



# PACIFIC LINGUISTICS

Research School of Pacific and Asian Studies

THE AUSTRALIAN NATIONAL UNIVERSITY

## PDF file

This .pdf (portable document file) document is an image file produced from a scan of an original copy of the book. It may not be reproduced without the permission of *Pacific Linguistics*.

*Pacific Linguistics* Home Page: <http://www.pacling.com>

---

### Language variation:

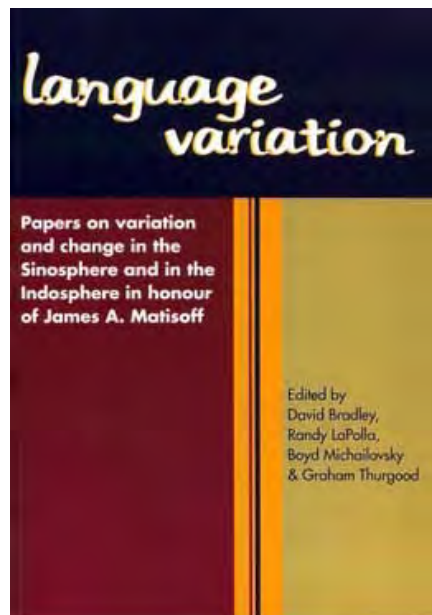
#### **Papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff**

edited by David Bradley, Randy LaPolla, Boyde Michailovsky and Graham Thurgood  
PL 555

2003 ISBN 085883 333 pp

Price: International: \$28.50, Australia \$31.35 incl. GST

---



This volume discusses the nature of variation and change in a number of East, Southeast and South Asian languages, especially of the Sino-Tibetan family, also extending to other languages, even as far afield as English. The papers honour the work of James A. Matisoff, in celebration of his 65th birthday.

There are nineteen papers by twenty authors concerning issues in phonology, morphology, syntax, language contact, orthography and language documentation.

Randy LaPolla provides a paper with broad theoretical implications, 'Why languages differ: variation in the conventionalisation of constraints on inference'.

Martha Ratliff writes on Hmong secret languages. Graham Thurgood and Fengxiang Li give an account of contact-induced variation and syntactic change in the Austronesian Tsat language of Hainan. Benji Wald's contribution considers verb compounding in English and East Asian languages.

The other papers in the volume concern Sino-Tibetan languages. Balthasar Bickel writes on prosodic tautomorphemicity in Sino-Tibetan word structure. Robert Bauer discusses the impact of English loanwords on the Hong Kong Cantonese syllabary. Eleven papers are on Tibeto-Burman topics. David Bradley discusses deictic patterns in Lisu and SE Tibeto-Burman. Jackson Sun describes tonal developments in Tibetan, while Michel Ferlus writes on borrowing from Middle Chinese into Proto Tibetan. Yasuhiko Nagano describes negation particles in Gyarong (Sichuan) and David Peterson agreement and grammatical relations in Hyow (Bangladesh). Jerold Edmondson provides data from Phù Lá, Xá Phó, Lô Lô, all located in Vietnam.

Five of the Tibeto-Burman papers concern languages of Nepal. Michael Noonan writes on recent language contact among Tibeto-Burman languages of the Himalaya. Carol Genetti provides some case studies on linguistic variation involving Newar. Boyd Michailovsky describes time-ordinals in Kiranti languages, Aimée Lahaussais ergativity in Thulung Rai, and Martine Mazaudon the discourse/grammar interface in Tamang.

Two papers deal with orthographic issues. R.K. Sprigg describes the Lepcha and Limbu (Tibeto-Burman) scripts of Nepal and Mark Hansell analyses variations in Chinese character choice in writing loanwords in Taiwan.

# *Language variation*

*Papers on variation and change in the  
Sinosphere and in the Indosphere  
in honour of James A. Matisoff*

# Pacific Linguistics 555

Pacific Linguistics is a publisher specialising in grammars and linguistic descriptions, dictionaries and other materials on languages of the Pacific, Taiwan, the Philippines, Indonesia, East Timor, southeast and south Asia, and Australia.

Pacific Linguistics, established in 1963 through an initial grant from the Hunter Douglas Fund, is associated with the Research School of Pacific and Asian Studies at The Australian National University. The authors and editors of Pacific Linguistics publications are drawn from a wide range of institutions around the world. Publications are refereed by scholars with relevant expertise, who are usually not members of the editorial board.

---

**FOUNDING EDITOR:** STEPHEN A. WURM

**EDITORIAL BOARD:** John Bowden, Malcolm Ross and Darrell Tryon (Managing Editors), I Wayan Arka, Bethwyn Evans, David Nash, Andrew Pawley, Paul Sidwell, Jane Simpson

---

## EDITORIAL ADVISORY BOARD:

- |                                                                           |                                                                                                 |
|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Karen Adams, <i>Arizona State University</i>                              | Lillian Huang, <i>National Taiwan Normal University</i>                                         |
| Peter Austin, <i>School of Oriental and African Studies</i>               | Bambang Kaswanti Purwo, <i>Universitas Atma Jaya</i>                                            |
| Alexander Adelaar, <i>University of Melbourne</i>                         | Marian Klamer, <i>Universiteit Leiden</i>                                                       |
| Byron Bender, <i>University of Hawai'i</i>                                | Harold Koch, <i>Australian National University</i>                                              |
| Walter Bisang, <i>Johannes Gutenberg-Universität Mainz</i>                | Frantisek Lichtenberk, <i>University of Auckland</i>                                            |
| Robert Blust, <i>University of Hawai'i</i>                                | Patrick McConvell, <i>Australian Institute of Aboriginal and Torres Strait Islander Studies</i> |
| Lyle Campbell, <i>Canterbury University</i>                               | William McGregor, <i>Aarhus Universitet</i>                                                     |
| James Collins, <i>Universiti Kebangsaan Malaysia</i>                      | Ulrike Mosel, <i>Christian-Albrechts-Universität zu Kiel</i>                                    |
| Bernard Comrie, <i>Max Planck Institute for Evolutionary Anthropology</i> | Claire Moyse-Faurie, <i>Centre National de la Recherche Scientifique</i>                        |
| Soenjono Dardjowidjojo, <i>Universitas Atma Jaya</i>                      | Bernd Nothofer, <i>Johann Wolfgang Goethe-Universität Frankfurt am Main</i>                     |
| Matthew Dryer, <i>State University of New York at Buffalo</i>             | Ger Reesink, <i>Universiteit Leiden</i>                                                         |
| Jerold A. Edmondson, <i>University of Texas at Arlington</i>              | Lawrence Reid, <i>University of Hawai'i</i>                                                     |
| Nicholas Evans, <i>University of Melbourne</i>                            | Jean-Claude Rivierre, <i>Centre National de la Recherche Scientifique</i>                       |
| Margaret Florey, <i>Monash University</i>                                 | Melenaite Taumoeofolau, <i>University of Auckland</i>                                           |
| William Foley, <i>University of Sydney</i>                                | Tasaku Tsunoda, <i>University of Tokyo</i>                                                      |
| Karl Franklin, <i>Summer Institute of Linguistics</i>                     | John Wolff, <i>Cornell University</i>                                                           |
| Charles Grimes, <i>Universitas Kristen Artha Wacana Kupang</i>            | Elizabeth Zeitoun, <i>Academica Sinica</i>                                                      |
| Nikolaus Himmelmann, <i>Ruhr-Universität Bochum</i>                       |                                                                                                 |

**Language variation**  
**Papers on variation and change in the**  
**Sinosphere and in the Indosphere**  
**in honour of James A. Matisoff**

David Bradley, Randy LaPolla, Boyd Michialovsky  
and Graham Thurgood (editors)



Pacific Linguistics  
Research School of Pacific and Asian Studies  
The Australian National University

Published by Pacific Linguistics  
Research School of Pacific and Asian Studies  
The Australian National University  
Canberra ACT 0200  
Australia

Copyright in this edition is vested with Pacific Linguistics

First published 2003

National Library of Australia Cataloguing-in-Publication entry:

Language variation : papers on variation and change in the  
Sinosphere and in the Indosphere in honour of James A. Matisoff.

Bibliography.  
ISBN 0 85883 540 1.

