

Thomas Owen-Smith, Nathan W. Hill (Eds.)
Trans-Himalayan Linguistics

Trends in Linguistics Studies and Monographs

Editor

Volker Gast

Editorial Board

Walter Bisang

Jan Terje Faarlund

Hans Henrich Hock

Natalia Levshina

Heiko Narrog

Matthias Schlesewsky

Amir Zeldes

Niina Ning Zhang

Editor responsible for this volume

Walter Bisang

Volume 266

Trans-Himalayan Linguistics

edited by
Thomas Owen-Smith and Nathan W. Hill

DE GRUYTER

ISBN 978-3-11-031074-0
e-ISBN 978-3-11-031083-2
ISSN 1861-4302

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

© 2014 Walter de Gruyter GmbH & Co. KG, Berlin/Boston
Typesetting: PTP-Berlin Protago-TEX-Production GmbH, Berlin
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
♻️ Printed on acid-free paper
Printed in Germany

www.degruyter.com

List of contributors

Roger Blench

Kay Williamson Educational
Foundation
8 Guest Road, Cambridge, CB1 2AL, UK
rogerblench@yahoo.co.uk

Tim Bodt

Institut für Sprachwissenschaft
Universität Bern, Switzerland
timotheus.bodt@isw.unibe.ch

Ilija Časule

Department of Linguistics
Macquarie University, Sydney,
Australia
ilija.casule@mq.edu.au

Scott DeLancey

Department of Linguistics
University of Oregon, USA
delancey@uoregon.edu

George van Driem

Institut für Sprachwissenschaft
Universität Bern, Switzerland
vandriem@isw.unibe.ch

Nathan Hill

Department of Linguistics
School of Oriental and African Studies
(SOAS),
University of London, UK
nh36@soas.ac.uk

Isao Honda

Department of English and
Communication
Nagoya College, Japan
honda@nagoyacollege.ac.jp

Christian Huber

Phonogrammarchiv
Austrian Academy of Sciences
Vienna, Austria
christian.huber@oeaw.ac.at

Gwendolyn Hyslop

ANU College of Asia and the Pacific,
Australia
gwendolyn.hyslop@anu.edu.au

Alexis Michaud

International Research Institute MICA
(HUST-CNRS-Grenoble INP), Hanoi,
Vietnam,
and LACITO-CNRS, Paris, France
alexis.michaud@mica.edu.vn

Jean Robert Opgenort

www.opgenort.nl
jeanrobert@opgenort.nl

Thomas Owen-Smith

Department of Linguistics
School of Oriental and African Studies
(SOAS),
University of London, UK
thomasowensmith@gmail.com

David Peterson

Program in Linguistics and Cognitive
Science
Dartmouth College, USA
david.a.peterson@dartmouth.edu

Nicolas Tournadre

Université de Provence, and CNRS
(LACITO), France
nicolas.tournadre@univ-amu.fr

Mark W. Post

Institut für Sprachwissenschaft
Universität Bern, Switzerland
markwpost@gmail.com

Content

Thomas Owen-Smith and Nathan W. Hill

Introduction — 1

George van Driem

Trans-Himalayan — 11

Scott DeLancey

Creolization in the Divergence of the Tibeto-Burman Languages — 41

Roger Blench and Mark W. Post

**Rethinking Sino-Tibetan phylogeny
from the perspective of North East Indian languages — 71**

Nicolas Tournadre

The Tibetic languages and their classification — 105

Isao Honda

Internal diversity in the Tamangic lexicon — 131

Gwendolyn Hyslop

A preliminary reconstruction of East Bodish — 155

Ilija Čašule

Burushaski kinship terminology of Indo-European origin — 181

Christian Huber

Subject and object agreement in Shumcho — 221

Alexis Michaud

**The tone patterns of numeral-plus-classifier phrases in Yongning Na:
a synchronic description and analysis — 275**

David A. Peterson

Rengmitca: the most endangered Kuki-Chin language of Bangladesh — 313

Jean Robert Opgenort

Initial Grammatical Sketch of Tilung — 329

Tim Bodt

Tshangla Phonology and a Standard Tshangla Orthography — 393

Index — 437

Thomas Owen-Smith and Nathan W. Hill

Introduction

The Himalayan Region is home to around 600 languages (Turin 2007: foreword), making it one of the most linguistically diverse regions of the world.¹ But it is not only this diversity which constitutes the region's great importance for linguistics. The typological profiles of Himalayan languages are also striking, with many displaying features which are rare in Eurasia and in the Old World in general. A large number of language families are present in the region, and complex histories of ethnic mixing and language contact have created a contemporary linguistic landscape which is as challenging as it is exhilarating to research.

The mountains, though not an absolute barrier, hinder transport and communications. Even today travel in the hills of Nepal or Northeast India is more difficult than on the Indo-Gangetic plain, or for that matter the Tibetan Plateau (see footnote 2). This rugged topography has throughout history weakened the assimilating and centralizing tendencies of large states, and allowed for the existence of small, isolated communities living with a high degree of economic self-sufficiency and political autonomy. Such a pattern is reflected in other mountains regions of the world (Braudel [1949] 1976: 30–43; Gellner 1969: 1–2; Scott 2001: 13–22 *inter alia*), and not surprisingly fosters linguistic and cultural diversity.

Bàrtoli points out that mountainous areas tend to preserve archaic forms of speech:

nella storia del linguaggio...le isole sono di norma più conservative che i continenti..., e più le montagne che le pianure e le marine..., e più certe aree laterali che le aree di mezzo..., e più i centri minori che i maggiori

[in the history of language...islands are generally more conservative than continents..., mountains more so than plains or the seaside..., certain marginal areas more so than central areas..., and minor centres more so than major ones] (Bàrtoli 1925: 4)

Historical linguistics handbooks refer to these isolated enclaves of linguistic conservatism as “relic areas” (Campbell 2004: 216; Hock 1991: 440; Dimmendaal 2011: 159). Nichols uses the term “accretion zone,” and notes that rather than

¹ Our definition of the Himalayan Region is a wide one, including the Himalaya proper, the Tibetan Plateau, and the various mountain ranges which flare off from these in all directions, including the Hindu Kush, Pamir and Tian Shan, the highlands of Sichuan and Yunnan, and the hilly tracts of Northeast India, Bangladesh and Burma.

simply conservatism in such areas, the influx of languages from other places also affects their character:

An accretion zone...is an area where genetic and structural diversity of languages are high and increase over time through immigration. Examples are the Caucasus, the Himalayas, the Ethiopian highlands and the northern Rift Valley, California, the Pacific Northwest of North America, Amazonia, northern Australia, and of course New Guinea... Accretion zones generally contain representatives of major stocks in the vicinity as well as some languages with no outside kin. (Nichols 1997: 369)

Mountains differ from lowlands in that lowland areas have historically been prone to large-scale spreads of particular languages or groups of languages (on a family or group scale examples include Iranian, Turkic, Indo-Aryan, Sinitic; on a language scale examples include Mandarin Chinese, Persian, Hindi-Urdu) which submerge earlier languages spoken in the same space; while when new languages move into the mountains they tend to add to the linguistic diversity of the region rather than obliterate it (see Nichols 1992).²

Linguistic diversity is manifested in the Himalayan Region in a number of ways. Firstly there is the density of discrete languages spoken there. There is then the diversity of linguistic stocks represented, which totals roughly nine depending on where the region's boundary is drawn: Sino-Tibetan (or "Trans-Himalayan" – see below), Indo-European, Austroasiatic, Hmong-Mien, Tai, Turkic and perhaps Mongolic, as well as the presumed isolates Burushaski and Kusunda.

Within several of these stocks, multiple branches are represented in the Himalayan Region. For instance, in Indo-European all three subbranches of Indo-Iranian are present: both Eastern (Pashto, the Pamir languages) and Western Iranian languages (Dari, Tajik, Parachi,Ormuri) are spoken, as well as several branches of Indo-Aryan,³ and the Nuristani languages, which are proposed to stand at the same phylogenetic level as Iranian and Indo-Aryan (see Nelson 1986). Worthy of special mention is the heterogeneous Dardic group spoken at the far Northwestern reach of the Himalaya. The Dardic languages are generally considered Indo-Aryan but have not participated in a number of developments

² By this definition, the Tibetan Plateau may share more features with lowland areas and prototypical spread zones: although situated at a high altitude it is relatively flat and easy to traverse, and has been dominated throughout recorded history by the spread of a single language, Old Tibetan. However, as Zeisler (2005) and Tournadre (this volume) point out, the degree of diversity between the modern varieties collectively referred to as "Tibetan" is more akin to that of a family of languages than dialects of one language. The languages of the Tibetan Plateau also display some of the unusual linguistic features which appear to delineate the Himalayan Region as a typological enclave.

³ The details of subgrouping within Indo-Aryan itself are disputed (see Masica 1991: 446–463).

which are general throughout the rest of Indo-Aryan, preserving some extremely conservative features while innovating others which are quite untypical of Indo-Aryan and Indo-European in general (see Bashir 2003).⁴

The Himalaya's diversity is even better exemplified by the profusion of Sino-Tibetan/Trans-Himalayan language groups spoken on both sides of the range. In contrast to the Indo-Iranian languages, which moved into the area from elsewhere (originally from Central Asia and later from the Indo-Gangetic plain), the preponderance of high level Sino-Tibetan/Trans-Himalayan taxa within the Himalayan Region may indicate that it has always been the centre of gravity of this family (see van Driem this volume; Blench and Post this volume).

Aside from their phylogenetic diversity, Himalayan languages display typological features rarely encountered in the Old World. From a survey of some 350 languages from around the world, Bickel and Nichols (2003) propose that the Himalaya and the Caucasus can be considered “typological enclaves,” due to the high frequency in these two regions (both mountainous relic areas/accretion zones) of certain features which are otherwise rare in Eurasia, and generally in most parts of the world. The unusual features which characterise the Himalayan enclave in particular are very high levels of morphological synthesis, polypersonal agreement on verbs,⁵ complex systems of evidentiality,⁶ bipartite verb stems, radical double marking (including agreement for more than one argument on the predicate and as well as case marking on the arguments themselves, and marking of possession on both the possessor and possessee), and multiple classes for possessive marking in the noun phrase.

The multitude of languages in the Himalayan region and their typological diversity therefore more than warrant the steady stream of research on these languages, of which the current volume presents a recent selection. The volume arises from papers given and topics discussed at the 16th Himalayan Languages Symposium, held in September 2010 at the School of Oriental and African Studies, University of London; and it includes chapters on historical linguistics, and new descriptive work and original data on under-researched Himalayan languages.

The majority of papers in the volume deal primarily with Sino-Tibetan/Trans-Himalayan languages. Since the earliest days of research on languages of this

⁴ Strand (1973) and Bashir (2003) indicate that the group “Dardic” is heuristic rather than genetic, arguing that the Dardic languages are highly diverse, and do not display shared innovations which delineate them as a subgroup.

⁵ Defined as “unconditionally obligatory verb agreement with more than one argument” in Bickel and Nichols (2003: 10).

⁶ These include patterns which have been analysed as conjunct/disjunct in some languages (see e.g. Hale 1980; Watters 2006a), although questions have also been raised about the applicability of the terms “conjunct” and “disjunct” to certain languages (see e.g. Tournadre 2008).

family, different classifications have vied for adherents. Numerous scholars have proposed classification and subgrouping schemes, including the early work of Hodgson ([1828] 1972, [1848] 1972) and Konow (1909), and more elaborate proposals from the mid-20th century onwards (Shafer 1955; DeLancey 1987; Bradley 1997, van Driem 2001 *inter alia*). Few of these taxonomic proposals have weathered the years free of controversy. The disputed names for the family, Indo-Chinese, Sino-Tibetan, Tibeto-Burman, Trans-Himalayan – all of which appear at some point in this volume – are one indicator of the controversial status of the *Stammbaum*.

As van Driem relates in the present volume, “Indo-Chinese” originally grouped together all languages of Asia and Oceania. Over the decades various stocks were respectively distinguished and then removed from this presumed genetic unity. The term “Sino-Tibetan,” which has a close association with the University of California at Berkeley, inherited its reference from “Indo-Chinese.” Although “Sino-Tibetan” has varied in meaning, it is now normally understood to reflect a family with two primary branches, “Sinitic” and “Tibeto-Burman,” i.e. a binary split at the highest level between Chinese and everything else (see Matisoff 2003: 5–6). But as no decisive evidence has been put forward to suggest that all non-Sinitic languages of the family share a common innovation (or indeed that Sinitic languages share a common innovation to set them apart from all the rest of the family), some scholars are now searching for terms which reflect an agnostic approach to subgrouping.

Although von Klaproth coined “Tibeto-Burman” in 1823 to serve as a phylogenetically neutral term for the family of which Chinese, Tibetan, and Burmese are members, the now common use of the same term for all non-Sinitic members of the family makes it difficult to use it with Klaproth’s original designation (i.e. the whole family) without causing confusion. Consequently, other terms have been proposed, including “Sino-Tibeto-Burman” (see Tournadre this volume; Matisoff 2011: i), and “Trans-Himalayan” (see van Driem 2007, this volume). The term “Sino-Tibeto-Burman” acknowledges the three most prominent literate civilizations, although as Blench and Post (this volume) point out, most languages in the family do not have literate histories, and there is no *a priori* reason to think that these three languages are of any special significance in the structure of the family or in the reconstruction of its ancestor. The term “Trans-Himalayan” has the advantage of being more geographically neutral and less culturally biased, and as a geographic term it parallels the names of other language families such as Indo-European and Austroasiatic. To instantiate the advantages of this neutral geographic term, *Trans-Himalayan* is an apt title for this volume.

