas,” it is precisely the legacy of political configurations such as the Ottoman Empire that created the conditions for the emergence of the Balkan sprachbund as it was identified by Trubetzkoy. It is here that the German Bund ‘union’ in Sprachbund (sozuz in Trubetzkoy’s 1923 Russian formulation) has misled scholars such as Stolz (2006), who suggests that since sprachbunds do not have clearly definable boundaries the concept should be discarded. His “all or nothing” methodology misses Trubetzkoy’s original point that the sprachbund is fundamentally different from a linguistic family, and it fails to take into account the basic historical fact that the “boundaries” of a sprachbund are consequences of on-going multilingual processes (Friedman 2012). In Hamp’s (1989: 47) words, they are “a spectrum of differential bindings” rather than “compact borders,” a point also alluded to in Hamp (1977: 282). It is also important to remember that Trubetzkoy first proposed the term at a time when the Sprachfamilie ‘language family’ was widely considered the only legitimate unit of historical linguistics, while resemblances that resulted from the diffusion of contact-induced changes were described in terms as those used by
Schleicher (1850: 143), who described Albanian, Balkan Romance, and Balkan Slavic as “agree[ing] only in the fact that they are the most corrupt in their families.” Trubetzkoy explicitly wanted to avoid the confusion more recently generated by conflations of areal and typological linguistics, although in his time the issues involved areal and genealogical linguistics.

Turning now to language ideology in the Balkans itself, the difference between Greek and other languages is striking. It is certainly the case that multilingualism itself does not guarantee the formation of a sprachbund. As Ball (2007: 7-25) makes clear, in the multilingual Upper Xingu, multilingualism, while necessary for dealing with outsiders, is viewed as polluting, and monolingualism is considered a requisite for high status. This endogamous region is quite different from exogamous ones in other parts of Amazonia, where multilingualism is an expected norm, and lexical mixing is viewed negatively, but morphosyntactic convergence is rampant (Aikhenvald 2006b, Epps and Michael, this volume). Consider also the vertical multilingualism that Nichols (1997) has identified in the Caucasus, which is similar to various Balkan multilingual practices, where specific types of multilingualism index different types of social status.31 Ideologies that consider contact-induced change as symptomatic of pollution and equate isolation and archaism with purity were at work in the nineteenth century too, as seen in Schleicher’s formulation above.

6 Answering the delineational and assessment issues 32

We are now in a position to address the last of the issues by which the notion of “sprachbund” has been problematized. Question (4d.vii) in a sense restates an issue dealt with earlier, namely that of setting the boundaries of the sprachbund and recognizing different degrees of “participation” in the sprachbund on the part of speakers of the various languages involved. The answer emerges from the foregoing discussion: the boundaries are as elastic as the micro-zones of convergence that add up to the larger convergence area. There is nothing fixed, and even political boundaries, while convenient, are relevant only insofar as they correspond to socio-historical realities that might promote the sort of contact necessary for sprachbund formation. Geography is no accident as far as the Balkans are concerned, in that for most of the features recognized as important regarding structural convergence in this region, the more geographically peripheral the language, the less likely it is to demonstrate fully the feature and the more centrally located a language or dialect is in the Balkans, the more fully it shows the feature. This is especially clear in the replacement of the infinitive by finite subordination. Particularly telling from a geographic standpoint are the comparisons in (5), where [+infinitival] means that the infinitive is alive — or remained alive longer — in the language to some (not insignificant) degree and [-infinitival] means that there

31 See footnote 17. Nonetheless, as Tuite (1999) makes clear, aside from shared glottalized consonants and some phraseological calques, a Caucasian sprachbund, when examined closely, vanishes like a mirage. Hamp (1977), too, noted that glottalization in Armenian and Ossetian, respectively, must have distinct areal diachronic explanations.

32 The discussion in this section draws on Friedman and Joseph (2016), especially Chapter 7.7.2 and Chapter 8.
essentially is no infinitive:33

<table>
<thead>
<tr>
<th>(+infinitival)</th>
<th>(-infinitival)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeyka Greek (eastern Turkey)</td>
<td>mainland Greek</td>
</tr>
<tr>
<td>Southern Italy Greek</td>
<td>mainland Greek</td>
</tr>
<tr>
<td>Arbëresh Tosk Albanian (So. Italy)</td>
<td>mainland Tosk Albanian</td>
</tr>
<tr>
<td>Geg Albanian</td>
<td>most of Tosk</td>
</tr>
<tr>
<td>West South Slavic (BCSM, Slovene)</td>
<td>East South Slavic</td>
</tr>
<tr>
<td></td>
<td>(Macedonian/Bulgarian)</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Maleshevo-Pirin, Lower Vardar Mac.</td>
<td>the rest of Eastern Mac. and all of</td>
</tr>
<tr>
<td>former Serbo-Croatian/Croatia</td>
<td>former Serbo-Croatian/Serbia</td>
</tr>
<tr>
<td>Non-Torlak former Serbo-Croatian</td>
<td>Torlak</td>
</tr>
<tr>
<td>Istro-Romanian</td>
<td>Balkan Romance (Aromanian,</td>
</tr>
<tr>
<td></td>
<td>Meglenoromanian, Romanian)</td>
</tr>
<tr>
<td>Romanian</td>
<td>Aromanian/Meglenoromanian</td>
</tr>
<tr>
<td>East Rumelian and Anatolian Turkish</td>
<td>Western Rumelian Turkish</td>
</tr>
<tr>
<td>Modern Indo-Aryan (e.g. Hindi)</td>
<td>Romani</td>
</tr>
<tr>
<td>non-Balkan Romani</td>
<td>Balkan Romani</td>
</tr>
</tbody>
</table>

The generalization emerging from such comparisons is that the more deeply embedded a language (or dialect) is in the Balkans, the weaker its category of infinitive is. Admittedly, there are some exceptions to this generalization, e.g. Cypriot Greek is geographically peripheral but lacks the infinitive to the same degree as mainland Greek, and Geg Albanian is relatively central in the Balkans but has an infinitive. However, Tosk is arguably more central and does not have an infinitive, or at least did not for some stretch of time. While there may be other forces at work in such cases that can help to explain exceptions to the generalization, even with them, it nonetheless holds true in the vast majority of cases and thus provides support for the notion that there is something characteristically Balkan about the loss of the infinitive and its replacement by finite forms.34

Peripherality in (5) is geographic in nature, but there can be chronological peripherality as well. Judezmo offers a striking example here. As noted in Friedman and Joseph (2014b), Judezmo entered the Balkans rather late, only after Sephardic Jews were expelled from Spain and Portugal after 1492, and some of the ways in which Judezmo appears to be non-Balkan can be attributed to this chronological dimension. For instance, Judezmo speakers arrived in the Balkans with a fully functioning preposed definite article so absence of the enclitic Balkan definite article can be explained by the fact that an article had already developed in Judezmo (from a Latin starting point without an article); that is, Judezmo came to the Balkans at a point when, in terms of its own development, the issue of a definite article had already been settled. The situation of Judezmo vis-à-vis

33 See Friedman and Joseph (2016: Ch. 7.7.2) for details on the various languages that inform this table; most of the relevant facts can be found also in Joseph (1983).
34 For instance, contact between Greeks of the mainland and Cypriot Greeks -- note that the Cypriots have a separate word for mainland Greek (kalamaristika), suggesting on-going contact -- may have helped to spread the lack of an infinitive to the Cypriots.
the infinitival developments is another, perhaps even stronger, case in point. That is, although the ultimate loss of the infinitive in some of the languages is late, or has not yet occurred -- Romanian, for instance, preserves the infinitive as an option even in contemporary usage -- it can be localized temporally in the sixteenth or seventeenth century for the languages that lack it most fully, especially Greek, Macedonian, and Tosk Albanian.\footnote{Tosk Albanian has developed a new infinitival construction, e.g. \textit{për të punuar} ‘(for) to work’, apparently composed of a preposition (\textit{për} ‘for’) with a nominalized participle (\textit{të punuar}), but this seems to be a relatively recent development, after a period without an infinitive in the language.} Thus the relative robustness of the infinitive in Balkan Judezmo, as compared to its linguistic neighbors in the Balkans, especially Greek and Macedonian, may in part be due to chronology. That is, the entry of Iberian Jews into the Balkans came at the end of the period when the infinitive was lost in this region. One can speculate, then, that being peripheral to the temporal period most associated with strong loss of the infinitive may have played a role in the survival of the Judezmo infinitive even into contemporary usage.

Another factor affecting the degree of participation in the convergence builds on the observation in section 5 concerning socio-historical factors. That is, there is also a social dimension to peripherality. Once again, this can be seen with Judezmo and the infinitive. As discussed in Friedman and Joseph (2014b, 2016: Chap. 7.7.2), Balkan Judezmo shows contradictory tendencies regarding the infinitive, with both innovative finite subjunctive usage and conservative infinitival usage. The sociolinguistics of Jewish languages provides a basis for an explanation here. Jewish languages in general are likely to preserve archaisms different from those of coterritorial languages (cf. Wexler 1981), and given the local and social segregation of Jewish communities, Jewish speakers would have had less exposure to linguistic innovations found in the usage of coterritorial non-Jewish speakers. The Judeo-Greek of sixteenth century Constantinople, for instance, shows archaic infinitival usage paralleling that of New Testament Greek (Joseph 2000b). Moreover, as documented in Friedman (1995), Jews were linguistically peripheral as shown by the absence of Judezmo from nineteenth century Macedonian code-switching anecdotes, where Jews, who speak Turkish, are the only ethnic group that does not switch into its own language. While it is true that non-Jewish merchants in the bazaars often had some knowledge of Judezmo (and, as with Yiddish, Hebraisms were used as cryptolectal elements in such circumstances; Benor 2009), multilingualism, as with Romani, tended to be unidirectional (Friedman 2000). Thus the persistence of the use of infinitives in at least some Balkan Judezmo varieties seems to be an important reflection of a lesser degree of contact between Jews and non-Jews in the Balkans than among the non-Jewish speakers of various languages in the region. By contrast, Romani was in intimate contact with Greek for hundreds of years beginning no later than the eleventh century or so. And while we cannot know what proto-Romani looked like at that time, Romani is as dependent as Greek and Macedonian on its analytic subjunctive to perform infinitival functions.\footnote{Relatively recently, Romani dialects outside the Balkans have begun developing new infinitival constructions in contact with European languages that do have infinitives (Boretzky 1996).}

Thus peripherality can be measured spatially, temporally, and socially, of which social factors are the most important. Geography can help or hinder such interactions, and
chronology is insurmountable but also instrumental in that speakers can only be in a place when they have come to that place. As the lesson of the *Pax Ottomanaica* shows, political conditions -- which are a macrosocial phenomenon in any case -- can also serve as a contributing factor. In the end, then, social factors, aided and abetted by various other external conditions, determine participation and boundaries.

Andriotis and Kourmoulis (1968: 30) maintain that the Balkan sprachbund is ‘une fiction qui n’est perceptible que de très loin’ and that the commonalities are ‘tout à fait inorganiques et superficielles’. Their view — no doubt influenced by Greek nation-state ideology — addresses the final question, (4e.viii), that is whether the Balkan sprachbund, or any sprachbund that might be identified, is an on-going concern or just an artifact of the past.

Contrary to the challenge posed by Andriotis and Kourmoulis, one can argue Balkan linguistic diversity occurs within the context of a set of structural similarities that constitute a framework of contact-induced change. Moreover, it is important to distinguish between ‘superficial’ and ‘surface’. As Joseph (2001b) argued, surface realizations constitute the locus of language contact, and therefore accounts that are concerned with typological aspects of universal grammar (including formalist descriptions) do not address the mechanisms of language contact. Also, surface realizations are by no means ‘inorganic’; they represent convergences that are evidence of the multilingualism that we know existed for centuries and even millennia, and which thrived and developed into the “Four-M” (multi-lateral multi-directional mutual multilingualism) set of conditions under Ottoman rule.

Moreover, and this may be the real source of Andriotis and Kourmoulis’s position, when one compares just the standard national languages of the Balkans (as does Asenova 2002, leaving out, however, Macedonian), one can get a sense of language states frozen in time, fixed in a form that might reflect a structural state of affairs from several centuries back, affected somewhat by the process of standardization. Or then again, the literary language might reflect a state of affairs that never existed in the spoken language, but has been borrowed from some prestigious language outside the Balkans. As for Greek, for instance, it is instructive to examine Thumb (1895) for the state of the language as recently as the late nineteenth century and to see regular Demotic forms given that exist today only in regional dialects, forced out of the emerging standard language by the onslaught of puristic pressures, overt and covert; for example, *práyma* ‘thing’, currently with a genitive *práymatos*, forms that match Ancient Greek except for the realization of the velar as a fricative, is given as *práma* (still an accepted variant today) but with a genitive *pramátu*, a form found now in such outlying dialects as Cappadocian but not in the standard language. And the effects of lexical purism affecting the ERIC loans are evident in a look through Grannes et al. (2002), where so many of the Turkisms recorded there for Bulgarian are artifacts of the nineteenth century, considered now, in present-day Bulgarian, to be obsolete or else serving only to lend an archaic flavor to literary works. Such is also the case in all the other Balkan languages (Friedman 1996, but see Friedman 2005 on the resurgence of Turkisms in ex-communist countries as a badge of democracy). These examples teach two important lessons.

First, there is the important injunction of Bailey, Maynor and Cukor-Avila (1989: 299) regarding vernaculars and standard languages:

---

37 Aikhenvald (2006b) makes a similar point.
...[T]he history of ... language is the history of vernaculars rather than standard languages. Present-day vernaculars evolved from earlier ones that differed remarkably from present-day textbook-varieties. These earlier vernaculars, rather than the standard, clearly must be... the focus of research into the history of... [languages].

This holds true for the Balkans, and the difference between the crystallized structure of the standard languages, with some Balkan convergent structure intact but not going anywhere, gives a sense of déjà vu to the sprachbund if one looks only at these standard languages. They are based on speech forms that were affected by the intense contact of the Ottoman period, and therefore show Balkanisms. But the “present-day vernaculars” are the regional dialects, so that the second key lesson is that the dialects and not the standard languages are where Balkan linguistic processes continue. The example in (1) of the ‘feels-like’ construction in Kastoria Greek but not elsewhere in Greek is a case in point. On-going contact among speakers of regional varieties of Greek, Macedonian, Balkan Romance, Romani, Albanian, and so on, continue to occur in northern Greece, in the Republic of Macedonia, in parts of Albania and elsewhere in the Balkans, in cities and in rural areas. This contact continues the “Four-M” model conditions (see Friedman 2011b). In that sense, the Balkan sprachbund is alive and developing as contact among speakers continues.

For sprachbunds that have no standard languages and no literary tradition, the vitality of the convergence zone as a living and on-going entity will depend on the health of the languages involved, i.e. of the speakers’ commitment to and ability to use their languages, and the extent to which contact and mutual multilingualism continue. Here the Republic of Macedonia presents a considerably more promising picture than the Hellenic Republic, although both countries are still home to speakers of languages from all the major Balkan linguistic groups. But in principle, a sprachbund is not a relic of the past; under the right conditions, the necessary type of contact renders the processes of convergence on-going.

We see the Balkan sprachbund as crucial to placing the differences in the context of the similarities. That is, there are certain cleavages within the Balkans that are particularly revealing, e.g. between ‘have’-based futures and ‘want’-based futures, or between the presence of systematic marking for evidentiality and the absence of such marking, that yield a crisscross pattern over the whole of the sprachbund area. Paying attention to such features and their distribution offers a more nuanced picture of the sprachbund, one based in part on the recognition that, like all language change, degrees of convergence can take place at varying speeds. And, as with other aspects of the present discussion, social conditions are the crucial factor.

7 Conclusion

The problematization and interrogation of the notion of “sprachbund” offered here comes down to a single issue: Do sprachbunds exist as a distinct entity in the context of contact-induced change?

Our answer is ‘yes’. Insofar as we can speak of ‘language’ as both a linguistic and a social fact, so too can we speak of a sprachbund as a linguistic and social result of a particular kind of language contact. Such contact reflects the effects of intense multi-
lateral multi-directional mutual multilingualism. A sprachbund is no more a linguist’s construct than is a language. Just as a language, in the external sense of a medium of communication shared in a speech community, is the creation of its speakers, so, too, speakers are involved in forging the convergent structures that define a sprachbund. The sprachbund is a well-instantiated and distinctly observable entity shaped by space, time and social conditions, and it is necessarily based on contact between speakers of distinct “languages” in a geographical space. This answer leads to an additional question: Are the Balkans an instance of a sprachbund? Again, the answer is ‘yes’. Whether the region that was once a sprachbund still is depends on whether one focuses on achronic comparisons involving current standard languages or instead looks to see the forces that shaped the standard languages structurally and that are still observable in on-going contact situations outside the standard varieties.

Finally, does a sprachbund represent the outcome of a special type of contact, different from that found in casual borrowing, trade contexts, learned borrowing, or creolization? The answer is still ‘yes’. A sprachbund reflects aspects of contact found in other contexts such as multilingualism, but the type of stable, relatively egalitarian multilingualism we discuss here, with intense and sustained contact on an everyday basis, is arguably diagnostic. The sprachbund, therefore, even if it raises important questions, is a useful construct, which contact linguistics needs to recognize within the overall scope of contact-induced linguistic developments.

References


Emeneau, Murray, B. 1956. India as a linguistic area. *Language* 32.3-16.


Friedman, Victor A. 1996. The Turkish lexical element in the languages of the Republic of Macedonia from the Ottoman period to independence. Zeitschrift für Balkanologie 32.1, 133-150


Friedman, Victor A. and Brian D. Joseph. 2014b. Lessons from Judezmo about the Balkan sprachbund and contact linguistics. International Journal of the Sociology of


Kodov, Hr. 1935 Ezikă t na trakijskite bălgari [The language of the Thracian Bulgarians]. Sofia: Trakijskij nauč institut.


Nuckols, Mark. 2000. *Examination of the case for a Central European Sprachbund of Czech, German, Hungarian and Slovak*. M.A. Thesis, The Ohio State University, Columbus, OH.


Vinožito 2006. *Bukvar/Anagnostiko* ['Primer’ Macedonian/Greek]. Thessaloniki: Batavia.
4 Areal sound patterns: From perceptual magnets to stone soup

Juliette Blevins

1 Defining areal sound patterns

Linguistic areas are geographic regions where languages share characteristics as a result of language contact and not as a consequence of shared inheritance, general linguistic tendencies, linguistic universals, or chance. Though there is controversy over the precise set of grammatical characteristics that can spread via language contact (see Campbell, this volume), recognition and study of linguistic areas in the modern era continues to reaffirm that sound patterns can and do spread in this way (Boas 1911, 1920, 1929; Trubetzkoy 1939; Emeneau 1956; Heath 1978; Aikhenvald 2002). Given this general outline, we can define areal sound patterns as in (1).

(1) Areal sound pattern: a definition

An areal sound pattern is a sound pattern shared by two or more languages in a designated geographic region that: (i) does not result from shared inheritance for at least one pair of languages; (ii) is not a consequence of general linguistic tendencies alone for at least one language; and (iii) cannot be attributed to chance for at least one language. An areal sound pattern results from language contact when speakers of a language that lacks a particular sound pattern come to acquire a sound pattern in their speech from extensive contact with a distinct language that has that pattern.

The notion of sound pattern in (1) is a general one. It includes: overall properties of contrastive sound inventories such as vowel, consonant, and tone inventories; patterns determining the distribution of sounds or contrastive features of sounds, including the distribution of stress, tone, length, laryngeal features, and consonant clusters; and the variable realization of sounds in different contexts that constitute phonological alternations. We will look at cases of each of these types in the sections that follow.

The definition in (1) defines an areal sound pattern as a sound pattern shared minimally by two languages, having arguably diffused from one into the other. Though linguistic areas as large as the Australian continent have been proposed (Dixon 2002),

1 Within some phonological frameworks, like Optimality Theory, linguistic universals are encoded as markedness constraints, while general linguistic tendencies are not (Prince and Smolensky 2004; Kiparsky 2006). In other frameworks, like Evolutionary Phonology, there is no clear categorical distinction between statistical tendencies and phonological universals, with apparent universals exhibiting extreme instances of statistical tendencies (Blevins 2004, 2006, 2009a, 2014). In the remainder of this chapter, I do not distinguish between general linguistic tendencies and linguistic universals, since, in both cases, what is relevant is a strong statistical tendency for a particular sound pattern to occur cross-linguistically, independent, for the most part, of inherited features. For cases where strong statistical tendencies do appear to correlate with other inherited sound patterns, see Blevins (2009b) and Blevins and Grawunder (2009).
looking only at sound patterns, we find many cases of areal sound patterns limited to small geographic areas, and to a small number of languages.²

Within the large and widely dispersed Austronesian language family, an areal sound pattern is found in a small region of central Taiwan as shown in Map 1.

Map 1. Pre-glottalization in a micro-area of central Taiwan

² Macro-areas and micro-areas for sound patterns may be more common than those for other levels of languages for the simple fact that the phonetics of one language is immediately accessible to speakers of another in a language contact zone, while aspects of morphology and syntax are not.
Three geographically contiguous Formosan languages, - Thao, Bunun, and Tsou -, show pre-glottalization of voiced stops /b/ and /d/ as [ʔb] and [ʔd] respectively (Blust 2009:52, 165, 641-2). Blust (2009: 641) suggests that the innovation occurred in Bunun, and spread to both Tsou and Thao. Pre-glottalization is not a feature of Proto-Austronesian, - the only common ancestor from which these three languages descend. Therefore, shared inheritance is ruled out. Assessing whether pre-glottalization of /b/ and /d/ in three distinct Formosan subgroups could be attributed to phonetic naturalness or chance is more complex. If we look at the distribution of pre-glottalization and implosion of voiced stops throughout the Austronesian language family, we find it occurring in approximately 20/1000 or 2% of languages, while cross-linguistically, based on modern sampling of the WALS database, implosives occur in 77/567 or approximately 13.5 % of languages, but in less than 10% of languages that do not also have ejectives or glottalized sonorants (Maddieson 2013a). The chance, then, of three randomly chosen Austronesian languages having this property is .00008, while the chance of any three randomly chosen languages having this property is .001. Pre-glottalization/implosion is also clearly under-represented within the Austronesian family. The occurrence of this feature in 3/24 or 12.5 % of known indigenous languages of Taiwan at a shallow time-depth (see footnote 3), and in a contiguous area, then, does not appear to reflect the general tendencies for these sounds to develop within Austronesian languages, but, as Blust suggests, is best attributed to diffusion of pre-glottalization from Bunun to neighboring languages of central Taiwan.

Since areal sound patterns must be distinguished from sound patterns resulting from shared inheritance, general linguistic tendencies, or chance, the establishment of an areal feature cannot be based simply on the association of a geographically defined region with a linguistic feature that crosses language boundaries, no matter how vast the geographic region is, or how uncommon the linguistic feature is outside of this region. In this context, consider Dixon’s (2002) proposal that the entire continent of pre-contact Australia is a linguistic area, including more than 200 indigenous Aboriginal languages. While this area might be defended on other grounds, proposed areal sound patterns do not meet the criteria set out in (1) because direct inheritance cannot be ruled out. The clearest example involves the contrast between retroflex and non-retroflex apical stops found in

---

3 I follow Blust’s (2009, 2013) Proto-Austronesian (PAN) reconstructions here. However, the possibility that PAN did indeed have (phonetically) preglottalized [ʔb] and [ʔd] that were retained in at least one Formosan language, Bunun, might be considered: Proto-Mon-Khmer is reconstructed with preglottalized stops */ʔb/ and */ʔd/ (Haudricourt 1965; Diffloth 1976) which were maintained in Mon, Palaungic, Katuic, and Bahnaric, but lost in Khmer, Pearic and Khmuic (Haudricourt 1965; Diffloth 1976).

4 Under Blust’s (1999, 2009) classification, Bunun is an immediate daughter of Proto-Austronesian, while Thao is a member of the Western Plains group, and Tsou a member of the Tsouic subgroup. Since no other Western Plains or Tsouic language has pre-glottalization, this feature appears to post-date the diversification of these subgroups.

5 Implosives are found in Bintulu, Lowland Kenyah dialects, languages of the western Lesser Sundas and languages of southeast Sulawesi (Blust 2009: 176, 188). If these additional cases are not included, the figure may be under one percent. However, even in the language count, implosion is likely over-represented, since, as noted by Blust, implosion appears to be an areal feature in the western Lesser Sundas and southeast Sulawesi as well.
more than two-thirds of the indigenous languages. Cross-linguistically, contrastive retroflexion for oral and nasal stops ranges from about 5-7% in the UPSID database sample of 451 languages (Maddieson 1984; Maddieson and Precoda 1989), so the much higher rates of this feature across Australia are suggestive. A further suggestion of areality is the band of Pama-Nyungan languages lacking retroflexes, stretching along the Eastern coast of the continent, from Cape York south, as shown in Map 2 (after Dixon 2002: 66).

![Map 2. Areas on the Australian continent lacking contrastive retroflexion](image)

However, contrastive retroflexion is reconstructed for Proto-Pama-Nyungan (Bowern and Koch 2004), the mother language of approximately 75% of Aboriginal languages. The remaining Aboriginal languages fall into approximately 27 families/isolates. While detailed historical work on many of these is in the early stages, at least some non-Pama-Nyungan proto-languages, like Proto-Tangkic, are reconstructed with contrastive retroflexion as well (Evans 1995; Round to appear). Given these historical proposals, the high frequency of contrastive retroflexion in Australian Aboriginal languages seems due

---

6 Australian languages with a retroflex contrast are often referred to as “double apical” languages in contrast to “single apical” languages that lack it. We return to consonant retroflexion as an areal feature in section 2.
to inheritance or shared inheritance within recognized language families. If any areal feature needs to be explained, it is the band of Pama-Nyungan languages on the East coast that appear to have lost this contrast.

In regions where significant historical work has been done, shared inheritance is usually the easiest factor to rule out as a source of phonological convergence. For example, in assessing whether the occurrence of glottalized consonants in Yurok, an Algic language of Northwestern California, may constitute part of an areal sound pattern in Northwestern California (Haas 1976), or part of a larger identifiable areal pattern of high frequency glottalized consonants on the Northwest Coast of North America (Maddieson 2013a), we can rule out shared inheritance. This areal pattern is shown on Map 3, where only the northernmost language, Alutiq (1), and two southern fringe languages, Karok (48) and Wiyot (51), lack ejectives.

7 In support of this, one can look at other apparent areal features of the Australian continent that are typologically unusual, like the occurrence of non-homorganic NC clusters. These are also arguably inherited, not diffused (Blevins 2004: 209-211).
### Languages of the Pacific Northwest coast with glottalized consonants.

Circles indicate languages without ejectives.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Alutiq</td>
</tr>
<tr>
<td>2</td>
<td>Eyak</td>
</tr>
<tr>
<td>3</td>
<td>Tlingit</td>
</tr>
<tr>
<td>4</td>
<td>Haida</td>
</tr>
<tr>
<td>5</td>
<td>Gitksan, Nishga</td>
</tr>
<tr>
<td>6</td>
<td>Coast Tsimsian and Southern Tsimsian</td>
</tr>
<tr>
<td>7</td>
<td>Haisla</td>
</tr>
<tr>
<td>8</td>
<td>Bella Coola</td>
</tr>
<tr>
<td>9</td>
<td>Hecateuk</td>
</tr>
<tr>
<td>10</td>
<td>Oowekeyda</td>
</tr>
<tr>
<td>11</td>
<td>Kwakwala</td>
</tr>
<tr>
<td>12</td>
<td>Comox</td>
</tr>
<tr>
<td>13</td>
<td>Sechelt</td>
</tr>
<tr>
<td>14</td>
<td>Puntluten</td>
</tr>
<tr>
<td>15</td>
<td>Nootka</td>
</tr>
<tr>
<td>16</td>
<td>Squamish</td>
</tr>
<tr>
<td>17</td>
<td>Nakomelem</td>
</tr>
<tr>
<td>18</td>
<td>Neahklit</td>
</tr>
<tr>
<td>19</td>
<td>Songhees, Saanich, Lummi, Samish, Clallam</td>
</tr>
<tr>
<td>20</td>
<td>Makah</td>
</tr>
<tr>
<td>21</td>
<td>Quileute</td>
</tr>
<tr>
<td>22</td>
<td>Nooksak</td>
</tr>
<tr>
<td>23</td>
<td>Lushootseed Salish</td>
</tr>
<tr>
<td>24</td>
<td>Chemakum</td>
</tr>
<tr>
<td>25</td>
<td>Twana</td>
</tr>
<tr>
<td>26</td>
<td>Quinault</td>
</tr>
<tr>
<td>27</td>
<td>Lower Chehalis, Sinkoyone</td>
</tr>
<tr>
<td>28</td>
<td>Upper Chehalis</td>
</tr>
<tr>
<td>29</td>
<td>Lower Chinook, Clatsop</td>
</tr>
<tr>
<td>30</td>
<td>Kwaloqua</td>
</tr>
<tr>
<td>31</td>
<td>Cowichan</td>
</tr>
</tbody>
</table>

- **Cathlamet**
- **Clatskanie**
- **Kwiksh**
- **Control**
- **Kwaquay**
- **Ye’kelle**
- **Tusat**
- **Tillamook**
- **Alsea**
- **Sluslaw**
- **Hanas**
- **Walluk (Coqu)**
- **Tsalme**
- **Uper**
- **Umpqua**
- **Gallie**
- **Applegate**
- **Tututti**
- **Chto**
- **To owa**
- **Karok**
- **Yunuk**
- **Uupa**
- **Chilla**
- **Whikut**
- **Wyt**
- **Vistole**
- **Waiakai**
- **Sinkoyone**
- **Lessik**
- **Chto**
- **Yuk**
- **Pomo**
- **Wappu**
Glottalized consonants are found in Yurok (49), as well as in neighboring Hupa (50) and Tolowa (47), two Athabaskan languages, Chimariko, an isolate, in slightly more distant Shasta, another isolate, and in Wintu, a Wintuan language. Since Yurok is unrelated to all of these languages, the glottalized consonants in Yurok are not due to shared inheritance. Further, Yurok is unique among Algic languages in having ejectives /p’, t’, k’, tʃ’/. The development of ejectives in Yurok is, apparently, unique within its language family.

However, ruling out general linguistic tendencies or chance may be more difficult in this case. Ejectives, or ejective-like consonants, are found in 92/566 or 16.3% of languages in WALS (Maddieson 2013a). Since Yurok is one of approximately 32 Algic languages, and the only one with ejectives, how can we rule this out as a chance event within Algic? If a particular sound change or sound pattern is common, or simply attested in some significant percentage of the world’s languages, how can we determine whether its occurrence in what could be a linguistic area is independent, or due to contact?

To appreciate the problems involved, let us consider a more common sound pattern, like final obstruent devoicing. Final obstruent devoicing refers to sound patterns where an obstruent voicing contrast is neutralized to the voiceless series in word-final position. Devoicing is widespread in the world’s languages, and has a well-understood natural phonetic basis (Blevins 2006). An area with a high frequency of final devoicing sound patterns is Europe as illustrated in Map 4.

---

8 The development of ejectives does not appear to be a natural one within the Algic family. No other languages have glottalized obstruents. Wiyot, a neighbor of Yurok, has some evidence of glottalized sonorants, but not obstruents. In Algonquian proper, aspiration and spread glottal gestures are more common accompaniments to the voiceless stop series.

9 The term is used to refer to other kinds of laryngeal neutralization in final position as well. In German, for example, many varieties show a contrast between fortis aspirated and lenis unaspirated series, with neutralization to the aspirated series in final position (Iverson and Salmons, 2006). See Jansen (2004) for phonetically-based definitions of obstruent voicing vs. VOT contrasts in a range of European languages. This is perhaps the best place to point out that the classic Balkan Sprachbund (Joseph 1983, 1992), on which the notion of linguistic areas is based, is not securely associated with areal phonological features as defined in (1). Some features (like *n > r in Tosk Albanian and Romanian) are found in only a small subset of the languages involved. Others, like the existence of a central unrounded vowel, are common sound patterns, and could therefore result from natural parallel developments, or chance. See Vaux (2002) for further discussion.
This sound pattern stretches from East Slavic (Russian) to West Slavic (Polish), north and west into Germanic languages (German, Dutch), and farther west and south into Romance (Picard, Romansh, Camuno). Proto-Indo-European, as well as Proto-Slavic, Proto-Germanic, and Proto-Italic are all reconstructed with voicing contrasts in initial, medial and final position. As a consequence, final devoicing sound patterns are not due to shared inheritance. Independent developments must be posited within each subgroup, and could reflect parallel evolution of a phonetically-based sound change, or diffusions. In cases like this, a useful heuristic for diffusion is the occurrence of closely related dialects or languages, where one, arguably within the contact zone, shows the areal feature, and another, outside it, does not. Within the swathe of European devoicing, at least a few cases in the literature show this profile. In the central area, for example, we find that Standard Albanian maintains voicing contrasts finally, while Northern Tosk and transitional Southern Geg dialects exhibit final devoicing, shared with adjacent Montenegrin dialects. In this case, Friedman (2004) attributes Albanian devoicing to diffusion, due to historical contact with Macedonian (South Slavic) speakers. Many cases of final devoicing in Romance are also attributed to contact, usually with Germanic languages. For example, final devoicing in Old French is attributed to contact with
In the discussion that follows, we will use the range of criteria in (1) to define areal sound patterns so as to ensure that the sound pattern in question is not likely to have resulted from (shared) inheritance, general linguistic tendencies, or chance. Though some of these criteria may prove too restrictive, eliminating some cases of areal sound patterns, they will provide us with a working base of areal patterns from which we may be able to extract interesting generalizations. It should be kept in mind that databases on sound pattern frequency are limited, so that all references to general distributions noted in the typological literature should be taken as preliminary. For general statistical approaches to linguistic typology, and the problem of ruling out chance in the assessment of areal features see Bickel (in press). For the purposes of this chapter, we classify a sound pattern as uncommon if it occurs in less than 25% of the world’s languages, and as rare if it occurs in less than 5% of the world’s languages based on the UPSID (Maddieson 1984) and the WALS (Dryer and Haspelmath 2013) survey data. A sound pattern like final obstruent devoicing, then, is not rare, and the areal devoicing illustrated in Map 4 could well be a consequence of parallel evolution, diffusion, or some combination of the two.

2 The evolution of areal sound patterns: stone soup and the perceptual magnet effect

How do sound patterns spread from one language to another? How does a language without consonant implosion or retroflexion acquire these features by osmosis? How do new sound patterns materialize as speakers, exposed to these sound patterns in another language, continue to speak their own?

Though it was once commonly believed that areal sound patterns diffused primarily through lexical borrowing, as early as Emeneau (1956), it was clear that a process of “acculturation” occurred, whereby native sounds became more like sounds of contact-languages, without, or independent of, lexical borrowing. A case where this is particularly clear is the linguistic area of Northwestern California, mentioned earlier with respect to the areal distribution of glottalized consonants illustrated in Map 3. The Northwestern California culture area is home to at least four distinct language families/isolates: Karuk (isolate), Yurok and Wiyot (Algic); Hupa and Tolowa (Athabaskan), and Chimariko (isolate). As detailed in Conathan (2004), hundreds of years of multilingualism, intermarriage, and cross-cultural exchange have led to linguistic convergence in many areas of grammar. However, in this linguistic area, lexical borrowing is very rare (Conathan 2004: 80-81). Not only are borrowings rare, but we are hard-pressed to find even one word that was borrowed into Yurok with /p'/ or /k'/; the most likely instances of borrowed words with ejectives involve /ɾ/ and /ɾʃ'/ (Blevins 2004),

---

10 Diffusion of apparent regular sound changes leads one to question their usefulness in subgrouping. In a recent study of Western Numic dialects, Babel et al. (2013) show that all of the six regular sound changes identified show evidence of diffusion. They conclude that, like parallel natural sound changes in distinct lines of descent, “phonological diffusion too can confound cladistic analysis (p. 482).”
Given that ejectives like /p'/ and /k'/ are thought to be areal features, and to have arisen in Yurok as a consequence of language contact, we cannot attribute the source of the sounds themselves to lexical borrowing. For at least some ejectives, an internal historical source has been identified. A sequence of regular sound changes takes *Ct clusters to ejectives as shown in (3) (Blevins 2002).

\[(2) \text{One source of glottalization in Pre-Yurok} \]
\[*Ct [Ct, C't, C't'] > C' > C'\]

Yurok ejectives, then, appear to be triggered by areal contact with neighboring languages that have ejectives. Yet, they appear to evolve via natural phonetically motivated sound change. This observation is just one instantiation of what we might call the “Areal Sound Pattern Paradox”: areal sound patterns are paradoxical, in that their primary catalyst is external, and yet, they often take the form of regular internally motivated sound change. How can this be?

A useful comparison is to the making of stone soup in the well known European folktale, *Stone Soup*:\n
A poor peasant has been travelling for days through rugged land. He is hungry. When he arrives at the next village, he begs for food, first at one door, then at the next, and the next. But no one obliges. “Ah” he tells them, “then I will have to make a delicious stone soup like no other”, and he collects wood, makes a fire, fills his black pot with water from the stream, sets it on the flames, and, with the villagers all looking on, drops one large stone into the pot and begins to stir. “Go away” he says to them, - it will take some time to cook. I will tell you when it is ready. But the villagers are curious, and, one by one, they return. First a woman comes by, “What kind of soup is this, made from a stone?” the woman asks. “Ah” he tells her, taking a taste, “you will see, it is a delicious soup, like no other, and only needs a bit of carrot to be just right”. The woman gives a carrot from her basket, and the peasant drops it in. Soon the woman’s daughter comes by, “What kind of soup is this, made from a stone?” the girl asks. “Ah” he tells her, taking a taste, “you will see, it is a delicious soup, like no other, and only needs a bit of onion to be just right”. The girl takes an onion from her skirt pocket, and the peasant drops it in. Soon the suitor of the girl arrives and asks “What kind of soup is this, made from a stone?”. “Ah” he tells the boy, taking a taste, “you will see, it is a delicious soup, like no other, and only needs a bit of mushroom to be just right.” The boy gives what he has gathered, and the peasant drops it in. Soon the father of the boy comes around, a farmer. “What kind of soup is this, made from a stone?” the farmer asks. “Ah” he tells the farmer, taking a taste, “you will see, it is a delicious soup, like no other, and only needs a bit of barley

11 The most likely borrowings with ejectives are Yurok /tʃʼɨʃ/ ‘bird’, /tʃʼɨʃ/ ‘wren’, /tʃʼɨʃ/ ‘dog’ (Blevins 2002: 11-12). Internal to Yurok, the primary source of glottal stop, and ejection, is Proto-Algic *t.
12 Indeed, Yurok pronominal prefixes ’ne-‘ first person, k’e-‘ second person, we-‘ third person, that are cognate with Wiyot, and more distant Algonquian languages, and that form the basis of the established genetic relationship of these languages, show glottalization, including the ejective /k'/ in the second person.
13 This folktale has many different variants, including those where the hungry travelers making the soup are peasants, beggars, monks, or soldiers, and where the initial ingredient is a stone, a piece of wood, a nail, or some other non-food stuff. The version here is the one I heard most often as a child.
to be just right.” The farmer digs in his bag for a pouch of barley, and the peasant drops it in. Soon the farmer’s wife arrives. “What kind of soup is this, made from a stone?” she asks. “Ah” he tells her, taking a taste, “you will see, it is a delicious soup, like no other, and only needs a bit of herbs to be just right.” The farmer’s wife gives what she has just picked, and the peasant drops it in. When the peasant declares the soup “just right”, he shares it with the entire village, and all agree that it is a most delicious soup, like no other. All are well fed and rejoice.

While there are many ways of reading this tale, the features that most versions share are that of a hungry stranger coming into new country with nothing. From nothing, he makes a delicious soup. In essence, the soup is made from contributions of the surrounding villagers, but since each contribution is small, none of them notice that the soup is made from their own contributions, and not from the original stone. Clearly the stone is not the central ingredient, in fact, in the culinary sense, it is not an ingredient at all. It is a tool which mobilizes external ingredients, and these are what make the soup.

This, I suggest, is a useful metaphor for how areal sound patterns arise, The “stone” in the linguistic scenario of language contact is the contact sound pattern that ultimately spreads: it is the main ingredient of something new, different, and delicious, that people speaking their own language want to experience, but, in reality, it is completely inert. Areal sound patterns are not the result of phonetic borrowing or contamination. Under the view outlined below, the contact language serves as catalyst, but salient features of the new sound pattern come from the indigenous language, naturally, and unknowingly. In the same way that stone soup has the same active ingredients as other soups, areal sound patterns have similar language-internal phonetic trajectories as non-areal sound patterns. The external stimulus of a contact sound pattern shifts the odds of the same pattern evolving in a neighboring language, but it could have evolved in that language independently of contact, only with much lower probability.14

To make this comparison more concrete, I propose the Areal Sound Pattern Hypothesis:

Areal sound patterns are due to perceptual magnet effects within one language, where the perceptual magnets themselves are sounds from another language. As a consequence, their evolution may mimic that of internal phonetically-based sound change.

The notion that language experience alters phonetic perception is central in the field of language acquisition, where it is often referred to as the “perceptual magnet effect” (e.g. Kuhl 1991; Kuhl and Iverson 1995; Kuhl 2000). The central finding in this research paradigm is that exposure to a specific language results in the warping of perceived distances or similarities between phonetic stimuli. In experiments with children and adults, listeners judge similar acoustic differences as being perceptually closer when the tokens include a prototype or best instance of a particular phonological category. In the course of language acquisition, as proto-categories are established, they function as

14 Taking some of the frequencies introduced earlier, recall that ejectives, worldwide, occur in approximately 13.5 of languages. In contrast, in the Pacific Northwest culture zone depicted in Map 3, 94% of languages have ejectives (3/50 do not). Since these languages represent many distinct families, inheritance cannot explain the discrepancy between 13.5% and 94% . The hypothesis is that the odds of a language like Yurok developing consonant glottalization are skewed because of the external stimuli of ejectives in neighboring languages, as detailed below.
perceptual magnets, making stimuli in their vicinity seem more like them, and drawing tokens into the evolving category.

Areal sound patterns can be viewed as long-term consequences of a special case of the perceptual magnet effect. In the case of an areal sound pattern, an external phonetic proto-type is internalized by a speaker on the basis of data external to the language being acquired. In contact situations where bilingualism is the norm, the phonetic proto-type established for one language may have a magnet effect in another. In the mind of the speaker, the two languages are distinct, however, phonetic categories may move towards each other due to the perceptual magnet effect. While this process may have the greatest impact in the early years of language acquisition, there is growing evidence that, at least for fluent bilinguals, these effects may continue into adulthood (Simonet 2014).

This model of areal sound patterns does not require lexical borrowing for sound pattern spread. Further, with phonetic proto-types from one language acting as magnets in another, a prediction is that areal sound patterns will evolve incrementally via one or more seemingly natural phonetically based sound changes, since perceptual magnet effects are natural movements of categories in the perceptual/acoustic space.

Three additional concrete conditions can be identified from the empirical case studies summarized in this chapter as stated in (3).

(3) Sound pattern spread by the perceptual magnet effect: three variables

i. Establishment of a phonetic proto-type requires phonetic saliency: the more salient the phonetic pattern, the more likely it will spread areally.

ii. Establishment of a phonetic proto-type requires significant exposure: the more intense the language contact, the more likely it will result in diffusion of a sound pattern. The perceptual magnet effect works on phonetic proto-types, and may take several generations to yield sound patterns that are recognizable instances of sound change. For this reason, exposure must not only be intense, but span several generations of language learners.

iii. The perceptual magnet effect requires phonetic proximity between the proto-type and language-internal tokens, and draws phonetically similar tokens closer to the phonetic proto-type. If there are such tokens in the neighborhood, sound change will appear to be natural and phonetically motivated, and indistinguishable from internal developments. If the phonetic proto-type is far from similar tokens, there may be no effect, and no new category will evolve.

Within this model, perceptual saliency is critical to sound pattern diffusion (3i). Articulation plays little role in the spread of sound patterns. If the difference between an apical dental stop and an apical alveolar stop is not phonetically salient, this articulatory feature is unlikely to diffuse. Other features that may be less likely to diffuse under (3i) are within-category differences: for example, if two neighboring languages both have a plain voiceless vs. voiceless aspirated contrast that is cued by positive VOT, but one language has significantly longer VOT than the other, the difference in positive VOT is less likely to diffuse than an unshared phonetic feature, because perception of VOT will already itself be active in warping the perceptual space.
At the same time, perceptual saliency is not enough for a sound pattern to diffuse, even if exposure is significant. If the phonetic proto-type is too far from native sounds, magnet effects will not result in new categories. One perceptually salient sound pattern that may resist diffusion for this reason is the occurrence of clicks. Clicks, or consonant sounds made with the ingressive velaric airstream mechanism, are rare sounds occurring in less than 1% of the world’s languages, and in 1.8% of the WALS sample of 567 (Maddieson 2013b). They are primarily restricted to the Khoisan languages (including Sandawe) where they are believed to be directly inherited, but also found in Hadza of Tanzania, and clearly borrowed into some unrelated Bantu languages like Xhosa and Zulu. Areally, however, it is clear that clicks did not spread widely outside of the Khoisan zone. The few languages like Zulu that show borrowed clicks have them through two processes: lexical borrowing, and *hlonipha* vocabulary, - a process of taboo word-formation where clicks can be replaced by non-clicks and vice versa (Faye 1923-25). In short, there is no evidence of perceptual magnet effects yielding diffusion of clicks. Within this model, the failure of clicks to establish themselves as an areal feature is not due to lack of phonetic saliency (3i) or to insignificant contact (3ii); on the contrary, both of these conditions are strongly met. It is the actual mechanism of sound change (3iii) that renders clicks ineffective as perceptual magnets in Bantu. Most Bantu languages reflect the simple Proto-Bantu consonant system with /p t c k b d j ɲ g/ combining into CV or NCV syllables. The problem seems to be that from this Bantu language type, no phonetic categories are particularly close to clicks. As perceptually robust as clicks are, there is nothing to draw into their sphere. They can be borrowed in specific lexical items, and they can replace sounds in taboo word-formation, but they remain inert within the phonological system due to their phonetic distance from other consonants.\(^\text{15}\)

In contrast, the ejective consonants of many North American languages seem to be close enough to distinct sound patterns to serve as perceptual magnets. Recall that Yurok is unique among Algonquian languages in having ejective or glottalized consonants /p’, t’, k’, tʃ’/.\(^\text{16}\) The evolution of ejectives via the sound changes in (2) appears to be due to contact. Glottalized consonants are an areal feature of Northwestern America, and Northwest California, the region where Yurok is spoken is a genetically diverse micro-area within this greater macro-area. How did Yurok acquire ejectives? What triggered the sound changes in (2)? Adhering to the model in (3), we suggest that ejectives, as pronounced in the American Northwest, are perceptually salient sounds (3i). Further, these ejectives were established as phonetic proto-types in Pre-Yurok due to sustained

---

\(^\text{15}\) Another segment type that rarely, if ever, takes part in areal sound patterns is pharyngeals. Pharyngeals and pharyngealized sounds are perceptually salient. However, significant contact between languages with pharyngeals and those lacking them is not enough to induce areal pharyngealization, since it is rare that a language will have a phonetic pattern that approaches the characteristic F2 lowering of adjacent vowels triggered by pharyngeals. For example, languages in contact with varieties of Arabic rarely show sound changes shifting other sounds to pharyngeals. One possible case of pharyngealization as an areal feature is nomadic Northern Songhay Tasawaq, that might have acquired pharyngealization via contact with Tuareg (Kossmann 2012).

\(^\text{16}\) The development of ejectives does not appear to be a natural one within the Algonquian family. No other languages have glottalized obstruents. Wiyot, a neighbor of Yurok, has some evidence of glottalized sonorants, but not obstruents. In Algonquian proper, aspiration and spread glottal gestures are more common accompaniments to the voiceless stop series.
contact and multilingualism with languages having these sounds (3ii). Finally, given certain contexts where consonants were produced with glottalization, perceptual space between these sounds and ejectives was compressed, giving rise to the sequence of changes in (2) where a glottal stop and an adjacent oral stop are produced as a single glottalized consonant (3iii).

One striking aspect of the evolution of areal sound patterns within this model is that the sound changes involved, like those in (2) above, are, for the most part, indistinguishable from phonetically-based sound changes elsewhere in the world that appear to be internally motivated. For example, the association of glottalization with /t/, and a sound change of *t [t, ʔt, tʔ] > ʔ has occurred in many varieties of English (Wells 1982), especially in the syllable coda, with no apparent external conditioning, while similar neutralization of glottalized codas to glottal stop is found in the transition from Middle Chinese to Fuzhou (Blevins 2004: 120-121). At the same time, the regular pre-Yurok fusional development Cʔ > C’ is the diachronic source of ejectives in widespread unrelated languages, including Caddo of the American central plains, and Yapese, an Austronesian language with a long independent history (Fallon 2002: 303).

If ejectives can propagate in the Pacific Northwest through the process described, then, given significant contact in other parts of the world, ejectives are expected to show areal distributions in these regions as well. A survey of the high-density pockets of languages with glottalized consonants in the world shows that this is indeed the case. For example, as illustrated in Map 5, in the Caucasas region, ejectives are found in four different language families: in Nakh-Daghestani languages (e.g. Archi, Lak); in Northwest Caucasian languages (e.g. Kabardian); in South Caucasian languages (e.g. Georgian, Laz); and in two distinct subgroups of Indo-European, - Armenian (Eastern Armenian) and Iranian (Ossetic).
While ejectives may be inherited in Caucasian languages, their occurrence in Armenian and Ossetic stands out in the same way Yurok glottalized consonants do: within the Indo-European language family, ejectives are only found in these two languages. In Ossetic, unlike Yurok, loanwords are the primary source of ejectives (Thordarson 2009). These include Caucasian loans, where ejection is preserved, and Russian loans, where voiceless stops can be replaced with ejectives.\(^\text{17}\) However, there is at least one environment in Ossetic where ejectives have arisen in native words: in final \(-SC\) clusters, e.g. `xuisk’, `xusk’ ‘dry’ (< Old Iranian *huʃka-). Again, it appears that perceptual magnet effects may be at work, drawing a final \([k]\) with heavy release and surrounding noise into the \(/k’/\) category.

The phonetic naturalness and regularity of internal sound changes giving rise to areal sound patterns (3iii) is central to Hamp’s (1996) account of the origins of the retroflex contrast in Sanskrit. As first detailed by Emeneau (1956) in his influential paper “India as a linguistic area”, the contrast between retroflex and non-retroflex apical stops is a striking feature of India and surrounding areas that cannot be attributed to direct inheritance and defies chance. Within the area, illustrated in Map 6, the contrast is found in languages of the four major language families represented: Dravidian, Indo-Aryan, Austro-Asiatic (Munda) and Sino-Tibetan (e.g. Ladakhi).

\(^{17}\) In the case of Russian loans, ejection may be used to mark words as foreign, and could be unrelated to perceptual magnet effects.
As one moves east and north through Pakistan, retroflexion can be found in Burushaski, an isolate, and in Iranian languages (e.g. Wakhi). Retroflexion is not reconstructed for Proto-Indo-European, Proto-Austro-Asiatic, or Proto-Sino-Tibetan, but it is a contrastive feature of Proto-Dravidian. For this reason, Emeneau (1956) and others attribute the feature in Indo-Aryan languages to early contact with Dravidian languages, which clearly had a wider distribution than they do at present.18

Map 6. General area of consonant retroflexion

18 Brahui, a Dravidian language, is spoken in the modern Balochistan region of eastern Iran and Pakistan, almost one thousand miles from the closest Dravidian speaking regions of southern...
In trying to pin down the precise source of retroflexion, Emeneau (1956) suggests that bilingualism allowed "allophones to be redistributed as retroflex phonemes". Hamp (1996) attempts to be more precise about the mechanisms giving rise to Sanskrit retroflexion, justifying the level of detail in terms of providing a clear explanation of "the development and spread of an important trait of the South Asian Sprachbund...[and] one of the clearest, most fully documented, and most internally complex cases known to us of what appears as areally induced phonological diffusion." Hamp’s focus is on the internal evolution of retroflexion by natural means (3iii). The evolution of retroflex consonants within Indic begins with the birth of Indo-Iranian *ʃ, which had several sources, including the "Ruki Rule", whereby *s > f / {r, u, k, i}__\[.\] A rule of voice assimilation in clusters results in a further split, so that Proto-Indo-Iranian has both *ʃ and *ʒ. In Indic, these sounds shift to retroflex articulations. In terms of the model in (3), exposure to Dravidian retroflexion, seems to draw the “closest” Indic sounds towards the retroflex category, giving rise to an internal context-free *ʃ, *ʒ > ś, z sound change.\[20\] After the evolution of [ʂ], [ʐ] in Indic, local and long-distance assimilations between /t, d, n, l/ and these retroflex sibilants give rise to retroflex oral stops, nasals and laterals. Since these assimilations, are, by their very nature, natural and phonetically motivated, the key step in the infiltration of retroflexion into the Indic system appears to be the initial *ʃ, *ʒ > ś, z sound change, itself a clear example of the perceptual magnet effect.\[21\]

Other languages with distinct sound patterns within this greater India area have alternative means of fabricating retroflex obstruents from scratch. For example, in Ladakhi, a Tibetan language of Ladakh, India, retroflex oral stops have developed from...
fusion of earlier *tr and *dr clusters, as in truk > ʈuk ‘six’, dre > ɖe ‘devil’, etc. (Sharma 2004:30).

An interesting question one can pose in this context is how common or rare a sound pattern is outside of cases where it appears to arise through contact. Consider, for example the case of front rounded vowels. Maddieson (2013c) finds that 37/562 languages or only 6.6% of languages in the WALS survey have front round vowels. This suggests that front rounded vowels are relatively rare features in the world’s languages. However, when one looks at the distribution of the vowels in the sample, one finds that the majority of languages with front rounded vowels (29/37 or 78%) are spoken in the north-central area of the Eurasian continent, and further, that outside of this region very few languages with front rounded vowels are found, and they are widely scattered. (Maddieson 2013c).

Within Eurasia front rounded vowels appear to be directly inherited within Uralic and Turkic, but elsewhere, the result of widespread diffusion. Map 7 illustrates some of the languages of Europe with front rounded vowels, and shows that they constitute a near-continuous stretch, from Hungarian in the east, north to Finnish, Norwegian and Swedish, south to Albanian and Greek, and west and south to Gallo-Romance, including Camuno, a Gallo-Italic language of northern Italy (Cresci 2013).

Map 7. Front rounded vowels in Central and Western Europe, including “edge” languages Breton (NW), Souletin Basque (SW), Albanian (SE), Finnish (NE)
From Gallo-Romance, front rounded vowels have diffused even farther west, from French to Breton, and farther south, from French to eastern varieties of Basque, including Souletin, where they are attested in the earliest written documents (Egurtzegi 2013, 2014). Of Proto-Indo-European, Proto-Urалic, and Proto-Basque, only Proto-Urалic is reconstructed with front rounded vowels. Within Indo-European, front rounded vowels arose in: Germanic (excluding Gothic) via umlaut; in Albanian; in South Slavic (Vermeer 1979); and in Gallo-Romance. We suspect, that many of these cases were also contact-induced, given the rarity of front-rounded vowels outside this zone. If we eliminate front rounded vowels that are clear areal features of Western Europe from the WALS database, the sound pattern falls from 6.6% to 5.6%, and if inherited cases are eliminated in Uralic and Turkic, the figure falls to below 5%. By our gross measures above, front rounded vowels would be rare features.

As with Indic retroflexion, details of the evolution of front rounded vowels support a model incorporating the perceptual magnet effect. Front rounded vowels are perceptually salient, and can result in phonetic proto-types when a speaker of a language without them has sustained and significant contact with a language making use of them. As perceptual magnets, front rounded vowels draw phonetically similar tokens closer to the proto-type. And, if there are such tokens in the neighborhood, sound change will appear to be natural and phonetically motivated (3). In Germanic and Albanian, front rounded vowels are a consequence of anticipatory coarticulation in \( C_0 \) sequences, while in Gallo-Romance and Souletin Basque, we see context-free changes of \( u: \) > \( y \), and \( u > y \) respectively; in Souletin and Camuno, subsequent vowel to vowel harmonic sound changes take more vowels to \( /y/ \). While all of these are natural, phonetically motivated changes, they appear to be extremely rare without pre-existing front rounded vowels as perceptual magnets.25

However, at least one areal sound pattern leads one to question whether perceptual saliency is central to all cases of phonetic diffusion. The feature of interest is pre-aspiration, an areal feature of medial tense voiceless stops in Northern Europe, especially Scandinavia and areas where Old Norse was once spoken (Hansson 1999; Helgason 2002). Pre-aspiration is a rare phonetic feature cross-linguistically, occurring in only 4/451 languages of UPSID, and is typically an allophonic variant of post-aspiration.

---

22 See Trudgill (2011: 120-121) for related discussion.
23 In South Slavic, front rounded vowels are still found in some Kajkavian dialects (Vermeer 1979). Greek is also thought to have had front rounded *y (Allen 1987). Under this view, it would be typologically highly unusual in lacking a back rounded counterpart.
24 Slavic speakers consistently unround rounded vowels when speaking German (Raymond Hickey, p.c.). Recall that contact must be sustained for the perceptual magnet effects hypothesized here to take effect.
25 Maddieson’s (2013c) remarks are in line with this hypothesis. With respect to the distribution of front rounded vowels, he remarks “…it is quite striking that their occurrence is so relatively concentrated in a particular geographical area. It seems likely that the hearing of sounds of this sort in some languages of the area may have given further support to phonetically natural processes in other languages [my emphasis:JB], with the end result being the addition of front rounded vowels to the inventory of more of the languages.”
in voiceless stops. It is even rarer as a contrastive feature of oral stops, with Ojibwe being the only UPSID language with this property. Bladon (1986) has suggested that the rarity of pre-aspiration is due to its low perceptual saliency. If this is the case, then, pre-aspiration should be less likely to diffuse by (3i) than other more salient phonetic properties. Helgason’s (2002) detailed study of pre-aspiration in Nordic languages, shown in Map 8, provides confirmation of the model in (3) in two ways.

Map 8. Principal locations of pre- and post-aspiration in Scandinavia and the North Atlantic (Helgason 2002: 3)

First, it demonstrates that the strongest instances of pre-aspiration, cases where it is a language-specific phonetic feature, or “normative” in his terms, are those directly inherited from Old Norse. In these cases, direct inheritance accounts for the common feature of pre-aspiration and its distribution. In cases where pre-aspiration has diffused

---

26 Tarascan (aka P’urhépecha), an isolate of Michoacan, Mexico, is classified by UPSID as a language with (post-)aspirated stops, though these are pre-aspirated in predictable environments (Friedrich 1971, 1975; Lluvia Camacho, 2013, personal communication). Counting Tarascan as a pre-aspirating language would make the figure 5/451 languages in UPSID.

27 Helgason’s definitions of normative vs. non-normative phonetic traits are as follows: “If the absence (or presence) of a particular phonetic trait leads to a pronunciation that is considered deviant by the speakers of a given dialect, that trait can be classified as normative (or normatively absent) in that dialect. Conversely, a trait whose absence or presence does not lead to deviant pronunciation can be classified as non-normative in that dialect” (Helgason 2002: 21). There may be some circularity in these terms, and I have taken the liberty of paraphrasing them in the text.
into other language families, it is either highly variable and non-contrastive, or strengthened to be more perceptually salient. In the first case, exemplified by Saami varieties and Tyneside English, the weak perceptual saliency of pre-aspiration results in spotty diffusion, with variability across speakers, and what Helgason refers to as “non-normative” phonetics. In contrast, a stronger percept is associated with Helgason’s category of “normative” phonetics. The spread of pre-aspiration from Northern Germanic languages into Celtic has given rise to many Scots-Gaelic dialects where Common Gaelic /p, t, k/ are realized as pre-aspirated stops. In areas where pronunciation appears to be the most “normative”, preaspirates are strengthened to velar fricatives, resulting in pronunciations like [xp], [xt], and [xk]. It appears then that where pre-aspiration diffuses and becomes a regular allophonic pattern, it is more perceptually salient than elsewhere.

3 Radical areal spread, and tone as an areal feature

To this point, the areal sound patterns under discussion have involved single features or segment types: pre-glottalization of voiced consonants; ejectives; retroflexion; click sounds; front-rounded vowels; and pre-aspiration. Recall that part of the reason for focusing on less common phonological features was to rule out chance, or parallel evolution (i.e. natural internal developments) of more common sound patterns. However, another way of ruling out chance or parallel evolution is to show that a group of features is shared between neighboring languages, and that this group of features cannot be the consequence of shared inheritance. In cases where multiple aspects of sound patterns spread, radical changes to the typological profile of a language are possible.

One of the most striking instances of this kind of radical areal spread is found in South-East Asia. Matisoff (2006) catalogues numerous instances of this kind of profound structural and prosodic influence, 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 influence of Mon-Khmer and Sinitic on the evolution of tone and register in Chamic, a subgroup of Austronesian. This last case is, perhaps, the best documented of all instances of radical phonological areal diffusion, being the subject of Thurgood’s (1999) monograph, From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change. The level of detail in terms of social and linguistic history is remarkable, and the stages and range of areal phonological features are audible to this day in the Chamic languages still spoken.

Chamic, and its sister Malayic, are subgroups of Western Malayo-Polynesian, within the expansive Austronesian language family. The history of Chamic-speaking

---

28 The phrase-final distribution of pre-aspiration in Tyneside English is consistent with perceptual magnet effects: in this variety, which lacks a geminate-singleton or tense/lax obstruent contrast, pre-aspirates, as perceptual magnates, might be expected to draw in phrase-final obstruents due to the effects of phrase-final spread-glottal gestures (Blevins 2006).

29 A final aspect of pre-aspiration that should be kept in mind in assessing its perceptual saliency is what it contrasts with. In typical cases, the phonetic contrast is between long voiceless stop closure preceded by modal voicing versus shorter voiceless stop closure preceded by a period of glottal friction, e.g. VtV vs. VhtV. The relevant question that must be asked then is how perceptually salient this contrast is for speakers of different language types.
people over the last 2,000 years involves movement from insular Asia to the Southeast Asian mainland as early as 500 B.C., and involves initial and sustained contact with speakers of Mon-Khmer languages, with the exception of Achenese speakers, who returned to Northern Sumatra [Map 9a].

Map 9a. The modern distribution of Chamic languages (Thurgood 1999)
Subsequent movements of distinct groups of Chamic-speaking peoples resulted in later intense contact with Vietnamese, and with Min speakers on the island of Hainan [Maps 9a, 9b] Thurgood’s (1999: 6) summary of the general tendencies in sound change under the influence of these languages are listed in (4).

(4) Radical areal sound patterns in Chamic

i. Templatic modification: disyllabic > sesquisyllabic > monosyllabic

ii. Restructuring of segment inventories
   a. New consonant series
   b. Proliferation of vowel contrasts

iii. Syllabic modification
   a. merger and loss of final consonants
   b. neutralization of voicing and vowel-quality in pre-syllable

iv. Prosodic innovation: neutral > register complex > tone
Table 1 illustrates the shift from disyllabic to sesquisyllabic to monosyllabic/tonal with Austronesian cognate sets. The shift to sesquisyllabic or iambic words was due to contact with Mon-Khmer languages that are sesquisyllabic; the further reduction to monosyllables in Phan Rang Cham was due to contact with Vietnamese, a monosyllabic language, while for Tsat, monosyllabism resulted from contact with the monosyllabic Sinitic languages of Hainan.

<table>
<thead>
<tr>
<th>PMP</th>
<th>Malay</th>
<th>Rade</th>
<th>Tsat</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>*mamaq</td>
<td>mamah</td>
<td>məmah</td>
<td>ma$^{15}$</td>
<td>chew</td>
</tr>
<tr>
<td>*qumah</td>
<td>Huma</td>
<td>hmah</td>
<td>ma$^{13}$</td>
<td>'dry field'</td>
</tr>
<tr>
<td>*tæran</td>
<td>Tanan</td>
<td>kæran</td>
<td>na$^{13}$</td>
<td>'hand; arm'</td>
</tr>
<tr>
<td>*pænaq</td>
<td>Panah</td>
<td>monah</td>
<td>na$^{11}$</td>
<td>'shoot (bow)'</td>
</tr>
<tr>
<td>*baseq</td>
<td>Basah</td>
<td>mosah</td>
<td>sa$^{13}$</td>
<td>'wet; damp'</td>
</tr>
<tr>
<td>*m-uda</td>
<td>Muda</td>
<td>moda</td>
<td>tʰa$^{11}$</td>
<td>'young; tender'</td>
</tr>
<tr>
<td>*daRaq</td>
<td>Darah</td>
<td>ãrah</td>
<td>sia$^{15}$</td>
<td>'blood'</td>
</tr>
<tr>
<td>*bulan</td>
<td>Bulan</td>
<td>mlan</td>
<td>-phian$^{11}$</td>
<td>'moon'</td>
</tr>
<tr>
<td>*qabu</td>
<td>Abu</td>
<td>hbaø</td>
<td>phis$^{11}$</td>
<td>'ashes' (p.183)</td>
</tr>
</tbody>
</table>

Table 1. Radical phonological change in two Chamic languages: Rade of South Vietnam, and Tsat of Hainan Island (based on Thurgood 1999)

As with other cases of areal sound patterns discussed above, Thurgood demonstrates that the contact-induced sound patterns that arise in Chamic do so through natural, phonetically-based regular sound change. While substantial loans undoubtedly played a role in seeding sesquisyllabic structure, register, and tone in these languages, the internal developments support the model in (3), where a perceptual magnet effect gives rise to new categories from old cloth. In the shift from disyllables to sesquisyllables, the iambic prosody of Mon-Khmer languages draws similar disyllables toward it, resulting in wholesale reduction of initial vowels as seen in Rade (Table 1). In Western Cham, contact with neighboring Mon-Khmer register languages, has additional effects. Proto-Chamic *b-, *d-, *g- and *j- and their transitions into following vowels are perceptually close to the breathy voiced register of contact languages, and it is this register which is found in syllables with these initials. In the Tsat case, Min tones act as perceptual magnets drawing F0 perturbations, originally due to laryngeal aspects of consonants, in to be reinterpreted as tones.

Thurgood’s (1999) case study of the areal diffusion of multiple sound patterns, including tone and register, has serious implications for general typological work on tone. It is widely recognized that about half of the world’s languages are tone
languages. Most of these tone languages are centered in three large zones: South-East Asia, New Guinea, and equatorial Africa, as illustrated in Map 10, where "simple tone systems" are those with a H vs. L contrast, and "complex tone systems" are all others (Maddieson 2013d).

Map 10. Tone as an areal feature of sub-Saharan Africa, South-East Asia, and New Guinea (Maddieson 2013d)

Within Africa, most of the tonal languages are found within the Niger-Congo family, and tone is assumed to be an inherited feature there. However, the situation in South-East Asia is very different. Though tone and register are pervasive, the largest language families in the area, Proto-Austro-Asiatic and Proto-Sino-Tibetan, are reconstructed as non-tonal languages. It could well be that tone has evolved through contact more than any other phonological feature, both within South-East Asia, and within the New Guinea area.

Potential support for the rarity of non-contact induced tonogenesis can be found in the wider history of the Austronesian family, a family of approximately 1,000 languages. Proto-Austronesian was thought to be spoken approximately 6,000 years ago in what is now Taiwan. Proto-Austronesian had no register or tone contrasts and fairly simply CVC(C)V(C) stem structure. Over the course of thousands of years, as populations moved and split and moved and split, thousands of sound changes occurred. However, nearly all cases of tonogenesis within the Austronesian family can be attributed to contact-induced change: tonogenesis in Tsat occurred under Min influence as outlined above; tone in a dialect of Moken, and in Pattani Malay are due to heavy contact with southern Thai (Blust 2009: 181); and there are scattered cases of tonogenesis in the New Guinea area, including the Raja Empat languages of western Papua (Remijsen 2003), Jabem and Bukawa of the coastal Huon Gulf area (Ross 1993), and Kara of New Ireland.

30 Maddieson’s (2013d) WALS figure is 58.2 non-tonal vs. 41.8 tonal, however, in the text he suggests that if Niger-Congo languages were not under-represented, the figure would be closer to 50/50.
(Hajek and Stevens, 2004), which occur in Papuan contact zones. The one apparent exception to contact-induced change is tonogenesis in New Caledonia, resulting in 5 (of 28) tone languages (Rivierre 1993). Assuming that the islands of this chain were uninhabited when settled by Austronesian speakers, tonogenesis in New Caledonia is the sole instance of an internal development leading to contrastive tone over thousands of years of development, and thousands of languages.

Implications of tone as a common areal sound pattern, however, go beyond linguistic typology. It has been hypothesized by Dediu and Ladd (2007) that the global distribution of tone languages is related to a negative correlation between the linguistic feature of tone and population frequency of the derived haplogroups of two brain size genes in human populations, ASPM and Microcephalin. Evidence that tone can and does spread areally, and that the tone languages of South-East Asia represent a linguistic area, seems incompatible with a hypothesis that the distribution of tone (or absence of tone) in the world’s languages is related to an inherited genetic feature of human brains.31

4 Concluding remarks

Areal sound patterns are easy to define but not always easy to identify. This is because they appear to mimic internal developments. The perceptual magnet model of sound pattern diffusion makes use of the same mechanisms needed in first language acquisition, and therefore, predicts the similarity between regular sound changes with internal perceptual magnets, and those whose perceptual magnets are external. Sound patterns that diffuse must be perceptually salient, but how salient? It was suggested that category-internal divisions are unlikely to propagate, but a more positive proposal is possible. Phonetic features that are not central to signaling contrast in a language are more likely to be co-opted into new category formation than others, and, the great majority of areal sound patterns are additive, as opposed to neutralizing. Because areal sound patterns require extensive periods of significant language contact, they could, in theory, be used to reconstruct ancient population movements and pre-historic contact zones. In this area, countless mysteries present themselves. Consider, for example, the geographic distribution of rare word-final pre-stopped (pre-ploded) nasals, found in some Austronesian languages of Borneo, Chamic languages, and some Mon-Khmer languages of the Malaysian peninsula (Blust 1991:149). Does this distribution reflect pre-historic contact between Austronesian and Mon-Khmer speakers in Borneo? Another mysterious but well studied sound pattern is the Proto-Aztecan sound change of *t > tɬ/a (Campbell and Langacker 1978). Since the evolution of lateral affricates from /l/ before low vowels appears to be unknown outside of this case, could areal influence from languages with lateral fricatives or affricates have played a role? Could this unusual sound change provide evidence for early contact between Proto-Aztecan speakers and speakers of Proto-Totonaco-Tepehua? The Areal Sound Pattern Hypothesis together with other tools of analysis offered in this volume should bring us closer to answering these questions, and many more.

31 A serious problem is raised by Wong et al. (2012), an fMRI study of brain function, where a positive relationship between ASPM and lexical tone perception is found, in contrast to the opposite pattern predicted by Dediu and Ladd (2007).
References

Babel, Molly, Andrew Garrett, Michael J. Houser, and Maziar Toosarvandani. Descent
and diffusion in language diversification: A study in Western Numic dialectology.
Bickel, Balthazar. in press. Distributional typology: statistical inquiries into the dynamics
of linguistic diversity. In: Bernd Heine and Heike Narrog (eds.) The Oxford
Blevins, James P. and Juliette Blevins (eds). 2009. Analogy in Grammar: Form and
Blevins, Juliette. 2002. Notes on Yurok glottalized consonants. In Laura Buszard-
Welcher and Leanne Hinton (eds.), Proceedings of the Meeting of the Hokan-
Penutian Workshop (Survey of California and Other Indian Languages, Report 11).
Cambridge: Cambridge University Press.
Blevins, Juliette. 2006. A theoretical synopsis of Evolutionary Phonology. Theoretical
Blevins, Juliette. 2007. The importance of typology in explaining recurrent sound
Klaas Willems and Ludovic De Cuypere (eds.), Naturalness and Iconicity in
Blevins, Juliette. 2009a. Another universal bites the dust: Northwest Mekeo lacks coronal
syncopae in Austronesian. In Alexander Adelaar and Andrew Pawley (eds.),
Austronesian Historical Linguistics and Culture History: a Festschrift for Bob Blust.
Canberra: Pacific Linguistics. 33-49.
Blevins, Juliette. 2014. Evolutionary Phonology: A holistic approach to sound change
typology. In, Patrick Honeybone and Joseph Salmons (eds), Handbook of Historical
Blevins, Juliette and Sven Grawunder. 2009. *K> T sound change in Germanic and
elsewhere: Descriptions, explanations, and implications. Linguistic Typology 13: 267-
303.


Emeneau, Murray B. 1956. India as a linguistic area. Language 32: 3-16.


Vaux, Bert. 2002. There was and wasn’t a Balkan Sprachbund. Paper presented at AIEA, Würzburg.
Convergence and divergence in the phonology of the languages of Europe

Thomas Stolz and Nataliya Levkovych

1 Introduction

This paper addresses the synchronic structural similarity and dissimilarity of languages in general by way of discussing selected phenomena of the segmental phonology of the languages of Europe in particular. Our guiding question is whether or not the distribution of certain properties of phonological systems provides at least circumstantial evidence of areal-linguistically meaningful patterns. We start from the assumption that the phonological profile of Europe provides sufficiently interesting data to feed the debate on the comparability of language structures.

Before we present, discuss and evaluate the phonological phenomena, it is necessary to clarify a small set of basic notions which are essential for the understanding of our line of argumentation. Similarity and dissimilarity are synchronic properties of many languages. The degrees of similarity and dissimilarity among languages serve as indicators of structural unity and diversity, respectively. Positive similarity is equivalent to the presence of a given phenomenon in the languages compared whereas the parallel absence of a given phenomenon constitutes a case of negative similarity, i.e. dissimilarity. Convergence and divergence, on the other hand, are dynamic concepts which presuppose processes of language change in the course of which the degree of similarity or dissimilarity of a given set of languages increases or decreases. Increasing similarity is tantamount to the strengthening of structural unity whereas decreasing similarity results from the growth of diversity. The convergence of languages A and B may trigger the divergence of languages A and C and vice versa. In this paper, we adopt an exclusively synchronic perspective which needs to be complemented by a diachronic interpretation in follow-up studies. Thus, discounting the occasional digression and the outlook section, we do not address directly the issue of language contact and independent parallel development in this contribution.

The identification of linguistic areas is not a primary goal of ours since we subscribe to Campbell’s (2006) critical assessment of the Sprachbund-debate. Instead we apply the principles laid down in Bechert ([1981] 1998: 14) according to which the determination of the full extension of isoglosses has to precede the construction of linguistic areas and not vice versa. In this study, only the outer boundaries of Europe impose restrictions on the possibility to track some of the isoglosses in their entirety. What we expect to find are bundles of isoglosses (isopleths in the sense of van der Auwera (1998)) and/or different statistically noticeable preferences in different parts of the continent under scrutiny. To this end, we frequently employ simple dichotomies such as periphery vs centre/interior, east vs west, etc. The geolinguistic interpretation of the results is supported by maps throughout this chapter.

As to the geographical limits of Europe, we follow the lead of the EUROTYPO-inspired maximalist interpretation in Kortmann and van der Auwera (2011) so that the westernmost part of Kazakhstan, Transcaucasia and Anatolia belong to Europe. To make the endeavour more interesting, we add Greenland as the northwesternmost extension to
the area under scrutiny. Our sample comprises Afroasiatic, Eskimo-Aleutic, Indo-European, Mongolian, Northeast Caucasian, Northwest Caucasian, South Caucasian, Turkic, and Uralic languages as well as the isolate Basque (cf. Appendix). For reasons of space, the focus is on consonantal phonemes and some of their prominent properties which we check for in an as yet unbalanced pan-European sample of 185 languages including nonstandard varieties. We discuss both selected qualitative and quantitative aspects of the systems of phonemic consonants.

In terms of phonological theory and methodology, we are indebted to functional typology (Velupillai 2012: 61-87) and to the “received” structuralist tradition (Hall 2011: 37-99). As to terminology and classification we largely adopt the distinctions introduced by Ladefoged and Maddieson (1996).

Notwithstanding the previous areal linguistically-minded research of phonological issues in European languages (cf. Section 2), the project into which this study is embedded is a first since there is as yet no comprehensive areal-linguistic account of the phonologies of the languages of Europe. Only one of the eight volumes which document the achievements of the EUROTYP project displays an orientation towards phonology, namely van der Hulst (1999) on word-prosodic systems. A distant forerunner, the volume on sandhi phenomena in the languages of Europe edited by Andersen (1986) covers an important aspect of phonotactics (also across word-boundaries). Proper segmental phonology however, is clearly understudied also in the many post-EUROTYP spin-offs which continue to refine our knowledge of the areal characteristics of the European linguistic landscape. According to Haspelmath (2001: 1493), there are two possible explanations for this apparent neglect. Either the phonological properties of the languages of Europe are too trivial to justify being studied in-depth or phonologists “have not looked hard enough” to discover geolinguistically meaningful patterns. The most likely reason for this gap is given by Haarmann (1976a: 115) who states that the phonological systems of the languages of Europe still need to be described according to one and the same model. Since the extant descriptions of the individual sample languages apply different phonological models, the compatibility of the results is severely restricted. This is why we have to reanalyze (many of) the sources in order to fit the data into a common format.

Problems arise also from the different goals and motivations of the scholars to argue for or against diversity or unity. For the now largely outdated pioneer Lewy (1964 [1942]: 107-108), for instance, it was important to establish the existence of structural diversity to prove his Völkerpsychologische idea that people in different parts of Europe think differently. Conversely, Décsy (2000b) needs structural unity to support his idea that there is a “common European mind-set”. In the light of van Pottelberge’s (2005) criticism of the ideology behind many areal-linguistic accounts of the European situation, we attempt to establish - as objectively as it is possible with a constructed category - the distribution of (classes of) phonemes across the European linguistic landscape. In brief, we assume that structural similarity applies if the phonological phenomena in different languages are identical.

There is a plethora of options for how to determine whether languages are structurally similar or dissimilar. For the purpose at hand, we stipulate three different levels on which similarity and dissimilarity can be checked. On the macro-level, phonological classes and other overarching principles of the organization of phonological systems are compared to each other. The meso-level is reserved for the comparison of phonemes as (moderately) abstract members of systems whereas the concrete realizations
of phonemes, i.e. the phonetics, are studied on the micro-level. Macro-level similarity does not automatically correspond to micro-level similarity. For the purpose at hand, we focus especially on the macro-level and check the meso-level only for primary articulations (cf. below) to the exclusion of the micro-level since the range of variation on the micro-level is considerably more complex than on the two higher levels and would demand separate investigation.

For genetically closely related languages it is assumed, as an initial premise, that they are structurally similar to each other. The data discussed in this study suggest that this assumption is not borne out by the facts. What we will often find instead are areally-based preferences which cut across genetic boundaries. Since two of the above three levels are checked for similarity and dissimilarity, we make a selection from the wealth of data to keep the presentation within reasonable bounds. This is why we comment on the heterogeneity or homogeneity of language families only in passing. A brief excursus highlights aspects of internal diversity for four selected families. Our main interest, however, is to determine to what extent diversity and unity in the realm of phonology is dependent upon areal considerations.

We are aware of the limitations of our endeavour. The study of a narrowly circumscribed fragment of the phonology of our sample languages does not allow us to classify these languages holistically. Even if it can be shown that certain languages are similar in their phoneme inventories, the chances are that the same languages differ markedly in other aspects of phonology (such as phonotactics). We intend to fill in the many gaps in the future and thus to complement the present explorative study.

The paper is organized as follows. In Section 2, we first report briefly on the role phonological issues have played in recent areal-linguistic research and then we zoom in on Europe. To this end we consider the views on the interrelationship between phonology and areal factors in the languages of Europe. With this background information, we can view the empirical side of the areal phonology of Europe in Section 3 and its subsections. The excursus on diversity in genetic families is presented in Section 4. Section 5 contains the conclusions including the general evaluation of our findings and an outlook on future research. Space restrictions do not allow us to do justice to all of the interesting hypotheses suggested for phonological systems in areal perspective and/or under conditions of language contact. In the subsequent parts of this paper, we therefore content ourselves with referring to a small selection of the contributions on this topic. We restrict the cartographic illustration of the distribution of phonological properties of our sample languages to a selection of those cases in which the isoglosses do not single out particularly small sub-areas of Europe but yield a different picture with two or more major sub-areas.

2 Areal phonology – with a special focus on Europe

Jakobson ([1930] 1971) coined the term *Phonologischer Sprachbund* already in the early 1930s. The existence of phonological areas – sometimes of considerable extent – is corroborated by the *World Atlas of Language Structures (WALS)* which contains nineteen chapters devoted to phonological problems (Maddieson 2005a-m; Anderson 2005; Hajek 2005; Goedemans/van der Hulst 2005a-d). The maps which accompany these chapters illustrate that most of the properties offer evidence of areal preferences some of which involve Europe or parts thereof. Phonological issues are relevant also for all of the eight
linguistic areas discussed in Matras, McMahon and Vincent (2006). Accordingly, one would expect that phonology also features prominently when it comes to determining areal properties of European languages. However, the situation is by no means as clear as that.

The prior work on areal-phonological issues in Europe can be divided informally into several strands of research. First of all, a distinction must be made between essentialist and descriptive approaches. The former aim at identifying the common (and, if possible, distinctive) phonological properties shared by all of the languages of Europe or a sub-region of the continent. On the descriptive side, scholars ideally take stock of all phonological phenomena which occur in (a sub-region of) Europe independent of their areal typicality. Both these perspectives are well represented in the relevant literature.