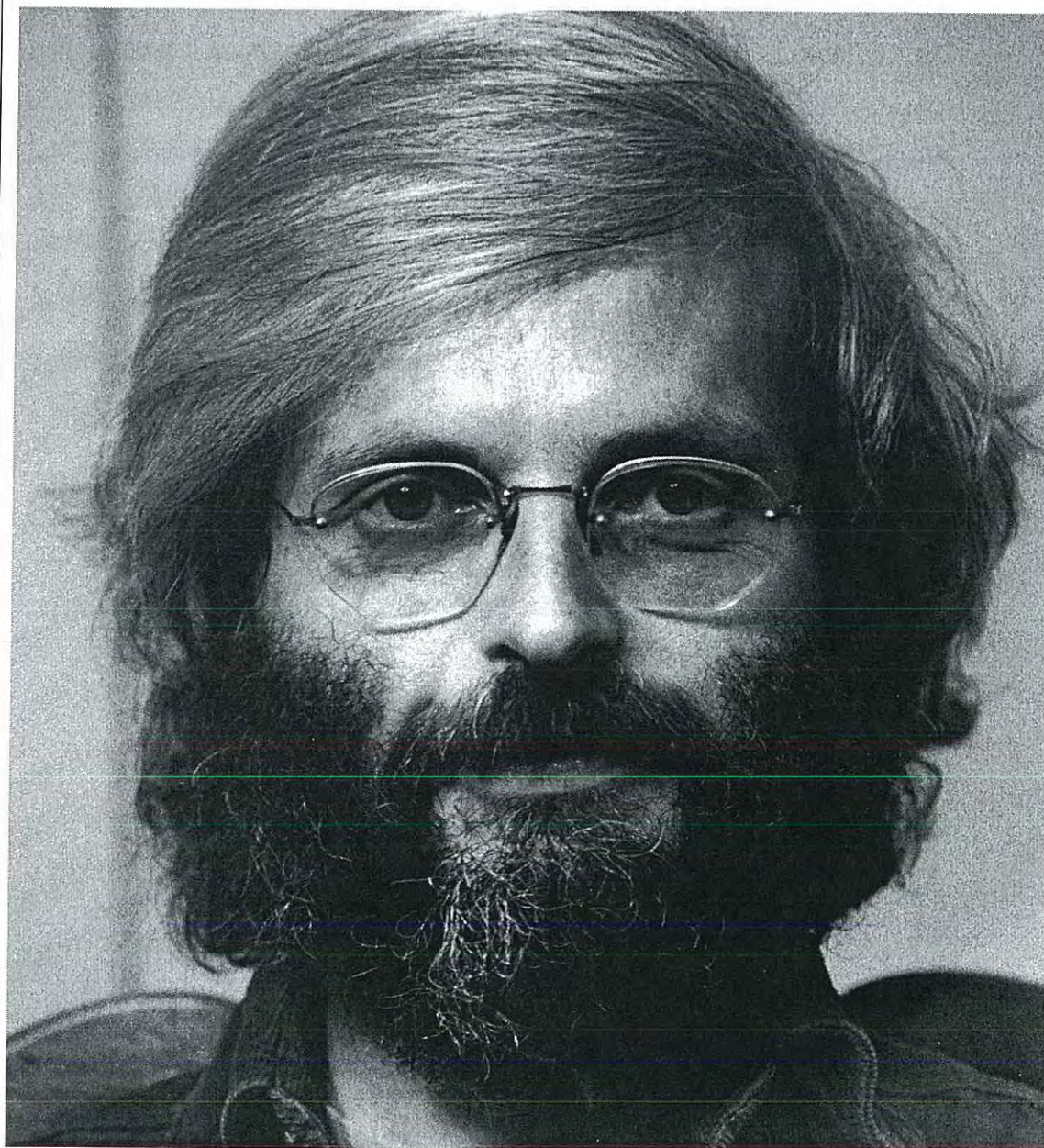
1. Language and languages - Variation. 2. Asia - Languages.  
I. Bradley, David, 1947- .

409.5

Typeset by Ms Jo Taylor, LaTrobe University  
Cover design by Emily Brissenden  
Printed and bound by Ciril's Printers, Fyshwick, Canberra

Variability is... as basic to linguistic structure as the countertendency toward order and organization... It is precisely this variability that lies at the heart of linguistic creativity.

Matisoff 1979[2000]:2



## *A brief biography of James A. Matisoff*

---

James (Jim) Alan Matisoff was born in Boston, USA, on Bastille Day (July 14), 1937. His father was an immigrant from a town near Minsk in what is now Belarus who sold fish for a living. Jim began his career in linguistics at Harvard, where he was awarded a Bachelor's degree (*magna cum laude*) in Romance Languages and Literatures in 1958, and a Master's degree in French Literature in 1959. For the next two years he lived abroad, first for a few weeks in the Soviet Union, then France, then India, and then Japan. In Japan he studied for a year at International Christian University, then in 1961 returned to Harvard on a teaching fellowship in Romance philology. While at Harvard he began studying Chinese with one of the daughters of the eminent linguist Y. R. Chao.

At Harvard he also met Susan Kimball, and married her in January of 1962. Susan's brother Jim was attending the University of California, Berkeley at the time, and suggested that Jim go to Berkeley to continue his studies. In the fall of 1962, Jim did just that, and there studied with luminaries such as Murray Emeneau, Wallace Chafe, Mary Haas, and Y. R. Chao. In the summer of 1963, Mary Haas received a letter from a Jinghpo anthropology student, Laraw Maran, saying he would like to act as an informant for someone interested in studying Jinghpo (a Tibeto-Burman language of Burma and China). Mary Haas suggested Jim take the opportunity, and this was the beginning of Jim's interest in Tibeto-Burman languages. By chance, that fall Wallace Chafe was investigating Burmese in his field methods class, and noticing the cognates shared by Jinghpo and Burmese Jim came to be interested in doing comparative work. He then applied to go to Burma to do fieldwork for his dissertation, but as Burma expelled all foreigners in 1964, Mary Haas suggested he go to northern Thailand instead. Before he left for Thailand he attended the LSA Summer Institute, and there studied Thai with a tutor, and took classes with William Comyn, Fang-kuei Li, Nicholas Bodman, and Edwin Pulleyblank, writing a paper comparing Burmese and Jinghpo for Prof. Pulleyblank.

In the fall of 1964, Jim, Susan, and their nine-month old baby, Nadja, left for Chiang Mai in northern Thailand. There he began his lifelong work on the Lahu language. While still in northern Thailand he received an invitation to go to Columbia University to teach. In 1966 he



moved to Columbia, where he taught three courses per semester. At the same time he was working on his dissertation, and finished it the following year. Jim and Susan also had their second daughter, Alexandra, during their first year at Columbia. At Columbia Jim met many eminent linguists who became great friends, such as Robert Austerlitz, William Labov, and Søren Egerod. While at Columbia Jim also met Paul K. Benedict, who had worked on a large-scale comparison of Sino-Tibetan languages from 1938-1941, and had produced a manuscript, *Sino-Tibetan: A Conspectus*, reconstructing hundreds of Proto-Tibeto-Burman and Proto-Sino-Tibetan forms. Jim worked with Benedict to edit the manuscript, and it was published in 1972. This publication was a major impetus to the field of Sino-Tibetan linguistics. While at Columbia Jim, along with Benedict, Egerod, and five other scholars from Princeton, Yale, Cornell, and Columbia, organized the first conference on Sino-Tibetan languages and linguistics, which began an annual tradition that went on to become the International Conference on Sino-Tibetan Languages and Linguistics. Over the years Jim has organized three other ICSTLL conferences, the second in 1969, the eighth in 1975, and the twenty-fifth in 1992. It is at the 35th ICSTLL in 2002 that we are presenting this volume to Prof. Matisoff. In 1969, Jim got a call from Wallace Chafe inviting him to go to Berkeley to teach, and this began a 33 year career of teaching and research at Berkeley.

Over the years Jim has done a vast amount of important research, and has been chair of twenty-five PhD committees. Aside from work on Lahu, producing *The Grammar of Lahu* (1973), *The Dictionary of Lahu* (1988), and numerous articles on aspects of Lahu language and culture, Jim has also published on Yiddish, on Jinghpo, and many other languages, and has also published many important theoretical works on grammaticalization, semantics, and typology. Jim was also a co-founder of the journal *Linguistics of the Tibeto-Burman Area*, which has become a major outlet for work on Tibeto-Burman languages and other languages of Southeast Asia. More than anyone else, Jim has been responsible for making the field of Sino-Tibetan linguistics what it is today. Though retiring this year (2002), Jim is still remaining active, carrying on with the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project. This year will see the publication of the *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*, an 800 page compendium on the theory and methodology of Tibeto-Burman linguistics.

The publications listed below appeared prior to Jim's retirement in July 2002. We all hope and expect that this phenomenal level of productivity and creativity will continue for many years to come!

---

# Table of contents

---

	<i>Biography</i> .....	vii
	<i>Contributors to this volume</i> .....	xi
1	<i>Introduction</i>	
	David Bradley .....	1
2	<i>Hmong secret languages: themes and variations</i>	
	Martha Ratliff.....	21
3	<i>Variiegated tonal developments in Tibetan</i>	
	Jackson T.-S. Sun.....	35
4	<i>Some case studies on linguistic variation and their implications</i>	
	Carol Genetti.....	53
5	<i>Recent language contact in the Nepal Himalaya</i>	
	Michael Noonan.....	65
6	<i>Prosodic tautomorphemicity in Sino-Tibetan</i>	
	Balthasar Bickel .....	89
7	<i>Ergativity in Thulung Rai: a shift in the position of pronominal split</i>	
	Aimée Lahaussais .....	101
8	<i>Why languages differ: variation in the conventionalisation of constraints on inference</i>	
	Randy J. LaPolla .....	113

9	<i>From discourse to grammar in Tamang: topic, focus, intensifiers and subordination</i> Martine Mazaudon.....	145
10	<i>Preliminary remarks on Gyarong negation particles</i> Yasuhiko Nagano.....	159
11	<i>Agreement and grammatical relations in Hyow</i> David A. Peterson.....	173
12	<i>Contact induced variation and syntactic change in the Tsat of Hainan</i> Graham Thurgood and Fengxiang Li.....	185
13	<i>Comparative notes on verb compounding in English and East Asian languages</i> Benji Wald.....	201
14	<i>Deictic patterns in Lisu and Southeastern Tibeto-Burman</i> David Bradley.....	219
15	<i>Suffix-runs and counters in Kiranti time-ordinals</i> Boyd Michailovsky.....	237
16	<i>The impact of English loanwords on the Cantonese syllabary</i> Robert S. Bauer.....	253
17	<i>On borrowing from Middle Chinese into Proto Tibetan: a new look at the problem of the relationship between Chinese and Tibetan</i> Michel Ferlus.....	263
18	<i>Phonetic fidelity vs. suggestive semantics: variations in Chinese character choice in the writing of loanwords</i> Mark Hansell.....	277
19	<i>Original and sophisticated features of the Lepcha and Limbu scripts</i> Richard Keith Sprigg.....	291
20	<i>Three Tibeto-Burman languages of Vietnam</i> Jerold A. Edmondson.....	305