The typological divergence of Trans-Himalayan languages leads not only to disagreements over nomenclature and subgrouping, but also to differences of opinion about the morphosyntactic profile to be reconstructed for the proto-

language. Whereas Rgyalrongic and Kiranti languages show a high degree of morphological complexity (e.g. synthetic verb inflection including agreement for more than one argument, double marking of possession on possessor and possessee), Sinitic and Lolo-Burmese languages are more analytic in structure (e.g. zero agreement on verbs and simple juxtaposition of nominal elements in a possessive clause in some languages). Some scholars, most prominently LaPolla (1989, 1992, 1994), argue that verbal agreement systems are independent innovations in various branches of the family, arising from the grammaticalization of erstwhile independent pronominal forms. In contrast, van Driem (1993), DeLancey (1989, 2010a) and Jacques (2012) propose a highly synthetic proto-language which has experienced varying degrees of phonological erosion in different subgroups. Fully accounting for the extremely divergent typologies of languages across the family naturally presents difficulties for both sides of this debate. In the current volume DeLancey proposes that the simpler typology of the analytic languages can be explained by the fact that these languages underwent a process of creolization when they emerged as *lingua francas*. This proposal could account for the correlation between analytic typology and the settings of urbanization and state centralization associated with languages such as Chinese, Burmese, Tibetan, Newar, and Meithei.

Work on a *Stammbaum* for Trans-Himalayan also stands in a certain tension to what appear to be extremely complex histories of migration, contact, acculturation, and ethnic and linguistic shift (see e.g. Thomason and Kaufmann 1988; Hickey 2010) across the regions where Trans-Himalayan languages are spoken. The survival of remnant language isolates like Kusunda (see Watters 2006b) reminds us that there may once have been many more linguistic stocks in the Himalayan Region which were ultimately submerged by acculturation to the larger families. Various authors have argued that large areas where Trans-Himalayan languages are now spoken are the result of intrusive migrations, or acculturation and shift from now lost languages to Trans-Himalayan forms of speech (e.g. Grierson 1909; LaPolla 2001, 2009; Zeisler 2005, 2009; DeLancey 2010b, 2011 *inter alia*), but it has not yet been possible to incorporate these observations systematically into the theoretical models about the linguistic history of the area. In this volume Blench and Post consider subsistence and environmental vocabulary from languages of Northeast India, and reiterate their position earlier stated in Post and Blench (2011) that the high levels of divergence amongst the languages of Arunachal Pradesh and surrounding areas have not been taken into account in any of the classification schemes for Trans-Himalayan. They further argue that some of these languages are so aberrant that there is a distinct possibility they represent the outcome of a situation in which speakers of unknown language stocks or isolates acquired a Trans-Himalayan lexicon through varying degrees of contact and acculturation.

Terminological and subgrouping controversies also beset the Trans-Himalayan subbranches. Drawing from the Tibetan autonym *Bod*, Shafer posits four “Bodish” subgroups for languages spoken in the region of the Tibetan Plateau: West, Central, Southern, and East Bodish (1966: 78–123), of which he derives only Central Bodish and Southern Bodish from Old Tibetan (1966: 87). Although some researchers still see West Bodish as a meaningful grouping (see Biellmeier 2004), it is becoming increasingly clear that all of these forms of speech except for the East Bodish languages (see Hyslop this volume) derive directly from Old Tibetan (see Hill 2010). Many authors refer to the languages descending from Old Tibetan as “Tibetan dialects” (e.g. Denwood 1999: 21–36), but Tournadre (this volume) presents a forceful argument that these languages should be called “Tibetic” languages to acknowledge the degree of their divergence from one another.

The key to a successful analysis of subgrouping is the identification of shared innovations. As LaPolla (2001: 245) notes, many of the proposed subgrouping schemes for Trans-Himalayan do not give the reasons for their groupings. In this process, the distinction of inherited features from borrowed features is paramount. Three papers in the current volume aim to sharpen the distinction between inherited and borrowed vocabulary in the languages they study. Honda examines patterns of lexical diversity in Tamangic languages, looking to differentiate inherited Tamangic words from later borrowings from Tibetan. Hyslop attempts a preliminary comparative reconstruction of the East Bodish languages. With this reconstruction in hand she is able to show that several grammatical morphemes which Kurtöp shares with Tibetan are not cognate, as might first appear to be the case, but as they are not reconstructible to proto-East Bodish they must be borrowings from Tibetan. This approach is a step forward from earlier work that compares Tibetan directly to words in East Bodish languages. Pursuing a line of research which he began in 1998, Čašule presents new evidence of an Indo-European affiliation for Burushaski; on this occasion he focuses on kinship terminology. Although many Indo-Europeanists will regard his contribution with scepticism, his engagement with the methods and literature of this field, the well-spring and the touchstone for the methodology of historical linguistics, reminds us that our knowledge of Trans-Himalayan historical phonology, in which regular correspondences across what is taken to be cognate vocabulary in the modern languages have still not been established, has far to go.

Historical linguistics, as well as areal linguistics and typology, are absolutely dependent on descriptive work on contemporary languages. Unfortunately, few Himalayan languages have been researched in depth and many remain virtually undocumented. This situation is partly due to the remoteness of many mountain localities which makes fieldwork difficult, but also to the periodic imposition of travel restrictions to certain areas. Four contributions to the current volume

provide fresh data from *in situ* fieldwork on under-described languages. Huber presents the strategies of person marking in Shumcho, where a fairly complicated verbal agreement system indexes participants of main and logophoric clauses. Michaud examines the tonal patterns in numeral plus classifier phrases in Yongning Na, whose complexity presents a number of challenges for a tonological analysis. Peterson describes the phonology and various aspects of the morphology of the highly endangered language Rengmitca, which with its conservative segmental phonology could contribute greatly to the reconstruction of Proto-Kuki-Chin, or indeed Trans-Himalayan itself. Opgenort provides an initial grammatical sketch of another highly endangered language, Tilung, which holds a unique position within the Kiranti group, though now appears to be falling into obsolescence due to the spread of Nepali.

All four of these languages are small and vulnerable. In the Himalayan Region as in other parts of the world, modernization is eroding traditional economies and lifestyles, and ancestral cultures and minority languages are (understandably) being fast abandoned for national languages such as Hindi, Chinese and Nepali which facilitate employment in the mainstream of the national economy. In such a context, the work of linguistic description and documentation often goes hand in hand with efforts in language development and education working with the community in which a fieldworker does their research. The long-term survival of a language across multiple domains of use in the modern world essentially necessitates that the language be written. Consequently, orthography development is an area in which a linguist can assist in promoting a language's long-term viability. Bodt discusses issues in the development of Roman and 'Ucen orthographies for Tshangla, in view of the phonological variation across the language's dialects, and the influence of literacy in a more prestigious language, Tibetan.

Despite such efforts, under the current rate of development, and rapid linguistic and cultural loss, it is highly probable that a great many languages used in the Himalayan Region today will no longer exist as spoken languages within a few generations. As many voices have already warned, thorough and sensitive research on the minority languages of the Himalayan Region is a matter of urgency.

References

- Bàrtoli, Matteo. 1925. *Introduzione alla neolinguistica*. Geneva: L. S. Olschki.
- Bashir, Elena. 2003. Dardic. In George Cardona and Dhanesh Jain (eds.), *The Indo-Aryan Languages*, 818–9. London: Routledge.
- Bradley, David. 1997. Tibeto-Burman languages and classification. In David Bradley (ed.), *Tibeto-Burman Languages of the Himalayas (Papers in Southeast Asian Linguistics No. 14)*, 1–71. Canberra: Department of Linguistics, The Australian National University.
- Braudel, Fernand. [1949] 1976. *The Mediterranean and the Mediterranean World in the Age of Philip II, Vol 1*. New York: Harper & Row, Publishers.
- Bickel, Balthasar and Johanna Nichols. 2003. *Typological enclaves*. Paper presented at the 5th Biannual Conference of the Association for Linguistic Typology, Cagliari, September 18, 2003.
- Bielmeier, Roland. 2004. Shafer's proto-West Bodish hypothesis and the formation of the Tibetan verb paradigms. In Anju Saxena (ed.), *Himalayan Languages Past and Present*, 395–412. Berlin: Mouton de Gruyter.
- Campbell, Lyle. 2004. *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- DeLancey, Scott. 1987. The Sino-Tibetan Languages. In Bernard Comrie (ed.), *The World's Major Languages*, 799–810. New York: Oxford University Press.
- DeLancey, Scott. 1989. Verb agreement in Proto-Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 52 (2): 315–33.
- DeLancey, Scott. 2010a. Towards a history of verb agreement in Tibeto-Burman. *Himalayan Linguistics* 9 (1): 1–39.
- DeLancey, Scott. 2010b. Language replacement and the spread of Tibeto-Burman. *Journal of the Southeast Asian Linguistics Society* 3(1): 40–55.
- DeLancey, Scott. 2011. The origins of Sinitic. In Zhuo Jing-Schmidt (ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23) 2011*, Vol. 1, 51–64. Eugene: University of Oregon.
- Denwood, Philip. 1999. *Tibetan*. Amsterdam: John Benjamins.
- Dimmendaal, Gerrit J. 2011. *Historical Linguistics and the Comparative Study of African Languages*. Amsterdam: John Benjamin's Publishing Company.
- van Driem, George. 1993. The Proto-Tibeto-Burman verbal agreement system. *Bulletin of the School of Oriental and African Studies* 56 (2): 292–334.
- van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, Containing an Introduction to the Symbiotic Theory of Language*. Leiden: Brill.
- van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1 (2): 211–270.
- Gellner, Ernest. 1969. *Saints of the Atlas*. London: Weidenfeld and Nicolson.
- George A. Grierson (ed.). 1909. *Linguistic Survey of India. Volume I, Part 1: Introductory*. Calcutta: Office of the Superintendent of Government Printing.
- Hale, Austin. 1980. Person markers: Finite conjunct and disjunct verb forms in Newari. In S.A. Wurm (ed.), *Papers in South-East Asian linguistics* 7: 95–106. Canberra: Department of Linguistics, The Australian National University.
- Hickey, Raymond (ed.). 2010. *The Handbook of Language Contact*. Singapore: Wiley-Blackwell.

- Hill, Nathan W. 2010. An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society* 108 (2): 110–125.
- Hock, Hans Henrich. 1991. *Principles of historical linguistics*. Berlin: Mouton de Gruyter.
- Hodgson, Bryan Houghton. [1828] 1972. Notices of the Languages, Literature and Religion of Nepál and Tibet. In Bryan Houghton Hodgson. *Essays on Nepal and Tibet*, Part I, 1–35. Amsterdam: Philo Press.
- Hodgson, Bryan Houghton. [1848] 1972. On the Aborigines of the Himálaya. In Bryan Houghton Hodgson, *Essays on Nepal and Tibet*, Part II, 29–37. Amsterdam: Philo Press.
- Jacques, Guillaume. 2012. Agreement morphology: the case of Rgyalrongic and Kiranti. *Language and Linguistics* 13: 83–116.
- Konow, Sten. 1909. The Tibeto-Burman family. In George A. Grierson (ed.), *Linguistic survey of India. Volume III, 1, Tibeto-Burman family*. Calcutta: Office of the Superintendent of Government Printing.
- LaPolla, Randy J. 1989. Verb agreement, head-marking vs. dependent-marking, and the ‘deconstruction’ of Tibeto-Burman morpho-syntax. In Kira Hall et al. (eds.), *Proceedings of the fifteenth annual meeting of the Berkeley Linguistics Society*, 356–67. Berkeley: Berkeley Linguistics Society.
- LaPolla, Randy J. 1992. On the Dating and Nature of Verb Agreement in Tibeto Burman. *Bulletin of the School of Oriental and African Studies* 55 (2): 298–315.
- LaPolla, Randy J. 1994. Parallel Grammaticalizations in Tibeto-Burman: Evidence of Sapir’s ‘Drift.’ *Linguistics of the Tibeto-Burman Area* 17 (1): 61–80.
- LaPolla, Randy J. 2001. The Role of Migration and Language Contact in the Development of the Sino-Tibetan Language Family. In R. M. W. Dixon and A. Y. Aikhenvald. *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*, 225–254. Oxford: Oxford University Press.
- LaPolla, Randy J. 2009. Causes and effects of substratum, superstratum and adstratum influence, with reference to Tibeto-Burman languages. In Yasuhiko Nagano (ed.), *Issues in Tibeto-Burman historical linguistics* (Senri Ethnological Studies 75), 227–237. Osaka: National Museum of Ethnology.
- Masica, Colin. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley, Los Angeles, and London: University of California Press.
- Matisoff, James A. 2011. Series Editor’s Introduction. In Christopher Button. *Proto-Northern Chin*. Berkeley: University of California.
- Nelson, David Niles. 1986. *The Historical Development of the Nuristani Languages*. Minneapolis and St. Paul, MN: University of Minnesota dissertation.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 1997. Modeling Ancient Population Structures and Movement in Linguistics. *Annual Review of Anthropology* 26: 359–384.
- Post, Mark W. and Roger Blench. 2011. Siangic: A new language phylum in North East India. Paper presented at the 6th International Conference of the North East Indian Linguistics Society, Tezpur University, Assam, India, 31 January – 2 February 2011.
- Scott, James C. 2009. *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia*. New Haven and London: Yale University Press.

- Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word, Journal of the Linguistic Circle of New York* 11: 94–111.
- Shafer, Robert. 1966. *An Introduction to Sino-Tibetan: Part 1*. Wiesbaden: Harrassowitz.
- Strand, Richard F. 1973. Notes on the Nuristani and Dardic Languages. *Journal of the American Oriental Society* 93 (3): 297–305.
- Thomason, Sarah G. and Terrence Kaufman (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tournadre, Nicolas. 2008. Arguments against the concept of ‘conjunct’/‘disjunct’ in Tibetan. In B. Huber, M. Volkart and P. Widmer (eds.). *Chomolangma, Demawend und Kasbek, Festschrift für Roland Bielmeier*, 281–308. Halle (Saale): International Institute for Tibetan and Buddhist Studies.
- Turin, Mark 2007. *Linguistic Diversity and the Preservation of Endangered Languages: A Case Study from Nepal*. Kathmandu: International Centre for Integrated Mountain Development (ICIMOD)
- Watters, David E. 2006a. The Conjunct-Disjunct Distinction in Kaike. *Nepalese Linguistics* 22: 300–310.
- Watters, David E. 2006b. Notes on Kusunda Grammar: A language isolate of Nepal. *Himalayan Linguistics Archive* 3: 1–182.
- Zeisler, Bettina. 2005. On the position of Ladakhi and Balti in the Tibetan language family. In John Bray (ed.). *Ladakhi histories: Local and regional perspectives*, 41–64. Leiden: Brill.
- Zeisler, Bettina. 2009. Reducing phonological complexity and grammatical opaqueness: Old Tibetan as a lingua franca and the development of the modern Tibetan dialects. In Enoch Aboh and Norval Smith (eds.). *Complex processes in new languages*, 75–95. Amsterdam and Philadelphia: Benjamins.

George van Driem

Trans-Himalayan

1 Astride the Himalayas

This Trans-Himalayan tale unites two narratives, an historical account of scholarly thinking regarding linguistic phylogeny in eastern Eurasia alongside a reconstruction of the ethnolinguistic prehistory of eastern Eurasia based on linguistic and human population genetic phylogeography. The first story traces the tale of transformation in thought regarding language relationships in eastern Eurasia from Tibeto-Burman to Trans-Himalayan. The path is strewn with defunct family trees such as Indo-Chinese, Sino-Tibetan, Sino-Himalayan and Sino-Kiranti. In the heyday of racism in scholarship, Social Darwinism coloured both language typology and the phylogenetic models of language relationship in eastern Eurasia. Its influential role in the perpetuation of the Indo-Chinese model is generally left untold. The second narrative presents a conjectural reconstruction of the ethnolinguistic prehistory of eastern Eurasia based on possible correlations between genes and language communities. In so doing, biological ancestry and linguistic affinity are meticulously distinguished, a distinction which the language typologists of yore sought to blur, although the independence of language and race was stressed time and again by prominent historical linguists.

2 The teetering trail from Tibeto-Burman to Trans-Himalayan

The Tibeto-Burman linguistic phylum was identified in 1823. However, the term “Tibeto-Burman” was subsequently used in two different meanings, one by scholars following Julius von Klaproth’s polyphyletic framework and another by scholars operating within the Indo-Chinese or Sino-Tibetan paradigm. The essential differences between the two lineages of thought are contrasted, and the evidence is weighed. The geographical distribution of major subgroups and the phylogeny of the language family provide clues to Tibeto-Burman ethnolinguistic population prehistory. Several alternative theories of linguistic relationship are discussed and the major subgroups are presented.

In 1823, Julius von Klaproth identified the Tibeto-Burman phylum in Paris in his polyphyletic view of Asian linguistic stocks. Klaproth’s model of many distinct Asian linguistic phyla was initially controversial because many scholars in

the West at the time entertained an undifferentiated view of Asian languages as all belonging to some nebulous all-encompassing language family. His Tibeto-Burman comprised Burmese, Tibetan and Chinese and all of the languages which could be demonstrated to be related to these three. He explicitly excluded languages today known to be Kra-dai or Daic (e.g. Thai, Lao, Shan), Austroasiatic (e.g. Mon, Vietnamese, Nicobarese, Khmer) and Altaic (e.g. Japanese, Korean, Mongolic, Turkic). The name Tibeto-Burman gained currency in English for the language family recognised by Klaproth and was widely used by scholars in the British Isles, e.g. Hodgson (1857), Cust (1878), Forbes (1878), Houghton (1896).

Some other scholars of the day followed the Indo-Chinese theory proposed by the Scots amateur John Casper Leyden, who died at the age of 35 after making a short but dazzling career in the British colonial administration in Asia during the Napoleonic wars. In 1807, Leyden proposed his exuberant but poorly informed Indo-Chinese theory to George Barlow, Governor General of India at Fort William, in which he claimed that all the languages in Asia and Oceania shared some “common mixed origin” (Leyden 1808).

This murky view held appeal to adherents of Biblical mythology who had been inclined to lump Chinese together with numerous other Asian languages into a grand Japhetic family, on the assumption that Chinese was one of the languages spoken by the descendants of Noah’s son Japhet, whilst some alternatively attempted to explain Chinese as an antediluvian language or as one of the “confounded” forms of speech with which Yahweh had afflicted mankind after the fall of the Tower of Babel. Klaproth was the first scholar to assign Chinese to its proper language family.

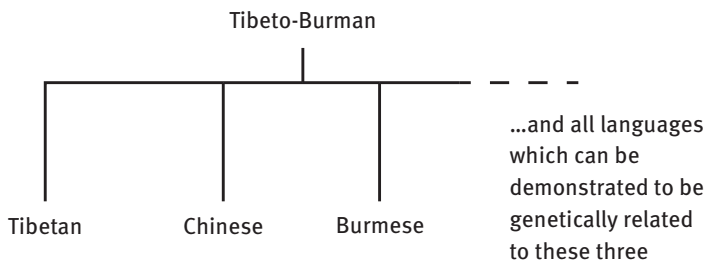


Fig. 1: Julius von Klaproth’s Tibeto-Burman family

The Biblically inspired Japhetic was not the only pan-Asian catch-all. Wilhelm Schott wrote personally to the famous scholar of Himalayan languages Brian Houghton Hodgson to warn him against the “Turanian” theory then being propa-

gated from Oxford. In 1856, Schott likewise published an essay warning against “Indo-Chinese” (Schott 1856). Schott foresaw that scholars who used the label would continue to think in terms of the mistaken phylogenetic model which the label designated. Yet the Indo-Chinese model became the favourite of racist language typologists who believed that Asian languages were generally more rudimentary and that Asian peoples were more primitive than their Western counterparts.

Grammatical typology inspired language typologists such as Heymann Steinthal (1850, 1860), Ernest Renan (1858), Arthur de Gobineau (1854–1855) and John Beames (1868) to rank Chinese and Thai together on the lowest rung of the evolutionary ladder of language development based on their “monosyllabicity” and lack of inflection. These scholars argued that Chinese and Thai must be closely related and that neither was part of Tibeto-Burman. James Byrne argued that “the causes which have determined the structure of language” lay in the varying “degrees of quickness of mental excitability possessed by different races of men” (1885: 45). Chinese and Siamese ostensibly mediated a rudimentary, less evolved way of thinking and so were assigned to the lowest rungs of Steinthal’s ladder of language evolution.¹ The following quote typifies this once widespread genre of scholarly discourse.

...la langue chinoise, avec sa structure inorganique et incomplète, n'est-elle pas l'image de la sécheresse d'esprit et de cœur qui caractérise la race chinoise? ...Suffisante pour les besoins de la vie, pour la technique des arts manuels, pour une littérature légère de petit aloi, pour une philosophie qui n'est que l'expression souvent fine, mais jamais élevée, du bon sens pratique, la langue chinoise excluait toute philosophie, toute science, toute religion, dans le sens où nous entendons ces mots. (Renan 1858: 195–196).

¹ Through the lense of historical hindsight, racist linguistic typology in the 19th century had its burlesque moments, as, for example, when some linguists contested Steinthal’s hierarchy on the basis of the argument that ‘Negeridiomen’ could not possibly be positioned on rungs that were higher on the typological tree of language evolution than Chinese or Siamese in view of the differences in the material cultures of the language communities concerned. Another ludicrous moment was the coinage of the term ‘analytic’ to characterise languages such as English and French, which were no longer flamboyantly flexional and must therefore have ostensibly evolved beyond the stage of perfection purportedly reflected by Sanskrit. To account for the contrast between the technological advancement of Chinese civilisation and the ostensibly low rung on the typological ladder of language evolution ascribed to the Chinese language, Comte de Gobineau invented a distinction between so-called male and female races, whereby “les races males” possessed “un langage plus précis, plus abondant, plus riche que les races femelles” (1854, i: 190). His explanation, therefore, was that the Chinese “race” was in some sense “male” despite the inferior status which he imputed to the typological traits of the Chinese language.

Such reasoning contrasted starkly with the older but more sophisticated tradition of linguistic relativity, developed by John Locke (1690), Étienne de Condillac (1746), Pierre de Maupertuis (1748, 1756) and Wilhelm von Humboldt (1822, 1825, 1836). Linguists following this scholarly tradition, notably Julius von Klaproth (1823), Jean Jacques Nicolas Huot (Malte-Brun 1832, i: 521), August Friedrich Pott (1856) and Friedrich Max Müller (1871, 1881), vehemently opposed the ideas of the racist language typologists, stressed that biological ancestry was independent of language, and argued that the relationship between language structure and human cognition was not at all so simplistic, but more subtle, more interesting and, then as today, still largely unexplored.²

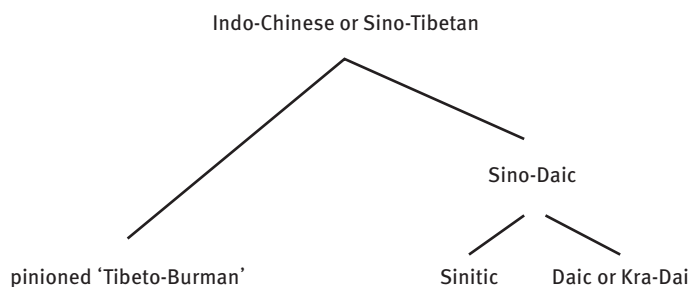


Fig. 2: The Indo-Chinese or Sino-Tibetan theory: Kradai or Daic has been excluded since the 1940s

At first, Indo-Chinese encompassed Asian languages from the Caspian Sea to Polynesia. This untenable construct embodied numerous misguided phylogenetic conjectures and so it came to be whittled down in successive stages. After Philipp von Siebold (1832) and Anton Boller (1857) presented their case for a distinct Altaic phylum, Ernst Kuhn (1883, 1889) attempted to remedy what was still wrong with the Indo-Chinese model by correcting the erroneous inclusion of Austroasiatic, but the resulting model still represented a false family tree. Yet

² The historical linguistic tradition of linguistic relativity was antagonistic to the racist tradition of the language typologists. Yet in the wake of the Second World War, the rejection of racism in most scholarly circles often went hand in hand with an unrefined, undifferentiated view of the distinct strands in the history of linguistic thought. Against this background, the backlash against the shortcomings in the writings of Benjamin Lee Whorf, who died in 1941, led to the view, dogmatically propounded in many introductory courses in general linguistics worldwide, that all languages are created equal. This smug spirit of linguistic equivalence would have been music to the ears of Pierre Maine de Biran (1815), but fortunately scholars such as George Grace (1989) continued to contest this post-war orthodoxy.

some scholars and several notable sinologists adopted the Indo-Chinese name and the false Indo-Chinese phylogeny, e.g. von der Gabelentz (1881), Forchhammer (1882), Conrady (1896), Laufer (1916), Wulff (1934).

In 1924, the French orientalist Jean Przyluski coined *sino-tibétain* as the French term for Indo-Chinese in the English and German sense (Przyluski 1924).³ This French term entered English in 1931 when Jean Przyluski and Gordon Luce co-authored an article on the root for the numeral hundred in “Sino-Tibetan” (Przyluski and Luce 1931). The new term did not catch on at once, but during the Great Depression in 1935 the American president Franklin Roosevelt instituted the employment scheme called the Works Progress Administration. Through WPA, the famous Berkeley anthropologist Alfred Kroeber, inspired by the enthusiasm of Robert Shafer, managed to raise funding for his Sino-Tibetan Philology project. Changing the name of the model of linguistic relationship to the new Gallic label helped to deflect the widespread criticism against Indo-Chinese.