Wagner (2002 [1964]) and Ewels (2009), for instance, scrutinize two contiguous regions – northern Europe and the Baltic – regarding the diffusion of selected phonological phenomena over groups of genetically diverse languages. In this way they contribute to characterizing two potential sub-areas of Europe phonologically. The accessibility of several other studies of this kind dedicated to further sub-areas of the continent notwithstanding, it is practically impossible to paint a linguistically satisfactory picture of the areal phonology of Europe by way of simply adding up the results of the investigations of specific sub-areas. Some of the phonological isoglosses cross the borders of, or occur in several (not necessarily contiguous) sub-areas, other properties are not shared by all members of a supposed sub-area and still other phenomena are passed over tacitly because they do not count among the characteristics of a given sub-area. For example, a stable site for word stress is considered a characteristic not only of the Danube-Sprachbund but also of the Baltic Sprachbund (Haarmann 1976b: 100-101 and 110).

Haarmann (1976a), Ternes (1998, 2010), and Décsy (2000a-b) are examples of investigations which combine essentialist and descriptivist elements. In their search for pan-Europeanisms, these authors come up with a variety of solutions. Haarmann (1976a: 108-109) postulates so-called Europemes the first four of which belong to the realm of phonology. These phonological Europemes can be shown to reflect language-independent universal trends and thus fail to be distinctive (Stolz 2006). This result is in line largely with the results of the two studies by Ternes (1998, 2010) who surveys the phonological typology of the languages of Europe with a view to formulating generalizations of pan-European dimensions. According to this author, the European languages do not stand out in any way in a cross-linguistic perspective (Ternes 2010: 594). Décsy (2000b: 341-53) tries to reconstruct the set of phonemes common to all contemporary languages of Europe without judging the distinctiveness of the set in comparison to non-European languages. In general, essentialist models do not inform us sufficiently about the areal phonology of the European continent because they gloss over all aspects of structural diversity to the benefit of unity.

On the other hand, Haarmann (1976a: 116) and Ternes (2010: 584-6) observe that there is variation on other parameters and some of the phenomena can be interpreted as being areally (and to some extent also genetically) biased. This observation suggests that phonologically, Europe is probably more interesting in terms of its internal diversity than it is in a global perspective. As Comrie (1993) maintains, it is important to linguistically study the variation that occurs within a relatively limited geographical space, for instance that in Europe in comparison to more sizable continents like Africa, Asia, and the
Americas. The extent of the structural diversity across Europe on the phonological level still needs to be determined.

In this context, studies like that of Stadnik (2002) on the areal aspects of palatalization correlations in Europe and Asia are especially helpful since they focus on well-defined structural phenomena without having to prove an idée fixe as to the areal subdivision of the European continent. Distributions which are independent of the putative boundaries of internal sub-areas are identified in this approach to the areal phonology of Europe. In a similar vein, Stolz (2007, 2010) and Stolz, Urdze and Otsuka (2010, 2011, 2012) check the distribution of the structures of monosyllables, sibilants and affricates, along with velar and postvelar fricatives, liquids, and phonological rara and rarissima in Europe. In most of the cases, areal patterns are discernible so that it is very probable that similarly meaningful results can be achieved by way of scrutinizing further phonological phenomena. This potential of our project is illustrated below by those aspects of the phonemics of European languages we discuss in Section 3.

3 The areal phonemics of Europe

This section contains two major subsections with one of each dedicated to the macro-level and the meso-level, respectively. In these subsections, we discuss the phenomena which are connected to structural unity and to structural diversity. These aspects are checked for areally remarkable results with the occasional comment on diversity and unity within genetically defined groups of languages. We look at both positive similarity and negative similarity without aiming at a balanced treatment of both. Furthermore, most of the data we analyze in this section will be evaluated quantitatively without addressing qualitative aspects systematically. Since our study looks exclusively at consonants, we do not make any statements about the consonant-vowel ratio of our sample languages – a topic that is important for Haarmann (1976a: 113-114) and Ternes (2010: 579-580). Furthermore, Ternes (2010: 578), for phonological reasons to be explained below, excludes the languages of the Caucasian region from his sample of European languages. Our study – which also covers the languages of the Caucasus - is meant to check whether or not Ternes’ decision is justified.

Throughout this chapter we provide maps which reflect the geolinguistic distribution of selected phenomena discussed in the subsequent paragraphs. To guarantee easy readability, the maps are kept as simple as possible, i.e. we restrict the number of features, whose distribution is represented on a given map, to two or three. More often than not the map is based on the dichotomy of the presence vs absence of a phonological phenomenon. For the purpose of areal-linguistic cartography, we focus on those features for which there is a statistically significant majority distribution and a (set of) undisputable minority distribution(s).

We start from a straightforward working definition of the segmental phoneme which we conceive of as a countable unit which serves to distinguish meaningful elements in a given language. We discount geminates because their phonological status is too controversial to allow for their consideration. The presentation follows largely a top-down pattern.

The frame of reference for the empirical part of this study is the chart of consonants of the International Phonetic Alphabet (IPA) provided by the International Phonetic Association (1999: ix). There are fifty-eight simple symbols for pulmonic
consonants, sixteen symbols for ejectives and implosives, and nine additional symbols including those for epiglottal consonants and clicks. Affricates are represented by complex symbols (mostly ligatures). On the basis of our prior investigations, we have identified a set of seventy-six simple IPA-symbols which are required for the languages of Europe. For convenience, we term the phonological units represented by these symbols ‘cardinal phonemes’, segments which involve primary articulations. In the IPA-system, secondary articulations or coarticulations are represented by diacritics which turn simple symbols into complex symbols. In our terminology, complex symbols of this kind represent extended cardinal phonemes. Cardinal and extended cardinal phonemes conform to the structuralist definition of the phoneme insofar as their feature specifications contain only those elements which are distinctive systemically. If the coarticulations are merely phonetics, they are ignored in this study.

4 Macro-level: phonological classes

In this subsection, the areal distribution of phonological classes is determined. For a phonological class to be attested in a given language, the presence of a single phonemic representative of this class is sufficient.

We start on the highest levels of the phonological matrix by way of checking whether or not the parameters of airstream mechanisms, manner of articulation, place of articulation, and phonation are subject to geolinguistic variation in Europe. This section comes to a close with the discussion of secondary articulations. Owing to the general systematicity of phonological phenomena, this sub-section bears resemblance to the organization of the studies by Ternes (1998, 2010).

4.1 Airstream mechanisms

The initial observation is trivial, of course, because it states that all 185 languages of our sample are equipped with pulmonic consonants. Since this is a universal property of human languages, the common behavior of our sample languages is predictable. What is much less trivial however, is the existence of additional non-pulmonic consonants in a sizable minority of the sample. Non-pulmonic classes such as clicks and implosives are unknown in Europe. In contrast, ejectives are attested in the Caucasian region i.e. in a narrowly delimited geographical part of Europe. Thus, the distinction of pulmonic vs non-pulmonic consonants is areally significant because it divides Europe into a large area of languages which only admit pulmonic consonants and a much smaller southeastern area in which pulmonic consonants coexist with non-pulmonic consonants (cf. Table 1).

<table>
<thead>
<tr>
<th>airstream mechanism</th>
<th>unity</th>
<th>diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pulmonic</td>
<td>+ pulmonic</td>
<td>-</td>
</tr>
<tr>
<td>non-pulmonic</td>
<td>- clicks</td>
<td>+ ejectives</td>
</tr>
<tr>
<td></td>
<td>- implosives</td>
<td>1. genetic:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Northwest Caucasian,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Northeast Caucasian,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• South Caucasian</td>
</tr>
</tbody>
</table>
Table 1: Unity and diversity with regard to airstream mechanism

The pulmonic consonants constitute an expected element of structural unity whereas the existence of non-pulmonic consonants is an element of structural diversity. The vast majority of our sample languages show only one airstream mechanism. There is a minority of languages in which two modes of airstream mechanisms are phonologically relevant.

It is exactly the existence of ejectives in the languages of the Caucasus which motivates Ternes (2010: 578) to exclude the Caucasian region from his account of the phonological typology of the languages of Europe because the Caucasian propensity for ejectives would jeopardize the supposedly homogeneous picture of the areal phonology of Europe. Ejectives are a common feature of all members of the genetically unrelated phyla Northwest Caucasian, Northeast Caucasian and South Caucasian. However, ejectives are also an areal feature of the Caucasian region since this airstream mechanism is phonemically relevant in Indo-European languages as well (Eastern Armenian and Ossetic) and regional varieties of Turkic languages (Kumyk, Azeri, Karachay-Balkar, and Turkish) spoken in the Caucasus as well as in the geographically slightly more remote Neo-Aramaic (i.e. Semitic) varieties of Urmia, Van and Mosul (Chirikba 2008: 44-45). This means that the diffusion of the feature of non-pulmonic airstream mechanisms across genetic boundaries has contributed to the structural unity of genetically unrelated neighbors and at the same time those non-Caucasian languages which have converged on this parameter with their Caucasian neighbors have become dissimilar to their next-of-kin outside the Caucasian region. Therefore, the increase of similarity in areal perspective triggers the decrease of similarity in genetic perspective.

In what follows we pay special attention to the behavior of the languages of the Caucasus if it is at variance with that of the bulk of our sample languages.

4.2 Manner of articulation

As to the manners of articulation, the distinction of obstruents vs sonorants is a common European feature because both these classes occur in all of our sample languages. This omnipresence holds also for the majority of distinctions on the next level. In terms of phonological classes, stops, fricatives, nasals, liquids and approximants are indeed attested across the board in Europe. There are three cases however, for which the analysis is not as straightforward as that.

The class of liquids can be further subdivided into laterals and rhotics. All 185 of our sample languages offer evidence of both these sub-classes. The phonemes which are often subsumed under the class of rhotics (= /ɹ/, /ɻ/, /ɿ/, /ɾ/, /ʁ/) do not form a natural class because by no means all of them share the necessary minimum of features

1 The Caucasian region is also excluded from the studies by Lewy ([1942] 1964), Haarmann (1976b) and Décsy (2000a-b) although mostly for extra-linguistic reasons some of which are of a dubious nature.

2 Not all of the varieties equipped with ejectives form part of our sample.
In terms of the manner of articulation, different rhotics belong to different phonological classes, namely approximants, taps/flaps, or trills. Given that approximants occur in all of our sample languages, only the distribution of taps/flaps and trills as classes is of interest for the topic of this section. With attestations in 167 languages (90% of the sample), trills represent the more widespread of the two classes. In contrast, taps are reported for a minority of 24 languages or 13% of the sample. 178 languages show either taps/flaps or trills which is equivalent to 96% of the sample. The erstwhile cover-term for the two classes of taps/flaps and trills is vibrants, i.e. 96% of our sample languages offer evidence of vibrants. Moreover, this means that there are eleven languages (6% of the sample) in which phonemic taps/flaps coexist with and are distinct from phonemic trills. Thus, there are four groups of languages, namely those which show trills, those which only allow for taps/flaps, those which contrast phonemic taps/flaps with phonemic trills, and finally those three languages (< 2% of the sample) from which these classes of rhotics are absent. Since 39 languages (21%) also show the voiced uvular approximant /ʁ/ (sometimes classified as a fricative), there must be 36 languages in which this approximant forms a phonological opposition with trills or taps/flaps or both (cf. Table 2). Map 1 shows the distribution of those sample languages which lack the class of trills.

Map 1: Absence of trills

<table>
<thead>
<tr>
<th>Rhotics</th>
<th>attested</th>
<th></th>
<th>unattested</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>share</td>
<td>absolute</td>
<td>share</td>
</tr>
<tr>
<td>taps/flaps</td>
<td>24</td>
<td>13%</td>
<td>161</td>
<td>87%</td>
</tr>
<tr>
<td>approximants</td>
<td>39</td>
<td>21%</td>
<td>146</td>
<td>79%</td>
</tr>
<tr>
<td>trills</td>
<td>167</td>
<td>90%</td>
<td>18</td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 2: Diversity in the area of rhotics

Areally significant results are yielded by the distribution of lateral fricatives and lateral affricates which are somewhat special in comparison to the lateral approximants (Ladefoged and Maddieson 1996: 197). All of the sample languages give testify to the phonological class of laterals. Only a relatively small segment of the sample however provides evidence of lateral fricatives, namely twenty-two (11%) with Welsh being the sole example of a language outside the Caucasian region to employ a lateral-fricative phoneme. The Caucasian monopoly is even more pronounced with the lateral affricates which are attested in sixteen languages (9%) all of which belong to either the Northwest Caucasian or to the Northeast Caucasian phylum. There is thus a clear pattern of structural diversity which is areally and genetically based at the same time. The above Northwest Caucasian and Northeast Caucasian languages display several phonemic laterals which belong to different manners of articulation (cf. Table 3).

<table>
<thead>
<tr>
<th>laterals</th>
<th>unity</th>
<th>diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>approximants</td>
<td>+ approximants</td>
<td></td>
</tr>
<tr>
<td>fricative</td>
<td>+ fricative</td>
<td></td>
</tr>
<tr>
<td>1. Areal:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Caucasus</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Genetic:
   - Northwest Caucasian
   - Northeast Caucasian
3. Isolate:
   - Welsh

<table>
<thead>
<tr>
<th>affricate</th>
<th>+ affricate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Areal:</td>
</tr>
<tr>
<td></td>
<td>- Caucasus</td>
</tr>
<tr>
<td></td>
<td>2. Genetic:</td>
</tr>
<tr>
<td></td>
<td>- Northwest Caucasian</td>
</tr>
<tr>
<td></td>
<td>- Northeast Caucasian</td>
</tr>
</tbody>
</table>

Table 3: Unity and diversity in the area of laterals

The geolinguistic facts are, however, different with affricates (Ternes 2010: 585) because only 160 of 185 sample languages offer evidence of phonemic affricates\(^3\). This corresponds to 86% of our sample. There is thus a bipartition of the sample into a majority formed by those languages which distinguish six phonemic manners of articulation and a minority of those languages which make do with five phonemic manners of articulation. Stolz (2010: 616-619) provides several maps which show that the distribution of affricates follows an areal logic in the sense that the languages which display phonemic affricates occupy a huge uninterrupted central territory around which those languages are placed which lack phonemic affricates. The latter are thus situated on the periphery (cf. Map 2).

Map 2: Absence of affricates

Both positive similarity and negative similarity cut across some genetic boundaries since we find members of the Germanic, the Romance, the Turkic and the Uralic phylum in both classes. In sum, European languages tend to have phonemic affricates if their neighbours have them. They tend to lack phonemic affricates if their neighbours lack them too. The distribution of the phonological class of affricates thus illustrates diversity. Genetic unity characterizes the Slavic, Baltic, Northwest Caucasian, Northeast Caucasian, and South Caucasian languages because all of these languages count affricates among their phonemes. On the other hand, none of the Celtic languages in our sample shows phonemic affricates. Some language families or branches thereof behave homogeneously whereas others do not.

4.3 Place of articulation

As to the places of articulation, Ternes (2010: 584) assumes that the globally most widespread categories are also represented throughout Europe. This view of the facts is

---

\(^3\) Haarmann (1976a: 116) does not recognize any areally significant pattern and erroneously assumes that Portuguese too has phonemic affricates.
corroborated by the statistics for the ternary distinction of labial, coronal, and dorsal places of articulation all of which are present in each of the 185 sample languages. This means that all of our sample languages make use of labial and lingual places of articulation. If we exclude the affricates and focus on the tongue as active articulator, it is evident that the root position does not score as high as either the coronal or the dorsal positions. There are twenty-four languages (13% of the sample) in which the root of the tongue is the active articulator (for pharyngeals and epiglottals). Table 4 shows that unity is characteristic of the places of articulation which involve the front half of the tongue as active articulator.

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>present</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>share</td>
</tr>
<tr>
<td>Coronal</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>Dorsal</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>Radical</td>
<td>24</td>
<td>13%</td>
</tr>
</tbody>
</table>

Table 4: Unity vs diversity with major places of articulation (tongue positions)

The picture becomes more varied if we include the non-tongue positions. On the one hand, the omnipresence of labial places of articulation fits in with the above preference for places of articulation which are located in the front part of the oral cavity. However, this tendency is counterbalanced by the remarkably high share of the glottal (laryngeal) place of articulation among our sample languages, cf. Table 5.

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>share</td>
</tr>
<tr>
<td>labial</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>laryngeal</td>
<td>119</td>
<td>64%</td>
</tr>
</tbody>
</table>

Table 5: Unity vs diversity with major places of articulation (non-tongue positions)

The high number of languages with laryngeal places of articulation notwithstanding, what the values in Tables 4 and 5 are suggestive of is a bipartition of the following kind. With regard to those places of articulation which are situated in the front part of the oral cavity, the languages of Europe behave homogeneously in the sense that all of them offer evidence of the phonological classes concerned. Structural diversity is largely a matter of those places of articulation which are located in the back part of the oral cavity or beyond. Especially the feature ‘tongue root’ is associated with the languages of the Caucasus whereas the absence of the feature ‘laryngeal’ is areally less clearly distributed.

For the subdivisions of the major categories of place, the picture is even more varied. The highest rates are reported for bilabials, alveolars and velars with a share of 100% for each of the three places of articulation. As Table 6 shows there are only a further three places of articulation – palatal, labio-dental and palato-alveolar – which are attested in significantly more than two thirds of the sample languages. The glottal place of articulation is attested in slightly less than two thirds of the sample languages. All other places of articulation are minority configurations none of which comes near the 40%-mark. The statistical underrepresentation of the latter classes is additionally highlighted by grey shading in Table 6. Map 3 indicates the location of those languages.
which realize one or several of the minority configurations given in Table 6. Map 4 informs us about the geolinguistic position of those languages which do not partake in one of the majority configurations.

Map 3: Attested minority configurations (place of articulation, Table 6)

Map 4: Absence of majority configurations (place of articulation, Table 6)

<table>
<thead>
<tr>
<th>Place of articulation</th>
<th>present</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>share</td>
</tr>
<tr>
<td>bilabial</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>alveolar</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>velar</td>
<td>185</td>
<td>100%</td>
</tr>
<tr>
<td>palatal</td>
<td>176</td>
<td>95%</td>
</tr>
<tr>
<td>labio-dental</td>
<td>165</td>
<td>89%</td>
</tr>
<tr>
<td>palato-alveolar</td>
<td>165</td>
<td>89%</td>
</tr>
<tr>
<td>glottal</td>
<td>119</td>
<td>64%</td>
</tr>
<tr>
<td>labio-velar</td>
<td>69</td>
<td>37%</td>
</tr>
<tr>
<td>uvular</td>
<td>63</td>
<td>34%</td>
</tr>
<tr>
<td>alveo-palatal</td>
<td>27</td>
<td>15%</td>
</tr>
<tr>
<td>pharyngeal</td>
<td>21</td>
<td>11%</td>
</tr>
<tr>
<td>interdental</td>
<td>19</td>
<td>10%</td>
</tr>
<tr>
<td>epiglottal</td>
<td>8</td>
<td>4%</td>
</tr>
<tr>
<td>retroflex</td>
<td>3</td>
<td>2%</td>
</tr>
<tr>
<td>labio-palatal</td>
<td>1</td>
<td>&lt; 1%</td>
</tr>
</tbody>
</table>

Table 6: Unity vs diversity with places of articulation (primary subdivisions)

The palatal place of articulation is absent from nine languages which do not belong together either genetically or areally. Seven of these languages are situated on the periphery of the continent. As to the absence of the labio-dental place of articulation, the overall distribution yields a very clear pattern. Apart from Basque, nineteen of the twenty languages which lack labio-dentals are located in the European east. With sixteen Northeast Caucasian languages and a South Caucasian language, there is an undisputable areal and genetic bias. No Indo-European language reflects this negative feature.

The palato-alveolar place of articulation is not made use of in a variety of Indo-European and Uralic languages. These languages form a string of languages on the periphery of a large and compact area of genetically diverse languages which are equipped with palato-alveolar phonemes.

Both the presence and the absence of the glottal place of articulation cut across the boundaries of language families. Members of the Indo-European family are found in both groups – the same division holds for Turkic, Uralic, Afroasiatic, and Northwest Caucasian languages. In terms of the areal distribution, it is possible to distinguish two regions from which the glottal place of articulation is absent, namely a long southern stretch along the Mediterranean and a huge territory which covers East-Central Europe and much of Eurasia. In contrast, the European North and the southeasternmost parts of
the continent host the bulk of those languages which make use of the glottal place of articulation for phonological distinctions.

Double articulation is a rarity in Europe. The labio-velar place of articulation stands out statistically in the sense that it is attested in slightly more than a third of the sample languages. In practically all of these cases, we are dealing with a voiced labio-velar approximant /w/. Of the 69 languages which offer evidence of labio-velars, 28 (41%) belong to the three Caucasian families. Another ten non-Indo-European languages display labio-velars so that the 31 Indo-European languages with labio-velars account for only 45% of this group of languages. The phenomenon is absent from most of the languages of East Central Europe and Eurasia.

Ternes (2010: 584) assumes that uvular and pharyngeal places of articulation are excluded from the repertoire of the languages of Europe (with the putative exception of Maltese). This statement needs to be revised. It is true that 37 of the 63 languages which have uvulars belong to the Caucasian language families. Another eight languages are situated outside of Ternes’s conception of Europe. However, there are eighteen languages (29%) which are located inside the area under scrutiny and which nevertheless display uvulars. Many of these languages show the uvular trill /ʁ/ or the uvular fricative /ʁ/. Ternes (2010: 586) concedes that the geolinguistic distribution of these uvular rhotics yields an areally significant pattern with a centre of diffusion in the so-called Charlemagne-Sprachbund (van der Auwera 1998).

Alveo-palatal places of articulation are reported for twenty-seven languages eleven of which are members of the Indo-European language family. With eighteen languages, the two-thirds majority of the languages equipped with alveo-palatal phonemes are well established in the European east. Six of these languages belong to the Northwest Caucasian family. In the western half of the continent, the languages with alveo-palatal phonemes are not connected to each other either genetically or areally.

Except for Cypriot Arabic, those languages which make use of the pharyngeal place of articulation are situated in the Caucasus and belong to the Northwest Caucasian and the Northeast Caucasian phyla. Similarly, the very small number of languages which attests to the epiglottal place of articulation – all of them are members of Northeast Caucasian – lends support to the idea that the back part of the oral cavity is disfavored as an area in which speech sounds are produced in the bulk of the languages of Europe.

The interdental place of articulation is largely absent from the huge central and eastern area. None of the Caucasian languages testify to phonemic interdentals. Similarly, Slavic and Baltic languages do not employ this place of articulation. Interdentals are attested in a variety of languages in the south (including Albanian, Greek and Cypriot Arabic) and in the northwest of the continent. There are also three languages on the eastern fringe of Europe. Germanic, Celtic, Romance, Semitic, Turkic, and Uralic languages behave heterogeneously with regard to the interdental place of articulation because some of them have interdental phonemes whereas their sister-languages lack this phonological class altogether, e.g. Welsh does but Irish does not (any more).

The two remaining places of articulation – retroflex and labio-palatal – are only marginally represented in Europe. For retroflexion, caution is necessary since it is not always clear whether we are dealing with a phonemic property or an allophonic realization, e.g. in Faorese. Ternes (2010: 584) assumes a Northwest-European isogloss of retroflexion whereas the labio-palatal place of articulation is unique to French which has phonemic /ɥ/. These rarities single out small numbers of languages from their families and thus contribute to the structural diversity of these families.
4.4. Phonation

Ternes (2010: 587) observes that the opposition voiced vs voiceless is very common throughout Europe. As a matter of fact, the feature [+voice] is present in each of the 185 sample languages. However, in two of the languages in the east – Chuvash (Turkic) and Karelian (Uralic) – this feature is not properly phonemic since there is no voiced counterpart. Both languages show /v/ but lack */f/. In 183 languages i.e. in 99% of the European languages, phonation is phonologically distinctive. It remains to be seen over which phonological classes the domain of the voice correlation extends. With reference to manners of articulation Ternes (2010: 587) states that the classes of stops and fricatives are the most likely manners of articulation which are subject to voice distinctions. This statement is corroborated by the behaviour of our sample languages. There are only eleven languages (6%) which do not have phonemic voice distinctions with fricatives. With just eight languages (4%) the group of languages which lack evidence of the voice correlation with stops is even smaller. It is thus absolutely exceptional for a language of Europe to lack the voice correlation in these phonological classes, the Karelian variety of Archangelsk and Chuvash being the only examples of languages which do not have phonemic voice distinctions in both of the classes. Affricates however tell a different story because 41 languages (22% of the entire sample) allow for voiceless affricates only. This is equivalent of a share of 26% of the 160 languages which show the phonological class of affricates.

In areal terms, the absence of the voice correlation with stops and fricatives is largely confined to languages which are situated on the periphery. In the case of the stops, there is an accumulation in the east with four Uralic and Turkic languages of Eurasia. For the fricatives, the Iberian Peninsula seems to be largely excluded from the voice-correlation isogloss. These two areal centres are also included in the geolinguistic domain characterized by the absence of voiced affricates (in languages with affricates). Twenty-nine or 71% of the 41 languages are located in the eastern half of the European continent. Another three are at home on the Iberian Peninsula. With fourteen languages, the North Caucasian family accounts for 34% of the languages without voiced affricates. Twenty-three (56%) of these languages are non-Indo-European (Northeast Caucasian, Uralic, Turkic, Basque). There is thus a clear areal ratio behind the distribution of the negative property under scrutiny.

In Europe, voiceless sonorants belong among the rara and rarissima (Ternes 2010: 587). Some of the reported cases are doubtful insofar as the voiceless units perhaps might be better analyzed as allophones. The indisputable cases are represented by the Uralic languages Erzya and Mokša Mordvin, Kildin and North Saami which show phonemic voice contrasts with laterals, nasals and rhotics. Chechen and Welsh illustrate the opposition of voiced and voiceless rhotics.

Table 7 provides a synoptical view of the findings regarding the distribution of the voice correlation over manners of articulation. The absence of voice contrasts is infrequent in general and especially rare with obstruents. This does not hold for affricates for which the reverse is true. Note however, that affricates themselves are not attested as widely as stops or fricatives, i.e. they constitute a crosslinguistically marked phonological class which displays marked behavior insofar as it does not allow for distinctions made with other obstruent classes.
Is there a correlation between place of articulation and the presence or absence of the voice? The various places of articulation behave differently as to the susceptibility to voice contrasts. If we ask whether or not the distinction of voiceless vs voiced is phonemic with a given place of articulation in the languages of our sample, we get the following results, cf. Table 8.

<table>
<thead>
<tr>
<th>manner of articulation</th>
<th>Present</th>
<th>absent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td>share</td>
</tr>
<tr>
<td>general</td>
<td>183</td>
<td>99%</td>
</tr>
<tr>
<td>stops</td>
<td>177</td>
<td>96%</td>
</tr>
<tr>
<td>fricatives</td>
<td>174</td>
<td>94%</td>
</tr>
<tr>
<td>affricates</td>
<td>41</td>
<td>22% (of 185)</td>
</tr>
<tr>
<td>sonorants</td>
<td>6</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 7: Unity vs diversity in the area of phonation (manner of articulation)

In Table 8, the places of articulation come in the order of increasing shares for those languages which lack voice contrasts. Since the languages differ as to the absence/presence of many of the places of articulation in their phonological system, the figures for sharing are calculated on the basis of the number of languages which make use of a given place of articulation. Except for the three categories at the bottom of Table 8, all the values are indicative of a preference for phonemic voice contrasts. The highest rate of absence of voice contrasts – apart from the final three places of articulation – exceeds a third of the languages involved only by a narrow margin. For the four places of articulation which are attested in over 90% of the sample languages, the absence of the

<table>
<thead>
<tr>
<th>place of articulation</th>
<th>absent</th>
<th>share of place</th>
<th>total of place</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>absolute</td>
<td></td>
<td>Absolute</td>
</tr>
<tr>
<td>alveolar</td>
<td>5</td>
<td>3%</td>
<td>185</td>
</tr>
<tr>
<td>palatal</td>
<td>7</td>
<td>4%</td>
<td>176</td>
</tr>
<tr>
<td>velar</td>
<td>8</td>
<td>4%</td>
<td>185</td>
</tr>
<tr>
<td>bilabial</td>
<td>8</td>
<td>4%</td>
<td>185</td>
</tr>
<tr>
<td>alveo-palatal</td>
<td>4</td>
<td>15%</td>
<td>27</td>
</tr>
<tr>
<td>palato-alveolar</td>
<td>31</td>
<td>19%</td>
<td>165</td>
</tr>
<tr>
<td>labio-dental</td>
<td>35</td>
<td>21%</td>
<td>165</td>
</tr>
<tr>
<td>interdental</td>
<td>5</td>
<td>26%</td>
<td>19</td>
</tr>
<tr>
<td>pharyngeal</td>
<td>6</td>
<td>29%</td>
<td>21</td>
</tr>
<tr>
<td>retroflex</td>
<td>1</td>
<td>33%</td>
<td>3</td>
</tr>
<tr>
<td>uvular</td>
<td>21</td>
<td>33%</td>
<td>63</td>
</tr>
<tr>
<td>epiglottal</td>
<td>3</td>
<td>38%</td>
<td>8</td>
</tr>
<tr>
<td>labio-palatal</td>
<td>0</td>
<td>100%</td>
<td>1</td>
</tr>
<tr>
<td>labio-velar</td>
<td>0</td>
<td>100%</td>
<td>69</td>
</tr>
<tr>
<td>glottal</td>
<td>0</td>
<td>100%</td>
<td>119</td>
</tr>
</tbody>
</table>

Table 8: Unity vs diversity in the area of phonation (place of articulation)
voice correlation is minimal. Among the eleven places of articulation with higher absence rates for voice contrasts, eight categories belong to the minority configurations (marked grey in Table 6 above). Voice is phonologically not distinctive in three cases. The labio-palatal approximant is a trivial case because, phonetically, it is exempt from voice contrasts. For labio-velars and glottals, however, it is phonetically possible to distinguish voiced from voiceless realizations. Nevertheless, in our sample, all labio-velars represent voiced /w/ (and never voiceless /ʍ/ except in conservative varieties of Scottish and rural Irish English which have both /w/ and /ʍ/, i.e. in varieties not included in our sample). For the glottals, there are examples of voiced and voiceless phonemes. However, they never co-occur in phonemic oppositions in one and the same language.

Unsurprisingly, the absence of voice contrasts with pharyngeal and epiglottal places of articulation has a relatively clear areal foundation. Since these places of articulation are almost a 100% monopoly of the Caucasian families, it is only logically that the eight of nine languages which lack voiceless-voiced pairs of phonemes with these places of articulation are spoken in the Caucasus. Six of them belong to the Northeast Caucasian family, two others – to the Northwest Caucasian family; the only non-Caucasian language without voiceless-voiced pairs within pharyngeal and epiglottal consonants is Cypriot Arabic (Afroasiatic phylum, Semitic branch). In all other cases, the genetic composition of the groups of languages which lack phonemic voice contrasts is considerably more heterogeneous. However, at least for some of the places of articulation, an areal factor is discernible. This is the case with the bilabials, for instance. The eight languages without an opposition of voiceless vs voiced bilabials are all situated on the periphery of the European continent, see following map.

Map 5: Absence of voice correlation with bilabials

This peripheral location also holds for the twenty-one languages which lack voice contrasts with uvulars, see Map 6.

Map 6: Absence of voice correlation with uvulars

The situation is similar though not identical for the places of articulation alveo-palatal, palatal, velar, and uvular which lack voice distinctions mostly in languages which are located on the northern, southern or eastern rim of the continents, see Map 7.

Map 7: Absence of voice correlation with alveo-palatals/palatals/velars

In each of these categories, however, there is also an example of a language – Danish twice, Swedish, Franco-Provençal and the German variety of Brigg once each – which is located closer to the interior of the area under scrutiny. For labio-dentals and palato-alveolars, the facts are more complex. Twenty-five i.e. 71% of the languages without voice contrasts in the class of labio-dentals are located in the eastern part of Europe, see Map 8.

Map 8: absence of voice correlation with labio-dentals
Twenty-one i.e. 60% of the same group of languages are non-Indo-European languages. There is also a majority of languages (63%) which are located on the periphery of the European map. In the case of the palato-alveolar place of articulation, the distribution is more balanced genetically as well as areally although there is a slight majority of 58% for peripheral languages, see Map 9.

Map 9: Absence of voice correlation with palato-alveolars

From an overall perspective, the contrast voiced vs voiceless is a majority situation. The size of its domain varies with the manners and places of articulation. Languages in the interior of Europe are more likely to employ phonemic voice contrasts whereas with peripheral languages the probability that the voiceless-voiced opposition is phonologically relevant diminishes.

4.5 Secondary articulation

According to Ternes (2010: 586-587), secondary articulations are not prominently represented among the languages of Europe. The author discounts labialization and glottalization as European features because he excludes the Caucasian region from his version of Europe. For our project however, labialization is a characteristic trait of the Northwest Caucasian and Northeast Caucasian families. Moreover “glottalization” (in Ternes’ terms) is treated as an equivalent of the non-pulmonic class of ejectives in the present chapter. This class is again typical of the Caucasian region (see discussion above).

Our own count yields 73 languages which make use of phonemic secondary articulations. This correspond to a share of 40% of the sample which is not a negligible quantity. The genetic affiliation of these 73 languages is given in Table 9. For convenience, Northeast and Northwest Caucasian languages are lumped together to form a fictitious phylum Caucasian. There is no evidence of secondary articulations with phonemic status in the South Caucasian languages of our sample. Map 10 captures the geographic distribution of languages which display phonemic secondary articulations.

Map 10: Phonemic secondary articulation

<table>
<thead>
<tr>
<th>Affiliation</th>
<th>absolute</th>
<th>share (of 73)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indo-European</td>
<td>40</td>
<td>55%</td>
</tr>
<tr>
<td>Caucasian (Northeast, Northwest)</td>
<td>16</td>
<td>22%</td>
</tr>
<tr>
<td>Uralic</td>
<td>12</td>
<td>16%</td>
</tr>
<tr>
<td>Turkic and Mongolic</td>
<td>5</td>
<td>7%</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 9: Genetic affiliation of languages with phonemic secondary articulations

Going by the standards in Ternes (2010), the only secondary articulation that yields statistically significant values is palatalization. Stadnik (2002) demonstrates that phonemic palatalization is an areal feature of the European east and Eurasia which is shared by Indo-European, Uralic, Turkic, and Mongolian languages of these regions (cf. also Haarmann 1976a: 116). The presence or absence of the palatalization correlation
splits genetic families in two branches, cf. the Baltic phylum: Lithuanian displays phonemic palatalization whereas its next-of-kin Latvian lacks palatalization. Outside the compact eastern area there is a much smaller island of palatalization which comprises the Goidelic languages Irish and Scots-Gaelic in the far Northwest (Ternes 2010: 586).

Table 10 proves the dominance of palatalization among the types of phonemic secondary articulations. Almost three quarters of all instances of phonemic secondary articulations are covered by palatalization alone.

<table>
<thead>
<tr>
<th>Secondary articulation</th>
<th>Absolute</th>
<th>Share (of 73)</th>
<th>Caucasian languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Palatalization</td>
<td>56</td>
<td>77%</td>
<td>7</td>
</tr>
<tr>
<td>Labialization</td>
<td>15</td>
<td>21%</td>
<td>15</td>
</tr>
<tr>
<td>Aspiration</td>
<td>11</td>
<td>15%</td>
<td>2</td>
</tr>
<tr>
<td>Velarization</td>
<td>6</td>
<td>8%</td>
<td>0</td>
</tr>
<tr>
<td>Pharyngelization</td>
<td>4</td>
<td>6%</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 10: Types of phonemic secondary articulations (statistics)

At the same time, the values featured in Table 10 are also indicative of genetically based preferences since labialization and pharyngelization are exclusive to Northeast and Northwest Caucasian languages whereas these two families account for relatively small shares with palatalization and aspiration – and do not partake in the velarization isogloss at all. It is worth noting however that even without the contribution of the eighteen Caucasian languages, there is a clear preference for phonemic secondary articulations to occur in languages of the European east. Of the 73 languages with phonemic secondary articulations, only seven (10% of this class) are located in the western half of Europe whereas 66 (90%) are to be found on the other side of the vertical dividing line. Those western languages which show phonemic secondary articulations are members of the Celtic branch of Indo-European.

The strong position of phonemic secondary articulations in the east is in line with Haarmann’s (1976a: 114) observation that the number of phonemes per language is noticeably higher in the European east than in the west.

As can be seen in Table 11, the different manners of articulation are affected by secondary articulations to relatively low degrees.

<table>
<thead>
<tr>
<th>Manner of articulation</th>
<th>Absolute</th>
<th>Share of manner</th>
<th>Total of manner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>61</td>
<td>33%</td>
<td>185</td>
</tr>
<tr>
<td>Fricatives</td>
<td>49</td>
<td>27%</td>
<td>185</td>
</tr>
<tr>
<td>Laterals</td>
<td>48</td>
<td>26%</td>
<td>185</td>
</tr>
<tr>
<td>Nasals</td>
<td>45</td>
<td>24%</td>
<td>185</td>
</tr>
<tr>
<td>Affricates</td>
<td>35</td>
<td>22%</td>
<td>160</td>
</tr>
<tr>
<td>Rhotics (trills, taps/flaps)</td>
<td>27</td>
<td>15%</td>
<td>178</td>
</tr>
<tr>
<td>(labio-velar) Approximants</td>
<td>6</td>
<td>9%</td>
<td>69</td>
</tr>
</tbody>
</table>

Table 11: Secondary articulations and manner of articulation

The upper limit is a third of the sample languages. For those manners of articulation which are not attested in all the sample languages, the share of secondary articulations is
even more restricted. The geographic picture is largely inconclusive in the sense that there is no clear areal preference for any manner of articulation. We therefore do not provide a separate map for this issue.

The question arises whether or not similar patterns can be observed in correlation to the places of articulation. Our sample languages do not allow secondary articulations with the following places of articulation: interdental, alveo-palatal, retroflex, labio-palatal and epiglottal. For the remaining ten places of articulation, the statistics yield results which are reminiscent of those of the manners of articulation, cf. Table 12.

<table>
<thead>
<tr>
<th>place of articulation</th>
<th>absolute</th>
<th>share of place</th>
<th>total of place</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar</td>
<td>67</td>
<td>36%</td>
<td>185</td>
</tr>
<tr>
<td>Velar</td>
<td>45</td>
<td>24%</td>
<td>185</td>
</tr>
<tr>
<td>Bilabial</td>
<td>42</td>
<td>23%</td>
<td>185</td>
</tr>
<tr>
<td>Uvular</td>
<td>14</td>
<td>22%</td>
<td>63</td>
</tr>
<tr>
<td>Labio-dental</td>
<td>25</td>
<td>15%</td>
<td>165</td>
</tr>
<tr>
<td>Palato-alveolar</td>
<td>15</td>
<td>9%</td>
<td>165</td>
</tr>
<tr>
<td>Labio-velar</td>
<td>6</td>
<td>9%</td>
<td>69</td>
</tr>
<tr>
<td>Pharyngeal</td>
<td>2</td>
<td>10%</td>
<td>21</td>
</tr>
<tr>
<td>Glottal</td>
<td>8</td>
<td>7%</td>
<td>119</td>
</tr>
<tr>
<td>Palatal</td>
<td>1</td>
<td>&lt;1%</td>
<td>176</td>
</tr>
</tbody>
</table>

Table 12: Secondary articulations and place of articulation

The three places of articulation which are attested universally in our sample also document the highest degrees of secondary articulation. The shares are comparable to those of the top-ranking manners of articulation in Table 11 above. Since palatalization is the most important type of secondary articulation, it is hardly surprising that the palatal place of articulation is subject to phonemic secondary articulation only marginally.

Glottal, pharyngeal and uvular places of articulation are susceptible to secondary articulations predominantly in Northeast and Northwest Caucasian languages. For the remaining places of articulation, the geolinguistic distribution is random, i.e. no areal foci can be identified.

5 Meso-level: Primary articulations

Haarmann (1976a: 115-116) makes several unsystematic observations about the geolinguistic distribution of certain individual phonemes. He mentions the absence of single phonemes from the phoneme systems of individual languages and remarks on the commonality of a variety of other phonemes. Ternes (2010: 583-587) comments briefly upon a selection of specific phonemes and their presence or absence in some of the languages of Europe. On closer inspection, the statistics of the phonemic units reveals rather interesting patterns of structural unity and diversity, cf. Tables 13-14.

For expository reasons, we have done statistics only for primary articulations (i.e. secondary articulations were not counted separately). On purely statistical grounds, the phonemes are divided into two classes. Table 13 provides the rank order of those phonemic primary articulations which are attested in more than half of our sample languages. This survey is complemented by Table 14 which contains the statistical
information about those phonemic primary articulations which are attested in less than half of the sample languages.

<table>
<thead>
<tr>
<th>rank</th>
<th>languages</th>
<th>share</th>
<th>phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>185</td>
<td>100%</td>
<td>/p/, /t/, /m/, /n/</td>
</tr>
<tr>
<td>5-6</td>
<td>184</td>
<td>99%</td>
<td>/k/, /s/</td>
</tr>
<tr>
<td>7</td>
<td>181</td>
<td>98%</td>
<td>/l/</td>
</tr>
<tr>
<td>8-9</td>
<td>177</td>
<td>96%</td>
<td>/l/, /d/</td>
</tr>
<tr>
<td>10</td>
<td>174</td>
<td>94%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>11</td>
<td>163</td>
<td>88%</td>
<td>/ʃ/</td>
</tr>
<tr>
<td>12</td>
<td>160</td>
<td>86%</td>
<td>/ɬ/</td>
</tr>
<tr>
<td>13-14</td>
<td>159</td>
<td>86%</td>
<td>/ʃl, /ʃy/</td>
</tr>
<tr>
<td>15</td>
<td>149</td>
<td>81%</td>
<td>/l/</td>
</tr>
<tr>
<td>16</td>
<td>146</td>
<td>79%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>17</td>
<td>140</td>
<td>76%</td>
<td>/ʃʃ/</td>
</tr>
<tr>
<td>18</td>
<td>136</td>
<td>74%</td>
<td>/ʃʃ/</td>
</tr>
<tr>
<td>19</td>
<td>133</td>
<td>72%</td>
<td>/ʃʃʃ/</td>
</tr>
<tr>
<td>20</td>
<td>117</td>
<td>63%</td>
<td>/ʌ/</td>
</tr>
<tr>
<td>21</td>
<td>109</td>
<td>59%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>22</td>
<td>107</td>
<td>58%</td>
<td>/ʃʃʃ/</td>
</tr>
</tbody>
</table>

Table 13: Statistics of phonemic primary articulations which occur in more than half of the sample languages

<table>
<thead>
<tr>
<th>rank</th>
<th>languages</th>
<th>share</th>
<th>phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>69</td>
<td>37%</td>
<td>/w/</td>
</tr>
<tr>
<td>24</td>
<td>66</td>
<td>36%</td>
<td>/ʃʃ/</td>
</tr>
<tr>
<td>25</td>
<td>52</td>
<td>28%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>26</td>
<td>48</td>
<td>26%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>27</td>
<td>44</td>
<td>24%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>28</td>
<td>40</td>
<td>22%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>29</td>
<td>39</td>
<td>21%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>30</td>
<td>38</td>
<td>21%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>31</td>
<td>36</td>
<td>19%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>32-33</td>
<td>35</td>
<td>19%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>18%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>35</td>
<td>27</td>
<td>15%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>36</td>
<td>25</td>
<td>14%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>37</td>
<td>24</td>
<td>13%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>38</td>
<td>23</td>
<td>12%</td>
<td>/ŋ/</td>
</tr>
<tr>
<td>39</td>
<td>21</td>
<td>11%</td>
<td>/ŋ/</td>
</tr>
</tbody>
</table>
What strikes the eye immediately is the huge gap of twenty-one points which separates the last entry of Table 13 from the first entry of Table 14. This clear-cut distinction of majority and minority configurations precludes the possibility of treating the above data as a continuum. Equally noteworthy is the surprisingly small number of phonemic primary articulations which can be termed truly pan-European since there are only four units which are reported for each of the 185 sample languages. This means that, already on the meso-level, structural unity is restricted. Diversity is connected to 72 out of 76 phonemic primary articulations (including affricates). 95% of the phonemic primary articulations are attested in less than 185 languages, i.e. crosslinguistic variation is the rule and not the exception. This observation holds both for positive and negative phenomena.

From an overall perspective, one may say that for the top-ranking ten percent in Table 13 there is a strong tendency towards positive structural unity. The other way round, the last ten percent of Table 14 are suggestive of an equally strong tendency towards negative structural unity because the phonemic primary articulations found at the bottom of Table 14 are avoided by 90% or more of the sample languages.

The group of languages which lack one or more of the ten most frequently attested primary articulations overlaps with that of the languages which attest one or more of the 35 least frequently attested primary articulations. Fifteen languages are represented in both groups. Seven of these languages are situated in the Caucasian region, eleven languages are non-Indo-European. Twenty-one of the 35 least frequently attested primary articulations are attested in Northeast and/or Northwest Caucasian languages, see Map 11.

Map 11: Presence of ranks 42-76 (Table 14)
Twenty-six of the same class are attested in non-Indo-European languages. On the European map, see Map 12, a relatively clear areal pattern emerges according to which those languages which lack one or more of the most frequently attested primary articulations are overwhelmingly peripheral.

Map 12: Absence of ranks 1-10 (Table 13)

Sixteen of these languages are located on the eastern or northern rim of the continent. Superficially, the geographic distribution of the languages which show rare primary articulations does not seem to be remarkable. However, if we concentrate on those languages which are repeatedly involved with infrequent primary articulations the situation changes drastically. Languages which show two or more of the rare primary articulations form a chain which encircles the large central region from which rare primary articulations are absent or where they are attested only occasionally. Map 13 accounts for the geographical location of those languages which combine the attestation of minority configurations with the absence of majority configurations.

Map 13: Attestation of minority configurations (ranks 42-76) and absence of majority configurations (ranks 1-10) in combination

Haarmann (1976a: 116) claims that most of the majority and minority configurations fail to yield an areally significant distribution. Only for a small number of individual phonemes does he assume that there are restrictions on their occurrence in the phonological systems of the languages of Europe such that their absence from certain regions gives rise to a negatively-defined area. Our own results, however, are indicative of a considerably higher degree of areal sensitivity of the phonological phenomena under scrutiny than Haarmann admits.

5.1 Excursus: Glimpses of phylum-internal diversity

In the previous sections, we have alluded repeatedly to aspects of phonological homogeneity vs heterogeneity within language families. To emphasize that genetic relatives do not necessarily behave absolutely the same phonologically, we provide statistical information about the structural variation within three branches of the Indo-European language family (Slavic, Romance, and Germanic) and the Uralic language family.

In Table 15, we determine how many of the members of a given phylum show a given manner of articulation. Since all of the languages have stops, fricatives, nasals and approximants and lack lateral fricatives as well as lateral affricates, we take account only of those manners of articulation for which variation can be expected. The languages are ordered according to their decreasing size. The values indicate how many of the languages of a phylum/family document the manner of articulation under scrutiny.

<table>
<thead>
<tr>
<th>family</th>
<th>trills</th>
<th>taps</th>
<th>affricates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slavic (27)</td>
<td>27 = 100%</td>
<td>0 = 0%</td>
<td>27 = 100%</td>
</tr>
<tr>
<td>Uralic (24)</td>
<td>24 = 100%</td>
<td>0 = 0%</td>
<td>22 = 92%</td>
</tr>
<tr>
<td>Romance (20)</td>
<td>20 = 100%</td>
<td>4 = 20%</td>
<td>17 = 85%</td>
</tr>
<tr>
<td>Germanic (17)</td>
<td>11 = 65%</td>
<td>1 = 6%</td>
<td>10 = 59%</td>
</tr>
</tbody>
</table>
There is a marked difference of eastern European languages and western European languages. The Slavic group offers evidence of homogeneity because either all of its members attest a manner of articulation or they lack it altogether. Uralic languages behave similarly regarding trills and taps but fail to reach full homogeneity with affricates. For Romance languages, homogeneity is restricted to the class of trills whereas there is variation in connection with the other two phonological classes. The Germanic branch of Indo-European is characterized by diversity in all three categories.

We conclude this section with a glance at the variation within language families concerning places of articulation. Since all the languages show bilabials, alveolars and velars while lacking evidence of pharyngeals and epiglottals, Table 16 informs us only about those ten places of articulation which are subject to dissimilarity.

<table>
<thead>
<tr>
<th>place of articulation</th>
<th>Slavic (27)</th>
<th>Uralic (24)</th>
<th>Romance (20)</th>
<th>Germanic (17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>labio-dental</td>
<td>27 = 100%</td>
<td>23 = 96%</td>
<td>20 = 100%</td>
<td>17 = 100%</td>
</tr>
<tr>
<td>interdental</td>
<td>0 = 0%</td>
<td>3 = 13%</td>
<td>2 = 10%</td>
<td>5 = 29%</td>
</tr>
<tr>
<td>Palato-alveolar</td>
<td>27 = 100%</td>
<td>21 = 88%</td>
<td>9 = 45%</td>
<td>14 = 82%</td>
</tr>
<tr>
<td>retroflex</td>
<td>0 = 0%</td>
<td>0 = 0%</td>
<td>0 = 0%</td>
<td>2 = 12%</td>
</tr>
<tr>
<td>alveo-palatal</td>
<td>3 = 11%</td>
<td>8 = 33%</td>
<td>6 = 30%</td>
<td>1 = 6%</td>
</tr>
<tr>
<td>palatal</td>
<td>27 = 100%</td>
<td>24 = 100%</td>
<td>20 = 100%</td>
<td>14 = 82%</td>
</tr>
<tr>
<td>Uvular</td>
<td>1 = 4%</td>
<td>0 = 0%</td>
<td>2 = 10%</td>
<td>5 = 29%</td>
</tr>
<tr>
<td>Glottal</td>
<td>8 = 30%</td>
<td>14 = 58%</td>
<td>7 = 35%</td>
<td>14 = 82%</td>
</tr>
<tr>
<td>labio-velar</td>
<td>7 = 26%</td>
<td>2 = 8%</td>
<td>7 = 35%</td>
<td>4 = 24%</td>
</tr>
<tr>
<td>labio-palatal</td>
<td>0 = 0%</td>
<td>0 = 0%</td>
<td>1 = 5%</td>
<td>0 = 0%</td>
</tr>
</tbody>
</table>

This time the picture is not as straightforward as with the manners of articulation. The Slavic languages are especially homogeneous as a group because in six out of ten categories they display absolute solidarity either by attesting or by failing to attest a given category. In the case of Uralic, solidarity is restricted to four categories. In three of these cases, it is the absence of the place of articulation which unites the languages of this family. Germanic (with two cases of solidarity) and Romance (with three cases of solidarity) tend towards heterogeneity.

The considerations here show that identical genetic affiliation does not guarantee that languages are very similar phonologically. The language families differ, however, as to the degree of structural variation of their members. In comparison to the Slavic languages, Germanic languages, for instance, are relatively dissimilar from each other.

6 Conclusions

The conclusions for this chapter are divided into two subsections the first of which provides a discussion of the major insights gained from evaluating the above data. The second subsection outlines a possible continuation of the project outlined in this study.
6.1 Evaluation

It is necessary to state that the languages of Europe are neither all the same phonologically nor do they offer evidence of an overdose of diversity in the realm of phonology. The distribution of phonological properties over the languages of our sample is not entirely random. On the contrary, the above presentation suggests that there are areal factors which shape the phonological diversity or unity of Europe.

We have to state the obvious by saying that the higher the degree of generalization and/or abstraction, the higher the degree of similarity. The more concrete the units are that are compared to each other the higher is the degree of dissimilarity. The languages of Europe thus are very similar to each other on the highest level whereas they tend to be dissimilar on the lowest level of comparison.

Diversity comes into play already on the highest possible level since languages with two airstream mechanisms can be distinguished from those with only one airstream mechanism. With every subsequent level the degree of structural variation over the sample increases to culminate at the level of primary articulations. Thus, it can be said that independent of the macro-level and meso-level, Europe is never fully homogeneous in phonological terms.

At the same time, there is a common core shared by all of the sample languages on all of the levels. This basis of structural similarity shrinks with every step we take in the direction of the micro-level. On the highest level, one of two options – pulmonic vs non-pulmonic – is chosen by 100% of the sample languages. With primary articulations, however, only four out of 76 options are represented in each of the sample languages. Dissimilarity is a crucial factor not only on the level of primary articulations. The differences of the sample languages are not random though.

The principal differences we observe reflect areal biases. First of all, Ternes’ (1998, 2010) assumption that the Caucasian languages display characteristic traits which separate them from the bulk of the European languages finds corroboration in a variety of cases starting with the non-pulmonic airstream mechanism via the pharyngeal and epiglottal places of articulation down to the many examples of rara and rarissima with primary articulations attested in the Northeast and Northwest Caucasian languages – either exclusively or vastly predominantly. In several cases, there is thus a bipartition of Europe with the Caucasus hosting the representatives of the minority configuration as opposed to the majority configuration which dominates outside this region.

However, this does not mean that the inclusion of the languages of the Caucasus distorts the picture of the areal-phonological composition of Europe. If we subtract the Caucasian languages from the sample, Europe still does not turn into a fully homogeneous entity. Mutatis mutandis the position of Greenlandic in our sample is similar to that of the Caucasian languages since this member of the Eskimo-Aleut language family behaves largely like other languages which are situated on the periphery of the continent.

In point of fact, the geolinguistic distribution of the phonological phenomena corresponds often to a distinction of periphery vs centre/interior. Minority configurations tend to occur in languages which are remote from the centre of the European continent. More often than not, these languages are situated on the outskirts or near to the boundaries of Europe. In contrast, majority configurations are found frequently in
languages which occupy a sizable portion of the interior of the continent. The territory these languages claim is not only larger than that of the minority configurations but normally also forms an extended and uninterrupted contiguous neighborhood whereas the languages which reflect minority configurations may not always be direct neighbors of each other on the periphery.

These distribution patterns are reminiscent of the isoglosses identified by Haspelmath (2001) in his search for evidence of the Standard Average European linguistic area. We are cautious to claim that our data confirm this supposed Sprachbund which has been postulated almost exclusively on the basis of non-phonological phenomena. Nevertheless, several of the phonological properties discussed above yield distribution patterns which reflect a cline from east to west in the sense that the eastern half of the continent is characterized by preferences which differ from those which are most typical of the western half and vice versa. This rule of thumb holds even if we discount the evidence provided by the languages of the Caucasian region. The overwhelmingly eastern basis of the secondary articulations is a showcase of the east-west divide.

The latter case is also a paradigm case of an areal preference that disregards genetically defined boundaries. The further to the west a Slavonic language is located, the less likely it is that it employs phonemic palatalization. In its eastern sister-languages, however, palatalization is the rule. Genetically unrelated or only distantly related languages in the Eurasian region behave like their Slavonic neighbors and distinguish plain vs palatalized consonants. On a minor scale, similar areal preferences of genetically mixed groups of languages can be identified also with other phonological properties. No language family is absolutely homogeneous. Their internal diversity can often be connected to areal factors.

Yet these are all comparatively vague observations which need to stand the test of dedicated in-depth studies in the future. We are still largely in the dark as to the possibility of identifying correlations and implications. It is of course clear that in the top-down architecture of phonology the lower levels and higher levels are logically connected to each other. However, how categories of the same rank – say, manner of articulation and place of articulation – interact is a problem that calls for closer scrutiny. The catalogue of open questions is long.

6.2 Towards contact phonology

If genetically unrelated or only distantly related neighboring languages share phonological properties which are unknown in their closest relatives, what comes to mind first is an explanation which assumes diffusion via language contact. In this study, we have refrained from any kind of explanation of the distribution patterns we have observed. That language contact must be a strong factor in the areal phonology of Europe is beyond doubt. However, there is as yet no fully developed framework within which phonological convergence can be studied systematically.

According to Thomason’s (2001: 70-71) borrowing scale, phonological systems are affected by language contact from stage 2 onwards. The processes become linguistically interesting mostly in the final stages, i.e. when the degree of bilingualism with the donor language is very high among replica-language speakers. There is ample evidence of the transfer of phonological properties in language contact situations worldwide (Matras 2007: 36-40). Yet, a full-blown linguistic paradigm of contact phonology
still needs to be developed. There is a research program termed loan phonology which mainly inquires into the mechanisms of phonological adaptation of loanwords (Calabrese and Wetzels 2009: 1; Uffmann 2014). However, the phonological treatment of loanwords does not exhaust the phenomenology of phonologies under the conditions of language contact.

When studying the areal diffusion of phonological properties, the distinction of direct and indirect borrowing introduced by Aikhenvald (2002: 33-50) is helpful. However, as Sakel (2007: 17-18) shows it is relatively difficult to disentangle (overt) matter-replication from (covert) pattern-replication in the realm of phonology.

The European facts outlined above invite an interpretation that is inspired by insights from the study of language contact. However, one should not mistake any kind of similarity as proof of contact-induced convergence. Owing to the universal constraints on phonological systems, there are far too many chances that structural similarities of languages of the same geographic region arise from general phonological principles and not from language contact. To determine the correct interpretation, it is urgently necessary that the similarities and dissimilarities of the phonologies (not only) of the languages of Europe be recorded and evaluated as completely as possible. It is hoped that this chapter has shown this to be feasible and promising for the field of areal linguistics in general.

Acknowledgements

This study is intended as a preliminary to our project The Areal-Phonological Atlas of the Languages of Europe (abbreviated to Europhonology). We are grateful to Darja Appelganz, Sonja Hauser, Julia Nintemann, Robin Okrongli and Benjamin Saade for their kind support in technical matters and logistics. We thank Raymond Hickey for inviting us to contribute to the Handbook of Areal Linguistics. The responsibility for what is said in this article is exclusively ours.

References


Maddieson, Ian. 2005. (a) Consonant inventories; (b) Vowel quality inventories; (c) Consonant-vowel ratio; (d) Voicing in plosives and fricatives; (e) Voicing and gaps in plosive systems (f) Uvular consonants; (g) Glottalized consonants; (h) Lateral consonants; (i) Front rounded vowels; (g) Syllable structure; (k) Tone; (l) Absence of common consonants; (m) Presence of uncommon consonants. In Martin Haspelmath et al. (eds.), The World Atlas of Language Structures. Oxford: Oxford University Press, pp. 10-41, 50-61, 78-85.


Appendix

Sample (185)
Indo-European (98)
Germanic (17): Danish (Standard), Dutch (Standard), English (Standard), English (Cockney), English (Cannock), English (Bolton area), Faroese (Standard), German (Standard), German (Cologne), German (Brig), German (Ladelund), Icelandic (Standard), Low German (East Frisian), Low German (Northern Low Saxon), Norwegian Bokmål (Standard), Swedish (Standard), Yiddish (Standard)
Romance (20): Asturian (Standard), Catalan (Standard), Corsican (Standard), Franco-Provençal (Standard), French (Standard), Friulian (Udine), Italian (Standard), Italian (Genovese), Ladin (Standard), Ladino (Standard), Portuguese (Standard), Romanian (Standard), Romanian (Meglelo), Romansch (Puter), Romansch (Surmeiran), Romansch (Sursilvan), Romansch (Sutselvan), Romansch (Vallader), Sardinian (Standard), Spanish (Standard)
Slavic (27): Belorussian (Standard), Belorussian (Dieveniškės), Belorussian (Gervjaty), Belorussian (Lazduny), Belorussian (Zetela), Bosnian (Standard), Bulgarian (Standard), Bulgarian (Dimitrovgrad), Croatian (Standard), Czech (Standard), Czech (Moravian-Slovak), Kashubian (Standard), Macedonian (Standard), Macedonian (Kostur-Korča), Polish (Standard), Polish (Dieveniškės), Polish (Gervjaty), Polish (Lazduny), Polish (Zetela), Russian (Standard), Serbian (Standard), Slovak (Standard), Slovene (Standard), Sorbian Lower (Standard), Sorbian Lower (Vetschau), Sorbian Upper (Standard), Ukrainian (Standard)
Baltic (7): Latvian (Standard), Latvian (Skrunda), Lithuanian (Standard), Lithuanian (Dieveniškės), Lithuanian (Gervjaty), Lithuanian (Lazduny), Lithuanian (Zetela)
Celtic (11): Breton (Standard), Breton (Léonais), Breton (Trégorois), Breton (Vannetais), Cornish (Standard), Irish (Standard), Irish (Northern), Irish (Western), Irish (Southern), Scottish Gaelic (Leurbost), Welsh (Standard)
Indo-Iranian (10): Kurmanji (Standard), Ossetic (Standard), Romaní Ajia Varvara (Standard), Romani Burgenland (Standard), Romani Kalderash (Standard), Romani Lithuanian (Standard), Romani North Russian (Standard), Romani Sepecides (Standard), Zaza (Northern)
 Others (6): Albanian (Standard), Albanian (Mandrica), Albanian (Salamis), Armenian East (Standard), Armenian West (Standard), Greek (Standard)
Non-Indo-European (87)
Uralic (24): Estonian (Standard), Finnish (Standard), Hungarian (Standard), Karelian (Archangel), Karelian (Tikhvin), Karelian (Livvi), Karelian (Valdai), Komi-Permjak (Jažva), Komi-Zyrian (Standard), Komi-Zyrian (Middle Sysola), Komi-Zyrian (Pécora), Komi-Zyrian (Udora), Livonian (Standard), Mari Hill (Standard), Mari Meadow (Standard), Mordvin Erzya (Standard), Mordvin Moksha (Standard), Nenets (Western Tundra), Saami Central-South (Standard), Saami Kildin (Standard), Saami Northern (Enontekiö), Udmurt (Standard), Veps (Standard), Votic (West)
Turkic (17): Azerbaijani (Standard), Azerbaijaní (Iranian), Azerbaijani (Khalaj), Bashkir (Standard), Chuvash (Standard), Crimean Tatar (Standard), Gagauz (Standard), Karachay-Balkar (Standard), Karaim (Galits), Karaim (Eastern), Karaim (Trakai), Kazakh (Standard), Kumyk (Standard), Noghay (Standard), Tatar (Standard), Turkish (Standard), Turkish (Trabzon)
South Caucasian (5): Georgian (Standard), Laz (Standard), Laz (Mutafi), Mingrelian (Standard), Svan (Standard)
Northeast Caucasian (28): Aghul (Standard), Akhvakh (Standard), Andi (Standard), Archi (Standard), Avar (Standard), Bagvalal (Standard), Bezhta (Tlyadal), Botlikh (Standard), Budukh (Standard), Chamalal (Standard), Chechen (Standard), Dargwa (Icari), Godoberi (Standard), Hinukh (Standard), Hunzib (Standard), Ingush (Standard), Karata (Standard), Khinalug (Standard), Khvarshi (Standard), Kryts (Standard), Lak (Standard), Lezgian (Standard), Rutul (Standard), Tabasaran (Standard), Tindi (Standard), Tsakhur (Standard), Tsova-Tush (Standard), Udi (Standard)
Northwest Caucasian (6): Abaza (Standard), Abkhaz (Standard), Adyghe (Standard), Kabardian (Standard), Kabardian (Baksan), Ubykh (Standard)
Others (7): Arabic (Cypriot/Kormakiti), Maltese (Standard), West Greenlandic (Standard), Kalmyk (Standard), Basque (Standard), Basque (Lekeitio), Basque (Marquina)
6 Word prominence and areal linguistics

Harry van der Hulst, Rob Goedemans, Keren Rice

1 Introduction

The goal of this chapter is to present an overview of the consequences of language contact, with the understanding that linguistic areas arise through contact (Hickey 2010), focusing on word prominence (i.e. stress and pitch-accent). Section 2 deals with preliminary issues which are relevant to the study of contact-induced change. In section 3 we briefly present some cases of convergence involving changes in word prominence attributable to contact. Section 4 makes the point that language contact often leads to hybrid systems. Realizing that language contact can lead to both convergence (linguistic areas) and divergence, section 5 focuses on divergence of an ancestral system, partly due to language contact, into a variety of closely related systems, taking Basque as a case study. Given the rather limited availability of systematic studies of contact-induced change in word prominence, in sections 6 and 7 we present two detailed original case studies. Section 6 deals with the languages of North America where, drawing on Rice (2010, 2013), but based on a larger sample of languages, various instances of language contact are studied, focusing on areal distributions that cut across language families. Section 7, drawing on Goedemans (2010a), offers a case study concerning the aboriginal languages of Australia in which we find variation between initial stress and penultimate stress in a geographically concise contact area. Section 8 presents some conclusions and directions for further research.

2 Preliminaries

2.1 What is a linguistic area?

The definition of ‘linguistic area’ as the domain of study of ‘areal linguistics’ has been the subject of much debate. We accept the conclusion of Campbell (2006) that the notion ‘linguistic area’ refers to the epiphenomenal result of an accumulation of linguistic changes that are due to language contact. Convergence of properties in ‘a couple of’ languages that are ‘not too closely related’ points the linguist in the direction of suspecting the influence of language contact and thus offers an opportunity to study the manner in which languages acquire properties laterally, or due to external factors, as opposed to vertically, or due to inheritance, as expressed in the family tree model.2

---

1 We would like to thank Ksenia Bogomolets, Matthew Gordon, Raymond Hickey, José Hualde, Larry Hyman, Sarah Thomason and Victor Friedman for valuable comments and suggestions.

2 Vertical and lateral effects are not mutually exclusive. An internally-driven change might be enhanced by contact with a language that has already implemented this change or is more advanced in implementing it.
2.2 Word prominence

The phrase ‘word prominence’ covers a wide variety of properties that can vary independently. A number of terms are used in the literature, including (word) prominence, (word) accent, tone, stress, and pitch accent.

(1) Dimensions of word prominence (see van der Hulst 1999b, 2012, 2014b)

a. Presence/absence of word accent (or word stress3)

b. Location of accent (word edge, weight-sensitivity, boundedness)

c. Phonetic correlates of accent (e.g., enhanced duration, amplitude, pitch; taken separately or together often referred to as ‘stress’ or, if pitch is prevalent, as ‘pitch accent’)

d. Phonotactic correlates of accent (e.g., richer vowel contrast, more complex syllable structure, tonal contrast exclusively or richer on the accented syllable)

e. Presence/absence of word rhythm

f. Type of rhythm (edge prominence, direction, weight-sensitivity)

Given this ‘deconstruction’ of word prominence, it is clear that changes in this aspect of language, whether attributed to internal or external causes, can be of many different kinds. Limiting our attention to effects of language contact, we expect languages to potentially influence each other in many different ways, even with reference to such a ‘limited’ aspect as word prominence.

Word prominence is an especially worthwhile area to study, as it has often been noted in the literature that it ‘proves particularly vulnerable to systemic reshaping within language contact’ (Salmons 1994:7). Similarly, Matras (2009:231) remarks that, within phonology, ‘prosody seems to be more prone to cross-linguistic replication in contact situations than segmental phonology, with stress figuring in-between the two’.4

2.3 The distribution of stress/accent systems in the languages of the world

Various examples of linguistic areas that display convergence of word prominence have been mentioned in the literature (e.g. Jakobson 1971; Salmons 1994; Hayes 1995);

---

3 Many linguist use the terms ‘accent’ and ‘stress’ interchangeable. We here follow a tradition that regards ‘accent’ as the abstract lexical mark and ‘stress’ as a particular bundle of phonetic properties (e.g. duration, amplitude, fundamental frequency) that realize accent. This gives a stress-accent language. If accent is realized mainly in terms of F0, we speak of a pitch-accent language. Languages that have fully predictable occurrences of prominence peaks can be said to have ‘stress’ that does not realize a lexical accent mark, although we believe that such cases are rare and, if seemingly occurring, ‘stress’ in those cases may be due to ‘edge prominence’ and phrasal rhythm.

4 ‘Word prominence’ can be subsumed under the more general phrase ‘word prosody’ which covers all phonetic and phonological properties that are dependent on the ‘word domain’ such as lenition and fortition processes that mark word edges, and vowel harmony.
Areas of convergence

a. The area around the Baltic with languages from different families having initial stress (Dogil 1999a)
b. Afro-Asiatic languages like Hausa (and other Chadic languages) which got tone from neighboring Niger-Kordofanian languages (Ruhlen 1975: 62)
c. Areas in Southeast Asia sharing tonal properties (Benedict 1948)
d. Occurrence of iambic word rhythm in North American languages (Hayes 1995)

Since Salmons’ study, which focuses on (2a), more typological surveys of word prominence systems have become available. One such source is StressTyp2, a database that contains information on stress systems of over 700 languages. For reference to general characteristics regarding the distribution of word prominence see Goedemans and van der Hulst (2005a, 2005b, 2005c, 2005d) and Goedemans (2010b) which are based on a subset of the StressTyp data set. Comparing these distributions with the geographical spread of language families, two types of findings can be reported. On the one hand, word prominence properties often do cut across language families that are spoken in contiguous areas (possible evidence of convergence), while on the other hand, we also see several examples of rich diversity within language families (evidence of divergence). In particular, van der Hulst and Goedemans identify a number of general areal tendencies (corroborated by the studies in van der Hulst, Goedemans and van Zanten (2010)) in the distribution of prominence:

Areas of convergence

a. European systems that are not sensitive to weight tend to have initial stress (see van der Hulst 2010)
b. Austronesian systems tend to have penultimate stress (see section 7 and Goedemans 2010b)
c. Arabic dialects are generally sensitive to weight (see Hayes 1995; van der Hulst and Hellmuth 2010)
d. South American languages are generally insensitive to weight, with stress near the right edge of the word (see Wetzels and Meira 2010)
e. Australian languages generally have initial stress (except in the north) and are not weight sensitive (see section 7 and Goedemans 2010b)

Section 7 offers an in depth discussion of Australian patterns. For additional information we also refer to van der Hulst, Goedemans and van Zanten (2010) which contains extensive chapters of word prominence patterns in all continents of the world, with occasional explicit references to areal characteristics.

2.4 Loan phonology

Before turning to linguistic areas in particular, it is worthwhile to mention some aspects of loan phonology (Kang 2011, 2011; Uffmann 2015), as the incidental patterns that we find in borrowing form a foundation for looking at more extensive patterning that might arise through contact and the formation of what we call linguistic areas. When words are borrowed, the pattern of the source language may be respected (preservative borrowing) or not (neglectful borrowing). We identify a number of possible scenarios that have been reported in the relevant literature (Broselow 2009; Davis, Tsujimura and Tu 2012; Kang 2010, 2011; Kubozono 2006) in (4).6

(4) Borrowing scenarios

A. Preservative borrowing:
   a. Pattern of source language fits the pattern of recipient language: words are adopted with no change
   b. Pattern of source language differs from pattern of recipient language: admit foreign words into recipient language as exceptional pattern

B. Neglectful borrowing:
   a. Pattern of source language and recipient language differs: Impose the recipient system on borrowings from source language7
   b. Create a default pattern specific to loans

The list in (5) presents illustrative examples that we collected from the literature (in particular Davis, Tsujimura and Tu 2012; Kang 2010, 2011), identifying the specific realization of word prominence (involving tone, pitch-accent and stress-accent) in both the source and the recipient language:8

(5) Borrowing word prosody

A: Source (tone) / Recipient (tone)
   Preservation of tone in tone language: Hausa > Gwari
   Neglect of tone in tone language: Mandarin > Lhasa Tibetan

B: Source (tone) / Recipient (pitch-accent)
   Preservation of tone in pitch accent language: Mandarin > Yanbian Korean
   Neglect of tone in pitch accent language: Chinese > Japanese

C: Source (tone) / Recipient (stress)
   Preservation of tone in stress language: not attested
   Neglect of tone in stress language: Chinese > English common

---

6 In this discussion we ignore the important distinction between contact among adults and among child language learners.
7 Sometimes foreign words may be clipped so that their stress pattern will fit the receiving language.
8 Here we mention these examples found in these sources, without taking issues with specific cases.
D: Source (pitch-accent) / Recipient (tone)
Preservation of pitch accent in tone language: not attested
Neglect of pitch accent in tone language: Japanese > Thai
E: Source (pitch-accent) / Recipient (pitch-accent)
Preservation of pitch accent in pitch accent language: not attested
Neglect of pitch accent in pitch accent language: Japanese > Korean
F: Source (pitch-accent) / Recipient (stress)
Preservation of pitch accent in stress language: not attested
Neglect of pitch accent in stress language: Japanese > English
G: Source (stress) / Recipient (tone)
Preservation of stress in tone language: English > Cantonese
Neglect of stress in tone language: French/English > Vietnamese
H: Source (stress) / Recipient (pitch-accent)
Preservation of stress in pitch accent language: not attested
Neglect of stress in pitch accent language: French/English > Japanese
I: Source (stress) / Recipient (stress)
Preservation of stress in stress language: very common
Neglect of stress in stress language: default stress English > Finnish

Where a property is said to be preserved from one prosodic type into another type, this implies that the location of a property is preserved but not necessarily its phonetic cues. Thus, if stress is preserved in a tone system, the stress cue is reinterpreted as a tone and vice versa. The summary in (5) is, we believe, useful in providing clear illustrations of the consequences of borrowing, which, then, might also be expected to show up in linguistic areas.9

2.5 A theory of word prominence change

In this section we briefly survey some of the kinds of changes that have been identified in word prominence, whether attributable to vertical or lateral factors. As pointed out in section 2.2, the holistic notion ‘word prominence’ comprises numerous smaller properties that each, independently, can be subject to change; we will consider some of these here.

Taking ‘primary accent/stress’ and rhythm to be the main components of many prominence systems, we first mention factors that give rise to the notion of primary accent/stress. It has been suggested (see for example Hyman 1977 and Gordon 2014) that word accent/stress may result as the reanalysis of intonational peak as properties of word edges. Penultimate accent systems are common in languages of the world (Goedemans and van der Hulst 2005a). Hyman (1977) suggests that the preponderance of this accent

9 In addition to loan phonology, another area of research that is relevant to the study of language contact is second language phonology. There is a considerable literature on how speakers perceive and produce word prominence patterns of languages that they are learning: see, for instance, Peperkamp and Dupoux (2002); Altman (2006); Kijak (2009); Altmann and Kabak (2011), and for a the relation between second language acquisition and phonological change, Eckman and Iverson (2015). The three areas of loan phonology, second language phonology and language contact share a number of interesting common results, but space limitations prevent us from discussing these here.
location may be due to the occurrence of an intonation pitch accent $H^+L$ on the right edge of the rightmost word in an intonational phrase. To avoid tonal crowding on the final syllable, the $H^L$ melody would spread out over the final two syllables, giving the penultimate syllable the appearance of prominence due to bearing the $H$-tone. This then may lead to the interpretation of this syllable as being accented at the word level. Gordon (2014) pays special attention to the development of word prominence from phrasal intonation and suggests that, even synchronically, some cases of alleged word accent may be more properly analyzed in terms of intonational structure.\(^{10}\)

Another source for the emergence of an accentual system is tone (see Ratliff 2015 about tonal change in general). Salmons (1994) in particular discusses the change from tonal to accentual systems, resulting from the fact that the distribution of specifically high tones may develop constraints (up to the maximal occurrence of one $H$ tone per word) which, then, trigger a reinterpretation of tone and pitch-accent and eventually stress-accent.

Once accent is present, changes in the factors that determine accent location may be related to, or caused by, changes in another component of the phonology or the grammar at large. In classical Latin, accent placement was sensitive to vowel length (among other factors). Contrastive vowel length was lost in Romance languages such as Italian and Spanish, but vowels that were long kept the accent. Accent placement thus became partially lexicalized, with specification of these unpredictable accents in the lexicon (see Roca 1999).

Another important factor that can change the determination of accent location is emergence of syllable weight, when intrinsic properties of syllables (in particular syllable rhymes) phonologize by becoming determinants in stress location. Penultimate heavy syllables (containing a long vowel or coda consonant) can influence the relocation of final accent (which is as such already somewhat disfavored) to penultimate accent.

Another kind of accentual change involves an edge switch for accent, as can be seen by reference to two Slavic languages, Polish and Czech. Polish has penultimate accent, while Czech has initial accent. Many languages with penultimate accent have an initial edge prominence (‘secondary stress’) on the first syllable, which in specific contexts might be the locus of an intonational pitch accent (see van der Hulst 2014b). This may motivate the learner to regard the initial rather than the penultimate syllable as accented. Such a change can then be said to be caused by the intonational system.

Regarding rhythm, we suggest that this aspect of word prominence may arise as the result of grammaticalizing low-level phonetic properties. For instance, a language can acquire ‘notable’ rhythm when low-level rhythm (which we take to be a universal property of all languages) is exaggerated and causes allophonic changes in phonemes. It is even possible for rhythm to fully lexicalize (and lose some of its ‘natural’ phonetic properties) becoming ‘abstract’, still guiding for example allomorphic choice and phonotactic asymmetries, while meanwhile the language has been subjected to a new kind of rhythm. Such mismatches, which reflect a form of hybridity and which

\(^{10}\) While the languages discussed above involve reanalysis where the changes appear to arise internal to the system, reanalysis may also be triggered from outside the system. Reanalysis may also lie behind the change from Germanic initial accent to right-edge accent, due to the presence of Latin loans (Lahiri, Riad and Jacobs 1999).
underscore the need for as ‘deconstructed’ analysis (see section 4), are discussed in Gordon (2014, to appear).

Changes in rhythm type may arise due to the influence of the same intrinsic properties of syllables that can affect the location or primary accent/stress.

Given the deconstruction of word prominence into several factors (see 1), several other patterns of change would have to considered in a fuller account. Many of these changes can be seen as motivated by grammar-internal factors or in terms of acquisition strategies. This leaves a small remnant for changes that could be said to rely crucially on language contact, but, of course, even internally motivated changes can be triggered or enhanced by external influences due to languages in which such factors play a role. Illustrating the potentially complex interaction of internal and external factors, Salmons (1992) offers extensive discussion of specific patterns of change that involve the transition from tonal systems to pitch accent systems to stress systems, which he sees as potentially resulting from internal factors and reanalysis, although language contact may contribute to such internally-motivated changes. Moravcsik (1978) also remarks that changes in word prominence, whether potentially internally driven or not, are often probably due to language contact, precisely, as we have already mentioned earlier, because they are the first to be affected by such contact.

3 Convergence

Salmons (1992) focusses on language contact in Early Northern Europe, especially between Germanic and Celtic, but including Italic, Finnish, Sámi and other Finno-Ugric languages which together, he argues, form an areal fixed initial stress group. He pays particular attention to the joint Germanic-Celtic shift to initial accent (shared with Italic). It has been debated whether this shift is attributable to language contact with an unknown substrate language in Northern Europe or to Finno-Ugric. Salmons opts for the latter. Salmons’ conclusion that the presence of initial stress in this area is due to language contact receives support, as he argues, when one considers the distribution of stress patterns in the languages of the world. Based on Ruhlen (1975, 1977), Salmons finds that initial accent is absent in 16 of Ruhlen’s 28 genetic groupings (including Austro-Asiatic, Nilo-Saharan, Sino-Tibetan and Paleosiberian). In many other families the average number of languages with initial stress is under 7%. The bulk of initial stress systems come from Australian (section 7) and Uralic (with 16 out of 23 languages listed in Ruhlen having initial stress). Interestingly, the 6 Indo-European languages in his survey (Latvian, Czech, Irish, Swedish, Danish, Dutch) all belong to the linguistic area that forms the subject of Salmons’ survey. Final stress also has a biased genetic distribution, with “63.9%” of the languages with this stress system coming from three

11 Another area of change concerns the distinction between bounded and unbounded systems which we omit here for reasons of space.
13 Only Northern and Western Irish have initial stress. Southern Irish has a complicated system determined by vowel length and syllable weight which attracts stress to non-initial syllables. This may have been due to contact with Anglo-Norman after the late twelfth century or to internal tensions due to the rise of long vowels in non-initial syllables in Middle Irish (or due to a combination of both). See Hickey (1997, 2014).
families (Penutian, Altaic, Indo-European) based on Ruhlen’s counts. The Indo-European languages in this group are French, Armenian, and Iranian languages, and the latter form part of a linguistic area with Altaic languages. Penultimate stress appears to have a much more equal distribution in language families (which correlates with its preponderance as a fixed accent type; see Hyman 1977; Goedemans and van der Hulst 2005a).

Results such as these show that accentual types can run in families (as might be expected). This makes resemblances across families more likely to be the result of language contact than of independent development toward certain common types. Thus, the widespread occurrence of initial stress in the Northern European initial stress area proposed by Salmons is therefore plausibly the result of language contact.14

Salmons includes a survey of examples of accentual change due to language contact; to these we add further examples in the list below.15 We suspect that a broader survey of the literature might reveal many more of these examples.

(6) Examples of contact induced changes in word prominence

a. Scandinavian tonal systems lost in the dialects spoken in Finland, a non-tonal language (Haugen 1970)
b. Within Bantu, languages range from fully tonal to restricted tone or pitch-accent systems, to stress systems (as in Swahili); Salmons (1992).
c. Loss of tone in African pidgins (Fanagalo, Town-Bemba, pidgin Hausa); Heine (1973)
d. Presence of fixed stress in African pidgins (Kituba); Heine (1973)
e. Loss of tone in some Caribbean pidgins and creoles (Alleyne 1980)16
f. Loss of tone in Danish in contact with low German (Gårding 1977)
g. Loss of tone in Swedish spoken in the United States (Ureland 1971)
h. Russian dialects bordering on Karelian-Olonec- Vesperian-speaking areas show a tendency toward initial stress (Veenker 1967)
i. Hungarian may have influenced the emergence of initial accent in Czech and Slovak (Veenker 1967; Thomason and Kaufman 1988)
j. Boanon, a Mongolian language, acquiring tone under the influence of Linxia, a Chinese language (Li 1986)
k. transfer of a stress system in Iroquoian languages Cayuga to Onondaga and Seneca (Michelson 1988)
l. influence of penultimate stress in a dialect of French spoken in Brittany as a possible influence on penultimate stress in Breton (Sommerfelt 1962)
m. final stress in Turkic languages, Armenian, and Iranian languages (Hyman 1977)
n. English influence on Stoney Dakota (Siouan) in having right edge stress patterns similar to English in addition to the expected second syllable Dakota stress pattern (Shaw 1985)

14 Peters (1994) in his review of Salmons (1992) criticized Salmons’ use of percentages to make this point.
15 We thank Sally Thomason for bringing some of these examples to our attention. See Thomason and Kaufman (1988) for other examples.
16 Island Caribbean. Sranan and Saramaccan in Suriname are exceptions.
o. Germanic initial primary accent added accent determined from the right edge, as in Romance with French loan-words (van der Hulst, Hendriks and van de Weijer 1999; 2010)
q. Shift in Hungarian dialects near the former Yugoslavia border from initial stress to fixed penultimate stress (Thomason 2001: 143) (note that Serbo-Croatian varieties generally have free stress, but in this dialect, there is also fixed penultimate stress)
r. Latvian complex accent system replaced by first syllable stress under influence of Livonian (Finnic) (Thomason 2001)

Clearly, a list like this does not allow for any general conclusions or hypotheses. What is missing in most cases is detailed analyses of the circumstances and mechanisms involved. Thus, while there is no shortage of reported cases of accentual change due to language contact, there is therefore a clear need for more detailed studies of specific cases and we will offer two such studies in section 6 and 7.