## *Contributors to this volume*

---

ROBERT S. BAUER  
Chinese and Translation  
Hong Kong Polytechnic University  
Hung Hom, Kowloon  
HONG KONG  
[ctbao@polyu.edu.hk](mailto:ctbao@polyu.edu.hk)

BALTHASAR BICKEL  
Linguistics  
University of Leipzig  
Beethovenstr. 15  
04107 Leipzig  
GERMANY  
[bickel@rz.uni-leipzig.de](mailto:bickel@rz.uni-leipzig.de)

DAVID BRADLEY  
Linguistics  
La Trobe University VIC 3086  
AUSTRALIA  
[D.Bradley@latrobe.edu.au](mailto:D.Bradley@latrobe.edu.au)

JEROLD A. EDMONDSON  
Linguistics  
University of Texas at Arlington  
Arlington TX 76019  
USA  
[jerry@ling.uta.edu](mailto:jerry@ling.uta.edu)

MICHAEL FERLUS  
CRLAO  
Centre National de la Recherche Scientifique  
54 boulevard Raspail  
75006 Paris  
FRANCE  
[Michel.Ferlus@ehess.fr](mailto:Michel.Ferlus@ehess.fr)

CAROL GENETTI  
Linguistics  
University of California, Santa Barbara  
Santa Barbara CA 93106  
USA  
[cgenetti@linguistics.ucsb.edu](mailto:cgenetti@linguistics.ucsb.edu)

MARK HANSELL  
Chinese  
Carleton College  
Northfield MN 55057  
USA  
[mhansell@carleton.edu](mailto:mhansell@carleton.edu)

AIMÉE LAHAUSSOIS  
LACITO  
Centre National de la Recherche  
Scientifique  
7 rue Guy Môquet  
94801 Villejuif  
FRANCE  
[aimee@vjf.cnrs.fr](mailto:aimee@vjf.cnrs.fr)

RANDY LAPOLLA  
Chinese and Translation  
City University of Hong Kong  
Tat Chee Avenue, Kowloon  
HONG KONG  
[ctrandy@cityu.edu.hk](mailto:ctrandy@cityu.edu.hk)

LI FENGXIANG  
English  
California State University Chico  
Chico CA 95929  
USA  
[Fli@csuchico.edu](mailto:Fli@csuchico.edu)

MARTINE MAZAUDON  
LACITO

Centre National de la Recherche Scientifique  
7 rue Guy Môquet  
94801 Villejuif  
FRANCE  
[mazaudon@vjf.cnrs.fr](mailto:mazaudon@vjf.cnrs.fr)

BOYD MICHAÏLOVSKY  
LACITO

Centre National de la Recherche Scientifique  
7 rue Guy Môquet  
94801 Villejuif  
FRANCE  
[boydm@vjf.cnrs.fr](mailto:boydm@vjf.cnrs.fr)

YASUHIKO NAGANO

National Museum of Ethnology  
Senri Exposition Park, Suita 565  
JAPAN  
[nagano@idc.minpaku.ac.jp](mailto:nagano@idc.minpaku.ac.jp)

MICHAEL NOONAN

English  
University of Wisconsin at Milwaukee  
PO Box 413  
Milwaukee WI 53201  
USA  
[noonan@csd.uwm.edu](mailto:noonan@csd.uwm.edu)

DAVID A. PETERSON

Linguistics and Cognitive Science  
Dartmouth College  
Hanover NH 03755  
USA  
[david.a.peterson@dartmouth.edu](mailto:david.a.peterson@dartmouth.edu)

MARTHA RATLIFF

English  
Wayne State University  
Detroit MI 48202  
USA  
[martha\\_ratliff@wayne.edu](mailto:martha_ratliff@wayne.edu)

R. K. SPRIGG

Linguistics  
School of Oriental and African Studies  
Thornhaugh St, Russell Sq  
London WC1H 0XG  
UNITED KINGDOM  
[sprigran@btinternet.com](mailto:sprigran@btinternet.com)

JACKSON T.-S. SUN

Institute of Linguistics  
Academia Sinica  
Nankang, Taipei 115  
TAIWAN, RoC  
[hstssun@gate.sinica.edu.tw](mailto:hstssun@gate.sinica.edu.tw)

GRAHAM THURGOOD

English  
California State University Chico  
Chico CA 95929  
USA  
[gthurgood@csuchico.edu](mailto:gthurgood@csuchico.edu)

BENJI WALD

5452 Barton  
Los Angeles CA 90038-3204  
USA  
[bjwald@earthlink.net](mailto:bjwald@earthlink.net)

# 1 *Introduction*

---

DAVID BRADLEY

## **Introduction**

This volume contains nineteen papers by students and colleagues of James A. Matisoff. These are all focused on the central themes of Jim's work: describing the variation that exists within and among languages, and investigating the processes of change that link them but make them differ.

The editors chose to focus this volume on the topic of variation and change in the East, Southeast and South Asian linguistic areas, where so much of Matisoff's contribution to linguistics has been made. All the studies included are directly related to his work and based on his insights, prepared by students and colleagues who have worked closely with Jim.

He is most widely known as the author of a large number of studies of Lahu, including a grammar (Matisoff 1973 [1982]) and a dictionary (Matisoff 1988). These have provided a model for all our work on related languages of the Tibeto-Burman (TB) family and for work on other languages in the same linguistic areas, East, Southeast and South Asia.

He has been highly creative in all areas of linguistics throughout his career, making a major contribution to the basic terminology of linguistics, coining many very widely-used terms; here, we cite only a few, chosen from various areas of linguistic structure.

Phonology, **sesquisyllabic**—the syllable-and-a-half word structure so widely found in languages of Southeast Asia

Phonology of tone, **tonogenesis**—the development and expansion of tonal systems, often conditioned by segmental features such as final glottal stop or other syllable-finals; and by voiced versus voiceless or other types of syllable-initials, coined in Matisoff (1970a) and further developed in Matisoff (1973c).

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 1–20.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

2 *David Bradley*

Morphosyntax, **versatile verbs**—the grammaticalised verbs which precede or follow the head verb in a structured way, described with characteristic brio and thoroughness in his studies of Lahu

Semantics, **psycho-ostensive expressions**—the culturally-determined proverbs and similar phrases which Jim has so amply documented and analysed in Yiddish

Historical linguistics, **allofam**—based on the key recognition that proto-languages could not have been any more monolithic and invariant than modern languages, the reconstruction of word families recognizes the likelihood of alternative related forms with different reconstructed forms

Sociolinguistics, the productive use of Greek compounds for language and group names such as **autonym** (a group's own name for itself) and so on, clarifying the terminology of a highly complex and problematic area, fully explored for the Tibeto-Burman languages in one of the STEDT project's first publications

Areal linguistics, the terms **Sinosphere** (area of linguistic and cultural influence of China) and **Indosphere** (area of linguistic and cultural influence of India) in the title of this volume

As the papers are all self-contained, this introduction will not attempt to summarise them. They include papers on every area of linguistics, including documentation, phonology, morphosyntax, discourse, sociolinguistics, orthography and so on. The organization of the volume follows the core areas of linguistics. The languages and areas represented range across the TB and Sino-Tibetan languages, various other languages of the same linguistic area, and even more widely, with one study relating the rapidly-expanding English pattern of two-verb compounds to Matisoff's Lahu versatile verbs and similar patterns in other languages of the Sinosphere.

Jim has approached linguistics throughout his career with a combination of rigor and creativity. His inspiration has started many others on the same path, and we are delighted to join in presenting him with this volume. We know that many others would have been able and willing to contribute, and we apologise to them and to Jim that they were not able to do so this time. But we are sure that there will be many other opportunities to honor Matisoff in the future!

The papers in this volume fall into various major areas of linguistics, phonology, morphosyntax, semantic fields, contact and orthographies, and surveys. The volume follows this arrangement. First, there are two papers on various aspects of phonology: details of secret languages in Hmong by Ratliff, and discussion of the development of tones in various varieties of Tibetan by Sun. This is followed by two papers which discuss variation in both phonology and morphosyntax: by Genetti on Newar and by Noonan on the languages of Nepal in general.

The morphosyntax section is the largest, with Bickel's study of tautomorphemicity, a study of a shift in the ergative split in Thulung which took place within the last thirty years by Lahaussais, a broad-ranging discussion of the constraints on inference by LaPolla, Mazaudon's paper on the grammaticalisation of discourse in Tamang, Nagano's presentation on negation in Gyarong, Peterson's study of the verb agreement morphology of Hyow Chin, Thurgood & Li's study of syntactic convergence toward Chinese, especially in Tsat, an Austronesian language of Hainan, but also in Oroqen, a Tungusic language of northeastern China, and in Mulam, a Tai-Kadai language of southwestern China. Then, Wald discusses how English is developing versatile-verb like compound verbs, parallel to those of Lahu and many other East Asian languages.

The two semantic fields represented here are demonstratives (in a paper by Bradley on Lisu and closely related Burmic languages) and time ordinals for days and years (Michailovsky on Kiranti languages). Both areas are particularly well-developed in a variety of TB languages.