Shafer effectively ran the project for Kroeber, but saw two things fundamentally wrong with “Sino-Tibetan.” In 1938, Shafer proposed to remove Kradai or Daic from the language family, but in the end he was not allowed to do so (Shafer 1955: 97–98). Shafer also put Sinitic on par with other divisions in the family. The two operations would effectively have heralded a return to Julius von Klaproth’s original Tibeto-Burman model. After Paul Benedict came to Berkeley in the winter of 1938–1939 to join the project, he traded in the name Indo-Chinese for “Sino-Tibetan.” Moreover, after the conclusion of the project in 1940, he took credit for removing Daic (Benedict 1942). Benedict (1972) also restored Sino-Tibetan to its original Indo-Chinese shape, again isolating Chinese as the odd man out.

Ironically, after the Cultural Revolution, Chinese scholars adopted as orthodoxy the Indo-Chinese model as repackaged in America. Sino-Tibetan became 漢藏語系 *Hàn-Zàng yǔxì*, notwithstanding its empirically unsupported phylogeny and its racist legacy. Historically, Sino-Tibetan is rooted in the fact that morphosyntactic typology had perplexed less enlightened linguists of 19th century into believing that Chinese and Thai represented an inferior developmental stage on a Steinthal’s ladder of language evolution. This view relied on the assumption that Sinitic languages had never evolved and that Chinese had remained typologically unchanged and “without inflection, without agglutination” for millennia, e.g. Chalmers (1866).

3 The need to coin a proper French term had become pressing, since in French *indochinois* referred politically and geographically to the French colonial dominions on the Indo-Chinese peninsula and linguistically to the Mon-Khmer-Kolarian or Mon-Annam linguistic phylum which Wilhelm Schmidt had renamed Austroasiatic at the beginning of the 20th century. Some British writers fond of terminological gallicisms also used the term ‘Indo-Chinese’ in the meaning Austroasiatic, e.g. Sir Richard Temple (1903: iii, 251–284).

By contrast, the informed historical linguistic view represented quite a different understanding of Chinese. Carl Richard Lepsius (1861: 492–496) proposed that Chinese tones had arisen from the merger of initials and the loss of finals based on correspondences between Chinese and Tibetan. He argued that entire syllables had been lost in Chinese and that Chinese ideograms once represented words which may often have contained more than just the root syllables whose reflexes survive in the modern pronunciations. The view of Chinese promulgated by Lepsius later inspired Bernhard Karlgren (1920, 1957) to conceive of Old Chinese as a *langue flexionnelle* and to undertake the reconstruction of Old Chinese in accordance with the principles of the comparative method.

Two models of phylogenetic relationship sought to defy the Sino-Tibetan paradigm propagated at Berkeley, i.e. Sino-Himalayan (Bodman 1973, 1980) and Sino-Kiranti (Starostin 1994). Although neither proposal gained acceptance, these sallies made the crucial point that to date no evidence has ever been adduced in support of the Sino-Tibetan phylogenetic model, defined by its truncated “Tibeto-Burman” taxon encompassing all non-Sinitic languages. Methodologically, attempts to define all non-Sinitic languages negatively in terms of Sinitic innovations which other languages lack or to invoke the argument of gross word order for Karen and Sinitic, as Benedict (1976) once did, are known to be phylogenetically meaningless. All comparative evidence amassed to date supports Julius von Klaproth’s 1823 minimalist Tibeto-Burman tree, which epistemologically therefore continues to represent the default model.

However, the history of the field has left us with an unfortunate nomenclatural legacy. Whereas Tibeto-Burmanists in Klaproth’s tradition used the name “Tibeto-Burman” for the family as a whole, Sino-Tibetanists continue to use the term “Tibeto-Burman” to denote all non-Sinitic languages as comprising a single taxon. In an attempt to escape this terminological morass, in 2004 the alternative name “Trans-Himalayan” was proposed for the linguistic phylum because the world’s second most populous language family straddles the great Himalayan range along both its northern and southern flanks (van Driem 2007a: 226).

This neutral geographical term is analogous to “Indo-European” and “Afro-Asiatic” in reflecting the geographical distribution of the language family. The term “Afro-Asiatic” was coined in 1914 and replaced the earlier “Hamito-Semitic” for similar reasons. Hamitic was shown not to be a valid subgroup, just as Sino-Tibetan, defined by its unitary non-Sinitic taxon, likewise denotes a false tree. The linguistic phylum is, of course, literally Trans-Himalayan in distribution. By far most of the roughly 300 different Tibeto-Burman languages and three fourths of the major Trans-Himalayan subgroups are situated along the southern flanks of the Himalayas (Figure 3), whilst by far most speakers of

Trans-Himalayan languages live to the north and east of the great Himalayan divide (Figure 4).

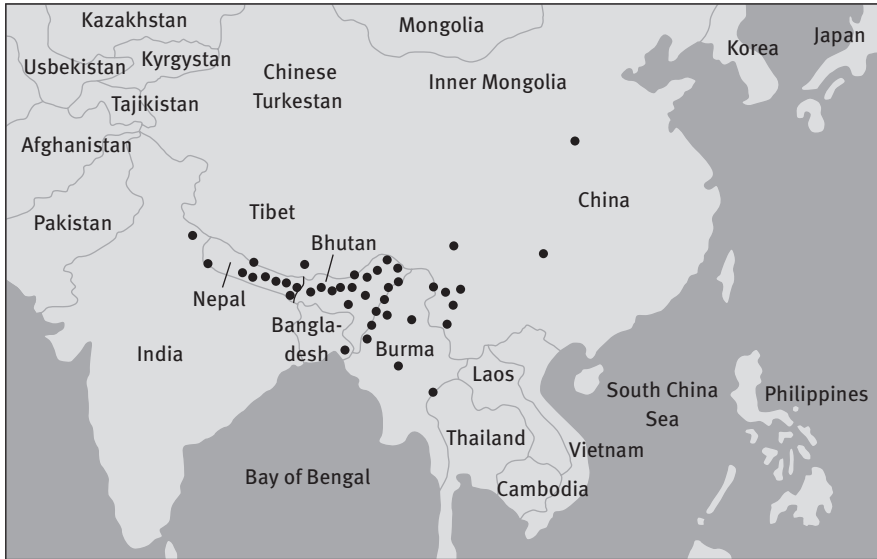


Fig. 3: Geographical distribution of the major Trans-Himalayan subgroups. Each dot represents not just one language, but the putative historical geographical centre of each of 42 major linguistic subgroups.

3 Towards a linguistic phylogeography

Much more is known about the Tibeto-Burman language family today than in the days of Klaproth. Today we can identify 42 subgroups for which there appears to be evidence and about which there is some degree of consensus. The 2012 version model of the Fallen Leaves model, shown in Figure 5, contains a number of groups not mentioned when this model was first presented (van Driem 2001). The Rgyalrongic subgroup was proposed and validated by Jackson Sun (2000a, 2000b). The Nàic subgroup, comprising Nàmùyì and Shǐxīng and the closely related Nàish languages, i.e. ¹Na²khi [na˧˥hi˧˥] (Nàxī), Moso (Mósuō a.k.a. [na˧˥˩]) and Laze [la˧˥ze˧˥], has been proposed by Jacques and Michaud (2011). Evidence for an Ērsūish subgroup has been presented by Yu (2011). The validation of lower-order groups not only enables the validation of correctly delineated higher-order groups, but will also give us a clear view of their internal phylogeny.

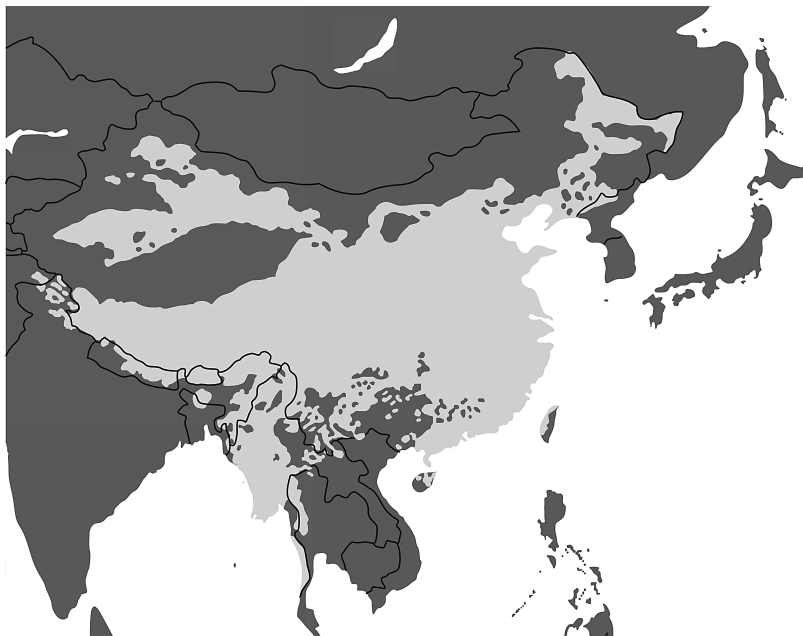


Fig. 4: Geographical distribution of Trans-Himalayan languages

Post and Blench (2011) presented evidence for Siangic, a group comprising Milang and Koro. At one level, Post and Blench envisage Siangic not as a Tibeto-Burman subgroup, but as an altogether non-Tibeto-Burman phylum which has left vestiges in Koro and Milang. A more conservative stance would be to treat Koro and Milang together as a Tibeto-Burman subgroup in their own right. In a similar vein, many scholars have recently publicly aired the view that Puroik, also known as Sulung, normally deemed to be a member of the Kho-Bwa cluster of languages, is not a Tibeto-Burman language at all. Despite the apparently aberrant nature of some of the lexicon, Puroik, Koro and Milang all exhibit a good share of Tibeto-Burman vocabulary. The history of Indo-European is instructive in this regard.

French shows a smidgen of Celtic lexicon that can be viewed as substrate (Lambert 1994), whilst the language itself is indisputably a Romance dialect. Words borrowed from the substrate language do not determine the linguistic affinity of a language. Until Ritter von Xylander (1835), Albanian was held to be a language isolate in Europe just like Basque. It is sobering to reflect that less is known today about Tibeto-Burman historical grammar than was known in 1835 about Indo-European historical grammar. The Gongduk language in Bhutan is analogous to Albanian, or for that matter much like Koro, Milang and Puroik,

in exhibiting much vocabulary which appears outlandish from a Tibeto-Burman perspective. Yet our perspective on Tibeto-Burman has been changing rapidly in recent years, as more becomes known about the less well documented languages of the phylum. Our understanding of what Starostin called “Tibeto-Burman in the narrow sense” is broadening to encompass a more informed and fine-mesh view.

The growing awareness in the field that the Tibeto-Burman analogues of Armenian, Hittite and Albanian all appear to be found within the eastern Himalayas highlights the fact that the language family’s centre of phylogenetic diversity lies squarely within the eastern Himalayas. The lexical diversity observed in many subgroups of the eastern Himalayas is just one residue of a complex and many-layered ethnolinguistic prehistory in a region of ancient human habitation.

The whereabouts and the names of the languages in the 42 leaves that have fallen from the Trans-Himalayan tree are listed below. The most obvious disambiguations are indicated with the symbol ≠ with additional elucidation. Realities on the ground are far more complex than any short list can show. Related but entirely distinct and mutually unintelligible languages sometimes go by the same name, e.g. Magar, Limbu, Chinese. So the roughly 280 language labels in this non-exhaustive list obscure a great deal of dialectal and linguistic diversity.



Fig. 5: The 2012 version of the agnostic Fallen Leaves model. Thirty out of 42 Tibeto-Burman subgroups lie south of the great Himalayan divide, seven to the north and east of the Himalayas, and five, i.e. Tshangla, Bodish, Nungish, Lolo-Burmese and Kachinic, are distributed on both sides of the Himalayas

Sometimes the ethnic designation and the mother tongue do not match, as when a community, for example, consider themselves Jǐngpō but speak the Lolo-Burmese language Zaiwa or when a community consider themselves Tibetan but speak a Rgyalrongic language. Some languages are extinct, e.g. Pyu, Dura, believed to be extinct, e.g. the Sak languages, or moribund, e.g. Barām. In fact, most Tibeto-Burman languages are endangered with imminent extinction. A more detailed account can be found in the handbook *Languages of the Himalayas* (van Driem 2001) and in the literature referenced therein.

Angami-Pochuri (southern Nagaland, northern Manipur, neighbouring portions of Burma and Assam): Angami, Chokri a.k.a. Chakri, Kheza, Mao a.k.a. Sopvoma, Pochuri, Ntenyi, Maluri a.k.a. Meluri, Sema, Rengma, Kezhama, Senkadong.

Ao (central Nagaland and neighbouring portions of Burma): Yacham, Ao Chungli, Ao Mongsen, Yimchungrü a.k.a. Yachumi, Sangtam a.k.a. Thukumi, Yacham and Tengsa, Lotha a.k.a. Lhota.

Bái (the area around Dàlǐ in Yúnnán province): Bái.

Black Mountain Mōnpa (the Black Mountains of Bhutan): 'Olekha, Riti, Jangbi and 'Wangling.

Bodish (Tibet, Pakistan, India, Nepal, Sikkim, Bhutan): Balti, Purik, Ladakh, Zanskar, Lahul, Central Tibetan (Dbus and Gtsang), Sherpa, Ölmo Sherpa, Lhomi, Jirel, Kagate, Mustang, Limirong, Mugu, Northern Kham, Eastern Kham, Amdo Tibetan, Brokpa, Dzongkha, Lakha, Dränjoke, Cho-ca-nga-ca-kha, Bumthang, Kheng, Mangde, Kurtöp, Chali, Dzala, Dakpa.