4 Hybridity

Although, as argued in section 2.5 in general, it is difficult to determine whether change in a word accent system is motivated by contact or not, we take hybridity to be an indication of contact-induced change. By hybridity we mean that a system clearly incorporates aspects of systems that are, under usual assumptions, compatible with one another (i.e. can be shown to follow from a single uniform analysis). Languages can be hybrid in their prominence system in many ways; several of these types of hybridity will be illustrated in our case studies in section 6 and 7:

(7) Types of hybridity

A. Competing patterns
   a. Due to loans, a regular pattern may have a limited number of exceptions
   b. Exceptions may increase in number and, if originating from a single language, form a second pattern, that may be restricted to an independently identifiable lexical stratum
   c. The new (default) pattern may start being applied to non-loans, but not all
   d. The change may lead to variability

B. Deviant patterns
   a. Due to contact a pattern may change one of its parameters, leading to a theoretically possible, yet more marked pattern
   b. Due to contact, words of different lengths may have different patterns

C. Incomplete patterns
a. Due to contact with other patterns, a pattern may display incompleteness in that certain parameters are not set, e.g. absence of primary stress.

An example of a stratal effect (8Ab) is found in European languages belonging to the Germanic group which, as a whole, have Romance-influenced right-edge stress, while also maintaining traits of the Germanic initial stress. The Romance influence on Germanic languages (partly via English) has given rise to a ‘pan-European right-edge stress pattern, from which only Icelandic has escaped’ (van der Hulst, Hendriks and van de Weijer 1999; Lahiri 2015).

While hybridity can be a sign of language contact, we must realize that hybridity can also result from the synchronic presence of different stages of historical developments. As mentioned in section 2.5, when a language changes its rhythmic bias, a synchronic accent of certain phonological regularities may require the postulation of the older pattern at a deeper level of analysis, while the new pattern is accounted for at a later stratum (see Gordon 2014). In a general sense, hybridity results from the presence of two (or more) systems, whether originating from different stages of a single language or from different languages that are in contact.17

5 Divergence: The case of Basque

In section 2.1 we noted that contact induced change can result in cases in which a set of closely related languages displays a wide variety of word prominence systems, although it is likely that such fractionation can also result without much external influence. There are several cases with a sometimes strikingly rich proliferation of word prominence types within one language family:

(8) Examples of divergence

a. Basque dialects (see below)


d. Arabic dialects (Hayes 1995; van der Hulst and Hellmuth 2010)

e. Scandinavian dialects (Gårding 1977; Riad 1998; Bye 2004).

Modern-day Slavic languages exhibit quite a bit of diversity (see Dogil 1999b; Salmons 1994: 49). To this, we add the languages spoken in Meso-America, which show a rich variety of tonal accentual systems, although in these cases several language families have been claimed to be involved (see van der Hulst, Rice and Wetzels 2010). Several more examples of divergence can be found in van der Hulst, Goedemans and van Zanten (2010), which offers an overview of the distribution of word prominence systems in all language families of the world.

In all these cases, a group of languages belonging to a single (macro-) family display an enormous amount of diversity, while preserving certain common properties. We suggest that it would be very interesting to look into the internal and external causes

17 Dresher and Lahiri (1991) speak of metrical incoherence to characterize such cases.
that result in the divergence processes that lead to such diversity. As an example of such an approach we discuss the case of Basque.

Hualde (1999) presents an overview of the different contemporary accentual systems found in Basque dialects.\(^{18}\) The Basque dialects present a great diversity of word-prosodic systems, especially when one takes into account the size of the area in which Basque is spoken (only about 135 x 35 km.). This area shows many patterns of convergence and divergence; see Hualde (2007). The Basque word-prosodic systems range from lexical pitch-accent and stress-accent systems in the Western dialects to weight-insensitive accent on the second syllable in some Central Basque dialects and weight-insensitive accent on the penultimate syllable in, for instance, the High Navarrese variety of Baztan. Thus, there is not only a distinction between pitch-accent and stress-accent systems and between weight-sensitive and weight-insensitive systems, but also, within the weight-insensitive accent systems, accent can be assigned from either the right or the left edge.\(^{19}\) Hualde distinguishes the following main accentual types:

(9) Basque word prominence systems

a. the WESTERN type, which is a lexical system with unaccented and accented stems and affixes, and prominence in accented words realized as pitch drop or stress, depending on region (types: Markina, Gernika-Gexto, , Antzuola)

b. the CENTRAL type in which accent is assigned from the left word-edge
c. the HONDARRIBIA/OLD LABOURDIAN type, in which accent seems to be assigned from the right word edge, with variety between dialects
d. the SOULETIN type (with unmarked final and marked penultimate stress) as a special type
e. the WESTERN NAVARRESE type represented by Goizueta, with both lexical stress and lexical pitch accent

In this section we focus on some changes that have occurred in Western varieties of Basque which Hualde (2003) attributes to contact with Spanish. This case reveals a number of subtle and perhaps unexpected responses to confrontation with a prominence system that is different from the influenced dialects.

As a point of departure, Hualde takes the Western Basque dialects (spoken in the provinces of Bizkaia and Gipuzkoa and neighboring areas), and, in particular, the Northern or Costal Bizkaian pitch accent type. A hypothesis for how the Western Navarrese (Goizueta) type may be historically related to Northern Bizkaian is developed in Hualde (2012). Western Basque is a system with accented and unaccented morphemes, the majority being unaccented. A pitch pattern is associated with phrases. In particular, Hualde proposes initial %=LH and final H*L accentual patterns, the former distributed over two syllables and the latter on the final syllable. A phrase final unaccented word in the Gernika-Gexto type receives a derived accent on the final syllable (the Markina type has a penultimate default; see 13). Here we illustrate the Gernika-Gexto type:

\(^{18}\) We thank José Hualde for additional information and helpful comments on this section of our chapter.

\(^{19}\) The scope and nature of this enormous variety in such a small territory is reminiscent of the situation in the Caucasus (see Kodsasov 1999).
The plural has a morphologically determined accent, in particular the plural suffix is pre-accenting; see Hualde (1999) for paradigms showing this. On the surface, all lexical accents are on a non-final syllable because the accented suffixes are actually pre-accenting. For lexical roots there are some tendencies for their location depending on the local dialect. For instance in 3-syllable accented roots, the accent is generally on the first syllable in some dialects (e.g. bélarri ‘ear’, tómate ‘tomato’, pátata ‘potato’), but on the second in other towns (belárrri, tomáte, patáta). We refer to Hualde (2012) for details.

Hualde (2003) shows how this Western type has developed into different kinds of systems under the possible influence of Spanish. All Western Basque speakers are Basque/Spanish bilinguals.

The Western pattern has given rise to two responses:

(12) Reanalysis

Response A: In Bilbao Basque (as well as in varieties of the Central type) the initial LH in (11) has been reanalyzed as the phonetic realization of a pitch accent on the second syllable, attributed to the fact that in Spanish a comparable pitch rise is an important cue to its stress-accent.

(13) Pitch rise => accent

Response B: In the Antzuola dialect, the initial H is suppressed, and there is a high pitch accent on the penultimate syllable followed by a low on the final syllable.

(14) Accent => pitch rise

Markina type             Antzuola type

<table>
<thead>
<tr>
<th>%L  H-     H*L</th>
<th>%L         H*L</th>
</tr>
</thead>
<tbody>
<tr>
<td>[la gu nen] [a ma ri]</td>
<td>[la gu nan tza kú]</td>
</tr>
<tr>
<td>‘to the friend-SING’s mother’</td>
<td>‘the one for the friend’</td>
</tr>
</tbody>
</table>

This change makes the Antzuola type sound much more like Spanish.

Both responses reflect an influence of the Spanish stress cue, a rise to high pitch on the stress-accented syllable. In response A it is the initial rise that is identified with the
accent (because rises identify accent location in Spanish), whereas in response B the accent location is identified with a rise (again because rises identify accent location in Spanish). Hualde suggests that the change in the Antzuola system with penultimate accent led to changes in Azkoitia and other cases, giving rise to third syllable stress accent [+3] (with non-finality, i.e. avoidance of final accent), a reinterpretation that can easily result from the ambiguity that is caused by words of up to 4 syllables where a penultimate location could be reinterpreted as a third syllable location. The difference would only be clear in words with 5 syllables or more.


Summarizing (broken lines indicate influence, solid lines indicate change):20

\[
\begin{align*}
\text{Gernika-Gexto} & \rightarrow \text{Spanish} \\
\text{Bilbao} & \rightarrow \text{Markina} \\
\text{Azkoitia} & \rightarrow \text{Getaria} \\
\end{align*}
\]

This case study shows how language contact, including contact between dialects, can lead to surprising consequences. In particular, Spanish contact triggered changes that in one case (response A) caused an analysis of a non-accentual pitch rise as accentual and in another case (response B) a change in the pitch realization of the accent (and the entire word contour). Both responses are based on identification of accents with pitch rises.21

6 North American linguistic areas: evidence for convergence?

Several linguistic areas have been identified in North America, but accent has not been considered as a factor that defines these areas. In this section we first present three case studies of accent patterns in areas that are argued to be linguistic areas, asking what these areas reveal about contact-induced change. We then look more broadly at the continent as a whole, asking whether there is support for the observation that North America might be a linguistic area.

6.1 Three areas of convergence

In this section we offer a study of three linguistic areas of California for which evidence of convergence has been given (see discussions in Campbell 1997 and Mithun 1999; see Haynie 2012 for the most recent discussion of two of these areas). The grounds for speaking of convergence for these linguistic areas are quite different. In two cases, there

\[+2\] means second syllable accent, etc. \[-2\] means penultimate accent.

21 See Hualde (2000) and Hualde et al. (2002) for additional details.
is evidence for the borrowing of words, while in the third, there is no such evidence, and the linguistic factors that define this third area are quite different. We will ask what these areas tell us about word prominence and areal phenomena more broadly rather than about contact-induced change specifically.

It is important to keep in mind that we do not have detailed descriptions of accentual systems for many of the languages addressed in this section. We can, generally, refer to the placement of ‘primary stress’, but often not to ‘secondary stress’. Also, there is no detailed discussion of the realization of accent available for many of the languages. As Golla (2011: 209) points out “Accentual uses of pitch and tonal phenomena occur widely in the languages of the northern part of the California region, but in most instances the documentation is poor.”

6.1.1 Clear Lake Linguistic Area (California)

One well-defined linguistic area in California is Clear Lake, located approximately 100 miles north of San Francisco. This area includes the following languages; the language family is given in parentheses: Lake Miwok (Utian), Patwin (Wintuan), Eastern Pomo (Pomoan), Southeastern Pomo (Pomoan), and Wappo (isolate, or Yukian, see Campbell 1997: 132, Mithun 1999: 554; Golla 2011: 192-193 for discussions and different perspectives). This list of languages is taken from Campbell (1997: 336); Mithun (1999: 317) includes examples from a second Wintuan language, Nomlaki.

Campbell and Mithun identify a number of phonological characteristics shared by languages of the Clear Lake area. Their particular focus is on how Lake Miwok, part of the Clear Lake area, differs from genealogically related languages. For instance, Lake Miwok, like its neighbors, has contrasting phonation types in stops and affricates, while other Miwok languages do not. Lake Miwok and some of the other languages in the area have a voiceless lateral and a lateral affricate; other Miwok languages do not; Lake Miwok and other Clear Lake languages all have five short and five long vowels, while other Miwok languages have six short and six long vowels. The source of innovation in Lake Miwok, discussed by Callaghan (1964), is borrowing: Lake Miwok borrowed words from Hill Patwin and from Southeastern Pomo, with the sounds not found in other Miwok languages introduced through those borrowings. Campbell (1997: 129) notes that Lake Miwok is geographically isolated from the other Miwok languages, and speakers had frequent contact with speakers of Eastern Pomo, Southeastern Pomo, Foothills Patwin, and Wappo, as reflected in loanwords. Golla (2011: 107) reports on the relationship between Patwin and Southeastern Pomo, noting that Patwin influence was strong in the Southeastern Pomo area, with Patwin-Pomo bilingualism common.

We now examine word accent in this area, asking if it too shows areal characteristics. In order to discuss word accent in most of these languages, it is necessary to say something about morphology, so information about this is included in the discussion below.

Word stress in Wappo appears on the first syllable of the stem. The language has some prefixes as well as derivational and aspectual suffixes.

We have not found sources that discuss stress in the Wintuan language Patwin. In Wintu (Pitkin 1984), stress is described as falling on one of the first two syllables of the stem – the first heavy syllable, with the first syllable if neither is heavy. Golla (2011:146)
notes that both Wintu and Patwin have prefixes indicating location and direction; both languages also have suffixes.

The position of word accent in the Pomoan languages is addressed in Moshinsky (1976). Moshinsky (1976: 56) proposes that stress was predictable in Proto-Pomoan, falling on the first stem vowel. The stem was frequently preceded by a monosyllabic prefix, so stress was often on the second syllable. Languages in the family developed in different ways. In Eastern, Central, and Northern Pomo, this pattern was usually maintained, with stress generally on the second syllable, although there is not always clear evidence for a synchronic morphological analysis. McLendon (1975:12) notes that in Eastern Pomo, stress falls on the root, with second syllable stress being the predominate pattern. Northeastern Pomo often moved stress on to a previously unstressed syllable.

Kashaya, another Pomoan language, has a complex system. Basically, primary stress can appear on one of the first three syllables of the word. Buckley (2013) treats the language as quantity sensitive and iambic, with the first syllable invisible for purposes of stress. Moshinsky (1976) notes that Southern Pomo has the most aberrant stress pattern, with stress on the penultimate syllable of a phrase; see Walker (2008), who also finds penultimate stress on words and phrases and secondary stress on the first syllable. Southeastern Pomo is regarded by Moshinky as having regular stress (1976:57), with word stress generally on the first syllable of nouns, verbs, and adjectives. While Moshinsky notes that there are a few exceptions where it does not fall on the first vowel – it is on the second vowel if there is a directional prefix or if the first vowel is a part of reduplication, Buckley (2013) argues that primary stress falls on the first syllable, while directional prefixes and the first vowel of the stem may have secondary stress.

While the Pomoan languages in question differ in detail, there is in general morphological sensitivity, with stress attracted to the root in most of the languages. The position of primary stress is determined from the left-edge, and is near the left-edge of the word in languages of this family. Even in Northern Pomo, there is secondary stress on the first syllable.

The Miwok languages are largely suffixing, with pronominal prefixes that fall outside the domain of stress. Word accent appears on one of the first two syllables of the root, being attracted to heavy syllables, with heaviness defined slightly differently depending on the language. Lake Miwok has word stress on the first or second syllable of the root; prefixes do not take stress (Callaghan 1963). In particular, primary stress falls on the second root syllable if it is heavy, and otherwise on the first syllable.

The languages of the Clear Lake area (Lake Miwok, Patwin, Eastern Pomo, Southeastern Pomo, Wappo) share several characteristics with respect to accent. Most noticeable is that primary stress occurs near the left edge of either the word or the root/stem. Given that these languages are largely suffixing (as are Miwok, Wintuan, and Pomoan languages that are not in the area), stress is quite likely to fall on one of the first two syllables. In general, all the languages under consideration have the root as the domain of stress (Southeastern Pomo is the exception). Some of the languages show quantity sensitivity (Lake Miwok, other Miwok, Kashaya), some do not (Eastern Pomo, Southeastern Pomo, Wappo).

The left-edge orientation of primary stress is clearly a characteristic that defines this linguistic area, and thus can be considered an areal phenomenon. Whether this is a result of contact is not certain, since the major languages that border the Clear Lake area are genealogically related to languages of the area and generally share the same
orientation of stress. It is possible that the languages of the area retained similar accent patterns because they form an area. Kashaya, geographically the most remote member, has the most divergent pattern in the Pomoan family, although it maintains much of the basic pattern of the related languages with the position of accent determined from the left edge. The homogeneity that is found in word accent could result from convergence, or from inheritance, with the areal effects inhibiting change.

6.1.2 Northwestern California

Another part of California that has been identified as a linguistic area is Northwestern California. Haynie (2012), based on Conathan (2004) and others, notes good evidence for areal feature spread in Northern California, and particularly in the Northwestern California area.

The Northwestern California area differs from the Clear Lake area in several ways. While in the Clear Lake area, there is evidence for lexical borrowing between languages, this is not so in the Northwestern California area. Conathan (2004) argues that the Northwestern California area is characterized by functional convergence, but not by actual borrowing or calquing.

Conathan defines the Northwestern California area as consisting of the following languages: Tolowa (Athabaskan), Hupa (Athabaskan), Karuk (isolate), Chimariko (isolate), Yurok (Algic), and Wiyot (Algic). Karuk and Chimariko are sometimes grouped together in the Hokan stock. Conathan notes that the language group is somewhat controversial, and that she could have included Shasta (Hokan) and Wintu (Wintuan) as well.

Haynie (2012) examines several phonological and morphological characteristics that are claimed to mark an even broader area, Northern California, arguing, based on a spatial autocorrelation technique that examines features likely to have diffused geographically rather than genealogically or by chance, to determine whether feature diffusion is a likely scenario in this area (Haynie 2012:88). Haynie concludes that the Northwestern California area, discussed in this section, and the Clear Lake area, discussed in the previous section, show evidence for feature diffusion, while Northern California as a whole is “more like a collection of smaller diffusion zones” (Haynie 2012:89).

The phonological characteristics of Northwestern California that Haynie discusses are summarized below, with some additional characteristics discussed in Jany (2009). Several of the languages have plain, ejective, and aspirated stops (Tolowa, Hupa, Chimariko); Yurok has plain and ejective stops; Wiyot plain and aspirated stops; and Karuk has plain stops. Hupa and Chimariko have a back velar; the others do not. Tolowa and Wiyot have a lateral fricative; Yurok has both a plain and ejective lateral fricative; Hupa has a lateral fricative and ejective affricate; Karuk and Yurok have neither. Hupa has a velar nasal; none of the other languages do. Tolowa, Hupa, Yurok, and Wiyot have labialized consonants, while the other languages do not (Jany 2009); Tolowa has retroflex consonants while the others do not (Jany 2009). The languages also differ in their vowel inventories (ranging from three to six vowels; some have length contrasts, some have nasalized vowels). There are clearly considerable phonological differences between the languages in this group, as Conathan (2004) and Haynie (2012) conclude.
Turning to word accent, there are also differences between the languages. In terms of domain, Tolowa, Karuk, Yurok, and probably Wiyot take the word as the domain, Hupa has the word as the domain of accent with the morpheme also playing a role, while Chimariko takes the root as the domain of accent. Wiyot determines the location of accent from the left edge, as does Hupa as recorded by Sapir; later Hupa has accent on the first long vowel, and on the root if there is no prefix long vowel (see Gordon and Luna 2004 on Hupa). Karuk is similar, with accent on the first long vowel, with default to the last vowel. In Chimariko, accent is generally on the penultimate syllable of the root. The various languages are quantity sensitive. In terms of morphology, Yurok and Wiyot are largely suffixing, with some prefixes; the others are largely prefixing; Karuk has both prefixes and suffixes.

Languages in this group may have lexical accent (Tolowa), or the position of accent may be predictable. Accent may take the word or the root as its domain and it might respond to weight (Yurok, Shasta). It may appear near the left edge (Hupa, Wiyot, Karuk), but stress can also be near the end of the word (Hupa default in the later recordings; Karuk default).

Jany (2009:32) comments extensively on similarities in stress systems in these, and a few other languages:

Stress systems are often described in detail in the grammars consulted. However, the phonetic correlates of stress are not always mentioned. In general, stress patterns show many similarities in Northern California. Immediate neighbors of Chimariko, Hupa, Shasta, and Wintu, all show weight-sensitive stress systems. While their weight hierarchies are slightly different, all have CVV as their heaviest syllable. Root stress, as well as penultimate stress and leftward attraction of stress, are also very common in the area. Shasta, for example, has penultimate stress, but moves the stress in longer sequences to the first preceding heavy syllable. Acoustic correlates of stress include pitch and intensity for Hupa. For Shasta, a high-low pitch tonal accent has been described. Hence the acoustic correlate of stress in Chimariko, which is pitch, is also attested in other languages of the area. Given that stress is easily transferred through language contact, it is likely that the languages in Northern California have shifted their stress patterns as a result of multilingualism in the area. For Chimariko it can be speculated that vowel length on stressed syllables was developing as a contact phenomenon given the weight-sensitive stress systems of neighboring languages with CVV as the heaviest syllable type.

Conathan (2004) takes a different perspective in her discussion of the area. She notes that this is an area of intense cultural interaction, but with little lexical borrowing. She says specifically that “Local convergence of phonological features is conspicuous in its absence” (2004: 167). Conathan remarks that the three languages Karuk, Hupa, and Yurok, all genealogically distinct, are the ‘core’ members of this area, and their inventories are quite different; she attributes this to lack of lexical borrowing between these languages (Conathan 2004: 169). Golla (2011) remarks that, given the close social and ceremonial ties, intergroup marriage, and a moderate degree of multilingualism, “it is surprising how few of the distinctive phonological, lexical, or grammatical features of Hupa-Chilula can be attributed to direct Yurok influence”, further noting that lexical borrowings are almost nonexistent. What Conathan finds most noteworthy about the
Northwestern California linguistic area is the existence of grammatical borrowing without lexical borrowing. Thus, she finds evidence of contact effects in certain components of the grammar – tense and aspect marking, classifier systems, second person prominence in argument marking, loan translations, word order – but she does not find loan words and phonological convergence. Mithun (2010) too notes a number of structural parallels in languages of this area (and in North America more broadly).

Conathan (2004: 175-179) identifies a variety of reasons for why there might be an absence of lexical borrowing, including the absence of a dominant language in terms of population size, the overall rarity of bilingualism, an overall egalitarian society with multidirectional bilingualism combined with the absence of language shift or a lingua franca, and efforts to avoid mixing languages. With respect to multilingualism, she points out that it arose largely from interlingual marriages.

These languages, unlike those of the Clear Lake area, do not show many similarities in their word accent systems. Perhaps the commonalities that Jany (2009) identifies – weight sensitivity, realization as pitch – are the phonological parallels to the kinds of contact effects that Conathan identifies in the Northwestern California languages. Within the phonological domain, a study of intonational patterns would be worthwhile, as this might be an area where convergence would be more likely. Matras (2009: 231-233) specifically notes the susceptibility of prosody, referring largely to word-level prosody, to shift under contact conditions. He suggests that prosody is peripheral in conveying meaning, being prototypically a form of expression of emotive modes, and thus operating at the level of the speech act rather than the word level. He further notes that “This allows speakers to mentally disconnect prosody more easily from the matter or shape of words”, making it prone to change.

While the Clear Lake area likely involves both areal and genetic changes, this is less clear in the Northwestern California area, where even between genetically related languages there are several differences, there is no clear evidence of areal convergence, and absence of clear genetic relationships between other languages makes it difficult to ascertain if there are divergences.

6.1.3 Yuman-Takic

Hinton (1991) presents a detailed study of a group of languages in southern California, making the case that Yuman languages and a class of Uto-Aztecan languages known as Cupan formed a linguistic area at some point. Hinton argues that Yuman languages had a major influence on the Cupan language group (1991: 148), with the Cupan languages (Luiseño, Cupéño, Cahuilla) evolving to resemble the Yuman languages phonologically under the influence of a Yuman substratum that was replaced by the Cupan languages. Hinton shows that there are several characteristics that are shared by Yuman and Cupan languages, but are not reconstructed for Proto-Uto-Aztecan and are generally not present in the neighbouring Esselen, Chumashan, and Salinan languages. The traits that Hinton examines are segmental, with a focus on unusual segments that are reconstructed for Proto-Yuman but not for Proto-Uto-Aztecan, and are found in Cupan languages but not in other Uto-Aztecan languages of the area. These include a kw/qw distinction, phonemic in Proto-Yuman and in Luiseño, and allophonic in Cahuilla and Cupéño. Other characteristics found in Proto-Yuman, not in Proto-Uto-Aztecan, and in Cupan languages but not in other nearby Uto-Aztecan languages include a distinction between s and
retroflex s (also in a nearby Uto-Aztecan language, Serrano), x̠, ñ (not in Luiseño), ly (not in Luiseño), r/l. Hinton further notes that Cahuilla and Cupeño share a small vowel inventory with Proto-Yuman (three vowels with allophonic variation), while Proto-Uto-Aztecan and most Uto-Aztecan have a larger vowel inventory (five vowels).

What about word accent? Yuman languages have accent near the right edge of the word. There is variation across Uto-Aztecan in terms of stress placement, but Cahuilla, Cupeño, and Luiseño, while they differ in their patterns, all have primary stress near the left edge of the word. Of the other Takic languages, a larger group in which the Cupan languages fall, Tübatulabal has no main stress, with the placement of stress determined from the right edge of the word, and Chemehuevi has stress on the second syllable. We have not found information on Serrano and Gabrielino, two other Takic languages. While there are clearly segmental influences of Yuman languages on the Cupan languages, as detailed by Hinton (1991), it does not appear that the accent system was subject to contact effects in terms of merging. Thus, this area suggests that, while stress is viewed as subject to borrowing (see Matras 2009), borrowing of segmental features may occur without borrowing of the stress system.

6.1.4 Summary: linguistic areas of California

We have surveyed three areas of California that have been identified as linguistic areas. The three different areas investigated lead us to different conclusions. In one, Clear Lake, word accent is a factor that contributes to the linguistic area, although its origin (genealogy, contact, both) is not clear. In another, Northwestern California, accent resembles segmental properties in not showing obvious areal effects. In a third case, Yuman-Takic, several segmental features suggest an area, but accent does not appear to be involved in defining this area.

6.2 A survey of accent patterns in North America

Hayes (1995) remarks on the distribution of iambicity in North America, noting that iambic patterns are more common in North America than elsewhere in the world. Mithun (2010) looks at a number of parallels in morphosyntactic patterns across North America, arguing that “The Americas provide rich examples of language contact” in terms of structural parallelisms “even in the absence of borrowed words and morphemes” (Mithun 2010: 691). In this section we examine some of the similarities in accent patterns in languages of North America in order to study whether there are any broad generalizations that might be drawn.

First we report on findings related to a study of word prominence systems in languages around the world, drawing on Goedemans’ (2010b) study of cross-linguistic patterns of stress, which are based on the StressTyp database.

We begin with fixed vs. variable (weight-sensitive) stress. Due to limitations of space, we do not provide details, but simply summarize findings.

Based on language family (a total of 40 families; within a family, languages may fall in both categories), 24 families have languages that are quantity insensitive (fixed stress), and 26 of the families have languages that are quantity sensitive (variable) with respect to primary stress. Goedemans (2010b: 651) reports that in the StressTyp survey,
278 languages (55.5%) showed fixed, or quantity insensitive, stress, while 222, or 44.5%, have variable, or quantity sensitive, stress. The difference between the numbers of North American families showing fixed and variable stress is probably not significant (recall that the count is based on families, not on languages).

The specific location of primary stress in quantity sensitive languages is also of interest. In the 222 quantity sensitive languages examined by Goedemans, primary stress is located on one of the leftmost two syllables in around 15% and on one of the rightmost two syllables in around 30% of the languages (2010b: 655); the next most common pattern is stress located on any syllable (unbounded), followed by stress on the right edge, but not restricted to the rightmost two syllables (~13%) and then by stress location being unpredictable (e.g., lexical, irregular, no primary stress). This suggests that the most common pattern is for variable stress to be determined from the right (around 43% of the languages in the survey).

The numbers for North America are by family unless there are differences between languages within the family, in which case the family is counted in both categories left and right. Languages that are considered to be count systems within StressTyp – foot assignment starts at one edge, but primary stress occurs on the other edge – are considered in terms of the position of primary stress, not in terms of the edge from which it is determined. Given this, 11 of the families in North America have languages with primary stress near the left edge, while six have it near the right edge. Since we are not counting individual languages, but rather patterns within a language family, it is not clear whether the numbers are comparable with those in Goedemans (2010b). If they are, languages of North America exhibit a different bias in terms of the placement of primary stress than Goedemans found, with a preference for left-edge primary stress in North America but right edge generally.

We next consider the edge at which primary accent occurs in the quantity insensitive languages, conflating word edge and root/stem edge. In the sample, 12 families have languages with stress on the final or penultimate syllable, while 11 have left-edge oriented primary stress. For quantity insensitive languages, Goedemans (2010b) includes information about whether stress is initial, second syllable, antepenultimate, penultimate, or final. In the StressTyp survey, at the left edge, stress is more likely to fall on the first syllable – around 32% of the languages – than on the second syllable – around 4% of the languages, while at the right edge, stress is more likely to fall on the penultimate syllable (around 37% of the languages) than the final syllable (around 17%).

In the North American survey, there are slightly more families with languages that mark stress at the end of the word, although the sample size is quite small and the difference is very small, and most likely not significant.

Compared then with the StressTyp survey, North America shows some differences. Approximately equal numbers of families have languages with quantity insensitive and quantity sensitive stress systems, while the StressTyp survey finds more quantity insensitive languages. Furthermore, there appears to be a preference for accent falling near the left edge in quantity sensitive North America, but near the right edge generally. Finally, in quantity insensitive languages, somewhat more languages show final stress, in keeping with patterns noted in StressTyp. Based on this evidence, it appears that North America as a whole shows some distinct patterns when compared with other languages in StressTyp; more careful study with counts based on similar criteria would clearly be worthwhile to determine whether accent patterns indeed define North America as a linguistic area.
7 Australia (the Arnhem Land clash)

A particularly striking example of an areal phenomenon related to stress (reported in Goedemans 2010a) can be found in Arnhem Land and neighboring areas in the Northern Territory of Australia. In this section, the case will be presented anew, drawing from additional examples found in continued research on the stress systems of languages in the area.

Australia is divided into two linguistic areas. Across most of the continent, languages from only one family are spoken. We refer to these as the Pama-Nyungan languages, a term coined by Ken Hale (cf. Dixon 1980) from the word for “man” in the languages spoken in the northeast and southwest of Australia. Opposed to this continent-wide unity we find an area in the northern parts of the Northern Territory and Western Australia in which linguistic diversity is abundant. Languages from no fewer than 15 different families, the so-called non-Pama-Nyungan languages, are spoken there. Moreover, a few isolated “pockets” of the Pama-Nyungan group can be found there as well.

7.1 Two different patterns: initial and penultimate stress

When we look at the stress systems of the languages in these two areas, we observe a striking difference. The dominant stress pattern for Australian Aboriginal languages is one of the most common patterns we find in the languages of the world (cf. Goedemans 2010b). Although minor variations and exceptions exist, almost all Pama-Nyungan languages place primary stress on the first syllable, and secondary stress on all odd syllables to the right of that primary stress (Initial stress: shorthand “I”). The pattern is illustrated in (15) with some examples from Djambarrpuyngu, a language spoken on Elcho Island off the northeast coast of the Northern Territory (Wilkinson 1991):

(15) Djambarrpuyngu

ˈputuru      'ear'
ˈŋurrupəŋdala    'bush apple’
ˈiithan,mara,nhamirr   ‘dry CAUS+FOURTH+PROP’

This language does not place a secondary stress on the final syllable, producing a final lapse in words with an odd number of syllables. In a common variant on the pattern exemplified in (15) a secondary stress is placed on the final syllable of words with an odd number of syllables. Maranunggu (Daly River, Tryon 1970) stresses the final syllable of láŋkarateti ‘prawn’, while in Djambarrpuyngu it remains unstressed in púthuru.

22 Although not geographically correct, we will refer to this area as Arnhem Land for the sake of convenience.
Among the non-Pama-Nyungan languages, however, many languages place the primary stress on the penultimate syllable, while secondary stresses appear on even syllables before the penult (Penultimate stress: shorthand “P”). The examples in (16) from Limilngan, another Arnhem Land language (Harvey 2001), illustrate the pattern:

(16) Limilngan
\[
\begin{align*}
\text{u'wagi} & \quad \text{‘fire’} \\
\text{ˌlatdin’yayan} & \quad \text{‘crocodile’} \\
\text{ˌuru,galitj’bagi} & \quad \text{‘bandicoot’}
\end{align*}
\]

The cases that are most interesting to us appear in the border areas between groups of languages with penultimate stress and those featuring initial stress. It is quite clear that in the hotbed of diffusional activity that Arnhem Land is (Heath 1978, Dixon 1980), languages influence each other on many linguistic fronts, and stress is no exception. We observe a host of patterns that seem to have features of both the patterns in (15) and (16), but which are in fact different from both. We will call these systems “hybrids” and we will show below that these hybrids do not randomly select features from contact stress patterns. Rather, there is a distinct order to their behavior which allows us to draw a continuum between the patterns in (15) and (16) with discrete steps, all but one of which can be filled with an example language from the Arnhem Land area. Moreover, we will show why these hybrids are of considerable theoretical importance. To understand this fully, we must first briefly explain the basics of the theory that students of stress use in their field.

7.2 Deriving the patterns

To represent the Djambarrpuynngu pattern in (15) and (16) we adopt the so-called metrical approach (Liberman and Prince 1977; Hayes 1995) in which patterns like those in (15) and (16) can be derived in two steps. First, syllables are grouped into binary feet (from right-to-left or from left-to-right). These feet can be left strong (trochaic) or right strong (iambic). In words that contain an odd number of syllables the left-over syllable at the end of the parse can be left unparsed or form a monosyllabic foot. After footing, a second step (called the End Rule) promotes the leftmost or rightmost strong syllable to the status of primary word stress. Given these parameters, the patterns illustrated in (15) and (16) can be analyzed as follows. For Djambarrpuynngu we assign trochees from left-to-right and apply the End Rule left. A final left-over syllable remains unparsed.

(17) Djambarrpuynngu
\[
\begin{align*}
\text{a.} & \\
\text{(* .) (* .) (* .)} & \\
\text{'lithan,mara,nhamirr} & \\
\text{b.} & \\
\text{(* .)(* .)} & \\
\text{ŋurrupaŋgal} & \\
\end{align*}
\]

Maranunggu is identical, except for the fact that a left-over syllable will be parsed as a monosyllabic foot. In Limilngan we assign trochees from right-to-left and apply the End Rule right. In this case, an initial left-over syllable cannot be parsed as a monosyllabic
foot because this would create a stress clash, which is universally disallowed (see van der Hulst 2014c). In (18), we summarize the three analyses:

(18) Parameters | Djambarrpuynu | Maranunggu | Limilngan
--- | --- | --- | ---
Footing (LR/RL) | LR | LR | RL
Foot type (trochee/iamb) | trochee | trochee | trochee
Monosyllabic foot (Yes/No) | no | yes | (no)
End rule (L/R) | L | L | R

Let us now turn to the hybrid systems. Looking at these parameters, we could envisage a continuum between true I languages like Djambarrpuynu and true P languages like Limilngan. A small change (as a result of diffusional pressure) in parameter settings with respect to the values that deliver true right-edge or true left-edge patterns in (18) will alter the stress system of the language in question, moving it towards the other end of the continuum, slightly changing the orientation of stress to one of the word edges. The striking observation with respect to the Arnhem Land contact area is that we find example languages for the full range of possible changes in parameter settings. Before we present those languages, we must address one more theoretical problem which we encounter when we consider a language like Nakara (Eather 1990). This language only deviates from Djambarrpuynu in that it assigns secondary stress from right to left instead of left to right. The primary stress remains firmly on the first syllable, as (19) shows.

(19) Nakara
ˈdijːaɾaˌbaga ‘he emerges’

We observe that, although the step that Nakara takes towards the right edge of the continuum is a minor one, we run into a theoretical problem. The Nakara pattern is not a logical option, given the set of parameters available to us. In a standard metrical system, we should assign feet starting at either the right or the left edge, and then promote one of the feet to primary stress. Nakara seems to indicate that the choice of edge for primary stress and the starting edge of footing for secondary stresses must be stated separately. Cases like Nakara, as well as many other considerations, have prompted van der Hulst (1996) to deviate from standard metrical theory, as explained above, by separating the algorithms for the assignment of primary and secondary stress. Contrary to standard metrical practice, he claims that primary stress is assigned first, after which secondary stress is assigned using its own set of rules (see Goedemans and van der Hulst 2014 for supporting arguments for this position). In Quantity-Insensitive languages primary and secondary stress may be assigned along the following lines.

(20) Parameters for primary stress: Create a bisyllabic domain ([.]...[])
Edge (L/R)
Type (trochee/iamb)

24 These are languages not using the internal make-up of syllables in the stress assignment rules.
25 These distinctions refer to the parameters we use in the StressTyp database to describe the stress patterns of the world’s languages. We ignore some parameters that we do not need in this section.
Parameters for secondary stress: ((…))

Footing (LR/RL)
Foot type (trochee/iamb)
Assign more than one foot (iterative: Yes/No)
Monosyllabic foot (Yes/No)

There is no need for an End Rule in this approach because as a matter of principle the strong syllable in the primary stress domain is the primary stress. Application of these principles to (19) leads to the representation in (21):

\[
(21) \quad |^*\text{ .} | (* \text{ .})
\]

ˈdiʃaʧa, baga

Adopting this alternative model, below we present the Arnhem Land hybrid stress languages in a logical step by step “tour” of parametrical changes from Djambarrpuyngu-type languages to Limilngan-type languages (each parameter represented in bold indicates the crucial change with respect to the languages above it; Iterative foot assignment is indicated by an asterisk * after the direction setting and we leave the settings for monosyllabic feet out, since they are not relevant for the discussion).26

Table 1. Arnhem Land hybrid stress languages

<table>
<thead>
<tr>
<th>Main</th>
<th>Secondary</th>
<th>Description, language and example</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/tr</td>
<td>LR*/tr</td>
<td>Primary stress on the first syllable, secondary on alternates after it.</td>
</tr>
<tr>
<td></td>
<td>Dyambarrpuyngu</td>
<td>ˈiithan, mara, nhamirr</td>
</tr>
<tr>
<td>L/tr</td>
<td>RL*/tr</td>
<td>Primary stress on the first, but secondary on the penult and alternates before it.</td>
</tr>
<tr>
<td></td>
<td>Nakara</td>
<td>ˈdiʃaʧa, baga</td>
</tr>
<tr>
<td>L/tr</td>
<td>RL/tr</td>
<td>Primary stress on the first, but only one secondary stress at the right edge.</td>
</tr>
<tr>
<td></td>
<td>Waanyi (Osbourne 1966)</td>
<td>ˈwabinbara, ulu ‘for turtles’</td>
</tr>
</tbody>
</table>

26 Languages in an area that is as diffusionally active as this one will have scores of exceptions to regular patterns. Large subsets of exceptions might be seen as “a first step” towards the other end of the continuum. There are example languages in this respect, but for reasons of space, we will not discuss them here. See section 7.3, however, for a discussion of exceptions as a sign of instability.
Secondary stress on the penult and alternates before it. Primary stress on the first or second, depending on which is stressed based on the RL rhythm. **Ngankikurrungkurr** (Hoddinott and Kofod 1988)

'weri, lepi 'cave’  a'nimpirr, mire ‘firefly’

Secondary stress on the first and alternates after it. No primary stress.

**Rembarrnga** (McKay 1975) kamu, nunku ‘white ochre’

Secondary stress on the penult and alternates before it and one secondary stress on the first (for which we use the primary stress domain).

**Anindilyakwa** (Leeding 1989) ningkwirri, pwikwi, rriwa ‘you three’

Secondary stress on the first and alternates after it and one secondary stress on the penult.

**Yanyuwa** (Kirton 1977) maṟuwa, ŋala ‘cousin’

Secondary stress on the penult and alternates before it. No primary stress.

**Wardaman** (Merlan 1994) ja, warrga ‘liver’

Primary stress on the penult, and only one secondary stress on the first.

**Umbargarla** (Davies 1989) no example available

Primary stress on the penult, secondary stress on the first and alternates after it.

**Nunggubuyu** (Hore 1981) rawu, rumugu'rumu plant species

Primary stress on the penult, secondary stress on alternates before it.

**Limilngan** ŋuru, galiŋj bagi ‘bandicoot’

We can see that the large variety of intermediate stress systems on our continuum can be straightforwardly analyzed once we adopt the theoretical separation of primary and secondary stress. Without it, the analyses for all these hybrid patterns form a much greater challenge. It would seem, then, that Arnhem Land is a show case for the validity of this separation.

### 7.3 Exceptions and other signs of instability

The first possibility that comes to mind when we think of languages that adopt specific features from neighboring languages involves plain exceptions due to the borrowing of

---

27 The unspecified main stress domain allows main stress to end up on the first or second syllable, whichever one is made “strong” by iterative secondary stress assignment from right to left (see Goedemans, van der Hulst and Visch 1996 for a discussion of this unusual type of stress system called a count system; see also note 26 and the discussion in 6.2). In our view, count systems are hybrids situated well towards the center of the left-to-right spectrum. One could say that languages like Ngankikurrungkurr are still within the left half, since the primary stress domain is located there, but they obviously occupy a niche one step further to the right than, for example, Waanyi in that they have gotten rid of the fixed initial primary stress. Interestingly, Ngan'gityemerri (Reid 1990), a dialect of Ngankikurrungkurr, is of the Djambarrpuyngu type, placing primary stress on initial syllables, while alternates thereafter carry secondary stress. However, that language also shows distinct signs of right-edge diffusional pressure, since in many five-syllable words secondary stress occurs on the penult or final instead of the third syllable.

28 Note that one logical variety is missing in the scheme. There are R and L mirror images for most patterns, but not for Ngangkikurrungkurr. This conspicuous absence is theoretically relevant and we will come back to it in 7.5.
words with preservation of their original stress pattern. When that effect becomes so large that significant parts of the lexicon exhibit the alternative pattern (or a common variant of it) we should include that pattern in the stress description of the language as belonging to a set of “regular” exceptions. In our view, abundance of exceptions in many languages in the area may point to high diffusional activity. We have seen in 7.2 that there is an abundance of hybrid stress types in Arnhem Land that form intermediate stages between true I and true P languages. Add to that the fact that most of the languages we looked at have scores of exceptions and we have all the ingredients for a boiling cauldron of metrical activity. This might mean that some (or all) of the hybrid patterns we have encountered are unstable; they are transitional, and “on the move” along the suggested continuum, and are therefore more weakly anchored in the phonology of the language than the patterns of languages with more common single-edge fixed stress locations. Such systems might be more susceptible to outside influence in the form of (a) relatively swift changes in the rules, (b) harboring a wealth of exceptions, (c) variability of stressing within single words, and (d) maybe even some unexplained metrical effects. Such instability has indeed been reported for Arnhem Land languages, and we review some cases below.

Many of the penultimate stress languages have scores of initial stress exceptions. Two cases in point are Ngalakan and Mangarayi (Merlan 1983, 1989). Some examples are presented in (23):

(22)  
\begin{align*}
\text{Ngalakan} & \quad \text{Mangarayi} \\
ˈdakbaˌrara & \quad ˈwuruˌmumu \quad \text{‘green tree frog’ ‘hornet’} \\
ˌmiliˈbalkiŋ & \quad ˌwarinˈjalan \quad \text{‘salt water’ ‘Exocarpus Latifolius’}
\end{align*}

Merlan notes that these languages even show exceptions that go beyond penultimate or initial stress. Antepenultimate stress is not at all hard to find. It is clear that the stress patterns of these two languages are anything but stable.

In another type of hybrid, stressing in longer words may be relatively uniform, while three-syllable words, often quite susceptible to variation, show alternation between the initial and the penultimate pattern. The pattern is exemplified in (24) by Alawa (Sharpe 1972) spoken in the southern end of the region, to the west of Yanyuwa. In Maung, a language from the north coast (Capell and Hinch 1970) however, a radically different solution is chosen. Three-syllable words in this language reportedly have primary stress on the initial and the penultimate syllable.

(23)  
\begin{align*}
\text{Alawa} & \quad \text{Maung} \\
aˈlawal & \quad ˈbaˈladji \quad \text{‘properly’ ‘bag’} \\
ˈparikal & \quad ˈmaˈninjə \quad \text{‘spear’ ‘clam shell’}
\end{align*}

The Maung examples exhibit what we called “unexplained metrical effects” above. Whereas languages without primary stress, only employing several equal secondary stresses, occur frequently, languages with ‘more than one primary stress’ would appear to be impossible from a theoretical point of view. We do not attempt to solve that issue here. Whatever is going on in Maung three-syllable words, it is clearly not common, and in our view, a sure sign of the instability of its hybrid stress pattern.
Another sign of the instability of hybrids we mentioned above is that, in some languages, changes in the stress pattern could occur relatively swiftly. Such changes are perhaps difficult to capture in grammatical descriptions, which may not be based on contacts with the speakers that is long enough to reveal metrical transitions. Yet, we might have cases in Ngalkbun (Capell 1962, Sandefur and Jenhan 1977) and Ndjébanna (McKay 1975, 2000), two languages spoken on the northeastern fringe of the penultimate area, near the border with the predominantly left-oriented Yolngu languages. Capell classifies Ngalkbun as a typical initial stress language that has the same pattern as Djambarrpuyngu. However, 15 years later, Sandefur and Jenhan write that the pattern in words longer than three syllables is to stress the penult and alternates before it. Three-syllable words also stress the penult, and place a secondary stress on the initial syllable that remains “stranded” in bisyllabic parsing. Some examples from Sandefur and Jenhan are presented in (25):

(24) Ngalkbun
   a. ˌwulkun'tjaŋŋan  'my younger sister'
   b. ˌpaʔjenjyenjtjuŋjan  'I will talk'
   c. na'komtutj  ~  'nakomtutj  'little boy'

(25a) shows the pattern claimed by Sandefur and Jenhan to apply to most Ngalkbun words. (25b) however, reveals the Umbagarla pattern, with only one secondary stress located on the left-hand side. (25c) shows that Ngalkbun is also exemplary of the last sign of instability we mentioned above: variation of stressing within the word. The language as described by Sandefur and Jenhan unmistakably shows signs of affinity with the edge Capell designated as the location for primary stress. Even though we might suspect that one of the descriptions is just plainly wrong, all the signs point at a much more enticing possibility: the sources have captured Ngalkbun at two different stages in its transition from a left-oriented to a right oriented stress system. Similarly, in a few scanty remarks about stress in Ndjébanna, McKay (1975) sketches a left-oriented system, while in a much later source (McKay 2000) a right-oriented pattern seems to dominate the scene.

With this discussion on exceptions and their signal function for instability we conclude the overview of Arnhem Land stress types. We have seen what all these hybrid patterns tell us theoretically. They strongly support van der Hulst’s claim that the algorithms for the assignment of primary and secondary stress must be separated. What we have not yet seen is how these languages are located with respect to each other. In the next section we will check whether the hybrid patterns have something to tell us when we look at their geographical locations.

7.4 Geographical distribution

To create a perspicuous map, we have taken the geographical language data for Google Earth that were created in the Autotyp project (http://www.autotyp.uzh.ch/). We have excluded all the languages for which we had no stress data, and divided the others into three categories:

(25) 1. Predominantly initial stress (Djambarrpuyngu type, white dots)
2. Predominantly penultimate stress (Limilngan type, black dots)
3. Hybrids (grey dots)

In categories 1 and 2, we have incorporated languages that are of the prototypical type but which do have exceptions. Even though we introduced these as having taken the first step towards the other edge, and therefore hybridity, we felt it would be more revealing to include only the languages for which hybridity is more prominently present in category 3. The cleaned-up Autotyp map is shown in Map 1.

Map 1. Arnhem Land area with P, I and Hybrids

What strikes us immediately is that the black dots split the area down the middle. A pocket of white initial stress languages is isolated in the northeast, while some other white dots are located in the west. In the southeast, not visible on the map, some Pama-Nyungan initial stress languages are the closest neighbors of Yanyuwa, Garawa and Waanyi. Most hybrids are located in the border areas, where the initial and penultimate stress systems collide. We already noted in the introduction that this contact area is where almost all hybrids appear, but their exact locations with respect to I, P and the other hybrids only now becomes clear. Perhaps most noteworthy in this respect is the fact that we do not see any hybrids in the south. We do not know the reason for this. There are Pama-Nyungan languages there, but diffusional forces between these and Djingulu and Gudanji, if any, have not (yet) resulted in any overtly hybrid stress patterns.

7.5 Going left, or going right? Innovation or remnant?
In this case study, we have discussed a journey along a continuum from initial (I) to penultimate (P) stress, as if that were the direction in which the languages in the Arnhem Land area are evolving. But we have in fact thus far made no claim regarding the issue of the direction of diffusion. This is, however, a most intriguing question, and we endeavour to answer it in this section.

What is really happening in Arnhem Land? Has the P stress innovation invaded the area, forcing back the native I pattern to the northeast and the west? Or was the P pattern once used across the board in all non-Pama-Nyungan languages and is it now being slowly mopped up by the Pama-Nyungan I forces? Or is our impression of a “journey” that languages can make to either end of our spectrum a false one? Maybe the intermediate hybrid “states” that we have found above are in fact stable stress patterns and no evolution of stress patterns to more left or right oriented versions has taken place at all. Even though it is difficult to be certain, given the highly volatile nature of most of the data we have found, we do believe that we can answer these questions. It is our firm belief that the metrical scene in the Arnhem Land area is in turmoil. Almost all the data we have found in grammars point at systems that are changing. A myriad of exceptions to patterns and rules reflects a situation of great instability. Also on theoretical grounds, as we noted in section 7.5, we could assume that hybrids will not always remain hybrids, but are on the move towards more common stress patterns. Therefore, firstly, we propose here that these languages are changing as the result of diffusional pressure, and secondly, that they are going rightward to become more and more P-like. For the second claim, we have not yet given the arguments. We now review what evidence is available.

Our first piece of evidence concerns the northeast of Arnhem Land, the isolated “white pocket” in Map 1. The suggestion that this group of Pama-Nyungan languages is slowly giving in to pressure of the P languages to the west turns out to be correct in the light of claims made by Dixon (1980). Dixon shows that the Yolngu tribes speaking the languages in the area moved there only recently, and notes that after the migration, considerable diffusional pressure towards the Yolngu languages (I languages) originated from the P languages to the west. It is therefore most likely that the stress patterns of languages like Nakara developed from true I patterns through influence of P languages, and not the other way around.

Evidence for the same direction of diffusion can be found in Yanyuwa (Kirton 1977). In long polymorphemic words, a tendency is developing to replace the stress on the initial syllable of a word-internal morpheme by a stress on the pre-antepenultimate syllable. Compare the alternatives in (27).

(26) Yanyuwa

\[
\begin{align*}
\text{gumba} & \quad \text{ramanda} & \quad \text{ninjdja} & \quad \text{‘he was hitting himself’} \\
\text{gumbara} & \quad \text{manda} & \quad \text{ninjdja}
\end{align*}
\]

The top pattern is the traditional one, the bottom word represents the tendency. Clearly, this tendency reflects a movement to a more rigid right-to-left way of assigning

---

29 Note that we can do nothing but make claims about tendencies here. In an area as diffusionaly active as this one, pressure may well work both ways, and some stress systems might evolve leftward, while others go to the right, and some may even go back and forth between directions before they make their final choice. The case made here will simply concern the direction in which the languages are generally going.
secondary stresses. This is the only language-internal evidence available in favor of movement towards penultimate stress.

A theoretical argument for the movement towards penultimate rather than initial stress can be found when we consider the absence of left-to-right count systems \(^{30}\) in the diffusional melting pot of Arnhem Land that we noted in section 7.2. In this chapter we cannot delve into the full arguments \(^{31}\), so we must summarise. To derive a right-to-left count system from a basic non-count pattern we can walk two very different paths: 1) we could start out with a basic initial stress pattern and switch the starting edge of footing from the left to the right side while leaving the primary stress domain where it is or 2) starting with the domain at the right edge in a penultimate stress system, we could derive the desired count system if we flipped the domain to the left while leaving the starting edge of footing at the right. However, in our view such a movement of the domain is far too drastic to constitute the first reaction of a language to diffusional pressure, and once other, less drastic, evolutionary steps have been taken, the right-to-left count system can no longer be derived. Therefore, we submit that I-systems are the point of departure for Arnhem Land hybrids. For left-to-right count systems, the argument would be the inverse of the above, with P-systems forming a point of departure. We propose, therefore, that the absence of such left-to-right count systems in the area can be interpreted as evidence for the claim that, in principle, Arnhem Land languages do not move from P to I-systems.

Two scenarios for the development of diffusional pressure in Arnhem Land were sketched above. P-systems could be relatively new on the scene, expanding their influence on the much older I-systems. Alternatively, P-systems could be remnants of a much older pattern that is on the retreat under the pressure of the I-systems. There is no doubt that P-systems are indeed an innovation with respect to Proto-Australian, which had simple initial stress across the board (Dixon 1980). But that may still mean that P-systems could have arisen long ago and are now slowly fading away. To choose between the two scenarios, we need to look for a trigger, a relatively recent development that could have caused the genesis of P-systems or be the agent of their demise. It so happens that the former is easy to find. Both Heath (1978) and Dixon (1980) report prefixing as a rather new phenomenon among the non-Pama-Nyungan languages. It is quite likely that the penultimate stress pattern arose as the result of the addition of prefixes to stress initial stems, many of which will have been bisyllabic. A penultimate, or at least a more right-oriented, primary stress, perhaps already with some secondary stresses on the chain of prefixes, might have been the result. \(^{32}\) This pattern may have spread rapidly among the non-Pama-Nyungan languages, in the wake of the prefixing innovation. In this scenario, penultimate stress has become more regular over time and is now exerting its influence over non-prefixing neighbouring languages. The fact that a trigger for P-innovation can be found so easily adds to the evidence for the claim that hybrids become more P-like through pressure from the penultimate stress systems.

The evidence presented above does not make a rock solid case for the claim that diffusion indeed pushes back the I-languages in Arnhem Land. It does, however, provide some fairly firm ground to walk on when we take new steps towards understanding the metrical upheaval that these languages are subject to. Much more scrutiny of linguistic

---

\(^{30}\) In a count system the starting edge for footing is opposite to the edge selected by the end rule; see van der Hulst (2014c).

\(^{31}\) The interested reader is referred to Goedemans (2010a).

\(^{32}\) See Goedemans (2010a) for a more in-depth discussion.
data is needed before we can say anything more definitive. We hope that such data will become readily available in the near future.

8 Conclusions

While this chapter has referenced different approaches and studies concerning contact-induced change, linguistic areas, and word prosodic systems, our main conclusion is that the study of these topics is largely an under-investigated area. There are challenges in identifying whether similarity of stress systems in an area is due to independent factors or is contact-induced. We definitely see areal effects and many very likely cases have long been identified. In this chapter we have provided information and references concerning the typological distribution of word prominence systems in selected languages / language families of the world in the hope that proper considerations of these will lead to the identification of additional linguistic areas. However, we believe that the study of contact-induced change in word prominence systems is still largely dominated by anecdotal reports and much is uncertain about the precise mechanisms that are at play. (A notable exception is the detailed study of Salmons (1994).) This, in part, is caused by a typical holistic approach to stress/accent that does not take into account that these notions more often than not cover a package of properties at both the phonological and the phonetic levels (see 2.2). We have also suggested that when moving toward an understanding of the mechanisms of contact-induced changes much can be learned from descriptive and experimental work on loan phonology and second language phonology (see 2.4).

By looking at three cases in more detail, we have tried to move beyond anecdotal reports. In section 5 we discussed the case of Basque dialects, based on the work of José Hualde. This shows how a prominence system can respond to language contact in various different ways. It would clearly be desirable to have more studies of this kind within the context of specific modular theories of word prominence systems which also take into account the interplay between word prominence and intonation (Gordon 2014). Section 6 discussed in detail the word prominence evidence for considering certain previously identified language clusters in North America as linguistic areas, showing that prominence is an areal indicator in some areas but not in others. Finally, section 7 offered a detailed case study of the consequences of language contact in northern Australia, clearly showing that a modular, parametric analysis of word prominence systems provides insights into the myriad of attested systems. Overall, we have argued that contact situations can create hybrid systems which provide clear descriptive and theoretical challenges for areal linguistics.

References


van der Hulst, Goedemans & Rice -- page 188 of 393


7  Semantic patterns from an areal perspective

Maria Koptjevskaja Tamm and Henrik Liljegren

1.  Introduction

The aim of this chapter is to define, exemplify and problematize the lexico-semantic aspects of areality cum language convergence. Just like with areal linguistics at large, this subfield is an intersection of a number of more general approaches: the study of language contact, language change and typological research. Particularly prominent in our treatment is the cross-fertilization of meticulous documentation of linguistic features in specific geographical regions or areas, often grounded in fieldwork, and large-scale cross-linguistic findings and generalizations.

Areal semantics, in its concern with the diffusion of semantic features across language boundaries in a geographical area, is a potentially vast field, spanning the convergence of individual lexemes, through the structuring of entire semantic domains to the organization of complete lexicons. For practical reasons, we have largely excluded the two extreme ends of that continuum. At the lower end, loanwords, or the spread of individual vocabulary forms, will only feature when it also involves co-lexicalization or the diffusion of area-specific concepts. At the higher end, we will largely exclude features defining the lexicon at large, such as very general derivational mechanisms or the proportion of verbs vs. nouns, sometimes referred to as the lexical profile of a language. The latter is related to an attempt to uphold a distinction between grammar and lexicon, also reflected in our decision to exclude grammaticalization per se from our discussion.

Our choice of focus for the chapter is to a large degree motivated by a desire to attend to the issues that have so far received relatively little attention in theoretical discussions in areal linguistics (cf. Ameka and Wilkins 1996, Matisoff 2004). As a consequence, we will hardly be able to report on generalisations comparable to those accumulated in research on “material” borrowings or on contact induced grammaticalisation, such as constraints on borrowability or correlations between the sociolinguistic parameters of contact situations and possible contact-induced change. The main part of this chapter is an overview of a number of lexico-semantic phenomena that have been shown, or at least suggested, to serve as indicators of areality. The phenomena dealt with are lexico-semantic parallels, in their turn subdivided into polysemy calquing and lexico-constructional calquing (Section 2), shared formulaic expressions (Section 3), area-specific lexicalizations and a shared or similar-looking internal organization of certain semantic domains (Section 4). Section 5 is a case study of semantic patterns in the Hindukush region. After that (Section 6) follows a discussion of causality and possible mechanisms behind obvious correlations between lexico-semantic phenomena and geographically contiguous areas: inheritance, diffusion, shared environment and independent innovation. Section 7 offers a conclusion.

2.  Lexico-semantic parallels (calques)
The two traditionally distinguished groups of contact phenomena in the lexicon are loanwords and calques – the distinction paralleled by contact phenomena at other levels (‘replication of matter’ vs. ‘pattern replication’ in Matras and Sakel 2007, cf. also Croft’s 2000 distinction between ‘substance linguemes’ and ‘schematic linguemes’, and Heine and Kuteva’s 2005 notion of ‘polysemy copying’). Calques encompass a broad group of lexico-semantic parallels between two or more languages with respect to the range of meaning(s) of the expressions that may count as interlingual matches (cf. Weinreich’s 1953: 7–8, 32 notion of ‘interlingual identification’), as well as their internal structure, as will be clarified and exemplified later.

2.1 Polysemy calquing

The first subtype of these phenomena includes polysemy calquing, or polysemy copying (cf. Heine and Kuteva 2003, 2005), also called “semantic borrowing”, “semantic loan”, “semantic shift” or “loan synonym” (cf. Urban 2012 for the terminology). In polysemy copying the same two meanings are expressed by the same form in each of the languages (normally by different forms in different languages), or are “colexified”, to use François’ (2008) term. For instance, the Spanish verde ‘green, unripe’, is semantically extended to also mean ‘raw’ in the speech of the Acatec on the model of the Acatec (a Mayan language of Guatemala) word yaaš (Smith-Stark 1994). Here the semantic extension is triggered by the interlingual identification of verde and yaaš by the bilingual Acatec–Spanish speakers on the basis of their two overlapping meanings – ‘green’ (colour) and ‘unripe’. However, since yaaš, in addition, has the meaning ‘raw’, verde has extended its semantic range to ‘properly’ align with its interlingual match, which leads to the convergence of the two languages with respect to the semantic structure of these two lexical items. The whole process is visualized in Figure 1, inspired by Gast and van der Auwera’s (2012) notion of ‘semantic map assimilation’:

![Figure 1: The process of polysemy copying – Spanish verde and Acatec yaaš in the speech of Acatec-Spanish bilinguals (after Smith-Stark 1994)](image)

In the verde-yaaš example we can easily identify the model and the replica language, given the general knowledge about both languages involved and the contact situations. Since things are often much more complicated, especially when many languages are involved, in the rest of the chapter we will be talking about “shared polysemies” or “shared colexification patterns”, rather than about “polysemy copying / calquing”, without making any further commitments about the possible reasons for these similarities. Examples (1–3) illustrate some patterns of polysemy shared among the languages spoken in the same area:
Polysemy sharing across languages

(1) ‘draw water’ = ‘copy, imitate’ in the languages of Ethiopia-Eritrea: 
    *k’ädda* in Amharic (Afro-Asiatic, Semitic), 
    *war[aabe]* in Oromo (Afro-Asiatic, Cushitic) and 
    *duuk k’ides* in Gamo (Afro-Asiatic, Omotic) (Hayward 1991, 1999)

(2) ‘child’ = ‘fruit’ in West-African languages: 
    *
    ñ
    * in Mandinka (Mande, West Mande) and in several other Mande languages, 
    *doom* in Wolof (Niger-Congo, Atlantic), 
    *țe* in Songhay (Nilo-Saharan, Songhai), 
    *fidju* in Kabuverdianu (Portuguese-based creole), 
    *obi* in Sɛlɛɛ (Niger-Congo, Kwa)

(3) ‘eat’ = ‘drink’ in many Papuan and Australian Aboriginal languages, 
    e.g. *ka*- in Manambu (Ndu) or 
    *a* in Kwoma (Kwoma-Nukuma) (Aikhenvald 2009), as well 
    as in a number of other languages of the world (Vanhove (ed.) 2007).

2.2 Lexico-constructional calquing

For the other subtype of calquing, which we will call lexico-constructional calquing (also 
    called “translation loans”), it is both the meaning and the internal structure of the 
    expressions, that is in focus. What matters here is that expressions that constitute 
    interlingual matches show the same semantic and structural patterning, i.e. are made up 
    of parts that match each other across the languages. The term “calque” was for a long 
    time exclusively or mainly used for exactly such cases, with only a recent inclusion of 
    polysemy copying into its scope. For instance, Singlish (or Singapore English) abounds 
    in lexicalized phrases that clearly replicate the model of the local languages, as in the 
    following two examples of loan translations from Mandarin, visualized in Fig. 2:

(4) Singlish vs. Mandarin (http://www.singlishdictionary.com/)
    a. *eat salt* vs. *chí ‘eat’ + yán ‘salt’ – ‘suffer a bitter or serious setback’
    b. *give face* vs. *gě ‘give, grant’ + miàn ‘face; reputation, prestige’ – ‘show due 
    respect for one’s feelings’

<table>
<thead>
<tr>
<th>Mandarin</th>
<th>Singlish</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>chí</strong> ‘eat’</td>
<td><strong>eat</strong></td>
</tr>
<tr>
<td><strong>yán</strong> ‘salt’</td>
<td><strong>salt</strong></td>
</tr>
<tr>
<td>‘suffer a bitter or serious setback’</td>
<td>‘suffer a bitter or serious setback’</td>
</tr>
</tbody>
</table>

Figure 2: Lexico-constructional parallels – Mandarin *chí yán* and Singlish *eat salt*

Again, since in most areal contexts it is difficult or even impossible to identify the model 
and the replica languages, we will be using a more neutral term “shared lexico- 
    constructional patterns”. Examples in (5)–(8) illustrate several types of lexico- 
    constructional patterns shared among the languages in the same area, ranging from 
    derivational patterns to collocations.

Lexico-constructional parallels across languages
‘need’ = causative of ‘want’ (a shared derivational pattern) in the Ethio-Eritrean languages: asfällägä (from fällägä) in Amharic, barbaachise (from barbaade) in Omoro and kosshides (from koyides) in Gamo (Hayward 1991, 1999)

‘sun’ = ‘eye of the day’ (a shared compounding pattern) in Mainland Southeast Asia and parts of Oceania: mata hari in Malay/Indonesian (Austronesian, Malay-Polynesian), wangere ma la'o ‘day POSS eye’ in Sahu (Papuan, North Halmahera), mata-ni-siga ‘eye/face-POSS-day/sun’ in Fijian (Austronesian, Malayo-Polynesian), masonândro (mäso-nändro) in Malagasy (Austronesian, Malayo-Polynesian) (Urban 2010, 2012, Blust 2011)


‘to obey someone’ = ‘to follow someone’s mouth’ (a shared collocational pattern) in the languages of Karkar island (Papua New Guinea): awa-n ga-ri ‘mouth-3SG.POSS 1SG.S-follow’ in Takia (Austronesian, Oceanic) vs. kury karotu-sam ‘mouth:3SG.POSS follow-1SG.S’ in Waskia (Nuclear Trans New Guinea, Madang) (Ross 2007: 122)

Polysemy sharing and lexico-semantic sharing are different manifestations of lexico-semantic parallels among languages, but there is no strict borderline between the two. Different meanings of a polysemous lexeme are normally associated with different constrictions, and polysemy in fact often arises due to the ellipsis of additional disambiguating material. To give a simple example, some languages share the polysemy ‘child’ = ‘fruit’, as in (2). In some other languages, however, the latter meaning requires a combination with ‘tree’, such as ‘child of tree’, resulting in a shared lexico-constructional pattern. However, even in the languages with the ‘child’=’fruit’ polysemy the latter meaning may also be prevalent in combinations with ‘tree’, which makes it difficult to draw a strict line between shared polysemy and shared lexico-constructional patterns. The common denominator here is the shared semantic association between ‘child’ and ‘fruit’ (cf. Matisoff 2004, Vanhove (ed.) 2008, Urban 2012).

2.3 Lexico-semantic parallels as areality indicators: Mesoamerica and Ethiopia-Eritrea

Examples of lexico-semantic parallels abound in the literature on contact phenomena; among others, they figure prominently in discussions of metatypy (e.g., Ross 2007). Some have been discussed or at least mentioned in connection with various regions of different sizes and significance. A few of these have been attested and described across a large number of languages. Hendery et al.’s (in prep.) careful study of the different colexification patterns between ‘tree’, ‘firewood’ and ‘fire’ includes 275 languages spoken in Sahul, i.e. Australia, New Guinea and surrounding islands; Aikhenvald’s (2009) sample for the ‘eat’= ‘drink’ polysemy includes 36 languages from the same region (28 Papuan and 8 Australian aboriginal languages). Other lexico-semantic parallels are often mentioned in a more implicit way. Our impression is that there is much
knowledge on such shared patterns among experts on particular languages, language families and linguistic areas, which often remains concealed to outsiders.

There is surprisingly little discussion of the role of lexico-semantic parallels in areal linguistics. Some of the exceptions, in addition to the above mentioned studies by Hendery et al. and by Aikhenvald, include Austin et al. (1976) and Evans (1992), who discuss or at least mention a number of semantic associations shared by the aboriginal languages of Australia, Matisoff (1978, 2004), who focuses on various semantic associations in South-East Asian languages, and Pardeshi et al. (2006), who deal with the meanings, functions and structure of expressions involving the verb ‘eat’ across the Asian languages. Vanhove 2008 (ed.) (cf. also http://www.typologie.cnrs.fr/spip.php?rubrique131) focuses on several cross-linguistically recurrent patterns in polysemy and semantic change across 45 languages. Heine and Leyew (2005) and Güldemann and Fehn (this volume) mention some lexico-semantic parallels of potential relevance for Africa, including the Kalahari basin, as a group of linguistic areas. Enfield’s 2003 excellent study on the multi-functionality of ‘acquire’ in the languages of mainland Southeast Asia is mainly concerned with patterns crossing the border between lexicon and grammar.

We are aware of only two linguistic areas where lexico-semantic parallels have been systematically used as areality indicators – the Meso-American and the Ethio-Eritrean languages. Each of the cases involves a list of lexico-semantic parallels that have been checked both for the languages belonging to the presumed linguistic area and for those that do not.

For the Meso-American languages the discussion of lexico-semantic parallels as evidence for linguistic contact and, further, as indicators of linguistic areas, goes back at least to the 1970s and has involved such prominent linguists as Kaufman, Campbell, Smith-Stark and Brown. In the present context, the two most relevant publications are Smith-Stark (1994) and Brown (2011).

Smith-Stark (1994) tests to what extent 52 lexico-semantic parallels that have earlier been detected in the Meso-American languages, may in fact count as strong areality indicators. The study embraces 46 languages, of which 25 come from Mesoamerica (MA) and represent its genetic and geographic diversity, 11 languages border the region, and 10 languages comprise five North-American and five South-American languages well removed from the Mesoamerican area. The procedure takes into consideration three parameters: to what extent the feature occurs throughout the region; its frequency in Mesoamerica (MA), and its limitation to MA (within the Americas). Fourteen lexico-semantic parallels turn out to be bound to MA, where they are both widely distributed. Among these nine occur in at least four language families: ‘boa’ = ‘deer snake’, ‘lime(stone)’ = ‘rock ashes’, ‘wrist’ = ‘neck (of hand)’, ‘molar’ = ‘grindstone’, ‘mouth’ = ‘edge’, ‘thumb’ = ‘mother (of hand)’, ‘finger’ = ‘child (of hand)’, ‘poor’ = ‘widow’ = ‘orphan’, ‘alive’ = ‘awake’, ‘marry’ = ‘find/meet’.

Cross-linguistic similarities are of course particularly valuable as areality indicators whenever it is possible to trace their origin and the details of their diffusion, and to link them to the history of the linguistic contacts in an area. While this is often a relatively easy task for lexical borrowings, where the forms may betray their origin, finding out the history of lexico-semantic parallels is a much trickier endeavour. This step is undertaken by Brown (2011) who attempts to account for the circumstances behind the shared lexical traits in Mesoamerica, such as which languages acted as donors, when the traits spread, etc. To this end he further tested thirteen of the suggested Mesoamerican lexico-semantic
parallels in a sample of 70 Native American languages in other areas. According to his conclusions, only five of these most probably originated in a single area language from where they diffused across Mesoamerica (among others, ‘boa’ = ‘deer snake’ and ‘lime(stone)’ = ‘ashes’). Brown attributes the decisive role here to Nahuatl, the major lingua franca in Mesoamerica (first in the Aztec empire and later in New Spain). Although this conclusion is far from certain, the study is an impressive attempt to put areal lexico-semantic parallels in a larger historical framework.

The systematic research on lexico-semantic parallels in the Ethio-Eritrean languages has so far been relatively modest in that it only covers four languages (see, however, Treis 2013, mentioned in Section 6, for an update). Hayward’s (1991, 1999) list contains 40 lexico-semantic parallels, among others ‘draw water’ = ‘copy, imitate’ in ex. (1), ‘need’ = causative of ‘want’ (5), or ‘foreign country’ = ‘land of man/person’. These are shared by three languages representing the three main groups of the Ethio-Eritrean languages – Amharic (Semitic), Oromo (Cushitic) and Gamo (Omotic), all belonging to the Afro-Asiatic phylum. The list has also been tested for another East Lowland Cushitic language, Somali, which is both closely related to many of the languages in the Ethio-Eritrean area and is also spoken in their vicinity. Strikingly, Somali shares only four of the 40 patterns in the list, which fits into the accepted view on the history of the Somali within the Horn of Africa. In a nutshell, Somali groups have not had enough time to develop close contact with the speakers of the other languages more deeply integrated in the Ethio-Eritrean sprachbund, since they arrived there relatively late (most probably from northern Kenya). In addition, most of their social networks have been within the larger Muslim context, rather than with the other ethnic groups within Ethiopia and Eritrea.

Summarising, Smith-Stark’s (1994), Brown’s (2011) and Hayward’s (1991, 1999) studies of the Mesoamerican and the Ethio-Eritrean languages suggest that lexico-semantic parallels are a powerful tool for measuring membership of a linguistic area. As emphasised in these studies, the idiosyncratic nature of lexico-semantic parallels, their potential multiplicity and logical independence of each other lead to a great increase in the number of quantifiable properties. This makes them potentially more advantageous as areality indicators than the seemingly more important structural properties that are often listed as indicators of language convergence, such as word order or phonological contrasts. These latter features may show various interdependencies and be in the long run reducible to too few independent properties for quantifying convergence. However, evaluating any cross-linguistic similarities as areal indicators is always a challenge, and the same goes for lexico-semantic parallels (cf. the discussion in Section 6). Similar lexico-semantic parallelisms may be found in other regions of the world for different language groups. What often counts is that they are found in contiguous zones and in bundles with others.

3. Shared formulaic expressions

Conventionalized formulaic expressions used for particular pragmatic functions (e.g., greetings, curses, proverbs, etc.) represent a special case among shared lexico-constructional patterns, cf. the familiar farewell expressions au revoir (French), auf Wiedersehen (German), på återseende (Swedish), do svidanija (Russian), näkemin (Finnish), that follow the same model across a number of European languages. Such
conventionalized formulae are occasionally mentioned as areality indicators, as, for instance, the curse ‘eat earth!’, included in Hayward’s (1991, 1999) list over the lexico-semantic parallels in the Ethio-Eritrean area. Ameka (2006, 2011) quotes several interactional routines including proverbs shared among the West African languages, primarily those in the Volta Basin, e.g. good-night wishes (‘sleep/lie well’) thanking or leave-taking expressions. The formula ‘When I die don’t cry (i.e., don’t mourn for me)’ as expression of extreme gratitude in (9) builds on the West African cultural requirement to publicly show sorrow during the funeral by crying and wailing, and people who do not participate in this public display of sorrow are viewed with greatest suspect. However, “when someone does something very good for you, by absolving them from doing the things that one is expected to do when other people die, one is saying that the favour that has been received is like the ultimate thing or even more than it” (Ameka 2011: 253). The leave-taking expressions in (10) reflect the common West African cultural model that a visitor cannot just leave without first asking the host for a permission to do so. Pre-closing requests (‘I am asking for a way/road’) are genuine requests that can be answered both positively and negatively, initiating a whole chain of further communicative exchanges between the host and the visitor (Ameka and Breedveld 2004: 171–172).

(9) Expressions of extreme gratitude in West Africa / in the languages of Volta Basin (Ameka 2011: 232, 254)

(a) Ewe (Niger-Congo, Kwa, Gbe; Ghana and Togo)
Né me-kū lá, me-ga-fa aví o.
COND 1SG-die TP 2SG:NEG-REP-shed cry NEG
‘When I die, don’t cry.’

(b) Akan (Niger-Congo, Kwa, Tano; Ghana)
Se ma-wu-a, n-su.
COND 1SG-die-TP 2SG:NEG-cry.
‘When I die, don’t cry.’

(c) Dagaare ((Niger-Congo, Gur, Oti-Volta; N Ghana, Burkina Faso)
Ka maa wa kpi tɔɔ kono.
If 1SG come die NEG:IMP cry
‘When I die, don’t cry.’

(d) Gã (Niger-Congo, Kwa, Ga-Dengme; SE Ghana)
Ke o-ŋu akɛ p-gbɛ-ɛ kaa fo,
COND 2SG-hear QUOT 1SG-die-TP NEG cry
‘When you hear that I am dead, don’t cry.’

(e) Moore (Niger-Congo, Gur, Oti-Volta; Burkina Faso)
M sã nki bi y ra yâb ye.
1SG cond die that 2:POLITE NEG cry NEG
‘When I die, please don’t cry.’

(10) Pre-closing requests in the languages of Volta Basin (Ameka 2006: 138–139)

(a) Likpe ŋ-tɔ ku-sù lɔ

(9) Expressions of extreme gratitude in West Africa / in the languages of Volta Basin (Ameka 2011: 232, 254)

(a) Ewe (Niger-Congo, Kwa, Gbe; Ghana and Togo)
Né me-kū lá, me-ga-fa aví o.
COND 1SG-die TP 2SG:NEG-REP-shed cry NEG
‘When I die, don’t cry.’

(b) Akan (Niger-Congo, Kwa, Tano; Ghana)
Se ma-wu-a, n-su.
COND 1SG-die-TP 2SG:NEG-cry.
‘When I die, don’t cry.’

(c) Dagaare ((Niger-Congo, Gur, Oti-Volta; N Ghana, Burkina Faso)
Ka maa wa kpi tɔɔ kono.
If 1SG come die NEG:IMP cry
‘When I die, don’t cry.’

(d) Gã (Niger-Congo, Kwa, Ga-Dengme; SE Ghana)
Ke o-ŋu akɛ p-gbɛ-ɛ kaa fo,
COND 2SG-hear QUOT 1SG-die-TP NEG cry
‘When you hear that I am dead, don’t cry.’

(e) Moore (Niger-Congo, Gur, Oti-Volta; Burkina Faso)
M sã nki bi y ra yâb ye.
1SG cond die that 2:POLITE NEG cry NEG
‘When I die, please don’t cry.’

(10) Pre-closing requests in the languages of Volta Basin (Ameka 2006: 138–139)

(a) Likpe ŋ-tɔ ku-sù lɔ
Formulaic expressions are of great potential value as areality indicators, in particular due to their idiosyncratic properties. First, since their meaning is normally not compositional in the normal sense, their chances of being shared as a result of independent innovation is in general lower than for many other expressions. Second, speakers have to learn additional conversational routines and conventions as to which formulae should be appropriate for which pragmatic functions. These are therefore normally shared via socialization and repeated communication across languages, which implies more than sporadic language contact.

Formulaic expressions are often permeated with shared cultural scripts and values and may bear testimony to the shared cultural history of the area. This is amply demonstrated in Piirainen’s (2012) study of European figurative idioms, or items of phraseology. Piirainen, in collaboration with numerous colleagues (about 250) and experts spread over Europe, collected data on idioms in 73 linguistic varieties spoken in Europe, 17 non-European languages and Esperanto. As the result of this unprecedented effort she was able to identify 380 widespread European idioms. These can be traced to various sources, such as texts of ancient writers, the Bible, post-classical literature, proverbial units of medieval and reformation times, and fables, tales and legends. The six most widely spread European idioms (the number of the languages using them are in square brackets) are night and day [69], to be/fight like cat and dog [68], to be someone’s right hand [64], to play with fire [64], to take someone under one’s wings [62], and to tear/pull one’s hair out [62]. Interestingly, the European area, as defined by its core lexicon of idioms, is organized somewhat differently from that defined by shared grammatical structures. Whereas structural features place West Romance in the prototypical Standard Average European area and Slavic in its periphery (cf. van der Auwera 2011), it is the Slavic languages that share the European figurative lexicon to a greater extent than West Romance. Likewise, while both Armenian and Arabic are quite non-European from the structural point of view, yes they turn out to be much more “European” in their idioms than the receding languages West Frisian and Kashubian, spoken in the core of Europe.

4. Area-specific lexicalizations and shared organization of semantic domains

Experts occasionally mention area-specific concepts, i.e. concepts that are lexicalized across languages in a particular area, but strike outsiders as very specific and curious. The extent to which these testify to language contact is, of course, not always clear. Some of them may be rooted in the shared physical environment, others bear witness to shared material culture and/or cultural values and practices, which may go hand in hand with language contact, but do not have to. To give a few examples, the languages of Ethiopia-Eritrea lexicalize the same four seasons of the year, similar categories of terrain, similar skin colour classification and particular kinds of artefacts (like a wooden chair). Another shared lexicalization is ‘die without ritual slaughter (of cattle)’, implying that the meat is
unfit for consumption – bäkkätä in Amharic, rak’e in Oromo and bowutides in Gamo (Hayward 1991, 1999). West African languages sometimes have a special unanalyzable lexeme for ‘joking-relative’, a notion which has no equivalent in European cultures. The joking relationship may be of three kinds: between ethnic groups, between clans, or between kinspersons of certain categories (for example, between a woman and the younger brothers of her husband) – all of them are covered by the word sêmówu in Mandinka (Niger-Congo, Mande; Senegal), gaammu in Wolof (Niger-Congo, Atlantic; Senegal), or bërrëd in Bomu (Niger-Congo, Gur; Mali, Burkina Faso) (Denis Creissels and Martine Vanhove p.c.). Matisoff (2004) quotes the Jingpo (Sino-Tibetan, Tibeto-Burman) expression my it ìtòwâm~my it ìtòwâm ‘to be deterred by feelings of respect, embarrassment, fear of offending; be generally restrained in one’s interpersonal behaviour by the knowledge that self-assertiveness is not socially approved’ as paralleled by Thai (Tai-Kadai, Kam-Tai) krêŋ-çaj, Burmese (Sino-Tibetan, Tibetan-Burman) ðà-nû, Japanese (Japonic) enryô suru, "reflecting a mind-set more typical of the [Southeast Asian] region than the more aggressive interpersonal ideal in Western competitive societies” Matisoff (2004: 369, see also Marlan 1979).