The contact and orthography section includes a fascinating paper by Bauer showing exactly which syllable types are shared with other Sinitic languages, which are specifically Cantonese, and which occur only in English loanwords. Ferlus shows how the process of blending can lead to apparently irregular correspondences, with examples primarily showing influence from Chinese on the form of Tibetan numerals. Hansell looks at the phonetic, semantic and graphic factors in the representation of English loanwords into Mandarin Chinese. Sprigg discusses how the Lepcha and Limbu orthographies are organized, showing how they reflect the different systemic value of final consonants. In the final section on language surveys, Edmondson provides various kinds of descriptive data on three TB languages of Vietnam.

As can be seen, a bit more than half of the contributions are on languages of the Sinosphere, and a bit less than half on languages of the Indosphere, with several more broad-ranging discussions which cover both areas and go even further. Most of the papers blend synchronic and diachronic approaches, as is characteristic of much of Matisoff's work. Many of the papers also shed light on issues in sociolinguistics and typology. Overall, the volume is focused on the core notion of variation.

All the studies in this volume cite the specific inspirations and publications of Matisoff which have contributed to their work. The STEDT phonetic font, created by Matisoff's STEDT project, is used for phonetic forms throughout the volume.

We hope Jim will be pleased with what we have done.



**James A. Matisoff: Publications**

- 1966 The phonology of the Lahu (Muhsur) language. Report to the National Research Council of Thailand, Bangkok. Mimeographed ms. (37pp.)
- 1967 A Grammar of the Lahu Language. PhD dissertation, University of California, Berkeley. Ann Arbor, MI: University Microfilms, Order No.67-11648. (697 pp.)
- 1968 Review of Robbins Burling, 1967, *Proto-Lolo-Burmese*. In *Language* 44/4:879-897.
- 1969a] Lahu and Proto-Lolo-Burmese. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*, vol. I, 117-221. Ann Arbor, MI: Department of Linguistics, University of Michigan.  
Reviews:  
Haudricourt, André-Georges, 1970, *Bulletin de la Société de Linguistique de Paris* 65/2:253-254.  
Jones, R.B., 1970, *Journal of Asian Studies* 30/1:230-231.
- 1969b] Verb concatenation in Lahu: the syntax and semantics of 'simple' juxtaposition. *Acta Linguistica Hafniensia* 12/1:69-120.
- 1969c] Lahu bilingual humor. *Acta Linguistica Hafniensia* 12/2:171-206.
- 1969d] Review of Paul Lewis, 1968, *Akha-English Dictionary*. In *Journal of Asian Studies* 28/3:644-645.
- 1969e] Review of Thomas Sebeok ed., 1967, *Current Trends in Linguistics II: Linguistics in East Asia and Southeast Asia*. In *Journal of Asian Studies* 28/4:835-837.
- 1970a] Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study. *Journal of the American Oriental Society* 90/1:13-44.
- 1970b] Review of Norman A. Mundhenk, 1967, *Auxiliary verbs in Myang of Northern Thailand*. In *Journal of Asian Studies* 29/2:492-493.
- 1970c] Note on the orthography of Lahu. In Anthony R. Walker *Red Lahu village society and economy in North Thailand*, vol. I, xxxiii-xxxv. Chiangmai: Tribal Research Center.
- 1971 The tonal split in Loloish checked syllables. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*, vol. II. Urbana, IL: Department of Linguistics, University of Illinois. (44 pp.)  
Review:  
Kun Chang, 1970, *Journal of Asian Studies* 31/4:988.
- 1972a] *The Loloish tonal split revisited*. Research Monograph No.7, Center for South and Southeast Asia Studies, University of California, Berkeley. (viii + 88 pp.)

## Reviews:

Coblin, W. South, 1974, *Journal of the American Oriental Society* 94/4:522–524.

Haudricourt, André-Georges, 1973, *Bulletin de la Société de Linguistique de Paris* 68/2:495.

Sedláček, Kamil, 1975, *Zeitschrift der deutschen morgenländischen Gesellschaft* 125/1:227–228.

Sprigg, R.K., 1974, *Bulletin of the School of Oriental and African Studies* 37/1:259–262.

Chinese translation with new introduction [2002a].

1972b] Tangkhul Naga and comparative Tibeto-Burman. *Tonan Azia Kenkyu [Southeast Asian Studies]* 10/2:1–13.

1972c] Lahu nominalization, relativization, and genitivization. In John Kimball ed. *Syntax and Semantics*, vol. I, 237–57. Studies in Language Series. Seminar Press, New York.

## Review:

Bendix, Edward H., 1975, *American Anthropologist* 77/4:957.

1972d] Review of Radoslav Katicić, 1970, *A Contribution to the general theory of comparative linguistics*. In *American Anthropologist* 74/1–2:96.

1972e] Review of D.N. Shankara Bhat, 1969, *Tankhur Naga vocabulary*. In *Language* 48/2:476–479.

1972f] Contributing editor of Paul K. Benedict, *Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.

Reviews by Bodman, Nicholas C., Chang, Kun, Chou Fa-kao, Coblin, W. South, Denwood, Philip, Egerod, Søren, Haudricourt, André-Georges, Hoffman, Helmut, Lehman, F.K., Miller, Roy Andrew, Roy, Gilbert, Sedláček, Kamil, Simon, Walter, and Sprigg, R.K.

Chinese translation [1984b].

1973a] *The grammar of Lahu*. University of California Publications in Linguistics, No.75. University of California Press, Berkeley, Los Angeles, London. (ii + 673 pp., map, photographs, 57 figures) Reprinted 1982.

## Reviews:

De Lancey, Scott, 1988, *Language* 64/1:213–14.

Denlinger, Paul B., 1979, Grammatical comparison in Sino-Tibetan: a review article of James A. Matisoff, *The Grammar of Lahu. Monumenta Serica* 33:300–308.

Haudricourt, André-Georges, 1974, *Bulletin de la Société de Linguistique de Paris* 69/2:373.

- Lehman, F.K., 1978, *Journal of the American Oriental Society* 98/3:296–297.  
Okell, John, 1975, *Bulletin of the School of Oriental and African Studies* 38/3:669–674.  
Sedláček, Kamil, 1979, *Zeitschrift der deutschen morgenländischen Gesellschaft* 129:441–442.

- 1973b] The annual Sino-Tibetan conferences: the first five years, 1968–1972. *Journal of Chinese Linguistics* 1/1:152–162.
- 1973c] Tonogenesis in Southeast Asia. In Larry M. Hyman ed. *Consonant types and tone*, 71–95. Southern California Occasional Papers in Linguistics, No.1. Los Angeles: UCLA.
- 1973d] Review of LaRaw Maran, 1971, *Burmese and Jingpho: a study of tonal linguistic processes*. In *Journal of Asian Studies* 32/4:741–743.
- 1973e] Notes on Fang-kuei Li's 'Languages and dialects of China.' *Journal of Chinese Linguistics* 1/3:471–474.  
Chinese translation [1985g].
- 1973f] Review of John Okell 1969 *A reference grammar of colloquial Burmese*. In *Journal of the American Oriental Society* 93/3:471–474.
- 1973g] The little crabs who walked zigzag. Fable translated from the Lahu. In Herbert R. Kohl ed. *Fables, level II*, 22–24. Boston: Houghton Mifflin Co.
- 1974a] Verb concatenation in Kachin. *Linguistics of the Tibeto-Burman Area* 1/1:186–207.
- 1974b] The tones of Jinghpaw and Lolo-Burmese: common origin vs. independent development. *Acta Linguistica Hafniensia* (Copenhagen) 15/2:153–212.
- 1974c] Sifrut ha-bituy b'Yidish: lashon psixo-ostensivit ba-dibur ha-amimi. Translation into Hebrew by Chana Kaufman (Kronfeld) of 1973 MS, Psycho-ostensive expressions in Yiddish. *Hasifrut* [Literature] 18–19:181–223. Errata in *Hasifrut* 21:168 (1975). The English original was published with revisions and additions as 1979a, and republished (2000a) by Stanford University Press.
- 1975a] Rhinoglottophilia: the mysterious connection between nasality and glottality. In Charles Ferguson, Larry M. Hyman, & John Ohala eds *Nasálfest: papers from a symposium on nasals and nasalization*, 265–287. Stanford, CA: Stanford University Language Universals Project.
- 1975b] Benedict's Sino-Tibetan: a rejection of Miller's Conspectus inspection. *Linguistics of the Tibeto-Burman Area* 2/1:155–172.
- 1975c] A new Lahu simplex/causative pair: 'study/train.' *Linguistics of the Tibeto-Burman Area* 2/1:151–154.