Brahmaputran a.k.a. Bodo-Koch and Northern Naga (West Bengal, Assam, Meghalaya, Arunachal Pradesh, northern Nagaland and adjacent portions of Burma): Chutiya, Kokborok, Tiwa, Dimasa a.k.a. Hills Kachāḍi, Bodo, Plains Kachāḍi, Meche, Garo, Atong, Pani Koch, Ruga, Rabha, Tangsa, Nocte, Wancho, Kuwa, Haimi, Htangan, Konyak, Ponyo, Phom, Chang, Welam, Nokaw.

Chepangic (central Nepal): Chepang, Bhujeli.

Dhimalish (eastern Nepalese Terai, western Bhutanese duars): Dhimal, Toto.

Digarish a.k.a. 'Northern Mishmi' (Dibang river valley, Lohit district, Arunachal Pradesh): Idu, Taraon a.k.a. Digaro.

Dura (central Nepal's Lamjung district): Dura.

Ĕrsūish (southern Sichuān, northern Yúnnán): Ĕrsū, Tosu, Lizu.

Gongduk (south central Bhutan): Gongduk.

Rgyalrongic (southern Sichuān): Situ, Japhug, Tsobdun, Zbu, Lavrung (inc. Thurje Chenmo and nDzorogs), Horpa (inc. rTau and Stod-sde).

Hrusish (western Arunachal Pradesh): Hruso a.k.a. Aka, Dhímmai a.k.a. Miji, Levai a.k.a. Bangru.

Kachinic a.k.a. Jinghpaw (northeastern India, northern Burma, southern Yúnnán): The various Kachin, Singpho, Jìngpō or Jinghpaw languages and the Sak a.k.a. Luish languages Sak, Kadu, Andro, Sengmai, Chairel.

Karbí a.k.a. Mikir (Mikir Hills or Karbí Anglóng, neighbouring districts of Assam): Karbí a.k.a. Mikir.

Karenic (lower Burma, the Tenasserim and adjacent Thailand coastal regions): Pa'o, Pwo, Sgaw, Kayah, Brek a.k.a. Bwe, Bghai.

Kho-Bwa (western Arunachal Pradesh): Khowa a.k.a. Bugun, Sherdukpen, Puroik a.k.a. Sulung, Lishpa.

Kiranti (eastern Nepal): Pāñcthare Limbu, Phedāppe Limbu, Tamarkhole Limbu, Chathare Limbu, Yakkha, Chiling, Āṭhpahariyā (inc. Belhare), Lohorung, Yamphu, Mewahang, Kulung, Nachiring, Sampang, Sam, Chamling, Puma, Bantawa, Chintang, Dungmali, Thulung, Jero, Wambule, Tilung, Dumi, Khaling, Kohi, Bahing, Sunwar, Hayu.

Kukish a.k.a. Mizo-Kuki-Chin (Mizoram and Indo-Burmese borderlands): Mizo a.k.a. Lushai, Lai, Sizang a.k.a. Siyin, Thado, Tiddim Chin a.k.a. Paite a.k.a. Sokte a.k.a. Kamhau, Haka, Chinbok, Laizo, Lakher, Ashö, Khumi Chin, Hmar, Anal, Lakher a.k.a. Mara, Falam, Vaiphei, Lamgang, Simte.

Lepcha (Sikkim, Darjeeling, Kalimpong): Lepcha.

Lhokpu (southwestern Bhutan): Lhokpu a.k.a. Doya.

Lolo-Burmese (southwestern China, Burma, Southeast Asia): Burmese, Zaiwa (≠ Midžuish Zaiwa) a.k.a. Atsi, Lăshi, Măru (≠ Mru in the Chittagong), Maingtha a.k.a. Achang a.k.a. Ngachang, Hpon a.k.a. Hpun, Dănu, Taungyo a.k.a. Tăru (≠ Danaw), Phunoi, Akha, Lahu, Lisu, mBisu, Ahsi, and various Yí languages.

Magaric (central Nepal): Syāṅgjä Magar, Tanahũ Magar, Pālpā Magar, Khām Magar a.k.a. Kham (≠ Tibetan Kham).

Meithei (Manipur): Meithei a.k.a. Manipuri

Pyu (extinct language of pre-Burmese epigraphy in Burma): Pyu.

Midžuish a.k.a. 'Southern Mishmi' (Lohit drainage, Lohit district, Arunachal Pradesh): Kaman a.k.a. Miju a.k.a. Mijhu, Zaiwa (spoken by the Meyöl clan near Walong ≠ Burmic Zaiwa).

Mru (in the Chittagong of Bangladesh): Mru a.k.a. Măru (≠ the Shan State Măru in Burma).

Nàic (southern Sìchuān, northern Yúnnán): ¹Na²khi (Nàxī), Moso (Nà, Mósuō), Laze, Nàmùyì, Shíxīng.

Newaric (central Nepal): Kathmandu Newar, Pahari Newar, Badikhel Newar, Chitlang Newar, Dolakha Newar, Barām, Thangmi.

Nungish (Yúnnán province, northern Burma): Trung, Ālóng, Răwang, Róuruò, Nung inc. Nūsū and Ānù (≠ the Daic Nung in northern Vietnam).

Qiāngic (southern Sìchuān, northern Yúnnán): Southern Qiāngic, Northern Qiāngic, Mǐ-ñag (Mùyǎ), Prinmi (Pǔmǐ), Choyo (Quèyù), Tangut (Xixià), Zhābā, Ěrgōng, Guìqióng.

Raji-Raute (western Nepal, Uttarakhand): Raji, Raute.

Siangic (Arunachal Pradesh): Koro, Milang.

Sinitic (China): Mandarin, Cantonese, Wú, Gàn, Xiāng, Hakka a.k.a. Kèjiā, Southern Mǐn (inc. Hokkien), Eastern Mǐn, Northern Mǐn, Central Mǐn, Càijiā, Wǎxiāng.

Tamangic (central Nepal): Tamang, Gurung, Thakali, Chantyal, Ghale, Kaiké, 'Narpa, Manangba.

Tangkhol (northeastern Manipur, neighbouring parts of Burma): Tangkhul, Maring.

Tani a.k.a. Abor-Miri-Dafla (Arunachal Pradesh, neighbouring portions of Assam): Apatani, Nyisu, Bengni, Nishing, Tagin, Yano, Sarak a.k.a. Hill Miri, Galo, Bokar, Ramo, Ashing, Pailibo a.k.a. Libo, Damu, Bori, Mishing a.k.a. Plains Miri, Padam, Shimong, Pasi, Panggi, Tangam, Karko, Minyong.

Tshangla a.k.a. Shâchop (eastern Bhutan, enclaves in Arunachal Pradesh and Tibet): Tshangla a.k.a. Shâchop or loconyms.

Tǔjiā (Húnán, Húběi and Guìzhōu provinces): Tǔjiā.

West Himalayish (Himachal Pradesh, Uttarakhand): Manchad, Tinan, Bunan a.k.a. Gari, Kanashi, Rangpo, Darma, Byangsi, Rangkas, Zhangzhung.

Zeme (southwestern Nagaland, northwestern Manipur, neighbouring portions of Assam): Mzieme, Liangmai a.k.a. Kwoireng, Zeme a.k.a. Empeo Naga a.k.a. Kacha Naga, Maram, Khoirao, Puiron, Rongmai a.k.a. Kabui a.k.a. Nruanghmei.

Some of the subgroups in the above list of 42 fallen leaves represent tentative subgrouping hypotheses that have yet to be subjected to closer scrutiny, e.g. Newaric, Qiāngic. By the same token, questions arise such as whether Bodish should include East Bodish as well as Bodish proper, and how East Bodish should otherwise be renamed, or whether Brahmaputran should encompass both the Bodo-Koch as well as the Northern Naga languages. In historical linguistics, it is preferable to work from the bottom up, i.e. starting with the tangible leaves that have fallen from the tips of the branches, and then moving upward to gain an understanding of the nodes in the tree. Yet many Tibeto-Burman languages are still poorly documented and scantily described.

The Fallen Leaves model is no definitive phylogeny by definition. Though agnostic about higher-order subgrouping, the model does not deny that there is a family tree whose structure must be ascertained by historical linguistic methods. The continuing identification of subgroups presents a challenge to the current generation and to future generations of historical linguists to reconstruct the

internal phylogeny of Trans-Himalayan on the basis of reliable data and regular sound laws and not to accept false family trees that we inherit from our predecessors or find in the literature without the support of historical comparative evidence. Two of Shafer's old "divisions" continue to lead robust lives of their own as higher-order albeit vaguely delineated subgrouping proposals, i.e. Bodic and Burmic.

Recently, Jacques and Michaud (2011) have proposed a higher-order subgroup called Burmo-Qiāngic, comprising Lolo-Burmese and a subgroup newly christened Nà-Qiāngic. Nà-Qiāngic essentially represents the same catch-all that used to be called 'Qiāngic' *sensu lato*. This constellation of subgroups has now been rendered less nebulous, however, by Sun (2000a, 2000b), Yu (2011) and Jacques and Michaud (2011), who have validated the Rgyalrongic, Ērsūish and Nàic subgroups respectively. In addition to these three subgroups, Nà-Qiāngic also contains Mi-ñag (Mùyǎ), Prinmi (Pǔmǐ), Choyo (Quèyù), Tangut (Xìxià), Zhābā, Qiāngic *sensu stricto* and perhaps Ērgōng and Guìqióng. The internal phylogeny of the latter medley of subgroups still has to be worked out, and the higher-order subgrouping hypotheses Nà-Qiāngic and Burmo-Qiāngic likewise require validation.

The Càijiā 蔡家 language was recently discovered in the northwestern corner of Guìzhōu (Bó 2004). Zhèngzhāng (2010) considers Càijiā to be a member of the same subgroup as Bái, whereas Sagart believes that both Càijiā as well as the Wǎxiāng 瓦鄉 dialect of western Húnán could represent the first sub-branches of the Sinitic subgroup to have split off from Proto-Sinitic, even before the Mǐn dialects (de Sousa 2012).

Another higher-order subgrouping hypothesis, Sino-Bodic, has a long history. Julius von Klaproth (1823) observed that Tibetan and Chinese appeared to be more closely related to each other than either were to Burmese. Simon (1927, 1928, 1929) and Forrest (1956, 1962) adduced lexical evidence which suggested a closer relationship between Chinese and Tibetan within the family. Although Shafer criticised Simon's work, Shafer (1955) too observed that a closer genetic affinity obtained between Sinitic and Bodic than between any other two divisions. Later Bodman (1973, 1980) too adduced evidence indicating a closer relationship between Sinitic and Bodic. The name "Sino-Bodic" was proposed for the hypothesis, and additional lexical evidence for this affinity was adduced (van Driem 1997). Matisoff (2000) protested, but most of the Sino-Bodic evidence still stands (van Driem 2005). Possible new evidence for Sino-Bodic has been adduced by Nathan W. Hill (2011) and Zhèngzhāng Shàngfāng (2011). Future research will determine whether any of these supergroups will survive the test of time.

4 Paternal patterns

Despite valiant efforts by David Bradley (2012), Blench's (2009) claim still appears to hold that no rice agricultural terminology can be confidently reconstructed for the Tibeto-Burman phylum. Instead the linguistic ancestors of the Austroasiatics and the Hmong-Mien appear to be the likeliest candidates behind the early cultivation and later the domestication of Asian rice (van Driem 2011, 2012). Rather, as has long been widely presumed, the ancient Trans-Himalayans probably cultivated foxtail millet *Setaria italica* and broomcorn millet *Panicum mileaceum*. Yet significant advances in linguistic palaeontology, supported by detailed descriptions and lexicographical documentation, in tandem with genetic work on these two cultigens may one day bring us closer to unravelling this portion of the Trans-Himalayan past.

A more obvious approach to tackling our prehistory than studying the link between languages and millet genes is the study of possible correlations between genetic markers in modern language communities and the phylogeography of the languages which they speak. However, from the beginning of the 19th century, when Jean-Baptist Lamarck elaborated his theory of evolution, to the Second World War, interdisciplinary approaches tying linguistics and human biological ancestry have had a chequered history. Since genes are always inherited by offspring from their parents, whilst the languages spoken by people are not necessarily those that were spoken by their parents or grandparents, correlations between languages and genes could only be probabilistic at best, and there need not be any relationship whatsoever.

Therefore, it is highly interesting that when geneticists began to look for correlations between genetic markers and the geographical distribution of language communities, they began to find statistically relevant correlations, not with genetic markers on the maternally inherited mitochondrial DNA but with genetic markers on the paternally inherited Y chromosome. Such a tendency, first recognised in the pioneering studies of Poloni et al. (1997, 2000), has repeatedly been observed that some correlation obtains between the most frequent Y chromosomal haplogroups of a community and the language which the people happen to speak. This correlation between a community's language and that community's prevalent paternal ancestries is what I called the Father Tongue hypothesis (van Driem 2002).