Shared areal lexicalizations pertaining to shared physical environment, material culture and cultural values and practices, are, of course, extremely valuable for areal studies. From a theoretical linguistic point of view it is, however, even more interesting to ask whether areal convergence can lead to similar lexicalizations within more ‘universal’ lexical semantic domains. Some of Hayward’s (1991, 1999) shared lexicalizations in the Ethio-Eritrean area may be included in this category, for instance, the adjectives t’àfäff yală in Amharic, kafaʃa in Oromo and ts’izà in Gamo, which all mean ‘dry enough for use’ (clothes that have been washed for wearing, a road for travelling, firewood to be used as fuel, etc.), or the two different verbs for ‘borrowing something to be returned in kind only (like money)’ and ‘borrowing which is itself to be returned’.

A fundamental issue here is whether and/or to what extent languages can in general lexically converge in a particular semantic domain. Ricca’s (1993) study of deictic verbs in the European languages is interesting here. The twenty languages in the sample fall into three groups depending on the extent to which they make a systematic distinction between verbs showing centripetal (to the deictic centre) vs. centrifugal (from the deictic centre) motion. Significantly, the distribution of the types across the sample is dependent on a combination of genetic and areal factors, where the fully deictic languages are mainly found in South-western and Southern Europe (Portuguese, Spanish, Italian, Albanian, Modern Greek, with the two Finno-Ugric outliers Hungarian and Finnish), the non-deictic languages are Western and Eastern Slavic and Baltic, while the predominantly deictic ones are Germanic, along with French and the two Southern Slavic languages Serbo-Croatian and Slovenian.

Another instructive example is provided by van der Auwera’s (1998) research on phasal adverbials, again in the European languages. These languages normally have systems that contrast the four adverbials still, no longer, not yet, already with each other, which is by no means universal and is also lacking in a few languages on the fringes of Europe. The languages with four adverbials fall further into two large groups, depending on how they express no longer. English, French (ne plus) or Swedish (inte längre) use a comparative here, which is also true for many (but not all) languages in Western and Central Europe. Spanish (ya no), Russian (uče ne) and most of the languages in Eastern Europe and on the Iberian peninsula show a different pattern, by combining ‘already’
with negation. Since both these large areas contain languages of distinct or at least distant genetic affiliation, the similarities in the semantic organization of their systems of phrasal adverbials are largely attributable to convergence. All in all, van der Auwera (1998) has twelve properties for identifying the typical Standard Average European system, and their different subsets are particularly useful for zooming in on various convergence areas among the European languages.

5. The Greater Hindukush: A case study

In this section we attempt to outline the lexico-semantic profile of one particular ‘linguistic area’, the Greater Hindukush, focusing on the groups of lexico-semantic parallels discussed in the preceding sections. The Greater Hindukush (henceforth GHK) is the mountainous region comprising northern Pakistan, north-eastern Afghanistan and the northern-most part of Indian Kashmir. There are 40-50 languages belonging to six different genera spoken here: Indo-Aryan (the largest component both demographically and in terms of the number of individual varieties), Iranian, Nuristani, Tibeto-Burman, Turkic and the isolate Burushaski. Areality research in the Greater Hindukush is only in its early stages, and although its significance as a convergence area has been suggested by several scholars for at least a few decades (Toporov 1970; Edelman 1980; 1983, 16; Bashir 1996; 2003, 823; Tikkanen 1999: 2008; Baart 2003), only a few, primarily phonological and grammatical, features have been studied in a more systematic fashion. GHK shares a number of features with the South Asian sprachbund at large, some others with languages spoken in the adjacent regions of Central Asia to the north of the Greater Hindukush, while still others might be typical of the region itself, or of a significant subdivision of it. The region has probably never been united politically in its entirety, but has for a long time been at the periphery of a few larger contending political forces, while also maintaining cultural homogeneity and collective identity (Cacopardo and Cacopardo 2001: 251-252). Areal features in the realm of semantics in the GHK are found on multiple levels, and it is likely that higher-level convergence is an effect of prolonged and massive lower-level convergence, and that higher-level convergence in its turn functions as a catalyst for even more pervasive lower-level convergence. The following inventory should be read as an indication of semantic convergence rather than as a set of firm results from empirical study; our main focus here is to point to features bundled with one another in contiguous geographical areas or zones. Neither does it exclude the possibility of similar parallelisms in other regions of the world – a point that we will return to at the end of this section. Indeed, all of the subfields exemplified would benefit from further research.

The most visible signs of areal convergence are seen in the identical, or at least significantly similar, semantic structure of individual lexemes (or lexicalized phrases) across languages. To begin with, the shared loan vocabulary comes often from more prestigious languages (primarily Persian, Arabic, Portuguese and English). Some examples of polysemy copying or polysemy sharing across languages in the GHK (and most likely well beyond) are:

\[(11)\] ‘take’ = ‘buy’ (widespread and old in the entire Indo-Aryan family (Turner 1966, 4236)): griik in Kalasha (Trail and Cooper 1999), ghína in Palula, gi-gig in Dameli (Perder 2013, 207), lenma in Balti (Read 1934), nāa in Kamviri (Nuristani (Strand
axistə in Pashto. Interestingly, the semantically related notions of ‘sell’ and ‘give’ are normally kept apart in the same set of languages: e.g. warkawə ‘give’ vs. xartsəwə ‘sell’ (Pashto).


(13) ‘heavy’ = ‘important, honourable’: drund in Pashto, gāraka in Kalasha (Trail and Cooper 1999).

Related to that are shared metaphorical extensions, such as Kashmiri tofən ‘cold’ → ‘hostile, unkind’. There are also a few widespread metonymical shifts, such as from ‘house’ to ‘household’ and by extension to a taboo-avoiding mechanism for inquiring about someone’s wife (which is considered inappropriate in an environment where women observe purdah, i.e. seclusion) without actually using the word ‘wife’. The phenomenon as such if of course even more widespread, particularly across the Muslim world.

Some of the shared lexico-constructional patterns in the region (and, again, most likely with an even wider distribution) are listed in (14)–(16):

(14) ‘parent’ = ‘mother-father’: mor-plaə in Pashto, yeei-baabə in Palula, yii-dadi in Dameli (pc Emil Perder), nan-tai in Khowar, nua-tot in Kamviri (Strand 2013), atəama (lit. father-mother; i.e. the reverse order) in Purik (Tibeto-Burman, pc Marius Zemp).


The following examples (also known from many other, individual, languages and parts of the world, see below) show a shared semantic association that manifests itself as polysemy in some languages and as a lexico-constructional pattern in others:

(17) ‘finger’ = ‘toe’: aaŋür in Kamviri (Strand 2013), cəməṭ in Khowar, aanguṭhi in Dameli (pc Emil Perder), senmo in Balti (Read 1934), zuu-zuγu in Purik (pc Marius Zemp), ɛpiʃ in Burushaski (Willson 1999)

(18) ‘toe’ = ‘foot’s finger’: pōngo cəməṭ in Khowar, kɔs onγeʃ in Kashmiri, khuray angui in Indus Kohistani (pc Claus Peter Zoller), kangmi senmo in Balti (Read 1934), yuutise ɛpiʃ in Burushaski, paaes anguuriikii in Degano Pashai (Indo-Aryan (Strand 2013)).

Many of the GHK languages have a basic kinship term covering both ‘father’ and ‘father’s brother’ (often lexically distinct from ‘mother’s brother’). However, the latter meaning is probably becoming lexicalized in combinations with qualifying adjectives ‘big’ and ‘small’, where BIG FATHER is one’s father’s older brother and SMALL FATHER is one’s father’s younger brother (cf. ex. 19–23). Similarly, there is a widespread polysemy pattern for ‘mother’ and ‘mother’s sister’, but again with ‘big’ and ‘small’ only used for ‘mother’s sister’.
(19) áya ‘father’, (uyúm) áya ‘(his) father’s older brother (lit. big father)’ (Burushaski, Parkin 1987, 161; Willson 1999)

(20) ata ‘father’, ata tsharma ‘father’s older brother (lit. big father)’, ata tshuntse ‘father’s younger brother (lit. small father)’ (cf. momo ‘mother’s brother’) (Balti, Read 1934)

(21) to ‘father’, aalto ‘father’s oldest brother (lit. big father)’, maŋam to ‘father’s middle brother (lit. middle father)’ krāsto ‘father’s youngest brother (lit. youngest father)’ (Kati, Nuristani, Strand 2013)

(22) bāabu ‘father’, gaaḍbāabu ‘father’s older brother (lit. big father)’, lhookbhāabu ‘father’s younger brother (lit. small father)’ (cf. momo ‘mother’s brother’; pitrí ‘father’s brother (generic)’) (Palula)

(23) dadi ‘father’, jëšṭadadi ‘father’s older brother (lit. older father)’, mažumadadi ‘father’s middle brother (lit. middle father)’, sureshadadi~ucuṭadadi ‘father’s younger brother (lit. young father, small father, respectively)’ (cf. mam ‘mother’s brother’; pitrí ‘father’s brother (generic)’) (Dameli, Perder 2013, 67)

Area-specific lexicalizations that do not easily translate into European languages (or even into any language outside the region) are often found in those domains which play a prominent role in the shared Hindukush culture, shaped by a similar mountainous environment, by traditional economies, and by similar-looking world views (some of them rooted in Islam, others in similar-looking and still pervasive pre-Islamic values and beliefs, yet others in a common pan-South Asian culture). Some of the more promising domains for further research are marriage, kinship, animal husbandry and supernatural beings.

Marriage in the region (and again, as in many other, primarily non-Western, societies) is never an individual affair; when you marry, you marry an entire family, and a large number of people are involved in the arrangements. A wedding, here exemplified from Khowar (Akhunzada and Liljegren 2009, 88–97), is a whole series of events or stages (vec̣ik ‘confirmation of engagement, betrothal’; žaqdoyo ‘the party arriving from afar prior to the first day of the wedding’; nikah/sotbal ‘the official signing of the marriage contract conducted by a religious scholar in the house of the bride’; šištu ‘the procession taking the bride to the groom’s house’, yeṭmisik ‘the groom’s visiting his father-in-law three days after the wedding’; mari ‘a party given by the groom to which the relatives of the bride are invited a few days after the wedding’). Many of those stages involve a specified kind of gift-giving or financial transaction (mahar ‘bride price, as negotiated between the families of the bride and the groom’; hosti c̣ahkanli ‘engagement gifts, partly monetary, handed over by the groom’s father to the bride and her family’; maal ‘food items for the wedding supplied by the groom’; šabokaan ‘wedding clothes (for the bride and groom) supplied by the groom’s family’; panduair ‘gifts from relatives and villagers’; išpeeiri ‘food-gifts (sweets, sugar, etc.) given to the new couple at their arrival at the groom’s place’).

An example of area-specific lexicalization (although far from being limited to the GHK) from the domain of kinship is a lexeme, mostly synchronically non-analyzable, referring to a ‘co-wife, a rival wife’ in a situation where a man has more than one wife: abeendi in Dameli (Perder 2013, 69), iauri in Kamviri (Strand 2013), son in Kashmiri (p.c. Khawaja Rehman), ambax in Degano Pashai (Strand 2013).

The domain of animal husbandry has a highly specialized vocabulary describing dairy practices and dairy products. Some examples (in this case from Nuristan, in the eastern

Area-specific lexicalizations that do not easily translate into European languages (or even into any language outside the region) are often found in those domains which play a prominent role in the shared Hindukush culture, shaped by a similar mountainous environment, by traditional economies, and by similar-looking world views (some of them rooted in Islam, others in similar-looking and still pervasive pre-Islamic values and beliefs, yet others in a common pan-South Asian culture). Some of the more promising domains for further research are marriage, kinship, animal husbandry and supernatural beings.

Marriage in the region (and again, as in many other, primarily non-Western, societies) is never an individual affair; when you marry, you marry an entire family, and a large number of people are involved in the arrangements. A wedding, here exemplified from Khowar (Akhunzada and Liljegren 2009, 88–97), is a whole series of events or stages (vec̣ik ‘confirmation of engagement, betrothal’; žaqdoyo ‘the party arriving from afar prior to the first day of the wedding’; nikah/sotbal ‘the official signing of the marriage contract conducted by a religious scholar in the house of the bride’; šištu ‘the procession taking the bride to the groom’s house’, yeṭmisik ‘the groom’s visiting his father-in-law three days after the wedding’; mari ‘a party given by the groom to which the relatives of the bride are invited a few days after the wedding’). Many of those stages involve a specified kind of gift-giving or financial transaction (mahar ‘bride price, as negotiated between the families of the bride and the groom’; hosti c̣ahkanli ‘engagement gifts, partly monetary, handed over by the groom’s father to the bride and her family’; maal ‘food items for the wedding supplied by the groom’; šabokaan ‘wedding clothes (for the bride and groom) supplied by the groom’s family’; panduair ‘gifts from relatives and villagers’; išpeeiri ‘food-gifts (sweets, sugar, etc.) given to the new couple at their arrival at the groom’s place’).

An example of area-specific lexicalization (although far from being limited to the GHK) from the domain of kinship is a lexeme, mostly synchronically non-analyzable, referring to a ‘co-wife, a rival wife’ in a situation where a man has more than one wife: abeendi in Dameli (Perder 2013, 69), iauri in Kamviri (Strand 2013), son in Kashmiri (p.c. Khawaja Rehman), ambax in Degano Pashai (Strand 2013).

The domain of animal husbandry has a highly specialized vocabulary describing dairy practices and dairy products. Some examples (in this case from Nuristan, in the eastern
part of GHK) are ‘buttermilk’ (niwo in Kati, trawə in Ashkun, wašip in Waigali), ‘clarified butter’ (anõu in Kati, uc in Prasuni, amaa in Ashkun), ‘butter churn made from inflated goat skin’ (üzuu in Prasuni, ma in Ashkun, mõka in Waigali), ‘cheese whey from milk curdled with rennet’ (šipuu in Ashkun/Waigali), ‘soft rennet cheese (type 1)’ (tsila in Ashkun, kilaa in Waigali), ‘soft rennet cheese (type 2)’ (aməṣ in Ashkun, amüṣ in Waigali), the difference between the latter two only in the boiling and use of the sack (Edelberg and Jones 1979, 74–91).

Some clearly pre-Islamic beliefs are still deeply rooted in the region, quite a few of them even amalgamated with the new religion or re-interpreted though its lens. This is particularly relevant from an areal-semantic viewpoint in the identification of a number of supernatural or ‘not-quite-human’ beings that have parallels in various languages across the region (Cacopardo and Cacopardo 2001, 140–143). First and foremost, there is the category of fairies, often, but not necessarily referred to with a word related to Persian parii ‘winged; a good genius, a fairy’ (Steingass 2000): peri in Dameli, pari in Khowar, peerëni or peerai in Palula, suic or pari in Kalasha, pari in Kalasha, or parii in Palula, suic or pari in Khowar (Akhunzada and Liljegren 2009, 106) and daatik in Dameli (p.c. Emil Perder). A third category comprises a witch-like creature, who, although often described as decidedly human, is also endowed with clearly supernatural powers or characteristics – rui in Shina-varieties, goorav in Khowar, sišek or possibly ranzi in Dameli and rukiṭi (i.e. rui woman) or sišaki in Palula (also a Shina variety). According to some “modernist” accounts, she is a real, living woman who has abandoned society and gone into the forest or, alternatively, has taken residence in the ruins of an old house. Other, probably more “traditional” accounts hold that she is a spirit who appears in the form of an old woman, sometimes seen with her feet turned backwards and with long breasts. Like the ṭhaṭéeki, she drinks blood and eats people (Cacopardo and Cacopardo 2001, 140–141, 170; Akhunzada and Liljegren 2009, 107–108).

Shared formulaic expressions also deserve further research. Common greetings, farewell expressions and a number of other interactional routines (exclamations, congratulations, etc.), are to a very large extent Islamicized, based on Arabic or Persian expressions, and are therefore often also formally identical or near-identical across language boundaries. However, many of these are supplemented with language-specific expressions that nevertheless follow a rather uniform pattern, with a quick succession of polite questions and answers. An older layer of greetings seems to contain a distinct element of congratulation and well-wishing. Before the conversion to Islam (and probably a long time even after), Palula speakers greeted each other with an initial whaatee ‘Didn’t you come down?’ if the other person came down from a higher elevation, or ukhaatee ‘Didn’t you come up?’ if he happened to have ascended from a place further down the valley or the slope; cf., Dameli ageepi ‘Have you come?’ (p.c. Emil Perder). Although framed as questions, these formulae were really expressions of congratulation to the other person having safely reached his destination. This is still reflected in the expression tu xeera sangi phediluee ‘Did you arrive safely (lit. with
happiness)?’ A simple kíi biáanu ‘Where are you going?’ or kíi gúum de ‘Where did you go?’ is still frequently heard, and is as much a greeting in itself as an actual inquiry for information. The polite, traditional, exchange between Pashto speakers meeting each other contains paired expressions of (more or less) set well-wishes, such as po xer raayle ‘May you come in peace!’, triggering the reply po xer use ‘May you live in peace!’ or stefay moše ‘May you not be tired!’ answered with xwaar moše ‘May you not be poor!’.

Similar to the observations about West African leave-taking in Section 3, the guest in the GHK also explicitly asks for permission to leave, not seldom by using the Pashto word ijaazat ‘permission’:

\[ \text{ijaazat day} \]

Another level of convergence has to do with shared organization of entire semantic domains, as for instance within calendrical expressions that include distinct lexical items for ‘three days ago’, ‘the day before yesterday’, ‘yesterday’, ‘today’, ‘tomorrow’, ‘the day after tomorrow’ and ‘three days from now’.

Table 1. Calendrical expressions in Kamviri (Strand 2013), Burushaski (Berger 1998, 103), Dameli (Morgenstierne 1942, 137–178 and p.c. Emil Perder) and Balti (Read 1934, 30–31).

<table>
<thead>
<tr>
<th>Kamviri</th>
<th>Burushaski</th>
<th>Dameli</th>
<th>Balti</th>
</tr>
</thead>
<tbody>
<tr>
<td>nuc̖t</td>
<td>ucooʃ cvooʃ diyoo</td>
<td>danma jaq</td>
<td>‘three days ago’</td>
</tr>
<tr>
<td>Nutri</td>
<td>yáarbulto</td>
<td>itrii</td>
<td>karchagla ‘the day before yesterday’</td>
</tr>
<tr>
<td>dos</td>
<td>sabaur</td>
<td>doos</td>
<td>dore ‘yesterday’</td>
</tr>
<tr>
<td>strák gaaʃar</td>
<td>khúulto</td>
<td>mudya</td>
<td>diring ‘today’</td>
</tr>
<tr>
<td>daalə</td>
<td>jìmale</td>
<td>bera</td>
<td>bela, haske ‘tomorrow’</td>
</tr>
<tr>
<td>aatri</td>
<td>hipulto</td>
<td>traida</td>
<td>snangla ‘the day after tomorrow’</td>
</tr>
<tr>
<td>aac̖u</td>
<td>máalto</td>
<td>cvooʃ cvooʃ ki</td>
<td>rzesla ‘three days hence’</td>
</tr>
</tbody>
</table>

Another example of organizational patterns, this one affecting an entire part of speech, is the structure of numeral systems. Virtually all languages in the region, regardless of genealogical affiliation, have a vigesimal system, or numeral systems that are primarily vigesimal, i.e. based on ‘twenty’, although they differ in the details and the extent to which more peripheral decimal features have been added (often affecting numerals above 100). Tibeto-Burman varieties related to Balti and those spoken further to the south and southeast are entirely decimal, and so are most of the major languages surrounding the region, such as Hindi-Urdu, Persian, Turkic, etc. Massive Urdu influence in today’s Pakistan leads to a gradual replacement of numerals above ten in the local languages by Urdu numerals.

Table 2. Numerals in Burushaski (isol.) (Berger 1998, 100–101), Balti (Tibeto-Burman) (Read 1934, 24–25; Bielmeier 1985) and Ashkun (Nuristani) (Strand 2013).

<table>
<thead>
<tr>
<th>Burushaski</th>
<th>Balti</th>
<th>Ashkun</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>tóorumo</td>
<td>phću</td>
</tr>
<tr>
<td>20</td>
<td>áltar</td>
<td>šiša</td>
</tr>
<tr>
<td>30</td>
<td>áltar-tóorumo</td>
<td>20 &amp; 10</td>
</tr>
<tr>
<td>40</td>
<td>altó-áltar</td>
<td>2 x 20</td>
</tr>
<tr>
<td>50</td>
<td>altó-áltar-tóorumo</td>
<td>(2 x 20) &amp; 10</td>
</tr>
<tr>
<td>60</td>
<td>iski-áltar</td>
<td>3 x 20</td>
</tr>
</tbody>
</table>
Areality shared patterns have a bearing on the overall organization of the lexicon. Dimensions in time, space and discourse make use of similar multi-dimensional and multi-degree deictic contrasts in many GHK languages (some exemplified in the simplified paradigm in Table 3).

Table 3. Sets of pro-forms in GHK languages. Morphemes entailing emphasis (either in combination with a non-emphatic pro-form or in contrast with it) are indicated within parentheses (Koul 2003; Bashir 2003; 2009; Berger 1998; Strand 2013; Zemp 2013)

<table>
<thead>
<tr>
<th>Proximate (visible)</th>
<th>Remote visible</th>
<th>Remote non-visible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kashmiri</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yi(hoy)</td>
<td>hu(hay)</td>
<td>su(y)</td>
</tr>
<tr>
<td>Palula</td>
<td>(ee)nu(fo)</td>
<td>(ee)so</td>
</tr>
<tr>
<td>Kalasha</td>
<td>(š)iya(šasa)</td>
<td>(ša)xe</td>
</tr>
<tr>
<td>Burushaski</td>
<td>in(é)khin(é)</td>
<td>khin(é)</td>
</tr>
<tr>
<td>Wakhi</td>
<td>(hai)em(hai)et</td>
<td>(hai)ya</td>
</tr>
<tr>
<td>Kamviri</td>
<td>(o)iná(ojá)</td>
<td>(o)azaská</td>
</tr>
<tr>
<td>Purik</td>
<td>dyu ao–(odoo)</td>
<td>eu</td>
</tr>
</tbody>
</table>

It seems that lexical material for such extensive (and to some extent symmetrical, and not seldom rhyming) sets of pro-forms has been recruited from various language-internal sources. A slightly more extended set is displayed for Kashmiri (Table 4); this set is, however, in no way unique to this particular language but shows a common structure found in most of the languages of the region, regardless of their genealogical classification.

Table 4. Extended (but incomplete) set of Kashmiri pro-forms (Koul 2003)

<table>
<thead>
<tr>
<th>Proximate</th>
<th>Remote within sight</th>
<th>Remote out of sight</th>
<th>Interrogative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>yt</td>
<td>hu</td>
<td>su</td>
</tr>
<tr>
<td>Location</td>
<td>yeti</td>
<td>hoti</td>
<td>tati</td>
</tr>
<tr>
<td>Direction</td>
<td>yapī’</td>
<td>bōpī’</td>
<td>tapī’</td>
</tr>
<tr>
<td>Manner</td>
<td>yithi kin’</td>
<td>hubhi kin’</td>
<td>tithi kin’</td>
</tr>
<tr>
<td>Quantity</td>
<td>yānāh</td>
<td>hūnāh</td>
<td>t’ānāh</td>
</tr>
</tbody>
</table>

The distinctions made in such sets (although not all of them are shared by all languages in the region) include emphatic vs. non-emphatic, proximate vs. remote, visible vs. invisible, interrogative vs. demonstrative vs. relative, nominal vs. adnominal. Some of the deictic values, usually the remote-within-sight ones, are further specified for distance, direction and angle, especially relevant in a mountainous region as this one (Bashir 2003); paár ajóo ‘over there where I point’, paár adí ‘right over there’, paár asdí ‘right
over there somewhere’, *pér adí* ‘over there (near, known but invisible)’, *pér asdí* ‘over there (out of sight)’ in Kohistani Shina (Schmidt and Kohistani 2008, 97–98). Many of those, e.g. in Indo-Aryan Palula and Gawri (Baart 1999: 175–196), in Iranian Wakhi (Bashir 2009: 230-232), in the isolate Burushaski (Berger 1998: 78-97), as well as in Tibeto-Burman Purik (Zemp 2013: 235-242) can be used adverbially (‘that place down there’), pronominally (‘she down there’) as well as adnominally (‘that x down there’), see Fig. 3. This might also be reflected in verbal specifications such as distinct lexical items for ‘come up’, ‘come down’, ‘come out’; ‘take up’, ‘take down’. The same, or a similar phenomenon, referred to as “vertical case” and “vertical verbs” (Noonan 2003), is characteristic of a group of languages spoken in the Nepal Himalayas, an adjacent mountainous region located to the east of the GHK.

Figure 3. Palula spatial pro-forms.

An example that is difficult to place in any of the hitherto mentioned categories, is the prevalence of co-lexicalized intensifiers: Burushaski (Berger 1998: 226–227): *qhal-matúm* ‘pitch black’; Gilgiti Shina (Carla Radloff, pc): *khutún šaróo* ‘full autumn’, *brang sang* ‘very bright’; Kalasha (Trail and Cooper 1999): *ṭam ḫḷira* ‘perfectly round’; Kashmiri (Khawaja Rehman, p.c.): *sia kriohon/kriohon sia* ‘pitch black’. Curiously, for Indo-Aryan Palula, such intensifiers are almost exclusively made up of a single closed syllable, CVC(C). Characteristically, the intensifier is a unique lexical unit, compatible with a single head, or in some cases with a limited set of those.

<table>
<thead>
<tr>
<th>Palula co-lexicalized intensifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>phaʂ̪</em> <em>pɑːʂ̪aːru</em></td>
</tr>
<tr>
<td><em>kham kɪʂ̪ɪʈʊ</em></td>
</tr>
<tr>
<td><em>cɑw lhóilu</em></td>
</tr>
</tbody>
</table>
As indicated in the beginning of this section, the inventory presented here should be taken with a certain degree of caution. For instance, neither of the lexical parallels ‘toe’ = ‘(foot’s) finger’, ‘take’ = ‘buy’ or ‘parent’ = ‘mother-father’, is unique to the Hindukush region. ‘Parent’ = ‘mother-father’ or ‘father-mother’ is mentioned by Wälchli (2015, 2005) as the most reliable single-word indicator for co-compounding (dwandwa-compounds) in Eurasia, in the sense that if a Eurasian language will have co-compounding at all, it will necessarily express ‘parent’ as ‘mother-father’ or ‘father-mother’. The sequence of the components differ areally in various unpredictable ways (cf. Map 6.2 in Wälchli 2005: 218 for the occurrence of this pattern in the languages of Eurasia). ‘Toe’ = ‘(foot’s) finger’ is found in such familiar languages as Russian (palec (na noge)), or Italian (dito (del piede)), but is also well attested elsewhere. The on-line database CLIS (http://clics.lingpy.org/direct.php, accessed on December 30th 2014) lists 51 occurrences of ‘toe’/’finger’ co-lexifications and 38 ‘take’ = ‘buy’ colexifications among the 221 languages in the database, both in Eurasia and in other parts of the world. Conspicuously most of the examples for ‘take’ = ‘buy’ come from the Northern Caucasus, but these languages are also heavily represented in the database.

Vigesimal systems, or numeral systems that are primarily vigesimal, are found in various parts of the world, but are also being replaced by decimal systems. In Comrie’s (2013) 196-language sample, Asia is indeed populated by decimal systems, with Burushaski as one of the very rare exceptions (none of the other Hindukush languages are included in the sample).

An evaluation of the systems of calendrical expressions as an isogloss is not straightforward. On the one hand, the systems with at least three distinctions in both the future and the past are not very rare and occur in 44 languages (28%) in Tent’s (1988) global 157-language sample. On the other hand, systems based on monomorphemic (non-transparent) expressions are only found in 16 languages (10%) within Tent’s sample, including Hindi. Some of these, however, have fewer semantic distinctions than the Hindukush languages. Interestingly, the calendrical expressions in the other Indo-Iranian language in Tent’s study, Farsi, involve transparent polymorphemic structures.

Evaluation of most of the other properties, suggested as areal indicators for the Hindukush languages, is currently not possible due to lack of reliable data and would require further systematic research, both within the region and outside of it.

6. Causation and mechanisms: inheritance, diffusion, shared environment or independent innovation?

A central issue in all research concerned with cross-linguistic comparison is to what extent an observed cross-linguistic similarity may be explained by accident, inheritance, contacts, universal tendencies, or a combination these factors. Certainly, when many languages belonging to different families within a more or less well-defined region share a property that is very rare in other parts of the world, language contact suggests itself as a likely explanation. However, as is evident from our discussion of the potential areal Hindukush traits, such an exercise is normally much more complicated than expected, in particular when most (or all) languages in the area are closely related to each other, which is in fact a frequent complication in presumed linguistics areas (most of the Ethio-
Eritrean languages are Afro-Asiatic, most of the languages in the Volta basin are Niger-Congo or even Kwa, etc.). François (2010) exhibits the intricacies embedded in such an enterprise for the Oceanic languages of Northern Vanuatu, while Epps (2013) demonstrates how a careful and well-informed argumentation may be used for evaluating the relative probability of the different factors behind the lexico-constructional similarities, in this case the morphologically complex numeral expressions, in the Amazonian languages of the Vaupes region (see Epps and Michael, this volume).

Like areal linguistics in general, a quest for areal lexico-semantic parallels faces various potential fallacies, which are often rooted in the researchers’ natural bias to notice and overestimate the importance of patterns that strike them as unusual. However, most contact-induced change is not particularly spectacular, and most isoglosses are probably neither unique to an area nor skewed in their distribution so much that they will ‘betray’ the area in a large-scale sample. Isoglosses rooted in language contacts will, thus, often ‘stand out’ only within a particular area but will not necessarily be noticeable from a large-scale typological perspective. To identify them as areal properties requires, among other things, that they have been systematically tested both across the languages of the region and across the languages outside of it, ideally across a large sample of the world’s languages (with a combination of micro- and macrotypological methods, cf. Koptjevskaja-Tamm 2011). A systematic large-scale investigation gives an important evaluation of how frequent or unique a particular property is among the languages of the world and provides indications of the areas in which might be found. By zooming in on a particular area with the help of a fine-grained sample it is possible to further estimate to what extent this property is systematically represented in the area rather than, say, appearing here and there.

In what follows we will illustrate the complexity of these issues with a few examples of relevant recent and on-going research related to the evaluation of particular lexico-semantic parallels as areal indicators.

Cross-linguistic research on semantic associations between verbs of perception and cognition provides an excellent example of how data based on different language samples may affect our understanding of what is universal, genetic and areal in a particular lexico-semantic field. In her influential study, mainly based on the Indo-European languages, Sweetser (1990) suggests that the metaphorical extension from vision to knowledge (‘know’, ‘think’) may “be fairly common crossculturally, if not universal” (Sweetser 1990: 45), while the cognitive extension of hearing may be limited to ‘understand’. These putatively universal connections are questioned by Evans and Wilkins (2000) in their study of about 60 Australian aboriginal languages, where the most frequent semantic extension is from hearing to cognition, and in particular, to knowledge (‘know’, ‘think’, ‘remember’). Vision, on the contrary, mainly gives rise to social interaction readings such as desire and sexual attraction, aggression and negative social interactions, etc. Evans and Wilkins provide numerous arguments for showing that the connection between perception and cognition in the Australian aboriginal languages is rooted in various social and cultural practices of their speakers, without, however, ruling out that some of these may find parallels in other parts of the world.

Vanhove’s (2008) typological follow-up demonstrates that Evans and Wilkins’s findings hold more generally. For instance, all the 25 languages of her genetically and geographically diverse sample have semantic extensions from hearing to cognition – to internal reception (‘heed’, ‘obey’), often to intellectual perceptions and cognition (‘understand’, ‘learn’, ‘know’), and, more rarely, to ‘think’ and ‘remember’. The shift to
cognitive meanings from vision is far less common, and most of the languages concerned have the same cognitive extensions for both ‘hear’ and ‘see’.

Summarizing, the semantic associations that were believed to be (almost) universal have later turned out to be mainly genetic, while those that were believed to be genetic / areal have later proved to be much more widespread.

The most significant large-scale investigations of cross-linguistically recurrent semantic associations are reported in Urban (2012, also 2009 and 2010). Urban’s goal is to detect the semantic and structural patterns among the terms for 160 nominal concepts contained in a global sample of 109 languages. To give an idea of what is at stake, consider the concepts ‘sun’, ‘moon’ and ‘day’. These are expressed by three different morphologically simple (indivisible) terms in English. Some languages, however, associate ‘sun’ and ‘moon’ in various ways, either by colexifying them within the same lexeme or by deriving the word for ‘moon’ from ‘sun’ in patterns like ‘night sun’. Many languages associate ‘sun’ and ‘day’ by either using the same word for both (sometimes colexifying all the three concepts) or by deriving the term for ‘sun’ from ‘day’. Languages differ therefore as to whether they express comparable concepts by unanalyzable terms and as to which concepts are semantically associated with each other, either by polysemy or by various lexico-constructional patterns. There are both remarkable cross-linguistic dissimilarities and striking cross-linguistically recurrent patterns, that may be accounted for by a complex interplay among universal tendencies and genetic vs. areal relations among the involved languages. A balanced global sample of this scale is, however, often too sparse a net for catching significant areal lexico-semantic parallels and should be complemented by a more fine-grained sample covering many language varieties within the relevant area.

To continue with the ‘sun’/’day’/’moon’ example, the pattern ‘sun’ = ‘eye of day’ is very rare in the balanced global 214-language sample, but shows a remarkably high presence in the Austroasiatic, Tai-Kadai and Austronesian languages of Southeast Asia and Oceania, as demonstrated in the additional augmented 154-language sample for this part of the world (Urban 2012, 2010). Urban himself suggests two possible contact scenarios for this situation, while Gil (2015) mentions this skewed distribution as one of the areal properties of what he calls the Mekong-Mamberamo linguistic area, consisting of Mainland Southeast Asia, the Nusantara archipelago in Indonesia and parts of New Guinea. However, to complicate the issue, the reliability of Urban’s data and his explanations are still contested by Blust (2011), who makes a case for the universality of this pattern by adding parallels from other languages.

The cross-linguistic distribution of the ‘sun’ = ‘moon’ colexification presents another challenge for historical linguistics. As Urban (2012, 2009) shows, the pattern is very unevenly distributed among the languages of the world. It is richly represented in the Americas, but is also found in the Paleosiberian languages (the indigenous languages of north-eastern Eurasia) and among the languages of New Guinea. Urban further argues that this distribution strongly correlates with the large Circum-Pacific area, suggested by Nichols and her collaborators (e.g., Nichols 1995, Bickel and Nichols 2006). This very old area is believed to have resulted from the large-scale population movement out of coastal Southeast Asia during the last glaciation. At such a great time depth “historical” relatedness among languages may be detected by means of careful statistical analysis of the observed cross-linguistic patterns (Nichols and Peterson 1996), but it is not possible to further distinguish between genetic and areal relatedness.
Our next example concerns formulaic expressions used for particular pragmatic functions. Although their chances to be shared by languages as a result of independent innovation is in general lower than for many other expressions, this possibility should not be neglected either. In his inspiring paper “Areal semantics — is there such a thing?” Matisoff (2004: 369) mentions typical for Southeast Asia greetings such as “Where are you going?” and “Have you eaten yet?” (cf. the examples in 25 below).

(25) Conventionalized direction “where”-greetings (Gil 2015)

a. Vietnamese (Austro-Asiatic, Mon-Khmer)
   Đi đâu?
   go where
b. Jakarta Indonesian (Austronesian, Malayo-Polynesian)
   Mau ke mana?
   want to where
   ‘How are you?’ (lit. ‘Where are you going?’)

Such greetings do not presuppose any precise reply, like “I am going to my mother”, in the same way as “How do you do?” does not presuppose any real report on the interlocutor’s life situation. Somewhat surprisingly, conventionalized greetings with “directional where” (i.e., Where are you going?” or “Where are you coming from?”) are much more spread than Matisoff suggests: they occur in 65% of the 363 languages in Gil’s (2015) global sample, including the Greater Hindukush languages, as shown in Section 5. It remains unclear, though, whether all these languages use such expressions to the same extent. As Gil comments, ‘where’-greetings are often restricted to outdoor contexts, but in some cases can also be heard indoors. Obviously, greeting someone outdoors often presupposes that at least one of the interlocutors is on their way from one place to another. Under such circumstances the ‘where’-questions have a reasonable source, and different languages may have independently developed their conventionalized greetings on the basis of this universally recurrent situation. However, as Gil (2015) argues, directional ‘where’-greetings are not equally well spread across the world: in some areas they are used just in few languages, whereas in others they are the rule. The latter is the case in the Mekong-Mamberamo area (cf. above), but this isogloss covers a larger area, extending, among other things, into the Himalayas in South Asia.

Our final reflections concern the issue of whether and/or to what extent languages can in general converge in their lexical systems used for a particular semantic domain. The phasal adverbs and deictic verbs in the European languages (Section 4), or the calendrical expressions in the Greater Hindukush region (Section 5) are excellent examples of convergence in the conceptual organization of entire lexical semantic systems. However, in all these cases the lexical systems are relatively closed and/or organized according to comparatively simple oppositions, bordering on grammatical phenomena. The question is to what extent convergence may involve more typical lexical systems, corresponding to complex semantic domains which in principle allow for greater freedom in categorization. The evidence so far is very sketchy and partly contradictory.

There are, in fact, some examples of shared organization in lexical domains that have been attributed to areal forces. Particularly inspiring for these suggestions is Viberg’s (1984) widely quoted lexico-typological study of perception verbs. Güldemann and Fehn (this volume) show that the conflation of hearing, touch, taste and smell in one verb as opposed to sight is found in languages of all the different families in Central
Kalahari, while the languages in other parts of the Kalahari basin show more elaborated systems. Also Treis (2010) provisionally suggests that some particular confluations of perception verbs (such as one verb for ‘hear’ and ‘listen’, or the use of the passive form of the hearing verbs for TOUCH and TASTE) and taste adjectives may be added to the lexico-semantic parallels within the Ethio-Eritrean area (cf. the discussion of Hayard 1991 and 1999 in Section 2). Enfield (2011: 31) asks whether language contact can “permeate the conceptual organization of lexical semantic domains” in relation to the domain of TASTE and FLAVOUR in two Southeast Asian languages – Lao (a south-west Tai-Kadai language spoken by 25 million people in Laos, Thailand and Cambodia), and Kri (an Austroasiatic language with 300 speakers in upland central Laos). His answer is positive, in that the two languages turn out to divide the domain in similar ways, i.e. organize it into similar categories. Since Tai languages were adopted by pre-existing Austroasiatic-speaking populations, Enfield (2011: 36) hypothesizes that perhaps “the descendants of modern-day Lao speakers once learned an earlier form of Lao as their second language, and in shifting toward using it as their first language introduced semantic distinctions that were encoded in their original language”.

However, while Enfield’s results are striking, they are also somewhat questionable, given the knowledge accumulated in cross-linguistic research into the lexicon. In a nutshell, the lexicon is subject to relatively rapid changes, which often lead to restructuring of lexical semantic domains. As a result, even closely related languages often display significant typological differences in how they categorize and organize similar cognitive domains, often by means of cognates (cf. Majid et al. 2007 for the CUT and BREAK domain in Germanic and Koptjevskaja-Tamm et al. 2011 for the AQUAMOTION in Germanic and Slavic). The same is also true for lexical systems in closely related languages that have been spoken in close contact with each other for any considerable time. For instance, the Timor-Alor-Pantar languages Abui and Kamang (Schapper 2015) or the Kwa languages Ewe, Likpe and Stilte (Ameka 2015, Agbetsoamedo and Di Garbo 2015) show significant dissimilarities in their categorization and lexicalization of the TEMPERATURE domain. Against this background it is hardly plausible that languages would keep the same lexical distinctions for such a long time period as is invoked by Enfield.

There may, however, be a “milder” version of lexical convergence having to do with the general level of elaboration within a particular lexical system, rather than with the distinctions themselves. For instance, according to Matisoff (2004: 366), the Southeast Asian lexico-semantic areal features include a rich lexicon of verbs of manipulation within such domains as CARRYING or CUTTING. And at least for CARRYING this is confirmed by Wälchli’s (2009: 35–39) study of ‘carry’ verbs in the parallel corpus of the Mark New Testament in 100 languages, where South-East Asia and Oceania show the highest presence of body-specific ‘carry’-verbs. These are largely absent from some other areas (such as Europe) or are sometimes restricted to ‘carrying on head’, as, for instance, in a number of African languages.

7. Final words

Lexical phenomena have of course a long standing record in research on language contact and linguistic areas. However, the recent developments in areal linguistics and areal typology have, with a few exceptions, mainly concerned grammatical phenomena. This is
not at all surprising given the central place of this research in modern linguistics of all
denominations, including typology, where the rapidly developing field of areal typology
has encouraged and facilitated serious research on the relative role of universal, genetic
and areal factors for many grammatical (and phonetic phenomena). Also loanwords have
been studied from a more systematic cross-linguistic perspective, where the core issue
has been the varying borrowability of various words, seen as belonging to different parts
of speech and/or coming from different semantic domains (cf. Haspelmath and Tadmor
eds. 2009, Wohlgemuth 2009). The interesting research angles here, as elsewhere in
research on contact phenomena and in (areal-)typological research (cf. Koptjevskaja-
Tamm 2011), are (i) possible outcomes of language contact in the realm of the lexicon
and (ii) the possibility of using lexical phenomena for reconstructing contact.

As we hope to have shown in this chapter, lexical semantics in language contact and
diffusion of lexico-semantic phenomena across language boundaries in a geographic area
has a great potential for historical and areal linguistics, but is still awaiting systematic
research. This is partly related to the relatively limited cross-linguistic research on lexical
issues in general, which may impede evaluation of particular lexico-semantic parallels as
areal indicators and obstruct informed attempts to find reasonable explanations for their
origin.

The following two enterprises directed at unveiling cross-linguistically recurrent
patterns of polysemy and lexico-constructional patterns deserve special mention.

1) The catalogue of semantic shifts in the languages of the world (http://semshifts.iling-
ran.ru/) at the Institute of Linguistics, Russian Academy of Sciences in Moscow is a
searchable computer database (not yet fully implemented online) that currently
contains more than 3000 semantic shifts found in 319 languages (Zalizniak 2008,
Zalizniak et al. 2011).

2) CLICS: Database of Cross-Linguistic Colexifications (List et al.,
http://clics.lingpy.org/main.php) is an online database of colexifications in 221
languages. It is based on four different freely available online resources and contains
16 239 different links among the 1280 concepts it operates with (all in all 45 667
cases of colexification).

The information accumulated in these databases provides a valuable cross-linguistic
complement to many of the lexico-semantic parallels that have been mentioned in
connection with different areas.

The good news is that lexical typology is currently on the rise (cf. Koptjevskaja-Tamm
et al. forthcoming for a recent review). We are therefore looking forward towards more
cross-linguistic research on the categorization of lexical semantic domains, polysemy
patterns, semantic associations and lexico-constructional patterns, complemented by
detailed case studies of these phenomena in languages in various contact situations. This
knowledge is essential for gaining a better understanding of what happens with semantics
in language contact.

References


Gast, Volker and Johan van der Auwera. 2012. What is contact-induced grammaticalization? Evidence from Mayan and Mixe-Zoquean languages, In Björn


Majid, Asifa, Marianne Gullberg, Miriam van Staden and Melissa Bowerman. 2007. How similar are semantic categories in closely related languages? A comparison of cutting and breaking in four Germanic languages. In Asifa Majid and Melissa


Part II: Case studies for areal linguistics
8 The Germanic languages and areal linguistics

Johan van der Auwera and Daniël Van Olmen

1 Introduction

The Germanic languages have most likely been ‘areal’ since prehistoric times. When Germanic split off from Indo-European, it was probably in part due to contact with non-Indo-European languages. And if we time travel directly to the present, we see that the most successful Germanic language, with success defined in terms of global societal relevance and the number of first or second language speakers, i.e. English, is also very ‘areal’ (Hickey ed., 2012). For one thing, English is very Romance, arguably also partly Celtic and, though it is a West Germanic language, it also North Germanic to an appreciable extent. And for another thing, due to its worldwide geographical spread, English is now giving rise to a large number of contact varieties. English is indeed areally unique, but the stories of the other Germanic languages have important areal dimensions, too, with e.g. Low German shaping Continental North Germanic, Danish splitting up Norwegian or Dutch giving way to Afrikaans. The current chapter sketches the research and issues in the areal linguistics of the Germanic languages. In section 2, some general notions are introduced. Section 3 deals with areality within Germanic, section 4 with areality involving Germanic and non-Germanic. Section 5 is the conclusion.

Note that the current chapter overlaps with several others, viz. that by Stolz and Levkovych on the phonologies of languages in Europe, that by Kortmann and Schröter on Varieties of English, that by Hickey on The British Isles and that by Mesthrie on South Africa. The volume-wide presence of English is partially a result of the areal success of the language, but not entirely. English is indeed the best studied language in the world but that does not mean that its areal linguistics was attended to sufficiently. In fact, for a long time, it was not. But recent years have seen a drastic change (e.g. Miller 2012; Hickey ed., 2012; Schreier and Hundt eds, 2013). Though this chapter cannot focus on English, the language cannot be absent either, for we may not always agree with our fellow linguists. We will also take special care to compare the areal linguistics of English with that of the other Germanic languages.

2 Concepts, dimensions and restrictions

Firstly, to describe the dimensions of contact change, various terms and concepts are at our disposal. The first is ‘areality’. We use the term in a rather strict but also standard way, i.e. to refer to similarities between languages used in the same or an adjacent geographical area which are due not (only) to chance, universal formal/functional tendencies or genetic affiliation but (at least partially) to contact. The more and the deeper the languages resemble each other, the more the similarity concerns grammar rather than the lexicon. The more the similarities are typologically unusual and the more

---

1 Thanks are due to Markku Filppula (University of Eastern Finland), Hartmut Haberland (Roskilde University) and Axel Holvoet (Universities of Vilnius and Warsaw).
languages the contact involves, the more interesting the phenomena are and the more linguists have been inclined to refer to the languages as a Sprachbund (Trubetzkoy 1930) or – the standard rendering in English – ‘linguistic area’. Of the linguistic areas discussed in the current literature, three are relevant for Germanic, viz. ‘Standard Average European’ (van der Auwera 1998b, 2011; Haspelmath 1998, 2001, Heine and Kuteva 2006), ‘Circum-Baltic’ (Koptjevskaja-Tamm and Wälchli 2001, Wälchli 2011), and ‘British Isles’ (Tristram 1999, Filppula, Klemola and Paulasto 2008, see also Hickey, this volume). These areas are old and the first two are very large, loose and multi-causal. In what follows we will refer to these linguistic areas, but the areality of Germanic languages is by no means confined to these or to the periods in which they arose. Areality is also found in Germanic linguistic enclaves (Sprachinseln), i.e., relatively small areas in which a Germanic vernacular – (nearly) always a non-standard variety – is surrounded by other languages, or in border areas, in which a Germanic language – again (nearly) always a non-standard variety – is adjacent to another language. For both situations it would be justified to bring in a notion of ‘linguistic area’, but it is not clear how much that would advance our knowledge (cf. Matras 2009: 265-277). Note that these two situations feature non-standard languages and it is increasingly stressed that areality studies should indeed focus on such varieties (e.g. Wälchli 2011: 339, Murelli and Kortmann 2011). This is not in conflict with the fact that the current chapter often discusses the standard varieties, the reason being that their areal properties arose in the period before they were standardized.

Secondly, the traditional notions of ‘superstrate’, ‘substrate’ and ‘adstrate’ (coming from Romance linguistics, Kontzi 1982: 4, 9-10) remain useful. They reflect differences of prestige, with ‘super’ pointing to more prestige, ‘sub’ with less prestige and ‘ad’ with equal prestige. The distinction also relates to registers, at least for the notions of superstrate and substrate, with superstrate influence being associated with the written and more formal registers and substrate with the spoken and more informal registers.

Thirdly, we will use ‘borrowing’ vs ‘shift’, and distinguish between them along the lines of Thomason and Kaufman (1988) and Van Coetsem (1988). On the level of the languages involved, these processes are alike: one language, the target language, changes through interference from another language, the source language. On the level of the speakers involved the processes are different, however. In borrowing the target language speakers are more active: they take something from the source language, either a relatively concrete linguistic entity like a sound or a morpheme or something more abstract like a word order pattern – to put it in the terms of Matras and Sakel (2007),

2 With this correlative statement (the more ... the more) we imply that the notion of linguistic area is a fuzzy concept (cf. Campbell 2006 and Stolz 2006).
3 Koptjevskaja-Tamm and Wälchli (2001) and Wälchli (2011) suggest that the concept of ‘linguistic area’ is not appropriate for the Circum-Baltic languages and propose ‘contact superposition zone’ instead, rather similar to Weinreich’s (1958b: 378-379) proposal of ‘convergence area’. We will gloss over these largely terminological proposals here (cf. Östman 2011: 374, van der Auwera 2001: 297-298).
4 They are not the only ones, see e.g. Hickey (1999) for considering Ireland as a linguistic area or Wehr (2001) for a ‘West-Atlantic’ linguistic area, comprising Celtic, French and Portuguese.
5 Even though there is a large literature on linguistic enclave and border situations, it generally focuses on sociolinguistic aspects and to the extent that it deals with language, it tends to focus on the lexicon (see the point below about the levels of contact influence).
either ‘MAT’ (‘matter’) or ‘PAT’ (‘pattern’). In shift, however, the source language speakers are the more active ones, as they learn the target language. They can then bring in features from the source language, but the very fact of acquisition, typically imperfect, may also give the target language new properties, which themselves are not found in the source language and which typically make the target language in some respects either more simple or more complex. Note that a language shift scenario can, of course, lead to the disappearance of the source language. In that case, one ends up not with a linguistic area but with a contact-influenced new language. Of course, if the shifting language is documented and/or a closely related language is spoken in the same region, we may still observe areality: similarities of the target language with the disappearing source language or a language related to the source language. Note also that when shift itself simplifies or complexifies a target source, there can be no areal effect, for the simple or complex properties of the target source are not part of the source language.

Characterizing a change in terms of borrowing vs. shift or superstratal vs. substratal vs. adstratal is, of course, a simplification. The source language may be partly adstratal and partly superstratal, and the contact change may partly involve borrowing and partly shift. Also, as already adumbrated with the definition of the notion of ‘areality’, a change may be partially contact-based and partially ‘something else’.

A few more general considerations are in order. Thus it is important to realize that there is a geographical dimension in the study of language which this chapter will not deal with. The Germanic family was once a dialect continuum stretching over an ‘area’ and this geographical dimension is reflected in the branching of the family into North, West and East Germanic. When the language family broke up into different languages, in part due to standardization, the languages that developed out of erstwhile neighboring dialects may still betray common features, either in the standard languages or in their dialects. A good illustration are the ‘Ingvaeonic’\textsuperscript{6} features of Dutch, Frisian and English. They include the absence of a nasal in English \textit{five}, Dutch \textit{vijf} and West Frisian \textit{fiif} and the presence of one in German \textit{fünf} and the existence of \textit{h}-pronouns such as English \textit{he}, Dutch \textit{hij} and Frisian \textit{hy} versus German \textit{er}. These kinds of commonalities, reflecting erstwhile – or for that matter, current – dialect continua, remain outside of the scope of this chapter. This does not mean, however, that when two languages are closely related and spoken next to each other, it is always easy to distinguish between similarities due to contact and similarities due to the fact they once formed a dialect chain.

Another kind of areality that we will largely leave undiscussed is the fact that a language may have acquired an international status through culture or religion and have influenced languages within that sphere. What we have in mind is Latin, later also French and German, mostly in Europe, and now English, not just in Europe, but in the entire world.

A further point concerns the number of languages that are involved in the contact situation. The paragraph clarifying the distinction between borrowing and shift was phrased in terms of ‘the’ target language and ‘the’ source language. It is correct that many contact scenarios figure just two languages. But this is a simplification: in any one area more than two languages could be spoken and influence each other and the speakers involved may be, to some extent, multilingual rather than bilingual. Thus the East Frisian

\textsuperscript{6} The term ‘Ingvaeonic’ is typical of Dutch and German linguistics and refers to the division of the Germanic tribes into Ingvaeones, Istvaeones and Erminones in Tacitus’ \textit{Germania}. One also finds the term ‘North Sea Germanic’ (Nielsen 1985).
contact situation involves East Frisian and both Low German and High German and for North Frisian the speakers closest to the Danish border may have Danish as the fourth language (Århammar 2007). Nevertheless, for any one item of structure that switches between languages, the transfer probably always just happens between two languages; it is not or rarely the case that during the transfer part of that item or structure comes from one language and another part from a third one.

Also, in the research reflected in this chapter, there is a strong focus on the target language. This is understandable. In the case of borrowing, it is only the target language that changes and in the case of language shift, the source language may disappear (at least the variant that is involved in the shift scenario) and leave few or no traces. But one must be aware of simplification here too. A transfer could be reciprocal, with e.g. German borrowing Danish and Danish borrowing German along the Danish German language border. In that case both Danish and German are target languages. For some such pairs, both scenarios have been studied, but often independently. In the case of language shift – as already remarked in the above – the disappearance of the source language by no means implies that its transitional stage is not documented. Thus the Norman French that influenced English during the Middle English period did indeed disappear – at least in England – but not without leaving an enormous literature (e.g. Wogan-Browne 2009).

A last point but one is this: areal and contact linguistics is often concerned with what could be called the linguistic level(s) at which matter or patterns converge between languages, and, very specifically, whether there is a hierarchy of ‘affectability’ or ‘borrowability’ (see the introduction to this volume, Matras 2009: 153-165). Thus, uncontroversially, lexical items, and particularly, content words are more malleable than inflectional morphemes and quite often also than syntactic patterns. In what follows we will try to give a general feeling for the levels affected in the various contact scenarios, but we will give pride of place to morphosyntax. It is more interesting, precisely because of their lower malleability and because its lesser visibility, it is also more controversial.

Finally and obviously, the discussion cannot be complete. Though the survey touches upon a great many contact situations, some will not even be mentioned, in part because not much is known about them (e.g. Crimean Gothic, Swedish Walloon or Ukrainian Swedish) and of the ‘great many’ that do pass the review, the discussion will be superficial, yet all in all this chapter should make clear that the Germanic languages offer an extremely rich data set, from Proto-Germanic prehistory until today. In what follows, the more detailed discussion will involve hypotheses about the areal modification of the grammars of current languages, viz. English, Mainland North Germanic, Dutch, Afrikaans and, surprisingly perhaps, French. A moderate amount will also be devoted to the three partially Germanic ‘mixed languages’.

2. Areality within Germanic

An important early inner-Germanic contact situation concerns the Viking settlement in the British Isles (in the smaller Northern Isles but also in a large Northern part of England called the ‘Danelaw’) during the fourth quarter of the first millennium. The languages of the Vikings (varieties of Old Norse) were close to Old and later Middle English, and they were predominantly adstratal. There was both borrowing and shift, with the speakers of Old and Middle English borrowing a lot of Old Norse lexicon and the Viking settlers taking Old Norse lexicon with them in the Old and Middle English that they were
switching to. The lexical interference was aptly described by Otto Jespersen more than a century ago: “[A]n Englishman cannot thrive or die or be ill without Scandinavian words: they are to the language what bread and eggs are to the daily fare’ (Jespersen 1905: 80). The italics are Jespersen’s and mark words taken from Scandinavian. These words include grammatical formatives such as they, and to that extent Old Norse influenced the grammar too. It is a matter of debate whether the substantial morphosyntactic simplification of English, as compared to its West Germanic sister languages, could, at least in part, be due to contact with the Vikings. Fischer (2013: 31-37) reviews this debate. From the fact that the loss of morphosyntactic variation happened faster in the North than in the South of England, she concludes that it is the imperfect learning in the shift from Old Norse to English that simplified English (a point already made by Jespersen 1922: 214 and repeated many times).

The influence of Old Norse on English is thus considerable, but because it is predominantly lexical and because the impact is not normally considered to make English very similar to North Germanic (yet see Emonds 2011 and Emonds and Faarlund 2014 for the opposite view), it would not be appropriate to consider English and (some of) North-Germanic to constitute a linguistic area. Probably during the time that Old Norse was given up in English, Scandinavianized English and Anglicized Old Norse did form a small linguistic area, but this Anglicized Old Norse is not documented very well (yet see Hofmann 1955). It is a bit different in the extreme North, for the Vikings also settled on Orkney, Shetland and the northern coast of Scotland. Here the switch happened much later, not to English but to Scots. The Old Norse, known as ‘Norn’ survived until the eighteenth century, in a progressively Scots-influenced version (Barnes 1998), and it left its mark on the local dialects, mostly their lexicons (Melchers and Sundkvist 2010).

While the preceding paragraphs briefly touched upon North Germanic impinging on West Germanic, in this paragraph we focus on the reverse progress. In the second half of the twentieth century the West Germanic language influencing North Germanic is English, but that is not very special given that English, due to its global role, exerts influence in the world at large. From the Reformation to the nineteenth century, it is High German that is important, but most impact dates back to the thirteenth to early sixteenth century, with Low German as the language of the Hansa League. This commercial enterprise took substantial numbers of speakers to mainland Scandinavian cities, more so in the Baltic South than in the North, both as transient and as resident traders. There Low German acted as superstrate but also as an adstrate, to the extent that Low German functioned as the Baltic lingua franca. Like adstratal Old Norse in England it contributed mainly to lexical MAT borrowing. The citation from Jespersen about the influence of Scandinavian on English is matched by one by his Norwegian colleague Didrik Arup Seip (1931: 102-103): “Two Norwegians cannot in our day carry on a conversation of 2-3 minutes without using L[ow] G[erman] loanwords ... of course without knowing that they are doing so. It is the same with Swedish and Danish”. As in the English scenario some of the material serves a grammatical purpose, particularly derivational affixes (cf. Diercks 1983). The influence also extended to morphology, with a change from a three-gender system to a two gender-system in the Hanseatic Norwegian town of Bergen (Östman 2011: 371), and to syntax, with e.g. the replacement of Swedish varda by the Low German loan bli(va) for the pseudo-copula ‘become’ (Markey 1969). Moreover, the

---

7 The original is in Norwegian. The translation is ours, but for the first sentence we follow Haugen (1976: 316), who is responsible for sharing this line with a larger audience.
contact itself, with language mixing and imperfect learning, has been held responsible for the simplification in the grammars of Mainland North Germanic, compared with Insular North Germanic, or in those of the southern Swedish dialects, as compared to the northern ones (Wessén 1929: 272; Haugen 1976: 285; Östman 2011: 371). But there is also evidence for complexification. For the replacement of varda by bliva Dahl (2009: 46) shows that it is complete in both the South, closer to the Hanseatic Baltic sphere of action, and the written standard, and that it is absent in the conservative Northern vernaculars. In the central dialects, however, Swedish has a complex system with varda and bliva competing depending on the form of verb it accompanies. Interestingly, Dahl (2009: 46) mentions that the original meaning of Low German bliva is not ‘become’, but ‘stay’ – as it still is in modern German and Dutch – and that a similar semantic shift can be found in Polish, Lithuanian, Latvian and Estonian. This puts the contact influence from Low German on North Germanic into the wider picture of the Circum-Baltic linguistic area. For this reason and also because the Low German influence had a considerable impact on morphology, one may invoke a notion of ‘linguistic area’ here, while there isn’t any reason for invoking one for aligning English and North Germanic.

Interestingly, both English and mainland North Germanic are morphologically simple as compared to German (for English) and Icelandic (for mainland North Germanic). In that sense, they have become similar. Although there is a contact explanation for both processes of simplification, involving the Vikings for English and Low German Hansa traders for mainland North Germanic, the similarities are not a matter of areal convergence: the morphological simplicity of either English or mainland North Germanic is not due to the morphological simplicity of the other language(s). Also, some of the parallels between North Germanic and English may also be due to the fact that both English and North Germanic are arguably not part of the core of the Standard Average European linguistic area and thus missed out on the convergence characterizing this core (see van der Auwera 1998a for an illustration involving phasal adverbials – already, yet, still, not yet, no longer).

What is also interesting about the Low German contact in the North is that there is some documentation of Scandinavian Low German. There may not be many studies, but see Carlie (1929) and Brattegard (1945) on the Low German of Bergen and Copenhagen, respectively. It seems that the influence of the North Germanic languages on Scandinavian Low German is modest in that it primarily concerned the orthography and the lexicon. At least part of the reason must be that Hanseatic Low German was after all a language of traders and travelers, not of migrants loosing contact with the homeland.

During the Middle Ages inner-Germanic contact influence in Scandinavia did not only involve Low German, but also the vernacular and standardizing North Germanic languages. The most important source language was Danish and the strongest contact influence, in the form of a superstrate, relates to Norwegian. Norway was ruled by Denmark from 1380 to 1814, and from the sixteenth century onwards the main written language in Norway was Danish. This written Danish-in-Norway gave rise to a spoken variety as well, and until today Dano-Norwegian or Bokmål ‘book language’, the current official term (earlier also Riksmål ‘kingdom’s language’) is one of the two official Norwegian languages. The other one is ‘New Norwegian’ (Nynorsk) (earlier landsmål ‘country language’), a standard created in the nineteenth century and based on Norwegian dialects. The twentieth century saw an attempt to unite the two standards – the Samnorsk ‘pan-Norwegian’ project. The two varieties did converge somewhat, but they nevertheless remain distinct (Jahr 2004) and in the spoken register, Nynorsk exists in
more than one variety. Through Bokmål Norwegian thus belongs to a true, but small linguistic area, a ‘Danosphere’\(^8\), a part of the Circum-Baltic area.

There is also a Danish contact story to be told for Swedish, but it is of a different nature. The dialects in the southern part of Sweden, Scania (Skåne), were more Danish than Swedish in the dialect continuum between the two languages. Politically, the area was Danish until 1658. After the annexation by the Swedes, a Swedification process started, with Swedish as the superstrate. This process appears to have been successful surprisingly fast (Haugen 1976: 351; Ohlsson 2004: 1362). Nevertheless, the Scanian dialects today still retain Danish features, and there is a Scanian variant of standard Swedish. The Danish dialects native to the area now only survive on the island of Bornholm – off the Scanian coast, closer to Sweden than to Denmark, but protected from Swedification because it belongs to Denmark. Scanian is thus not Swedish that fell into the Danosphere, the way Norwegian did. It is the opposite: it is Danish that entered the ‘Swedosphere’.\(^9\)

Danish also served as the superstrate language for Faroese and Icelandic. Iceland was under Danish sovereignty from the Middle Ages until 1943 while the Faroe Islands are still part of the Kingdom of Denmark, though with ‘home rule’ since 1948. Due to its isolation, its larger population, its earlier and complete independence from Denmark as well as its early writing tradition (twelfth to fourteenth century), the impact of Danish is less strong in Icelandic than in Faroese (Warren 2002: 2037-2038) and less strong than that of Low German on Mainland Scandinavian, thus placing Faroese in the middle of a West to East cline from Iceland to Mainland Scandinavian (O’Neil 1978: 277-281; Van Coetsem 1988: 52). The Faroese also have a distinct Faro-Danish variety, which differs from Danish not just in the lexicon but also in the grammar. The Danosphere therefore extends to the West too. Does it extend to the South as well?

In the South, North Germanic Danish borders on West Germanic German and Frisian, but, as the North vs. West Germanic division suggests, there is no dialect continuum, probably because the region had no or few inhabitants at the end of the period of the Germanic migrations (Walker 2001: 266). There are, however, various border effects, maybe already in the fifteenth century (Pedersen 2009: 312) and definitely in the nineteenth and twentieth centuries, with Petuhantendeutsch or Petuhsnak. This variety was named after ladies (Tanten ‘aunts’) that had Partout-tickets on the ferries. It has a predominantly (High) German lexicon and a Low German and Danish (standard as well as dialectal) grammar (Faatz 2009; http://petuhschnacker.de) The border area, Schleswig/Slesvig, was partitioned between a German and a Danish part in the aftermath of the First World War. As a result there is now a small German minority in the Danish part and a somewhat bigger Danish minority in the German part. Both the German and the Danish vernaculars show influence from their respective coterritorial languages (High German, Low German, Standard Danish and dialectal Danish, North Frisian) in lexicon, pronunciation and grammar (Pedersen 2000: 212-214, 2009: 306-307; Kühl and Petersen 2009).

Within West Germanic there is more contact influence and there are sometimes true contact varieties, on the borders between North and East Frisian and (Low) German (various chapters in Munske 2001), between West Frisian and Dutch (see below),

---

\(^{8}\) The term is inspired by Matisoff’s (1991) use of the terms ‘Sinosphere’ and ‘Indosphere’.

\(^{9}\) In prehistorical times, though, central Swedish may well have been in the Danosphere – to the extent that Dahl (2001) ventures the flippant question why Swedes speak Danish.
between Low and High German, between German and Dutch (Smits 2009) and between German and Western Yiddish (Reershemius 2007). The most interesting variety might well be the one called Stadsfries in Dutch, literally ‘Town Frisian’. It is not actually Frisian, and, in a sense, not really Dutch either. It is indeed a language of towns, more particularly of some of the towns in the province of Frisia in the Netherlands. Van Bree and Versloot (2008) convincingly argue that it arose in the sixteenth century when native West Frisian speakers learnt Dutch, the more prestigious language, but imperfectly so. To the extent that the original speakers intended to speak Dutch, their language can be considered to be a variety of Dutch with Frisian as a substrate. However, this substrate is so strong that Stadsfries became a language of its own, a ‘mixed language’ (Matras and Bakker eds 2003). Its accent, syntax, regular inflection and function words are mainly of West Frisian origin while its derivation, composition, irregular inflection and content words are primarily Dutch. No wonder there is some disagreement as to whether the language is ‘basically’ Frisian or ‘basically’ Dutch (Van Bree and Versloot 2008: 35-46). Stadsfries thus has a contact-based similarity to both Dutch and Frisian, but interestingly, it did not lead to one Stadsfries area, but to several ones, viz. the various Frisian towns.

Interestingly, the grammar of Dutch, as compared to that of English and German, has also been approached from a contact perspective. Comparing the three languages, English is morphologically simplest, German most complex and Dutch is in between. According to some, the deflexion of English is due to language contact – the Vikings. Weerman (2006) (and earlier but less explicitly also O’Neil 1978: 281) agrees and claims that the moderate deflexion seen in Dutch is also due to contact, viz. the influx of migrants in sixteenth- and seventeenth-century Amsterdam, the time and the place where the modern standard finds its origin. We thus see a geographical West to East gradient in West Germanic, similar to the North Germanic one (from Iceland over the Faroese Islands to mainland Scandinavia). Note, however, that both the deflexion of English and that of Dutch, though each arguably resulting from contact, do not constitute a case of areal convergence in the sense employed in this chapter. They are independent processes, just like those of mainland North Germanic and Faroese.

So much for a brief discussion of at least the most prominent areality effects involving the Germanic languages in Europe. But, of course, when the Europeans colonized the world, they took their languages with them and they interacted not only with the indigenous population but also with each other. In this way Germanic languages met outside of Europe too. Usually we find contact scenarios with English as one of the partners, especially in the United States and in South Africa. In the United States most contact scenarios resulted in the disappearance of the ‘other’ Germanic language, although this process often took a few generations and the ‘life and death’ of such languages is sometimes well documented (e.g. the 200 years of Texas German, Boas 2009). The two Germanic languages that resist having their speakers shift to English are Pennsylvania German and, less successfully, Yiddish. It is religion and culture as well as the size of the community (compare the size and the fate of the two West Virginia communities studied by Van Ness (1990) to the size and the fate of the Pennsylvania ones studied by Meister Ferré (1994)) that keep the languages alive and the contact influence from English relatively modest. American English did not undergo much

---

10 They typically took different dialects with them. This interdialect contact – contact again! – then led to dialect leveling and the formation of a new dialect, possibly out of line with the regular dialect continuum (?) back home.
influence from them either – perhaps the most visible influence being some Yiddish-derived lexical items (‘Yinglish’, Gold 1985). In South Africa, it is Afrikaans, the ‘colonial offshoot’ of Dutch, that has resisted the language shift to English. The two languages have been in contact (of various types) for more than two centuries and have influenced each other at the lexical and the grammatical level (Donaldson 1991, Bowerman 2004). But the Afrikaans impact on South African English phonology and morphosyntax, though significant, should not be exaggerated (Lass and Wright 1986) (e.g. progressive busy X-ing was once considered a pure calque of Afrikaans besig om te but has recently been linked to Settler English; Mesthrie 2002). In the same vein, it is tempting to attribute English-like aspects of Afrikaans to influence from English but such similarities may well be the result of independent developments (e.g. Colleman 2013): as the Germanic ‘contact languages par excellence’ (Van Coetsem 1988: 137), English and Afrikaans have both undergone extreme deflection (see also section 3.6) and are not unlikely to have developed, and perhaps continue to develop, in comparable ways.

3 Areality with Germanic and non-Germanic

In one sense, contact-based similarities between Germanic and non-Germanic should be easier to detect than inner-Germanic ones: the intrusion of non-Germanic sounds, words or patterns should be clearly visible, and vice versa for the intrusion of Germanic into non-Germanic. However, to describe the finer details one needs a knowledge of (at least) two language families, e.g. Germanic and Celtic, see Hickey (2007: Chapter 4) for a detailed discussion of the case for contact between Irish and English in Ireland during the past few centuries. Some instances of language contact may also date back very far, which makes the detection of the similarity very difficult, as we see in the first contact situation to be discussed now.
3.1. Semitic

In prehistoric Europe (from the fifth millennium BCE extending into our era), Vennemann claims—in two collections (2003, 2012) with the titles Europa Vasconica – Europa Semitica and Germania Semitica echoing Gamillscheg’s Romania Germanica (1934-1936) – speakers of Semitic languages (probably Phoenician) were present on the Atlantic coast of Europe and exerted a superstrate influence on Indo-European. This influence particularly targeted the languages that were to become ‘Germanic’. The argumentation is based on cultural and mythological similarities as well as linguistic ones. Thus Germanic verbal ablaut (i.e. the Germanic strong verb system) and a good part of the ‘sizable’ non-European lexicon of Germanic would be due to this superstrate influence. According to Vennemann (2003, 2012), Indo-European also underwent a prehistoric, substratal influence from Vasconic (the family that Basque is the sole survivor of), but this influence did not specifically target Germanic. Despite the great depth of the hypothesized contact influence, the argumentation is rather detailed, but also speculative and it competes with Nostratic attempts to explain Indo-European Semitic similarities (Bomhard and Kerns 1994).

A modern Semiticized Germanic language is Yiddish, with superstrate input mainly in the vocabulary of about 12% to 20% (Jacobs et al. 1994: 417) and marginally also in morphology (with the nominal pluralizer –im in profesorim ‘professors’, a pejorative alternative for profesorn; Jacobs 1994: 402). One would not, however, resort to the notion of ‘linguistic area’ to explain the similarity. The Semitic superstrate effect does not, of course, come from an adjacent or coterritorial group of speakers of Aramaic and Hebrew. It is a subset of the speakers of Yiddish themselves that also speak or at least use Aramaic and Hebrew. The Semitic effect on Yiddish is thus similar to the Latin and Greek effect on European languages.12

11 The size of this part of the lexicon is unclear: for Polomé (1972: 45) it is ‘substantial’, for Hawkins (1987: 74-75) it could be a ‘full one third’ and for Vennemann (e.g. 2012: 257) even ‘close to two thirds’.

12 The areal effect of Latin and Greek on European languages and that of Hebrew and Aramaic on Yiddish is not quite the same though. Thus Kortmann (1988) does give Latin and Greek a role in the formation of linguistic areas, i.e. the Balkan one for Greek and Standard Average European for Latin, and Drinka (2013) does the same with Carolingian Latin in relation to the Charlemagne core of Standard Average European.
3.2. Finno-Ugric

Vennemann’s hypothesis that earliest Germanic was a contact language relates to Vasconic and Semitic. But there is another view on the contact language status of Germanic, based on phonetics, that takes Germanic to have involved a Finno-Ugric substrate and a language shift from Finno-Ugric speakers (Wik 1997). This view is also contested, though not the less radical claim that Finno-Ugric and Germanic have been in contact for a long time (Koivulehto 2002), to some extent as part of the Circum-Baltic area (Wälchli 2011: 326). And instead of contact, a common origin, in Nostratic, might explain some of the similarities.

In historical times, Finno-Ugric interacted with both North-Germanic and West-Germanic. Both Norwegian and Swedish interacted with Sami languages (e.g. Bergsland 1992) and with Finnish. In addition, Swedish interacted with Estonian and some minor Finnic languages, which in turn (except for Sami) interacted with Low German and later High German – and often also a Baltic and/or Slavic language. In Norway and Sweden, the Finno-Ugric languages are minority languages, and in Finland and earlier also Estonia, it is the other way round. The Finnish-Swedish contact is described best (see Lagman, Ohlsson and Voodla eds 2002, Ohlsson 2004 for references to other situations) and it is also rather special, in the sense that Swedish used to be the only ‘official’ and high-prestige language in Finland and that the minority that speaks it as a native language is not so small (some 290,000 speakers). In its written form Finland Swedish is very similar to Sweden Swedish. The spoken variants are rather different, both because of their dialect base and because of the influence from Finnish. The latter is especially clear in the pronunciation, prosody (e.g. Helgason, Ringen and Suomi in print) and the lexicon, with so many intrusions from Finnish that they have been the target of purists for more than a century (Östman 2011: 366). This variant of Swedish is the one that is best described. Others include the Finland Swedish used by native speakers of Finnish and the Swedish spoken by the sizable Finnish immigrant community in Sweden.

The other major Finno-Ugric language of Europe, i.e. Hungarian, interacted with Germanic, too, in this case, German. Most of the Hungarian Germans trace their ancestry to migration waves of the late seventeenth and most of the eighteenth century taking settlers into present-day Hungary, but also into present-day Romania and Serbia, when these areas fell under the Habsburg monarchy. The villagewise settlement created linguistic enclaves, a phenomenon for which German linguistics has appropriated the term Sprachinsel (lit. ‘language island’). In these enclaves contact influence went in both directions, mostly, as expected, involving lexicon and pronunciation (Manherz and Wild 1987, Knipf-Komlósi 2006).

3.3. Romance

English is the most Romance of the major Germanic languages. This is due in part to Latin, which influenced its lexicon – but this happened for all the Germanic languages – but primarily to French, which acted as a superstrate as a consequence of the invasion by William the Conqueror in 1066 and the ensuing colonization of the British Isles. The influence is visible in the lexicon (with an estimated 38% of the 4000 most common words – Scheler 1977: 70-77, Lutz 2002: 147, Durkin 2014: 254-280) and in derivational
morphology (e.g. –ment or –(a)tion-; Dalston-Puffer 1996). Syntax has largely remained untouched, as Miller (2012: 185-186) and Fischer (2013: 38-40) argue after a scrutiny of the literature. The impact on the lexicon, it should be noted, was partially a matter of replacing Anglo-Saxon and partially a matter of enriching it. As a result, the lexicon became more complex. The now obligatory choice between *sheep* and *mutton* is a case in point: the former comes from Anglo-Saxon, the latter from French. Another example is the stylistic difference between Anglo-Saxon *help* and ‘French’ *aid* or *assist*. The variety of French that crossed the channel with the Normans was ‘Norman French’, itself a contact variety of French with minor lexical input from Old Norse. Once settled on the British Isles, it can be called ‘Anglo-Norman’ and it initially functioned as the mother tongue of the invaders and it progressively became a language learned through instruction. Because the language left a large body of texts from the Conquest in 1066 until the fifteenth century, we can see it change throughout the centuries. Thus Ingham (2012) focuses on the grammar and shows how Anglo-Norman diverged from continental French first in the phonology, then the morphology and finally and only marginally the syntax.

With this clear areal influence from (a kind of) French on English and back there is some temptation to ‘gild’ it with the concept of ‘linguistic area’. But this is not done. English, different from French, is also strongly influenced by North Germanic (see above) and probably also by Celtic (see below). Interestingly, there are two parallels to the English-French contact scenario just sketched. For the first, we have to go to the Channel Isles of Jersey, Guernsey and Sark, where something like Anglo-Norman French still survives – but probably not for very long anymore. These dialects are similar to the ones of the Norman mainland, but they have more Anglicisms, given that the standard superstrate language of the islands is English and that of the mainland is French (Spence 1984). They also colored, as a substrate, the English spoken there, ‘Channel Island English’ (Jones 2010). For the second parallel, we have to go to the French mainland.

If English is the most Romance of the Germanic languages, French has been called ‘the most Germanic of the Romance languages’ (Walter 1994: 225, Hélix 2011: 12). This ‘Germanization’ is a result of the migrations of Germanic tribes into the (Gallo-)Romance lands that were part of the Roman Empire, involving Goths, Lombards, Burgundians, Saxons and Franks. The effect on the Romance lexicon and toponymy is well described (Gamillscheg 1934-1936). The language most affected by the migrations was French and as the name of this language already suggests the crucial Germanic language was Frankish (Franconian), most clearly surviving as Dutch and the Franconian dialects of German. The Franks conquered parts of present-day France, settled in the North amidst the Gallo-Romance population, and ruled a mixed Romance-Germanic area, with more Germanic in the North, from the end of the fifth century. In this bilingual area Frankish was a superstrate – we have to wait until king Hugh Capet of the second half of the tenth century to have a ruler that did not have Frankish as a mother tongue (Picoche and Marchello-Nizia 1994:13) – and probably also an adstrate, in the Northern areas with

13 If the morphological attrition of English is associated with language contact, it is probably the contact with Old Norse that is relevant here, as mentioned above. But Norman French is also associated with it (e.g. Wessén 1929: 272, Mossé 1952: 44).
14 These languages are West-Germanic (Frankish, Lombardic, Saxon) or East-Germanic (Gothic, Burgundian). But as mentioned earlier, North-Germanic Old Norse also infiltrated Romance, viz. Norman French, and this way, to a minor extent, later standard French.
the denser Frankish settlement. The Frankish influence on the Gallo-Romance of the North (the langue d’oïl as opposed to the langue d’oc or Occitan of the South), which was to become standard French, did not only concern toponymy and the lexicon. According to Mitterand (1968: 18), about 35 of the 1000 most common current French words come from Frankish or some other older Germanic language. But the influence is also evident in phonology (e.g. the diphthongisation of Latin tres ‘three’ into French trois) and grammar. For grammar the literature usually mentions the more frequent anteposition of the attributive adjective in Old French – still reflected in toponyms (e.g. Northern Neufchatel ‘new castle’ vs. Southern Châteuneuf) – and the entry of some morphology (e.g. the suffix –ard). The most interesting ‘Germanic-like’ features are (1) the non-pro-drop character of French, i.e. the requirement a non-emphatic subject pronoun in French je t’aime (lit) ‘I you love’ (compare Italian ti amo (lit.) ‘you love.1st-person-singular’), and (2) the fact that Old French was a ‘verb-second’ language, requiring a verb in the second slot of the sentence. Interestingly, the view that these two features are not only superficially Germanic but at least in part due to contact influence from Germanic (with a language shift scenario) has not been given much attention in recent years. Thus it is not discussed by Picoche and Marchello-Nizia (1994) or Marchello-Nizia (1995)15, for instance, or scholars are content with observing the similarity between French and Germanic (Rainsford, Guillot, Lavrentiev and Prévost 2012: 175). This stance contrasts with that of earlier linguists (particularly von Wartburg 1946: 66; 1967: 102-104; but also Hilty 1968: 507-511) and of linguists that have tried to define ‘Standard Average European’ and even posit a Charlemagne area as its core (van der Auwera 1998, implicitly Haspelmath 1998, 2001 – cf. also Drinka 2013). The parallel with the French influence on English is clear, but partial. In both cases the language that arrived later, French in England and Frankish in France, came in as a superstrate. But in the North of France the incoming language must have been adstratal as well. It led to a wider bilingualism and to a slow language shift. In that sense the scenario is closer to that of the North Germanic contact influence on English, which was also adstratal and involved bilingualism and language shift. Interestingly, in France this kind of contact led to the calquing of syntax. This is not clear in England, but the reason is perhaps that the syntax of Old and Middle English and Old Norse was already very similar. In English, the language shift is taken to have led to morphological simplification. Interestingly, such a claim is not unreasonable for French either. In many ways French is phonetically and morphologically simpler than the other Romance languages. To some extent this difference is similar to the way English is phonetically and morphologically simpler than the other West Germanic languages (Lamiroy 2011) and – a further parallel – to the way Mainland North Germanic is simpler than Icelandic and Faroese. If the simplification of English and that of Mainland North Germanic are at least partially due to language contact, it is tempting to investigate whether a similar claim would make sense for French.

The interaction of Frankish and French did not stop when Hugh Capet ascended the French throne more than a millennium ago. Dutch descends from Frankish and it never stopped interacting with French. A very special case is Luxembourgish, now the national

15 Marchello-Nizia (1995: 61-65) surveys the various explanations of the V2 character of Old French, starting with Le Couthret (1875), considered typical of the period in which linguists were discussing to what extent French was ‘Germanized’. For later research, she does not mention the contact influence hypothesis (Marchello-Nizia 1995: 62).
language of Luxembourg (functioning alongside French and German as ‘administrative languages’): it is a Frankish variety with a lexicon particularly receptive to French. Going south along the Romance Germanic language border and into Switzerland and Italy brings us to many more German-Romance contact varieties. Some are German influenced Romance, such as the Rhaeto-Romance of the Swiss Canton of Graubünden/Grischun, which has a verb second word order, at least partially due to German (Kaiser and Scholze 2009) and there is also Romance influenced German, such as Cimbrian. Cimbrian refers to German enclaves in the North of Italy, with a strongly Italian influenced Bavarian of the originally eleventh- and twelfth-century migrants (Bidese, Dow and Stolz (eds.) et al 2005). In the Balkans German interacted with Romanian (Gadeanu 1998) and out of Europe, English and French interact(ed) in Canada and in the US (e.g. King 2000 on English influence on Acadian French morphosyntax; Poplack 2008 and Boberg 2012 on the predominantly lexical influence of French on Quebec English), in the US English interacts with Spanish (e.g. Roca and Lipski 1993; Otheguy and Zentella 2012 on the presence/absence of subject pronouns in Spanish in New York), and in Latin America the Low German of the Mennonites interacts with Spanish and Portuguese (Kaufmann 2011).