- 1975d] Review of D. Haigh Roop, 1972, *An introduction to the Burmese writing system*. In *Journal of the American Oriental Society* 95/3:536–537.
- 1975e] Translation (from Russian) of K. B. Keping, Subject and object agreement in the Tangut verb. *Linguistics of the Tibeto-Burman Area* 2/2:219–31.
- 1976a] Lahu causative constructions: case hierarchies and the morphology/syntax cycle in a Tibeto-Burman perspective. In Masayoshi Shibatani ed. *The grammar of causative constructions*, 413–442. New York: Academic Press.
- 1976b] Austro-Thai and Sino-Tibetan: an examination of body-part contact relationships. In Mantaro J. Hashimoto ed. *Genetic relationship, diffusion, and typological similarities of East and Southeast Asian languages*, 256–289. Tokyo: Japan Society for the Promotion of Science.
- Review:  
Benedict, Paul K., 1976. *Computational Analyses of Asian and African Languages* 6:93–4. Chinese translation [1985h].
- 1977a] Malediction and psycho-semantic theory: the case of Yiddish. *Maledicta (the international journal of verbal aggression)* 1/1:31–39. [Reprinted as 1996e].  
Review Posner, Sara 1981, *Journal of the North Shore Jewish Community* (Salem, MA), vol. 5/10:13. 9
- 1977b] Introduction to the Written Burmese Rhyming Dictionary. *Linguistics of the Tibeto-Burman Area* 3/1:iii–x.
- 1977c] Waga Nippon hakken no san-dankai. In *Nihonjin-tte omoshiroi desu ne*, 12–16. Tokyo: Kokusai Kyooiku Shinkookai. (Speech delivered at First Asahi Shimbun Speech Contest for Foreigners, Tokyo, June 1961.)
- 1978a] Mpi and Lolo-Burmese microlinguistics. *Monumenta Serindica*, No.4. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. (36 pp.)
- 1978b] *Variational semantics in Tibeto-Burman: the 'organic' approach to linguistic comparison*. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*, vol. VI. Philadelphia: Institute for the Study of Human Issues (ISHI). (xviii + 331 pp.)
- Reviews:  
Burling, Robbins, 1980, *Language* 56/4:888–891.  
Haudricourt, André-Georges, 1980, *Bulletin de la Société de Linguistique de Paris* 75/2:405–406.  
Thurgood, Graham 1982, *Indian Linguistics* 40/3:116–117.
- 1979a] *Blessings, curses, hopes, and fears: psycho-ostensive expressions in Yiddish*. Philadelphia: Institute for the Study of Human Issues (ISHI). (xx + 140 pp.)

Reviews:

Anon. 1980, *Kirjath Sepher [Bibliographical Quarterly]* (Jerusalem), vol. 55/2: item #2473.

Aman, Reinhold A., 1980, *Maledicta* IV:151–152 (August).

Fishman, Joshua A., 1979, *Library Journal* (June 15).

Keveson-Hertz, Bertha, 1980, Yiddish in a novel key. *The Jewish Spectator* (Santa Monica, CA), 52–53 (Fall).

Lester, Elenore, 1980, *The Jewish Week - American Examiner*, New York (week of Feb. 10).

Tumin, Israel, 1980, *Jewish News*, Livingston, NJ (August).

Wexler, Paul, 1990, In Paul Wexler ed. *Studies in Yiddish Linguistics*, 175–177. Tübingen: Max Niemayer Verlag.

- 1979b] Problems and progress in Lolo-Burmese: Quo Vadimus? *Linguistics of the Tibeto-Burman Area* 4/2:11–43.
- 1979c] Trickster and the village women: a psychosymbolic discourse analysis of a Lahu picaresque story. *Berkeley Linguistics Society* 5:593–636.
- 1979d] Editor and annotator of Paul K. Benedict, Four forays into Karen linguistic history. *Linguistics of the Tibeto-Burman Area* 5/1:1–35.
- 1979e] Review of Alfons Weidert, 1975, *Componential analysis of Lushai phonology*. In *Journal of the American Oriental Society* 99/3:496.
- 1979f] Translation (from Japanese) and annotation of Tatsuo Nishida, *The structure of the Hsi-hsia (Tangut) characters*. *Monumenta Serindica* 8. Tokyo: Institute for the Study of the Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies. (42 pp.)
- 1980a] Stars, moon, and spirits: bright beings of the night in Sino-Tibetan. *Gengo Kenkyu* (Tokyo) 77:1–45.
- 1980b] Review of Edward R. Hope, 1974, *The deep syntax of Lisu sentences*. In *Journal of the American Oriental Society* 100/3:386–387.
- 1981] Review of N.J. Allen, 1975, *Sketch of Thulung grammar*. In *Journal of the American Oriental Society* 101/4:435–436.
- 1982a] Proto-languages and Proto-Sprachgefühl. *Linguistics of the Tibeto-Burman Area* 6/2:1–64.
- 1982b] Second Printing of *The grammar of Lahu* [1973a].
- Review:
- Scott DeLancey, 1988, *Language* 64/1:213–14.

- 1982c] Conjugal bliss: an Indo-Aryan word-family pair /yoke/ join in Tibeto-Burman. *South Asian Review* VI/3:42–50. [Reprinted as 1987a].
- 1982d] Foreword to William Kuo, *Teaching grammar of Thai*. Berkeley: Center for South and Southeast Asia Studies, University of California Berkeley.
- 1983a] Linguistic diversity and language contact in Thailand. In John McKinnon & Wanat Bhruksasri eds *Highlanders of Thailand*, 56–86. Kuala Lumpur: Oxford University Press.
- Review:  
Turton, Andrew, 1985, *Bulletin of the School of Oriental and African Studies* 48/1:185–186.
- 1983b] Translucent insights: a look at Proto-Sino-Tibetan through Gordon H. Luce's Comparative Word-list. *Bulletin of the School of Oriental and African Studies* (London) 46/3:462–476.  
Chinese translation [1989g].
- 1983c] (with Reinhold A. Aman) A Yiddish Minnie-legend: Pokayente funem Shvartsfus-sheyyvet. *Maledicta* VII:259–268.
- 1984a] Translation (into Chinese) by Zhao Yansun of Some problems in the orthography of Lahu as Lāhù-yǔ wénzì fāng'àn zhong de ruògān wèntǐ *Minzu Yuwen* 1984/3:27–38. The English original is unpublished.
- 1984b] Translation into Chinese of 1972f by Le Saiyue & Luo Meizhen. Translation checked by Qu Aitang & Wu Miaofa. Beijing: Institute of Nationalities, Chinese Academy of Social Sciences.
- 1984c] [in Chinese] Historical linguistics and research on the Sino-Tibetan family. Interview taped and translated by Xu Tongqiang, published in *Yuyanxue Luncong* 23:213–34. (Also in this issue are Xu's interviews with Y. Malkiel, Chang Kun, and W. S-Y. Wang), under the general title Interviews with American linguists on historical linguistics. 200–258. Beijing: Department of Chinese, Beijing University.
- 1985a] God and the Sino-Tibetan copula, with some good news concerning selected Tibeto-Burman rhymes. *Journal of Asian and African Studies* (Tokyo Foreign Languages University) 29:1–81. With an Appendix by Richard Kunst.
- 1985b] (with Graham Thurgood & David Bradley) eds, *Linguistics of the Sino-Tibetan Area: the state of the art: papers presented to Paul K. Benedict for his 71st birthday*. Canberra: Pacific Linguistics. (vii + 498 pp.)
- 1985c] Paul K. Benedict: an appreciation. In Thurgood, Matisoff, & Bradley eds, 16–20.

- 1985d] New directions in East and Southeast Asian linguistics. In Thurgood, Matisoff, & Bradley eds, 21–35.
- 1985e] Out on a limb: ARM, HAND, and WING in Sino-Tibetan. In Thurgood, Matisoff, & Bradley eds, 421–450.
- 1985f] Foreword to Linda Young, *Shan chrestomathy*, vii. Washington, DC: University Press of America.
- 1985g] Translation into Chinese of 1973e by Liang Min as *Duì Lǐ Fāngguì Zhōngguó de yǔyán hé fāngyán yì-wén de pīpíng*. *Minzu Yuwen Yanjiu Qingbao Cailiao-ji* 6:136–138 and 98.
- 1985h] Translation into Chinese of 1976b by Wang Dewen & Hu Tan as *Aò-Tài yǔxì hé Hàn-Zàng yǔxì yǒu guān shēntǐ bùfen-cǐ jiēchù guānxi de jiǎnyàn*. *Minzu Yuwen Yanjiu Qingbao Cailiao-ji* 6:1–20.
- 1986a] The languages and dialects of Tibeto-Burman: an alphabetic/genetic listing, with some prefatory remarks on ethnonymic and glossonymic complications. In John McCoy & Timothy Light eds *Contributions to Sino-Tibetan studies*, presented to Nicholas C. Bodman, 3–75. Leiden: Brill.
- Review:  
Sagart, Laurent 1987, *Cahiers de Linguistique Asie Orientale* 16/2:289.
- 1986b] Labiovelar unit phonemes in Lolo-Burmese? A case to chew over: Lahu bê ‘chew’ < PLB \*N-gwya<sup>2</sup>. *Linguistics of the Tibeto-Burman Area* 9/1:83–88.
- 1986c] Hearts and minds in Southeast Asian languages and English: an essay in the comparative lexical semantics of psycho-collocations. *Cahiers de Linguistique Asie Orientale* 15/1:5–57.
- 1987a] Conjugal bliss: an Indo-Aryan word-family pair /yoke/ join in Tibeto-Burman. In *Wang Li memorial volumes (English Volume)*, 309–320. Hong Kong: The Chinese Language Society of Hong Kong. [Reprint of 1982c].
- 1987b] Review of Denise Bernot, 1978/1984, *Dictionnaire Birman – Français*, fascicules 1–8. In *Bulletin of the School of Oriental and African Studies* 50/1:191–195.
- 1987c] Review of Franklin E. Huffman, 1986, *Bibliography and index of mainland Southeast Asian languages and linguistics*. In *Journal of Asian Studies* 46/2:451–453.
- 1987d] Translation (from Chinese) of Sun Hongkai, A brief account of my research work. *Linguistics of the Tibeto-Burman Area* 10/1:117–125.