There are a number of reasons why we might expect this outcome. Initial human colonisation of any part of the planet must have involved both sexes in order for a population of progeny to establish itself. Once a population is in place, however, subsequent migrations could have been heavily gender-biased. Subsequently, male intruders could impose their language whilst availing themselves

of the womenfolk already in place. Presumably, tribes of Amazons could have spread in a similar fashion. If so, however, then the tell-tale correspondences between mitochondrial lineages and the distribution of linguistic phyla should certainly have been detected by now, but any correlation between maternal lineages and linguistic phylogeography discerned to date has been underwhelming. The Father Tongue hypothesis suggests that linguistic dispersals were, at least in most parts of the world, posterior to initial human colonisation and that many linguistic dispersals were predominantly later male-biased intrusions.

If we infer that a mother teaching her children their father's tongue has been a recurrent, ubiquitous and prevalent pattern throughout linguistic history, then some of the mechanisms of language change over time are likely to be inherent to the dynamics of this pathway of transmission. Such correlations are observed worldwide. The correlation of Niger-Congo languages with Y chromosomal haplogroups is a striking example (Wood et al. 2005). Likewise, the martial and male-biased historical spread of Hà Chinese during the sinification of southern China, recounted in painstaking detail in the Chinese chronicles, is clearly reflected in the genetic evidence (Wen et al. 2004). A recent common ancestry between native Americans and indigenous Altaians is also based preponderantly on the shared Y chromosomal heritage and is not quite as well reflected in the mitochondrial lineages (Dulik et al. 2012).

Whilst father tongues may predominate globally, mother tongues certainly do exist in the sense that there are areas on the planet where the linguistic affinity of a community corresponds more closely to the maternally transmitted mitochondrial lineage which the speakers share with other linguistically related communities. In this sense, in the north of today's Pakistan, the Balti speak a Tibetic mother tongue but profess a paternal religion that was first propagated in this area as early in the 8th century by men who came from the Near East, although the wholesale conversion of Baltistan to Islam is held to have begun only in the 14th century. The most prevalent mitochondrial DNA lineages amongst the Baltis are shared with other Tibetan communities, whereas the prevalent Y chromosomal haplogroups probably entered Baltistan during the introduction of Islam (Zerjal et al. 1997, Quintana-Murci et al. 2001, Qamar et al. 2002).⁴

At the same time, a jarring disconnect is sometimes seen between the occurrence of a highly salient genetic marker and the linguistic affinity of a community's language. Hungarians lack the TatC deletion defining the Y chromosomal

⁴ Ironically, the Balti call their language ཨ་སྐད་ *phaskat* 'father tongue' (Roland Bielemeier, personal communication, 10 September 2012), just as they call their homeland ཨ་ཡུལ་ *phayul* 'father-realm' and birthplace ཨ་སྤ་ *phasa* 'father-land' (Sprigg 2002: 127).

haplogroup N1c,⁵ despite the sheer prevalence of this marker amongst all other Uralic language communities (Lì et al. 1999). So, it deserves to be repeated that the linguistic ancestors of a language community were not necessarily the same people as the biological ancestors of that community. In fact, some of them could not have been the same people.

It also merits repeating that the time depth accessible to population geneticists studying polymorphisms on the genome is vastly greater than the reach of the linguistically reconstructible past. The wave of anatomically modern humans who introduced the proto-languages that were later to give rise to today's Asian linguistic phyla and language isolates can be dated to between 25,000 to 38,000 years ago (Rasmussen et al. 2011), and the antiquity of Y chromosomal haplogroups such as O1 or O2 has been calculated to be greater than 10,000 years (Yan et al. 2011). Historical linguists, on the other hand, generally estimate the linguistically reconstructible past to be shallower than 10,000 years. This temporal gap must temper and inform all speculations regarding correlations between linguistic and genetic affinity.

With such caveats in place, how can we address the question formulated at the beginning of this section? On the 28th of June 2006, at a symposium held at the École Française d'Extrême-Orient at Siem Reap, I identified the Y chromosomal haplogroup O2a (M95) as the marker for the spread of Austroasiatic on the basis of the then available genetic data (later published in van Driem 2007b). This view has been corroborated by subsequent genetic studies, e.g. Kumar et al. (2007), Chaubey et al. (2010). In the latter article, we concluded that Austroasiatic speakers in India today are derived from a dispersal from Southeast Asia, followed by extensive sex-specific admixture with local populations indigenous to the Subcontinent.

The autosomal data also reflect the distinction between two components in the genome, one represented by the predominantly indigenous maternal lineages and the other by the intrusive paternal O2a lineage that correlates with the linguistic affinity of the Austroasiatic language communities in the Indian subcontinent. These findings go well beyond Robert von Heine-Geldern's model of a Southeast Asian homeland and envisage a father tongue spread of Austroasiatic, borne to the Indian subcontinent by predominantly male speakers from mainland Southeast Asia, but also involving a complex sociolinguistic prehistory of bidirectional gene flow across the Bay of Bengal (Chaubey et al. 2010). In many parts of the world, the mitochondrial DNA lineages often appear preponderantly to reflect older resident maternal lineages.

5 The 2008 Y Chromosome Consortium haplogroup labels are used here.

to the reconstruction of linguistic population prehistory in eastern Eurasia, operating on the assumption of the veracity of the Father Tongue hypothesis.

The available genetic data also enabled us to identify a correlation of the Y chromosomal haplogroup O3a3b (M7) with the spread of Hmong-Mien, whilst our genetic samplings throughout the Himalayan region had established a correlation between Trans-Himalayan and the paternal lineage O3a3c (M134) (Parkin et al. 2006, 2007, Kraaijenbrink et al. 2007a, 2007b, 2009, van Driem 2011). The Y chromosomal haplogroup O is becoming ever more minutely mapped, and most recently the phylogenetic positions of mutations P164 and PK4 within the haplogroup have been revised (Yan et al. 2011). Yet the antiquity calculated for many of these mutations is generally greater than the time depth that most historical linguists are willing to ascribe to the major language phyla.

Let us venture into the twilight beyond the linguistically reconstructible past to a time just after the Last Glacial Maximum, when the Y chromosomal haplogroup O (M175) had split up into the subclades O1 (M119), O2 (M268) and O3 (M122). Based on what is known about linguistic phylogeny and about the geographical distribution of modern linguistic communities today, the three subclades can putatively be assigned to three geographical loci along an east-west axis. For the sake of argument and schematic representation, and without any claim to geographical precision or veracity, I shall assign the haplogroup O1 (M119) to the drainage of the Pearl River and its tributaries in what today is the Chinese province of Guǎngdōng. I shall situate haplogroup O2 (M268) in southern Yúnnán and O3 (M122) to the area where today's northeastern India, southeastern Tibet and northern Burma adjoin.

Since we have associated O2a (M95), which is a derivative clade of haplogroup O2 (M268), with the Austroasiatic language phylum, we might conjecture that Asian rice, perhaps both *japonica* and *indica* rice, was first domesticated roughly in the general area hypothetically imputed to O2 (M268) here. Whilst the bearers of the O2a (M95) haplogroup became the *Stammväter* of the Austroasiatics, the other derivative paternal subclade O2b (M176) spread eastward, where they introduced rice agriculture to the areas south of the Yangtze. Though the bearers of the O2b (M176) haplogroup continued to sow seed as they continued to move ever further eastward, they left little or no linguistic traces, except maybe an Austroasiatic name for the Yangtze river, as proposed by Pulleyblank (1993), reflected as the toponym borrowed by Old Chinese as 江 *kʰron (jiāng).

Meanwhile, back in southern Yúnnán, the early Austroasiatics spread from this locus initially to the Salween drainage in northeastern Burma and to the area that today is northern Thailand and western Laos. In time, the Austroasiatics would spread as far as the Mekong delta, the Malay peninsula, the Nicobars and later even into eastern India, where they would introduce both their language and their paternal lineage to indigenous peoples of the Subcontinent.

At the locus putatively assigned to the haplogroup O3 (M122), the bearers of this marker gave rise to the paternal lineages O3a3c (M134) and O3a3b (M7). Whilst the bearers of the polymorphism O3a3c (M134) stayed behind in the area comprising northeastern India, southeastern Tibet and northern Burma, the bearers of the O3a3b (M7) paternal lineage migrated eastward to settle in the areas south of the Yangtze. On their way, the early Hmong-Mien encountered the ancient Austroasiatics, from whom they adopted rice agriculture. The intimate interaction between ancient Austroasiatics and the early Hmong-Mien not only involved the sharing of knowledge about rice agriculture technology, but also left a genetic trace in the high frequencies of haplogroup O2a (M95) in today's Hmong-Mien and of haplogroup O3a3b (M7) in today's Austroasiatic populations.

On the basis of these Y chromosomal haplogroup frequencies, Cai et al. observed that Austroasiatics and Hmong-Mien “are closely related genetically” and ventured to speculate about “a Mon-Khmer origin of Hmong-Mien populations” (2011: 8). More precisely, the incidence of haplogroup O3a3b (M7) in Austroasiatic language communities of Southeast Asia appears to indicate a significant Hmong-Mien paternal contribution to the early Austroasiatic populations whose descendants settled in Southeast Asia, whereas the incidence of haplogroup O3a3b (M7) in Austroasiatic communities of the Indian subcontinent is undetectably low. The incidence of haplogroup O2a amongst the Hmong-Mien appears to indicate a slightly more modest Austroasiatic paternal contribution to Hmong-Mien populations than vice versa.

As the Hmong-Mien moved eastward, the bearers of haplogroup O2b (M176) likewise continued to move east. Even further east, the O1 (M119) paternal lineage gave rise to the O1a (M119) subclade, which moved from the Pearl River drainage eastward to the Mǐn river drainage in the hill tracts of Fújiàn province and across the strait to Formosa, which consequently became the *Urheimat* of the Austro-nesians. Back west in the easternmost spurs of the Himalayas, the bearers of Y chromosomal haplogroup O3a3c (M134) expanded eastward into Sìchuān and Yúnnán, north and northwest across the Tibetan plateau as well as westward into the Himalayas and southward into the Indo-Burmese borderlands. In the west and south, the early Trans-Himalayans encountered Austroasiatics, who had preceded them.

Linguistic research on Trans-Himalayan languages can inform a chronologically layered view of ethnolinguistic prehistory. Not only do historical linguistics and genetics present two distinct and independent windows on the past. Even on a logarithmically distorted time scale the time depth accessible to historical linguistics can be seen to be far shallower than the prehistorical depth accessible to human population genetics. The human population genetic data from beyond the linguistically reconstructible past embolden us to speculate that there

must have been an early eastward and northward spread into East Asia, possibly including the linguistic ancestors of modern Trans-Himalayan language communities, who may have been the first bearers of the Y chromosomal haplogroup O3a3c (M134). After this post-glacial colonisation, there must have been a number of discrete expansions in different directions at different times in the past.

To recapitulate the chronology of possible movements: (1) a post-glacial northward wave of peopling at a time depth beyond what is generally held to be linguistically reconstructible by historical linguists, (2) a northeasterly spread of a subset of the ancient Trans-Himalayans to the putative early locus of Sino-Bodic, (3) incremental spread of diverse ancient Trans-Himalayan groups throughout the Himalayas, where there appears to be both linguistic and genetic evidence of pre-Trans-Himalayan speaking populations, (4) a southward spread of Sino-Bodic, suggested by archaeology, genes and language, bringing Sino-Bodic groups, including Sinitic, into contact with the ancient Hmong-Mien, the early Austroasiatics, the Austronesians and a number of other Trans-Himalayan groups, (5) a Bodic spread across the Tibetan plateau spilling over into the Himalayas, as evinced by the distribution of Bodish, East Bodish, Tamangic, West Himalayish and several other groups, and (6) the spread of Trans-Himalayan groups from Yúnnán into Southeast Asia, e.g. Karen, Pyu and later Lolo-Burmese.

Following these tentatively reconstructed prehistoric stages of peopling, there were the historically attested ethnolinguistic dispersals: (7) the historically documented Hàn spread, clearly evinced in linguistics and genetics, probably assimilating non-Trans-Himalayan as well as other Trans-Himalayan groups, and (8) the historically documented spread of Bodish (i.e. Tibetic) across the Tibetan plateau.

The relative frequencies of the Y chromosomal haplogroup O2a (M95) in various Trans-Himalayan populations of the Indian subcontinent (Sahoo et al. 2006, Reddy et al. 2007) suggest that a subset of the paternal ancestors of particular Trans-Himalayan populations in northeastern India, e.g. certain Bodo-Koch communities, may originally have been Austroasiatic speakers who married into Trans-Himalayan communities or were linguistically assimilated by ancient Trans-Himalayans. At the same time, median-joining network analyses of haplogroup O2a (M95) microsatellites have suggested a division in the Indian subcontinent between Trans-Himalayans vs. Austroasiatic and Dravidian language communities. Austroasiatics and Dravidians show greater Y chromosomal microsatellite diversification than Trans-Himalayan language communities, and the highest frequency of the O2a haplogroup is found in tribal populations in Orissa, Chattisgarh and Jharkhand (Sengupta et al. 2006).