3.4. Celtic

When Germanic tribes, in this case Saxons, Angles and Jutes, settled in the British Isles in the fifth century the population they met spoke Celtic languages. The two groups mixed and by the end of the first millennium Anglo-Saxon had replaced Celtic in most of England. This process would continue – but not to completion – in Ireland, Scotland, and Wales. Celtic was the substrate, responsible for toponyms and a very small set of lexical items and no more needs to be said. At least, that was the agreed opinion until recently and to some extent it still is: when Fischer (2013) discusses the role of contact for syntax in the Old and Middle English periods Celtic is not mentioned and Townend (2012: 104) writes that ‘sadly the possible influence of Celtic on English (besides the handful of loanwords [...] ) remains obscure and disputed’. But the view that Celtic did play an important role in shaping English grammar has gained ground (Filppula, Klemola and Pitkänen eds. 2002; Hickey 2012), with McWhorter (2008: 61) echoing Jespersen on North Germanic in English and Seip on Low German in North Germanic when he describes English as a language ‘whose speakers today use Celtic-derived constructions almost every time they open their mouths for longer than a couple of seconds. Do you want to leave now? What’s he doing? Did he even know? What are you thinking? I don’t care. She’s talking to the manager.’ The constructions he refers to are do periphrasis and the progressive (van der Auwera and Genee 2002; Filppula 2002). The list is not restricted to these two features: White (2002: 169-170) lists as many as 52 possible ‘Brittonicisms’17, as good a prospect as any for daring to advance a Sprachbund hypothesis (Tristram 1999, Filppula, Klemola and Paulasto 2008: 228-229) and for keeping both English and Celtic out of the Standard Average European core (van der

16 The study of the synchronic and diachronic grammar of Luxembourgish has only just started (Moulin and Nübling eds. 2006).

17 This list includes morphological attrition, more commonly associated with Scandinavian contact (see above), rarely with Norman French contact (note 13). See also Tristram (2002).
Auwera 1998b). Similar to the Frankish influence on French syntax, the process deemed responsible is language shift (Hickey 2007: Chapter 4), with Celtic learners of English calquing features of Celtic.

Again, just like Romance-Germanic contact on the continent, Celtic and English continue to interact until today, in the regional varieties of English spoken close to or in areas with modern or earlier speakers of Welsh, Irish, Scots Gaelic, Cornish and Manx (Filppula, Klemola and Paulasto 2008: 135-220). The opposite contact influence is also very real but it is less well described (e.g. Jones 1973 for Welsh and Stenson 1993 for Irish).

3.5. Balto-Slavic

We have already discussed that the Baltic area was the stage for an interaction between Low German and later High German with North Germanic and Finno-Ugric. The contact also involved Slavic and Baltic languages and among the latter Latvian stands out, a contact which gave rise to the Latvian (and Estonian) influenced Baltendeutsch of the German settlers and the Halbdeutsch of the mother tongue speakers of primarily Latvian and Estonian but other languages too (e.g. Polish, Russian). Best described are the influences in the pronunciation and the lexicon (see Polanska 2002 for Latvian-in-German and Jordan 1995 for German-in-Latvian). Further east and south, German interacted with Slavic languages, especially North and East Slavic, but also South Slavic (especially Slovenian), not necessarily coterritorial, with mutual lexical enrichment, also in the realm of function words (e.g. the modal verb müssen ‘must’, Hansen 2000). The influence can also be more subtle as when one expresses uncertainty with either a verb meaning ‘may’ – the Western, Standard Average European strategy – or with an adverb meaning ‘maybe’ – the Eastern strategy (van der Auwera, Schalley and Nuyts 2005). Apart from the contacts along the German-Slavic line, there is also the typically German Sprachinsel phenomenon. Thus German settlers created small and large German enclaves in Russia, starting in 1765, the most well-known one that of the Volga Germans, creating varieties of Russian German (Berend and Jedig 1991). A particular twist to the contact history of this variant is that many of the ‘Russian Germans’ have now remigrated to Germany, allowing their Russian German – as well as their German Russian – to be influenced by varieties of ‘German German’ (Berend 1998, 2009). There are also two Slavic enclaves in German speaking areas, viz. the West Slavic enclaves of Sorbian in the eastern part of Germany (Scholze 2008) and the South Slavic enclaves of Burgenland Croatian in eastern Austria (e.g. Pawischitz 2009). We should also mention Yiddish, in its Eastern variety, interacting with Baltic (especially Lithuanian) and Slavic (especially non-standard varieties of Polish, Ukrainian, Belarusian and Russian), mainly in the lexicon – including grammatical formatives such as the question particle czy and the conjunction abi ‘if only’ – but also in the phonology and morphosyntax (Weinreich 1958a, Eggers 1998, Jacobs 2010) – with e.g. the Slavic style reanalysis of verbal prefixes (Talmy 1982, Kiefer 2010). There is also a claim that Yiddish is basically Slavic (Sorbian, Ukrainian and Belarusian), relexified into High German and, as the process continued, Yiddish-in-statu-nascendi (Wexler 2002).

3.6. Indo-Aryan
A particularly interesting case is the contact of Germanic with the Indo-Aryan language Romani, the language of the Roma. The Romani varieties of groups that lived in or passed through Germanic lands show influence from the Germanic languages, particularly German. But in two cases the contact also led to mixed languages, called ‘Para-Romani’ in Romani scholarship. One is Angloromani (Hancock 2010) and the other Scandoromani (Hancock 1992, Lindell and Thorbjörnsson-Djerf 2008). In both cases, the grammar is strongly Germanic while the lexicon is Romani. In this respect the languages are like Stadsfries and a further similarity is that the languages are not confined to one area, but to many small ones, not in cities this time but in diaspora. But there is an interesting difference too. For Stadsfries the substrate (Frisian) is prominent in the grammar, for Para-Romani the substrate (Romani) is prominent in the lexicon. This difference may be related to their different origin. Whether Para-Romani is a late stage of the abandonment of the ‘real’ Romani, surviving mostly in lexical items (Matras 2009: 295) or a consequence of an effort to create a secret language (Hancock 1992: 43-44) or even partially both (Hancock 2010: 374), these scenarios are irrelevant for Stadsfries and Frisian.
3.7. Other

Germanic interacted with other families too, e.g. Eskimo-Aleut languages (Danish in Greenland, and English in Alaska and Canada or American Indian languages in the US and Canada. In South Africa we find Dutch interacting with a number of language families. When Dutch was stabilizing and nativizing there, it was in immediate contact not only with the Khoisan language of the indigenous population but also with the languages of the Asian and African slaves that were brought to the Cape, Malay and Creole Portuguese in particular (as well as with the French, German and English of other migrants from Europe). Later on, it also interacted with the Bantu languages spoken in the rest of South Africa (Van Rensburg 2012). The lexical influence on the language cannot be denied (e.g. the Afrikaans words dagga ‘cannabis’, baie ‘very/much’, mielie ‘corn/maize’ and babelas ‘hangover’ come from Khoisan, Malay, Creole Portuguese and Bantu respectively). The impact on the grammar, however, has been the subject of debate for a long time, as it relates directly to the question of the origin of Afrikaans: is it (like) a creole (e.g. Hesseling 1923) or has it developed out of Dutch autonomously (e.g. Smith 1952), possibly with some dialect leveling and convergence? These two extreme positions seem untenable, though. First, Afrikaans exhibits just a few of the ‘typical’ features of creoles (Markey 1982) and has preserved the Dutch word order rules of V-second in main clauses and V-final in subordinate clauses, unlike ‘real’ creoles based on the language (Biberauer 2009: 9). Second, the autonomous evolution view cannot explain the drastic changes that Dutch underwent between 1652, when the settlers arrived, and the end of the eighteenth century, when the Europeans as well as the Khoi and the slaves already spoke something very Afrikaans-like (Raidt 1983). A more likely account involves the interaction of, on the one hand, the Dutch that was spoken and transmitted by the Europeans (a more conservative force) and served as the superstrate of the Cape Colony and, on the other hand, the Dutch that the Khoi and the slaves learned ‘imperfectly’ and ended up switching to en masse (a more progressive force) (e.g. Van Coetsem 1988: 129-144, Den Besten 1989). This process resulted not only in morphological simplification (similar to what happened to English, French and Mainland North Germanic) but also in some ‘complexification’. The varieties of the Khoi and the slaves were probably heavily influenced by the languages that they initially spoke – in fact, the three main dialects of present-day Afrikaans still reflect the varying degrees of influence from Khoisan (Orange River Afrikaans), Malay (Cape Afrikaans) and Settler Dutch (Eastern Cape Afrikaans) (Van Rensburg 1983) – and the origin of particular complexities of Afrikaans has been a matter of considerable discussion among scholars (e.g. Den Besten 1986 and Bernini and Ramat 1996: 51-81 on the Dutch, Khoisan and/or Creole Portuguese basis of the double negation with the clause-final marker nie). Note, finally, that the contact between Afrikaans and Bantu has also given rise to new varieties (e.g. Flaaitaal ‘Fly Language’, a township argot primarily used by black men in urban areas; Makhudu 2002).

4 Conclusion

Studying the areal linguistics of Germanic has a good side to it as well as a bad side. The good side is that both the linguistic and the extralinguistic phenomena are reasonably well documented, at least in comparison with most other language families. The bad side
to the study of Germanic areality is that many Germanic languages as well as the non-Germanic ones that they have been in contact with have been standardized, and standardization inhibits contact influence. For this reason the more interesting areality effects are to be found in either the period before these languages were standardized or in the non-standard varieties usually found in borders regions, including the borders of linguistic enclaves. These two situations are also what we have focused on in this chapter. A further focus was on the areality of grammar, the reason being that grammar is taken to be rather stable, as compared to the lexicon. In this respect the Germanic languages are quite special: “[...] the Germanic group is only with some difficulty and chiefly historically kept together from the point of view of their grammars’ (O’Neil 1978: 281) and this divergence is arguably due to contact, at least in part. This is most strongly arguable for the morphosyntactic simplification of English due to contact with North Germanic (and perhaps Celtic too) as well as for the morphosyntactic simplification of Mainland North Germanic due to contact with Low German. For both West and North Germanic languages it is also argued that the intensity of the contact resulted in East West clines, from English to Dutch (and Frisian) and then German, and from Icelandic to Faroese and then Mainland North Germanic. Further parallels are the simplification of Afrikaans relative to Dutch arguably due to contact with Khoi and other languages, and that of French arguably due to contact with Frankish. In each of these four cases the process would primarily involve language shift and imperfect learning, factors that are also claimed to have resulted in at least three ‘mixed languages’, viz. Angloromani, Scandoromani and Stadsfries. In some cases these areal convergences feature as elements in the argumentation for linguistic areas. Thus the Hanseatic separation of the Mainland North Germanic languages happened in the Circumbaltic linguistic area, the ‘Germanization’ of French in the Standard Average European linguistic area, and the Celticization of English in a British Isles linguistic area.

References


Fischer, Olga. 2013. ‘The role of contact in English syntactic change in Old and Middle English periods’, in Schreier and Hundt (eds), 18-40.


Melchers, Gunnel and Peter Sundkvist. 2010. ‘Orkney and Shetland’, in Schreier, Daniel, Peter Trudgill, Edgard W. Schneider and Jeffrey P. Williams (eds.), 17-34.


9 Britain and Ireland

Raymond Hickey

1 Delimiting geographical areas

Discussions of areality and linguistic areas generally assume a circumscribed spatial dimension to the topic: candidates for such areas are all geographically delimited in some particular way. This delimitation can be realised by a topographical feature such as a mountain range or a major river which can often act as an obstacle to population movements, or at the very least such a topographical feature may induce a perception among the members of a population that it represents a dividing line between them and others. The delimitation issue might seem less controversial if the area being considered consists of an island separated from a larger mainland. But there is nothing per se which makes an island a linguistically separate area from any nearby mainland. It is the behaviour of populations and various political and cultural events in history which may increase the linguistic separateness of islands. Relative geographical isolation does lead to linguistic conservatism as can be seen, for example, with Newfoundland vis à vis mainland Canada or Iceland vis à vis mainland Scandinavia. The larger the islands in question and the larger and more heterogeneous their populations, the more linguistically complex they are and the less they can simply be regarded as relic areas. When it comes to islands with the size and population of Britain the relationship to a nearby mainland is more or less irrelevant. There are a few linguistic traits which the south of England shares with the Low Countries of Europe, e.g. initial fricative voicing (in Dutch), but the European continent does not now have any linguistic influence on language in England.

The islands of Britain and Ireland are roughly comparable in size: Britain at 209,000 square km is somewhat over twice as large as Ireland at 84,000 square km. The remaining islands of the archipelago, above all the Isle of Man between England and Ireland, the Hebrides off the north-west coast of Scotland and the Shetland and Orkney Islands off the north-east coast of Scotland are, in size, only a fraction of the two main islands. The present-day political division of the islands is as follows: Britain consists of three political entities, England, Wales and Scotland, each constituent parts of the United Kingdom of Great Britain and Northern Ireland. The island of Ireland is divided in two politically: the Republic of Ireland consists of the southern two thirds of the island with Northern Ireland, the fourth constituent part of the United Kingdom, occupying the remaining third of Ireland in the north-east of the island.

---

1 The terminology surrounding references to the two large islands off the north-west coast of Europe is fraught with political and cultural sensitivities and the older geographical term “The British Isles” is viewed with disfavour in the Republic of Ireland and is increasingly avoided in Britain as well (see Davies 1999). After considerable reflection the names of the two islands were chosen for this chapter as they imply two separate entities but the conjunction ‘and’ (intentionally) highlights their relatedness, including their linguistic connections.
1.1 Areality in Britain and Ireland: some caveats

When dealing with the areality of language in Britain and Ireland, no claim is made about whether Britain or Ireland are linguistic areas in the classical sense (see Campbell, this volume; Hickey 2012a). But the issues discussed, specifically the factors promoting or demoting areality in Britain and Ireland are concerns shared with the other studies in this volume. When considering areality in Britain and Ireland some caveats apply which are not pertinent to similar considerations in, say, native South America or Africa. In both islands English has a long documented presence going back to the late sixth century and the early fourteenth century for Britain and Ireland respectively. This gives us a time depth which is not normally found in parts of the world outside Europe. For nearly all the language scenarios considered in this volume the treatments are present-day snapshots of developments which reach far back into the past but for which there is no documentation. In the case of English one cannot ignore the large textual record of the language which relativises many features which one finds in the present-day language.

The second major caveat is that the effects of supraregionalisation and the standardisation of English cannot be ignored as these processes serve to mask feature distributions which existed before both supraregionalisation (Hickey 2013a) and standardisation (Hickey, ed. 2012) set in, essentially leading to a levelling of strongly local features. \(^2\)

Supraregionalisation is a process by which vernacular features are removed from local varieties and more general features take their place. Consider the case of the FOOT-STRUT split, a reference to the split of Early Modern English /u/ into [U] and [V] (Britain 2012: 26). This occurred in the seventeenth century in southern England (but not in the north) with the general unrounding and lowering of /u/, e.g. in cut [kU:] > [kV:] (Dobson 1968: 585-593). The high vowel was retained in rounded environments, e.g. before [S], before a velarised [S] and sometimes after a labial stop which is why push, pull and put still have [U]. \(^3\) Given that the FOOT-STRUT split originated in the south of England and affected speech in London and the Home Counties later standard varieties of English incorporated this shift, leaving the North of England as a region which contrasted with standard English in this respect. Furthermore, in local Dublin English the [U] > [V] shift did not take place either, as seventeenth-century southern English had no influence on this vernacular, and so it did not experience the FOOT-STRUT split. However, an [V]-pronunciation was introduced for STRUT-like words (cut, cud, done, dove, etc.) through supraregionalisation, that is the local pronunciation without a split was abandoned with the rise of an educated middle class in Ireland, probably in the early nineteenth century, and the English pronunciation of STRUT-like words with [V] was adopted. Hence the split is found in all non-vernacular forms of English in Ireland, not because of an internal change but due to the adoption of a pronunciation from England.

Standardisation processes are closely connected to prescriptive notions of language.

---

\(^2\) An approach which examines vernacular features but without considerations of areality is found in discussions of putative vernacular universals, see Szmrecsay and Kortmann 2008 for a typical discussion.

\(^3\) In the word foot Middle English /o:/ was shifted to /u:/ as part of the Great Vowel Shift, then shortened to [U] but because the shortening took place after the [U] > [V] shift the word was not affected (contrast blood [blVd] in this respect).
Such notions came to the fore in eighteenth-century England with such figures as Robert Lowth, John Walker and Thomas Sheridan (Hickey 2010). This led to the removal of certain features from non-vernacular usage. For instance, the use of demonstratives as personal pronouns in the third person plural is a feature of many traditional dialects of England and Ireland, *e.g.* *Them boys out on the street, Them shoes are too big for me.* This feature was also found in official registers before the eighteenth century in England (Claridge and Kytö 2010: 30-32) but did not become part of the later standard. The same is true for certain verbs which previously had two forms but now have three in non-vernacular varieties, *e.g.* see (*see: seen : seen*), I seen him, and do (*do : done : done*), *I done the work* (Anderwald 2009).

1.2 The textual record of English

The documentation and orthography of English can help one recognise previous instances of change which have been completed, *e.g.* the loss of initial /kn-/ and /gn-/ clusters, where only the orthography betrays the pre-change situation, *cf.* *know* and *gnaw* respectively. Despite the considerable historical documentation of mainland English it is sometimes necessary to look to lesser-known sources to determine the time depth of certain changes. An example is TH-fronting, the realisation of /2, 3/ as /f, v/, a stereotypical feature of Cockney which has been and is still spreading to urban vernacular varieties throughout Britain (Foulkes and Docherty 1999: 11). Certain peripheral documents, such as the glossaries of the archaic and now extinct dialect of Forth and Bargy in south-east Wexford, Ireland have instances such as *brover* ‘brother’, *aulaveer* ‘altogether’ (Hickey 2007: 77) which indicate that TH-fronting has a considerable vintage in English. Another instance is the LAUGH-shift, the historical shift of /x/ to /f/ (Hickey 1984a), compare English *laugh* and German *lachen* with /-x-/, or southern English *dough* with northern *duff* ‘steamed pudding’. This shift appears to have survived at least into the seventeenth century and to have been found in the south and east of England because it appears in the documents of the Salem Witch Trials, court petitions and depositions from Massachusetts, New England in 1692 and 1693 (Rosenthal et al. 2009), *e.g.* *dafter (daufter) for daughter and thof for though*.

1.3 The historical sociolinguistics of language contact

The information one has about societies throughout history can be useful in accounting for how contact arose and the effects it had on the languages involved (Hickey 2012d). Indeed one and the same language can have different contact effects in differing sociolinguistic scenarios. An example of this is provided by Anglo-Norman in medieval Britain and Ireland. Forms of Northern French were taken to England in the second half of the second century as a result of the Norman invasion of 1066. Over a century later, in 1169, Norman warlords from the Welsh marches (present-day south-west Wales) started a large-scale military campaign in Ireland thus introducing their forms of Anglo-Norman to Ireland in the late twelfth century. But the social situation of Anglo-Norman in Britain and in Ireland in the Middle Ages was different: in Britain the French took over leading positions in England society at the time, at the court, in the administration, the military and in the higher clergy. The representation of French in upper levels of English society
meant that the borrowings into English were part of more formal registers and to this day
the split of English vocabulary can be seen in words pairs like freedom and liberty, work
and labour, etc. In Ireland the situation of the Anglo-Normans was different. Their
primary concern was gaining land and to this end they built fortified castles in the
countryside, so-called keeps, and settled down there, interacting with the surrounding
Irish population in the process. The linguistic result of this is that many loans from
Anglo-Norman appear in Irish and not a few of them are ‘core’ vocabulary items like the
words for ‘boy’ (garsín < Anglo-Norman garçon) and ‘child’ (páiste < Anglo-Norman
page). Given that Anglo-Norman was the superstrate in the medieval Irish society why
should the Irish have borrowed such core vocabulary items as ‘boy’ or ‘child’? The
answer would seem to lie in the manner in which these words entered Irish. Consider that
the Anglo-Normans used such words in their variety of Irish. It is a historical fact that the
Anglo-Normans lived in the countryside among the Irish and gradually shifted to their
language. During the shift period an intermediate variety was spoken by the Anglo-
Normans in which they used words from their own language like garçon and page.
Because of the status the Anglo-Normans enjoyed in Irish society, the native Irish
adopted core vocabulary items of their variety of Irish and, for example, the negation
structure Níl puinn Gaeilge agam [is-not point Irish at-me] ‘I cannot speak Irish’ which
shows the negative use of French point (Rockel 1989: 59). The likelihood of this scenario
can be strengthened by considering that the Anglo-Norman loans in Irish did not
necessarily replace the native Irish words. For instance, the Anglo-Norman loan páiste
exists side by side with the original Irish word leannbh [læn'b̪h̪] ‘child’; for ‘boy’ Irish has
two words: garsín (with the later form gasúir) and the original buachaill [bu'k̪æil‘], but
without the stylistic distinction similar word pairs have in English. To use a metaphorical
term, the Anglo-Normans ‘imposed’ elements of their second-language variety of the
Irish language on those who surrounded them in late medieval rural Irish society. The
conclusion to be drawn from these considerations is that the results of contact depend
crucially on varying sociolinguistic scenarios which must be taken into account where
historical information about them is available.

1.4 Remnants of former distributions

When viewed areally, Britain and Ireland show geographically discontinuous attestations
of features which would suggest that previously there was a much wider occurrence of
features which has been reduced to a few disconnected locations. For instance, the T-to-R
rule, which allows the realisation of /t/ as /r/ intervocally or between prosodically
linked words, as in bottom [b̪rʊm] or get up [gər-u] is found in the North of England
(Buchstaller, Corrigan, Holmberg, Honeybone and Maguire 2013) and in local Dublin
English (Hickey 2005: 41). Another instance is the uvular r which was common in
Northumbria in the north-east of England (called the Northumbrian burr). In Ireland
uvular r was very common, in both Irish (Hickey 2011: 162-166) and English, and is still
found in the town of Drogheda (north of Dublin) and in many vernaculars of the east and
south (Hickey 2004: 78-79) as well as occasionally among speakers of Irish.

The NURSE-TERM distinction, a continuation of the Middle English distinction
between /U/ and /ə/ before tautosyllabic /i/, falls into this category. This is now only
found in traditional dialects of the British Isles, especially in Scotland and Ireland.
Elsewhere the sounds in these words have merged, usually to a central schwa-type vowel.
However, it is not restricted to specifiable locations as would be the case with the vestiges of second person singular pronouns (Hickey 2003; Raumolin-Brunberg 2005). For example, in rural Yorkshire as well as on Orkney and Shetland, the original second person singular pronoun thou, with the oblique form thee, is still available. In the West Midlands and in the South-West thee serves as a nominative form (Upton and Widdowson 1996: 66-67).4

The lowering of /e/ to /a/ before /r/ is widely attested in the history of English and can be termed SERVE-lowering as it occurred in this common word (Hickey 2014: 279-280). In some cases there has been an orthographic adjustment so that certain instances, like barn (< ME bern), dark (< ME derk), harken (< ME herken) or marsh (< ME mersh) are no longer obvious. In England such lowering is recognisable from county names, such as Derbyshire, Berkshire, Hertfordshire (Ekwall 1975: 27). With common nouns, those instances which were not adjusted in spelling have been reversed, with the exception of clerk /kl<ck/ (Lass 1987: 277). Dialects in England retain this lowering in conservative pronunciation to the present day, e.g. in East Anglia, cf. har (= her), garl (= girl), etc. (Trudgill 2002: 37; 2008: 183). Historically, there is evidence of a much wider distribution of the lowering before /f/, e.g. in the South-West (Wakelin 1988: 628).

There are a few cases of this lowering in standard English where the original and the lowered form are found, e.g. thresh and thrash; wreck and rack in the phrase to rack and ruin (in these cases the lowering is after /f/). This type of lowering in Irish English has practically disappeared although it used to be very widespread (Hickey 2008), though it may still be found occasionally in rural forms of English in Ulster, e.g. nerves [narvz].

Vowel epenthesis in heavy syllable codas, i.e. those consisting of two sonorants, is generally associated with Irish English which shows it in unregistered instances such as film [filqm]. But there is textual evidence that this feature was widespread in England in previous centuries. Shakespeare, an author from the West of England, has in his plays no fewer than 114 instances of alarm (from Old French alarum), all written alarum [al<rm], showing schwa epenthesis in the sonorant coda /rm/.

Grammatical gender has long ceased to exist in English though masculine and feminine pronouns can be used for reference to inanimate objects, e.g. She’s a beaut, the new car. Occasionally, there are dialect forms which are derived historically from gender-distinctive pronouns, e.g. /@/d for ‘him/it’ in southwest England deriving from the masculine accusative of Old English hine (Wakelin 1984: 81), e.g. Tom saw un (= ‘him’). Also in this area one can find gendered pronouns used for inanimate objects, e.g. Pass the loaf – he’s over there (Wagner 2012, Ihalainen 1994).

1.5 Diagnostic value of features

The diagnostic value of vernacular features for considerations of areality depends on their relative unusualness. If a feature is statistically uncommon and nonetheless occurs in geographically adjacent languages or varieties then there is a greater likelihood that this arose through the influence of one language on the other. For instance, in colloquial forms of English the number of consonants is reduced in word-final position, e.g. task [tæs], best [bEs] (Schreier 2005: 126-197). Such deletions are normally sensitive to style

4 This situation is different in principle from the passive knowledge of thou/thee, present with many speakers from religious contexts.
with the greatest amount evident in informal registers. There are frequently specific conditions for deletion, e.g. in post-nasal, or more generally in post-sonorant, positions, e.g. *pound* [paUn]; *field* [fi:l], *hard* [ha:r]. This may also occur word-internally, e.g. *twenty* [tweni], *plenty* [plen]. The deletion may apply only to voiced consonants, e.g. *bold* [boUl] but *bolt* [boUlt]. Clusters consisting of three elements are commonly reduced to two, e.g. *facts* [f&S] or to a single long consonant if a cluster has two similar elements separated by a third, *tests* [tes;]. But these developments are very common and are of little or no assistance in determining feature areality.

2 Language change in Britain and Ireland

The extensive documentation of English shows that some changes, which have occurred in the last century or so and which are, in some locations, still taking place at present, form part of long term developments which have roots deep in the history of the language. For instance, the loss a voice distinction for approximants (Hickey 1984b) due to the voicing of the labio-velar [W] <wh> is of late modern origin (Jespersen 1940 [1909]: 374-375). Historically, word pairs like *which* [Wht] and *witch* [wlt] were distinguished consistently. This is just the final stage in the loss of voiceless sonorants in the history of English, cf. the voiceless nasals and liquids of Old English which were lost in Middle English, e.g. *hnutu > nut(u)* 'nut', *hlaf > laf* 'loaf', *Hrothgar > Roger*.

The increasing urbanisation of Britain in the past two centuries or so has led to different patterns of language change which need to be borne in mind for areal studies. The main one is cascade diffusion which takes place by going from one urban centre to another without necessarily encompassing the intervening countryside. An instance is the spread of TH-fronting to urban centres around Britain which are far from London without the rural areas in between being affected.5

2.1 Internally-motivated change

Areal features generally arise as a result of language contact and are instances of externally-motivated change (Hickey 2012b). But there are also instances of internally-motivated change which can arise and then spread areally in which case they would be relevant to the present study. An example would be analogical change where one element changes to another on the basis of a similar pattern which already exists, the latter providing the model for the change, e.g. a plural *fishes* on the basis of *dish : dishes*. Another instance would be the vernacular plural *youse* [ju:z] formed by adding the plural morpheme {S} to *you*. The similarity in such cases is usually phonetic. There are also instances of analogy from syntax, for instance where a regular structural pattern provides a model. In Irish English positive epistemic *must* provided the model for the negative by the addition of a negator, i.e. *mustn’t* means ‘it cannot be the case’, as in *He mustn’t be Welsh* ‘He can’t be Welsh’. Analogical change may have the effect of masking earlier

---

5 The second type of diffusion is counterhierarchical and involves the spread of features from a rural to an urban setting. It is labelled ‘counterhierarchical’ because this is the opposite of what usually happens. A few cases exist such as the spread of rural *fixin’ to into urban areas of Oklahoma.
changes in a language and must always be considered when reconstructing historical forms.

The alternation of two or more forms in a variety to render their distribution more regular is another type of internally-motivated change, e.g. the use of *were* for the entire past of *be*, or the distribution of *was* for positive statements and *were* for negative ones. Anderwald (2002: 17) offers a cognitive explanation for the remorphologisation according to polarity, i.e. it offers marking of negation via the verb form, additionally to the negator *not*.

2.2 Differing degrees of manifestation

Areality may be characterised not only by the presence or absence of features but by the degree to which a feature is present. Consider the widespread phenomenon of lenition, the weakening of consonants under certain phonetic conditions.

Table 1. Types of consonant lenition

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1| Degemination  
V CC [voiceless] V,# > V C [voiceless] V,# |
| 2| Fricativisation  
V C [voiceless] V,# > V F [voiceless] V,# |
| 3| Tapping  
| 4| Debuccalisation  
V C [alveolar] V,# > V C [glottal] V,# |
| 5| Voicing  
V C [voiceless] V,# > V C [voiced] V,# |
| 6| Deletion  
V C V,# > V Ø V,# |

Lenition is a weakening in articulation, usually leading to a change in a consonant from voiceless to voiced, from stop to fricative, from oral to glottal stop, or to the vocalisation of a consonant entirely. The following are two types of lenition attested in different parts of Britain and Ireland.

1) T-glottalisation, a feature of many vernaculars in Britain whereby /t/, outside word-initial position, is preferentially realised as [ʔ], e.g. *putty* [pWt], *put* [pUʔ]. There may be phonotactic conditioning, e.g. [ʔ] may occur post-vocalically/ pre-pausally but not intervocally, e.g. *putty* [pWt], *put* [pUʔ]. This distribution is typical of colloquial varieties intermediate between Received Pronunciation and Cockney. T-glottalisation is also found in Scots, Ulster Scots and local Dublin English. Given that the process of debuccalisation, i.e. the removal of the oral gesture for stops, is common, the diagnostic value of T-glottalisation in areal studies is very slight.
2) Stop fricativisation is a type of lenition which leads to a change in articulation but with retention of the oral gesture. In traditional Liverpool English lenition is widespread with voiceless stops /p, t, k/ realised as corresponding fricatives, i.e. [f, ɣ, x] (Honeybone 2007; Watson and Clark 2016). This kind of lenition, though only for alveolars, is also typical of Irish English (Hickey 2009).

In its manifestation type (2) above is quite different from type (1) and so its appearance in three locations in Britain and Ireland, which are historically connected with each other, would suggest a contact explanation. It is known that in the nineteenth century there was considerable Irish input to both Liverpool and Middlesbrough so that it would appear to be no coincidence that vernacular speech in these cities should show this type of lenition given that it has existed in Ireland for centuries (Hickey 2009). However likely it is that the Irish input was responsible for lenition in Liverpool and Middlesbrough, this only provided an impetus for a development which later continued on a trajectory somewhat different from Irish English with the generalisation of lenition from /t/ to all voiceless stops, hence its occurrence with /p/ and /k/.

2.3 Constraint hierarchies

A well-researched feature of non-standard English syntax is the ‘Northern Subject Rule’. Basically this is a co-occurrence pattern involving verbs and preceding subjects: -s is absent when the verb is immediately preceded by a pronoun, but not otherwise. An illustration would be We meet and talks together. The occurrence of –s is also favoured by preceding nouns as in The workers gets extra time-off at Christmas, possibly a priming effect due to the plural -s on the noun. This agreement pattern is well-attested in northern Middle English and Middle Scots. In middle and southern England the distribution is uneven (Klemola 2000: 336). East Anglia favours zero marking on all persons, while other southern dialects tend to show a freer use of -s with adjacent pronominal subjects than in the north.

The areal distribution of a feature or construction can be determined by a variety of factors. One of these is the hierarchy which specifies the order in which certain forms are likely to occur. Consider in this context the use of suffixal -s outside the third person singular, a prominent characteristic of many vernacular varieties of English. These vary in the extent to which the -s occurs. At least three determining factors can be recognised and placed in a hierarchy: 1) person and number of verb, 2) relative weight of subject (pronoun, noun, noun phrase), 3) syntactic distance of subject from verb. The ordering in terms of likelihood of occurrence varies and parallels in a constraint hierarchy have been used (Poplack and Tagliamonte 2001) to demonstrate the historical relatedness of varieties. In some varieties the factors just listed do not seem to cause different inflectional behaviour, e.g. in Tyneside singular concord is found with third person plurals, irrespective of the factors (1) - (3), see Beal (1993: 194). This also applies to forms of southern Irish English, particularly on the east coast, but in the north the bare plural pronoun does not use inflectional -s (Harris 1993: 155). Lack of concord can also apply to the past in the case of the auxiliaries have and be: They was going to buy the house.

Various proposals have been put forward concerning the origin of non-standard
verbal concord. Klemola (2000) considered the possibility of influence on English dialects from northern forms of Brythonic, the language spoken in England by the Celtic population at the time of the Germanic invasions and also spoken in Cumbria, Westmoreland and other parts of the far north. In his opinion this would have led to the distribution of verbal -s just outlined. Such agreement rules are rare cross-linguistically (Klemola 2000: 337). However, in P-Celtic languages, notably in Welsh (but not in Irish) there is an agreement rule whereby plural forms of the present are only used with verbs when the pronoun nhw ‘they’ follows immediately. In all other cases the singular is found (Klemola loc. cit.; Williams 1980: 94-95). In essence this is the Northern Subject Rule: the plural forms of Welsh (maen ‘is’-PL) correspond to the s-less forms of English, the Welsh verbal singular (mae ‘is’-SG) is the equivalent to the s-full forms: Maen nhw’n dysgu Cymraeg ‘They are learning Welsh’, Mae Trevor a Sian yn dysgu Cymraeg ‘Trevor and Sian are learning Welsh’. Klemola stresses that the Northern Subject Rule is most widespread in regions of northern Britain which were bilingual with Brythonic and Anglian in the Old English period.

Internal arguments for the Northern Subject Rule are also possible: the decline in inflectional morphology, which had probably set in during late Old English (masked by the West Saxon koiné), meant that language learners in the Middle English period were faced with difficulties analysing the remaining inflections in some systemic way and must have hit on alternative interpretations. In the northern areas, either close to Brythonic-speaking districts, or indeed bilingual, there would probably have been an embryonic Northern Subject Rule which could then have spread to monolingual English speakers and hence become entrenched. However, the standard wisdom on the matter is that it was originally a feature of Scots which spread into the north of England. Further south, different situations may obtain, e.g. generalised -s may be found as in the south-west (Ihalainen 1994: 214) or there may be no -s at all as in East Anglia, a situation which may be the result of contact with Dutch speakers in this area in previous centuries (Trudgill 1998).

3 The historical dimension

The areality of the English language can be examined with considerable time depth. From the time of the first settlements of Germanic tribes in England there is evidence of contact influence from Celtic given that a large proportion of the Celts switched to English in ensuing generations. This led to certain features spreading from Brythonic and appearing in the textual record of late Old English (Hickey 2012c). The scholarly opinion that Brythonic had a significant effect on the development of English is known as the ‘Celtic hypothesis’ (Filppula and Klemola eds, 2009; Filppula and Klemola 2012). Supporters of this hypothesis criticise the view that because there are only a few loanwords from Celtic there was no other influence (Filppula, Klemola and Paulasto 2008: 25). However, if contact persists over many generations, then the substrate can have a gradual and imperceptible influence on the superstrate, leading in some cases to grammatical change at a later time. This can be termed ‘delayed effect contact’ (Hickey 2010) and may well be the source of syntactic features in English which the latter has in common with Celtic (Poussa 1990, Hickey 1995, Vennemann 2002). If speakers of two languages live in close proximity then child language learners in one group can pick up features from the other
group, irrespective of which group is superstrate relative to the other. This type of infection through contact applies to speech habits, such as phonetic realisations, prosodic patterns, alternative exponents of identical grammatical categories alongside the use of lexical items from the other group.

Consider that in the fifth and sixth centuries in England the Germanic invaders probably shared the same environment with most of the Celts who they had subjugated, at least in the south of England. Several facts point to this. There is no record of a single battle in which the Germanic invaders were victorious over the Celts. The notion that they banished the latter into highland areas in the north and west of England and down to the remoter parts of the south-west is an assumption based on the later distribution of P-Celtic languages in England — Welsh and Cornish (descendants of Brythonic). In addition, Old English wealth meant ‘foreigner’ but also ‘Celt’. The word came to be used for ‘servant, slave’ (cf. wielen/wiln ‘female servant, slave’ with the same root, Holthausen 1974: 393) which apparently indicates the status of the Celts vis à vis the Germanic settlers. Lutz (2009: 239-240), drawing on work by the historian David Pelteret, emphasises that the meaning of ‘servant, slave’ was predominantly used in West Saxon, although they were other words with a similar meaning. This would suggest that the subjugation of the Celts was most marked in the south of England where the concentration of the Germanic settlers, and hence language contact, was greatest. Furthermore, the view that the Celts, left in the south and east of England, would have had to shift to English is subscribed to by scholars investigating the Celtic hypothesis (Lutz 2009: 228). Earlier scholars of English highlighted this fact, e.g. Chadwick (1963) and Tolkien (1963), but their work was regarded as peripheral and was not given consideration in ‘mainstream’ works such as Campbell (1959) or Mitchell and Robinson (2007).

That the majority of the population of sixth- and seventh-century West Saxony consisted of Celts who had shifted to English is evident from the numerical relationship of the Celts to the Germanic tribes in the early Old English period. Estimates vary here: the number of Germanic settlers during the fifth century has been put at anything between 10,000 and 200,000 (Filppula, Klemola and Paulasto 2008: 15). But given a population of Britain just before the Germanic arrivals of approximately one million then the relationship would have been anything from 1:100 to 1:5. The latter ratio is hardly likely as it would have implied a huge movement from the continental North Sea rim to England. A figure somewhere in the middle, say 1:20, would still imply that the Celts greatly outnumbered the Germanic settlers, assumed to be about 50,000 with this ratio. In the generations following the initial Germanic settlements, most Celts in contact with the new settlers would have given up their native Brythonic, speaking shift varieties of the newcomers’ dialects during the transition.

These issues are discussed in detail in Trudgill (2010) where the concern is primarily to account for grammatical simplification which is apparent in the later Old English period. Trudgill refers (2010: 29) to features traced to contact by other authors (many of the features are discussed in the present chapter) but returns (2010: 30-35) to the issue of simplification. He sees the contact between Old English and Brythonic speakers are having been most intense and prolonged in the North and that, along with the contact with Old Norse, it may have combined to make Northern English of the Middle Ages the most innovative form in England.
3.1 The rise of the internal possessor construction

English is the only Germanic language which categorically uses possessive pronouns with instances of inalienable possession, e.g. *My head is sore (contrast with German Mir tut der Kopf weh lit. ‘me.DATIVE does the head hurt’). Structures like *The head is sore to me are not possible in present-day English. In Old English the dative of a personal pronoun was found with the head of the noun phrase accompanied by a determiner. In the Old English Poem of Judith one can see this use:

(1)  þæt him þæt heafod wand
    forð on ða flore
    lit: ‘that him-DATIVE that head ...’
    ‘that his head rolled forth on the floor’

Such structures are labelled ‘external possessor’ constructions because possession is expressed by an oblique case pronoun which is outside the semantically related noun phrase. Nowadays the possessive element is internal to the noun phrase – a determiner modifying the head noun – hence the expression ‘internal possessor construction’.

Why did English develop this construction? Consider that external possessor constructions apply in most European languages. Usually they involve a dative-like case for the possessor (in function and commonly in form as well, Haspelmath 1999: 110-111). This is an areal feature in Europe (Haspelmath 1999: 116-117) for, while most languages with this feature are Indo-European, some of these are from outside this family, such as Basque, Hungarian and Maltese.

Remarkably, English, Welsh and Irish have NP-internal possessors as the norm. Put in areal terms, one has internal possessor constructions in Britain and Ireland and generally external possessor constructions in continental Europe. Only in the extreme south-east are internal possessor constructions found widely again, in Turkish and Lezgian, a Caucasian language.

A contact explanation would assume that the use of possessive pronouns in instances like (1) diffused from language shift varieties of the Germanic dialects into inherited varieties and became established there. This transfer did not affect the existence of a possessor construction, but it changed the exponence of the category, a frequent effect of language contact, especially in language shift situations (Hickey 2007: 133-137).

3.2 The twofold paradigm of ‘to be’

West Saxon is known to have had a twofold paradigm of ‘to be’. One paradigm, began with a vowel in the singular and with s- in the plural, and a further paradigm existed with forms in b-; both paradigms were inherited from Indo-European.

---

7 A very small set of verbs expressing physical contact still allow the external possessor, e.g. *He hit him in the face, She patted the child on the head. All such verbs take two objects, the first animate and the second prepositional referring to a part of the body.
Table 2. West Saxon present-tense forms of ‘to be’ (after Campbell 1959: 349)

<table>
<thead>
<tr>
<th>Existential present tense</th>
<th>Habitual present tense</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td><strong>Plural</strong></td>
</tr>
<tr>
<td>1. eom, 2. eart, 3. is</td>
<td>sindon, sint</td>
</tr>
</tbody>
</table>

The Germanic languages, apart from English, have combined the two paradigms to yield just one, cf. German *ich bin, du bist, er ist, wir sind*, etc. The West Saxon double paradigm is thus remarkable in the Germanic context, but not when considering Brythonic with which it co-existed in the early Old English period. Here one also finds two paradigms with a similar syntax and semantics.

Table 3. Middle Welsh present-tense forms of ‘to be’ (Evans 1976: 136)

<table>
<thead>
<tr>
<th>Existential present tense</th>
<th>Habitual present tense</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td><strong>Plural</strong></td>
</tr>
<tr>
<td>1. wyf, oef</td>
<td>ym</td>
</tr>
<tr>
<td>2. wyt</td>
<td>ywch</td>
</tr>
<tr>
<td>3. yw, (y) mae, (y) taw, oes</td>
<td>ynt, (y) maent, y maen</td>
</tr>
</tbody>
</table>

The Old English double paradigm is attested from the eighth century which, as Lutz (2009: 233) rightly highlights, was almost three centuries after the coming of the Germanic settlers. This suggests that both ‘be’ paradigms were entrenched in Old English from earlier transfer, probably by Celtic speakers shifting to the language of the invaders.

3.3 The development of the progressive

There are basically three views on the development of the progressive in English: (i) it developed independently (Visser 1963-73), (ii) it arose under Latin influence, perhaps via French (Mossé 1938), and (iii) it resulted from contact with Celtic (Keller 1925, Dal 1952, Preußler 1956, Wagner 1959, Braaten 1967).

---

8 The influence of Latin on languages in the British Isles has been treated variously in the literature, cf. Genee (2005) for possible influence of Latin on Old Irish.
A type of progressive structure in which a gerund existed sporadically in Old English: *ic wæs on huntunge* ‘I was hunting’ (Braaten 1967: 173). This also occurs in vernacular German, with an infinitive, as in *Ich bin am Schreiben* [I am at write.INFINITIVE] ‘I am writing’ and in Dutch *Hij was aan het schrijven* [he was on the write.INFINITIVE].

In this context a further, typological consideration is necessary (Mittendorf and Poppe 2000: 120-122). Progressive aspect is frequently expressed — in many unrelated languages — via a locative structure meaning to be ‘at’ or ‘in’ an activity. Furthermore, the step from structures like *ic wæs on huntunge* to *I was hunting* is small, involving only the deletion of the preposition and the shift of gerund to non-finite verb form. The fully developed progressive appears in Middle English, but the apparent time delay between contact with Brythonic and the later progressive can be attributed to the strong written tradition in Old English (Dal 1952: 113).

The progressive is found in all Celtic languages and can be seen in a Modern Welsh sentence like *Mae John yn torri coed* [is John in cut-VERBAL NOUN wood] ‘John is cutting wood’ (Jones and Thomas 1977: 63). This shows a prepositional expression for the progressive aspect and is structurally parallel to Old English *ic wæs on huntunge*.

In summary one can say that in both Old English and Brythonic the semantic category of progressive existed. Both languages maintained this category; English lost the locative preposition, increasing the syntactic flexibility and range of the structure, perhaps under the supportive influence of Celtic contact. A further point is that with habitual and progressive verb forms both English and Brythonic had aspect structures of a process-oriented nature in contrast to the goal-oriented nature of aspect types in other Germanic languages, such as German, which are largely telic in nature. This general orientation of aspect can be viewed as an areal feature of languages in Britain and Ireland.

### 3.4 The rise of periphrastic *do*

The syntax and semantics of the verb *do* is one of the most researched matters in the history of English and there are several opinions concerning why its development took the course it did (Garrett 1998, Klemola 2002). The consensus among the different views is that English developed a causative use of *do* involving a direct object followed by an infinitive and that this structure was not inherited from earlier forms of (West) Germanic.

(2) *Pe bispoc of Wincestre … did heom cumen þider.*

‘The bishop of Winchester … had them come thither.’

In the course of the Middle English period variants of this structure developed which had no direct object. There is some dispute about whether the object-less structure was causative or not (van der Auwera and Genee 2002: 293) but agreement exists about the resulting periphrastic construction involving *do* + lexical verb, a developmental path also found in German and Dutch.

Contact was already appealed to by those supporting the above development, e.g. Ellegård (1953) and Denison (1985), but it is contact with Latin and possibly with French *faire*. For other accounts geography plays a role. While causative uses of *do* are common in eastern texts, periphrastic uses were first observed in western texts. In the east (and
south-east) of England French survived longest and in the west (and south-west) Celtic was spoken most widely (Welsh and Cornish, both from Brythonic).

Although it is difficult to provide a cast-iron case for a Celtic origin of English periphrastic *do*, one can mention the line of argument which claims that contact situations do tend to give rise to auxiliaries and so high-contact areas of England (west/south-west) would be where periphrastic *do* would be expected to surface first (Poussa 1990: 412; Tristram 1997: 415). Furthermore, *do* can be used with nouns and so does not require that speakers know sets of corresponding lexical verbs. In a language contact situation, consisting in its early stages of adults shifting from Celtic, this device would improve communication despite being inflectionally less complex. Furthermore, this use of *do* + noun as equivalent to a lexical verb is an established feature of Celtic, e.g. Modern Irish *Rinne sí iarracht é a sheachaint*, lit. ‘did she try-Noun him to avoid’. In addition, using *do* for emphasis may well have been part of the pragmatic mode of adult Celtic second language learners. With the removal of stress from *do* the periphrastic use would have remained.

The Celtic-English contact situation involves the issue of directionality. It is not certain that Celtic first had periphrastic *do* and thus supplied the model for English. So the question of which direction the transfer went is open. Tristram (1997) assumes mutual influence and views contact as an areal phenomenon with bidirectional transfer. There are other structures in Celtic and English which are now areal features, e.g. the internal possessor construction (see above), the progressive and habitual aspect (in vernacular varieties of English). However, the use of *do* as a type of auxiliary is found in other West Germanic languages so that there could well have been language-internal input involved in the rise of periphrastic *do*.

Bearing all such factors in mind can help avoid monocausal contact assumptions such as that by McWhorter who has recently suggested that periphrastic *do* came from Cornish (McWhorter 2009: 168). This single-view approach ignores the frequent cases in West Germanic languages where *do* is a tense carrier used with a lexical verb in non-finite form, e.g. in West German and Low German dialects (van der Auwera and Genee 2002: 286-288) as in *Ich tue dir das morgen bringen*, lit. ‘I do you.DATIVE that tomorrow bring’. The inherent semantics of *do* as a verb denoting direct action also makes it a likely candidate to be employed as tense carrier or aspect marker.

3.6 Dental fricatives in the history of English

A central part of the Germanic Sound Shift is the change of a strongly aspirated /th/ into a dental fricative /θ/, e.g. *thin* /θin/ from an earlier *thin*, in stressed onsets (not preceded by /s/). The individual Germanic languages all lost this fricative later, except Gothic, Icelandic and English.9 Gothic did not survive long enough to be relevant here. However, both present-day Icelandic and English have dental fricatives from Germanic. But for different reasons: Icelandic has changed little over time, so inertia is the major force maintaining dental fricatives. English, however, has experienced great phonological change over the centuries, e.g. it has lost consonantal length, has acquired phonemic voiced fricatives and has developed contrastive word stress under Romance influence. So

---

9 Danish developed a voiced dental fricative through the lenition of /d/ seen in *mad /d/ ‘food’*, for instance.
why does a language with so much phonological change still show dental fricatives? Especially given that these are relatively rare cross-linguistically: the friction of dental fricatives is much less prominent than with /s/, for instance.

A contributory factor could be the existence of dental fricatives in Brythonic (still found in Welsh). Assuming that much of the later Old English population consisted of Celts who had shifted from Brythonic, dental fricatives would have been natural to them. This could be one of the reasons why Welsh, English and Scots still have dental fricatives. Contact as a contributory factor can also be appealed to when considering the loss of front rounded vowels in English. The Celts who shifted to English would not have had front rounded vowels from Brythonic and would in all likelihood not have used rounded realisations of Old English rounded vowels (which arose through i-umlaut), much as say Slavic speakers of German today do not have rounded realisations of /y, ø/.

3.7 Assessment of early contact with Celtic

The contact between speakers of Brythonic and those of Germanic dialects taken from the continent to England as of the fifth century may well have led to the transfer of features from the former to the latter (Ahlqvist 2010). In addition, such contact may have played a supportive role in the maintenance of features in the input Germanic dialects as these gradually evolved into forms of later Old English. Without wishing to overstate the case for contact one can list seven features whose existence in English may possibly be related to the contact with British Celts and to the language shift which large numbers of this section of the population underwent in the generations following the initial Germanic invasion and settlement.

Table 4. Possible transfer features from Celtic (Brythonic) to English

<table>
<thead>
<tr>
<th>Feature</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Internal possessor construction</td>
<td>Morphosyntactic</td>
</tr>
<tr>
<td>2. Twofold paradigm of ‘to be’</td>
<td></td>
</tr>
<tr>
<td>3. Progressive verb forms</td>
<td></td>
</tr>
<tr>
<td>4. Periphrastic ‘do’</td>
<td></td>
</tr>
<tr>
<td>5. Northern subject rule</td>
<td></td>
</tr>
<tr>
<td>6. Retention of dental fricatives</td>
<td>Phonological</td>
</tr>
<tr>
<td>7. Loss of front-rounded vowels</td>
<td></td>
</tr>
</tbody>
</table>

These features vary in the extent and the timing of their documentation and listing them in a single table does not imply that they have an equal status as transfer features. Furthermore, for features (3) to (5) there are cogent internal arguments for their development so that contact can have provided support for their development but can hardly have been the sole origin.

Paradoxically, the consideration of internal factors may strengthen the case for contact accounts as the operation of two factors in principle provides more evidence for the genesis of specific features. So while contact alone may, in an instance like the Northern Subject Rule, be viewed by scholars as too weak a source, the additional reinterpretation of decaying present-tense verbal inflections in medieval English by
language learners offers, in tandem with possible transfer from Brythonic, a more plausible scenario for the rise of this feature.

4 Historical remnants in low-contact areas

The regions of Britain and Ireland can be characterised according to the amount of language contact they have experienced in history (Kortmann and Schröter, this volume; Klemola 2013). The low-contact areas of England are typically those of central and southern England, along with the North of England after the Viking era. The east of England is generally a low-contact area, but East Anglia did experience considerable contact with Low Country inhabitants in Norwich in previous centuries (Trudgill 1998; Goose et al., eds, 2005). In the north of Britain, Scotland presents a mixed picture. The Lowlands and the Borders region have been low-contact areas since the Viking area, much as Northern England. The north of Scotland and the islands off the west coast (the Hebrides) as well as those off the north-east (Orkney and Shetland) have had sustained contact, mostly with Scottish Gaelic for the former and Old Norse (later in the form of Norn, the variety spoken on Orkney and Shetland) in the latter case.

The central lowland region of Scotland is the core area of Scots and shows many historical developments not shared with varieties of English to the south. The morphology of Scots, e.g. that of modals and negators, has developed from northern forms of the Anglian dialect of Old English. In its phonology, Scots is noted for an absence of phonemic vowel length, unusual in the context of first-language varieties of English. In varieties of Scots (McClure 1994: 50-51), Scottish English (McClure 1994: 80-81) or Ulster Scots (Harris 1984: 119-123), vowel length may be similar in word pairs like full and fool, both [fɔl]. Despite the absence of contrastive vowel length in word pairs such as that just cited, the varieties in question do have long and short vowels. However, their occurrence is determined by a nature of the syllable-coda consonant. The constraints in question are known collectively as the Scottish Vowel Length Rule (Maguire 2012: 59). The rule specifies that in stressed syllables all vowels before /r, v, D, z, g/ before another vowel and before a morpheme boundary are long. In other environments the vowels are generally short. Diphthongs also vary in their quality according to the rule, e.g. sight has a raised onset while size has a lowered and lengthened one.11

Archaic features of morphology are also found in Scotland (and rural Ulster). For instance, there are remnants of a nasal plural, the older weak plural in -n, above all in Scots and Ulster Scots (Montgomery and Gregg 1997), e.g. with eyen < eyes (Burchfield 1994: 9) or /ʃu, ʃu/ for shoes in Scots (McClure 1994: 69).

Moving south to Northern England one finds such features as definite article reduction (DAR, Jones 2002; Rupp and Page-Verhoeff 2005; Britain 2012: 26) in which the schwa in the /D@/ is lost leaving either (i) the initial interdental fricative, (ii) a glottalised fricative or (iii) a glottal stop. There is historical evidence that this was also typical of West Country accents further south again, e.g. Shakespeare displays the feature

---

10 For a general treatment of this complex in England, see Kortmann, Herrmann, Pietsch and Wagner (2005); Hernández, Kolbe and Schulz (2011).
11 This is sometimes called Aitken’s Law in honour of the Scottish linguist A. J. Aitken who was the first to describe the rule in detail. See Aitken (1981).
abundantly. It is indicated in writing as th’ or ’t (Hickey 2015: 12).

In the south of England a prominent phonological feature is initial fricative voicing, the change of a voiceless fricative at the beginning of a word to a voiced one. This happened in the early history of southern English dialects, yielding pronunciations such as *say [zel] shilling [ZIlIN], father [vA;D@] (Wakelin 1986, 1988).

The south and south-east of England are known for the widespread occurrence of zero pronouns with subject reference. While this is found with object reference in standard English, e.g. *The woman – he knows has come, it is legal in the south/south-east of England when the referent is the subject as in *The woman – lives here has come; That’s the woman – taught me. This is a feature of London and Home County English (Edwards 1993: 228-229).

In the south-west of England pre-nasal Z-stopping is found whereby /z/ is realised as [d] when it occurs before /n/, typically in the contracted verb forms *isn’t [IdnT] and *wasn’t [wQdnT] but possibly also in nouns like *business [bldn@]. This feature also occurs in south-east Irish English and some varieties of southern American English (Thomas 2008: 109) and in both cases it would appear to be due to contact by emigration. Although the south-west of England has a Celtic contact past the origin of pre-nasal Z-stopping does not seem traceable to transfer from Cornish.

South-western dialects of English also evince the rare phenomenon of pronoun exchange which involves the use of a subject pronoun in object position or that which would demand the oblique case, e.g. *Well, if I didn’t know they, they knewed I (Wagner 2012). This feature is generally thought to stem from an emphatic usage of the pronouns. The varieties which show this usage (Devon, part of Cornwall and a small part of west Somerset) can also have object pronouns in subject position, though this is less common, Him isn’t coming today. Pronoun exchange is found to a limited extent in other vernaculars, e.g. in Northern English for the first person plural, e.g. You can come with we to that as well; Us’ll do it. (Beal 2008: 377). The use of us together with a subject noun is more common colloquially, e.g. Us Irish often work abroad.

5 Possible Celtic areality

The features discussed in the previous section are, so to speak, a negative image of contact-induced areality given that they arose in situations of little or no contact and so must have been internal to the relevant speech community. For the present section a number of features are to be discussed which show noticeable occurrence patterns in the Celtic regions of Britain and Ireland12. These can be possibly traced to structural transfer (Winford 2005) from the source Celtic language to target English during the historical language shift. This shift applies13 to both that in Ireland and that of the west of Scotland, especially the Hebrides where a shift from Scottish Gaelic to English took place.

---

12 For a consideration of shared features among different forms of Insular Celtic and their areal significance, see Matasović (2007).

13 For reasons of space the influence of Welsh, and possibly Cornish, on English over the centuries has not been considered in this chapter. These influences have been slight by comparison with those in Ireland and western Scotland, see Penhallurick (2008a, b).
and is still taking place in a manner similar to that in Ireland from Irish to English.14

5.1 Overuse of definite article

In varieties of English which historically have been in contact with Celtic languages the definite article is found in generic senses. This usage may be extended to those varieties which in turn have been in contact with Celtic varieties of English (Harris 1993: 144-145).15 The following is a brief list of contexts in which the definite article appears contrary to more standard usage: 1) Generic statements The life there is hard. 2) Institutions: She’s gone to the hospital. The young ones are going to the school already. 3) Diseases: The child has got the measles. 4) Seasons: We left in the spring. Additionally, there are further contexts in which the definite article is found in highly vernacular varieties, above all in Ireland: 5) Quantifiers: He asked the both of them. 6) Abstract nouns, including languages and objects of study: Well, I think she likes the languages. 7) Parts of the body, afflictions: There’s nothing done by the hand anymore. It nearly broke the leg on me. I always had problems with the ol’ back. 8) Relatives, spouses, in-laws: Go in now to see the mother. 9) Days of the week, months, seasons, occasions: So we went into town on the Saturday. Well, how did the Christmas go for you? Wagner (2008: 418-419) suggests that the overuse of the definite article in the South-West of England could be due to Celtic influence.

5.2 Cleft sentences

These sentences are used for topicalisation purposes and involve moving an element to the left and placing it in a dummy main clause with it as subject, e.g. It’s tomorrow we’re leaving for Spain. Such sentences occur with moderate frequency in standard English but in some varieties, such as Scottish and Irish English (Harris 1993: 175-176), they are much more common. In both Irish and Scottish English the number and kind of topicalised elements is greater than in other forms of English (It’s to Dublin he’s gone today. It’s her brother who rang up this morning).

5.3 Subordinating and

---

14 General literature on contact forms of Scottish English, essentially Hebridean English, is difficult to come by, but see Sabban (1982, 1985) and Shuken (1984). Not infrequently the variety is simply not recognised, e.g. in the comprehensive overview volume, Kortmann and Lunkenheimer (eds, 2012), where all varieties of English in Scotland are treated together and classified as ‘low-contact’ Kortmann and Lunkenheimer (eds, 2012: 2) and hence end up in Cluster 1a and not Cluster 1c in their NeighborNet diagram Kortmann and Lunkenheimer (eds, 2012: 921).

15 The definite article in forms of American English has been viewed as a legacy of Irish influence (Montgomery 2001: 133; Butters 2001: 337).
This is a syntactic feature of Irish English in which a subordinate clause (usually concessive or restrictive in nature) is introduced by and plus a continuous form of a verb, e.g. They went for a walk and it pouring rain. The structure is probably a calque on Irish, cf. Chuaigh siad ar siúlóid agus é ag cur báistí, lit. ‘Went they on walk and it at putting rain’ (Hickey 2007: 261-265). This structure is also found in Hebridean English, cf. But many’s a time I was along with my auntie on the loom, and her weaving (Filppula 1997: 950).

5.4 Extended now

The structure being referred to by this label (Filppula 1999: 90, 122-128) can be seen in a sentence like I know Brian and Sheila for many years now. Essentially, the present tense is used in contexts where the time span is from some point in the past to the present. In these situations, standard English uses the present perfect, i.e. the sentence just quoted would be I have known Brian and Sheila for many years. The use of the simple present in contexts which conceptually stretch back into the past is a widespread feature of English in the entire island of Ireland and in contact varieties of English in Western Scotland (Sabban 1985: 128; Filppula 1997).

The question of origin is difficult to answer conclusively as Filppula (1999: 123-124) rightly notes. The use of the present in English has a long history and is probably the older Germanic type, still seen in present-day German, e.g. Ich kenne ihn seit mehreren Jahren, lit. ‘I know him since many years’. This type may have continued well into the early modern period and so been present in the input varieties of English during the language shift in Ireland. On the other hand auxiliary have did not, and does not, have a formal equivalent in Irish or Scottish Gaelic and so it is more than likely that Celtic speakers in a language shift situation would have ignored this form. The Irish equivalent to the present perfect is expressed quite differently as seen below.

(3) Tá aithne agam ar Bhrian agus ar Shíle le blianta anuas anois. [is knowledge at-me on Brian and on Sheila with years down now]

The acceptance of extended now in present-day Irish English was tested in A Survey of Irish English Usage (Hickey 2007: 144-145) and the rates were consistently high. There was a bias towards the south of Ireland with Wexford in the south-east scoring the highest value, considerably higher than the Ulster Scots core areas of Antrim and Down in the north-east of the country. This may be due to the very early settlement of the east coast before the present perfect had become established in English.

5.5 Aspectual categories in Celtic Englishes

Both varieties of English and the Celtic languages are known for the aspectual distinctions which they evince. For a discussion of these at least three types of aspect need to be recognised: (i) perfective, which signals that an action is completed; (ii)

---
16 Klemola and Filppula (1992) examined the occurrence of the structure in non-Celtic Englishes and found it to be very rare or non-existent.
progressive, which indicates that an action is on-going and (iii) habitual, which highlights the repeated occurrence of an action. Creoles are noted for having complex aspectual distinctions (Singler ed., 2000) and those which have developed from English have a more nuanced aspectual system than the original input. Second in line, in terms of aspectual complexity, are language shift varieties of English (Hickey 1997; Mesthrie 1992), followed by low-contact varieties which generally show the aspectual distinctions known from standard English today. The following table offers a classification of aspectual types with appropriate subdivisions for varieties of English (see also Kortmann and Schröter, this volume).

### Table 5. Aspectual distinctions for varieties of English

<table>
<thead>
<tr>
<th>Information about action</th>
<th>Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>duration</td>
<td>progressive</td>
</tr>
<tr>
<td>repetition</td>
<td>habitual</td>
</tr>
<tr>
<td>completion</td>
<td>perfective</td>
</tr>
<tr>
<td><strong>Sub-division of perfective</strong></td>
<td><strong>Types</strong></td>
</tr>
<tr>
<td>very recent completion</td>
<td>immediate perfective</td>
</tr>
<tr>
<td>completion of planned action</td>
<td>resultative perfective</td>
</tr>
<tr>
<td><strong>Sub-division of habitual</strong></td>
<td><strong>Types</strong></td>
</tr>
<tr>
<td>repeated shorter action</td>
<td>iterative</td>
</tr>
<tr>
<td>repeated longer action</td>
<td>durative</td>
</tr>
</tbody>
</table>

The progressive is established in all varieties of English, e.g. *She was singing when he arrived home* and is currently expanding (Leech, Hundt, Mair and Smith 2009: 118-143), e.g. encompassing verbs like want, e.g. *I’m wanting to leave that matter be*. Perfective aspect is widespread in the world’s languages and in some, such as the Slavic languages, pairs of perfective and imperfective verbs are found.

#### 5.5.1 Immediate perfective

In Celtic Englishes a distinction is common between an immediate perfective and a resultative perfective. The immediate perfective with *after* is a calque on Irish *tar éis* (Harris 1993: 141) which is used for the same purpose. A similar structure exists in

---

17 In Kortmann and Lunkenheimer (eds, 2012: 915) this perfective is classified as ‘present in’ Irish, Scottish English and Newfoundland English along with occasional occurrences in Sri Lankan English and Palmerston Island English. Szmrecsanyi in the same volume lists the *after* perfective under ‘Features rare or very rare in L1 varieties’ (2012: 828-829). A similar distribution pattern across varieties of English world-wide is found for the resultative perfective, classified in Kortmann and Lunkenheimer (eds, 2012) and elsewhere as the ‘Medial-Object
Hebridean English (Sabban 1982: 155-168; 1985: 134-135; Filppula 1997). The position in Gaelic is considered by Sabban (1982: 162) and she notes that Gaelic has a construction with *air* ‘after, behind’ as in *tha e air buadaladh* [is he after striking] ‘he is after striking’ (Adger 2010). The form *air* used to exist in this function in Irish (*iar*) but was later replaced by *tar éis*/*ndiaidh* (both meaning ‘after’) a long time after the transportation of Irish to Scotland (Hickey 2013b).

5.5.2. Resultative perfective

The resultative perfective indicates that something planned is now completed. In Irish English the word order ‘Object + Past Participle’ is used (Harris 1993: 160; Hickey 2007: 208-213), e.g. *I've the book read* ‘I am finished reading the book’ which contrasts with *I've read the book* ‘I read it once’ (the O+PP word order has a precedent in the history of English but also an equivalent in Irish in which the past participle always follows the object: *Tá an leabhar léite agam* lit.: ‘is the book read at-me’).

5.5.3 Habitual aspect

The habitual exists in English by contrast with the progressive, e.g. *He’s meeting the students* (now) (progressive) versus *He meets the students* (every Thursday morning) (habitual). In addition, many traditional dialects of English have explicit marking of the habitual, with the verb *do*, by using *be* + inflectional *-s* or combinations of these, e.g. *He does be drinking a lot*, *He bees drinking a lot*, *He does drink a lot* (with unstressed *does*). The dialects with an explicit habitual usually occur in Celtic areas, e.g. Ireland or southwest England (former Cornish area), and by extension Newfoundland, a fact which strongly suggests that language contact has played a role in its genesis. There are two types of habitual, one indicating a repeated brief action and one referring to a repeated but longer action. In Irish English, especially of the south-east, an inflectional *-s* expresses the iterative habitual, e.g. *They calls this place City Square*, while *does be / duh [d@ be]* is found for the durative habitual, e.g. *She does be worrying about the children*.

Table 6. Classification of English habitual aspect in Britain and Ireland

<table>
<thead>
<tr>
<th>Structure</th>
<th>Example</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suffixal -s on lexical verb stem</td>
<td><em>I meets my sister on a Friday afternoon.</em></td>
<td>Denotes the iterative habitual in the south and south-east of Ireland.</td>
</tr>
<tr>
<td>Suffixal -s on <em>be</em> or uninflected <em>be</em></td>
<td><em>The men bees at home at the week-ends.</em></td>
<td>Expression of the durative habitual in Northern Irish English.</td>
</tr>
</tbody>
</table>

Perfect’. The classification of both perfectives in the present chapter derives from their function and not their form.
Suffixal -s on do or uninflected do plus be, with the lexical verb in the progressive form

<table>
<thead>
<tr>
<th>Suffixal -s</th>
<th>do or uninflected do</th>
<th>Progressive Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>He does/do</td>
<td>[d@ be]</td>
<td>buying and selling old cars.</td>
</tr>
</tbody>
</table>

Uninflected, unstressed do is found in southern and south-eastern Irish English for the durative habitual.

As indicated in the right-most column above, some varieties, such as vernacular southern Irish English, can distinguish between a durative habitual and an iterative habitual. The former (durative habitual) refers to a repeated action which takes some time and is expressed via do + be + V-ing, e.g. They do be fishing on the lake while the latter (iterative habitual) refers to a short repeated action I calls me little one the baby. (Hickey 2007, Chapter 4).

5.5.4 Category and exponence

When shifting to another language, temporarily or permanently, adults, to begin with, expect the grammatical distinctions in the target which they know from their native language. To this end they search for equivalents in the target to categories they are familiar with (Hickey 2013b). This process is an unconscious one and persists even with speakers who have considerable target language proficiency. If the categories of the outset language are semantically motivated then the search to find an equivalent in the target is all the more obvious. Apart from restructuring elements in the target, speakers can transfer elements from their native language. This transfer of grammatical categories is favoured, if the following conditions apply.

(4)  a. The target language has a formal means of expressing this category.
    b. There is little variation in the expression of this category.
    c. The expression of this category is not homophonous with another one.
    d. The category marker, in the outset language, can be identified and easily extracted from source contexts.

The following table illustrates the formal correspondence of the habitual in Irish and Irish English. In this instance the match is poor and the exponence of this aspectual type is different in both languages (except perhaps type (2) in the left-hand column). However, the syntactic means which early modern English provided – unstressed do + verb and/or verbal inflection – were sufficiently transparent semantically for speakers across the community to grasp the usage and adopt it themselves.

Table 7. Category and exponence in Irish and Irish English

<table>
<thead>
<tr>
<th>Category</th>
<th>Exponent in Irish English</th>
<th>Exponent in Irish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) do(es) be + V-ing</td>
<td>They do be fighting a lot.</td>
<td>bíonn + non-finite verb form</td>
</tr>
<tr>
<td></td>
<td>[is-HAB they at fighting often]</td>
<td></td>
</tr>
<tr>
<td>2) bees (northern)</td>
<td>The lads bees out a lot.</td>
<td>Bíonn siad ag troid go minic.</td>
</tr>
<tr>
<td>3) verbal -s (first person)</td>
<td>I gets tired of waiting for things to change.</td>
<td></td>
</tr>
</tbody>
</table>
Such structures go through a period of mixture and fluidity – much as in New Dialect Formation scenarios – and only gain a clear linguistic profile somewhat later. This means that some features present in the early stages of a language shift variety are not continued by later generations of speakers. Furthermore, a particular feature may itself undergo a development in the shift variety after the latter has become independent of the original source language which may or may not be still extant. It would appear that only those features which have a fairly good structural match between outset and target language, and which are semantically transparent, have a chance of survival in a later focused form of the transfer variety. Take, for instance, the autonomous in Irish. This is a finite verb form – but without a personal pronoun – where relevance is expressed via a preposition, e.g. *Rugadh mac di.* ‘She gave birth to a son.’ [born-PAST son to-her]. This structure is (and was) never transferred to Irish English, probably because of the difficulty of mapping it onto English syntactically.

5.6 Assessment of later contact with Celtic

Present-day varieties of English in Ireland and Scotland, which have arisen through language shift from Irish and Scottish Gaelic respectively, still show significant syntactic features which are probably derived from the Celtic source language in question. The large degree of agreement in transfer features between Ireland and the Hebrides in Scotland testifies to the considerable degree of syntactic areality between Ireland and the West of Scotland (Sabban 1982: 99-116) as can be seen from the following table.18

<table>
<thead>
<tr>
<th>Feature</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitant aspect</td>
<td>Convergence with South-West English input on east coast, possibly with influence from Scots via Ulster. Otherwise transfer of category from Irish.</td>
</tr>
<tr>
<td>Immediate perfective aspect with <em>after</em></td>
<td>Transfer from Irish, Scottish Gaelic</td>
</tr>
<tr>
<td>Resultative perfective with OV word order</td>
<td>Possible convergence with English input, primarily from Irish/Scottish Gaelic</td>
</tr>
<tr>
<td>Subordinating <em>and</em></td>
<td>Transfer from Irish/Scottish Gaelic</td>
</tr>
<tr>
<td>Variant use of suffixal <em>-s</em> in present</td>
<td>South-west input to Ireland in first period on east coast</td>
</tr>
<tr>
<td>Clefting for topicalisation</td>
<td>Transfer from Irish/Scottish Gaelic, with some possible convergence</td>
</tr>
<tr>
<td>Greater range of the present tense</td>
<td>Transfer from Irish/Scottish Gaelic, with some possible convergence</td>
</tr>
</tbody>
</table>

18 There are a few more features common to the Ireland and Western Scotland in this respect, e.g. focus constructions such as those formed with an unbound reflexive pronoun as verbal subject, seen in a sentence like (Irish) *Himself and his wife were buried ... himself was buried twelve months ago* and (Scottish) *And himself and his brother was up in the Orkney Isles* (Filppula 1997: 951).
One significant feature which does not appear in Hebridean English is an explicitly expressed habitual via *do* + verb or via a verbal inflection, as in forms of Irish English (see Table 7 above). This was already noted by Filppula (1997: 952-953) who quotes an explanation by Bliss (1972) concerning a putative relationship between dependent verb forms in Irish and the appearance of *do* in Irish English. Bliss’s suggestion does not bear linguistic scrutiny and is not capable of accounting for the complexities of habitual aspect exponence in varieties of Irish English. It is more likely that during the language shift in Scotland unstressed periphrastic *do* was not available to speakers to anything like the degree that this was true for Irish speakers and so its use as an exponent of a formally marked habitual in contact English in Scotland never reached a sufficiently high level for it to become established in the speech community of the time.

### 6 Overall assessment and conclusion

Areality in Britain and Ireland presents a complex and shifting picture throughout the attested history of language in these islands. The internal interaction of speakers, and the linguistic results this engendered, is separate for both islands until the late twelfth century after which the political, social and linguistic fate of Ireland became inextricably linked with Britain. The influence exerted since the late Middle Ages has been almost exclusively from Britain to Ireland, bar nineteenth-century emigration from Ireland to parts of England and Scotland – Liverpool, Middlesbrough, Glasgow – to mention the most obvious cases, which resulted in some diffusion of Irish features into English in these cities or at least triggered developments which continued after immigration from Ireland ceased, as in the case of Liverpool (Honeybone 2007).

Those features which can be labelled as areal in Britain and Ireland need to be classified at least in terms of the time during which they occurred and the manner in which they arose. The earliest features are those which appeared in later Old English and which were probably the result of shift varieties of English arising among the descendents of speakers of Celtic in the south of England. This interpretation requires a revision of the standard view that the Celts were banished to the highland areas of the west and north of England (Trudgill 2010; Hickey 2012c). Strong features, such as the rise of the internal possessor construction (essentially the use of possessive pronouns in cases of inalienable possession), testify to the areality of Celtic and early English. The later influence of Celtic on English is largely confined to shift varieties of English in Ireland and Scotland. There are significant parallels between later Irish English and Hebridean English in the transfer features from Irish and Scottish Gaelic respectively. These testify to the similar outcome of contact and shift along the western seaboard of Britain and Ireland suggesting an areal spread from the north-west of Scotland to the south-west of Ireland, a high-contact zone contrasting with a low-contact zone in recent centuries reaching from central Scotland down to the south of England.
References


10 Varieties of English

Bernd Kortmann and Verena Schröter

1 Introduction

This chapter is solely concerned with areal patterns in the domain of morphosyntax in varieties of English around the world. Thus it tries to discern areality on that structural level of language where it is notoriously hardest to find; moreover its focus is on areality, i.e. “the areal concentration of linguistic features” (Hickey 2012: 2), on a global (world region) scale, not on a more delimited regional scale as is usual in dialectology or dialectometry. It adds to the line of research pursued in, e.g., Anderwald (2012), Lunkenheimer (2012), Huber and Brato (2012), Kortmann and Wolk (2012), Kortmann (2013), and various chapters offering synopses of individual world regions published in the Mouton World Atlas of Varieties of English (WAVE; Kortmann/Lunkenheimer 2012a). The data set used here is the one in eWAVE 2.0 (www.ewave-atlas.org), which includes ratings, examples, and interactive maps for 235 morphosyntactic features in 50 L1 and L2 varieties of English as well as 26 English-based pidgins and creoles (see Table 1). The only major difference between the data sets in eWAVE 2.0 and WAVE is that for the former (launched in November 2013) Philippine English and Cape Flats English have been added. For more information on the varieties, their classification into different variety types, the complete feature set, methodology, rating system, informants, etc. see Kortmann and Lunkenheimer (2012b: 1-6) or http://ewave-atlas.org/introduction.

Table 1. 76 L1 and L2 varieties, pidgins and creoles represented in eWAVE 2.0

<table>
<thead>
<tr>
<th>L1 (32)</th>
<th>high-contact L1 (22)</th>
<th>L2 (18)</th>
<th>P (7) &amp; C (19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>low-contact L1 (10))</td>
<td>Irish E (IrE), Welsh E (WelE), Manx E (ManxE), Channel Islands E (ChlsE), Maltese E (MltE)</td>
<td>Liberian Settler E (LibSE), White South African E (WhSAIE), White Zimbabwean E (WhZmE), Cape Flats English (CFE)</td>
<td>Ghanian E (GhE), Nigerian E (NigE), Cameroonian E (CamE), Kenyan E (KenE), Tanzanian E (TznE), Ugandan E (UgE), South African E (BISAIE), Indian South African E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>British Isles (11):</th>
<th>Irish E (IrE), Welsh E (WelE), Manx E (ManxE), Channel Islands E (ChlsE), Maltese E (MltE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orkney and Shetland E (O&amp;SE), North of England (North), SW of England (SW), SE of England (SE), East Anglia (EA), Scottish E (ScE)</td>
<td>Liberian Settler E (LibSE), White South African E (WhSAIE), White Zimbabwean E (WhZmE), Cape Flats English (CFE)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Africa (17):</th>
<th>Liberian Settler E (LibSE), White South African E (WhSAIE), White Zimbabwean E (WhZmE), Cape Flats English (CFE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ghanaian E (GhE), Nigerian E (NigE), Cameroonian E (CamE), Kenyan E (KenE), Tanzanian E (TznE), Ugandan E (UgE), South African E (BISAIE), Indian South African E</td>
<td>Ghanian Pidgin (Ghp), Nigerian Pidgin (NigP), Cameroonian Pidgin (CamP), Krio, Vernacular Liberian E (VLibE)</td>
</tr>
<tr>
<td>Region</td>
<td>Languages</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>America (10):</td>
<td>Newfoundland E (NfldE), Appalachian E (AppE), Ozark E (OzE), Southeast</td>
</tr>
<tr>
<td></td>
<td>American Enclave dialects (SEAmE)</td>
</tr>
<tr>
<td></td>
<td>Colloquial American E (CollAmE), Urban</td>
</tr>
<tr>
<td></td>
<td>African American Vernacular E (UAAVE), Rural</td>
</tr>
<tr>
<td></td>
<td>African American Vernacular E (RAAVE), Earlier</td>
</tr>
<tr>
<td></td>
<td>African American Vernacular E (EAAVE)</td>
</tr>
<tr>
<td></td>
<td>Chicano E (ChcE)</td>
</tr>
<tr>
<td></td>
<td>Gullah</td>
</tr>
<tr>
<td>Caribbean (13):</td>
<td>Bahamian E (BahE)</td>
</tr>
<tr>
<td></td>
<td>Jamaican E (JamE)</td>
</tr>
<tr>
<td></td>
<td>Jamaican C (JamC), Bahamian C (BahC), Barbadian C (Bajan)</td>
</tr>
<tr>
<td></td>
<td>Belizean C (BelC), Trinidadian C (TrinC), Eastern</td>
</tr>
<tr>
<td></td>
<td>Maroon C (EMarC), Sranan, Saramaccan (Saram), Guyanese C (GuyC), San</td>
</tr>
<tr>
<td></td>
<td>Andrés C (SanAC), Vincentian C (VinC)</td>
</tr>
<tr>
<td>South and</td>
<td>Colloquial Singapore E (CollSgE)</td>
</tr>
<tr>
<td>Southeast Asia</td>
<td>Indian E (IndE), Pakistan E (PakE), Sri Lanka E (SLkE), Hong Kong E</td>
</tr>
<tr>
<td>(8)</td>
<td>(HKE), Malaysian E (MaIE), Philippine E (PhilE)</td>
</tr>
<tr>
<td></td>
<td>Butler E (ButlE)</td>
</tr>
<tr>
<td>Australia (5):</td>
<td>Aboriginal E (AbE), Australian E (AusE), Australian Vernacular E (AusVE)</td>
</tr>
<tr>
<td></td>
<td>Torres Strait C (TorSC), Roper River C (RRC Krioll)</td>
</tr>
<tr>
<td>Pacific (8):</td>
<td>New Zealand E (NZE)</td>
</tr>
<tr>
<td></td>
<td>Colloquial Fiji E (CollFijiE), Acrolectal Fiji E (FijiE)</td>
</tr>
<tr>
<td></td>
<td>Hawaiian C (HawC), Bilama (Bisi), Norfolk Island/Pitcairn E (Norf'k),</td>
</tr>
<tr>
<td></td>
<td>Palmerston E (PalmE), Tok Pisin (TP)</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>St. Helena E (SIHE), Tristan da Cunha E (TdCE), Falkland Islands E (FkE)</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
</tr>
</tbody>
</table>
The Anglophone world regions considered in this chapter are, in alphabetical order, Africa, America, Asia (South and Southeast), Australia, the British Isles, the Caribbean, and the Pacific. (Given the extreme geographical dispersal and isolation of the relevant varieties – Falkland English, St Helena English, and Tristan da Cunha English – it would not make sense to also include in our discussion the eighth world region in WAVE, South Atlantic). As argued in a range of previous publications by the first author, most recently in Kortmann and Wolk (2012: 920-927), the overall morphosyntactic (or: typological) profiles of varieties of English and English-based Pidgins and Creoles primarily pattern according to variety type (high-/low-contact L1, L2, Pidgin, Creole) and not according to the relevant world region where they are spoken. In other words, the socio-historical conditions in which a given variety emerged have a far stronger impact on the morphosyntactic feature set constituting the typological profile of that variety than geography. This does not mean that geographical (or, alternatively, areal) clusterings cannot be observed, but they are only secondary to the primary patterning of morphosyntactic profiles according to variety type. This is shown very clearly in Figure 1, which offers a NeighborNet diagram (or: phenogram) allowing us to visualize the degree of typological similarity and distance between any pair of the 76 varieties/Pidgins/Creoles considered here, similarity/distance being determined by the number of co-presences and co-absences of morphosyntactic features for any two varieties in the WAVE data set that are compared (for details of the method cf. Kortmann and Wolk 2012: 919-920). As can be seen, the NeighborNet diagram falls into four main clusters, i.e. the members of each of these four clusters share more morphosyntactic properties with each other than with any member of any of the other three clusters. Most importantly, these clusters pattern rather neatly according to variety type: Cluster 1 is the L1 cluster (notably including all low-contact, or: traditional, L1s), Cluster 2 is the L2 cluster, Cluster 4 is the Pidgin and Creole cluster, and even the three main sub-branches of the more diffuse Cluster 3 are fairly coherent as regards variety type: Creoles (3a), high-contact L1s (3b), and so-called creoloids (3c).