- 1987e] Zàng-Miǎn-yǔ yánjiū duì Hàn-yǔ shǐ yánjiū de gòngxiàn. (The contribution of Tibeto-Burman research to historical research on the Chinese language.) Lecture presented at Beijing University (April 28, 1983), simultaneously translated by Mei Tsu-lin. Written translation by Yè Fēishēng from audiotape and notes, published in *Yuyan Yanjiu Lun Cong* (Collected Essays on Linguistic Research), 61–68. Tianjin: Nankai University Press.
- 1988a] Universal semantics and allofamic identification – two Sino-Tibetan case-studies: STRAIGHT / FLAT / FULL and PROPERTY / LIVESTOCK / TALENT. In Akihiro Sato ed. *Languages and history in East Asia: festschrift for Tatsuo Nishida on the occasion of his 60th birthday*, 3–14. Kyoto: Shokado.
- 1988b] *The dictionary of Lahu*. University of California Publications in Linguistics, vol. 111. University of California Press, Berkeley, Los Angeles, London. (xxv + 1436 pp., 80 plates, 12 frontispiece photos, 8pp. booklet).  
 Reviews:  
 Zgusta, L., 1990, *American Reference Books Annual*, Entry 1050. Libraries Unlimited, Goldwood, CO.  
 Egerod, Søren, 1991, *Language* 67/2:373–377.  
 Lyman, Thomas Amis, 1992, *Zeitschrift der deutschen morgenländischen Gesellschaft* 142/2:434–436.  
 Okell, John, 1994, *Bulletin of the School of Oriental and African Studies* 57/2.
- 1988c] Proto-Hlai initials and tones: a first approximation. In Jerold A. Edmondson & David Solnit eds *Comparative Kadai: linguistic studies beyond Tai*, 289–321. Publications in Linguistics, #86, Summer Institute of Linguistics and University of Texas (Arlington).  
 Review:  
 Tehan, Thomas M, 1992, *Notes on Linguistics* 58:40–44.  
 Chinese translation [1988d].
- 1988d] Translation into Chinese of 1988c by Ouyang Jueya as Yuánshǐ Lí-yǔ de shēngmǔ hé shēngdiào yí chūbù jìnsì gòunǐ. *Minzu Yuwen Yanjiu Ziliao* 10:1–25.
- 1989a] A new Sino-Tibetan root \*d-yu-k BELONG / TRUST / DEPEND / ACCEPT / TAKE, and a note of caution to megalo-reconstructionists. In David Bradley, Eugénie J.A. Henderson, & Martine Mazaudon eds *Prosodic analysis and Asian linguistics: to honour R. K. Sprigg*, 265–269. Canberra: Pacific Linguistics.
- 1989b] The bulging monosyllable, or the mora the merrier: echo-vowel adverbialization in Lahu. In Jeremy Davidson ed. *South-East Asian linguistics: essays in honour of Eugénie J.A. Henderson*, 163–197. London: School of Oriental and African Studies.



Review:

Clark, Marybeth, 1990, *Journal of the Asian Studies Association of Australia* 14/2:268–269.

- 1989c] Review of Jerry Norman, 1987, *Chinese*. In *Journal of Asian Studies* 48/4:836–838.
- 1989d] Review of G.H.Luce, 1985, *Phases of Pre-Pagan Burma: languages and history*. In *Bulletin of the School of Oriental and African Studies* 52/3:599–602.
- 1989e] Tone, intonation, and sound symbolism in Lahu: loading the syllable canon. *Linguistics of the Tibeto-Burman Area* 12/2:147–163. [Reprinted as 1995e].
- 1989f] Introduction to the STEDT Monograph Series. In Randy J. LaPolla & John B. Lowe, *Bibliography of the International Conferences on Sino-Tibetan Languages and Linguistics I-XXI*, pp. ix–xi. Berkeley: Center for Southeast Asia Studies, University of California.
- 1989g] Translation into Chinese of 1983b by Xu Shixuan as Dǒng wēi zhú yōu: cóng Huádēng H. Lúsī de Duìyīng Cíhuì kàn yuánshǐ Hàn-Zàng-yǔ. *Minzu Yuwen Yanjiu Ziliaoji* 11:77–90.
- 1990a] Zàng-Miǎn yǔzú yǔyán yánjiū yǔ zhǎnwàng: Mǎtísuǒfù Jiàoshòu fāngwèn-jì. (Current research and future prospects in Tibeto-Burman linguistics) *Minzu Yuwen* 1990/1:1–8. Chinese translation by Jackson Tianshin Sun of an interview with J. Matisoff, conducted at Berkeley by Dai Qingxia in November, 1989.
- 1990b] On megalocomparison. *Language* 66/1:106–120.
- 1990c] Nihongo to Chibetto-Biruma shogo [Japanese and the Tibeto-Burman languages] In Sakiyama, Osamu ed. *Nihongo no keisei [The formation of the Japanese language]*, 54–73, followed by Discussion, 74–9. Proceedings of a Symposium held at the National Museum of Ethnology (Kokuritsu Minzokugaku Hakubutsukan), Osaka. Tokyo: Sanseido.
- 1990d] Bulging monosyllables: areal tendencies in Southeast Asian diachrony. *Berkeley Linguistics Society* 16:543–559.
- 1991a] Syntactic parallelism and morphological elaboration in Lahu religious poetry. In Sandra Chung & Jorge Hankamer eds *A festschrift for William F. Shipley*, 83–103. Santa Cruz, California: Syntax Research Center.
- 1991b] Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth C. Traugott & Bernd Heine eds *Approaches to grammaticalization*, vol. II, 383–453. Amsterdam: Benjamins.
- 1991c] Jiburish revisited: tonal splits and heterogenesis in Burmo-Naxi-Lolo checked syllables. *Acta Orientalia* 52:91–114.

- 1991d]** Sino-Tibetan linguistics: present state and future prospects. *Annual Review of Anthropology* 20:469–504.  
Chinese translation [1993c].
- 1991e]** Endangered languages of mainland Southeast Asia. In R. H. Robins & E. M. Uhlenbeck eds *Endangered languages*, 189–228. Published with the authority of the Permanent International Committee of Linguists. Oxford and New York: Berg.  
Review:  
Daniels, Peter T, 1993. *Language in Society*. 22/4:584–588.
- 1991f]** Lexicography of other Tibeto-Burman languages. In F. J. Hausmann, O. Reichmann, H.E. Wiegand, & L. Zgusta eds *Dictionaries: an international encyclopedia of lexicography*, vol.3:2555–2560. Berlin and New York: de Gruyter .
- 1991g]** Notes on the Fifth International Yi-Burmese Conference and subsequent peregrinations. *Linguistics of the Tibeto-Burman Area* 14/2:179–183.
- 1991h]** The mother of all morphemes: augmentatives and diminutives in areal and universal perspective. In Martha Ratliff & Eric Schiller eds *Papers from the first annual meeting of the Southeast Asian linguistics society (SEALS)*, 293–349. Tempe, AZ: Program for Southeast Asian Studies, Arizona State University.
- 1992a]** Southeast Asian languages. In William Bright & Bernard Comrie eds *International Encyclopedia of Linguistics*, Vol IV:44–48. New York and Oxford: Oxford University Press.
- 1992b]** The Lahu people and their language. In Judy Lewis ed. *Minority cultures of Laos*, 125–247. Rancho Cordova, CA: Southeast Asia Community Resource Center, Folsom Cordova Unified School District in consultation with the California State Department of Education Bilingual Education Office.
- 1992c]** A key etymology. *Linguistics of the Tibeto-Burman Area* 15/1:139–143.
- 1992d]** Following the marrow: two parallel Sino-Tibetan etymologies. *Linguistics of the Tibeto-Burman Area* 15/1:159–177. [Reprinted as 1994h].
- 1992e]** Preface. *Berkeley Linguistic Society, special session on the typology of tone languages*, viii-xii.
- 1992f]** Siamese *jaai* revisited, or ¡Ay, madre!: a case study in multiple etymological possibilities. In Carol J. Compton & John F. Hartmann eds *Papers on Tai languages, linguistics, and literatures, in honor of William J. Gedney on his 77th birthday*, 111–117. Occasional paper #17, Center for Southeast Asian Studies. De Kalb, IL:Northern Illinois University.

- 1993a] Review of William J. Gedney, 1989, *Selected papers on comparative Tai studies*. In *Language* 69/1:178–182.
- 1993b] Sangkong of Yunnan: secondary verb pronominalization in Southern Loloish. *Linguistics of the Tibeto-Burman Area* 16/2:123–142. [See also 1994c]
- 1993c] Translation into Chinese of 1991d by Fu Ailan as Hân-Zàng-yǔ yǔyánxué de xiànzhàng yǔ wèilái. *Waiguo Yuyanxue* (Part I) 1993/3 22–28, (Part II) 1993/4:25–31, 43.
- 1994a] Regularity and variation in Sino-Tibetan. In Hajime Kitamura, Tatsuo Nishida, & Yasuhiko Nagano eds *Current issues in Sino-Tibetan linguistics*, 36–58. Osaka: National Museum of Ethnology.
- 1994b] On departing from verb-final word order. In Hajime Kitamura, Tatsuo Nishida, & Yasuhiko Nagano eds *Current Issues in Sino-Tibetan Linguistics*, 81–97. Osaka: National Museum of Ethnology.
- 1994c] Sangkong of Yunnan: secondary verb pronominalization in Southern Loloish. In Hajime Kitamura, Tatsuo Nishida, & Yasuhiko Nagano eds *Current Issues in Sino-Tibetan Linguistics*, 588–607. Osaka: National Museum of Ethnology. [Same as 1993b]
- 1994d] Introduction to Second Edition, In Randy J. LaPolla & John B. Lowe eds *Bibliography of the International Conferences on Sino-Tibetan Languages and Linguistics I-XXV*, ix–xi. STEDT Monograph Series, No.1–A. Berkeley: Center for Southeast Asia Studies, University of California.
- 1994e] Protean prosodies: Alfons Weidert's *Tibeto-Burman tonology*. *Journal of the American Oriental Society* 114/2:254–258. [Reprinted as 1994i].
- 1994f] Watch out for number ONE: Jingpho ṅāi '1' and ləṅāi 'one' (with some speculations about Jingpho number TWO). *Linguistics of the Tibeto-Burman Area* 17/1:155–165. [Reprinted as 1997d].
- 1994g] On keys and wedges : comment on Sagart's discussion note. *Linguistics of the Tibeto-Burman Area* 17/1:169–171.
- 1994h] Following the marrow: two parallel Sino-Tibetan etymologies. In Karen L. Adams & Thomas J. Hudak eds *Papers from the second annual meeting of the Southeast Asian linguistics society (SEALS 1992)*, 213–234. Tempe, AZ: Program for Southeast Asian Studies, Arizona State University. [Reprinted as 1992d].
- 1994i] Review of Alfons Weidert, 1987, *Tibeto-Burman Tonology*. In *Acta Orientalia* 55:284–292. [See also 1994e].