We must bear in mind that Y haplogroups are subject to selection and that frequencies change over time. As stressed above, haplotype frequencies by them-

selves are not a sufficient criterion. A rooted topology of the Y chromosomal tree and its subsidiary clades provides key evidence. Moreover, the ethnolinguistic significance of paternal lineages becomes even more manifest when other portions of the genome are scoured for correlations with linguistic phylogeography. At the same time, our understanding of what constitutes neutral diversity has been tempered by mathematical modelling. Simulations have shown that a normally low-frequency allele could surf on a demic wave of advance and so attain high frequency across a vast area. Gene surfing during a spatial expansion is likely to result in distinct geographical sectors of low genetic diversity separated by sharp allele frequency gradients.

The result of recurrent bottleneck effects during range expansion into newly colonised territories can mimic complex phylogeographical patterns of adaptation and segregation into clades in post-glacial niche refugia. Likewise, the massive introgression of resident genes into the incursive population can also be misinterpreted as the result of a selective process (Excoffier and Ray 2008, Excoffier et al. 2009). Surfing on the crest of a demic wave of expansion confers a selective advantage when compared to alleles left behind in the core area (Klopfstein et al. 2006, Moreau et al. 2011). Both the dynamics of sex-biased dispersals as well as the process of the sexually asymmetrical introgression of resident alleles into incursive populations can be modelled in terms of hybridisation during range expansions (Petit and Excoffier 2009, Currat and Excoffier 2011).

An observed state of affairs for which a particular model of population prehistory has been advanced may in many cases very well be either the result of demography or of selection on genome diversity (Fagundes et al. 2007). However, we must keep in mind that a scenario that has been computed to be the statistically more likely scenario may not necessarily correspond to prehistorical reality. Though presumably paternal lineages may often preferentially enjoy the benefits of surfing, incursive Y chromosomal lineages can go entirely extinct, as the linguistic evidence⁶ would suggest may very well have happened with the Y chromosomal haplogroup N1c in Hungary.

We must also not lose sight of the fact that these speculations are based on correlations between language and Y chromosomal haplogroups and that these too are interpreted in the light of the assumed veracity of the Father Tongue hypothesis over a vast stretch of time. This assumption may not hold true for all times in the past. Furthermore, correlations may be due to different kinds of cir-

⁶ The presence of the Hungarian language in the region that was once Pannonia represents incontrovertible linguistic evidence of the advent of Uralic linguistic ancestors, a fact which is historically attested at any rate, but the hypothetical correlation of the Y chromosomal haplogroup N1c with the Uralic linguistic phylum, of course, remains conjectural.

cumstances other than causation or direct relationship. So, whilst we are free cautiously to develop arguments which buttress a speculative model of ethno-linguistic prehistory, such as the one outlined here, we must not lose sight of the essential distinction between the facts and our assumptions and inferences as well as the precise nature and limitations of the empirical basis for our speculations.

Confronted with the overwhelming growing body of evidence in support of the Father Tongue hypothesis, Forster and Renfrew impute the spread of language families to “emigrating agriculturalists” who “took local wives” (2011: 1391). This interpretation is a transparent attempt to succour Bellwood and Renfrew’s embattled First Farmers hypothesis, which seeks to ascribe the founding dispersals of language families to the spread of agriculture (Bellwood and Renfrew 2002). At the same time, in order to buttress Renfrew’s widely doubted hypothesis of an Indo-European homeland in Asia Minor, Forster and Renfrew also propose a correlation of Indo-European with the Y chromosomal haplogroup J2a. In fact, it remains moot whether any part of Y chromosomal phylogeography correlates well with the spread of the Neolithic horizon.

Not every population movement led to the spread of a language phylum, and population movements are not uniform in nature. Whether during the exodus of anatomically modern humans out of Africa or at the shallow time depth of the colonisation of Oceania by Austronesian populations, the colonisation of previously uninhabited lands invariably involved both sexes and the introduction of a language phylum. During the Neolithic horizon, the spread of farming was necessarily a sedentary and incremental process, which likewise must mostly have involved both sexes. Early farmers might only have been able to spread their language at times of great surplus and concomitant population growth, perhaps sometimes involving the establishment of agricultural colonies elsewhere. By contrast, the modern ethnolinguistic composition of Asian populations must be understood, at least in part, as having resulted from male-biased linguistic intrusions, whether motivated by conquest, land grab or the urge to seek out new habitats.

In my argument against the premises and the reasoning behind the hypothesis of the founding dispersals of language phyla having been mediated by the spread of farming, I proposed the telic and more complex Centripetal Migration theory (van Driem 2007b). I shall not repeat that exposition here, but, with reference to Forster and Renfrew’s wilful interpretation of the Y chromosomal haplogroup J2, I shall reiterate that, in the context of the Indian subcontinent, “the J2 haplogroup... appears to emanate from the Arabian Peninsula and, unlike haplogroups N and R1a, attains no high frequency in Ceylon” and “probably reflects the historically attested male-borne eastward spread of Islam,” whereas Y chro-

mosomal haplogroups of the R subclades spread to the Subcontinent “from the northwest along with Indo-Aryan language across northern India and to Ceylon” (van Driem 2007b: 5). The spread of various Y chromosomal R subclades is likely to be linked to the dispersal of Indo-European from an original homeland in the Pontic-Caspian steppe, whilst the current geographical distribution of the Y chromosomal lineage L provides the likeliest candidate for a vestige of an earlier patrilingual dispersal of Elamo-Dravidian emanating from a region which encompassed the Bactria and Margiana of later prehistory.

References

- Beames, John. 1868. *Outlines of Indian Philology*. London: Trübner and Company.
- Bellwood, Peter, and Colin Renfrew (eds.). 2002 *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- Benedict, Paul King. 1942. Thai, Kadai, and Indonesia: A new alignment in southeastern Asia. *American Anthropologist* 44: 576–601.
- Benedict, Paul King. 1972. *Sino-Tibetan: a Conspectus*. Cambridge: Cambridge University Press.
- Benedict, Paul King. 1976. Sino-Tibetan: another look. *Journal of the American Oriental Society* 96 (2): 167–197.
- Blench, Roger Marsh. 2009. If agriculture cannot be reconstructed for Proto-Sino-Tibetan, what are the consequences? Paper presented at the 42nd Conference on Sino-Tibetan Languages and Linguistics at Chiang Mai, 3 November 2009.
- Bó Wénzé. 2004. Cǎijǐàhuà gàikuàng. *Mínzú Yǔwén* 2004 (2): 68–81.
- Bodman, Nicholas Cleaveland. 1973. Some Chinese reflexes of Sino-Tibetan s- clusters. *Journal of Chinese Linguistics* 1 (3): 383–396.
- Bodman, Nicholas Cleaveland. 1980. Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. In Frans van Coetsem and Linda R. Waugh (eds.). *Contributions to Historical Linguistics: Issues and Materials*, 34–199. Leiden: E.J. Brill.
- Boller, Anton. 1857. Nachweis, dass das Japanische zum ural-altaischen Stamme gehört. *Sitzungsberichte der philosophisch-historischen Classe der kaiserlichen Akademie der Wissenschaften, Wien*, 23: 393–481.
- Bradley, David. 2012. Proto-Tibeto-Burman grain crops. *Rice* 4 (3–4): 134–141.
- Byrne, James. 1885. *General Principles of the Structure of Language* (2 volumes). London: Trübner and Company.
- Cai Xiaoyun, Zhendong Qin, Bo Wen, Shuhua Xu, Yi Wang, Yan Lu, Lanhai Wei, Chuanchao Wang, Shilin Li, Xingqiu Huang, Huí Lǐ and the Genographic Consortium. 2011. Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *Public Library of Science* 6 (8): e24282.
- Chalmers, John. 1866. *The Origin of the Chinese: An Attempt to Trace the Connection of the Chinese with Western Nations in their Religion, Superstitions, Arts, Language and Traditions*. Hong Kong: De Souza & Co.

- Chaubey, Gyaneshwer, Mait Metspalu, Ying Choi, Reedik Mägi, Irene Gallego Romero, Siiri Rootsi, Pedro Soares, Mannis van Oven, Doron M. Behar, Siiri Rootsi, Georgi Hudjashov, Chandana Basu Mallick, Monika Karmin, Mari Nelis, Jüri Parik, Alla Goverdhana Reddy, Ene Metspalu, George van Driem, Yali Xue, Chris Tyler-Smith, Kumarasamy Thangaraj, Lalji Singh, Maito Remm, Martin B. Richards, Marta Mirazon Lahr, Manfred Kayser, Richard Villems and Toomas Kivisild. 2010. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution*, 28(2): 1013–1024.
- de Condillac, Étienne Bonnot. 1746. *Essai sur l'origine des connoissances humaines: Ouvrage où l'on réduit à un seul principe tout ce qui concerne l'Entendement Humain* (2 volumes). Amsterdam: Pierre Mortier.
- Conrad, August. 1896. *Eine indochinesische Causativ-Denominativ-Bildung und ihr Zusammenhang mit den Tonaccenten: Ein Beitrag zur vergleichenden Grammatik der indochinesischen Sprachen, insonderheit des Tibetischen, Barmanischen, Siamesischen und Chinesischen*. Leipzig: Otto Harrassowitz.
- Curat, Mathias, and Laurent Excoffier. 2011. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proceedings of the National Academy of Sciences* 108 (37): 15129–15134.
- Cust, Robert N. 1878. *A Sketch of the Modern Languages of East India*. London: Trübner and Company.
- van Driem, George. 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies* 60 (3): 455–488.
- van Driem, George. 2001. *Languages of the Himalayas: An Ethnolinguistic Handbook of the Greater Himalayan Region, containing an Introduction to the Symbiotic Theory of Language* (2 volumes). Leiden: Brill.
- van Driem, George. 2002. The Father Tongue hypothesis: Sexually dimorphic dispersals in prehistory. Paper presented at the 17th conference of the Indo-Pacific Prehistory Association, held at the Academia Sinica, Taipei, 9–15 September 2002.
- van Driem, George. 2005. Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In Yogendra Prasada Yadava, Govinda Bhattarai, Ram Raj Lohani, Balaram Prasain and Krishna Parajuli (eds.). *Contemporary Issues in Nepalese Linguistics*, 285–338. Kathmandu: Linguistic Society of Nepal.
- van Driem, George. 2007a. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1 (2): 211–270.
- van Driem, George. 2007b. Austroasiatic phylogeny and the Austroasiatic homeland in light of recent population genetic studies. *Mon-Khmer Studies* 37: 1–14.
- van Driem, George. 2011. Rice and the Austroasiatic and Hmong-Mien homelands. In Nick J. Enfield (ed.). *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, 361–389. Canberra: Pacific Linguistics.
- van Driem, George. 2012. The ethnolinguistic identity of the domesticators of Asian rice. *Comptes Rendus Palevol*, 11 (2): 117–132.
- Dulik, Matthew C., Sergey I. Zhadanov, Ludmila P. Osipova, Ayken Askapuli, Lydia Gau, Omer Gokcumen, Samara Rubinstein and Theodore G. Schurr. 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between native Americans and indigenous Altaians. *American Journal of Human Genetics* 90: 1–18.
- Excoffier, Laurent, Matthieu Foll and Rémy J. Petit. 2009. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution and Systematics* 40: 481–501.

- Excoffier, Laurent and Nicolas Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution* 23 (7): 347–351.
- Fagundes, Nelson J. R., Nicolas Ray, Mark Beaumont, Samuel Neuenschwander, Francisco M. Salzano, Sandro L. Bonatto and Laurent Excoffier. 2007. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences* 104 (45): 17614–17619.
- Forbes, Charles James F. S. 1878. On Tibeto-Burman languages. *Journal of the Royal Asiatic Society of Great Britain and Ireland* 10: 210–227.
- Forchhammer, Emile. 1882. Indo-Chinese languages. *Indian Antiquary* 11: 177–189.
- Forrest, R. A. D. 1956. Towards common Tibeto-Burman. *Wennti* 10: 1–18.
- Forrest, R. A. D. 1962. The linguistic position of Róng (Lepcha). *Journal of the American Oriental Society* 82 (3): 331–335.
- Forster, Peter, and Colin Renfrew. 2011. Mother tongue and Y chromosomes. *Science* 333: 1390–1391.
- von der Gabelentz, Hans Georg Conon. 1881. *Chinesische Grammatik mit Ausschluss des niederen Stiles und der heutigen Umgangssprache*. Leipzig: T.O. Weigel.
- de Gobineau, Joseph Arthur. 1854–5. *Essai sur l'inégalité des races humaines* (4 volumes). Paris: Firmin-Didot et Compagnie.
- Grace, George. 1989. *The Linguistic Construction of Reality*. London: Croom Helm.
- Hill, Nathan W. 2011. Multiple origins of Tibetan. *Language and Linguistics* 12 (3): 707–721.
- Hodgson, Brian Houghton. 1857. Comparative Vocabulary of the Languages of the broken Tribes of Népál. *Journal of the Asiatic Society of Bengal* 26: 317–371.
- Houghton, Bernard. 1896. Outlines of Tibeto-Burman linguistic palæontology. *Journal of the Royal Asiatic Society* 1896: 23–55.
- von Humboldt, Wilhelm. 1822. Ueber das vergleichende Sprachstudium in Beziehung auf die verschiedenen Epochen der Sprachentwicklung. *Abhandlungen der Königl. Preussischen Akademie der Wissenschaften, historisch-philosophische Klasse*, aus den Jahren 1820 und 1821: 239–260.
- von Humboldt, Wilhelm. 1825. Über das Entstehen der grammatischen Formen und ihren Einfluß auf die Ideenentwicklung. *Abhandlungen der Königl. Preussischen Akademie der Wissenschaften, historisch-philosophische Klasse*, aus den Jahren 1822 und 1823: 401–430.
- von Humboldt, Wilhelm. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Berlin: Königl. Akademie der Wissenschaften.
- Jacques, Guillaume and Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28 (4): 468–499.
- Karefet, Tatiana M., Fernando L. Mendez, Monica B. Meilerman, Peter Underhill, Stephen L. Zegura and Michael F. Hammer. 2008. New binary polymorphisms reshape and increase resolution of the Y chromosomal haplogroup tree. *Genome Research* 18: 830–838.
- Karlgren, Bernhard. 1920. Le proto-chinois, langue flexionnelle. *Journal Asiatique* (XI^e Série) 40: 205–232.
- Karlgren, Bernhard. 1957. *Grammata Serica Recensa* (i.e. *Bulletin of the Museum of Far Eastern Antiquities* 29: 1–332).
- von Klaproth, Julius Heinrich. 1823. *Asia Polyglotta*. Paris: A. Schubart.
- Klopfstein, Seraina, Mathias Currat and Laurent Excoffier. 2006. The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* 23 (3): 482–490.