However, and this is the starting-point for the present chapter, there are also some areal patterns which can be identified in Figure (1). These, though, can only be determined within each of the four major, variety-type-driven clusters: for example, there are sub-clusters of American varieties to be found in Clusters 1 (Am1) and 3 (Am2), sub-clusters of African varieties in Clusters 2 (Af1, Af2), 3 (Af3) and 4 (Af4), and sub-clusters of Caribbean varieties and creoles in Clusters 3 (Car1) and 4 (Car2, Car3). What is even more welcome for a chapter focusing on areality in a large-scale comparison of the morphosyntactic properties covering most of the variation and varieties in the Anglophone world is the following observation: quite a number of these areal sub-clusters really are formed by varieties spoken in geographically contiguous areas of the relevant world region. Just take the L2 Cluster 2, where the sub-cluster As1 is formed rather neatly by Southeast Asian Englishes; or consider the Pidgin/Creole Cluster 4, where sub-cluster Af4 consists exclusively of the four West African Pidgins/Creoles (Ghanaian Pidgin, Nigerian Pidgin, Cameroon Pidgin plus Krio, the only West African creole) and where its immediate neighbour branch, sub-cluster Car 3, is formed by the three radical creoles spoken in Suriname (Sranan, Eastern Maroon Creole, Saramaccan)\(^1\). But yet again a cautionary note needs to be struck: not all sub-clusters are to be

\(^1\) For a discussion of areal patterns among the three major Anglophone Pidgin and Creole regions (West Africa, Caribbean, Pacific/Australia) see Kortmann (2013: 177-184).
(exclusively) interpreted areally. They may rather document historical links, such as in sub-cluster BrIs1 between Irish English and the English Southwest as donor varieties of Newfoundland English, in sub-cluster BrIs2 between the Southeastern dialects of England and Falkland English, or in sub-cluster Am2 between Earlier AAVE and its modern descendants (Rural and Urban) AAVE. Alternatively, such clusters may at least foster the hypothesis that historical links and directions of influence may have existed between certain varieties (e.g. between Earlier AAVE and Bahamian English). A recent example of this line of research into historical language contact is Eberle and Schreier (2013), who, based on the WAVE data set and their own morphosyntactic survey following the WAVE feature catalogue, explore the “Caribbean connection” for African Bermudian English and selected Caribbean varieties and Creoles.
Figure 1. NeighborNet diagram: Typological and areal clusters of 76 varieties, Pidgins and Creoles in eWAVE 2.0.

Hence the major aim of the present chapter is to make areality an even stronger factor in the interpretation of the WAVE data, focusing NOT on the overall typological profiles of the 76 varieties/Pidgins/Creoles, but on the individual features. The central task will be to identify features, and in some cases feature sets, that are highly distinctive of, possibly even diagnostic for, a given Anglophone world region. In this feature-based approach, it is in section 2 that we will explore each of the twelve domains of grammar from which the 235 WAVE features are taken. The internal structure of that section will follow the order in which the relevant grammar domains are listed in Table 2:

Table 2. Domains of grammar covered in WAVE (235 features in all)

<table>
<thead>
<tr>
<th>Grammatical domain</th>
<th>Feature totals</th>
<th>% of 235</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronouns</td>
<td>47</td>
<td>20.0%</td>
</tr>
<tr>
<td>Noun phrase</td>
<td>40</td>
<td>17.0%</td>
</tr>
<tr>
<td>Tense and aspect</td>
<td>33</td>
<td>14.0%</td>
</tr>
<tr>
<td>Modal verbs</td>
<td>7</td>
<td>3.0%</td>
</tr>
<tr>
<td>Verb morphology</td>
<td>26</td>
<td>11.0%</td>
</tr>
<tr>
<td>Negation</td>
<td>16</td>
<td>6.8%</td>
</tr>
<tr>
<td>Agreement</td>
<td>15</td>
<td>6.4%</td>
</tr>
</tbody>
</table>
While in section 2 we will solely be interested in areality by grammar domain, it is in the synoptic section 3 that the seven Anglophone world regions will take centre stage. More exactly, it will offer generalizations, and possible explanations, for the areal patterns identified in this chapter. Section 4 will briefly address the issue of areality within individual world regions (e.g. between the varieties of the North and the South of the British Isles, between the varieties of English in South and Southeast Asia). The concluding section 5 will put the findings from the preceding sections in perspective, and make suggestions for future research with the promise of a further strengthening of areality as a factor influencing and shaping not only morphosyntactic usage but, ultimately, morphosyntactic systems, typological profiles and, not least, local norms in World Englishes.

2. Top diagnostic features in morphosyntax for individual Anglophone world regions

2.1 Defining the thresholds for areally distinctive and diagnostic features

Most of the findings in this section will be presented in the form of tables, all of which are cut to the same pattern and fairly self-explanatory. For each of the features in a given grammar domain it was determined whether its geographical distribution across the seven world regions shows a certain bias, i.e. whether it is overrepresented in one given world region compared with all others. As thresholds for features finding their way into the following tables we defined (a) the 50% mark, i.e. all of the features must be attested in at least 50% of all the L1/L2 varieties, Pidgins or Creoles in the relevant world region, and (b) the 30% mark, i.e. only those features are listed in section 2 whose attestation rate (AR) in one (and only one) world region exceeds the AR of this feature in the rest of the world (RoW) by at least 30% (calculus: AR region minus AR RoW). As diagnostic for a given world region we consider features where this AR difference reaches or exceeds the 60% mark. The relevant features are marked in bold print. The major ordering principle for each of the tables in this section is by Region (alphabetical order), and within the individual world region in decreasing order by the rightmost column ‘AR diff(ERENCE) region – RoW’. Thus in Table 3 it turns out that WAVE feature F9 (Benefactive ‘personal dative’ construction), as in I got me a new car (Urban AAVE), is a diagnostic feature of America (i.e. in the WAVE data set: US and Newfoundland). Note that for reasons of space no maps will be used in order to show geographical distribution and, especially, bias. For many of the features discussed below maps will be found in Kortmann and Lunkenheimer 2012 (see especially pp. xv-xx for the map survey ordered

---

2 Since F9 is attested in 10/10 North American varieties (AR region 100%) and in 23/66 varieties outside this region (AR RoW 34.85%), this results in a difference in attestation rate (AR diff) of 65.15%.
by feature numbers). In general, maps for every single feature can easily be produced with a few clicks by means of eWAVE (www.ewave-atlas.org).

While the attestation rates are based on the mere presence vs. absence of features, the regional pervasiveness (rp-score) of a feature is captured by integrating its relative frequency in a given region based on the WAVE rating system. The rp-score of individual features is calculated as follows: a numerical value is assigned to the WAVE-ratings A (pervasive/obligatory) = 1, B (neither pervasive nor rare) = 0.6, C (attested but rare) = 0.3, D (not attested), X (not applicable), ? (unknown) = 0. The sum of these is divided by the number of varieties in the respective region. With an rp-score of 0.73, F9 is not only a diagnostic, but also a pervasive feature of the North American varieties.3

2.2 Areality in a dozen domains of grammar

Let us begin with that grammar domain represented by the largest number of features (47) in WAVE, namely Pronouns. As shown in Tables 3a and 3b, about half of the pronominal features have a pronounced geographical signal. This is especially strong for the Caribbean (for 11 of the 17 features in Table 3a), but very noticeable, too, for the Australian/Pacific region: F36 and F37 are found in more than 60% of the Australian/Pacific varieties, but either nowhere else (!) in the Anglophone world (F37, as in Palmerston English We two is going) or in only one other variety of English (F36, as in Aboriginal English mela new teacher gotta come 'our [exclusive] new teacher will come'). Feature 13 (Subject pronoun forms serving as base for reflexives), as in She say it for sheself (Trinidadian Creole), is characteristic for the entire American and Caribbean area. There are also two areoversals to be observed in Table 3a: F9 (Benefactive 'personal dative' construction) is found in all 10 WAVE varieties of America, and F43 (Subject pronoun drop: referential pronouns, rp-score: 0.76) is strongly attested in all eight WAVE varieties of South and Southeast Asia, as in When I come back from my work___ just travel back to my home (Indian English).

Table 3a. Pronouns: top diagnostic features for individual world regions

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Benefactive “personal dative” construction</td>
<td>Am</td>
<td>100.00%</td>
<td>34.85%</td>
<td>65.15%</td>
</tr>
<tr>
<td>39</td>
<td>Plural forms of interrogative pronouns: using additional elements</td>
<td>Am</td>
<td>70.00%</td>
<td>27.27%</td>
<td>42.73%</td>
</tr>
<tr>
<td>13</td>
<td>Subject pronoun forms serving as base for reflexives</td>
<td>Am, Car</td>
<td>56.52%</td>
<td>20.75%</td>
<td>35.77%</td>
</tr>
<tr>
<td>43</td>
<td>Subject pronoun drop: referential pronouns</td>
<td>As</td>
<td>100.00%</td>
<td>45.59%</td>
<td>54.41%</td>
</tr>
<tr>
<td>40</td>
<td>Plural forms of interrogative pronouns: reduplication</td>
<td>As</td>
<td>50.00%</td>
<td>4.41%</td>
<td>45.59%</td>
</tr>
<tr>
<td>37</td>
<td>More number distinctions in personal pronouns than simply singular vs. plural</td>
<td>Aus, Pac</td>
<td>61.54%</td>
<td>0.00%</td>
<td>61.54%</td>
</tr>
</tbody>
</table>

3 For example in the 10 American varieties, F9 receives 4 A, 5 B and 1 C rating. \((4*1 + 5*0.6 + 1*0.3)/10 = \text{rp-score F9} = 0.73\).
36 Distinct forms for inclusive/exclusive first person non-singular

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>You as (modifying) possessive pronoun</td>
<td>Aus, Pac</td>
<td>92.31%</td>
<td>22.22%</td>
<td>70.09%</td>
</tr>
<tr>
<td>19</td>
<td>Subject pronoun forms as (modifying) possessive pronouns: first person plural</td>
<td>Car</td>
<td>76.92%</td>
<td>14.92%</td>
<td>62.64%</td>
</tr>
<tr>
<td>23</td>
<td>Second person pronoun forms other than you as (modifying) possessive pronoun</td>
<td>Car</td>
<td>84.62%</td>
<td>22.22%</td>
<td>62.40%</td>
</tr>
<tr>
<td>17</td>
<td>Creation of possessive pronouns with prefix fi- + personal pronoun</td>
<td>Car</td>
<td>61.54%</td>
<td>3.18%</td>
<td>58.36%</td>
</tr>
<tr>
<td>18</td>
<td>Subject pronoun forms as (modifying) possessive pronouns: first person singular</td>
<td>Car</td>
<td>69.23%</td>
<td>17.86%</td>
<td>51.37%</td>
</tr>
<tr>
<td>21</td>
<td>Subject pronoun forms as (modifying) possessive pronouns: third person plural</td>
<td>Car</td>
<td>76.92%</td>
<td>31.75%</td>
<td>45.18%</td>
</tr>
<tr>
<td>46</td>
<td>Deletion of it in referential it is-constructions</td>
<td>Car</td>
<td>76.92%</td>
<td>31.74%</td>
<td>45.18%</td>
</tr>
<tr>
<td>18-27</td>
<td>Alternative forms as (modifying) possessive pronouns: subject and object forms, first, second, third person</td>
<td>Car</td>
<td>63.08%</td>
<td>19.77%</td>
<td>43.31%</td>
</tr>
<tr>
<td>1,2</td>
<td>She/her; he/him used for inanimate referents</td>
<td>BrIs</td>
<td>63.64%</td>
<td>33.08%</td>
<td>30.56%</td>
</tr>
</tbody>
</table>

What is also worth noting are smaller or larger sets of pronominal features which exhibit a clear regional bias. Consider the three feature sets in Table 3b, which vary considerably in the strength of their geographical signal, with ‘AR region’ representing the average attestation rate for all features in the relevant set taken together.

Table 3b. Pronouns: top diagnostic feature sets for individual world regions

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>42, 43, 44</td>
<td>Pronoun drop: Object, referential subject, dummy subject</td>
<td>As</td>
<td>95.83%</td>
<td>34.80%</td>
<td>61.03%</td>
</tr>
<tr>
<td>18-27</td>
<td>Alternative forms as (modifying) possessive pronouns: subject and object forms, first, second, third person</td>
<td>Car</td>
<td>63.08%</td>
<td>19.77%</td>
<td>43.31%</td>
</tr>
<tr>
<td>1,2</td>
<td>She/her; he/him used for inanimate referents</td>
<td>BrIs</td>
<td>63.64%</td>
<td>33.08%</td>
<td>30.56%</td>
</tr>
</tbody>
</table>

Let us comment only on the first and areally most striking feature set. It is formed by the three pronoun deleting features F42 (Object pronoun drop), as in Do you say to Paul? Yes, I told Ø already (Hong Kong English), F43 (see the Indian English example above) and F44 (Dummy subject drop), as in Ø Raining so no tennis (Sri Lankan English). The concentration of these three deletion features in Asia is of course reminiscent of Mesthrie and Bhatt’s (2008) hypothesis of a broad dichotomy among World Englishes of ‘deleters’ vs. ‘preservers’, with the Southeast Asian varieties allegedly representing the ‘deleting’ type and African varieties like Black South African English representing the ‘preserving’ type. The strong limitations of this dichotomy in an aggregate approach have been discussed elsewhere (e.g. in Lunkenheimer 2012: 865-869, Kortmann 2013: 174-177). At the same time it must be acknowledged that there are individual deletion features, like in this case pronoun dropping/deletion, which send a very pronounced Asian signal. What
we can also see, though, is that F42-44 are not only extremely characteristic of Englishes in Southeast Asia, but of South Asian Englishes, too.

Asia also plays a prominent role in the second largest grammar domain in WAVE, namely NP features.

Table 4. Noun phrase: top diagnostic features for individual world regions (AR difference region – rest of world >= 30%; bold for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>Double determiners</td>
<td>Af</td>
<td>64.71%</td>
<td>20.34%</td>
<td>44.37%</td>
</tr>
<tr>
<td>71</td>
<td>No number distinction in demonstratives</td>
<td>Af</td>
<td>76.47%</td>
<td>33.90%</td>
<td>42.57%</td>
</tr>
<tr>
<td>83</td>
<td>Comparatives and superlatives of participles</td>
<td>Am</td>
<td>60.00%</td>
<td>9.09%</td>
<td>50.91%</td>
</tr>
<tr>
<td>70</td>
<td>Proximal and distal demonstratives with ‘here’ and ‘there’</td>
<td>Am, Car</td>
<td>95.65%</td>
<td>37.74%</td>
<td>57.92%</td>
</tr>
<tr>
<td>63</td>
<td>Use of zero article where StE has indefinite article</td>
<td>As</td>
<td>100.00%</td>
<td>47.06%</td>
<td>52.94%</td>
</tr>
<tr>
<td>81</td>
<td>Much as comparative marker</td>
<td>As</td>
<td>62.50%</td>
<td>14.71%</td>
<td>47.79%</td>
</tr>
<tr>
<td>62</td>
<td>Use of zero article where StE has definite article</td>
<td>As</td>
<td>100.00%</td>
<td>52.94%</td>
<td>47.06%</td>
</tr>
<tr>
<td>82</td>
<td>As/to as comparative markers</td>
<td>As</td>
<td>50.00%</td>
<td>14.71%</td>
<td>35.29%</td>
</tr>
<tr>
<td>64</td>
<td>Use of definite article where StE favours zero</td>
<td>As</td>
<td>75.00%</td>
<td>42.65%</td>
<td>32.35%</td>
</tr>
<tr>
<td>86</td>
<td>Zero marking of degree</td>
<td>As</td>
<td>50.00%</td>
<td>17.65%</td>
<td>32.35%</td>
</tr>
<tr>
<td>61</td>
<td>Use of indefinite article where StE has definite article</td>
<td>As</td>
<td>50.00%</td>
<td>19.12%</td>
<td>30.88%</td>
</tr>
<tr>
<td>76</td>
<td>Postnominal phrases with bilong/blong/blo to express possession</td>
<td>Aus</td>
<td>60.00%</td>
<td>7.05%</td>
<td>52.95%</td>
</tr>
<tr>
<td>51</td>
<td>Plural marking via postposed elements</td>
<td>Car</td>
<td>69.23%</td>
<td>23.81%</td>
<td>45.42%</td>
</tr>
<tr>
<td>77</td>
<td>Omission of genitive suffix; possession expressed through bare juxtaposition of nouns</td>
<td>Car</td>
<td>92.31%</td>
<td>49.21%</td>
<td>43.10%</td>
</tr>
<tr>
<td>66</td>
<td>Indefinite article one/wan</td>
<td>Car</td>
<td>92.31%</td>
<td>57.14%</td>
<td>35.17%</td>
</tr>
</tbody>
</table>

Out of the 15 NP features in Table 4, almost half are characteristic of Asia. Noteworthy in light of the immediately preceding discussion on ‘deletion’ features is that the only two areoversals in Table 4, i.e. features attested in every single variety of the relevant world region, concern the deletion of definite (F62) and indefinite articles (F63) in Asia. Consider examples like Malaysian English *Vocabulary is very important at Ø elementary level* (F62) or Indian English *We decided to rent Ø apartment* (F63). Moreover, both features are attested pervasively in the respective varieties; with rp-scores of 0.8 and 0.75 it is much more likely to encounter zero articles than either definite or indefinite articles in the Asian Englishes. To these two, we may add as a third deletion feature in the NP which exhibits an areal clustering in Asia F86 (Zero marking of degree), as in Hong Kong English *Chemistry is one of the Ø interesting subjects when I was in secondary school*. A near-areoversal (95.65%) of the widest reach in Table 4, namely for America and the Caribbean, is F70 (Proximal and distal demonstratives with ‘here’ and ‘there’), as in *this here car* (Chicano English), *Knock that there spoon back* (Earlier AAVE), or *Don’t touch this here, use that there* (Bajan, or Barbadian Creole). Two features deserve special mention because they are so rare outside the relevant Anglophone world region where they are attested in the majority of varieties. These are F76 (Postnominal phrases
with *bilong*/*blong*/*long*/*blo* to express possession), as in Aboriginal English woman *belong* friend ‘the woman’s friend’, and F83 (Comparatives and superlatives of participles) as in Appalachian English *Daddy said he was the fightingest little rascal he ever hunted or He was the singingest man this side of Turnpike* (Montgomery, 2008: 452).

Let us now turn to features of the Verb Phrase, more exactly to the domains of Tense and aspect (Table 5), Modal verbs (Table 6), and Verb morphology (Table 7). We will see that almost all WAVE features in the domain of Modal verbs exhibit striking areal distributions. In the T&A domain it is again Asia which is represented by the largest number of areally distinctive features, closely followed by the varieties spoken in America. Consider Table 5:

Table 5.  Tense and aspect: top diagnostic features for individual world regions (AR difference region – rest of world >= 30%; **bold** for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td><em>Come</em>-based future/ingressive markers</td>
<td>Af</td>
<td>58.82%</td>
<td>13.56%</td>
<td>45.26%</td>
</tr>
<tr>
<td>105</td>
<td><em>Compleative/perfect have/be + done + past participle</em></td>
<td>Am</td>
<td>70.00%</td>
<td>7.58%</td>
<td>62.42%</td>
</tr>
<tr>
<td>90</td>
<td><em>Invariant be</em> as habitual marker</td>
<td>Am</td>
<td>60.00%</td>
<td>16.67%</td>
<td>43.33%</td>
</tr>
<tr>
<td>92</td>
<td>Other non-standard habitual markers: <em>synthetic</em></td>
<td>Am</td>
<td>60.00%</td>
<td>19.70%</td>
<td>30.30%</td>
</tr>
<tr>
<td>104</td>
<td><em>Compleative/perfect done</em></td>
<td>Am, Car</td>
<td>78.26%</td>
<td>13.20%</td>
<td>65.06%</td>
</tr>
<tr>
<td>100</td>
<td>Levelling of the difference between present perfect and simple past: present perfect for StE simple past</td>
<td>As</td>
<td>100.00%</td>
<td>36.76%</td>
<td>63.24%</td>
</tr>
<tr>
<td>109</td>
<td><em>Perfect marker already</em></td>
<td>As</td>
<td>75.00%</td>
<td>26.47%</td>
<td>48.53%</td>
</tr>
<tr>
<td>89</td>
<td>Wider range of uses of progressive <em>be + V-ing</em> than in StE: extension to habitual contexts</td>
<td>As</td>
<td>87.50%</td>
<td>44.12%</td>
<td>43.38%</td>
</tr>
<tr>
<td>119</td>
<td><em>Would</em> for (distant) future in contrast to <em>will</em> (immediate future)</td>
<td>As</td>
<td>62.50%</td>
<td>22.06%</td>
<td>40.44%</td>
</tr>
<tr>
<td>113</td>
<td><em>Loosening of sequence of tenses rule</em></td>
<td>As</td>
<td>100.00%</td>
<td>61.76%</td>
<td>38.24%</td>
</tr>
<tr>
<td>95</td>
<td><em>Be sat/stood</em> with progressive meaning</td>
<td>BrIs</td>
<td>63.64%</td>
<td>6.15%</td>
<td>57.49%</td>
</tr>
<tr>
<td>102</td>
<td><em>Be</em> as perfect auxiliary</td>
<td>BrIs</td>
<td>54.55%</td>
<td>23.07%</td>
<td>31.48%</td>
</tr>
<tr>
<td>114</td>
<td><em>Go</em>-based future markers</td>
<td>Car</td>
<td>100.00%</td>
<td>36.51%</td>
<td>63.49%</td>
</tr>
</tbody>
</table>

Three T&A features turn out to be true areaversals. In terms of areal distinctiveness, the weakest of these is F113 (*Loosening of sequence of tenses rule*), which is found in all Asian varieties (e.g. in Pakistan English, as in *Even though I lost, I was happy that I participated in it*), but in about 62% of the varieties elsewhere in the Anglophone world as well. A much more forceful Asian areaversal is F100 (*Present perfect for StE simple past*), as in Sri Lankan English *We’ve been there last year* or Hong Kong English *I have learned to play piano but now I forget*. This feature even crosses the 60% threshold in the rightmost column, which makes it truly diagnostic of the South and Southeast Asian varieties. In the same league we find the Caribbean areaversal F114 (*Go*-based future markers, rp-score 0.83), as in Barbadian Creole *I gun go town tomorrow* or Trinidadian Creole *We go do that next week*. Areally diagnostic are also the two features involving
done in the coding of the completive/perfect category; while the simpler construction
F104 (Completive/perfect done), as in Appalachian English Uncle Mingus was done
dead, Gullah Uh done eat dat one already, or Jamaican Creole Sharon don riid di buk is
diagnostic of both America and the Caribbean, the more complex feature F105
(Completive/perfect have/be + done + past participle), as in He is done gone or ah fin'
dat somebody has done talk about me, is diagnostic only for the American varieties. At
the same time F105 is found very rarely outside of America, a property that it shares with
only one more T&A feature, namely F95 (Be sat/stood with progressive meaning). This
feature is highly distinctive of the British Isles. Consider examples like I was sat at the
bus stop for ages (North) or He was stood on the corner (East Anglia). However, even in
their home regions, these two features are encountered rather sporadically (rp-score F105:
0.39, F95: 0.25).

There are only seven features in WAVE relating to Modal verbs, but six of them
clearly display an areal clustering in one particular part of the English-speaking world,
with America figuring prominently in this respect. Consider the following table.

Table 6. Modal verbs: top diagnostic features for individual world regions
(AR difference region – rest of world >= 30%; bold for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>New quasi modals: core modal meanings</td>
<td>Am</td>
<td>80.00%</td>
<td>15.15%</td>
<td>64.85%</td>
</tr>
<tr>
<td>121</td>
<td>Double modals</td>
<td>Am</td>
<td>80.00%</td>
<td>22.73%</td>
<td>57.27%</td>
</tr>
<tr>
<td>122</td>
<td>Epistemic mustn’t</td>
<td>Am</td>
<td>50.00%</td>
<td>19.70%</td>
<td>30.30%</td>
</tr>
<tr>
<td>126</td>
<td>New quasi modals: aspectual meanings</td>
<td>Am, Car</td>
<td>69.57%</td>
<td>11.32%</td>
<td>58.24%</td>
</tr>
<tr>
<td>127</td>
<td>Non-standard use of modals for politeness reasons</td>
<td>As</td>
<td>87.50%</td>
<td>17.65%</td>
<td>69.85%</td>
</tr>
<tr>
<td>123</td>
<td>Present tense forms of modals used where StE has past tense forms</td>
<td>As</td>
<td>87.50%</td>
<td>33.82%</td>
<td>53.68%</td>
</tr>
</tbody>
</table>

The emergence of new quasi-modals like ‘counterfactual’ like tolliketa (‘almost’,
‘nearly’), supposeta/sposeta, useta (F125) or the aspectual finna/fixin to (‘prepare, get
ready/be about to’; F126) is highly characteristic of varieties in America and additionally,
but only for aspectual meanings (F126), of the Caribbean. Here are some examples from
Appalachian English (Montgomery 2008: 437): The measles like to kill me, I like to never
in the world got away, I’m fixin’ to leave now. Another prominent modal feature of
America is F121 (Double modals), as in Chicano English We might could do that, which
is a well-known feature especially of Texas and Appalachian English. The areally most
diagnostic feature in Table 6 is F127 (Non-standard use of modals for politeness reasons),
which, apart from Malaysian English, is found in all Asian varieties and can be
exemplified as follows: I would be visiting your place tomorrow (‘I will visit your place
tomorrow’, Indian English), Shall I use your phone? (‘May I use your phone?’), Sri
Lankan English), or Must I give you some water? (‘Should I may I give you some
water?’, Hong Kong English).
The third and last VP-related domain concerns verb morphology. A cursory glance at Table 7 suffices to see the dominance of the Caribbean, which is due to the very widespread use of serial verbs in this world region, on the one hand (F148-151), and other forms or phrases that Caribbean varieties and Creoles use for the copula ‘be’ (F140, F141), on the other hand:

Table 7. Verb morphology: top diagnostic features for individual world regions

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>a-prefixing on elements other than ing-forms</td>
<td>Am</td>
<td>70.00%</td>
<td>12.12%</td>
<td>57.88%</td>
</tr>
<tr>
<td>144</td>
<td>Use of gotten and got with distinct meanings (dynamic vs. static)</td>
<td>Am</td>
<td>60.00%</td>
<td>13.24%</td>
<td>46.76%</td>
</tr>
<tr>
<td>128</td>
<td>Levelling of past tense/past participle verb forms: regularization of irregular verb paradigms</td>
<td>Am</td>
<td>90.00%</td>
<td>57.58%</td>
<td>32.42%</td>
</tr>
<tr>
<td>133</td>
<td>Double marking of past tense</td>
<td>As</td>
<td>75.00%</td>
<td>24.24%</td>
<td>50.76%</td>
</tr>
<tr>
<td>143</td>
<td>Transitive verb suffix -em/-im/-um</td>
<td>Aus</td>
<td>60.00%</td>
<td>5.63%</td>
<td>54.37%</td>
</tr>
<tr>
<td>129</td>
<td>Levelling of past tense/past participle verb forms: unmarked forms</td>
<td>BrIs</td>
<td>90.91%</td>
<td>60.00%</td>
<td>30.91%</td>
</tr>
<tr>
<td>141</td>
<td>Other forms/phrases for copula ‘be’: before locatives</td>
<td>Car</td>
<td><strong>84.62%</strong></td>
<td><strong>14.29%</strong></td>
<td><strong>70.33%</strong></td>
</tr>
<tr>
<td>150</td>
<td>Serial verbs: come = ‘movement towards’</td>
<td>Car</td>
<td>76.92%</td>
<td>15.88%</td>
<td>61.04%</td>
</tr>
<tr>
<td>149</td>
<td>Serial verbs: go = ‘movement away from’</td>
<td>Car</td>
<td>76.92%</td>
<td>17.46%</td>
<td>59.46%</td>
</tr>
<tr>
<td>151</td>
<td>Serial verbs: constructions with 3 verbs</td>
<td>Car</td>
<td>76.92%</td>
<td>17.46%</td>
<td>59.46%</td>
</tr>
<tr>
<td>148</td>
<td>Serial verbs: give = ‘to, for’</td>
<td>Car</td>
<td>61.54%</td>
<td>11.11%</td>
<td>50.43%</td>
</tr>
<tr>
<td>140</td>
<td>Other forms/phrases for copula ‘be’: before NPs</td>
<td>Car</td>
<td>69.23%</td>
<td>19.05%</td>
<td>50.18%</td>
</tr>
</tbody>
</table>

Here are some examples for serial verb constructions from the Caribbean: F148 (give = ‘to, for’) They fry fowl egg, many cake give ‘for’ him (Bahamian Creole), F149 (go = ‘movement away from’) go a St Mary go ‘to’ look pon di children dem (Jamaican Creole), F150 (come = ‘movement towards’) Him say him was guwane come look for mi dis morning (Jamaican English), and F151 (Serial verbs: constructions with three verbs) Come here go see if Olga home (Bahamian Creole). The most diagnostic verb morphology feature of the Caribbean is F141 (Other forms/phrases for copula ‘be’: before locatives), as in me deh a mi yard a wait pon him (Jamaican Creole), I de here every night (Bahamian English), or Mi pikni de a skuul (Vincentian Creole). But in fact it is not only before locatives (F141) that Caribbean varieties and Creoles use alternative forms for copula ‘be’; this is also true for the position before NPs (F140), as in Evride a krismuhs (Vincentian Creole). The rarity of F135 (a-prefixing on elements other than ing-forms) outside America makes it highly distinctive of this region, in spite of its low rp-score in the respective varieties (0.27). Appalachian English is likely unrivalled in its range of options, including he comes a-nigh me, a pass a-toward him, go back down a-Sunday, I didn’t do it a-purpose, ride a-horseback, he was a-just tearing that window open (all from Montgomery 2008: 440-441). One other feature in Table 7 worth mentioning is F143 (Transitive verb suffix -em/-im/-um), as in Mi baiim kaikai (Torres Strait Creole), faksim ‘to fax’, imelim ‘email (someone)’, since it is extremely rare outside Australia.
In the domain of negation (see Table 8), America is prominently represented with a true 100%, highly pervasive areoversal (F158: Invariant don’t for all persons in the present tense, rp-score 0.88) and the ain’t cluster F155-157, i.e. ain’t as negated form of be (as in Earlier AAVE but I ain’t ‘bleevin in nothin but de good Lawd), have (as in Gullah You ain fuh go ‘you don’t have to go’), and do, i.e. before main verbs (as in Gullah You ain like um?). The latter use of ain’t is distinctive of both America and the Caribbean and, at the same time, very rare elsewhere in the Anglophone world. The latter is also true of the only negation feature distinctive of Australia, namely F162 (no more/nomo as negative existential marker), as in Aboriginal English Nomo nating insai dea ‘There isn’t anything in there’.

Table 8. Negation: top diagnostic features for individual world regions

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>156</td>
<td>ain’t as the negated form of have</td>
<td>Am</td>
<td>90.00%</td>
<td>34.85%</td>
<td>55.15%</td>
</tr>
<tr>
<td>158</td>
<td>Invariant don’t for all persons in the present tense</td>
<td>Am</td>
<td>100.00%</td>
<td>63.64%</td>
<td>36.36%</td>
</tr>
<tr>
<td>157</td>
<td>ain’t as generic negator before a main verb</td>
<td>Am, Car</td>
<td>53.08%</td>
<td>7.69%</td>
<td>45.39%</td>
</tr>
<tr>
<td>162</td>
<td>no more/nomo as negative existential marker</td>
<td>Aus</td>
<td>60.00%</td>
<td>7.05%</td>
<td>52.95%</td>
</tr>
<tr>
<td>163</td>
<td>was – weren’t split</td>
<td>BrIs</td>
<td>54.55%</td>
<td>12.31%</td>
<td>42.24%</td>
</tr>
<tr>
<td>161</td>
<td>not as a preverbal negator</td>
<td>Pac</td>
<td>50.00%</td>
<td>19.12%</td>
<td>30.88%</td>
</tr>
<tr>
<td>155, 156, 157</td>
<td>ain’t as negated form of be; have; before main verb</td>
<td>Am</td>
<td>83.33%</td>
<td>28.28%</td>
<td>55.05%</td>
</tr>
</tbody>
</table>

The tables for the following domains of grammar are so short that they hardly deserve special comment. As for Agreement (Table 9) the Caribbean (recall F140 and 141 in Table 7 above) is again noticeable for two features relating to copula be (F177 and F178). Only this time what is striking for this world region is that the vast majority of Caribbean varieties and Creoles delete copula be altogether: before Adjective Phrases (F177), as in Jamaican Creole wen tings slow di man lazy ‘When things are slow the man is lazy’, and before locatives (F178), as in Jamaican English Mary in the garden. With an rp-score of 0.86, zero copula before adjective phrases can be considered an almost categorical feature of Caribbean Englishes. Examples illustrating the other two features will follow Table 9.

Table 9. Agreement: top diagnostic features for individual world regions

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>173</td>
<td>Variant forms of dummy subject there in existential uses, e.g. they, it or zero</td>
<td>Am</td>
<td>80.00%</td>
<td>43.94%</td>
<td>36.06%</td>
</tr>
<tr>
<td>180</td>
<td>was/were generalization</td>
<td>BrIs</td>
<td>90.91%</td>
<td>46.15%</td>
<td>44.76%</td>
</tr>
<tr>
<td>177</td>
<td>Deletion of copula be: before AdjPs</td>
<td>Car</td>
<td>92.31%</td>
<td>42.86%</td>
<td>49.45%</td>
</tr>
<tr>
<td>178</td>
<td>Deletion of copula be: before locatives</td>
<td>Car</td>
<td>76.92%</td>
<td>37.77%</td>
<td>39.15%</td>
</tr>
</tbody>
</table>
They were saying that they had a lot of problems at Garner because it was a lot of fights and stuff. (Chicano English)

When you come home fae your honeymoon if you had one, you was 'kirkit'. (Scottish English)

Interestingly, the only three WAVE features in the domain of relative clauses exhibiting areality, at all, cluster in the same world region, namely America (see Table 10). The strongest areal signal of all is found with F186 (which for 'who'), as in With (last name), which is my relatives also (Newfoundland English).

Table 10. Relativization: top diagnostic features for individual world regions
(AR difference region – rest of world >= 30%; bold for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>186</td>
<td>which for 'who'</td>
<td>Am</td>
<td>80.00%</td>
<td>27.27%</td>
<td>52.73%</td>
</tr>
<tr>
<td>193</td>
<td>Gapping/zero-relativization in subject position</td>
<td>Am</td>
<td>90.00%</td>
<td>56.06%</td>
<td>33.94%</td>
</tr>
<tr>
<td>198</td>
<td>Deletion of stranded prepositions in relative clauses ('preposition chopping')</td>
<td>Am</td>
<td>60.00%</td>
<td>28.79%</td>
<td>31.21%</td>
</tr>
</tbody>
</table>

There's some people come here to look at water (Newfoundland English)

T's only ting weh covetin happier ('It's the only thing that coveting is happier than.' Gullah)

The Caribbean dominates Table 11, which offers the areally most distinctive features in the domain of complementation. The following features qualify as diagnostic of the Caribbean: F206 (Existentials with forms of have), as in Bahamian Creole (Turtle, is there much turtle now?) No, don't have turtle round, and F201 (for-based complementizers), as in Barbadian Creole It hard fo /fi get ova wha he do.

Table 11. Complementation: top diagnostic features for individual world regions
(AR difference region – rest of world >= 30%; bold for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>203</td>
<td>for (to) as infinitive marker</td>
<td>Am</td>
<td>90.00%</td>
<td>36.36%</td>
<td>53.64%</td>
</tr>
<tr>
<td>207</td>
<td>Substitution of that-clause for infinitival subclause</td>
<td>As</td>
<td>50.00%</td>
<td>17.65%</td>
<td>32.35%</td>
</tr>
<tr>
<td>206</td>
<td>Existentials with forms of have</td>
<td>Car</td>
<td>76.92%</td>
<td>12.69%</td>
<td>64.23%</td>
</tr>
<tr>
<td>201</td>
<td>For-based complementizers</td>
<td>Car</td>
<td>84.62%</td>
<td>23.81%</td>
<td>60.81%</td>
</tr>
<tr>
<td>210</td>
<td>Non-finite clause complements with bare root form rather than -ing form</td>
<td>Car</td>
<td>61.54%</td>
<td>26.98%</td>
<td>34.56%</td>
</tr>
</tbody>
</table>

The attentive reader may have noticed that Africa has hardly figured in any of the tables so far (only in Table 4 on NP structure, with F59 (Double determiners) and F71 (No number distinction in demonstratives)). In the domain of adverbial subordination, however, Africa is the Anglophone world region where the feature with the highest degree of areality is found. This is F214 (Conjunction doubling: clause + conj. + conj. +
clause), as in White Zimbabwean English: *I’m happy here but still I’m homesick.* Characteristic of Australia is F211 (Clause-final *but* = ‘though’), as in Aboriginal English *Yeah...a few smashes but ‘Yes [we had] a few fights though’, while chaining constructions linking two main verbs (F213) are slightly (by 31.14%) overrepresented in the Caribbean. Consider, for example, *dis man tel dem gwain du dat* ‘this man told them (he was) going to do that’ (Jamaican Creole).

Table 12. Adverbial subordination: top diagnostic features for individual world regions (AR difference region – rest of world >= 30%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>214</td>
<td>Conjunction doubling: clause + conj. + conj. + clause</td>
<td>Af</td>
<td>70.59%</td>
<td>25.42%</td>
<td>45.17%</td>
</tr>
<tr>
<td>211</td>
<td>Clause-final <em>but</em> = ‘though’</td>
<td>Aus</td>
<td>60.00%</td>
<td>21.12%</td>
<td>38.88%</td>
</tr>
<tr>
<td>213</td>
<td>No subordination; chaining construction linking two main verbs (motion and activity)</td>
<td>Car</td>
<td>69.23%</td>
<td>38.09%</td>
<td>31.14%</td>
</tr>
</tbody>
</table>

Africa figures again in the domain of Adverbs and prepositions, or rather postpositions (F217), as for example in *He came to office yesterday before* (Ugandan English), *We can do it in morning time* (Nigerian English), or *He lives Gwanda way* (White Zimbabwean English).

The top diagnostic feature in Table 13, however, is F218 (affirmative *anymore ‘nowadays’*), as in *Anymore they have had a hard time protecting things like that* (Appalachian English; Montgomery 2008: 455). This is one of the most strongly US-affiliated features in WAVE (i.e. excluding Newfoundland English; cf. also Hickey 2012: 8). This is also the feature which gives away Gullah as a distinctively US American Creole.

Table 13. Adverbs and prepositions: top diagnostic features for individual world regions (AR difference region – rest of world >= 30%; **bold** for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region – RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>217</td>
<td>Use of postpositions</td>
<td>Af</td>
<td>58.82%</td>
<td>28.81%</td>
<td>30.01%</td>
</tr>
<tr>
<td>218</td>
<td>Affirmative <em>anymore ‘nowadays’</em></td>
<td>Am</td>
<td>70.00%</td>
<td>3.03%</td>
<td><strong>66.97%</strong></td>
</tr>
</tbody>
</table>

The last domain of grammar to be checked for areal features is Discourse and word order (Table 14). From an areal perspective, this domain is dominated again by America and the Caribbean.
Table 14. Discourse and word order: top diagnostic features for individual world regions (AR difference region – rest of world >= 30%; **bold** for >= 60%)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>AR region</th>
<th>AR RoW</th>
<th>AR diff. region - RoW</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>‘Negative inversion’</td>
<td>Am</td>
<td>90.00%</td>
<td>12.12%</td>
<td>77.88%</td>
</tr>
<tr>
<td>23</td>
<td>Superlative marker <strong>most</strong> occurring before head noun</td>
<td>Am</td>
<td>50.00%</td>
<td>18.18%</td>
<td>31.82%</td>
</tr>
<tr>
<td>22</td>
<td>Other options for clefting than StE</td>
<td>Am, Car</td>
<td>65.22%</td>
<td>28.30%</td>
<td>36.92%</td>
</tr>
<tr>
<td>23</td>
<td>Either order of objects in double object constructions (if both objects are pronominal)</td>
<td>BrIs</td>
<td>72.73%</td>
<td>13.85%</td>
<td>58.88%</td>
</tr>
<tr>
<td>22</td>
<td>Sentence-initial focus marker</td>
<td>Car</td>
<td>76.92%</td>
<td>22.22%</td>
<td>54.70%</td>
</tr>
<tr>
<td>22</td>
<td>No inversion/no auxiliaries in <strong>wh</strong>- questions</td>
<td>Car</td>
<td>100.00%</td>
<td>65.46%</td>
<td>34.54%</td>
</tr>
<tr>
<td>23</td>
<td>Presence of subject in imperatives</td>
<td>Pac</td>
<td>75.00%</td>
<td>38.24%</td>
<td>36.76%</td>
</tr>
</tbody>
</table>

Clearly the strongest American feature of all in Table 14 is F226 (‘Negative inversion’), as in the following examples: *Didn’t nobody show up* (Colloquial American English), *Ain nobody ga worry wid you* ‘Nobody will worry with you’ (Gullah), *There’s an old house up here, but don’t nobody live in it, Hain’t nobody never set [the trap] for any bears since* (Appalachian English; Montgomery 2008: 444). The strongest Caribbean feature is F225 (Sentence-initial focus marker), as in *Is now where you does find we corn coming* (Bahamian Creole), *Is she love me* (Trinidadian Creole), or *Da uman him de luk* (San Andrés Creole). While Caribbean areoversal F228 is also rather common in other regions, the rp-score for this feature is remarkable – 0.94 means it is an obligatory feature in almost all Caribbean varieties. The British Isles are represented by only one feature (F232: Either order in double object constructions for two pronominal objects), as in *Give it me/Give me it* (North of England), *He couldna gae him it* (Orkney and Shetland English), or *She’d give us it* (Southwest England), but this feature nearly reaches the 60% threshold necessary for qualifying as a truly diagnostic feature in areal terms.

Before reviewing and perspectivizing the main observations in this section from an areal perspective (see section 3), let us briefly look at the preceding tables from the point of view of areality per grammatical domain (Table 15).

Table 15. Overview: Diagnostic areal features per grammatical domain (sorted by proportion of areal features per domain)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of diagnostic areal features (total: 93)</th>
<th>Proportion</th>
<th>Total number of WAVE features per domain (total: 235)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modal verbs</td>
<td>6</td>
<td>85.71%</td>
<td>7</td>
</tr>
<tr>
<td>Adverbial subordination</td>
<td>3</td>
<td>60.00%</td>
<td>5</td>
</tr>
<tr>
<td>Discourse organization and word order</td>
<td>7</td>
<td>53.85%</td>
<td>13</td>
</tr>
</tbody>
</table>
Table (15) is ordered by the column “Proportion” of WAVE features per grammatical domain which have been found to be distinctive or even truly diagnostic of one particular Anglophone world region. This ordering is meant to relativize the no doubt correct observation that the largest number of features exhibiting areal concentrations (some 60%) fall into the domains of the usual suspects, namely, Pronouns (17 features), NP structure (15 features), Tense and aspect (13 features), and Verb morphology (12 features). Judged against the total of WAVE features for each of the twelve grammatical domains, however, it emerges that domains represented by far fewer features may exhibit a much stronger areal clustering. This is particularly pronounced for modal verbs (recall Table 6 above).

3 Synopsis: Areality in morphosyntax across the Anglophone world

What are the major lessons to be learnt concerning areal clusterings in each of the seven world regions across the 76 varieties / Pidgins and Creoles investigated here? The following table, listing by world region the most distinctive areal features in the tables of the previous section, helps us to answer this question:

Table 16. Overview: List of most distinctive areal features per world region
(AR difference region – rest of world >= 30%; **bold** for >= 60%; * for rp-score >= 0.5, ** for >= 0.75)

<table>
<thead>
<tr>
<th>Region</th>
<th>Feature Total</th>
<th>Attested in 100% of all varieties per world region</th>
<th>Attested in 0-5 varieties in RoW</th>
<th>All other areal features with AR difference region - RoW &gt;= 30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>America + Caribbean</td>
<td>6</td>
<td>157 (4 varieties)</td>
<td></td>
<td>13, 70*, 104*, 126*, 223</td>
</tr>
<tr>
<td>Asia</td>
<td>18</td>
<td>43**, 62**, 63**, 100*, 113*</td>
<td>40 (3 varieties)</td>
<td>61, 64*, 81, 82, 86, 89, 109, 119, 123, 127*, 133, 207</td>
</tr>
<tr>
<td>British Isles</td>
<td>6</td>
<td>95 (4 varieties)</td>
<td></td>
<td>102, 129*, 163, 180*, 232</td>
</tr>
</tbody>
</table>
The strongest geographical signals we receive are for America, the Caribbean and (South and Southeast) Asia. About 80% of the 93 morphosyntactic features listed in Table 16 fall into one of these three Anglophone world regions, with the Caribbean clearly taking the lead (34), followed by America (28) and, at some distance, Asia (18). The Caribbean also exhibits the highest rp-scores overall, 21 of 28 exclusively Caribbean features are rated pervasive or even highly pervasive. Among the top diagnostic features for the Caribbean and America, there are six which are diagnostic of the entire American and Caribbean area. It does not appear to be a coincidence that these three belong to those Anglophone world regions which are clearly dominated by one variety type, and thus qualify as (fairly) homogeneous world regions: America is dominated by L1 varieties (8 out of 10), Asia by L2 varieties (6 out of 8), and the Caribbean by Creoles (11 out of 13).

At the same time it is noticeable that for the only other homogeneous Anglophone world region, namely the British Isles (10 out 11 varieties are L1 varieties), only six features exhibit a high degree of areality. Additionally, the distinctively British features are overwhelmingly rated optional or rare. This may be interpreted, on the one hand, as a reflex of (a) the British Isles as the historical homeland of English and the origin of its spread across the world from the seventeenth century onwards, which left few distinct British Isles features, and (b) as the varieties in the British Isles having developed only few morphosyntactic properties in the last 100-150 years or so which have not developed elsewhere in the Anglophone world, too. Remarkable for South and Southeast Asia is the fact that 5 out of its 18 areally distinctive features are found in 100 %, i.e. in all 8 varieties. This is the highest percentage of true areoversals for any Anglophone world region. Striking for the Australia and the Pacific is their high proportion of features (5 out of 8) hardly or not at all found elsewhere in the Anglophone world, the two most important instances of which (F36 ‘Distinct forms for inclusive/exclusive first person non-singular’ and F37 ‘More number distinctions in personal pronouns than simply singular vs. plural’) are diagnostic for the entire Australian and Pacific region. Such features, rarely or not at all found in other parts of the Anglophone world, which, in turn, send out an extra-strong geographical signal, most likely to be explained by very strong substrate influence. A last point worth noting is that Africa, that world region represented by the largest number (17) of varieties, Pidgins and Creoles in WAVE, has only five areally distinctive features, none of which (see the fourth column in Table 16) would qualify as a rarum in the rest of the Anglophone world (RoW - found in at most 10% in RoW), let alone rarissimum (found in at most 5% in RoW).

Even though the aggregate approach to morphosyntactic variation on a global scale, as underlying the NeighborNet diagram in Figure 1, leads to the conclusion that areality can be attributed only secondary importance compared with the explanatory potential of variety type, the approach adopted in this chapter has shown that the factor ‘areality’, i.e. the geographical signal, can indeed be strengthened. The overall

<table>
<thead>
<tr>
<th>Region</th>
<th>Varieties</th>
<th>Table 16 Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>5</td>
<td>59, 71*, 116, 214, 217</td>
</tr>
<tr>
<td>Australia</td>
<td>4</td>
<td>76 (5 varieties) 211</td>
</tr>
<tr>
<td>Pacific</td>
<td>2</td>
<td>143 (4 varieties) 161, 233*</td>
</tr>
<tr>
<td>Australia +</td>
<td>2</td>
<td>162 (5 varieties) 37 (0 varieties)</td>
</tr>
<tr>
<td>Pacific</td>
<td></td>
<td>93/235 (39.57%)</td>
</tr>
</tbody>
</table>

The strongest geographical signals we receive are for America, the Caribbean and (South and Southeast) Asia. About 80% of the 93 morphosyntactic features listed in Table 16 fall into one of these three Anglophone world regions, with the Caribbean clearly taking the lead (34), followed by America (28) and, at some distance, Asia (18). The Caribbean also exhibits the highest rp-scores overall, 21 of 28 exclusively Caribbean features are rated pervasive or even highly pervasive. Among the top diagnostic features for the Caribbean and America, there are six which are diagnostic of the entire American and Caribbean area. It does not appear to be a coincidence that these three belong to those Anglophone world regions which are clearly dominated by one variety type, and thus qualify as (fairly) homogeneous world regions: America is dominated by L1 varieties (8 out of 10), Asia by L2 varieties (6 out of 8), and the Caribbean by Creoles (11 out of 13).

At the same time it is noticeable that for the only other homogeneous Anglophone world region, namely the British Isles (10 out 11 varieties are L1 varieties), only six features exhibit a high degree of areality. Additionally, the distinctively British features are overwhelmingly rated optional or rare. This may be interpreted, on the one hand, as a reflex of (a) the British Isles as the historical homeland of English and the origin of its spread across the world from the seventeenth century onwards, which left few distinct British Isles features, and (b) as the varieties in the British Isles having developed only few morphosyntactic properties in the last 100-150 years or so which have not developed elsewhere in the Anglophone world, too. Remarkable for South and Southeast Asia is the fact that 5 out of its 18 areally distinctive features are found in 100 %, i.e. in all 8 varieties. This is the highest percentage of true areaversals for any Anglophone world region. Striking for the Australia and the Pacific is their high proportion of features (5 out of 8) hardly or not at all found elsewhere in the Anglophone world, the two most important instances of which (F36 ‘Distinct forms for inclusive/exclusive first person non-singular’ and F37 ‘More number distinctions in personal pronouns than simply singular vs. plural’) are diagnostic for the entire Australian and Pacific region. Such features, rarely or not at all found in other parts of the Anglophone world, which, in turn, send out an extra-strong geographical signal, most likely to be explained by very strong substrate influence. A last point worth noting is that Africa, that world region represented by the largest number (17) of varieties, Pidgins and Creoles in WAVE, has only five areally distinctive features, none of which (see the fourth column in Table 16) would qualify as a rarum in the rest of the Anglophone world (RoW - found in at most 10% in RoW), let alone rarissimum (found in at most 5% in RoW).
Typological profile of a given variety of English may not be determined by geography, i.e. by the part of the world where it is spoken, but this does not prevent a large number of morphosyntactic features from exhibiting distinct areal biases. In fact, it is highly astonishing that 40% of all 235 features in the WAVE data set (93, i.e. the sum total of all features listed in Tables 3-15) exhibit a noticeable areality in the sense that they are overrepresented by at least a 30% margin in one of the Anglophone world regions compared with the rest of the world. For 16 of these 93 features (marked in bold in Table 16) the attestation rate (AR) difference between the relevant Anglophone world region and all others is higher than 60% (e.g. F201 ‘for-based complementizers’ is attested in 84.62% of all 13 WAVE varieties in the Caribbean whereas its attestation rate in all other 63 varieties, Pidgins and Creoles in the WAVE data set is 23.81%). These are the 16 features listed in Table 17 below.

Table 17. Diagnostic morphosyntactic features per Anglophone world region (AR difference region – rest of world >= 60%; * for rp-score >= 0.5, ** for >= 0.75)

<table>
<thead>
<tr>
<th>No</th>
<th>Feature</th>
<th>Region</th>
<th>Example from world region</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Benefactive ‘personal dative’ construction</td>
<td>Am</td>
<td>They found them an apartment (ChcE)</td>
</tr>
<tr>
<td>105</td>
<td>Completive/perfect have/be + done + past participle</td>
<td>Am</td>
<td>He is done gone (EAAVE)</td>
</tr>
<tr>
<td>125</td>
<td>New quasi modals: core modal meanings</td>
<td>Am</td>
<td>He belongs to come here today ‘he ought to come…’ (AppE)</td>
</tr>
<tr>
<td>218</td>
<td>Affirmative anymore ‘nowadays’</td>
<td>Am</td>
<td>Anymore they have a hard time protecting things like that (AppE)</td>
</tr>
<tr>
<td>226</td>
<td>‘Negative inversion’</td>
<td>Am</td>
<td>Ain nobody ga worry wid you (Gullah)</td>
</tr>
<tr>
<td>104</td>
<td>Completive/perfect done</td>
<td>Am, Car</td>
<td>He done gone (Bajan)</td>
</tr>
<tr>
<td>22</td>
<td>you as (modifying) possessive pronoun</td>
<td>Car</td>
<td>Tuck in you shirt (TrinC)</td>
</tr>
<tr>
<td>23</td>
<td>Second person pronoun forms other than you as (modifying) possessive pronoun go-based future markers</td>
<td>Car</td>
<td>Tek out unu buk! (SanAC)</td>
</tr>
<tr>
<td>114</td>
<td>‘go-based future markers</td>
<td>Car</td>
<td>Mi go pik dem uhp (VinC)</td>
</tr>
<tr>
<td>141</td>
<td>Other forms/phrases for copula ‘be’: before locatives</td>
<td>Car</td>
<td>Mi de na mi mama oso. (Sranan)</td>
</tr>
<tr>
<td>150</td>
<td>Serial verbs: come = ‘movement towards’</td>
<td>Car</td>
<td>Run come quick (TrinC)</td>
</tr>
<tr>
<td>201</td>
<td>for-based complementizers</td>
<td>Car</td>
<td>I haad fi kraas di riba (JamC)</td>
</tr>
<tr>
<td>206</td>
<td>Existentials with forms of have</td>
<td>Car</td>
<td>You have people that own big piece a land (BelC)</td>
</tr>
<tr>
<td>100</td>
<td>Levelling of the difference between present perfect and simple past: present perfect for Ste simple past</td>
<td>As</td>
<td>Ben has return back the product yesterday. (MalE)</td>
</tr>
<tr>
<td>127</td>
<td>Non-standard use of modals for politeness reasons</td>
<td>As</td>
<td>Must I give you some water? (HKE)</td>
</tr>
<tr>
<td>37</td>
<td>More number distinctions in personal pronouns than simply singular vs. plural</td>
<td>Aus, Pac</td>
<td>We two is going... (PalmE)</td>
</tr>
</tbody>
</table>
Figure 2 illustrates the different layers of areal distinctiveness. The strength of the geographical signal increases towards the centre. Circles I and II represent the 16 core areal features that surpass a 60% difference in attestation rate (cf. Table 17); the five features in the Circle I are additionally either areoversals or exceedingly rare in the rest of the world (cf. Table 16, columns 3 and 4). Areoversals and rara with an AR-difference of 30-60% are found in Circle III. In the two more peripheral layers, areally distinctive features (Table 16, column 5) are distinguished according to their regional pervasiveness (Circle IV: rp-score ≥ 0.5, Circle V: rp-score < 0.5).
Figure 2. Strength of the geographical signal

Taken together, this clearly shows that North America, i.e. in WAVE the US varieties joined by Newfoundland English, the Caribbean and Asia are those anglophone regions exhibiting the highest degree of areality (i.e. the largest number of features with a strong “local” signal). Of these regions, the Caribbean is most distinctively marked by both diagnostic and highly pervasive features, whereas the majority of American and Asian features are found in the peripheral layer.

4 Smaller-scale areal patterns within selected world regions
Could the geographical signal for morphosyntactic variation be strengthened even more on the basis of the WAVE data? Yes it could, but only to a limited extent. Reasons of space prevent us from presenting the results of our relevant studies in detail here, but the following can be said. Zooming in from the set of Anglophone world regions on sub-regions within each of them, one can try to determine, for example, whether it is possible to find (bundles of) morphosyntactic features separating, for example, the North from the South in the British Isles, South Asia from Southeast Asia, or West Africa from East and South(African) Africa. It is indeed possible to do that. For the British Isles (see, for example, Kortmann 2012: 697-698) it turns out, for instance, that there are more distinctively Northern features (the North being constituted by Orkney and Shetland English, the dialects of the North of England, Scottish English, Irish English and Manx English) than there are distinctively Southern features (the South comprising the dialects of the Southwest and Southeast of England, East Anglia, Channel Island English and Welsh English). The only two features which are found exclusively in the North and not at all in the South are F69 ‘yon/yonder indicating remoteness’ (attested in all 5 Northern varieties), as in *yon oil company* (O&SE) or *it’s allus light in yandhar place* (Manx English), and F124 ‘want/need + past participle’, as in *That shirt wants washed* (North England dialects). *It needs cleaned out* (Northern Irish English), or *Does your floor need hoovered?* (Scottish English). F124 is only attested in exactly these three of the five varieties of the North, though. There are no such exclusively Southern British Isles features. Zooming in even further into the South of England, though, one receives a pronounced geographical signal for British Creole, which, socio-historically and typologically, is the obvious outlier in the British Isles since it is the only creole in an otherwise exclusively L1 world region. It can be taken as a crucial piece of evidence of areal impact on a given variety (or Pidgin, Creole, language for that matter) where the feature pool of this variety includes one or more morphosyntactic features characteristic of its (part of a) world region, but not documented at all for the typologically most closely related varieties/Pidgins/ Creoles in other parts of the world. For British Creole the most important test is, of course, Jamaican Creole, from which it derives historically. It turns out that British Creole has F 155 (ain’t for negated be) and F 156 (ain’t for negated have), both of which are typical of East Anglia and the dialects of the Southeast of England, but neither of which is attested in Jamaican Creole (nor in Jamaican English). A similar ‘areal case’ for certain morphosyntactic properties as for British Creole can, for example, be made for the West African Pidgins, i.e. for Cameroonian Pidgin, Ghanaian Pidgin, Nigerian Pidgin, and Vernacular Liberian English. Thus F59 ‘Double determiners’ and F116 ‘come-based FUT/INGRESS markers’ are documented in all four West African Pidgins as well as in Krio (the only West African Creole), but virtually unattested in Pidgins and Creoles elsewhere in the world. Moreover, it is important to note that exactly these two features are (with only one exception for each feature) also found in the West African L1 and L2 Englishes, only rarely in the East African L2s and not at all in the South African varieties.

F 59  double determiners (elsewhere: only in Saramaccan)

Dí yò tú bóy pikín dèm chóp. (NigP)
This 2SG POSS two boy child 3PL eat
These your two boys ate.
F 116  come-based FUT/INGRESS markers (elsewhere: only in Bislama and Butler English)

\[ ìm \quad kóm \quad yélo \quad wélwél. \quad (NigP) \]

3sg.SJ become be.pale thoroughly
He/she/it became very pale.

What the areal clusters for individual Anglophone world regions –, especially for the dominantly or exclusively non-L1 world regions, as well as the areal sub-clusters briefly addressed above – are taken to indicate is the importance of language contact and thus substrate influence. The substrate argument can, and has been, forcefully made for a range of properties of West African Pidgins (e.g. by Faraclas 2012 for Nigerian Pidgin which he sees as clearly influenced by the structural properties of the languages of Southern Nigeria). The argument can also be made for deletion features in the varieties of English spoken in Asia (recall our discussion concerning Tables 3 and 4 on pronoun and article deletion), and for features like F36 ‘Distinct forms for inclusive/exclusive first person non-singular’, F37 ‘More number distinctions in personal pronouns than simply singular vs. plural’ and F143 ‘Transitive verb suffix -em/-im/-um’, all of which are almost uniquely found in the Australia and the Pacific region, thus constituting a powerful reflex of the structural properties of the local languages.

5  Conclusion

We are convinced that future studies are bound to reveal more, and possibly more powerful, cases of substrate impact on the morphosyntax of varieties of English and English-based Pidgins and Creoles, thus at the same time identifying more and, possibly, stronger geographical, even downright local, signals in the observable variation worldwide. In order for this to happen, though, different data from those compiled in WAVE are necessary (for a full account of this line of argument see Kortmann 2013: 185-190). WAVE offers a typologically informed survey at a macro-level. It cannot zoom in on, e.g., the entire range of specific (free or bound) markers used for the coding of individual grammatical categories, or of specific constructions like certain verb- or adjective-specific complementation patterns, found in the extremely complex lexis-syntax interface. Least of all does WAVE take into consideration local usage (including pragmatic and stylistic) preferences and constraints, which are known to play an important role, however, in the formation and stabilization of Postcolonial Englishes (cf. Schneider 2007: 87). For the exploration of morphosyntactic variation at such a micro-level, other research tools and data sets need to be exploited (or to be compiled in the first place). The available corpora on many varieties of English are certainly a good starting-point. However, what the WAVE-based areal features, which we have identified in this chapter (see Tables 16 and especially 17), are a powerful indicator of is the path the Englishes in the different Anglophone world regions are going to take in the development of distinctly pan-continental (or: world-regional) Englishes (e.g. in the development of a distinctly Asian, African, or Pacific forms of English).
References


Wolk, Christoph 2013. Integrating aggregational and probabilistic approaches to dialectology and language variation. PhD thesis, University of Freiburg, Germany.
11 Slavic languages

Alan Timberlake

A concrete question: how and when, and in what context, did those languages contiguous with Eurasian territory acquire or develop or preserve an opposition in consonants of soft [palatalized] vs. hard [unpalatalized]?

Roman Jakobson (1931: 185)

1 Introduction

The precipitous expansion and subsequent history of the Slavs brought them into linguistic encounters with other peoples and languages, including some that might qualify as linguistic areas, in the sense of areas of intense convergence among typologically diverse languages. The discussion below examines encounters of Slavs with others to determine not only what the results of convergence are but also “in what context” convergence occurred: the modes of convergence and the communicative situations in which people and their languages interacted.¹

2 The Southeastern frontier

In an area we might call the Ukrainian Mesopotamia – from the Dnestr River over to the Dnepr, and from the southern edge of the basin of the Pripyat River down to the littoral of the Black Sea – culture complexes of pottery shaped like a tulip bulb, subterranean dwellings, and burial by cremation were continuous from 1500 BCE to the beginning of the common era. Because the same traits show up later among people we know to be Slavs (in Bohemia, Danube basin, Greece), we can identify the people with this culture as Slavs.

From 500 BCE on various peoples intruded on this Slavic homeland: Iranians (Scythians c. 500 BCE, then Sarmatians c. 200 BCE), Greeks on the Black Sea, Romans on the lower Danube. Goths arrived in 166 CE but were chased out by the nomadic, westward-bound Huns in 376. During the Gothic interlude Slavs apparently moved to the periphery of their homeland, establishing settlements in the west on the upper Dnestr, in the center in the Pripyat basin (previously inhabited by Balts), and on the east on the far side of the Dnepr (Barford 2001: maps II, III). It was from these newer outposts on the periphery of the ancient homeland that Slavs then moved to other areas (Gimbutas 1971:80).

¹ The approach here resonates with Bickel and Nichols (2006), who would define language areas “based on a theory of population and language spread.”
During the period from 500 BCE to 500 CE, Slavs had contact with others and borrowed lexicon: from Goths, the lexicon of technology and household (*stǻ̄vъ < Gothic *hlaifs ‘loaf’, *kotlъ < Germanic *kātil(u)s ‘kettle’, *dulgs < Gothic *dulgs) [Gimbutas 1971:77]) and from Iranians, possibly the lexicon of religion (*bogъ ‘bestower of bounty, god’ and *bogatъ ‘blessed by beneficence, rich’, akin to Avestan *baya ‘lord’, Sanskrit bhagas ‘donor’). Aside from lexical borrowings there is no evidence of areal convergence between Slavs and either Goths or Iranians.

3 Western frontier

From the periphery of their homeland in the Ukrainian Mesopotamia Slavs moved in at least three streams: south and west along the Danube and over to the Adriatic, around the Carpathians into Bohemia, and up the Dnepr north to the Baltic region.

Along the western frontier Slavs came into contact with other peoples with developed political structures, commerce, and confession. In the south of the western frontier, Slavs came into unsurprising contact with Romans and their descendants who colonized the coast. In the middle section of the Western frontier, in Slovenia, Moravia, Bohemia, and Sorbia, Slavs were subject to pressure to convert from new sees in Regensburg, Salzburg, and Freising. Confessional conversion went hand-in-hand with the declaration of political fealty – as in the 890s when Bohemians adopted Roman Christianity and made an alliance with Bavaria. Such dual conversion brought in its wake colonization by western peoples and led to language contact and, not infrequently, to a shift in language from Slavic to German.

Possible contact effects of the first five hundred years of contact between Czech and German have been reviewed by Berger (2014). Czech lost palatalization as a binary phonological opposition. Czech diphthongized non-front high long vowels, as in


Middle High German durfen and müzezen were borrowed into Czech directly (as drbiti and musiti), enriching the class of modal verbs. Native míti ‘have’ extended its range of usage as a modal and as the means of expressing predicative possession. Originally Czech expressed predicate possession using the dative (Vulgate *et huic erat soror > Dresden Bible (1360) *A tej bieše sestra ‘and to her there was a sister’ [Lk 10:39]) but under German influence switched to míti ‘have’ (Greek καὶ τῇ δε ἀδελφῇ > Kralice Bible (1579–96) *A ta měla sestru ‘and she had a sister’ [Lk 10:39]) (McAnallen 2011:27–32).

In syntax, Czech reduced certain uses of cases, probably under German influence: the instrumental case for predicative nouns and adjectives and adverbal genitives with verbs like pozorovati ‘observe’, ždáti ‘await’, přáti ‘wish for’, brániti ‘defend’. In

---

2 It is usual to cite Slavic etyma in the form they had in Late Common Slavic (c. 800 CE), using some idiosyncratic symbols: Roman «y̥ ę» [i xe], «c š ź ě» [ts ź tf] and the Cyrillic “jer” letters «ъ» [у], «ь» [и].
Russian, the preposition v ‘into’ with an illative sense of entrance into a domain is vastly more frequent than the allative preposition do ‘up to, into the vicinity of’ (84% v vs. 16% do). The preference is reversed in Czech (85% do vs. 15% v), under influence of German nach.

The Czech construction in (1) – in which the patient of an infinitive is the subject and agrees with the copula and adjective – is likely calqued from German constructions like (2):

(1) Práce_{fem.sg} jest krásn{fem.sg} hlédati

work is beautiful to observe

‘the work is beautiful to observe’

(2) Emotionsgeladene Erinnerungen sind schwer zu vergessen.

‘Emotional memories are difficult to forget.’

4 Northern frontier

From the Elbe to the Oder the Slavic population was trapped between Germans on the west and Poles on the east. The Liutici, a confederation of Slavic tribes, attempted to fend off the outsiders (983 to the middle of the 1100s), but were ultimately subdued (Brüske 1955), leaving the Slavic population exposed to settlement from the West and language shift to German. The last Slavic speaker on the Slavic island of Rugen, Frau Gulitzin, died in 1404. The shift from Slavic to German is substantiated by names like Ivan von Belaw – both first and last names of Slavic origin (1411, Witte 1906: 158)–and lexemes recorded in the Stralsund glossary (1460s): iuche ‘soup’, stupe ‘pillory’, bruche ‘belly’, seze ‘trawl’ (Damme 1987). Such lexical relics imply that, during language shift there was a period of bilingualism when speakers of the indigenous language, and then speakers of the intrusive language, used and preserved lexicon (and syntactic constructions) of the receding language.

The teleological process of colonization in Mecklenburg-Vorpommern was continued to the east by two institutions. The Hanseatic league of cities developed out of the charismatic practice of trade and marauding of the Vikings in the Baltic region (first centered in Wisby on Gotland off the coast of Sweden and later in Lübeck). The Teutonic Knights, formed from the residue of the failed Third Crusade of 1190–91, engaged in the sacred task of converting the heathens (Slavs, Balts, Fennic peoples), from the Oder across Pomerania through West and East Prussia all the way to Livonia. The commercial and confessional intrusion was supplemented by populating the newly converted lands with settlers.

The language of the Hanse is usually said to have been the Low German of Lübeck, but more important than the specific origin is the fact that Low German functioned as a lingua franca for the Baltic region. This lingua franca was receptive to the influence of other languages, Scandinavian (especially Danish) and Livonian; it accepted borrowings in lexical fields such as: goods (fur, clothing), currency, transport and shipbuilding, administration (Otsmaa (1975: 107) – that is, the terminology of trade, which after all was the primary communicative function of this lingua franca. Writing (correspondence

3 According to the ducal chronicler Thomas Kantzow writing in the 1530s (1816: 436–37).
4 See Dahl 21001 on the role of Danish as an early Baltic lingua franca.
or records) was originally in Latin. Latin was superseded by Low German by 1400, which in turn yielded to High German by 1600. All the way from the Elbe to East Prussia, the indigenous rural population on the littoral largely adopted Low German. When a Slavic pocket did survive, the grammar converged with German; Slovincian developed new perfects using auxiliaries ‘have’ and ‘be’ (Lorentz 1903: 11). In contrast to Pomerania and Prussia, indigenous languages survived in Lithuania, Latvia, and Estonia.

In the Livonian cities of Riga and Reval, locals provided services for the German elite and staffed their households; to do so, they learned Low German to some extent (Johansen and von zur Mühlen 1973: 372). After 1494 (when Ivan III closed the Hanse outpost in Novgorod) Russian trade had to go overland through Riga and Reval, making these cities intermediaries in the lucrative trade between the Hanse and Russia. Inhabitants included Danes, Swedes, and Finns. The cities were multilingual; there was an “accommodationist”\(^5\) disposition to multilingualism – an acceptance, even embrace, of linguistic pluralism.

The vernaculars of Riga and Reval were influenced by the Hanse lingua franca. Latvian lost final syllables and introduced a “broken” tone with a glottal catch, features reminiscent of Danish. Estonian experienced Low German influences such as: verbal prefixes (ära- ‘through’, ümber- ‘around’, üle- ‘over’) – atypical of Fennic languages; the demonstrative see used as an article and borrowing of the conjunction und – two core grammatical functions considered resistant to borrowing; and borrowing of doch ‘indeed’, trotz ‘notwithstanding’ – dialogic discourse operators which tell the interlocutor how to evaluate information (Johansen and von zur Mühlen 1973:370–71). These features presuppose a complex linguistic dynamic: speakers of the recipient language imitate Low German as they attempt to speak it, but more than that, they internalize patterns from the donor language and extend those patterns to innovations in their own language. Livonia continued to be an area of multilingualism, when, later in history, it was subjected to the Russian imperial language and circum-Baltic areal developments.

The easternmost outpost of the Hanse was Novgorod. The relation between the Hanse merchants and Novgorod was unique in the Hanse. Merchants did not colonize the foreign city. Because Hanse and Russian merchants did not learn each other’s language, the conduct of trade required translators, trained in Novgorod or later in Reval.

In contrast to Livonian languages, Novgorodian Russian was subjected to only modest influence from Low German. Lexical borrowings are not extensive (Koškins 1996;87). Some calquing did occur. A bilingual agreement from 1392 guarantees access to Novgorod with its flat landscape \(\text{goroju i vodoju} \) ‘by hill and by water’ (pp. 94–95). The phrase evidently calques Low German \(\text{to lande und to water} \) (itself based on Latin \(\text{per aquam ant per terram} \)). Novgorod commerce used two phrases for complex arithmetic calculations. One is diminution of a round number, for which Low German used \(\text{min} \) (\(\text{minn, myn} \)), glossed by Lübben (1888: s.v. \(\text{min} \)) as “weniger, geringer, minder”; from correspondence, note \(128 \text{ mark myn } 1 \text{ fr} \) (‘128 marks less one franc’). Novgorod has (3):

\[
(3) \quad \text{о: нёшку } \text{дова дисяь } \text{брковько } \text{бз брковько}
\]

\[
\text{from N twenty bundles less bundle}
\]

‘from N [I took] twenty bundles less one’ [Zaliznjak 2004:B13, #630, XII\(^1\) c.]

\(^5\) Substituting a different term for what Hill 2001 calls “distributive.”
Another calque may be the expression of intermediate quantity, as in *polo tratiu dojsjatinu* ‘half of the third decade’ = ‘25’ (Zaliznjak 2004: §4.11), parallel to Low German *achtehalf dasent rynsche Guldene* ‘eight half thousand [= 7500] Rhine guilders.’

The history of the Baltic region shows the power of a supra-regional lingua franca – initially developed for trade but invigorated by confessional conversion and outright colonization – to spread norms of language. In the west and center the indigenous languages were simply replaced by Low German. In Livonia the norms of the lingua franca were internalized by the local languages, Latvian and Estonian, and, more broadly, the lingua franca fostered a linguistic disposition of multilingualism and convergence that persisted. In Novgorod there was minimal convergence with the lingua franca or areal trends.6 The importance of linguae francae will reappear in the Balkan linguistic area below (§§8–9).

5 Eastern Baltic region, I

The eastern Baltic region, broadly understood, has been and continues to be an area rich in language encounters. The encounters are of different ages, and of different kinds.

The oldest layer is a pair of uses of case shared by Baltic, Slavic, and Fennic. These languages use a case other than accusative – genitive in Baltic and Slavic, partitive (distinct from accusative or genitive) in Fennic – for objects of predicates expressing restricted effect, including: a portion of a mass (traditional partitive meaning), static relations with no agency (emotions, perceptions), partial but incomplete activity, and negated predicates (both transitive verbs and existential verbs). In Slavic the genitive of negation with intransitive existential verbs is well preserved, while the object genitive of negation is not equally attested in Slavic languages. It has declined in Czech, somewhat in East Slavic, not at all in Polish, and it is gone from southwestern Slavic (Willis 2013).

In Indo-European the use of the genitive for objects of negated verbs is unique to Baltic and Slavic. In Fennic languages (including Finnish) the partitive is used widely, even more than the genitive in Baltic and Slavic; thus the partitive can be used with a single definite patient when the event is incomplete. Also, it is not usual for linguistic features originating in Baltic or Slavic to penetrate Finnish. This suggests that the genitive (/partitive) of negation was indigenous to Fennic and was borrowed by Baltic and Slavic. When is unclear.

The second instance is the use of cases other than nominative – essive and translative – for predicatives. In Finnish, with the copula and a predicative noun (‘s/he is/was a teacher’), the nominative is used overwhelmingly in the present tense (93%) and predominantly in the past (66%).7 In its local sense, the essive case in [-na] expresses

---

6 The oldest Russkaja Pravda might have arisen to adjudicate conflicts between Novgorodians and Jaroslav’s unruly Varengian (Scandinavian) entourage around 1015. (If so, it would date to the pre-Hanse period.) Because the code is formulated from the perspective of Novgorodians, it is generally assumed that the Russkaja Pravda (law code) is the codification of an autochthonous oral tradition. Yet it shows similarity to Western law codes, specifically that of Saxony (Feldbrugge 2009:53–54, with references).

7 Google searches conducted 24.09.2014.
location within a container; with a predicative noun, the essive indicates a relationship contained within a time interval, as for example Artturi Laitinen’s teaching post (4):

(4)  

\[ \text{he was a teacher in the Utajärvi parish school from 1914 to 1943} \]

The translative (in {-ksi}) marks a temporal boundary: *maanantaiksi* ‘by Christmas’

With a predicative, it indicates a state with (an initial) boundary – “the state into which anyone or anything passes” (Eliot 1890: 157); it is used with verbs that signify transformation, as in *Hän tuli opettajaksi* ‘S/he became a teacher’ (439 of 2571, or 17%).

In Lithuanian, the instrumental indicates transformation. In Russian it indicates temporal or modal limitation on the state. The case has been productive in Polish. It is not well-developed in South Slavic, suggesting only partial development in Slavic. In fact, the instrumental was rare in early East Slavic texts (twelfth-century Kievan Chronicle) even with verbs of transformation. Thus the development of non-nominative case is restricted in Baltic and partial and late in Slavic, while in Fennic non-nominative predicative cases developed organically and autochthonously from local senses of the cases.

These cases uses – genitive (/partitive) of negation and non-nominative predicative – might seem to be areal, but Czech, which lies outside of the Baltic area, has both; they must be due to contact convergence with Fennic. When is unclear. One might think of an encounter deep in prehistory (Thomason and Kaufman 1988: 238–251), except that both cases are inconsistent and recent in Slavic. Another hypothesis (Andersen 1996) is suggested by a phonological feature shared by Baltic and Slavic. A number of Slavic etyma begin with the reflex of a short non-high vowel (j)e or o. Examples are: ‘lake’, Russian *ozero*, Czech *jezero*; ‘deer’, Russian *olen*; ‘alder’, Russian *ol’xa*, Czech *olše*, Polish *olcha*, Slovenian *jelša*, *olša* (dialect. *olša*), Serbo-Croatian *joxa*, Bulgarian *elxa*. The *e* is original (and *j* is a prothetic glide); there must have been a change of *#e* > *#â* > #o. To judge by these examples, the change was frequent in East Slavic, less so in central Slavic (Czech, Slovenian), and infrequent on the peripheries (southeast Bulgarian, northwest Slovincian-Kashubian). The challenge is to account for the variation.

Internal accounts (Ablaut variants or some version of *je > jo*) do not work. It turns out there is an analogous change of *#e* > *#â* in Baltic, and this change also shows variation: ‘ice-hole’, Lithuanian *akte* (dialect. *aket*); ‘bass’, Lithuanian *ažegys*; ‘lake’, Lithuanian *ažeras* (dialect. *ežeras*), Latvian *ezars*, Old Prussian *Assaran*. Slavic borrowed this change and variation from Baltic.

When and how? Modern Slavic has preserved etymologically Baltic hydronyms from the southern edge of the Pripjat basin (51ºN) up to the Western Dvina (Daugava) River (55ºN). The fact that such Baltic hydronyms emerge in Slavic implies a specific historical scenario. Balts once inhabited that area; Slavs moved in and absorbed the Baltic population. During a transitional phase of asymmetric bilingualism (as is characteristic of a substratum), hydronyms and other features were absorbed by Slavic. The geographical gradation in reflexes of *#e* > *#â* > #o – most in East Slavic, fewer in Czech, least on the peripheries – suggests the need for a more articulated model of migration than simply an all-or-nothing process: instead, scouts explore and set up outposts; subsequent migrants strengthen outposts and fill in gaps; return to the origin is possible (Anthony
1990). In such a model, the innovation *#ē > *#ā > #o could take place in the center at the beginning of migration and then reach the mid zones in weakened form and appear minimally on the peripheries. This model could also be used to explain the gradation in case usage if we posit two layers of substrata: case uses (genitive of negation, predicatives) were adopted by Baltic in Belarus from a prior Fennic substratum; then Slavs took these case uses from Balts, not directly from Fenns. The geographical gradations are due to substratum plus migration, not areal spread.

6 Eastern Baltic region, II

A slightly younger development which spread throughout the East Baltic area (Lithuania, Latvia, Estonian, Finnish, North Russian dialects) is the construction called the nominative patient (or object) with infinitive. We illustrate with North Russian. Patients of personal verbs – finite personal verbs (5) and infinitives dependent on finite personal verbs (6) – are accusative:

(5) *ikonu(acc) pogrebli bjaju sъ mertvecemъ
icon had.buried with deceased
‘an icon_{acc} they had buried with the deceased.’ [VoprKir 37]

(6) *I povelèbdati opitemju_{acc}
And he.ordered do penance
‘And he ordered [him] to do penance_{acc}.’ [VoprKir 59]

In contrast to the banal use of the accusative with personal verbs, patients of “impersonal infinitives” – independent infinitives with modal force not dependent on any matrix predicate (7) and infinitives dependent on impersonal matrix verbs (8) – are nominative:

(7) *Kako dvržati imъ opitemja_{nom}?
How observe them penance
‘how it is [necessary] for them to observe penance_{nom}? ’ [VoprKir 59]

(8) *Dostoito li popu svojei ženě molitva_{nom} tvoriti vsjakaja?
fitting Q priest own wife prayer make any
‘is it fitting for a priest to say any sort of prayer_{nom} for his own wife?’ [VoprKir 29]

---

8 Gimbutas (1971: 93, 97) insists that there is no archeological record of Slavs in northern Belarus (Dvina basin) before the ninth century, which would be too late for these innovations to have spread. The transfer must have occurred earlier and then in the south, in Polesia and the Pripyat basin. For the distribution of Baltic hydronyms, see Toporov and Trubachev 1962: maps 16, 18.

9 Examples cited from the oldest manuscript. In fact, the use of the nominative patient is predominant but not exclusive; it is used on 14 of 19 occasions. It is unclear whether this mixed usage represents originally optional usage or reversion to accusative in copying. Later manuscripts are known to preserve more archaic morphology (Gippius 1996), which use the nominative consistently.
Impersonal infinitive constructions often include a dative, which is both the goal of modal force and the tacit agent of the infinitival action – *popu 'priest*' in (8). In the same context, animate nouns (which substitute genitive for accusative) take the accusative (9):

(9) Ne lučė Bogamoacc moliti
not better God beseech
‘Is it not better to beseech God* _amoacc_ ’ [VoprKir 53]

Lithuanian dialects use nominative for patients of impersonal infinitives, accusative for other patients. Latvian has a reflex of the nominative patient construction in its debitive.