- 1994j] How dull can you get?: *buttock* and *heel* in Sino-Tibetan. *Linguistics of the Tibeto-Burman Area* 17/2:137–151. [Reprinted as 1998d].
- 1994k] General Editor, *Bibliography of the International Conferences on Sino-Tibetan Languages and Linguistics I-XXV*, by Randy J. La Polla & John B. Lowe. STEDT Monograph Series #1A. Berkeley: Center for Southeast Asia Studies, University of California.
- 1994l] Lùn hóngguān yǔyán bǐjiào. Chinese summary of 1990b (On megalocomparison) by Qiū Fùyuán (a.k.a. Lama Ziwo). *Guowai Yuyanxue* 1994/2:40–41.
- 1995a] Lahu. In Franklin Ng ed. *The Asian American Encyclopedia*, 956–960. New York: Marshall Cavendish.
- 1995b] Sino-Tibetan numerals and the play of prefixes. *Bulletin of the National Museum of Ethnology* (Osaka) [Kokuritsu Minzokugaku Hakubutsukan Kenkyuu Hookoku] 20/1:105–252. [Republished in revised form as 1997c]
- 1995c] (with Yoshio Nishi & Yasuhiko Nagano) eds *New horizons in Tibeto-Burman morphosyntax*. Senri Ethnological Studies #41. Osaka: National Museum of Ethnology.
- 1995d] Sino-Tibetan palatal suffixes revisited. In Yoshio Nishi, James A. Matisoff, & Yasuhiko Nagano eds *New Horizons in Tibeto-Burman Morphosyntax*, 35–91.
- 1995e] Tone, intonation, and sound symbolism in Lahu: loading the syllable canon. In Leanne Hinton, Johanna Nichols & John Ohala eds *Sound symbolism*, 115–129. Cambridge: Cambridge University Press. [Reprinted as 1989e].
- 1996a] Review of Anna Wierzbicka, 1991, *Cross-cultural pragmatics: the semantics of human interaction*. In *Language* 72/3:624–630.
- 1996b] Contact-induced change, genetic relationship, and scales of comparison. In Suwilai Premsrirat et al. eds *Pan-Asiatic linguistics: proceedings of the fourth international symposium on languages and linguistics (Bangkok)*, vol. V, 1591–1611. Nakorn Pathom, Thailand: Institute of Language and Culture for Rural Development, Mahidol University at Salaya.
- 1996c] *Languages and dialects of Tibeto-Burman*. With Stephen P. Baron & John B. Lowe. STEDT Monograph Series #2. Berkeley: Center for Southeast Asia Studies, University of California. (xxx + 180 pp.)
- 1996d] General Editor, *Phonological inventories of Tibeto-Burman languages*, by Ju Namkung. STEDT Monograph Series #3. Berkeley: Center for Southeast Asia Studies, University of California.

- 1996e] Malediction and psycho-semantic theory: the case of Yiddish. In Reinhold A. Aman ed. *Opus maledictorum*, 5–13. New York: Marlowe and Co. [Reprint of 1977a].
- 1996f] The cognate noun/verb construction in Lahu. *Linguistics of the Tibeto-Burman Area* 19/1:97–101.
- 1996g] Remembering Mary Haas's work on Thai. In Leanne Hinton ed. *The Hokan, Penutian, and J.P. Harrington conferences and the Mary R. Haas memorial*, 105–113. Survey of California and Other Indian Languages, Report No.10. Berkeley: University of California.
- 1997a] Primary and secondary laryngeal initials in Tibeto-Burman. In Anne O. Yue & Mitsuaki Endo eds *In memory of Mantaro J. Hashimoto* [Hashimoto Mantarō kinēn chūgoku gogaku ronshū], 29–50. Tokyo: Uchiyama.
- 1997b] Introduction to the Bwe Karen texts. In Eugénie J.A. Henderson, *Bwe Karen dictionary, with texts and English-Karen word list*, edited by Anna J. Allott, vol. I, xiv–xxiii.
- 1997c] *Sino-Tibetan numeral systems: prefixes, protoforms and problems*. Canberra: Pacific Linguistics. (xi + 136 pp.)
- 1997d] Watch out for number ONE: Jingpho ṅāi '1' and ləṅāi 'one' (with some speculations about Jingpho number TWO). In Arthur S. Abramson ed. *Southeast Asian linguistic studies in honour of Vichin Panupong*, 161–169. Bangkok: Chulalongkorn University Press. [Reprint of 1994f].
- 1997e] (with Victor Golla & Pamela Munro) Obituary: Mary R. Haas. *Language* 73/4:826–837.
- 1997f] Remembering Mary R. Haas's work on Thai. *Anthropological Linguistics* 39/4:594–602.
- 1997g] Tonal correspondences in the checked syllables of Proto-Yi, Proto-Burmish, Naxi, and Jingpho: evidence against the 'monogenetic' theory of tonogenesis. *Yí-Miǎn-yǔ yǎnjiǔ* [Studies on Yi-Burmese Languages], 126–170. (In English, with Chinese abstract 126–7). Chengdu: Sichuan Nationalities Publishing House.
- 1997h] Editor's Preface to Fu Maoji, A descriptive grammar of Lolo. *Linguistics of the Tibeto-Burman Area* 20/1:1–ix.
- 1997i] In memoriam: Paul K. Benedict (1912–1997). *Linguistics of the Tibeto-Burman Area* 20/2:1–8.
- 1998a] Dàyáng Pumi phonology and adumbrations of comparative Qiangic. *Mon-Khmer Studies* 27:171–213.

- 1998b] Aspects of aspect, with special reference to Lahu and Hebrew. In Yasuhiko Nagano ed. *Time, language, and cognition*, 171–215. Senri Ethnological Studies, No.45. Osaka: National Museum of Ethnology.
- 1998c] Dào Bǎoluó Běnnídíkètè. *Minzu Yuwen* 1998/2:80. Chinese translation by Jiang Di of In Memoriam: Paul K. Benedict, presented at ICSTLL 30, Beijing (August 1997).
- 1998d] “How dull can you get?” *buttock* and *heel* in Sino-Tibetan. In Pierre Pichard & François Robinne eds *Etudes Birmanes en hommage à Denise Bernot*, 373–383. Etudes Thématiques 9. Paris: Ecole Française d’Extrême-Orient. French title (p. 448): Jusqu’où pouvez-vous être ennuyeux?: *fesse et talon* en sino-tibétain. [Reprint of 1994j].
- 1999a] Tibeto-Burman tonology in an areal context. In Shigeki Kaji ed. *Proceedings of the symposium ‘Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics’*, 3–32. Followed by Atsuhiko Kato, Comments, 33–35. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- 1999b] In defense of Kamarupan. *Linguistics of the Tibeto-Burman Area* 22/2:173–182.
- 1999c] Méiguó zhùmǐng Zàng-Miǎn yǔyánxuéjiā Mǎtísuǒfù Jiàoshòu chéngtòng dàoniàn Mǎ Xuéliáng Jiàoshòu. Translation by Xu Shixuan of a message of condolence sent on the death of Ma Xueliang. *Zhōngguó Mínzú Yǔyán Xuéhuì Tōngxùn* (Newsletter of the Association of Chinese Minority Language Studies) 1999/2:4.
- 2000a] *Blessings, curses, hopes, and fears: psycho-ostensive expressions in Yiddish*. Republication of 1979a, with a new introduction by the author. Hardcover and paperback. Stanford, CA: Stanford University Press. (xxx + 160 pp.)
- 2000b] An extrusional approach to \**p/w*- variation in Sino-Tibetan. *Language and Linguistics* 1/2:135–186. Taipei: Academia Sinica, Institute of Linguistics.
- 2000c] On the uselessness of glottochronology for the subgrouping of Tibeto-Burman. In Colin Renfrew, April McMahon, & Larry Trask eds *Time depth in historical linguistics*, 333–371. Cambridge: The McDonald Institute for Archaeological Research.
- 2000d] On ‘Sino-Bodic’ and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63/3:356–369.
- 2000e] Three Tibeto-Burman/Sino-Tibetan word families: set (of the sun); pheasant/peacock; scatter/pour. In Marlys Macken ed. *Proceedings of the tenth annual conference of the Southeast Asian linguistics society (SEALS)*, 215–232. Tempe, AZ: Program for Southeast Asian Studies, Arizona State University.