- Kraaijenbrink, Thirsa, Emma J. Parkin, Denise R. Carvalho-Silva, George van Driem, Guido Barbujani, Chris Tyler-Smith, Mark A. Jobling and Peter de Knijff. 2009. Genetic and linguistic borders in the Himalayan region. In Francesco d'Errico and Jean-Marie Hombert (eds.). *Becoming Eloquent: Advances in the Emergence of Language, Human Cognition and Modern Cultures*, 181–201. Amsterdam: John Benjamins.
- Kraaijenbrink, Thirsa, Jean Robert M.L. Opgenort, Nirmal Man Tuladhar, George van Driem and Peter de Knijff. 2007. Allele frequency distribution for 21 autosomal STR loci in Nepal. *Forensic Science International* 168 (2–3): 227–231.
- Kraaijenbrink, Thirsa, Karma Tshering of Gaselô and Peter de Knijff. 2007. Allele frequency distribution for 21 autosomal STR loci in Bhutan. *Forensic Science International* 170: 68–72.
- Kuhn, Ernst. 1883. *Ueber Herkunft und Sprache der transgangetischen Völker. Festrede zur Vornehmsten Geburt- und Namensfestes Seiner Majestät des Königs Ludwig II., gehalten in der öffentlichen Sitzung der Königlichen Akademie der Wissenschaften zu München am 25. Juli 1881*. München: Verlag der Königlichen Bayerischen Akademie.
- Kuhn, Ernst. 1889. Beiträge zur Sprachenkunde Hinterindiens. *Sitzungsberichte der Königlichen Bayerischen Akademie der Wissenschaften* (München), Philosophisch-philologische Classe II: 189–236.
- Kumar, Vikrant, Arimanda N.S. Reddy, Jagadeesh P. Babu, Tiprisetti N. Rao, Banrida T. Langstieh, Kumarasamy Thangaraj, Alla G. Reddy, Lalji Singh and Battini M. Reddy. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BioMed Central Evolutionary Biology*, 7: 47 (14 pp.) www.biomedcentral.com/1471-2148/7/47
- Lambert, Pierre-Yves. 1994. *La langue gauloise*. Paris: Éditions Errance.
- Laufer, Berthold. 1916. The Si-hia Language, a Study in Indo-Chinese Philology. *T'oung Pao* 17: 1–126.
- Lepsius, Carl Richard. 1861. Über die Umschrift und Lautverhältnisse einiger hinterasiatischer Sprachen, namentlich der Chinesischen und der Tibetischen. *Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin aus dem Jahre 1860*, 449–496.
- Leyden, John Casper. 1808. On the languages and literature of the Indo-Chinese nations. *Asiatic Researches* 10: 158–289.
- Lì Jīn, and Peter A. Underhill, Vishal Doctor, Ronald W. Davis, Peidong Shen, L. Luca Cavalli-Sforza and Peter J. Oefner. 1999. Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proceedings of the National Academy of Sciences of the United States of America* 96: 3796–3800.
- Locke, John. 1690. *An Essay concerning Humane Understanding, in Four Books*. London: Tho. Basset.
- Maine de Biran, François-Pierre-Gonthier. [1815] 1932. Note sur les réflexions de Maupertuis et de Turgot au sujet de l'origine des langues. In Pierre Tisserand (ed.). *Œuvres de Maine de Biran* (14 volumes), vol. 10, 315–338. Paris: Librairie Félix Alcan.
- Malte-Brun, Conrad [posthumous]. 1832. *Précis de la géographie universelle, ou description de toutes les parties du monde sur un plan nouveau, d'après les grandes divisions naturelles du globe: Nouvelle édition, revue, corrigée, mise dans un nouvel ordre et augmentée de toutes les nouvelles découvertes par J.J.N. Huot* (6 volumes) Bruxelles: Th. Lejeune.
- Matisoff, James Alan. 2000. On “Sino-Bodic” and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63 (3): 356–369.

- de Maupertuis, Pierre Louis Moreau. 1756a [1748]. *Réflexions philosophiques sur l'origine des langues et la signification des mots*. In *Œuvres* (nouvelle édition corrigée et augmentée, 4 volumes), vol. 1, 258–309. Lyon: Jean-Marie Bruyset.
- de Maupertuis, Pierre Louis Moreau. 1756b. *Dissertation sur les différents moyens dont les hommes se sont servis pour exprimer leurs idées*. In *Œuvres* (nouvelle édition corrigée et augmentée, 4 volumes), vol. 3, 435–468. Lyon: Jean-Marie Bruyset.
- Moreau, Claudia, Claude Bhérer, Hélène Vézina, Michèle Jomphe, Damian Labuda and Laurent Excoffier. 2011. Deep human genalogies reveal a selective advantage to being on an expanding wave front. *Science* 334: 1148–1150.
- Müller, Friedrich Max. 1871. *Lectures on the Science of Language* (2 volumes). London: Longmans, Green and Company.
- Müller, Friedrich Max. 1881. *Selected Essays on Language, Mythology and Religion* (2 volumes). London: Longmans, Green and Company.
- Parkin, Emma J., Thirsa Kraaijenbrink, Jean Robert M.L. Opgenort, Nirmal Man Tuladhar, George van Driem, Peter de Knijff and Mark Jobling. 2007. Diversity of 26 Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas. *Forensic Science International* 166 (2–3): 176–181.
- Parkin, Emma J., Thirsa Kraaijenbrink, Karma Tshering of Gaselô, Peter de Knijff and Mark Jobling. 2006. 26-locus Y-STR typing in a Bhutanese population sample. *Forensic Science International* 161 (1): 1–7.
- Petit, Rémy J., and Laurent Excoffier. 2009. Gene flow and species delimitation. *Trends in Ecology and Evolution* 24 (7): 386–393.
- Poloni, Estella Simone, Ornella Semino, Giuseppe Passarino, A.S. Santachiara-Benerecetti, I. Dupanloup, André Langaney and Laurent Excoffier. 1997. Human genetic affinities for Y chromosome P49a,f/*TaqI* haplotypes show strong correspondence with linguistics. *American Journal of Human Genetics* 61: 1015–1035 (cf. the erratum published in 1998 in the *American Journal of Human Genetics* 62: 1267).
- Poloni, Estella Simone, Nicolas Ray, Stefan Schneider and André Langaney. 2000. Languages and genes: Modes of transmission observed through the analysis of male-specific and female-specific genes. In Jean-Louis Dessalles and Laleh Ghadakpour (eds.), *Proceedings: Evolution of Language, 3rd International Conference 3–6 April 2000*, 185–186. Paris: École Nationale Supérieure des Télécommunications.
- Post, Mark William, and Roger Marsh Blench. 2011. Siangic: A new language phylum in North East India. Paper presented at the 6th International Conference of the North East Indian Linguistics Society, Tezpur University, 31 January to 2 February 2011.
- Pott, August Friedrich. 1856. *Die Ungleichheit menschlicher Rassen hauptsächlich vom sprachwissenschaftlichen Standpunkte*. Lemgo und Detmold: Meyer'sche Hofbuchhandlung.
- Przyłuski, Jean. 1924. Le sino-tibétain [and] Les langues austroasiatiques. In Antoine Meillet and Marcel Cohen (eds.), *Les Langues du Monde*, 361–384 [and] 385–403. Paris: Librairie Ancienne Édouard Champion.
- Przyłuski, Jean and Gordon Hannington Luce. 1931. The number “a hundred” in Sino-Tibetan. *Bulletin of the School of Oriental Studies* 6 (3): 667–668.
- Pulleyblank, Edwin George. 1983. The Chinese and their neighbours in prehistoric and early historic times. In David N. Keightley (ed.), *The Origins of Chinese Civilization*, 411–466. Berkeley: University of California Press.

- Qamar, Raheel, Qasim Ayub, Aisha Mohyuddin, Agnar Helgason, Kehkashan Mazhar, Atika Manoor, Tatiana Zerjal, Chris Tyler-Smith and S. Qasim Mehdi. 2002. Y-chromosomal variation in Pakistan. *American Journal of Human Genetics* 70: 1107–1124.
- Quintana-Murci, Luíís, C. Krausz, Tatiana Zerjal, S. Hamid Sayar, Michael F. Hammer, S. Qasim Mehdi, Qasim Ayub, Raheel Qamar, Aisha Mohyuddin, U. Radhakrishna, Mark A. Jobling Chris Tyler-Smith and Ken McElreavey. 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *American Journal of Human Genetics* 68: 537–542.
- Rasmussen, Morten, Xiaosen Guo, Yong Wang, Kirk E. Lohmueller, Simon Rasmussen, Anders Albrechtsen, Line Skotte, Stinus Lindgreen, Mait Metspalu, Thibaut Jombart, Toomas Kivisild, Weiwei Zhai, Anders Eriksson, Andrea Manica, Ludovic Orlando, Francisco de la Vega, Silvano Tridico, Ene Metspalu, Kasper Nielsen, María C. Ávila-Arcos, J. Víctor Moreno-Mayar, Craig Muller, Joe Dortch, M. Thomas P. Gilbert, Ole Lund, Agata Wesolowska, Monika Karmin, Lucy A. Weinert, Bo Wang, Jun Li, Shuaishuai Tai, Fei Xiao, Tsunehiko Hanihara, George van Driem, Aashish R. Jha, François-Xavier Ricaut, Peter de Knijff, Andrea B. Migliano, Irene Gallego-Romero, Karsten Kristiansen, David Lambert, Søren Brunak, Peter Forster, Bernd Brinkmann, Olaf Nehlich, Michael Bunce, Michael Richards, Ramneek Gupta, Carlos Bustamante, Anders Krogh, Robert A. Foley, Marta Mirazón Lahr, François Balloux, Thomas Sicheritz-Pontén, Richard Villems, Rasmus Nielsen, Wang Jun, and Eske Willerslev. 2011. An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334 (6052): 94–98.
- Reddy, B. Mohan, B.T. Langstieh, Vikrant Kumar, T. Nagaraja, A.N.S. Reddy, M. Aruna, K. Thangaraj, A.G. Reddy and Lalji Singh. 2007. Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. *Public Library of Science One* 2 (11): e1141. doi:10.1371/journal.pone.0001141
- Renan, Ernest. 1858. *De l'origine du langage* (deuxième édition, revue et considérablement augmentée). Paris: Michel Lévy, Frères, Librairies-Éditeurs.
- Ritter von Xylander, Joseph. 1835. *Die Sprache der Albanesen oder Schkipetaren*. Frankfurt am Main: die Andreäische Buchhandlung.
- Sahoo, Sanghamitra, Anamika Singh, G. Himabindu, Jheeman Banerjee, T. Sitalaximi, Sonali Gaikwad, R. Trivedi, Phillip Endicott, Toomas Kivisild, Mait Metspalu, Richard Villems and V. K. Kashyap. 2006. A prehistory of Indian Y chromosomes: Evaluating demic diffusion scenarios. *Proceedings of the National Academy of Sciences of the United States of America* 103 (4): 843–848.
- Schott, Wilhelm. 1856. Über die sogenannten indo-chinesischen Sprachen insonderheit das Siamische. *Abhandlungen der Königlischen Akademie der Wissenschaften zu Berlin aus dem Jahre 1856*, Philosophisch-historische Klasse: 161–179.
- Sengupta, Sanghamitra, Lev A. Zhivotovsky, Roy King, S.Q. Mehdi, Christopher A. Edmonds, Cheryl-Emiliane T. Chow, Alice A. Lin, Mitashree Mitra, Samir K. Sil, A. Ramesh, M.V. Usha Rani, Chitra M. Thakur, L. Luca Cavalli-Sforza, Partha P. Majumder and Peter A. Underhill. 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *The American Journal of Human Genetics* 78: 202–221.
- Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word, Journal of the Linguistic Circle of New York* 11: 94–111.
- von Siebold, Philipp Franz Balthazar. 1832. Verhandeling over de afkomst der Japanners. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 13: 183–275.