(10) man {Baumanisnom / viņšnom / teviacc } jādbtv-dzird!
to.me {Baumanis / he / you } listen-to
‘it is necessary for me to listen to dbtv {Baumanisnom / himnom / youacc }!’

The debitive, derived from an independent infinitive (Endzelin 1901), is an impersonal verb whose patient is nominative unless it is a first- or second-person pronoun (10).

The nominative patient with infinitive (or debitive) is unusual from the perspective of Indo-European languages, so much so that some investigators are inclined to derive it from a more palatable construction like that of Czech (1) above. But unlike the Czech construction, in which the patient really is the matrix subject and triggers agreement in the predicative adjective (‘the work is beautiful to look at’), the nominative patient is a not the subject but an object, as various syntactic properties show: 10 (a) a matrix predicate (adjective, copula) does not agree with the noun; (b) the patient is genitive under negation, whereas the subject of copular sentences like Czech (1) could never be genitive; (c) some patients with infinitives – pronouns or animates – are accusative and not subjects; by parallelism, neither are the nominative nouns. One could say that the nominative was subject at some prehistoric stage, but that would lead to positing a change from nominative subject to nominative object – in other words, a change towards the very construction deemed anomalous.

In truth, the construction does seem anomalous from an Indo-European perspective. But rather than denying the construction, we might rather look for its origin in convergence; convergence can lead to results uncharacteristic of the borrowing language (such as German-style prefixes in Estonian). In East Slavic territory, north of the belt of Baltic place-names, there is a belt of hydronyms with Fennic etymology. This belt starts above the Valdai Hills at 55°N, the dividing line between watersheds and between northern and central dialects of East Slavic. The non-native (from the point of view of Slavic) hydronyms show the existence of a substratum here. Indeed, considerable literature has pointed to the existence of a Fennic substratum for northern dialects of Russian (notably Veenker 1967). Such evidence includes *cokan’ë* (non-distinction of affricates *ʦ* and *ʧ*); absence of atonic vowel reduction (prominent in former Baltic areas of Belarus); existential-possessive sentences using the adessive case or preposition (McAnallen 2011). On the adessive in Finnish, Eliot (1890:151) says it is used (of a person) to ‘signify with, near, or in the house of’; he even glosses his example with

10 Arguments are presented in Timberlake 1974: chapter 2.
French *chez poika on meillä* ‘le garçon est chez nous’. He goes on to say, “from an extension of this use [the adessive] comes to denote possession,” citing the example *on-ko teillä saksta?* ‘have you scissors?’. Both meanings of the adessive – sphere of influence, possession – are expressed by the preposition *u* plus genitive in modern Russian: *v gostjax u Ivanovyx* ‘…as guests *chez* the Ivanovs’, *u Ivanovyx est’ kvartira v Kaliningrade* ‘the Ivanovs have an apartment in Kaliningrad’. In Slavic the extended use of the adessive preposition *u* appeared earliest in Novgorod and subsequently spread to Moscow and the south, but not beyond Russian; its original locus was North Russian.

Against this background we might consider patients in Finnish (only singular nouns below, since nominative and accusative are not distinguished in the plural). The accusative is used for patients in structures headed by a personal verb (one with a subject), either a finite verb itself (11) or an infinitive dependent on a personal finite verb (12):

(11) *[tractor], jolla hän keväisin kyntti maan*acc.
    "[tractor], by which he plowed the land in the spring"

(12) *Hän haluaa kyntää*inf *maan*acc
    ‘He wants to plow the land*acc*

But the patient of an infinitive dependent on an impersonal verb is nominative (13):

(13) *(Minun) täytyy kyntää maan*nom
    ‘It is necessary for me to plow the land*nom*

Pronouns in Finnish (third- as well as first- and second-person) have a distinct accusative form (distinct from the genitive), which is used even for patients of impersonal infinitives. Compare (14) with (13) above.

(14) *Minun täytyy kutsua {sinu*acc* / hänet*acc* / teidät*acc* }*
    ‘For me it is necessary to invite {you / her / you all }’

The nominative patient is layered form of differential object marking: the accusative is used except when transitivity is reduced (Hopper and Thompson 1980) by the necessary absence of a subject in systematically impersonal contexts (Timberlake 1974). But as a subrule, normal (accusative) case marking is invoked even in impersonal contexts if the object is highly individuated (pronominal or, in Russian, animate). The nominative is also used in Fennic for the object of an imperative, when the subject is predictable from the verb form itself and the predicate is thereby of lower transitivity.

There is every reason, then, to believe that the nominative with infinitive in North Russian, Lithuanian, and Latvian derives from Fennic substrata, not areal influence. Given that the construction is geographically limited in Slavic, the convergence must be have come directly from Fennic population in the very north of Russia, not from the intermediate Balts.
7 Eastern Baltic region, III

Other, more recent, contact in the eastern Baltic region looks more like areal spread. Lithuania was able to deflect the pressure of western colonization. It defeated the Teutonic Knights in 1238, then joined Poland in a dynastic alliance (Grand Duchy of Lithuania), which became the Polish-Lithuanian Commonwealth in 1569, including Belarus and Ukraine. This alliance defeated the Knights at the Battle of Tannenberg in 1410, setting the Knights on a downward spiral. The Commonwealth was broken up in the partitions of Poland in the 1790s, when much of the region came under tsarist control.

The Grand Duchy, then Commonwealth, became a small linguistic area in its own right. Notable is the development of an impersonal passive of transitive verbs, impersonal in the sense that the patient remains in an object case: accusative in Polish, nominative in (older) Lithuanian. The competing personal passive reports a change of state in an independently established patient, while the impersonal present a holistic view of the occurrence of an event including the patient. Thus in Lithuanian, (15) describes what is proper treatment as measured in terms of a specific type of horse; the passive participle agrees with ‘horse’:

(15) nūo 4–4,5 metų /uarklys/nom.msc.sg
    gali būti /jojamas/nom.msc.sg
    trumpais /periodais

    ‘from 4 to 4.5 years the horse /nom.msc.sg may be ridden /nom.msc.sg for short
    periods’

(16) Žiūriu – jo /arklys/msc.sg
    jojama /ntr.sg
    į ganyklą.

    ‘I look –by him there’s horse -riding into the pasture’ [Jablonskis 1922: §176D]

In (16) the focus is on the speaker’s observing an event, not on the horse. The patient is
nominative, but the participle does not agree with the nominative noun.

Polish and Lithuanian form impersonal passives from intransitive verbs as well.
Passives of intransitives are frequent in Lithuanian (Jablonskis 1922: §177D), with the
agent in the genitive; they have the overtone of evidentiality, as in (17), where tracks in
the dirt invite an inference:

(17) Jo /gen.sg ėia
    per grijovą /šokta/ntr.s
    per griovą /šokta/ntr.s
    /him here over ditch jumped/ntr.sg
    ‘by him it was apparently jumped over the ditch here’

As an exuberant extension (18), a passive of copular ‘be’ (būta) can be formed:

(18) Bilieto /msc.gen.sg
    būta /ntr.sg
    laimmingo /msc.gen.sg
    ticket be happy
‘The ticket evidently turned-out-to-be happy’ [Eugenijus Ališanka (21.01.2015)]

The impersonal passive (of transitive verbs) is seen in northwestern Russian dialects (Pskov, Novgorod and its outposts), as in the idealized example in (19) (inspired by Kuz’mina and Nemčenko 1971 27, 35, 38, 85). The patient is expressed in the nominative in areas with the “nominative with infinitive” construction but in the accusative elsewhere. The participle and copula (if present) are neuter singular with either case marking; in this sense the construction is impersonal:

(19) (U nas) {kartoškařem.nom.sg / kartoškuřacc.sg } zakopanoňtr.sg
byloňtr.sg
(chez nous) potato planted was
‘(by us/among us) there was potato-planting’

Fennic languages have a morphologically invariant verb form expressing unspecified agency whose patient is nominative. It is difficult to say if this Fennic impersonal passive is responsible for the construction in the Commonwealth area; the counter-indication is that the impersonal passive skips Latvian (normally included in areal innovations of the Baltic region) but includes Ukrainian (not normally part of Baltic areal innovations). Apparently it was a feature not of the larger eastern Baltic region but specifically of the Commonwealth.

A shared development of the Eastern Baltic region is the expression of mediated epistemology of events (quotation or sometimes inference based on evidence) by means of participles that are not attached to finite verbs. This modus relativus occurs in the two Baltic languages (it is used repetitively in Latvian folktales) and in Estonian and Finnish. Its origin is hard to date, but it does seem to be a genuinely areal phenomenon.

A possibly related Russian development is the use of (morphologically invariant) participles as autonomous predicates without a finite verb:

(20) Prišli bumagi, čto ubivši pcp ego
arrived papers that killed him
‘documents have arrived, [saying] that [they] had killed him’
[Kuz’mina and Nemčenko 1971:128]

(21) Kotorye byli ušotci pcp v Ščepcy tye pogibši pcp
whoever had gone to Ščepcy those perished
‘whoever had gone to Ščepcy, those have perished’
[Kuz’mina and Nemčenko 1971:116]

While (20) might qualify as quotative (it reports the content of the papers), this construction seems rather to have a perfect or resultative sense (21). It is found most intensively in Russian dialects from latitude 56º up to 60ºN and from longitude 28º to 36ºE (Kuz’mina and Nemčenko 1971: karta-sxema 1). It appears to be a Baltic areal phenomenon spread into adjacent areas of Russia.

Numerous developments are shared by Belarus dialects with adjacent areas, at least local convergences, conceivably areal convergences (Wiemer, Seržant and Erker 2014).
8 Southwestern Slavdom, I

The Balkans, or Southeastern Europe, or the southwest corner of Slavdom, has taken from Eurasia the title of language area par excellence. It covers more continuous area; it is more uniform not just in single features like consonant palatalization or evidentiality but in a whole range of features. Joseph 2010 gives a list, including: reduction of case, conversion of infinitive to finite complements, future tense formed with ‘want’, definite article, analytic comparative adjective, adnominal possession by clitic dative pronouns evidentiality, conversion of infinitive to finite complements, pleonastic (doubled) object clitic pronouns. All of these eight have to do with creating a particular typology of syntax organized as phrases. These are shared by all but two peripheral members of the area. The breadth and number of features and the degree of uniformity are remarkable.

The list of uniformly spread features gives the impression of timelessness and inevitability. Correlated with this, it is tacitly assumed that the spread of Balkan features was omnidirectional; indeed, it is sometimes asserted that the search for a source is futile (Tomić 2004). The danger here is that the emphasis on the synchronic results of convergence might obscure the processes and interactions that have given rise to the current situation. The discussion below attempts to develop a more articulated history.

At least two Balkanisms arose earlier in Greek than elsewhere: the analytic substitute for infinitives (Joseph 1983) and the ancient article, native to Greek. The fact that the article is pre-prepositional in Greek and post-positional in Slavic is minor. As the Slavic post-positive demonstrative gradually acquired the functions of an article, there was no reason to change the order.

Let us consider pleonasm. Balkan languages (using Bulgarian as example) can express a direct object in three ways. First, as clitic pronoun without a noun – twice in (22):

\[
\text{(22) } \text{Sled kato go dovārša šte go pousna pod drugo zaglavie} \\
\text{After it finish want it release under another title} \\
\text{‘After I finish it I will release it under a different title’} \]

The second and third options are a noun with or without a clitic pronoun. In examples adapted from Schick 2000, a clitic pronoun is not used if the referent is indefinite non-specific (23). The pronoun is “optional” with specific indefinite and definite readings of the noun (24–25) and obligatory with a pre-posed (and unique in reference) noun (26):

\[
\text{(23) } \text{Tārsjax (edin) lekar} \\
\text{look.1sg (one, a) doctor} \\
\text{‘I am looking for [any person who fits the description of being] a doctor’} \\
\]

\[
\text{(24) } \text{Rada {go / 0} tārsi edno pismo} \\
\text{Rada {it / 0} look.3sg a certain letter} \\
\text{‘Rad is looking for a (certain) letter’} \\
\]

11 За истинските фенове на Токио Хотел! (25.01.2015).
(25) Rada {go<cl / *0} tarsi pismo
ti

Rada {it / *0} look.3sg letter.def
‘Rada is looking for the letter’

(26) Ivo {go<cl / *0} obića Rada
Ivo {him / *0} love.3sg Rada
‘Ivo, Rada loves him’

Abstracting from these specific examples, it is possible to say (in the spirit of Friedman
(2008)) that pleonasm is favored in proportion to: discourse prominence
(topicalized > non-topicalized), referential autonomy of the noun (definite > indefinite
specific > indefinite), and predicate-internal prominence (objects of impersonal
experiential verbs > ordinary transitives). These principles can be subsumed under a
general rubric of referential autonomy (Laca 2006); they are relevant as well to pleonasm
in medieval (Laca 2006) and contemporary Spanish (Bull and Carmen 2004:§11.6).

As always, it is instructive to find a moment when a construction is just coming
into being. For pleonasm the Bulgarian damaskini are useful. Orzechowska (1973)
examined usage in the Koprištenski and Trojanski damiskini, manuscripts from the early
seventeenth century that go back to common ancestor perhaps a century before.
Orzechowska focused on the stylistic value of pleonasm – pleonasm is informal and oral
while absence of pleonasm has overtones of formal and written – and did not analyze the
distribution of pleonasm in terms of syntax or discourse. Nevertheless she did cite ten
examples of the total of 23 tokens of pleonasm from the Koprištenski damaskin. If these
ten are representative of the whole text, then it is telling that, in nine of these ten
sentences with pleonasm, the noun is pre-posed, and six of those nine have the
conjunction da – at this time a major boundary between the noun and the verb. See (27):

(27) němu sн страх i снс радост da mu vздрчаваме
to-him with fear & with joy that him rejoice
‘to him, it is with fear and with joy that we rejoice to him’

To judge by Orzechowska’s examples, the locus of diffusion for pleonasm was
topicalized arguments. This suspicion finds confirmation in Greek (de Boel 2008).
Pleonasm was found in personal correspondence on papyrus up through the seventh
century CE, when the supply of papyrus was interrupted (p. 94). Pleonasm shows up
again in texts in the twelfth century in a new tradition of writing courtly poetry in the
vernacular (Alexiou 2002:100), notably in the twelfth-century Ptochoprodromic Poems
(=PtP) and Digenis Akritis, Escorial manuscript (=DA/E). Two examples cited by de
Boel (pp. 97, 96):

(28) Τὴν θίλασσαν τὴν μὲ ἔφερες, γνωρίζεις, ἔπαρέ τὴν
the red-petticoat it me you-brought you-know take it
‘the red petticoat you brought me it, you know, take it’ [PtP I.58]

(29) Τὴν δὲ προῖκα μου τὴν πολλὴν δὲ τὴν ἔξον οἱ γυναικάδελφοι μου
the but dowry my the rich let it have the wife-brothers my
‘the dowry of mine, the rich one, let my brothers-in-law have that’ [DA/E 1007]
On the basis of Orzechowska and de Boel, we can hypothesize that pleonasm began as a resumptive pronoun associated initially with a syntactically and intonationally detached topic. That is to say, pleonasm has its origins in actual speech activity, not in an abstract change of grammar; moreover, it is striking that many of the examples cited by de Boel involve imperatives, the ultimate dialogic form. Pre-posing the object to guarantee clarity and comprehension is an adaptation speakers make when speaker or addressee do not have full and equal command over the same language. On that reasoning, pleonasm is linked to a specific kind of interlingual communicative situation. This link to interlingual speech is reminiscent of the adaptations of Estonian to Low German (demonstrative as article, discourse particles), which imitate the speech of the interlocutor.

Following that logic, we would now want to identify situations of interlingual communication involving Greeks and Slavs. There were many, and they stretched over a long time. Slavs appeared in the sixth century in the Danube valley raiding forts south of the Danube and eventually settling south of the Danube, to judge by characteristic Slavic pottery and subterranean houses (Văzarova 1965). The Avars captured Sirmium in 582 and then enlisted the Slavs as allies or vassals for further conquest; together they laid siege to Thessaloniki using sophisticated siege techniques (in 597, according to Vryonis 1981). Subsequently – especially in the interval of 610 to 626 – Slavs conducted raids throughout Greece, all the way to the Peloponnesus south of the Gulf of Corinth.

The Slavs not only raided, but to judge by Slavic pottery and burial by cremation (Vryonis 1981:379), they settled. Settlement is also indicated by numerous Slavic place-names preserved in Greek (Vasmer 1941), among them names invoking land-clearing (Τερπίτσα < *terbica ‘clearing’ [p. 76]) and cultivation (Σβάρνα < *borna ‘harrow’ [p. 289]) (Curta 2011: 211–12). The oldest place-names fix a stage of Slavic earlier than any written source; for example, liquid diphthongs like Τερπίτσα and Σβάρνα still have the Common Slavic shape of *CVRC with the vowel before the liquid, whereas the rest of Slavic modifies that order. The mere fact that Slavic place-names could be preserved in Greek implies a process analogous to the process discussed above for Mecklenburg-Vorpommern (§4): the newer settlers (here Greeks) interacted with the sedentary population (Slavs); the new settlers and the resulting mixed community adopted linguistic relics of the disappearing ethnos.

In response to Slavic settlement, Byzantine authorities initiated a program of reclaiming territory by setting up administrative units (“themes”) and converting the pagan Slavs. Civil administration, conducted in Greek, was rationalized. Some themes were established at the end of the seventh century, others thereafter. As Obolensky (1971:78) summarized,

Thus, by the end of the ninth century, the Byzantines had succeeded in establishing administrative control over a string of themes which formed an almost continuous edging around the Balkan peninsula. In some of these districts, notably in Thrace and in Southern Macedonia, Byzantine power extended far inland.

The military and administrative aspects of this policy of “reconquest” were accompanied by a renewal of Byzantine Christianity and of the importance of the Greek language.

In the Slavonic lands now reintegrated into the framework of the Byzantine provincial administration, Greek was not only the idiom of the church but the
language of civil service, of the armed forces, of polite society; a knowledge of Byzantine Greek became henceforth the necessary key to social status and a successful career.

The combination of forces is strikingly similar to the combination of military, political, confessional, and linguistic pressures observed in Mecklenburg-Vorpommern and Pomerania. Thus there was extended linguistic interaction between Slavs and Greeks, and it was not unidirectional: Slavs invaded and inhabited Greek territory but then under pressure Slavs learned Greek, learning enough (or switching entirely) to participate in the up-and-coming prestige life of Greeks. (One is reminded of Constantine and Methodios who were said to know Slavic and whose father was a drangarios – could their parents have been assimilated Slavs?) Recall that, in Livonia, ethnic Estonians learned the Hanseatic lingua franca to some degree and allowed this language to influence patterns in their own language (prefixes, particles, word-level prosody). Conceivably the same thing happened here in the ninth or tenth century, when Slavs learning some Greek incorporated Greek patterns into their Slavic and Greeks dealing with imperfect speakers modified their language.

9 Southwestern Slavdom, II

The question then becomes how pleonasm, and all the other features attributed to the Balkan Sprachbund (Friedman and Joseph, this volume), were spread so extensively in Southeastern Europe. One possibility is a variation of wave theory, which, applied to this case, might attribute the overall uniformity to recursive processes of convergence from one village to the next, all throughout the Balkans. There are certainly micro-instances of micro-contact and convergence, which continue to this day. Yet it is fair to ask, is that how it happened? Would recurrent micro-convergences have been sufficient to produce the uniformity and geographical reach of Balkan features? In the case of Western Romance, there was a lingua franca that was responsible for spreading relatively uniform norms throughout Western Romance; similarly, in the Baltic area, from 1150 to at least 1550, some variety of Low German served as a lingua franca. These parallels invite us to consider the possibility that, in southeastern Europe as well, there was a lingua franca (or linguae francae) used across the Balkans. It is well-known that the Balkan area was an extremely fertile area for trade, as is discussed in the classic study by Stoianovich (1960). Trade grew especially in the interval from the fifteenth century through the seventeenth century, in concert with territorial expansion of the Ottoman Empire which continued until the Treaty of Carlowitz (1699).

Trade moved towards the capital of Constantinople, especially from the eastern provinces, but also in the other direction, from Constantinople towards Vienna, even reaching the Netherlands. Trade was controlled by the state. A patent was required to engage in trade: “Fur traders, salt farmers, purveyors of beef and mutton, and buyers and exporters of silk, wool, cotton, coffee, rice, oil, grains, wax, copper, lead, and saltpeter were all required to purchase government patents authorizing them to engage in the commerce of their choice” (Stoianovich, p. 241).

Evidently the merchants interacted with peasants and (in cities) with artisans. Stoianovich comments (p. 241), “Merchants charged with satisfying the needs of the
capital and state at first made their purchases in the port towns. Soon, however, they extended their operations into the interior.” A debt economy developed, whereby peasants would agree to low prices for future crops in order to have money to live on (p. 303). There was a route of seasonal fairs at least in Thessaly. This is all to say that the itinerant merchants interacted directly with the sedentary population.

We cannot be sure what language merchants and their clients used. There is more than one possibility. It is certainly possible that there were regional Slavic or Romanian trade languages. But for all the diversity of ethnicities involved in trade, the one constant throughout the whole Balkan region was Greek traders. It is plausible that merchants, on the whole, used Greek or a simplified form of Greek. “Trade Greek” would have had compromise syntactic structure, which is to say, it was highly phrasal.

Such a lingua franca of trade would have provided the region with (informal) norms to imitate. If so, the Sprachbund of the Balkans would look more like Livonia, where the resident urban population was multilingual and, if only for economic reasons, had an “accommodationist” disposition to imitate, internalize, and use the norms of the lingua franca. Unlike Western Romance, this attitude did not come at the expense of abandoning the indigenous languages.12

10 Eurasian linguistic area?

All study of areal linguistics harks back to Roman Jakobson’s study (1931) of the Eurasian Sprachbund [Evrazijskij jazykovoj sojuz]. In this study he focused on phonemic palatalization of consonants (as in Russian); yet at points he seems to include more banal palatalizations like \( ki > ji > \text{ʧi} (\text{ʦi}) \) or \( ti > \text{ʧi} (\text{ʦi}) \) and, further, palatalization is subsumed under synharmony (agreement of frontness and backness of vowels and consonants).

Jakobson had two broader concerns. He announces early on (p. 146) the importance of “correspondence – the tight principled connection among phenomena of different spheres.” In his poetic studies, he blurred the boundary between ancient and modern culture (Merrill 2012); he does the same here with geographical boundaries. Secondly, it is no accident that the language union discussed here is Eurasia; Eurasianism – the belief that Russia represents a unique synthesis of Orient and Occident – was in vogue among emigrés after the Russian Revolution (N.S. Trubetzkoy, D.S. Mirsky, P.N. Savicky). The quest for connectedness and the investment in Eurasianism may have influence Jakobson’s analysis.

Jakobson begins with the clear case of Slavic. Leaving aside banal palatalizations (velars before front vowels or *j, dentals before *j), dentals and labials were palatalized before front vowels. Palatalization became contrastive as a result of apocope of final high lax vowels seen in pairs such as *plotъ [plotU] ‘raft’ > Russian /plot/ vs. *plotь [plotᵢ]

12 There is a parallel in early Scandinavian. Dahl (2001) argues that the uniformity of Scandinavian languages cannot be explained by preservation of original uniformity from an Ursprache, and also not by an extended down-the-line convergence from village to village throughout Scandinavia (which would create difference). Rather, a more powerful mechanism is needed: a koine (lingua franca) that developed in tandem with regional economic activity which spread norms. The situation is analogous in the Balkans: it is uncertain whether recursive down-the-line convergence could produce the kind of structural homogeneity we see over the enormous space of the Balkan Sprachbund.
Distinctive palatalization occurs in many environments in Russian (less in Ukrainian and Bulgarian); in other Slavic languages palatalization has either disappeared (Czech, Serbo-Croatian) or has been transformed into affrication (Polish).

Jakobson then moves around the Eurasian landmass noting languages with palatalization. In Finno-Ugric, Permic languages and Mordva have palatalization (pp. 171–72). (A recent study says Mordva has both allophonic palatalization of labials and velars in position before front vowels and a distinctive series of palatal consonants contrasting with dentals, namely /zʲt, dʲl, sʲl, zʲl, cʲl, rʲl, nʲl/ opposed to /t, dl, etc. [Zaicz 1997:185–86].) Even Samoyed has palatalization. But Hungary was evidently too far to the west and Lapp too far to the north to participate in Eurasian trends.

Jakobson cites Lithuanian and Latvian as border languages which have developed palatalization under Russian influence (p. 181). Estonian fits in this category. Estonian (Viitso 1997: 123) has extensive palatalization, more in the south than in the north; it has become phonemic through apocope: *hullu > [hull] ‘crazy’ vs. *kulli > [kuːl] ‘hawk’. How old the distinctive palatalization is in Estonian is not clear. It might reflect a historically deep Finno-Ugric disposition (a possibility considered by Abondolo 1997b). Or Eurasian influence. Or, just as likely, more immediate contiguity and convergence with Russian.

Jakobson devotes much attention to Turkic languages, replacing palatalization with syllable synharmony: consonants and vowels harmonize as high tonality [= front] or low tonality [back]). The earliest Turkic writing, from the Orkhon and Yenesei valleys, does indeed distinguish two kinds of syllables (p. 187). He notes distinctive palatalization in Mongolian proper (Khalkha) and certain other Mongolic languages (Buryat, Kalmyk); his geography and chronology of “XIII–XIVth” centuries are consistent with contemporary work (Svantesson, et al. 2005). Jakobson goes a step further and projects synharmony back to Proto-Altaic and talks of Altaic languages realizing their Altaic destiny step-by-step. Here he slips in an aside that the southern neighbor of Altaic – Chinese (= Ch) – developed an opposition of palatalization (p. 181). Evidently Jakobson has in mind developments between Early Middle Chinese (=EMCh, Qieyun rime tables, 601 CE) and Late Middle Chinese (=LMCh, mid twelfth century), when labials were palatalized by a medial *j and then bifurcated depending on the following vowel (Baxter 1992: 203): front vowels absorbed the palatalizing agent: OCh *pjin ‘guest’ > EMCh pjìn > LMCh pjìn > bīn (?); in contrast, before back vowels, the palatalized labial became a labiodental: OCh *pjjang ‘region’ > EMCh pjjang > LMCh fāng > Ch fāng (??). Jakobson’s mention of geographical contiguity suggests that he thought Mongolian was responsible. But unless the Orkhon writing is meant, that would not fit chronologically, since Chinese labial palatalization occurred well before Mongolian palatalization.

These concerns registered, it is still true that it is hard to not be daunted by Jakobson’s documentation of palatalization or synharmony spread over an enormous area. The difficulty of pinning down interlingual relatedness and chronologies fits with modern understanding of linguistic areas as a force operating with less immediate and less visible contact than ordinary contact. It may also be that some features are easier (or require less contact) to transfer than others – classifiers, tone (Emeneau 1956), discourse particles, and phonetic biases like synharmony; these can all be noticed and imitated by non-native speakers.
11 Slavic linguistic areas?

Above we examined encounters of Slavs with others on the edges of Slavic territory, some of which look like areal phenomena. The interest was both typological – to see how these encounters fit in the spectrum of possible encounters – and genetic – to see how linguistic areas evolve out of other encounters.

In some instances examined above, encounters where features are distributed over multiple languages turn out to be better analyzed as substrata (Baltic, Slavic and Fennic interaction). Other encounters involved a lingua franca, which can indeed leave an imprint on multiple languages, but it is not necessarily a linguistic area. Both instances of linguae francae above are unusual, if not quite singular. The Baltic lingua franca occurred during dramatic growth of commerce, confessional conversion, and colonization. The Balkan lingua franca likewise was driven by commerce. These linguae francae evolved in the direction of becoming linguistic areas. In the Baltic region that development was partial. In the Balkans the development, promoted by centuries of shifts in language allegiance between Greek and Slavic, was complete. In the mature stage, the lingua franca recedes; multiple languages participate; and grammar is so homogenized it appears to have no origin.

In these encounters, all participant languages were affected to some extent, reflecting a dialogic quality to language contact and convergence. Superstratal languages seem to involve a power asymmetry, but even there, the receding language leaves a mark, implying accommodation of superstratum to substratum. Linguae francae not only absorb lexicon from local vernaculars but often (usually?) become compromise codes – speakers adjust to the imperfections of their interlocutors. The compromise code is obvious in the Balkans and Western Romance; Low German looks stripped down at earliest attestation. The compromise character of linguae francae arises from accommodations made in speech, such as the pleonastic – resumptive – pronouns at a distance and periphrastic futures; these accommodations become conventionalized features of the language and get extended.
References


12 The Caucasus

Sven Grawunder

This chapter seeks to describe the area of the Caucasus Mountains as a linguistic contact area, as has been suggested by a number of authors (e.g. Comrie 2008; Catford 1977; Chirikba 2008a). After sketching the migration history, historical contacts, and contact scenarios, including multilingualism, I will primarily focus on the occurrence and co-occurrence of contrastive sound features in language varieties described in the area so far (e.g. Alekseev 1999; Job 2004; Kibrik & Kodzasov 1990; Hewitt 1989; Harris 1991; Smeets 1994). Additionally, the chapter attempts to demonstrate how the frequency of these occurrences as well as cross-linguistic comparison of the phonetic realization of features can improve our understanding of the areal characteristics of phonological contrasts.

1 Defining the linguistic area – the Caucasus as an area of contact

A necessary condition for a linguistic area is the existence of similarities, e.g. in terms of shared linguistic features. Since these similarities can result either from common descent, from contact induced assimilation, or simply from chance and the general make-up of human languages (universalities), the emphasis lies on similarities that can only be explained by contact. By using areality as a vehicle for explaining higher or lower degrees of similarity, we should be able to infer degrees of contact with some caution. Like the Himalayas, the Caucasus, too, is categorized as a Eurasian enclave, i.e. an area of accretion (Nichols 1997), which is surrounded by areas of migration. This means that “(i)intrusive languages... do not replace other languages or families but are added to them. Thus the Caucasus tends to increase in genetic and typological diversity over time” (Nichols 1992: 16). From this perspective, the high mountainous areas work as barriers against migration and cause a difference in population shift and mobility between highlands and lowlands; therefore, different depths of contact can be proposed. Linguistically, the Caucasus area today contains a fair amount of genealogical diversity, with at least five different language families (Nakh-Dagestanian, Adyghe-Abkhaz, Kartvelian, Altaic, and Indo-European) comprising over 150 described varieties of around 45 languages. At least for the three inner Caucasian language families (Kartvelian = South-Caucasian, Nakh-Dagestanian = Northeast-Caucasian, and Adyghe-Abkhaz = Northwest-Caucasian) along the Great Northern Crest – a mountain ridge across the borders of Russia, Georgia, Armenia, and Azerbaijan – a very long period of contact can be assumed, which allowed the develop-

<table>
<thead>
<tr>
<th>Family</th>
<th>Sub-family</th>
<th>Branch</th>
<th>Subgroup</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>South-Caucasian</td>
<td>East Kartvelian</td>
<td>Georgian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>West Kartvelian</td>
<td>Mingrelian, Laz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Svan</td>
<td>Balian, Svan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northwest</td>
<td>Abkhaz</td>
<td>Abkhaz, Abaza</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>Circassian</td>
<td>Kabardian, Adyghe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Adyghe-Abkhaz)</td>
<td>Ubykh</td>
<td>Ubykh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Northeast Caucasian (Nakh-Dagestanian)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(North)</strong></td>
<td>Nakh</td>
<td>Veynakh</td>
<td>Chechen, Ingush</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avar-Andic-Tsezic</td>
<td>Avar</td>
<td>Avar, Zakhatli</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Andic</td>
<td>Andi, Akhvakh, Tindi, Bagvalal, Chamalal, Godoberi, Karata</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tsezic</td>
<td>Tsez, Bezhta, Hinukh, Khvarshi, Inkakhkhvari, Hunzib</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lak</td>
<td>Lak</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Central)</strong></td>
<td>Dargwic</td>
<td>North-Dargwa</td>
<td>Megeb, North Dargwa, Kadar, Maira</td>
<td></td>
</tr>
<tr>
<td></td>
<td>South-Dargwa</td>
<td>Kaitak, Sirhwa, Kunki, Lower Vurqni</td>
<td>Kubachi-Ashhi, Chirax-Amaq</td>
<td></td>
</tr>
<tr>
<td><strong>(South)</strong></td>
<td>Lezgic</td>
<td>Lezgi, Tabasaran, Aghul, Rutul, Tsakhur, Udi, Kryz, Badukh</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Khinalug</td>
<td>Khinalug</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Archi</td>
<td>Archi</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indo-European</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iranian</td>
<td>North-Eastern</td>
<td>Ossetic (Digor, Iron)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>North-Western</td>
<td>Talysh, Kurdish, Zazaki</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>South-Western</td>
<td>Tat, Persian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Armenian</td>
<td>Armenian</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slavic</td>
<td>East Slavic</td>
<td>Russian</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greek</td>
<td>Northern</td>
<td>Pontic Greek</td>
<td></td>
</tr>
<tr>
<td><strong>Altaic</strong></td>
<td>Turkic</td>
<td>North-Western (Kyrgyz)</td>
<td>Nogai, Kumyk, Karachay-Balkarian</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>South-Western (Oghuz)</td>
<td>Turkic, Azerbaijan</td>
<td></td>
</tr>
<tr>
<td><strong>Afroasiatic</strong></td>
<td>Semitic</td>
<td>(Central)</td>
<td>West-Central</td>
<td>Neo-Aramaic (Aysor)</td>
</tr>
</tbody>
</table>

Table 1. Languages of the wider Caucasus, combined from Hewitt (1981) and Koryakov (2006). Italic terms are either dialects or reflect a non-established division. Note that the depths of the branches may not be the same in all places.

The mention of wide-ranging structural similarities. Northeast-Caucasian (NEC) languages and Northwest-Caucasian (NWC) languages are often grouped together as North Caucasian languages for reasons of typological similarity. The Transcaucasia area at the southern edge leads from the South Caucasian (SC) languages into a transitional area of its own (cf. the ARAXES area proposed by Stilo 1994) that carries traits from the compact Caucasus area of the North into the adjacent areas in Turkey, Armenia, Azerbaijan and Iran. With Semitic Aysor (Neo-Aramaic, Assyrian) even a sixth language family comes into play (see Table 1) for this southern Caucasian area. And for the Northern plains, the Mongolian language Kalmyk-Oirat could also be added, representing another branch of Altaic.
1.1 Divergence vs. convergence – homogeneity vs. diversity

More or less independently of the now rejected hypothesis of a common Caucasian language family (e.g. the Ibero-Caucasian hypothesis by Yavakhishvili 1937, cf. Tuite 2008), a Caucasian sprachbund had been suggested by Dirr (1928) and Klimov (1994) but was rejected by Tuite (1999) and Comrie (2008), and recently reinforced by Chirikba (2008a). Given the long-standing contact, which in a number of cases lasted more than two millennia, such a sprachbund would imply a range of commonalities, similarities, and overlaps. So one might end up arguing for contact in a circular fashion based on (structural) similarities to be found in today’s language setup, whereas linguistic similarities are explained by possible contact, without considering shared innovations or even homologies, i.e. independent changes with an equal outcome. Therefore a clear picture of the genealogy of the Caucasian families is needed to circumvent arguments about contact-based similarities instead of genealogical traits. Contact-induced changes occur more or less “naturally” due to mechanisms of reanalysis, reinterpretation, grammaticalization, and accommodation (Aikhenvald and Dixon 2001) and were found already in the beginning of language research in the Caucasus in the nineteenth century. Klimov (1994) refers here, for example, to Baudouin de Courtenay (1963), who had found “common traits irrespective of original relatedness”. Despite the fruitful typological perspective, one that is more abstract and comparatively oriented, Klimov (1994) also predicts that if provided with an areal-linguistic perspective, one might gain a better understanding of the convergence processes under circumstances of language
contact. The idea of also integrating dialectology (cf. Bisang 2004) needs to be taken into account since it will allow one to attain a more fine-grained mesh of contrasts on the one hand and to gain a better understanding of the particular changes that may spread through the entire language continua or be retained and lead to higher inner diversity on the other hand. Comrie (2008) locates the languages of the Caucasus at a middle position on a diversity scale, compared, for example, with Papua New Guinea, but he proposes a zone of convergence with regard to certain grammatical structures. From their general typological profile, agglutinative, mainly prefixing, morphology and marking of the syntactic agents, as well as a three-way phonation contrast in the obstruent systems (cf. Comrie 2008) can be of value here. Klimov (1994) relegates, in terms of their lexical properties, a rich layer of “internationalisms” of the Muslim world, i.e. loanwords from Arabic, Persian, or Turkic which are shared by all groups of Caucasian languages. But even beyond that, Klimov (1994), suggesting a sprachbund, talks of a possible core vocabulary of the Inner Caucasus languages – containing terms like those for ‘cherry’, ‘chicken’, ‘sister/daughter-in-law’, or ‘plough’ – that seems to be similar enough in all three Inner Caucasian families (cf. Chirikba 2008a).

Also for reasons of structural similarities one can even argue for a wider conglomeration of languages that comprise especially the southern edges. The latter would concur with the notion of a Transcaucasian cultural and linguistic contact zone spanning along and around the Araxes river from Turkey to the Caspian sea, as in Stilo (2004), where he actually describes a much greater spreading of linguistic features, with Iran as a buffer zone. Stilo (2006) finds here, for example, a North-South transition in the use of morphological marking from prepositions via circumpositions into postpositions. For the Caucasus, Chirikba (2008a) suggested a division into core languages (comprising NWC, NEC and SC languages), peripheral languages (Ossetic, Armenian languages), marginal (Karachay-Balkar, Kumyk, Azeri, Nogay, some varieties of Anatolian Turkish, and Iranian Tat), and contiguous languages (Kurdish, Talysh, historically Persian, Pontic Greek, Trukhmen Turkic, etc.), which are spoken in adjoining areas. Nonetheless one wants to find at least continua of structural similarities if there are no total overlaps, so that feature spread would cover the entire linguistic area in question and allow for adjacent linguistic areas to touch and overlap.

1.2 Contact scenarios – conditions for processes of convergence

1.2.1 Historical spheres of influence

Over the centuries a number of non-Caucasian languages played major roles in those periods when certain empires gave rise to a particular dominance and prestige of a speech community and their language. Similarly connected is a chronology of foreign invasions into the area that was already inhabited by the different Caucasian peoples, of which only the more “recent” part can be sketched here (following mostly Wixman 1980 and Coene 2009, see there for more extensive historical portraits).

Iranian nomads (Scythians, Sarmatians), ancestors of the Ossetians, settled in the fifth century in the foothills of the central North Caucasus. Other Iranian languages, like Persian, became more important due to the relations with the Kartvelian (Georgian) and Armenian empires. A Jewish Tat-(Iranian) speaking military colony south of Derbent was founded in the sixth century. The invasions of Hunnic tribes in the sixth and those of the
Khazars in the seventh to the ninth century could exemplify the very early and long standing influence of Turkic languages on the northern bounds of the Caucasus. The Khazar Empire was established in the eighth century in the northeastern steppes, and their people are regarded as ancestors of the Kumyks. The invasion from the east in the ninth to tenth century by the Pechenegs, likewise Turkic, was followed by an invasion of the Kypchaks (Kypchaks, Kumans, Polovtsy) tribes in the eleventh to thirteenth century. Remnant groups of these remain in the North Caucasus (especially Karachay). After that, Oghuz Turks, the ancestors of today’s Azerbaijanis, occupied the coastal plain of Dagestan as far as the city of Derbent. Hereafter, the Turkic Nogai established themselves in the area of the Lower Volga in the fifteenth century and were pushed southwards into northern Dagestan by the Mongolic Kalmyks and the Turkic Turkmen (Trukhmen) in the seventeenth century. Some Nogai (Lesser Nogai) occupied the Kuban region in the Northwest Caucasus and were pushed further into the Caucasus by Crimean Tatars. Finally, the Ottoman Turks began to settle along the eastern Black Sea coast in the seventeenth to eighteenth century (cf. Johanson 2006b). During the flourishing of the kingdoms (seventh to ninth century) of Orthodox Christian Georgia and Armenia, these languages became important as those of local powers. The Armenians seem to have been in the area for at least 2,500 years. Around the end of the fifth century, Armenia was incorporated into the (Persian) Sassanid Empire. Schulze (2005) and Nichols (2013) also mention a ruler situation, particularly in the sixth or seventh century, when Christian Caucasian Albania came into power with what is today called Caucasian Albanian (Alwan/Old Udi) as its dominant language; this presumably motivated many speakers to shift their language or to become bilingual.

In the ninth century, the incoming troops of the Umbayyads and Abassids introduced Islam and, at this point, brought some Arabic from the South into the area of Dagestan. Friedman (2009) assumes that, finally, after the spread of Islam in the seventeenth century, Arabic would serve as a written lingua franca. This adds another aspect to language influence, namely writing, which at least for North Caucasian had been to a small degree in Arabic script since it was part of religious life and also found use in other contexts (Dobrushina 2013). The Southern Caucasus with Georgian and Armenian writing traditions are comparably ancient but not necessarily as pervasive due to the religious borders. The first Russian settlements were presumably those of the Cossaks (Terek and Grebenskoy armies) in the steppes of the east-central North. A later settlement was started by the Ukrainian Kuban Cossaks in the western Caucasus. With the expansion of the Russian empire in the eighteenth and nineteenth centuries, those Russian populations also grew larger and certainly made Russian a useful language for trade and commerce. Wixman (1980) points out that since the 1930s Soviet Union language politics was mainly oriented towards the complete dominance of Russian. Before that it was official policy to support the smaller languages and to develop a standard of literacy with Russian as a role model. After a very short flourishing of alphabets in Latin scripts, local conventions of writing in Cyrillic were established in the 1930s and 1940s.

This means that there was presumably no single lingua franca for the entire Caucasus area at any time except in very recent history, the second half of the twentieth century particularly, when Russian became a major language of the area, serving not only as the language of administration, education, and research but also slowly replacing the traditional regional lingua francas. Avar in the Northwest of Dagestan, and Lak and Lezgi served this purpose in the central part. Among the Northwest Caucasus languages, the West Circassian Adyghe and some Turkic languages (Karachay, Nogay, Ottoman) served
as a lingua franca, whereas for Kartvelian, Georgian was used in this respect (cf. Chirikba 2008: 31 for a more extensive list).

1.2.2 Population genetics, migrations, and language contacts

In general, human-genetic research corroborates the view of relative immobility in the center of the Caucasus and suggests low genetic diversity with drift in the male population of isolated populations, and otherwise recently acquired similarity with Turkic and Iranian populations of West-Asia (Nasidze et al. 2004). Yunusbayev et al. (2012) characterize the “Caucasus as an asymmetric semipermeable barrier to ancient human migrations”, and this is the overall picture that emerges from studies of Y-chromosomal DNA. They state “signals of regional Y chromosome founder effects distinguish the eastern from western North Caucasians. Genetic discontinuity between the North Caucasus and the East European Plain contrasts with continuity through Anatolia and the Balkans, suggesting major routes of ancient gene flows and admixture...” (Yunusbayev et al. 2012). The authors conclude that “irrespective of the Early Upper Paleolithic presence of anatomically modern humans both south and north of the Caucasus... the combined autosomal and gender-specific genetic variation of the Caucasian populations testifies to their predominantly southern, Near/Middle Eastern descent. Y-chromosomal variants under strong founder events, seen in particular among populations inhabiting the northern flank of the High Caucasus Mountain Range, appear to never have expanded to the East European Plain, while the nomadic people of the latter, once settled down predominantly on the northern slopes of the Caucasus, have likely preserved, to a different extent, some of their earlier genetic heritage.” With regard to the population-specific profiles, Balanovsky et al. (2011) found that each “linguistic group” “…ended up with one major haplogroup from the original Caucasus genetic package, whereas other haplogroups became rare or absent in it. The small isolated population of the Kubachi (Dargwa), in which haplogroup J1*-M267(xP58) became virtually fixed..., exemplifies the influence of genetic drift there.” Genetic drift is the frequency change of an allele due to random variation vis-à-vis natural selection. “During population differentiation, haplotype clusters within haplogroups emerged and expanded, often becoming population specific. The older clusters became characteristic of groups of populations. Many younger clusters were specific to individual populations (typically speaking different languages).” (Balanovsky et al. 2011) All in all, population genetic research corroborates the general characterization of Caucasian societies as predominantly endogamic (see e.g. Tuite 1998; Coene, 2009: 64), which allows only slow drift in L1 and only small influence by direct contact through exogamy (L2-speakers). This is then reversed in those “moments” of invasion and/or larger language shift.

The situation prior to the Russo-Circassian war (1830-1864) can, especially for the smaller communities with no writing tradition, only be reconstructed by means of reports from outsiders like Jacob Reineggs (1744–1793) or Johann Anton Güldenstädt (1745–1781), in addition to the long spanning Georgian, Armenian, and Persian sources and other deeper historical research by means of archaeology and the like. On the other hand, there are a number of recent historical events precisely known, such as invasions and conquests of the Russian or Ottoman Empire, for example, or other detrimental shifts which eventually led to deportations and other forced migrations of populations like those of the Ubykhs in 1864 and other speaker populations like the Circassians and the
Abkhazians in the 1860s or like those of the Armenians in Anatolia in 1915/16 and earlier. The majority of these languages have been maintained until today but have undergone structural changes due to the inevitably strong influence of Turkish (see Höhlig 1997) or other languages of contact. When analyzing the historical geographical spread of particular features, it will often be necessary to leave out most of the current diaspora varieties. In cases like Ubykh, extinct since 1992, where the entire population shifted to Turkey, one will need to “replace” the language in its native territory at the Black Sea around Sochi in order to account for areal effects.

In a number of studies, individual contacts between groups of languages in the Caucasus were focused upon, where the Dagestanian languages have often been taken as a whole. This makes sense because here particular words or structures may have been “copied” several times (cf. Johanson 2006b). Khalidova (2006) studied Avar-Andi contacts with mainly Avar as the donor to the smaller languages. Of those investigating non-Caucasian-Caucasian contacts, the study of Zabitov and Efendiev (2001) on Arab and Persian lexemes in Lezgian or Dzhidalaev (2010) on particular traits of Azerbaijani in Lak can be taken as examples of studies focusing on lexical borrowings and loanword adaptation strategies (e.g. Khalilov 2004). The perspective of Caucasian languages moving out of the area (e.g. Höhlig 1997) can also be turned around: Johanson (2006b) points out that the speakers of West Armenian were, for approximately seven centuries, in contact with Turkish, so that those structures could be copied. To what degree, for example, Kurds, Greeks, and other smaller ethnicities, like the Assyrians, were also involved in this interwoven multilingual-net of the past centuries reaching from Anatolia and Kurdistan north to the Caucasus seems to be fairly hard to estimate accurately.

1.2.3 “Isolated” communities, multilingualism, and the “vertical archipelago” phenomenon

Obviously there are a number of cases that allow one to study at least relatively recent contact influence as it would take effect in situations where a small community is surrounded by speakers of a distantly related language or languages. These situations may have arisen due to migration of the community itself or due to remnant settlements (enclaves) of a former larger territory. Of this kind we find, for example, the Zakatly Avar dialect in Azerbaijan, the (long documented) Kisti Chechen dialect in Eastern Georgia, or the Udi dialects of the villages Nizh and Vartashen (today Oguz) in Azerbaijan (Schulze 2005; Gippert 2008). The variety of Udi spoken in Zinobiani in Georgia should be close to the one of Vartashen since this settlement was established and populated by Vartashen inhabitants in 1988/89 during the secession of the Mountain Karabakh region as part of the Armenian-Azerbaijani conflict (cf. Schulze 2005). The frequency with which these situations led to multilingual settings can be illustrated by further examples given in Chirikba (2008a: 31). According to the author, the Megeb Dargwas, who live in the predominantly Avar-inhabited Gunib district, speak Dargwa, and also Avar, Lak, and Russian. Many Abazas:

...in the Karachay-Cherkes Republic, speak, beside Abaza and Russian, also the distantly related Kabardian, and some also Turkic Karachay, the languages of their more numerous neighbours. The Ubykhs, who were numerically smaller than neighbouring Circassians and Abkhazians, were either bilingual with Circassian or Abkhaz as a second language, or tri-
lingual Ubykh-Circassian-Abkhaz. In the Muslim Georgian area of Adzharia in the southern Caucasus adjacent to Turkey, Turkish has traditionally been the second language... besides, nearly all Adzharians speak Russian... The very small Laz and Abkhaz communities in Adzharia, are to some degree quadrilingual since they speak... also Georgian, Russian, and Turkish... Avar was often used in the past by various Dagestani communities as a lingua franca, and even a special form of Avar called bolmac ('public language') had developed... To some extent Turkic Kumyk was also used in some North Caucasian communities as a kind of lingua franca, a role now overwhelmingly taken over by Russian. (Chirikba 2008a: 31)

In terms of descriptions of exile varieties we also find examples in Vaux (1998) of Armenian in Yerevan, Tiblisi, or Julfa and in Chirikba (2008b) of Armenian spoken in Abkhazia, illustrating the influence of various linguistic contexts. Finally, as mentioned above, there are a number of communities with a great number of speakers of Caucasian languages scattered all over Central Turkey (cf. Koryakov 2006), some of them already having lived there for more than one and a half centuries. Here we find limited descriptive data, mostly only for languages of Northwest Caucasian, like Turkish Kabardian (Gordon and Applebaum 2006) or Turkish Adyghe (Höhlig 1997) in Turkey. Particularly the two Circassian languages, whose speakers are usually found under the name Cherkess, are spread all over the Middle East, including Jordan, Syria, and Israel but also Eastern Europe (Kosovo) and the USA.

Social factors, such as religious, ethnic, historical, and political groupings also limit or facilitate potential intermarriages or closer interaction. The major religious groups in the area today are Sunni Muslims and the Georgian and Armenian Orthodox Christian groups. Historically, the situation was presumably more heterogeneous, since local traditional beliefs, habits, and folklores point to Zoroastrian, Mithraic, and other substrates (Coene 2009). It is also assumed that Judaism spread as result of the Khazar settlement from the seventh or eighth century into northeastern Dagestan but disappeared by the thirteenth century (cf. Wixman 1980: 68). Smaller groups built their own enclaves, such as the Mountain Jews in Georgia and Azerbaijan, who speak Juhuri, a variety of Tat (cf. Authier 2012).

The physical geography of the main two mountain crests in terms of inhabitable valleys, migration-blocking mountain ridges, glaciers, water sheds, etc. certainly plays a major role in explorative migration, usage of pastures, settlement sizes, and trading relations. In this way geography will have a regulating influence on the degrees of freedom of possible contact and interactions. With respect to the topographical situations in a populated high mountain area, Nichols (2013) used the term “vertical archipelago”. In particular Dagestan is characterized by a high degree of multilingualism especially among the male speakers and among those of smaller languages. Here – as a vertical component – the highland villagers usually know the languages that the lowlanders speak but not vice versa. The male population would spend a major part of their working life in the lowlands trading and working part-time, especially during the summer. A more recent sociolinguistic study of Dobrushina (2013) can provide support for Nichols’ idea. Based on fieldwork surveys aiming at the acquired multilingualism of informants and similar to that of their parents, grandparents, and great-grandparents, Dobrushina (2013) was able to sketch a picture of multilingualism that spans several generations back into the nineteenth
century. Furthermore, Dobrushina describes a network of hosts (Archi: xe\text{e}le), a social network which can at least be ascribed to the Southeast outer boundaries, i.e. today’s Southern Dagestan and Azerbaijan, but which can certainly be verified for and extended to other areas of the Caucasus. These host relations comprise people who travel from the highlands down to market places and seasonal working places, trading their local goods. Such social networks would also include a form of exchange where the labor of young adult men is in part reciprocated by the transfer of a more widespread, useful, and prestigious language of the lowlands. Dobrushina undertook the survey in three neighboring villages with a different L1 (Archi, Lak, and Avar) language each, differing in the size of the total speaker population and in prestige. L1-speakers of languages of higher prestige showed lower degrees of multilingualism and vice versa. Nichols (2013) and Gippert (2008) draw attention to the fact that in recent decades vertical relations have been disrupted, multilingualism is fading out, and parental language transmission of these small languages is on the decline, which all leads rapidly to endangerment. The predominant language shifts of these younger generations follow the borders of the major national languages (Russian, Georgian, Armenian, Azerbaijani, Turkish).

For a wider picture of the historical sociolinguistic situation, uncertainties remain concerning the number of speakers and semi-speakers, i.e. L2/L3-speakers, which is to some degree more important in order to account for overlaps and language-particular contact phenomena. The use of lingua francas and interethnic contacts, within religious or ethnic boundaries, relates to a mechanism that was proposed as one of the most potent sources of contact-induced language change, as observable in unsupervised L2/L3/…acquisition, and which leads in most cases to incomplete acquisition (Lupyan and Dale, 2010). Subsequently, larger groups of L2/L3 speakers are more easily facilitating the mechanism of misinterpretation, reanalysis, overgeneralization, and fossilization, i.e. the establishment of structural errors. Linguistic concepts (sound categories; morphological categories; semantic categories) are more or less consciously projected onto the L2 in analogy to the L1. As a consequence, languages used as lingua franca and other more widely used “bigger” languages tend to be simplified. Accommodated speech plays a significant role here; acquisition of this type of speech results in a number of structural parallelisms (including phonetic similarity and functional calques) as well as semantic transparency.

2 Phonological and phonetic areal typology of the Caucasus

2.1 Phonological and phonetic areal typology

A phonological areal typology aims at a distribution of attributes and characteristic features (see Hyman 2009 for an excellent discussion). A phonetically oriented areal typology considers the diversity and variability that occurs with the phonetic implementation of a phonemic contrast. If we take the voicing contrast of stops for example, we find in languages like Spanish or Russian that a \textipa{/b/ in word-initial position is pronounced with an early voice onset, i.e. vocal fold vibration starts before the release of the bilabial closure. The voiceless counterpart \textipa{/p/ would have a voice onset that starts slightly after the lip opening. In contrast other languages like German would have a “voiced” bilabial stop \textipa{/b/ starting about the same time as the \textipa{/p/ in Spanish. The voice onset time (Lisker and Abramson 1964), i.e. the distance in time of closure release and
voice onset, serves here as a function of the abstract (phonemic) 2-way contrast. From this perspective, it makes the most sense to adopt an understanding of phonetics and phonology as inseparable entities as, for example, argued for by Ohala (1990) and Blevins (2006). Such a view allows one to talk about the phonetic implementation of abstract contrasts in different phonetic domains (segmental, intersegmental, suprasegmental) as well as about diachronic processes and universals at the same time. From this viewpoint, most sound change relevant for phonemic distinctions would be driven by the listener’s misinterpretation of articulatory variations on the part of the speaker. These variations can be caused by a number of (external) mostly speaker-related factors (cf. Hickey 2012b). Such could be the influence of the actual L1 or L2, L3, and LX in terms of competing systems, which all interfere and compete (cf. e.g. Flege et al. 2003; Blevins and Wedel 2009). With respect to articulatory and auditory capabilities, physical factors such as denture, vocal tract morphology, and hearing (loss), etc. also need to be accommodated by a speaker/listener and create peculiar means of compensation. These processes add to redundancy in phonetic substance, i.e. the way the contrast is expressed in the language and finally the way it is realized by the speaker, especially in fast speech. Regular speaker variations can also reflect social group norms and thus social barriers (Hickey 2012a). The tendencies towards regularity and systematicity driven by analogy and accommodation within a population of subgroups would then still lead to an areally homogenous pattern, but presumably only when these subgroups are in contact. Trudgill (1996: 13) states that “dense multiplex networks typical of relatively closed, stable, non-fluid communities are more likely to lead to conformity in linguistic behaviour and to the maintenance of group norms as well as the successful carrying through of ongoing linguistic change.”

However, we are looking here at the level of forms, in particular, sounds. What would the phonological and phonetic level tell us vis à vis the other linguistic levels? A number of authors point out (Trudgill 2011: 2; Hickey 2012a; McMahon 1994) that with regard to time depth, phonetic and phonological changes happen rapidly. Therefore we would also need to consider other mechanisms like accommodation, where, especially in mutual bilingual situations, a convergence of L2 varieties, i.e. the varieties spoken by bilinguals, has been observed (see e.g. Flege and Port 1981 for Arabic to English). Most of the changes are fairly subtle for a particular point in time but accumulative given that, on the one hand, a number of speakers are experiencing similar influences within the community and, on the other hand, certainly repetitive input occurs. Eventually such phonetic variation can result in phonological changes which involve one or two speaker generations. Nonetheless, a basic assumption for sound change (Ohala 1990; Blevins 2006) would include the misinterpretation of a sound by a listener as one of the key mechanisms. Social hierarchies, concepts of prestige, traditions of loanword adaptation, and the like can reinforce this perception-driven process. Lexical borrowing may enforce the borrowing of segments and other characteristics such as alternative patterns of word prosody. The notion of frequency or re-occurrence of words or functional structures seems to predict the subsequent changes in phonetic forms (Bybee 2003) so that frequency also comes out as the major factor for markedness (Haspelmath 2006). Thomason and Kaufman (1988) already noted that marked features are harder to acquire in a contact situation with late acquisition patterns. Other more production-driven changes, which involve continuous biases, for example, of biomechanical couplings of articulators and thus preferences of temporal coordination, lead only to slow changes within the population of speakers (see e.g. Lancia and Grawunder 2013). Given the
longstanding presence of the three Caucasian families and the persistent contact of non-
Caucasian in the area, stabilization and enhancing of local features can be assumed. In
fact, similar to other major languages within a large multilingual environment such as, for
example, English in India or French on the African West Coast, Daniel et al. (2011)
demonstrate that Russian has also undergone some changes in Dagestan.

2.2 Phonological features – structure and occurrence

A “feature” in this context is used as a more familiar term as it is often met in the
literature (Russ. признак). With respect to the recent discussion of the nature of features
by Ladd (2014), the term “attributive features” is used here in order to account for this
wider view. The number of occurrences in a given phonological inventory, which reflects
how often a particular attributive feature is applied, needs to be combined with the actual
degrees of freedom that are made use of in the inventories of contrasts and that are thus
much more revealing. Amongst the most challenging issues when accounting for the
occurrence of features are the individual approaches of different authors towards the
phonemic status of particular contrasts (Maddieson 1984: 4) – a general problem with
typological approaches but also of other areal linguistic research (see e.g. Maddieson
1984; Gilles and Siebenhaar 2010). For the languages of the Caucasus, this concerns in
particular the status of contrasts involving secondary articulation such as labialization,
gemination, palatalization, pharyngealization consonants as well the status of nasalized,
long, pharyngealized vowels, and in general the status of vowel combinations.

2.2.1 Inventories and sub-inventory sizes

Caucasian languages are well-known for their large phoneme inventories with 50 or more
phonemes, especially consonants. This is true in the case of the NWC languages and to
some degree also for some of the northern NEC languages. Indeed, before the description
of many KhoiSan languages, Ubykh, with over 80, was considered the language with the
highest number of contrasts, i.e. the biggest phoneme inventory in the world’s languages
(cf. Catford 1983). Among the SC languages, Georgian, the biggest Caucasian language
in terms of speakers, uses a relatively low number of consonant and vowel contrasts (see
Figure 1).
However, Mingrelian and Laz also have more medium inventories, as well as Svan, with the only exception being its northern dialects. The vowel inventories of neighboring Nakh and Dagestani languages appear to be comparably large and larger. The vowel systems of Bezhta (Kibrik and Kodzasov 1990), Lower Chechen (Comrie 1981), and Upper Bal, a dialect of Svan (Gippert 2008; see Figure 2) serve as examples of languages with long and umlauted vowels. The NWC family forms is the opposite with very low vocalism but high consonantism. Already Trubetzkoy (1958) suggested the number of
vowel contrasts in relation to the number of consonant contrasts as a means of assessing the linguistic profile of a language as a basis for comparison. Maddieson (1984) links this consonant-to-vowel ratio to a general typological perspective. For the Caucasus however (see Figure 1), a vowel-consonant ratio of a sample of 47 doculects, one would expect a stable figure, if there is something like a homogenous setup, but the ratio seems to correlate with consonant inventory size ($r^2=0.57; r=0.76, t=7.88, df=45, p<0.00001$). Furthermore, there appears to be a growing complexity of smaller languages that are closer to the main Caucasus crest, where we can describe a “stacking” of contrasts in terms of feature combination. This generally involves consonants in the NWC languages, where we see labialization, palatalization, and pharyngealization (see section 2.2.3), and appending of a retroflex ($\tilde{t}$, $\tilde{d}$, $\tilde{z}$, $\tilde{ʒ}$) and alveolo-palatal ($\breve{t}$, $\breve{d}$, $\breve{z}$, $\breve{ʃ}$) fricative series and also, in some cases, the development of a (unaspirated voiceless) “tensed” stop series. NWC and some of the adjacent Andic languages of the NEC also include fricatives in the glottalic contrast. For the NEC languages, the contrasts extend on to the vowels, where nasalization, pharyngealization, length, and diphthongization (vowel + vowel or vowel plus glide combinations) are employed. In addition, a geminated series of stops and fricatives come into play for the northwestern part of Nakh-Dagestanian, and a “tensed” stop series in the central and southeastern part – similar to NWC.

Trudgill (2004) reintroduced a discussion with regard to a relationship between inventory size and speaker population (community) size, given the various expected effects of social-network density, contact intensity, adult L2 speaker proportion, etc. This question was also addressed in relation to the Caucasus by Nichols (2013), who applied her own measure of inventory size, including additional morpho-syntactic features as well. Similar to her results, for a sample of 74 doculects, only a weak relationship can be presumed here at all for log-transformed population size (numbers taken from ASJP1, Brown et al. 2013) and consonant inventory size ($r=-0.31; r^2=0.098; t=-2.81, df=72, p=0.006$). Thus population size could only explain 10% of the trend in the consonant inventory size. Arguably, the average speech community size and its correspondence to social network size and density and again the L2 user proportion would, in contrast, need to be taken into account, although these are harder to assess.

**Vowel inventories** On the other hand, we can take the vowel system of Bezhta (see Figure 2), another NEC language from the Avar-Andi-Tsezic branch that is also described as having unlauteled vowels, length contrast, and nasalized vowels. If we add these contrasts into the vowel scheme, it appears even bigger than the Chechen system of monophthongs (even when adding the six diphthongs). This is similar to the closely related language Hunzib, which Isakov and Khalilov (2012) describe as having at least 24 vowel phonemes, based on 8 basic vowel qualities (/i, e, ə, a, u, o, ɒ/) (cf. Van den Berg, 2005) with the addition of long (/iː, ēː, ɨː, aː, uː, oː, ɒː/) and nasalized vowels (/ĩ, ě, ɨ̃, ə̃, ã, ū, ŵ, ɒ̃/). For Bezhta the combined nasal and long vowels also come into play, adding another series of contrasts. However, most descriptions comprise only

---

1 This is the Automated Similarity Judgment Program by Søren Wichmann and his colleagues, see, http://asjp.cilid.org.
the “basic” vowel system and add those with contrasts of secondary articulation (nasalization, pharyngealization) or length separately. Umlauted vowels would be considered basic too, whereas in their formant structure, these only seem to group for Lezgi and Khinalug with front rounded vowels (cf. Catford 1994). Other samples for Lak (Catford 1994) and Bezhta (Grawunder 2015) seem to suggest that umlauted vowels and front rounded vowels as in Azerbaijani cannot easily be equated since they show different formant characteristics.

Often not all possible vowel contrasts are analyzed in the surveyed descriptions, which is especially true for contrastive vowel length. Although there is still discussion about the vowel inventories of some NWC languages, these vowel systems can, with Catford (1992) and others before, be characterized as vertical since these systems seem to mainly contrast in vowel height. Most descriptions (cf. Hewitt 2005; Colarusso 1988) suggest two vowel qualities /ə, a/ for Ubykh, Abaza, and Abkhaz as well as two vowels /ə, a/ with an additional length contrast /aː/ (Höhlig 1997) sometimes for the varieties of Adyghe (West Circassian) and a three-vowel system (/ə, e, a/ or /ɪ, ə, ɐ/) for Kabardian (East Circassian) (cf. Catford 1994; Gordon and Applebaum 2006). From a phonetic standpoint, however, one can clearly observe cardinal vowel qualities (like [i, a, u]), and the contextual predictability of such vowels may not be that transparent for a second language learner.

Klimov (1994) and Catford (1994) also state that Caucasian vowel systems employ no diphthongs, with Chechen and Ingush (Nichols in Smeets 1994: 1) as the only exceptions, where the sequences that correspond to the long vowels are analyzed as having a long initial vowel portion instead of a glide, like those of the short vowels (/e > je/ vs. /e: > ie/; /ü > üi/ vs. /ü: > üi/; /ɔ > wo/ vs. /ɔ: > wo/). Here one would need to argue against all similar monophonemic analyses in favor of a biphonemic contrast of glide plus vowel or vowel plus glide (/N+w/; /N+j/; /N+u/), e.g. in Svan (Gippert 2008) or Tsez (Job 2004) or recently Tabasaran (Vadzhibov 2012). Kodzasov (1999) argues that the Andic language Godoberi is the only one with diphthongs.

2.2.2 Commonalities and overlapping of attributive features in the Caucasus

For the following discussion of the Caucasus, those attributive features are taken that span across the Caucasian families in the core area (Figure 2 and Figure 3).

**Number of initiation (phonation) contrasts** As already pointed out above, all languages of the three (Inner) Caucasian language families employ at least a three-way contrast in the obstruent system. Therefore, we find as default contrasts in all three Caucasian families: ejective (glottalic) vs. aspirated/non-aspirated (pulmonic) vs. voiced (pulmonic) obstruents. These ternary contrasts are additionally found in non-Caucasian peripheral Ossetic (Iron and Digor) and varieties of Eastern Armenian. A ternary contrast with voiced, voiceless unaspirated, and voiceless aspirated stops is also found in the Transcaucasian periphery with Kurmanji Kurdish, Zazaki, and some Assyrian (Neo-Aramaic) dialects (e.g. Burwar; Khan 2008: 29). Binary (two-way) systems with (voiced vs. voiceless) pulmonic obstruents are otherwise found in Altaic languages (Turkic: Nogai, Kumyk; Mongolic: Kalmyk) and Indo-European (Russian) (see Figure 2).

Especially for languages of the Lezgian group (Lezgi proper, Tabasaran) and for Lak
but apparently also for Chechen (Nichols in Smeets 1994), we find analyses of so-called “tensed” stops, unaspirated pulmonic stops with a long closure. However, in Lezgi (Hapnelmath 1993), Lak (Anderson 1997), and varieties of Dargwa (e.g. Isari, Sumbatova and Mutalov 2003), these (voiceless unaspirated) “tensed” stops do also occur word-initially and build, for a number of authors, a fourth series of non-aspirated voiceless stops. Among the NWC languages, this is the case for the Adyghe dialects of Shapsug (Shapsegh), Arnavir (Temirgoy), and Bzhedukh (Colarusso 1988). In these languages, aspiration would then need to be counted as a contrastive feature, whereas in other ternary (3-way) contrasts, including glottalic obstruents, aspiration is just an additional (redundant) attribute. Hence, for a number of NEC languages, like Chechen, Tsova-Tush, Bezhta, or Hinukh, these long segments seem only to occur in word-medial position. But since these involve pulmonic and glottalic obstruents, including affricates and fricatives but also nasals and liquids, an analysis as geminates is certainly justified. The involvement of glottalic stops in this gemination seems to occur mainly among the Avar-Andi-Tsezic languages.

**Ejectives (glottalic initiation)** Tuite (1999: 217) pointed out that a quest for common traits in all three Caucasian families would result presumably in only one shared feature, the ejectives. Other authors (Catford 1977; Chirikba 2008a; Klimov 1994) saw more overlap of other non-phonological features like complex agglutinative morphology, the marking of the agent (ergative and ergative-like constructions), or verbal conjoining and relativization.

However since the occurrence of glottalic stops (ejectives) has also been described for non-Caucasian languages of the Caucasus area like Ossetic (Testen 1997), Eastern Armenian (Vaux 1998), and Kurmanji Kurdish (Jastrow 1977) and, according to Chirikba (2008a: 45), even for varieties of Turkish and for Neo-Aramaic in Urmia, Van, and Mosul (Iraq), and since this feature seems to be phonetically well definable (Grawunder et al. 2010; Grawunder 2013), it is taken here as one of the core features of the Caucasus area.

The occurrence of the glottalic initiation feature, mainly in the form of ejective stops and affricates, ends outside the geographical area in the realm of peripheral languages and can therefore also serve best for delimiting the linguistic area. Contrastive ejective fricatives are only found in NWC languages. However, phonetically there seems to often be more of a “fade-out” in the periphery in terms of a faint phonetic realization of ejective stops, as is observed for varieties of Udi (Nidzh variety, Schulte 1994), some coastal varieties of Georgian (Authier p.c.), and Laz (Lacroix 2009). Simultaneously, there are also regular sound changes in the form of deglottalization taking place (cf. Fallon 2001: 217).
Figure 2. Types of phonation contrasts in obstruent series (N=85); the quartiles of the circles correspond to the number of types of contrast (2-, 3-, 4-way contrasts shown as black quartiles of circles)

On the other hand, even such varieties of non-related Indo-European languages as Ossetic Iron and Digor (IE, Iranian: Testen 1997), Eastern Armenian (IE, Armenian: Vaux 1998), and Kurmanji Kurdish (IE, Iranian: Jastrow 1977) were described as employing ejectives or at least having ejective-like characteristics, the presumed results of contact. These slight categorical shifts lead only to small deviancies within a category but not to a new
contrast, i.e. a new phonemic category. Most likely we see here the influence of L2
speakers (re)introducing another contrast (glottalic – nonglottalic). Such a process would
leave the aspirated stop as “marked” because it is phonetically marked. Aspiration of
stops was shown to be one of the most easily “transferable” features in speaker contact
(cf. e.g. Babel 2009). Chirikba (2008: 43) states that for the non-Caucasian languages,
glottalization in the Caucasus is attested in Ossetic, the East Armenian dialects (e.g.
Tiflis, Artvin), in dialects of Kumyk (e.g. Kaytag), northern Azeri (e.g. Zakatala-Kakh),
and Karachay-Balkar (e.g. Malkar). Ossetic has glottalization not only in words borrowed
from the Caucasian languages but also in the native IE vocabulary (e.g. st’aly ‘star’). The
same is claimed for the Malkar dialect of Karachay-Balkar (e.g. k’ordum ‘I saw’). In the
Kumyk dialects, glottalization is explained by the NEC substrate. Chirikba (2008a) also
mentions that the population of several Kumyk villages, whose ancestors are known to be
speakers of Avar originally, represents a separate Kumyk dialect, displaying such non-
Turkic traits as the presence of glottalized consonants and the violation of the vowel
harmony rules, etc.

Places of articulation Generally we find five places of articulation: bilabial, alveolar,
velar, uvular, and laryngeal. Additionally, pharyngeals and laterals would come into play
as two other places that are employed across families, but these are only employed in
NWC and NEC languages. The laterals comprise a series of obstruents that occur in the
order of lateral fricative /ɬ/ > lateral affricate /tɬ/ > lateral affricate ejective /tɬ'/ in
addition to the lateral approximant /l/. To be precise, this distinction actually concerns the
direction of airstream in the oral cavity (lateral vs. central). The actual place of
articulation varies, both contextually and speaker specifically, especially with regard to
tongue contact in the stop part of the affricates between (coronal) alveolar [tɬ] or
postalveolar [tɬʰ] and (dorsal) palatal [cʎ̥] or even velar [kʟ̥] position. However, the
coronal realization [tɬ] is the one most commonly observed.

The SC and NEC languages typically show no labio-dentals (/f, v/) but do show labio-
vulars [w] in native vocabulary, with a few exceptions for /f/ described for Laz (Holisky
in Harris 1994), Rutul, and Budukh (Alekseev in Smeets 1994). On the other hand, all
NWC languages seem to additionally employ voiced and voiceless labio-dentals.
Kabardian, Abaza, and Abkhaz also have a glottalic fricative [f’] (Hewitt 2005). The lack
of labio-dentals is also seen in Turkic but not in Armenian and the Iranian languages.

A phonemic glottal stop occurs in most of the SC and NEC languages. For the
inventory of Lak by Alekseev (1999) a glottal stop is not analyzed as phonemic, although
other authors like Kibrik and Kodzasov (1990) or Anderson (1997) included it. For the
NWC languages, most authors agree with Colarusso (1988) that the glottal stop is sub-
phonemic in Abaza and not phonemic in Abkhaz and Ubykh either.
Figure 3. Presence of common contrastive features: uvular place of articulation [uvu] – labialization [lab] – glottalic initiation (ejectives) [ejec] – pharyngealization [phar] (number of features shown as black quartiles of circles).

_Uvulars_ There is an almost ubiquitous occurrence of uvular stops (affricates) and fricatives in the languages of the Caucasus, which was considered as another commonality by Klimov (1994) and Chirikba (2008a). Catford (1977: 288) states: “[i]f there is only one type of dorsal fricative it is always uvular, not velar”. This seems to hold for Ossetic (/q, qʷ/; Testen 1997), Kurmanji Kurdish (/q/; Jastrow 1977), and
Armenian (/ʁ/, /ʁ/; Vaux 1998). The use of the uvular place of articulation follows the following implicational scale: uvular fricative > uvular stop > labialized uvular fricative > labialized uvular stop. For the Avar-Andi-Tsezic branch but also other NEC languages, the uvular stops are often realized as affricates, e.g. in Chamalal (Magomedova in Job 2004), Tsez (Alekseev and Radjabov in Job 2004), Hinukh (Isakov and Khalilov in Job 2004), Tsakhur (Talibov in Job 2004). In the Turkish language Nogai, the uvular stop maintains at least a subphonemic status, whereas for Karachay-Balkar and Kumyk (Tolstoj 1997), the uvular stops are described as phonemic, which is not uncommon for Turkic in general. Furthermore, if we look at Kumyk near the Caspian Sea in the Northeast, we observe that uvular articulation can also be easily perceived as slightly glottalized (ejective), especially in contexts where a final back vowel+/q/ is adjacent to a velar or uvular stop onset. However, the occurrence of uvulars fades out in the contiguous languages on the outer periphery, where uvulars are not described as being unconditionally contrastive for the standard languages Azerbaijani/Azeri (Schönig in Johanson and Csató 1998: 249), Kalmyk (Bläsing 2003), and Turkish (Kornfilt 1997) but where they seem to occur as allophonic variants, e.g. of /k/ in the context of back vowels.

The voiced uvular stop [ɡ] is, according to Kibrik and Kodzasov (1990), found only in such NEC languages as Andi, Dübek Tabasaran, Rutul, Tsakhur, Kryz, Budukh, and Khinalug, “though it is reconstructed for both NC groups and for Proto-NC” (Chirikba 2008: 48). However, since the voiced uvular does not occur unanimously in all languages (going on the descriptions in the literature), for example, not in Andi and Rutul in Alekseev (1999), and not in all Dargwa dialects (Kibrik and Kodzasov 1990), it seems to keep its “typologically rare character” across the Caucasus area.

Pharyngeals Chirikba (2008) extends the “uvular argument” to a generalization of a “rich postvelar consonant system”, a notion that might serve as an additional “pan-Caucasian” characteristic. This postvelar attribute would involve the occurrence of uvulars, pharyngeals, and laryngeals. Pharyngeals, in terms of voiced and voiceless fricatives, can be found in most of the NWC and NEC languages but not in SC languages. In fact they are not described in Andi and Botlikh (Alekseev 1999: 228) and not in NWC Ubykh (Hewitt 2005; Colarusso 1988). Furthermore, they are also not found in Ossetic and Armenian but are found in Iranian Zazaki, Kurmanji Kurdish, and Neo-Aramaic. (see Figure 5). Although the fricatives are transcribed as “pharyngeal” (voiced) [ʕ] and voiceless [ħ], their articulation is characterized as involving epiglottal constriction (Ladefoged and Maddieson 1996; Kodzasov 1990). Kibrik and Kodzasov (1990) also analyzed two additional “epiglottal” fricatives [ʜ] and [ʢ] for Agul. However, it seems that these occur conditionally and thus count more as allophonic variants (cf. also Ladefoged and Maddieson 1996: 168).

Secondary articulation is usually referred to as modifying the primary articulation of a segment. Here one can expect the most varying points in the descriptions since a decision about a segmental or intersegmental or suprasegmental secondary articulation entails a number of attributes that again are meaningful for the number of contrasts in the segmental inventory. It turns out to be more useful to assign the attributive feature to a word or syllable instead of just a particular segment since these suprasegmental features have a larger domain, often spanning over the adjacent syllables. Nonetheless, this typological survey can only reflect the way it is laid out in the descriptions.

Labialization is seen here as an alteration of a C(C)V sequence achieved by inserting an intersegmental labial element (> C(C)WV), most typically a labio-dental [v] or labio-
velar [w] but also a labio-palatal [ʄ] and bilabial [β, ɸ]. Although such consonant alteration can be seen as cross-linguistically widespread, it appears to occur consistently throughout the inventories of the Caucasus, even including labial consonants for some languages. In the SC languages and a number of northern NEC languages (Tsez, Hunzib, Hinukh), this is analyzed as a separate segment that can follow a stop, e.g. /k’u, qw, ɡv, ɬw/, or fricative, /ʁw/ (cf. e.g. Alekseev 1999). Other languages, especially in the NWC and NEC languages to some degree, are described as using labialization extensively as a secondary feature, which would usually mean that there is a contrastive bilabial glide component at the burst phase of the stop (cf. Grawunder et al. 2010), typically transcribed with [ˀ] in IPA or [°] in traditional Caucasianist studies (cf. Klimov 1994). In the grammar of the standard literary variety of Western Circassian Adyghe (Rogava and Kerasheva 1966) we even find /pˀw/, /mw/, or /ˀw/ along with other labialized stops or fricatives of their plain, i.e. aspirated, ejective, or voiced series. In the table for consonantal phonemes for Literary East Circassian (Kabardian) by Hewitt (2005), only one series of labialized velar stops (voiced, aspirated, ejective) occurs vis-a-vis a uvular series of non-labialized and labialized stops. Similarly, in Hewitt’s presentation of the phonemes in Ubykh, the non-labialized velar stop series seems to only have allophonic status. Hence, this feature alternates on the descriptive level between a secondary and primary feature. Hewitt (2004) illustrates the resulting inventories for the two different approaches for the Andi language Bagvalal as provided by Gudava (1964) and by Kibrik (2001: 41). Whereas the articulatory and acoustic outcome may only shift gradually from a fricative to an approximant [v ~ v ~ β ~ ɬ ~ w] and finally to a partial rounding of a following vowel, the analysis again results in a different number of segments and thus a different inventory profile. The strongest structural argument for the monophonemic analysis appears to be the existence of labialized obstruents in word final position, like Bagvalal /kitˀw/ ‘cat’, /ʁandˀw/ ‘crow’, or /haʃw/ ‘voice’ (Kibrik 2001: 38). A similar example is given by Tsakbur (Hewitt 2004: 52). Apart from Caucasian languages, Armenian also employs contrasts involving an obstruent + labial element, however these are of low frequency. Also Testen (1997) decided on Iron and Digor Ossetian for a secondary articulation, i.e. monophonemic analysis, of the labialized uvular stops.
2.2.3 Areal pockets

Some of the attributive features seem to occur only in circumscribable sub-areas, sometimes crossing the genealogical lines defined by the three families (Figure 4).

*Nasalization* Properties such as nasalization or pharyngealization, which can occur...
sub-phonemically or only in a restricted set of the lexicon or in particular speech registers, may be easily overlooked, and we can therefore assume that the current list is still incomplete. Nasalization appears contrastively in a number of NEC languages, especially in the Avar-Andic-Tsezic branch, such as, Bezhta (Kibrik and Kodzasov 1990), Hunzib (van den Berg 1995), Khwarshi (Khalilova 2009), and Inkhokhvari (Kibrik and Kodzasov 1990), but also the southern Andi languages (Akhvakh, Chamalal, Tindi, Karata) (cf. Catford 1994). Nasal vowels are also described for the Gjunej dialect of Lezgi (Mejlanova 1970). Furthermore, Kibrik and Kodzasov (1990a) describe conditional nasalization occurring between “emphatics” (e.g. [hɑʔ] ‘stool, chair’; [ʊfɪrə] ‘struma’) in the Antsuh dialect of Avar. It appears sporadically in the NEC Nakh branch; Nichols (1997) describes nasalization for Chechen as conditionally occurring, e.g. at the coda of infinitives in /-an/, and similarly for Ingush (Nichols 2011). Otherwise, nasalization seems to be absent from the rest of the central and eastern NEC languages as well as from the NWC and SC languages (cf. Catford 1994).

Laterals The main area for the occurrence of lateral fricatives and affricates can be described for an area in the North Caucasus including only members of the NE-Caucasian and NW-Caucasian families. For NE-Caucasian, only languages of the Andi-Avar-Tsezic group are involved. One exception here is Archi, a small one-village language belonging to the Lezgian branch but in geographical adjacency to villages with Avar and with Lak as a first language. From ethnographic work (Dobrushina 2013) it becomes clear that Archi speakers are very likely to also speak at least the major language Avar, which has frequently occurring lateral fricatives ([ɬ]) and affricates ([tɬ, tɬː]), As Van den Berg (2005) observes, Akhvakh shows the complete set of seven laterals /tɬ, tɬ, tɬː, ɬ, ɬː, l/, whereas Andi and Avar are in between – Andi lacking /tɬː, tɬː:/ and Avar lacking the opposition for gemination (traditionally called “intensity” in the Russian literature, see 2.2.1) in the lateral affricates; the singleton voiceless affricate occurs only in some Avar dialects.

Sibilants The NWC languages in particular employ a large number of fricatives, especially sibilants. This enlargement of the fricative inventory is brought about by involvement in contrasts of secondary articulation labialization, pharyngealization, and palatalization but also by the inclusion of gemination and glottalic initiation (ejection). On the one hand, the palatalized and labialized series in NWC languages have ‘developed’ into two additional series of sibilants, a retroflex (apico-palato-alveolar), and a lamino-palato-alveolar series (see Colarusso 1988). On the other hand, we can see that the sibilants, in an area that comprises both NWC languages and the geographically close Andi sub-branch of the NEC languages, take part in the pulmonic vs. glottalic contrast. In the Andi languages, in addition, the sibilant inventory is enlarged by their geminated (intense) counterparts so that we find, for example, in Bagvalal (Kibrik et al. 2001) these segments: /s, s’, sː, z, f, f’, fː, ʒ, l/, labialized versions /sːw, sːw’, s’w, z’w, f’w, fːw, ʒw, žw, žw’, žw’/ plus most of the corresponding affricates /ts, ts’w, ts’w’, tf, tf’w, dʒ, tf’/. Some of the Lezgic languages of Nakh-Dagestanian also employ palatalized versions so that Kibrik and Testelec (1999: 16) identify /ts, ts’, sʃ, sʃ’, žʃ/ for Tsakhur.

Pharyngealization Pharyngealization is characterized by an epiglottal constriction, which is auditorily accompanied by a kind of vowel fronting (cf. Catford 1977, 1983). Pharyngealization as a contrastive feature is described for NEC languages in Tsez (Maddieson et al. 1996), Rutul (Kibrik and Kodzasov 1990), Lak (Anderson 1997), and
Phonetically, it is primarily realized via vowels, and a phonemic contrast would imply that pharyngealized vowels occur unconditionally, i.e. not adjacent to pharyngeals. Although Kibrik and Kodzasov (1990) call pharyngealization a suprasegmental feature, they (and others) assign it to a particular segment, which follows the traditional custom in the grammars. Depending on the predictability by consonantal context, the feature is described as secondary articulation, attached to either a vowel or the particular consonant it occurs with.

Among the NWC languages, pharyngealization is attested phonemically in Ubykh and Bzbedukh Adgyhe (Hewitt 2005; Colarusso 1988). Remarkable here is that there are, unlike the other NWC languages, no attested pharyngeals in Ubykh. As a non-contrastive feature, pharyngealization can be found, for example, in Bezhta (Khalilov p.c.), Hinukh (Forker 2013), and other NEC languages such as Chechen (Kingston and Nichols 1986). Though pharyngealization acts in a similar way to labialization, i.e. influencing the transition from the given obstruent into the following vowel (cf. Grawunder et al. 2009), both secondary articulations can occur together, as in the uvular series of Ubkyh (see Figure 6). But we can also find oppositions like /ʁwʕakʷu/ ‘hook’ and /ʁwankʼi/ ‘deer without antlers’ in Tsez. Comrie (2003) and others before have pointed at the correspondence of umlauted vowels with pharyngealized vowels in Tsez (cf. Tsez: /ʕatɬ/ Bezhta: /ätɬ/ ‘village’). Similarity in formant structure and articulatory posture on a small-scale basis (Grawunder 2015) seems to support this hypothesis. Although some pharyngealized lexemes without uvulars can be identified as Arabic loanwords, it can in general be ruled out that this is the main source of this contrast rather than resulting from the loss of a previous pharyngeal. Furthermore, there seems to be a related process of vowel fronting in certain parts of the lexicon; this applies mostly to loanwords from Arabic, which are presumably brought in by non-native speakers of Arabic. Low front vowels ([æ~a]) occur in these loanwords in a number of Nakh-Dagestanian languages such as Tsez, Bezhta, Lezgi, or Udi. Finally, we can also observe pharyngealization, manifested as “emphatic” obstruents, in non-Caucasian languages of the area, as in Kurmanji Kurdish (Jastrow 1977) and Zazaki (Keskin 2008) and Neo-Aramaic (Cereteli 1976).

Palatalization Palatalization is understood here as a secondary articulation of obstruents involving a palatal glide-offset and/or tongue body fronting. Among the NEC languages is it mainly described for languages of the Lezgi group such as Tsakhur (Kibrik and Testelec 1999: 16), Agul, and Rutul, but also for dialects of Dargwa and Lak (Kibrik and Kodzasov 1990). And besides a palatalized lateral approximant [ɬ] in Khwarshi (Khalilova 2013) and Khinalug (Smeets 1994: 370), palatalization seems not to appear further as a contrastive feature in the other Dagestanian languages.

Among the NWC languages, palatalized series of velar and uvular stops and fricatives occur in the inventories of Abzhywa Abkhaz and T’ap’anta Abaza and Ubykh (Hewitt 2005), and likewise in the Shapsegh and Hakuchi dialects of West Circassian (Colarusso 1988). The non-palatalized velar stop series in Ubykh is actually considered to be non-phonemic (Hewitt 2005). Outside the core area, this feature is found in Western Armenian (Vaux 1998), Kumyk (Berta in Johanson and Csató 1998: 302), and, of course, Russian.

Clusters Phonotactics allow, for example, Georgian monomorphic and monosyllabic clusters with up to eight segments. This is one extreme, whereas other languages in the Caucasus usually allow only two segments. In general, the clusters are
classified into ‘decessive’ clusters, such with places of articulation going from more front to more back (e.g. /bd, dg, px, t’k/) and, the other way round, ‘accessive’ clusters, from more back to more front (e.g. /db, gd, xp, k’t’/). On the whole harmonic clusters consist of consonants with identical manner of phonation (voiced, aspirated, glottalized), e.g. /dg, tx, t’q’/ (cf. Boeder 2005; Chitoran 2002; Klimov 1994). Harmonic and decessive clusters with a post-dorsal second component (pχ, txs, p’k’) seem to be preferred. With regard to legal initial obstruent clusters of the decessive type, one can observe that these apply for a number of NEC languages (see e.g. Chechen: Nichols 1994; Batsbi: Holisky and Gagua in Smeets 1994; Lezgi: Haspelmath 1996) and similarly for the NWC languages Kabardian, Abkhaz (Hewitt 2005), and Abaza (Colarusso 1988). In NWC languages these clusters have the same distribution as single consonants (Klimov 1994; Hewitt 2005), i.e. they occur in all positions. In Vaux's underlying phonological analysis (Vaux, 1998) Armenian is considered as having clusters of up to 10 segments, with an epenthetic vowel [ə] in the surface realizations.

2.3 Frequency of occurrence

Although attributive features occur in Caucasian languages, their productivity is still unknown, which leaves an uncertainty about the likelihood of spread (disregarding here sociolinguistic and extra-linguistic factors). According to Testen (1997), obstruents of the glottalic ejective series (/p’, t’, ts’, k’/ and Iron /tʃ’/) occur with lower frequency in Ossetic than in neighboring Caucasian languages and the majority of the forms in which they occur are Caucasian loanwords. He gives Iron k’utra, Digor k’otær ‘bush’ from Chechen k’otær, dʒit’ri ‘cucumber’ from Georgian k’it’ri as examples. Glottalic stops also occur in Russian loanwords, where they reflect Russian voiceless (unaspirated) stops, like in bulk’on ‘colonel’ < polkovnik, bap’iroz ‘cigarette’ < papiros (Testen 1997). While one cannot find out about the actual realization by looking at the frequency of occurrence, for example, in texts, one can still gain an insight into the degree of systemic usage and the degree of embedding of particular contrasts in lexical and functional morphology. Due to the apparent scarcity of large corpora for the area, I propose here an admittedly skewed shorthand in looking at the type frequency as we gain it from occurrences in a wordlist. Though this needs to be validated by token frequencies, type frequencies give an impression of how often a particular structure is employed in the make-up of lexical forms for a given variety.

The data used here mainly for (NEC) Nakh-Dagestanian are taken from Comrie and Khalilov (2010), and frequencies extracted for the use of a particular contrast and its related features are approximated by graphemes. With respect to the prevalent uvular place of articulation in the systems on the one hand and the relation of frequency and markedness on the other, one can approach some questions with respect to the general markedness of uvulars cross-linguistically. Do uvular stops occur more frequently in the lexicon? It turns out that for the given sample of 76 doculecs of Nakh-Dagestanian languages, uvular glottalic stops occur more often than velar glottalic stops (t=3.9, df=75, p=0.0002), whereas the relative frequency (words with feature occurrence/words in
sample) of velar pulmonic stops is still higher than that of uvular stops ($t=-2.6$, $df=85$, $p=0.0103$). With respect to the distribution of these relative frequencies across the area of their spoken varieties, one can observe that we gain here a pattern that supports the idea of a core vs. periphery distinction, even within the Nakh-Dagestanian languages. On a South-to-North axis, along lines of latitude, we observe an increase in relative frequency for uvular glottalic stops (Figure 5). In a simplistic analysis applying a poisson error distribution, one can test the $z$-transformed latitudinal coordinates ($X_1$) against the frequency distribution (p-values for frequencies~$X_1$, $X_1$: $p=0.0008$, and frequencies~$X_1+I(X_1^2)$, $X_1$: $p<0.0001$, $I(X_1^2)$: $p<0.001$). These results would primarily allow one to support a hypothesis of a non-random distribution of these type frequencies. More advanced statistical methods of geographically weighted regression seem to be more promising, as demonstrated by Kirby and Brunelle (this volume).

2.4 Expression of features – phonetic realization

The acoustic parameters found in phonetic descriptions, which cover durational measure, spectral measures of vowel formants, or fricative noise characteristics, need to be generalized for the entire speaker population and to be compared with measures of the same sort for other populations or other areas with different varieties. This implies that the significance of the measure for describing a particular contrast is reasonably established.
Figure 5. A type frequency account of glottalic uvular stops in wordlist sample (N=86) of Nakh-Dagestanian languages; plot with frequency (as percent per doculect)

Nonetheless, linguistic phonetic analysis focuses on common physical properties and characteristics within a variety and uses methods to generalize over speakers/listeners. Sociolinguistic variation, of course, needs to be monitored as well and certainly adds certainly to the pool of co-factors to be controlled for. According to the H&H-theory of Lindblom (1990), we may expect, depending on the communicative situation, a range of
realizations across different phonetic dimensions.

A comparison across varieties requires comparable data sets. Whereas natural speech data collections are valuable in their own right, these are also much harder to assess in terms of validity (representativeness). Moreover, we face a data situation for comparable audio material that only allows for small samples, i.e. a handful of speakers per language which makes it all the more necessary to focus on such measures or indices that are not speaker-dependent. With a small-scale sample of 10 languages, the differences in phonetic expressions of the same phonological contrast can already be illustrated, namely that of pulmonic vs. ejective (glottalic) stops. However, the relativity of the measure has to be defined for each phonetic attribute. Two possible measures are suggested that should allow the initiation (phonation) contrasts of stops in the area to be characterized: (1) F0 (fundamental frequency) of the following vowel with discrete attributes like “higher”, “lower”, and “equal” (cf. Warner 1996 for Ingush) with respect to (pulmonic) contrasts in the same position, and (2) VOT (voice onset time), which is the time from stop release (burst start) to the onset of the following vowel (voicing onset) (Abramson and Lisker 1964 in general; Catford 1992 for Caucasian stops). Especially for VOT, a vast amount of literature exists on conditional effects like place of articulation, position in the word and phrase, and speaking rate (cf. Cho and Ladefoged 1999). Likewise, due to the possible effect of intrinsic F0 for the suggested F0 measure, the following vowel (fronting) may actually play a role. These are conditions that need to be controlled for in the sample. Other measures, like the closure duration, the total duration (= duration of closure plus VOT), or the recently developed intensity slope (Grawunder 2013), may also work here to differentiate stops and, specifically, phonation/initiation contrasts. Although these may eventually serve as diagnostics as well, we measure here pre-established categories arising from descriptions and grammars. In this preliminary dataset, we can already observe for VOT a split behavior of languages that exhibit VOT in aspirated stops that is higher than ejective stops, or the other way round (see Figure 6). This is incongruent to the compared VOT as observed by Catford (1992) since he averaged over place of articulation and suggested a uniform behavior in terms of aspirated stops being longer than ejectives. We can in turn presume that this split results from a difference in the way the contrasts are “organized” in the language, perhaps with respect to the way aspiration as a marker is implemented for the voiceless pulmonic stops on the one hand and of the gemination contrast (“intense” vs. “weak”) in these languages. F0 of the following vowel does not, on the other hand, follow the split above, but more importantly, it exhibits another split, which is not immediately predicted but which points to another dimension of the pulmonic/glottalic contrast implementation.
Figure 6. Relative measures for F0 of the following vowel of a stop (lower panel); relative measures for VOT (voice onset time) with pulmonic, i.e. aspirated (asp), vs. ejective (ej) stops

Grawunder et al. (2009) observed that aspiration in voiceless aspirated stops in Georgian overlaps onto the beginning of the following vowel. This implies a subsequent lowering of F0 with respect to unaspirated stops. The F0 difference that Grawunder et al. (2009) observed for the Ingush female speaker was around one to two semitones and
therefore has to be counted as a potential “signal” that is presumably also picked up by a listener. Nonetheless, the described F0-parameter and its general trend for an entire speaker population must be called into question since Grawunder et al. (2009) (eight speakers) and Vicenik (2010) (five speakers) found no clear tendencies.

3 Conclusion

It has been argued here that the Caucasus is a clear linguistic area and an area of contact – one that can be circumscribed by means of phonetic/phonological attributes, which appear to occur more often within the area and fade out towards its periphery. These attributes mostly concern feature density in both the sound system and across the area. The latter is driven by medium linguistic diversity and long-standing contact, slow intertwining migration patterns and prevalent exogamy. Supporting areality, various forms of functional multilingualism with incomplete L2/L3 acquisition come into play. The core of phonological features in the Caucasus area is built around a minimum three-way phonation/initiation contrast that includes ejective (glottalic) stops, which in addition are most likely to involve the uvular place of articulation. These obstruents often contrast with the presence or absence of segments that may involve subphonemic or phonemic labialization. In addition, one finds pharyngeal/epiglottal articulation either as a pharyngeal stop or as a fricative, along with possible suprasegmental processes. Another trait is seen in the process whereby the density of otherwise common features in the obstructed system of contrasts, such as labialization, palatalization, and gemination, leads to higher complexity in terms of feature embeddings per possible slot in the system. From an areal perspective, this includes the combination and stacking of features, a process which represents a consistent expansion of (feature) components that are already in place. Nonetheless, linear feature combination, i.e. the clustering of consonants, forms an integral part of the areal profile of the Caucasus.

The genealogical borders of the three main Caucasus families overlap with different “front lines”, which can often be drawn showing particular groupings for individual features like nasalization or umlauted vowels. In other words, we do not see a process of complete assimilation. There is no vast area with completely the same setting. However, a number of structural components remain typical for the core area. The proposed area can and needs to be tested against its periphery (cf. Bickel 2008 and Bickel, this volume), whereas, given the availability of data for a particular area, those frequency distributions are more suitable for fine-grained mapping and for demonstrating particular areal shifts. In order to enable solid quantitative hypothesis testing, more data is necessary, especially for those languages of the wider contact pool such as Tat (Muslim/Judeo) and Talysh, but also for dialectal and sub-varieties of languages that are otherwise often treated as homogeneous, like Georgian, Kurdish, Armenian, and Azerbaijani. This relates to the known problem of density of data points, which relates directly to population density (number of settlements, homogeneity of varieties, etc.).

Apart from the historical and socio-demographic situation and apart from general trends observable through phonemic systems today, it seems possible to enhance our understanding of areal feature distributions by means of type frequencies, as they are found in dictionaries. The other suggested shift in the typological approach, focuses on the sound level and emphasises phonetic realizations rather than higher systemic levels. Feature expression seems far more appropriate (in the Caucasus area) as we often find...
similar phenomena, or transitional and gradual phenomena, that can surface as different features via interaction with other levels (syntax, prosody, lexicon). The underlying (phonetic) mechanisms need to be understood beforehand so that appropriate analytical measures can be chosen and clear predictions can be made with regard to feature expression. This phonetic approach opens up wider possibilities for substantiated hypotheses about historical sound change, be it induced through contact on the one hand or through “internal” mechanisms of analogical expansion and optimization of sound systems on the other.

Acknowledgements

I wish to express my gratitude to the editor Raymond Hickey for his infinite patience and assistance during the genesis of this chapter in its manifold versions. Furthermore, I would like to thank Bernard Comrie and Brent Reed for their very helpful and valuable comments. My thanks also go to Hans Jörg-Bibiko, Lisa and Natalie.

References


Dzhidalaev, Nurislam S. 2010. Tjurksko-dagestanskie jazykovye kontakty (na primere lakskogo jazyka) [Turkic-Dagestanian language contacts (on the example of Lak)]. Makhachkala: Russian Academy of Science in Dagestan.


Haspelmath, Martin. 2006. Against markedness (and what to replace it with). Journal of Linguistics, 42(01), 25–70.


Hickey, Raymond. 2012a Areal features of the Anglophone world, in: Hickey, Raymond
Grawunder -- page 360 of 393


Moscow: RAN.


in the Lezgian Language]. Makhachkala: Dagestan State University.
13 Western Asia: Anatolia as a transition zone

Geoffrey Haig

1 Introduction

What is loosely referred to here as “East Anatolia” covers the region of Turkey eastward of the town of Sivas (see Figure 1). However, the cultural and linguistic traits that characterize the region do not cease at the borders of the Turkish state, but permeate outwards in all directions: into western Anatolia, Armenia and the Caucasus, west Iran, North Iraq, and Syria. For this reason alone, attempts to define an East Anatolian ‘linguistic area’ are generally not compelling - if indeed the concept of ‘linguistic area’ has any intrinsic value at all (cf. Campbell 2006); this discussion is taken up in section five below. Topographically, the region encompasses the upper catchment region of the Tigris and Euphrates rivers; it is ruggedly mountainous in the north and east, descending gradually towards the lower elevation of the plains of Mossul in North Iraq.

The earliest attempt at characterizing language contact across the region, Haig (2001), was based on data from just four languages, and the conclusions were correspondingly tentative. Since then, the amount of available data has increased dramatically: for North Eastern Neo-Aramaic (NENA), Geoffrey Khan and his associates have produced rich documentations for at least six varieties, for Laz we now have the monumental grammar of the Arhavi variety (Lacroix 2009); several Arabic dialects of the region have been documented (cf. Talay 2011 for an overview), material on Homshetsma (West Armenian) is now more accessible (Vaux 2007), likewise on Domari of Aleppo (Herin 2012). For Kurmanji Kurdish, see now Haig and Öpengin (2014) and Öpengin and Haig (2014). Furthermore, conditions for field work in the region have vastly improved since 2000, and a number of local academic initiatives have been established in Eastern Turkey, in particular at Artuklu University in Mardin. These are, of course, exceedingly welcome developments, but they also mean that an overview chapter such as this cannot achieve anything like comprehensive coverage. In what follows, I will present what I consider to be the most salient indicators of language contact across the region, and assess their relevance against what is known from language history, and from related languages. In doing so, it will become evident that the notion of Anatolia as a coherent areal unit is probably not tenable.

The chapter is organized as follows: in section two, the notion of Anatolia as a transitional zone is outlined, and the varieties discussed in this chapter are introduced. Section three takes up selected structural features and discusses their distribution across the region, beginning with phonology. In section four, word order and adpositions are discussed, while section five critically assesses attempts at defining an Anatolian linguistic area, and proposes an alternative view, involving two sub-regions. Section six sums up the main points of the chapter.

2 East Anatolia as a transitional zone
To appreciate the dynamics of language contact in the region, it is worthwhile beginning with a macro-linguistic perspective: East Anatolia straddles the intersection of several major linguistic macro-areas: to the south, Anatolia fades into the Arabian peninsula and North Africa, dominated by VO, prepositional Afro-Asiatic languages; westwards it blends into the Mediterranean / Western European region, with their specific peculiarities (cf. Zeldes 2013); eastward we move into the central Asian region of OV and postpositional languages (Turkic, Eastern Iranian, Indo-Aryan etc.). Finally, the northeastern fringe of Anatolia borders on the Caucasus, and outlier Caucasian languages are also found in Anatolia (e.g. Laz, Kartvelian). The impact of all these regions are evident in various ways, which are discussed below.

As a consequence, what characterizes East Anatolia as a region is not the bundling of unique features emanating from a discernible geographic core, but the presence of multiple overlaps from the neighbouring regions. Nichols (1992) introduced the dichotomy between “spread zones”, and “residual zones” to account for global patterns of feature distribution. Spread zones are large areas, with a low density of genetic variation among the languages of the region. Such zones are created through the expansion of a dominant group that imposes its language(s) over a large region, typically a region with few geographic obstacles that would inhibit the expansion of the dominant group. By contrast, a residual zone is characterized by high genetic density, which typically arises in geographically isolated and often inaccessible regions. A prototypical residual zone is the Caucasus, characterized by three indigenous (and probably unrelated) language families.

East Anatolia, however, is neither a spread zone, nor a residual zone. It is a transitional, or overlap zone. The relatively high degree of linguistic diversity in the region (see below), while typical of a residual zone, is not the result of a gradual accumulation of linguistic diversity through long-standing geographic isolation (the hallmark of Nichols’ residual zones), but arises from the fact that East Anatolia is at the cusp of a number of distinct macro-regions. All the languages of East Anatolia have close relatives in neighbouring regions; its linguistic diversity is thus not indigenous, but a secondary product of its transitional status. Nevertheless, East Anatolia is more than just the sum of its contributing regions. It can be argued that some of the linguistic features characterizing the languages of the region cannot be explained in terms of diffusion from neighbouring regions, but are specific compromise responses to the conflicting typological profiles of the neighbouring languages (cf. Stilo 2005, 2012). This is particularly true in the realm of word order, and adpositions, discussed in section four below.

Finally, a brief note on the sociolinguistic situation. Over the last two millennia, the region has been at the intersection of two imperial epicentres, the Byzantine and Ottoman empires centred at Istanbul, and successive empires of Iran, e.g. the Achaemenid and Sassanid. Although different parts of the region nominally belonged to one or other of the dominant empires at various times, it is important to realize that up until the twentieth century, there was no single dominant “official” language across the region, no compulsory mass education in that language, and no pressing reason, or indeed any particular advantage, for speakers to abandon their own language in favour of another. Although Islam was spreading throughout the region from the tenth century, numerous non-Muslim groups remained in situ and continued practising their religions. It was not until the upheavals of the early twentieth century that their situation deteriorated dramatically, unleashing an exodus that continues to the present day. But prior to this
period, language was an emblematic feature of group identity, often coupled with membership in religious groups, or tribal affiliation. And while some languages enjoyed more prestige than others, the vernacular Turkish of Anatolia and west Iran was notably not among them (cf. Bulut 2002: 55). Thus the general pattern that can be assumed over most of the period is the acquisition of neighbouring languages if required for trade and other purposes, coupled with retention of one’s own.

With the founding of the Turkish Republic in 1923, a policy of wholesale Turkification was imposed on the entire country, leading to what can only be described as a catastrophic rupture in the delicate multi-lingual ecology of East Anatolia. In fact, the destruction of Anatolia’s linguistic diversity had begun prior to the founding of the Republic, with the wholesale eradication of the Armenian minority, and the forced re-settlements of thousands of Neo-Aramaic speaking Jews and Christians. Over subsequent decades, Turkey’s one-state-one-language policy has resulted in massive language shifts towards the official language, Turkish, so that now, only the largest minority language communities have retained any degree of viability (cf. among others Skutnabb-Kangas and Bucak 1995, Haig 2004, Öpengin 2012 on Turkey’s policies towards its minority languages).

In principle, then, it is important to distinguish the results of (probably) gradual convergence among the languages of East Anatolia over centuries of multi-lingual co-existence, with each group maintaining its own language (cf. Noorlander 2014), from the situation after the founding of the Turkish Republic, where Turkish has been imposed as the dominant language, and the speech of minority groups is characterized by imperfect first-language acquisition (or complete shift to Turkish), and the code-switching practices evident in many of the growing urban centres of the region.

2.1 Overview of varieties

The following list introduces the twelve varieties to which reference is made in this chapter. The numbering is reflected in Figure 1, which indicates the respective locations.

1. Hemshinli (also called Homshetsma), a dialect assigned to Western Armenian and spoken in Artvin province of northeastern Turkey on the Georgian border (Vaux 2007). The speakers converted to Sunni Islam at least two centuries ago.

2. Laz, a Kartvelian language spoken on the Black Sea coast of Turkey in a few towns and villages just short of the Georgian border; data here is taken from the Arhavi variety described in Lacroix (2009).

3. Northern Kurdish or Kurmanji Kurdish. Northwest Iranian, Indo-European. The most widely-spoken minority language in East Anatolia. It is also spoken in the Urmuye region of Iran, in northeastern Syria and north Iraq.

4. Central Kurdish (Sorani), Northwest Iranian, Indo-European, is spoken southwards of Kurmanji, and in the southeast towards Mahabad in Iran.
5. The so-called Şêxbizî dialects of Kurdish, the language of the descendents of southern Kurdish tribes re-settled in various parts of Anatolia in the sixteenth century. With the exception of a short description in Lewendî (1997), further data on these varieties are currently unavailable (data here from the Haymana variety).

6. Domari, the language of the Romani communities of much of the Middle East (Indo-Aryan). There are Dom communities scattered throughout much of East Anatolia, reputedly several thousand speakers in the Diyarbakir region. However, reliable information is extremely difficult to obtain; the most relevant recent source is Herin (2012) on the Aleppo variety (Syria); cf. Matras (2012) for an overview of Domari.

7. Zazaki, Northwest Iranian, and culturally often considered part of “Kurdish”, is considered by Iranian philologists to be a distinct language; cf. Paul (1998)

8. Turkish, Turkic. The sole official language of the Turkish state since 1923, and the basis for Ottoman Turkish, the language of the Ottoman Empire. There is a chain of dialects across Anatolia spreading into present-day Azerbaijan and Iran (Bulut 2002, 2006), but Turkish influence in the far southeast (south of Lake Van) was not particularly strong prior to the twentieth century. Material for this chapter is from Tabriz (Kural 2001), Erzurum (Gemalmaz 1995) and Turkmen of Iraq (Bulut 2007).

9. North Eastern Neo-Aramaic (NENA). A divergent group of dialects spoken by Christian and Jewish communities east of the Tigris river in Turkey, West Iran and North Iraq. NENA belong to the West Semitic languages. The varieties included here are Jewish Urmi (Khan 2008), Bohtan (Fox 2009), Ashitha (Borghero 2005), Barwar (Khan 2008).

10. Central Neo-Aramaic, spoken further west but closely related to NENA, represented here by the variety of Turoyo (Jastrow 1992).


3 Selected structural features

3.1 Phonology

The vowel systems of the Iranian and Semitic languages of eastern Anatolia are strikingly similar. They typically involve 7-9 vowel phonemes, of which five are long vowels, and two or three are short. However, vowel length by itself is not a distinctive feature, and most vowels can be distinguished phonemically through vowel quality (short vowels are invariably more centralized than the long vowels). Table 1 shows the vowel phonemes of four varieties; the transcriptions of the sources have been normalized:

<table>
<thead>
<tr>
<th>Semitic</th>
<th>Iranian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td></td>
</tr>
<tr>
<td>[i]</td>
<td>[i:]</td>
</tr>
<tr>
<td>[e]</td>
<td>[e:]</td>
</tr>
<tr>
<td>[a ~ æ]</td>
<td>[a:]</td>
</tr>
<tr>
<td>[o]</td>
<td>[ɔ:]</td>
</tr>
<tr>
<td>[u]</td>
<td>[u:]</td>
</tr>
<tr>
<td>Short</td>
<td></td>
</tr>
<tr>
<td>[ɔ]</td>
<td>[a]</td>
</tr>
<tr>
<td>[a]</td>
<td>[e]</td>
</tr>
<tr>
<td>[æ ~ ɛ]</td>
<td>[ɔ]</td>
</tr>
<tr>
<td>[ʊ]</td>
<td>[u]</td>
</tr>
</tbody>
</table>

Table 1. Typical vowel systems in Iranian and Semitic languages of Anatolia
Turkic languages of the region likewise typically exhibit approximately eight vowel phonemes, and length is also generally not distinctive. However, Turkic varieties also include the front rounded vowels [y] and [œ]. In some varieties of Northern Kurdish, these vowels may occur as allophonic variants of the long vowels [u:] and [o:], and in loanwords. In the southeastern dialects (Hakkari (Kurd. Colemerg), Şemdinli (Kurd. Şemzinan), [u:] tends to deround and front towards [i:], and [y] is often an intermediate stage in this process.

Turning to the consonants, a striking feature in both Zazaki and Northern Kurdish is a three-way distinction on the stops and affricates with regard to Voice Onset Time (VOT). For Northern Kurdish, the following is fairly typical:

<table>
<thead>
<tr>
<th>articulation</th>
<th>bilabial + velar</th>
<th>alveolar + affricates</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiceless aspirated:</td>
<td>[pʰɔ:r] ‘hair’</td>
<td>[tʰæv] ‘together’</td>
</tr>
<tr>
<td>voiceless, unaspirated</td>
<td>[p:o:z] ‘nose’</td>
<td>[te:vr] hoe, mattock’</td>
</tr>
<tr>
<td>voiceless aspirated:</td>
<td>[kʰa:r] ‘work, matter, concern’</td>
<td>[ʧʰɪma:] ‘why’</td>
</tr>
<tr>
<td>voiceless, unaspirated</td>
<td>[ka:j] ‘old man’</td>
<td>[ʧɛm] ‘stream, brook’</td>
</tr>
<tr>
<td>voiced</td>
<td>[g:a:v] ‘step, time’</td>
<td>[ʤɛm] ‘by, beside’</td>
</tr>
</tbody>
</table>

Table 2. Three-way VOT-distinction on bilabial, alveolar and dental stops, and affricates in Northern Kurdish

The nature of the phonetic distinction between the two voiceless series has been the subject of some controversy. What are termed here “unaspirated” have been claimed to possess an ejective quality, and in some dialects, there is an undeniable perceptual similarity with (weak) ejectives, particularly with the unaspirated affricate [ʧ]. However, it is also the case that there is a significant difference in VOT, and most phonetic investigations have focussed on this feature (Grawunder et al. 2013), while not excluding the possibility of an additional pharyngeal articulatory gesture. There is no plausible historical source for the additional series within Iranian, and it seems therefore most likely that they result from contact, most probably with East Armenian (cf. Gippert 2007/2008 for discussion), though Neo-Aramaic has also been suggested as a possible influence (Kahn 1976). The additional series is also possibly preserved in the Armenian loanwords in the Erzurum dialect, though the interpretation of the material is extremely difficult (Menz 2010). However, the additional series is completely absent in Central Kurdish in Iraq, and our impression is that the frequency and saliency of the effects diminish across Anatolia towards the southeast, reflecting a diminishing East Armenian influence.

All varieties of Kurdish have adopted pharyngealized consonants in some measure from neighbouring Semitic languages. However, the adoption of loan Semitic consonants in Kurmanjí is subject to certain constraints: interdental consonants are never borrowed, and geminate consonants in loanwords are generally de-geminated (cf. Kahn 1976: 80). In the only detailed study of loan phonology in Kurmanjí, Kahn (1976:89) notes a tendency for the over-generalization of pharyngealization in Arabic loans in Kurmanjí.
Thus Iraqi Arabic [ʤa:hi:l] is rendered [ʤa:xel] ‘young’, [zulal] as [zæːlal] ‘(clear) water’. In the Kurmanji investigated by Kahn, pharyngealization in loanwords thus becomes a general signal of “Arabic origin”, rather than the actual rendering of the source phonology. It can, however, spread to items of the inherited lexicon; in Kurmanji Kurdish of the Iran/Turkey border region, Kahn (1976: 34) notes characteristic “pharyngeal formant bending” for the first vowels in [sæd] ‘100’ and [zawa] ‘bridegroom’, indicative of the pharyngeal quality of the initial consonants. Similarly, [ḥ] is found in native words such as [hæft] ‘7’ or [mæhin] ‘mare’.

The Semitic languages of the region are not just the donors of loan consonants, but have also undergone considerable phonological restructuring themselves under contact influence. For the Mesopotamian dialects of Arabic, Talay (2011: 912) notes the addition of /č/ (= [ʧ]), /g/, /p/ and /v/ to the native phoneme systems, clearly through the contact influence of Turkic and Iranian. Similarly, most of NENA have adopted /č/ (= [ʧ]) and /ʤ/, and sporadically /ʒ/ (Kapeliuk 2011). Talay (2007:181) also notes that some Anatolian Arabian dialects (e.g. Hasköy) have lost pharyngealization on some consonants, e.g. tare < *ṭarīq ‘way’ and sēbī < *sabīy ‘boy’. Furthermore, in the Kozluk/Sason dialects, the characteristic Semitic interdental fricatives /θ/ and /ð/ (and their pharyngealized counterparts) may be realized either as sibilants or as stops. The latter are also lost in most varieties of NENA. Kapeliuk (2011) also notes the shift of interdental fricatives to dark-l in Jewish NENA of Azerbaijan, a development that has clear parallels in the Iranian languages of the region. The net result or these changes is to align the phoneme system of the local varieties of NENA and Arabic quite closely with those of their long-standing Kurdish neighbours.

3.2 Morphosyntax

3.2.1 Comparatives

All languages of Anatolia express the standard of comparison through a ‘source’ adposition or case marker (lit. ‘he is bigger from-me’), either an ablative case suffix as in Turkish (-den/-dan/-ten/-tan) or in Laz (-şen, Lacroix 2009: 93), or a preposition with a local meaning ‘from’, e.g. ji in Kurmanji, or mān (or variants thereof) in NENA and Arabic. This appears to be a universal feature across Anatolia.

The languages differ, however, as to how the adjective itself is marked as expressing the comparative degree. Three options are available: (i) There may be a particle preceding the adjective, as in Turkish daha, or NENA biš (Khan 2008: 229); (ii) the adjective may be left unmarked (Laz, Lacroix (2011:93); this is also an option for Turkish); (iii) the Iranian languages have an inherited comparative suffix with most adjectives, as in Kurmanji drevj-tir ‘long-er’. However, some Kurmanji speakers use the Turkish particle daha to express the comparative degree of an adjective. Conversely, the
Kurdish comparative suffix –tir (or variants thereof) has been borrowed into Domari (Herin 2012:23) and into Turoyo Neo-Aramaic (Noorlander 2014). The cognate comparative suffix in Persian has also been borrowed into the Azerbaijani Turkish of Tabriz in Iran (Kural 2001: 19). The Turkish superlative particle en (en güzel ‘most beautiful’) is sometimes heard in Kurmanji Kurdish, and is reported for Domari (Herin 2012), Laz, and Zazaki (Paul 1998: 58). In general, the morphosyntax of comparison is highly susceptible to contact influence, as noted for other contact regions (e.g. the Circum-Baltic languages, see Koptjevskaja-Tamm and Wälchli 2001: 684).

3.2.2 Copular constructions

Most (all?) of the languages of Anatolia have a common pattern of expressing identity, or feature attribution, in copular clauses. The construction has been mentioned as a pan-Anatolian feature in Matras (2009: 270), and it is worth looking at in more detail. It involves a (usually) clitic copular element that follows the nominal predicate, yielding the pattern illustrated schematically in (1), see Table 3 for examples:

(1) (this man) clever=is ‘this man is clever’

<table>
<thead>
<tr>
<th>Language</th>
<th>non-verbal pred.</th>
<th>copular, 3sg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turkish of East Anatolia</td>
<td>güzel ‘beautiful’</td>
<td>=di(r)</td>
</tr>
<tr>
<td>Northern Kurdish</td>
<td>jwan ‘beautiful’</td>
<td>=e</td>
</tr>
<tr>
<td>Turoyo Aramaic (Talay 2006/2007:183)</td>
<td>şefir ‘beautiful’</td>
<td>=o</td>
</tr>
<tr>
<td>Arab. (Kinderib) (Talay 2006/2007:183)</td>
<td>malîh ‘beautiful’</td>
<td>=we</td>
</tr>
<tr>
<td>Neo-Aramaic, Khabur dialect (Talay 2008:201)</td>
<td>xamte ‘beautiful’</td>
<td>=la</td>
</tr>
<tr>
<td>NENA (Urmi, Khan 2008)</td>
<td>jwän ‘beautiful’</td>
<td>=ile</td>
</tr>
<tr>
<td>Laz (Lacroix 2009:148)</td>
<td>genci ‘young’</td>
<td>yên</td>
</tr>
<tr>
<td>Domari of Aleppo (Herin 2012:8)</td>
<td>girân ‘heavy’</td>
<td>=e</td>
</tr>
<tr>
<td>Şêx Bizinî, Haymana dialect (Lewendi 1997)</td>
<td>rind ‘good’</td>
<td>=e</td>
</tr>
</tbody>
</table>

Table 3. Copular constructions, third person singular, present indicative

There are two important commonalities here: first, an overt copula in the third person singular, indicative present, is obligatory, or at least extremely frequent. Second, its position in pragmatically neutral clauses is clause-final. However, we can only meaningfully assess the plausibility of a contact scenario by considering the histories of the different languages individually.

Turning first to the Iranian languages, the two features (clause-final position, and obligatoriness) are inherited, and have characterized West Iranian throughout its attested history. Thus contact has not necessarily played a role in the copulas of these languages. For Turkic of the region, the presence of an obligatory third person copular in the
indicative present can plausibly be attributed to Armenian and Iranian influence. In vernacular standard Turkish, there is no “overt marking of the third person singular on predicates” (Göksel and Kerslake 2005: 128), thus to say ‘my mother is tired’, the two words ‘mother-my’ and ‘tired’ are simply juxtaposed, with no overt copular element: Anne-m yorgun. Spoken standard Turkish does have a clitic element =dir (with vowel-harmonic variants), which may occur with third person nominal predicates, but it has specific pragmatic and modal functions, and is considered to be “nowadays largely confined to formal language” (Göksel and Kerslake 2005: 341). What we observe in spoken Turkish from East Anatolia is a perceptible increase in the frequency of =dir, though this has not been noted in the literature nor systematically investigated. In the Azeri dialect of Tabriz (Kiraz 2001) and in the Turkmen dialects of North Iraq (Bulut 2007), both under heavy Iranian influence, the use of an overt copula for the 3sg. -di, appears to be obligatory. Thus Turkish has moved closer to the pattern of Table 3 through a decrease in pragmatic markedness of an available option (=dir), and its consequent increase in frequency, rather than by any obvious restructuring.

Moving to the Semitic languages, it is here we find the most radical re-structuring of inherited copular constructions. Historically, present-tense indicative clauses with non-verbal predicates lack an overt copular element; subject and predicate nominal are simply juxtaposed, and where overt copulas are found (e.g. in perfective aspect), they are not clause final. This situation is maintained in most of modern Arabic dialects outside Anatolia (Jastrow 1980: 148). Thus in colloquial Arabic of the Gulf and Saudi Arabia, we have the following in the present tense:

(2) iš-šārīka kabīra
DEF-company.FEM big.FEM
‘The company (is) big’ (Holes 1992: 41)

In Arabic dialects of Anatolia, however, an overt copula has become obligatory, though its position varies; in the Siirt dialect we find the order (Subject) - Cop - Non-verbal predicate (Jastrow 1980: 148), which appears to be the only counter-example to the general Anatolian pattern of clause-final copulas (but see below):

(3) āvi l-bont iye bōš malīḥa
this DEF-girl COP.3SG very good.F
‘This girl is very good’

In the other qalṭu-dialects of East Anatolia, we find the clause-final type of Table 3, illustrated for the Mardin dialect in (4):

(4) hādi l-bont kṯīr malēḥa=ye
this DEF-girl very good.F=COP.3S

Example (5) would match word-for-word with the translational equivalent in neighbouring dialects of Kurmanji:

(5) ev keč gellek jwan=e
Most varieties of NENA also have enclitic clause-final copulas, as in the Bohtan variety (Fox 2009:62):

(6) \( \text{awa} \) \( \text{ṭowé=le} \)  
\( \text{good=COP.3M.SG} \)  
'\( \text{This is good} \)'

In the area of copular constructions then, the Semitic languages of Anatolia have undergone major internal restructuring of their inherited Semitic constructions to bring them in line with the general Anatolian pattern of obligatory clause-final copulas, presumably due primarily to Iranian influence. However, in southeast Anatolia a couple of wrinkles disturb the overall pattern and these are worth pointing out briefly. They suggest that the Siirt pattern of medial copula, illustrated in (3) above, is not completely isolated.

A number of NENA dialects have an additional copular construction with a particular pragmatic force, described in Borghero (2005: 167) as being used to “draw attention to a specific situation or to a non-permanent property of the subject”. This kind of copula is clause-medial, and in the Ashitha dialect, may also occur together with non-finite full verbs to express present continuous action. Similar forms are noted for several varieties of NENA including Bohtan (Fox 2009: 63) and Hertevin (Jastrow 1988: 27), who refers to the ‘aktuelle Kopula’. In the Kurmanji dialects of southeast Anatolia and North Iraq, there is also a medial copula, based on the ezafe-particle. It is formed by an ezafe particle which agrees in number and gender with its antecedent, but rather than link an adnominal modifier to the head, it introduces the predicate, in particular a non-verbal predicate (Haig 2011). The following example has a prepositional phrase as predicate (Kurmanji of Amadiye, North Iraq, Blau 1975:106):

(7) \( \text{Xatûn=a} \) \( \text{min} \) \( \text{ya} \) \( \text{l hîvî=ya} \) \( \text{te} \)  
\( \text{lady=EZ.FEM} \) \( \text{my EZ.FEM} \) \( \text{at expectation=EZ.FEM your} \)  
'My lady is waiting for you (right now)' (lit. My lady which at your-waiting)

Another typical example is found in the name of a song, sung by a singer from Zakho, North Iraq (the Roman alphabet transcription in the original is \( \text{tuyal bîra min} \), which is revealing from the point of view of speakers’ perceptions of word boundaries, but which obscures the underlying morphology; in the glossed example I have added the morphological segmentation). Note that the noun \( \text{bîr} \) is feminine, hence the feminine form of the final ezafe, which is thus simply coincidentally identical to the first ezafe:

(8) \( \text{tu} \) \( \text{ya} \) \( \text{l bîr=a} \) \( \text{min} \)  
\( \text{2SG E2.FEM in memory=EZ.FEM my} \)  
'you (fem.) are on my mind' (lit. you which on my mind)

As in the Ashitha NENA case mentioned above, the use of these clause-medial copular elements with non-verbal predicates typically expresses ‘immediacy’, and non-permanent attribution. Again like the NENA medial copulas, the Kurmanji particles may occur with full verbs to express the immediacy and continuity of the event. This is illustrated for the Bohtan dialect of NENA (Fox 2009: 63) in (9):
Example (9) would translate word-for-word into the southeastern dialects of Kurmanji. For example the Şemzinan dialect of Kurmanji (Ergin Öpengin, p.c.) likewise uses a clause-medial ezafe particle combined with a verb to indicate immediacy:

\[(10)\] dayk-a axê ya dnûvî-t

mother-EZ.FEM Agha.OBL EZ.FEM sleep.PRS-3SG

The Kurmanji case is in fact considerably more complex than the brief discussion here might suggest (cf. MacKenzie (1961: 205-208) and Haig (2011) for details). But the presence of a medial copular form, inflecting for gender, and etymologically derived from a pronominal form (the demonstrative/relative ezafe), with a clear semantic connection to immediacy, appears to be a unique development within Kurdish, and restricted to just those varieties that have been in close contact with NENA. It certainly seems reasonable to assume mutual influence in the developments sketched here, which illustrate rather nicely how within Anatolian, sub-regions may develop quite divergent phenomena, which do not necessarily diffuse across the entire region.

3.2.3 The clitic additive, or “recalled topic” marker

All languages known to me in the region possess an enclitic particle with a pragmatic function signalling some notion of contrast, addition, or the re-introduction of a previously topical element. It has no exact correspondence in English; according to context, it could be translated with ‘and X’, ‘X too’ or ‘as for X’. It may also follow a predicate, in which case it has the entire event in its scope. The relevant marker is =ţī in Kurmanji and Zazaki, =ṭi in Laz, =da/de in Turkish, and =se in NENA of Bohtan (Fox 2009: 103). An example of typical usage from Zazaki is the following, where the preceding context describes how the friends of a young boy used to call him by the nickname of Gukulah (Paul 1998:232):

\[(11)\] Lājîkī=ţī enā leqamdâ xwi-râ zûf xûy kerdlînî

boy=ADD this nickname.of his=ABL very annoyance did

‘As for the boy, he was very annoyed about this nickname of his’

See Haig (2001:208) for further exemplification of Kurmanji, Laz and Turkish. In Central Kurdish, the same function is expressed through =ṭiš, which has been borrowed directly into Domari of Aleppo (Herin 2012: 47), and some dialects of Neo-Aramaic, for example in the dialect of Suleimaniye and Halabja (cf. Khan (2004:399-408) for extensive discussion), and Urmi (West Iran):

\[1\] The use of gender-sensitive particles following the subjects of copular constructions in dialects of Kurmanji spoken around Maraş in Turkey is quite similar, but these occur together with the normal clause-final copula (cf. Öpengin and Haig 2014).
However, Iranian languages further east such as Persian (h)æm, or Hawrami to the southeast, also have similar particles, as do some languages of the Caucasus and indeed from much further afield. Thus although this appears to be a feature of (possibly all?) languages of East Anatolia, it is also cross-linguistically quite common.

4. Word order: verbs and objects, and adpositions

For Standard Turkish, Laz, and Eastern Armenian, both direct and indirect objects generally precede the verb, though the relative ordering of the two objects (if both are present) may vary according to their information status (for example, indefinite direct objects tend to immediately precede the predicate). Objects are flagged via case-suffixes or enclitics, though the presence of an overt case marker may be mediated by Differential Object Marking. This is given schematically in (13); the term ‘Goal’ here refers to recipients, addressees as well as local goals, and is intended to refer only to full NPs, rather than pronominal objects (Haig 2014b). The symbol ‘~’ links two elements whose ordering may be reversed:

\[(13) \quad \text{Goal-case} \sim \text{Object-case} \quad \text{Verb}\]

The following examples illustrate the preverbal position of direct object and goal for Eastern Armenian (14), and Laz (15), with glosses adapted and simplified:

\[(14)\quad \text{dasaxos-ē nor girk-ē mi usanok-i talis ē} \]
\[\text{lecturer-DEF new book-DEF INDEF student-DAT givePTCP.PRS COP.3S} \]
\[\text{‘The lecturer gives the new book to a student’ (Dum-Tragut 2009: 566)}\]
For the Semitic languages, on the other hand, it can reasonably be assumed that historically, in pragmatically neutral clauses, both kinds of object follow the verb, and if flagged at all, then via prepositions, yielding (16):

(16) Verb (prep-)Object ~ prep-Goal

This is illustrated for NENA, the dialect of Barwar (17) and of Hertevin (18):

(17) ʔu-máxa xa-mášxa gu-be²éne dîye
and-they.put INDEF-oil on-forehead of.him
‘and they put oil on his forehead’ (Khan 2008: 882),

(18) drélele ḫalwe-hen l=maška
they.poured milk-POSS.3PL into-butter.churn
‘they poured their milk into the goatskin butter churn’ (Jastrow 1988: 108)

The two word-order profiles illustrated in (13) and (16) represent diametrically opposed types, and it is of considerable interest to observe the various accommodation strategies that have emerged through contact between them. Two changes can be observed in the languages of the region.

First, among some of Semitic languages we find a shift from VO to OV, notably in the Bohtan dialect of Christian NENA speakers, and the Urmi variety of NENA (Jewish speakers, Khan 2009). Examples (19-20) are from Bohtan:

(19) brota axost-aw yawó-la l-jambali
girl ring-POSS.3SG.FEM give.PST-3SG.FEM to-Jambali
‘The girl gave her ring to Jambali’ (Fox 2009: 101)

(20) danw-i mmun ʔqlax-le
tail-POSS.1SG why cut_off.2SG.FEM-OBJ
‘Why did you cut off my tail?’

However, OV word order in Bohtan is not the only option, and Fox provides several examples with post-predicate objects. Thus OV word order of the Bohtan dialect, unlike the OV order of neighbouring Kurdish dialects, is not a categorical rule, but rather a preference, “the most common order” of verb and object (Fox 2009: 107).

Example (19) also illustrates the second point: the goal argument (l-jambali) is post-verbal, yielding the order OVG. This kind of word order is typologically unusual, and has received very little attention in the literature on word-order.² It has a clear areal

² One of the few typological contributions is Hawkins (2008), who considers the relative ordering of direct object and “oblives”. However, Hawkins’ notion of ‘oblique’ is not co-extensive with ‘Goal’, as used here. Hawkins’ obliques do not include Recipients, and do not extend to the local goals of intransitive verbs.
distribution, characterizing those dialects of NENA that have shifted to OV (e.g. Bohtan, or Urmni, Khan 2008: 332), and all varieties of Kurdish spoken in southeast Anatolia and North Iraq, i.e. precisely those varieties where contact with Semitic must have been most prolonged and intense. An example of Kurmanji from Amadiye (North Iraq) is the following:

(21) to vī qat’ā ‘ardī nā-da-ya min
2SG this.OBL piece land-OBL NEG-give.PRS-DRCT 1SG.OBL

‘Won’t you sell me this piece of land?’ (MacKenzie 1962: 340)

I refer to the order illustrated in (19) and (21) as the Mesopotamian OVG word order. It can be seen as a specific areal response to the conflicting demands of the two types illustrated in (13) and (16) above, and appears to be characteristic of all OV languages of North Iraq and southeastern Anatolia, including for example Iraqi Turkmen, a Turkic language under heavy Kurdish influence (Bulut 2007). Schematically, this is displayed in Table 4, whereby the respective ordering of Goal and Object is not rigidly fixed when both are on the same side of the verb:

<table>
<thead>
<tr>
<th>Exs</th>
<th>Type</th>
<th>Order of verb, object and goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(14), (15)</td>
<td>ORIGINAL TURKIC</td>
<td>Object ~ Goal Verb</td>
</tr>
<tr>
<td>(19), (21)</td>
<td>MESOP. OVG</td>
<td>Object Verb Goal</td>
</tr>
<tr>
<td>(17), (18)</td>
<td>ORIGINAL SEMITIC</td>
<td>Verb Object ~ Goal</td>
</tr>
</tbody>
</table>

Table 4. Word orders for direct object and goals in Anatolia

It seems reasonable to seek the source of the post-predicate goals of the otherwise OV languages in the Semitic languages, which generally position their goals after the verb. Northern Kurdish is particularly instructive for investigating the areal distribution of post-verbal goals because it is spoken across a large area, and subject to the influence of different contact languages. In the Mesopotamian region, it has been subject to heavy Semitic influence, while Semitic influence progressively diminishes as one moves northwards and westwards into Anatolia. The differences turn out to be at the level of differences in textual frequency (see Haig and Thiele 2014, Haig 2014b for details), but there is one type of goal which shows a more categorical change in position and flagging type, depending on the location of the dialect. This is the addressee of verbs of speech. There are three options for marking addressees, involving different positions relative to the predicate, and different kinds of flagging:

---

3 Note that OVG refers to full NP objects; ordering of pronominal objects may differ.
1. Post-predicate, with a preposition (or a directive (DRCT) particle on the verb, which is etymologically derived from a preposition). This possibility is illustrated in (22), from the Midyat region of Southeastern Turkey (Haig 2014b):

(22) gōt-a Ŭsufšā, xatūn-ē gō...
     say.PST-DRCT Usufsha, lady-OBL.FEM say.PST
     ‘(she) said to Usufsha, the lady said...’

2. Pre-predicate, introduced with a circumposition in (23):

(23) Wezîr ji jin-a xwe ra got-î-ye
     Vizier ADP wife-ez.f REFL ADP say.PST-PTCP-COP.3SG
     ‘The Vizier said to his wife’
     (Muş dialect of Northern Kurdish, central Anatolia)

3. Pre-predicate, addressee flagged with a postpositional particle

(24) mi xewn-ek dît-î-ye ez=ê
     1SG.OBL dream-INDF see.PST-PTCP-COP.3SG 1SG=FUT
     xewn-a xu te-ra bêj-im
     dream-EZ.FEM REFL 2SG.OBL-to say.PRS.SUBJ-1SG
     ‘I had a dream, I will tell you my dream’
     (Northern Kurdish of Armenia, Djelil and Djelil 1978: 294, also typical of Tunceli, cf. Haig 2006)

The same pattern is also found in Zazaki, Central Anatolia:

(25) šofird-ē-xo=rē vano
     driver-EZ-REFL=to say.PRS.3SG
     ‘(He) says to his driver’ (Paul 1998: 218)

This is essentially identical to the Turkic / Laz / Armenian pattern of pre-verbal addressees, marked with case suffixes or postpositions, illustrated for Laz in (26):

(26) tuti-s at’sumeski
     bear-DAT she.said
     ‘She said to the bear’ (Lacroix 2009: 846)

What we see in Northern Kurdish is a progressive shift from a post-verbal, prepositional type of addressee, to a pre-verbal, postpositional one, roughly following a north/south cline through Anatolia. The patterns are illustrated schematically in Table 5.
5. **East Anatolia as a linguistic area?**

Haig (2001) initially tackled the question of whether East Anatolia would qualify as a “linguistic area”, concluding that a yes/no answer was premature at that stage. However, since then a good deal more data has become available, so that a reassessment appears to be overdue. In fact, the main problem is not lack of data, but the notion of linguistic area itself. On the whole I share the view of Campbell (2006) that linguistic areas are not particularly meaningful entities; at best they are secondary inferences derived from an investigation of individual contact phenomena (see Haig 2014a). Nor is East Anatolia clearly delineated as a geographic region, and depending on the extension of the definition, rather different results emerge. For example, Haig (2001) and Matras (2009: 270) both include the Kartvelian language Laz in their notion of East Anatolian linguistic area, but as will be seen below, such a Pan-Anatolian approach in fact misses several important generalizations. In what follows, I will assess attempts to defend a Pan-Anatolian linguistic area, before presenting a more finely-grained alternative solution below.
5.1 The indicative/progressive prefix

Matras (2010) refers to an East Anatolian “linguistic area”, citing as one major diagnostic the presence of a “progressive-indicative aspectual prefix” (2010: 75) in a number of the languages of the region: Kurdish, Levantine Arabic, Persian and Armenian, see also Matras (2007: 45) and (2009: 260) for similar claims. The proposed template for the “present tense finite verb in languages of East Anatolia” (Matras 2009: 260) is provided in (27):

(27) ASPECT - PRESENT_STEM - PERSON_AGREEMENT
   e.g. Kurmanji di-bêj-im ‘I say/am saying’

The claim is potentially interesting because it is concerned with bound verbal morphology, on most accounts the grammatical domain least susceptible to contact influence. But precisely because of its intrinsic interest, it requires careful assessment (see Haig 2014a for details). The first problem with this feature – at least as a diagnostic for an Anatolian linguistic area - is the glaring absence of a progressive indicative prefix in a number of languages of the area, notably all varieties of Turkic, in Zazaki, in the Aleppo variety of Domari (Herin 2012), and in Laz.4 Turning to those languages which do exhibit such a prefix, they are either Semitic, or Iranian. With regard to the Iranian languages, a present indicative prefix is a feature of West Iranian generally, found in Iranian languages outside of Anatolia including Hawrami, Vafsi, Balochi, and the Central Plateau languages. The presence of this feature in Kurdish, then, is hardly the result of shared areal distribution, but simply of genetic inheritance.

Turning now to the Semitic languages, for the qalitu Arabic dialects of Kurdistan and Anatolia contact influence appears to be likely. They have innovated a present tense prefix ku- attached to the imperfect stem; similar formations are found in dialects throughout North Iraq, and it seems highly probable that Kurdish was influential in this development. For NENA, however, the facts are far from clear. Where such a prefix is found, it is not always just a general present indicative marker, but may contribute additional aspectual (e.g. progressive or habitual) nuances. For example, in the Hertevin dialect of NENA, the present tense generally lacks a prefix: me Ɂā miton ‘what say.you.PRS.INDIC’ ‘what do you think (say)?’ (Jastrow 1988: 54), but a prefixed form is available for future or habitual senses. Or a prefixed present tense indicative is available, but only for a restricted number of lexical verbs (e.g. vowel-initial stems in Bohtan, Fox 2009: 55), or it may be entirely absent, as in Ashitha Neo-Aramaic, spoken in the village of Lower Tiyare in southeastern Turkey on the Iraqi border, squarely within East Anatolia. All this is quite different to the Kurdish case, where the prefix is present (in various forms) in all dialects, but lacks any aspectual value whatsoever; it is simply the default indicative present marker.

In sum, it is quite probable that the verbal system of the Semitic languages of southeast Anatolia (Arabic dialects and NENA) was strongly influenced by neighbouring

---
4 Homshetsma, the sole surviving dialect of West Armenian in Anatolia, has an indicative prefix in the present tense, but only with vowel-initial verb stems (Vaux 2007: 263). East Armenian lacks an indicative prefix in the present entirely. Just where “West Armenian” should be geographically located is not a simple question, so assessing the specific areal relevance of the West Armenian data is difficult; Matras (2010) does not provide the source for West Armenian.
Iranian languages, most significantly by Kurdish. The parallels are deep, and have led to
developments in NENA that set it apart from all other Semitic languages. In particular,
the development of ergative alignment based on participial forms of past-tense verbs, and
the use of clitic pronouns to index the agent, are matched exactly by the developments in
Middle Persian and Parthian, and in today’s central Kurdish, and few scholars now doubt
that the development of ergative structures with past transitives in NENA was in part the
result of influence from Iranian (Khan 2007, Kapeliuk 2011, Noorlander 2014). Against
this backdrop, it would not be surprising if Iranian patterns of mood/aspect marking were
also replicated in NENA and local varieties of Arabic, yielding the shared template given
in (27). But viewed in this light, we have a straightforward case of contact influence of
one language (or group of languages), Kurdish, on neighbouring groups, NENA and
Arabic. But this is not a feature that has diffused in Anatolia as such, as other languages
of Anatolia (e.g. Zazaki and Turkish) are totally unaffected by it. Thus the relevance of
this feature as a diagnostic for an Anatolian linguistic area is questionable.

5.2 Finite subjunctive modal complements after WANT

Another feature mentioned by Matras (2009: 270) as characteristic of East Anatolia is
“final subjunctives in modal complements”. An example of the pattern concerned is as
follows, illustrated by Neo-Aramaic and Kurmanji (Noorlander 2014) with the
complement of WANT:

<table>
<thead>
<tr>
<th></th>
<th>PRESENT INDICATIVE</th>
<th>PRESENT SUBJUNCTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘I’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kurmanji</td>
<td>Ez-di-xważ-im</td>
<td>φ-her-im malê</td>
</tr>
<tr>
<td>Turoyo</td>
<td>Ńono-k-abš-ono</td>
<td>(d-)zz-î(on) l-d=baθo</td>
</tr>
</tbody>
</table>

Table 6. Patterns of WANT-complementization

Superficially, NENA and Kurmanji appear to show close parallels. But to what extent
the parallelism evident in Table 6 is contact-induced, and to what degree it diffuses
beyond the specific NENA/Kurdish contact situation across Anatolia as a whole needs to
be investigated.

Let us consider first the probable initial states of Kurmanji and NENA respectively by
investigating related languages outside the region. We know that the pattern shown in
Table 6 is also widespread across West Iranian, and is not restricted to East Anatolia. Its
presence in Kurmanji is thus predictable, regardless of any areal influence. Turning to
Semitic, we find that Semitic languages outside of Anatolia also exhibit a pattern of
WANT-complementization quite similar to Table 6, with a person-marked verb occurring

5  Note that the form given in Table 6 for ‘Kurmanji’ is actually not found in those Kurmanji
dialects spoken in the southeast of the Kurmanji region, in particular North Iraq (e.g. Zakho).
Here the translational equivalent of Table 6 has a non-canonical subject (Oblique) form for the
‘wanter’ (i.e. min dvêt bičime mal ‘to-me is desirous I-go home’). Thus the common template
suggested by Table 6 applies to a sub-set of Kurmanji dialects.
after the modal. Example (28) is from Gulf Arabic, see also Brustad (2000: 234) on Egyptian and Moroccan Arabic:

(28) ḥābib  arūh  is-sīinema  
I.want  I.go  DEF-cinema  
‘I like going to the cinema, I would like to go to the cinema’  
(Holes 1992: 231)

Thus although we can reasonably assume that Kurdish supplied the model for Table 6, NENA did not actually have to accommodate very drastically to arrive at the common form: from the Semitic predecessors of NENA we can assume that the linear order of WANT and its complement was already aligned with the Iranian model, and having person-marked verbs in the complement clause is also a feature widespread in Semitic. Another language of the region that shares this feature is Domari of Aleppo (Herin 2012: ex. 9). In other words, the common pattern shown in Table 6 is largely attributable to the genetic inheritances of Semitic and Iranian respectively, with contact influence achieving some fine-tuning, but little more.

However, the pattern found in Table 6 is also found in those varieties of Turkic under Iranian influence. It is generally assumed that historically in Turkic, complements of WANT are (i) non-finite, and (ii) placed before WANT. Thus standard Turkish for (28) is:

(29) [sinema-ya  git-mek]  isti-yor-um  
cinema-DAT  go-INF  want-PROG-1SG  
The complement of WANT is pre-posed, and non-finite. Turkic varieties under Iranian influence, however, develop postposed complement clauses containing a finite (optative) verb-form, as in the following example from Azeri Turkish from Tabriz, Iran:

(30) vä  män istird-im  [ğiğişir-am]  
and 1SG want.PST-1SG  scream.OPT-1SG  
‘and I wanted to scream’  
(Kural 2001: 82)

What we are seeing, then, is the spread of this pattern wherever Iranian influence is strong. For Semitic languages, such as NENA, synchronizing with the Iranian pattern does not involve major re-structuring, whereas for Turkic it does. On the whole, the facts can be rather simply accounted for in terms of Iranian influence on neighbouring languages. Thus we have a case of bilateral contact influence, rather than the result of multi-lingual, pan-Anatolian diffusion. In the north of Anatolia, for example, this pattern is not prevalent. In Laz, the complement of WANT is non-finite, and generally pre-posed:

(31) yahudi-s  [k’işi-şi oçk’omo]  unt’u  
Jew-DAT  bird-GEN  eat.NONFIN  want.PST.3SG  
‘The Jew wanted to eat the bird’  
(Lacroix 2009: 516)

East Armenian has non-finite modal complements, but like Iranian, they are post-verbal. This in fact looks like a compromise solution between the Iranian and Turkic strategies,
with Turkic providing the non-finite form of the complement, and Iranian the post-verbal position:

(32) \textit{Anuš-ē uz-um ē [ašxat-el gradaran-um]}
\textit{Anuš-DEF want-PTCP.PRES COP.3SG work-INF library-LOC}

‘Anuš wants to work in the library’

Finally, the remnant dialects of Romeyka, Greek dialects spoken near Trabzon, have actually \textbf{retained} non-finite complements of \textit{WANT} despite the general drift in Greek to abandon them in favour of finite, subjunctive complements. In short, the shared pattern of subjunctive modal complements largely reduces to a matter of Iranian influence on neighbouring languages, with quite a clear directionality, and can be more cogently analysed in those terms, rather than in terms of a multi-directional diffusion.

5.3 An alternative solution: two spheres, and an intermediate zone

From the preceding it follows that there are good grounds to take a more differentiated view of Anatolia. Essentially, we can identify two areas of denser contact, with an overlap zone between them:

5.3.1 The Mesopotamian region

This centres on North Iraq and Turkey south of Lake Van, and is historically dominated by Central and Northern Kurdish, interspersed by a series of NENA and Arabic-speaking speech communities. Across this region we find massive parallels at all levels: in the lexicon, for example, Kahn (2007: 209) gives the following figures for Kurdish borrowings in the lexicon of the Sulemaniyya NENA of North Iraq: nouns: 67%, adjectives: 48%, particles: 53%, verbs: 15%. We have already noted the parallels in the vowel system (cf. 3.1 above), shared word order patterns (OVG), shared morphology (the aspectual suffix –\textit{awa} (and phonetic variants), not discussed here), the remarkable tolerance of complex syllable onsets in Central Kurdish, and the southeastern dialects of Northern Kurdish (not discussed here), as well as similar complementation strategies, and a shared present-tense morphological template, to name but a few.

5.3.2 Caspian/Caucasian region

Northeast Anatolia, north of Lake Van, dominated by Turkic, dialects of East Armenian, and dialects of Northern Kurdish. Extending westward along the Black Sea coast, this region includes languages such as Laz and Romeyka. The languages of this region exhibit diametrically opposed values on a number of features when compared to those of the Mesopotamian region.

Between these two regions there is a large area that shows affinities with both. The languages spoken here include the central Anatolian and northwestern varieties of Kurmanji Kurdish (generally referred to as Serhad-dialects), Zazaki, dialects of Turkish, and various varieties of Armenian.
An impression of the differences between the Mesopotamian and the Caucasian/Caspian spheres emerges when we plot a selection of morphosyntactic features across the languages, as shown in Table 4.

<table>
<thead>
<tr>
<th></th>
<th>CK</th>
<th>Ar.</th>
<th>NENA</th>
<th>Dom</th>
<th>NK</th>
<th>Za</th>
<th>Tu</th>
<th>Hom</th>
<th>Laz</th>
<th>EA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>*?</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-/+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>*?</td>
<td>+/-</td>
<td>+</td>
<td>(+) (+) (+) (+)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>*?</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+/</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+/</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>+</td>
<td>+/-</td>
<td>*?</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(+)</td>
</tr>
<tr>
<td>9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+/-</td>
<td>?</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 7. Morphosyntactic features in languages of Anatolia. Abbreviations: Dom=Domari of Aleppo, EA=Eastern Armenian, Hom=Homshetsma. +/- varies according to local varieties; (+) not the unmarked option, but available as a pragmatically marked variant in all varieties; *? feature attested, but insufficient data on frequency of occurrence; ? source unavailable

Features listed in Table 7:

1. Local relations (in, at, from etc.) expressed through prepositions
2. Same set of clitic pronouns attaching both to verbs as object indices, and to nouns as possessors
3. Noun - Possessor, Noun - Adjective word order in the NP
4. Post-predicate recipients of GIVE
5. Lack of postpositional marking of addressees of verbs of speech
6. Indicative/aspectual prefix on present tense verb forms
7. Grammatical gender
8. VO word order
9. Finite complement of WANT

It will be seen that NENA, Central Kurdish, and Arabic line up closely (barring two features of Central Kurdish), and that Turkish, Eastern Aramaic, Laz and Homshetsma likewise exhibit close similarities. But assigning Northern Kurdish, Domari of Aleppo, or Zazaki to either type is problematic. For Northern Kurdish, the problem is simply the size and geographical spread of the speech community; the southeastern dialects align with the languages of Mesopotamia, the northern dialects show greater affinities with the Caspian/Caucasian languages. The position of Domari is rather special, and reflects both the Indo-Aryan origins of the language as well as traces of the speakers’ movements prior to their current location.

The data thus generally support the assumption of two areal epicentres, with an intermediate zone, rather than a pan-Anatolian linguistic area. Nevertheless, a few candidates remain that could be considered shared features of a pan-Anatolian region. I offer the following list, taken from morphosyntax, to which a number of shared discourse
markers and (probably) the obligation particle *lazim* (in various phonetic variants, originally from Arabic) could be added:

i. the enclitic recalled-topic marker (cf. section 3.2.3 above)

ii. an obligatory clause-final copula (cf. section 3.2.2 above)

iii. ‘either-or’ constructions based on *ya(n)* ... *ya(n)* (Haig 2001)

iv. a grammaticalized indefinite article, accompanying indefinite singular NPs (not investigated here, but potentially of some interest)

v. echoic reduplication (Haig 2001, Matras 2009)

vi. the use of a general complementizer *ki* (with variant vowel values, Haig 2001)

Table 8. Possible shared features of a pan-Anatolian region

Whether the above features constitute grounds for assuming a linguistic area depends on how one defines “Anatolia” (many of the above-mentioned features are geographically distributed far beyond the conventional boundaries of Anatolia, for example i. and v.), and the concept of “linguistic area” itself. One could obviously tweak both definitions until they match the data, but that does not seem a particularly meaningful endeavour. What can be said with some degree of certainty is that two sub-regions of conspicuous structural parallels can be identified, the southeastern Kurdish/NENA/Arabic “Mesopotamian” region, and the northern Turkic/Kartvelian/Armenian “Caspian/Caucasian” region. Much of Anatolia lies between the two, and disentangling the respective contact influences is correspondingly difficult.

6 Conclusions

East Anatolia lies at the overlap of several major macro-linguistic areas. It is thus neither a spread zone, nor a residual zone, but a transition zone. Its linguistic diversity results from the overlay from its neighbouring areas, rather than through gradual accumulation of indigenous variation. Because different parts of the region show affinities to different neighbouring regions, it is extremely difficult to identify significant shared features covering the entire region. I suggest that it is more meaningful to identify the sub-regions, rather than simply assume the existence of an Anatolian linguistic area. The most prominent sub-region is in the southeast (south of Lake Van into North Iraq), which I refer to as the Mesopotamian region; most of the shared features here can be relatively straightforwardly attributed to Kurdish influence on local Semitic languages (NENA and Arabic), but Kurdish itself has also been affected (pharyngeal consonants, medial copula in some varieties, the greater tolerance of complex syllables in Central Kurdish when compared to the northernmost dialects of Northern Kurdish). Languages of the northern part of Anatolia are only marginally implicated, and in fact show quite different values on a number of features. I refer to them as the Caspian/Caucasian region; here, it seems that Turkish is the dominant language, and much of what can be found in languages like Laz can reasonably be attributed to Turkic influence.

A small number of pan-Anatolian features can nevertheless be identified, but their significance is not fully clear; they are either typologically commonplace, hence possibly coincidental, or represent inroads of larger shared areal patterns, extending well beyond
the geographical boundaries of East Anatolia. On balance, the assumption of an
Anatolian linguistic area now appears doubtful, though this in no way detracts from the
overall interest of the region as characterized by multiple contexts of long-standing
language contact among genetically unrelated languages, which has left its impact on all
levels of language structure.

References

Blau, Joyce, 1975. Le Kurde de ’Amâdiya et de Djabal Sindjâr : Analyse Linguistique,

Cambridge: University of Cambridge.

Brustad, Kristen. 2000. The Syntax of Spoken Arabic. A Comparative Study of Moroccan,
Press.

Bulut, Christiane. 2002. Evliya Çelebi as a linguist and dialectologist: seventeenth
Century East Anatolian and Azeri Turkic dialects. In: Nuran Tezcan and Kadir
Atlansoy (eds.), Evliya Çelebi ve Seyahatname. Doğu Akdeniz Üniversitesi
Yayınları, 49–64.

Bulut, Christiane. 2006. Turkish elements in spoken Kurmanji. In: Hans Boescheton and
Lars Johanson (eds.), Turkic Languages in Contact. Proceedings of the Wassenaar

Bulut, Christiane. 2007. Iraqi Turkman. In: John Postgate (ed.). Languages of Iraq,
Ancient and Modern, 159-187. Cambridge: British School of Archaeology in Iraq.

Campbell, Lyle 2006. Areal linguistics: A closer scrutiny. In: Yaron Matras, April
McMahon and Nigel Vincent (eds.). Linguistic Areas: Convergence in Historical

Tradition of the Kurds]


Press.

Ankara: Türk Dil Kurumu. [The Spoken Varieties (of Turkish) in Erzurum:
Analysis, Texts, Lexicon, Glossaries]

107.

Routledge

Stop Contrasts in Kurmanji Kurdish—Testing the Ejectives Hypothesis. Paper Held
at the Fifth International Conference on Iranian Linguistics (ICILS), University of
Bamberg, August 2013.

Haig, Geoffrey. 2001. Linguistic diffusion in present-day east Anatolia: From top to
bottom. In: Robert Dixon and Alexandra Aikhenvald (eds.). Areal Diffusion and


Kahn, Margaret. 1976. Borrowing and Regional Variation in a Phonological Description of Kurdish. Ann Arbor, Michigan: Phonetics Laboratory of the University of Michigan.


**Index (Part 1)**

Anatolia, East, 343  
geography, 343  
Anatolia, East  
as a transitional zone, 343  
language varieties, 345  
morphosyntax  
clitic additive, 353  
comparatives, 349  
copular constructions, 350  
Arabic dialects, 351  
Iranian languages, 350  
NENA dialects, 352  
marking addressee of verbs of speech, 358  
word order, 354  
direct objects and goals, 356  
Semitic languages, 355  
shift from VO to OV, 355  
North Eastern Neo-Aramaic (NENA), 343  
phonology, 347  
Kurdish, 348  
Semitic languages, 349  
voice onset time, 348  
vowel systems, 347  
sociolinguistic situation, 344  
spread vs. residual zone, 344  
status as linguistic area, 358  
Caspian/Caucasian region, 362  
evaluation, 364  
finitive subjunctive modal complements, 360  
indicative/progressive prefix, 359  
Mesopotamian region, 362  
pan-Anatolian features, 364  
Areas and universals, 36  
areal signals in Eurasia, 42  
assessment, 44  
frequencies of features, 37  
methods of analysis, 39  
predictive areality theories, 39  
generalised linear models, 40  
role of universal forces, 36  
triggers of language change, 37  
event-based, 38  

functional, 38  
Balkans  
genealogical groups, 58  
historical documentation, 58  
infinitive, 66  
Britain and Ireland  
areality, 229  
caveats, 230  
Celtic, 245  
aspect, 247  
category and exponence, 250  
habitual, 249  
perfective  
immediate, 248  
resultative, 249  
Brythonic, 237, 243  
Celtic hypothesis, 237  
cleft sentences, 246  
early contact  
assessment, 243  
extended ‘now’, 247  
later contact  
assessment, 251  
overuse of definite article, 246  
subordinating ‘and’, 246  
historical dimension, 237  
dental fricatives, 242  
development of possessive, 240  
internal possessor construction, 239  
twofold paradigm of ‘to be’, 239  
Northern Subject Rule, 237  
rise of periphrastic ‘do’, 241  

English  
diagnostic value of features, 233  
historical sociolinguistics, 231  
remnants of former distributions, 232  
NURSE-TERM distinction, 232  
SERVE-Lowering, 233  
vowel epenthesis, 233  
textual record, 231  
geography and political divisions, 229  
language change  
constraint hierarchies, 236  
degrees of manifestation, 235  
internal motivation, 234  
issues, 234
lenition, 235
low-contact areas, 244
Caucasus, 307
contact scenarios, 310
defining Caucasus as a linguistic area, 307
divergence and convergence, 309
historic settlement, 310
historical sociolinguistic situation, 315
isolated communities, 313
languages of the wider Caucasus, 308
Northeast-Caucasian (NEC), 308
Northwest-Caucasian (NWC), 308
South-Caucasian (SC), 308
multilingualism, 313
phonology
areal typology, 315
assessment, 335
consonant clusters, 329
inventory sizes, 317, 318
latterals, 328
nasalisation, 327
palatalisation, 329
pharyngealisation, 328
phonation contrasts, 320
ejectives, 321
phonetic realisation, 331
place of articulation, 323
labialisation, 325
pharyngeals, 325
uvulars, 324
pockets of areality, 327
sibilants, 328
sound frequencies, 330
structure and occurrence of features, 317
voice onset time, 333
vowel inventories, 319
commonalities and overlapping, 320
physical geography, 314
population genetics and migrations, 312
reasons for structural similarities, 310
vertical archipelago phenomenon, 313
English, varieties of, 260
areal patterns, 262
eWAVE
Africa, 260
America, 261
Australia, 261
British Isles, 260
Canada
Newfoundland, 263
Caribbean, 261
sub-varieties, 262
Pacific, 261
South and Southeast Asia, 261
South Atlantic, 261
varieties contained, 260
high/low-contact varieties, 262
morphosyntax
areality, 266
assessment, 282
diagnostic features, 265, 266
adverbial subordination, 274
adverbs and prepositions, 274
agreement, 272
complementation, 273
counterfactuals, 270
diagnostic areal features, 275
discourse and word order, 275
modal verbs, 270
negation, 272
noun phrase, 268
pronouns, 267
relativisation, 273
tense and aspect, 269
verb morphology, 271
distribution by region, 278
smaller-scale areal patterns, 280
strength of the geographical signal, 280
synopsis, 276
domains of grammar, 264
Pidgins and Creoles, 262
typology, 262
NeighborNet diagram, 264
ERIC loans (‘Essentially Rooted In Conversation’), 59
EUROTYP, 62, 106
Germanic languages
areality, 202
Afrikaans, 210
assessment, 218
Circum-Baltic area, 208
concepts, dimensions and restrictions, 202
contact with Finno-Ugric, 212
contact with Romance languages, 212
contact with Semitic, 211
Yiddish, 211
English and North Germanic, 207
English-Irish contact, 210
French-Frankish contact, 214
Frisian, 208
contacts with Dutch, 209
German contact with Romanian, 215
Germanic contact with Balto-Slavic, 216
Baltendeutsch, 216
Halbdeutsch, 216
Russian German, 216
Germanic contact with Celtic, 215
do periphrasis, 215
Germanic contact with Indo-Aryan, 217
Angloromani, 217
Stadsfries, 217
Germanic contact with other languages, 218
Ingvaenic features, 204
internal considerations, 205
Low German in Norwegian, 206
Luxembourghish, 214
mixed languages, 205
non-Germanic contacts, 210
Old Norse, 206
Romance character of English, 213
Scandinavian Low German, 207
status of languages, 204
Viking contact with English, 205
West Germanic contacts, 208
linguistic areas
Balkans, 109
borrowing vs. shift, 203
British Isles, 203
cause-based questions, 60
Chaco region, 18
active-stative verb alignment, 27
assessment, 29
directional verbal affixes, 28
further traits, 28
gender, 27
genitive classifiers, 27
lack of verbal tense, 28
set of demonstratives, 27
SVO word order, 26
Charlemagne sprachbund, 62, 64, 117, 214
Circum-Baltic, 203
contact superposition zone, 203
convergence areas, 58
defining concept, 22
defining criteria, 22
number of language families, 23
shared traits, 22
trait bundling, 24
trait sprawl area (TSA), 25
trait weight, 23
definition, 18, 24, 135, 203
delineational and assessment issues, 65
different types, 20
earlier scholarship, 107
feature-based issues, 56
functional typology, 107
Gran Chaco region, 26
have perfects, 63
identification, 106
interrogating the ‘sprachbund’, 51
language-based issues, 52
language family, 64
levels of language, 205
linguistic area sensu stricto
(LASS)definition, 25
linguistic enclaves, 203
northern Europe and the Baltic, 109
origin of term, 50
Phonologischer Sprachbund, 108
proposed areas, 18
Balkans, 20
Baltic, 20
Mesoamerica, 20
North-West Coast of North America, 20
South Asia, 20
reassessment, 48, 69
role of contact, 55
Standard Average European, 203
standards and literary traditions, 69
superstrate, substrate, adstrate, 203
Linguistic areas, 18
metatypy, 59, 176
phonology
areal sound patterns, 77
areal sound pattern hypothesis, 86
assessment, 100
contrastive retroflexion, 80
definition, 77
ejective consonants, 88
Caucasus, 89
Pacific Northwest, 89
evolution, 84
final obstruent devoicing, 83
front rounded vowels
Europe, 93
glottalised consonants, 81
perceptual magnet effect, 86
phonetic proximity, 87
phonetic saliency, 87
significant exposure, 87
Index -- page 393 of 393

preaspiration
Northern Europe, 94
radical areal sound patterns, 97
retroflex consonants
South India, 90
shared inheritance, 79
stone soup metaphor, 85
tone as an areal feature, 96
South East Asia, 96
world-wide distribution, 98
areal typology of Europe, 108, 110
airstream mechanisms, 111
contact phonology, 128
evaluation, 127
lateralis, 114
manner of articulation, 112
phonation, 117
phylum-internal diversity, 125
place of articulation, 114
primary articulations, 123
rhotics, 113
secondary articulation, 120
genetic affiliations, 121
convergence and divergence, 106
phonological classes, 111
word prominence, 135
assessment, 164
Australia
Arnhem Land, 155
deriving patterns, 156
distribution, 161
initial and penultimate stress, 155
innovation or remnant, 162
signs of instability, 159
convergence, 141
North American areas, 147
accent patterns, 153
Clear Lake Linguistic Area, 148
Northwestern California, 150
Yuman-Takic, 152
dimensions, 136
divergence
Basque, 144
hybridity, 143
loan phonology, 138
theory of change, 139
world-wide distribution, 136
semantics
areal patterns, 173
area-specific lexicalisations, 180
calques, 173
lexico-constructional calquing, 175
lexico-semantic parallels as areality indicators, 176
polysemy calques, 174
causation and mechanisms, 189
conventionalised direction 'where'-greetings, 192
evaluation, 193
shared formulaic expressions, 178
The Greater Hindukush calendrical expressions, 186
case study, 182
cos-lexicalised intensifiers (Palula), 188
numerals, 186
pro-forms, 187
catalogue of semantic shifts in the languages of the world, 194
Database of Cross-Linguistic Colexifications, 194
lexical typology, 194
Slavic languages, 285
eastern Baltic region, I, 289
language contact
Fennic, 289
eastern Baltic region, II, 291
nominating patient with infinitive, 291
eastern Baltic region, III, 294
impersonal passives, 294
Eurasian linguistic area, 300
palatal/non-palatal contrast, 300
northern frontier, 287
language contact
German during the Hanse period, 287
German in the Baltic region, 288
Slavic linguistic areas
evaluation, 302
southeastern frontier, 285
borrowed lexis, 286
southwestern Slavdom, I, 296
contact in the Balkans, 296
pleonastic structures, 296
southwestern Slavdom, II, 299
lingua franca, 300
western frontier, 286
language contact, 286
German into Czech, 286
Standard Average European, 207, 214
World Atlas of Linguistic Structures (WALS), 39, 42, 63, 79, 82, 87, 92, 108