- 2001a] Yòng xiēzi qiào kāi wèntí : Using a wedge to pry open a problem. *Yuyan Yanjiu* (Language Research) 42:106–127.
- 2001b] The interest of Zhangzhung for comparative Tibeto-Burman. In Yasuhiko Nagano & Randy J. LaPolla eds *New research on Zhangzhung and related Himalayan languages*, 155–180. *Bon Studies* 3. *Senri Ethnological Reports* #19. Osaka: National Museum of Ethnology.
- 2001c] Yuánshǐ Hàn-Zàng-yǔ / Yuánshǐ Zàng-Mián-yǔ gòunǐ de xiànzhuàng. Chinese translation by K. K. Luke (Lu Jingguang) & Zhang Zhenjiang, of The present state of PST/PTB reconstruction: can we even write a fable in Proto-Lolo-Burmese? *Yǔyánxué wèntí jíkān* (*Problems in linguistics*) 1/1:41–58. Published jointly by Heilongjiang University and Hong Kong University Departments of Linguistics. Jilin: People's Publishing Co. English version to appear 2002.
- 2001d] Review of Shobhana L. Chelliah, 1997, *A grammar of Meithei*. In *Anthropological Linguistics* 43/2:246–251.
- 2001e] Genetic vs. areal linguistics in Southeast Asia: prosodic diffusibility in Southeast Asian languages. In Alexandra Y. Aikhenvald & R.M.W. Dixon eds *Areal diffusion and genetic inheritance*, 291–327. Oxford: Oxford University Press.
- 2002a] *Zài lùn Yí-Miǎn yǔzhī de shēngdiào yǎnbiàn: gǔ bìyīnjiécí de shēngdiào fēnhuà*, (1972a) translated by Lin Ying-chin, with a new introduction by the author. Taipei: Institute of Linguistics, Academia Sinica.
- 2002b] Wedge issues. *Linguistics of the Tibeto-Burman Area* 25/1:137–164.
- 2002c] The present state of PST/PTB reconstruction: can we ever write a fable in Proto-Lolo-Burmese? *Linguistics of the Tibeto-Burman Area* 25/2:225–242.
- 2003a] Lahu. In Graham Thurgood & Randy LaPolla eds *The Sino-Tibetan Languages*, 208–221. London: RoutledgeCurzon.
- 2003b] Southeast Asian languages. In William Frawley & Bernard Comrie eds *International encyclopedia of linguistics*, Second edition. New York and Oxford: Oxford University Press.
- 2003c] Aslian: Mon-Khmer of the Malay peninsula. *Mon-Khmer Studies* 33/1:1–57.
- 2003d] *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley & Los Angeles: University of California Press.

**Submitted for publication:**

1. [submitted Sept. 1997]  
Tibeto-Burman languages and linguistics. MS: text, 29 pp; references, 17 pp; 6 figures. To appear in the next edition of *Encyclopedia Britannica*. Submitted by invitation.
2. [submitted 2001]  
On the linguistic position of Bai within Tibeto-Burman. To appear in Chinese translation by Fu Jingqi.
3. [submitted 2002]  
Is there such a thing as an areal semantics of Southeast Asia? In Anju Saxena & Jadranka Gvozdanovic eds *Himalayan languages, past and present*.
4. [submitted 2003]  
Appreciating the diversity of prosodic systems, and of linguists' approaches to them. In Shigeki Kagi ed. *Proceedings of the third international symposium on cross-linguistic studies of tonal phenomena*. Tokyo University of Foreign Studies 17–19 December 2002.

**Festschrift:**

- 2003] David Bradley et al. eds *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*. Canberra: Pacific Linguistics.

**In Preparation:**

1. [to be submitted 2003]  
*English-Lahu dictionary*. With the assistance of J.B. Lowe, Zev Handel, David Mortensen, Richard S. Cook & Mischa Park-Doob. A computerized reversal of 1988b. To be submitted to University of California Press.
2. [to be submitted 2003]  
The linguist's dilemma: l- / d- interchange in Sino-Tibetan. To be submitted to *Journal of Chinese Phonology*.
3. [to be submitted 2003]  
A modest start toward the sorting of materials for the reconstruction of Proto-Qiangic.



---

20 *David Bradley*

4. [in preparation]

*Languages of mainland Southeast Asia*. To be published in the Cambridge Language Surveys series, by Cambridge University Press.

5. Chinese translation of 1997h (editor's preface to Fu Maoji, *A descriptive grammar of Lolo*) by Fu Jingqi.

---

## 2 *Hmong secret languages: themes and variations*

---

MARTHA RATLIFF

It seems fitting to me to contribute an article on secret languages, which, in order to serve their primary function, must be created anew and with a difference for each new purpose, to a volume honoring Jim Matisoff and his life-long respect for the richness and diversity of human language. The role of secret languages differs from culture to culture (for those cultures which can boast of them) but whether used for play or serious subterfuge, they would not be worthy of the name 'secret' if they were immediately understandable to a native speaker not privy to its rules. The calcified Pig Latin of English fools no one, because its rules are well understood. But in a culture where secret languages do more work, it is important that speakers continually find ways to use 'secret language formation rules' to create new disguises so that sensitive communications can indeed stay secret.

In the closely related White and Green dialects of Hmong (Hmong Daw and Mong Leng) as spoken in Laos, Thailand, and communities of displaced speakers from Southeast Asia in the West, there seems to be no upper limit to the number of secret languages that can be produced. They are most widely referred to as *lus rov* in White Hmong or *lug rov* in Green Hmong, which means 'turned-back language'. Although less isolated Hmong people in both Southeast Asia and in the West are losing this oral ability as they are losing other aspects of the oral tradition, including their variety of oral poetry, reminiscent in its structure to the oral epic poetry of western cultures (Mottin 1980), some older Hmong people in the United States can still use secret languages, explain their function, and, to some extent, even explain what they are doing when they perform these acts of linguistic contortion.

In this paper I will categorize and exemplify the Hmong secret languages I have collected in interviews with members of the Hmong community in Detroit, Michigan.

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 21–33.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

Hmong secret languages have already been described in a number of published papers in English (Ai 1972, Chu 1972, Derrick-Mescua et al. 1982, Catlin 1997) and I know of one brief record of four secret languages published in Hmong (Vang 1994). The Derrick-Mescua paper is especially rich, describing ten different types of *lug rov*, all similar to, but not identical to those described here.<sup>1</sup> In addition to making more of the same type of data available to a wider audience, I also reproduce translations of excerpts from the interviews in which the speakers talk about secret language function and creation. Finally, I will point out what these languages add to our understanding of Hmong phonology, as well as what issues they raise for the phonologist and the psycholinguist.

I am indebted to Xing Zer Kue and Ia Kue, prominent members of the Greater Metropolitan Detroit Hmong community, who located older members of the community who still could use secret languages, assisted in the interviews, and did most of the translations. Pang Foua Her, one of my linguistics students, interviewed her mother for a class term paper on secret languages, and her data have also been included (1994). Finally, I would like to thank these women who are proficient in secret languages and were willing to talk with us: the cousin of Yee Moua, Kalia Vang Vue, Chue Vang Yang, Mao Her Chang, the mother-in-law of Shoua Lao, and Za Her.<sup>2</sup>

## 1 The function of secret languages

From published descriptions and interviews, it seems that secret languages are used whenever it is important to keep conversations private: in courtship, between friends, when parents do not want their children to understand them, or when children do not want their parents to understand them. Two contexts of use were described to us in more detail by Mao Her Chang. They may be useful in demonstrating that secret languages were occasionally used for serious business as well as for privacy, for which reason I prefer the term 'secret language' to the equally familiar term 'language game'. For example:

When our ancestors came from China, everyone knew *lug rov*. They learned it so another group or tribe could not know what they were saying ... [W]e had to use *lug rov* so the Chinese and other groups could not understand. It does not seem like a good thing to do, but we did not want them to know. For example, if we wanted to poison the fish, then we would use *lug rov* so that no one would know.

(Mao Her Chang, translation Ia Kue)

<sup>1</sup> The word for 'language' is *lug* in Green Hmong and *lus* in White Hmong. For simplicity's sake, I have chosen to represent this word exclusively as *lug*.

<sup>2</sup> Although all of the people we interviewed were women, it has been reported to us and to others that the use of Hmong secret languages is not limited to women.

Courtship is playful, but it is serious business as well. Traditionally, young people both initiated and concluded their search for marriage partners during the few days of the New Year's festival at the end of the year, during which time games were played to test the skills of potential partners (Mottin 1979).

About fifty per cent of my mother's generation knew *lug rov*. Their generation did not use it with the same purpose as the older generation. Their use was mainly for girlfriend-boyfriend talk. Someone would speak *lug rov* to you, and then you must speak back to them in *lug rov*. But in answering back you must come up with a new *lug rov* and so on ... If you don't know, then you lose. That is why we have a saying:

*Koj txhais tsis tau koj tso tes plaus*

*Koj txhais tsis tiav ces koj tso kev khiav*

'If you cannot interpret this, you must release my hand;

If you don't finish interpreting this, then take the road and run.'

(Mao Her Chang, translation Ia Kue and Martha Ratliff)

## 2 Hmong morpheme structure and orthography

As is typical of the area, the morpheme in Hmong is monosyllabic, and word derivation is primarily by compounding. There are fifty-eight possible onsets in White Hmong and fifty-six possible onsets in Green Hmong. These onsets include prenasalized stops and affricates, clusters with medial /l/, and voiceless (preaspirated) sonorants. Aspiration is distinctive, but voicing is not.<sup>3</sup> There are phonemic distinctions at eight different places of articulation: labial, alveolar, retroflex, palato-alveolar, palatal, velar, uvular, and glottal. In the Romanized Popular Alphabet used in this paper, /s/ is symbolized with *x*, /ʃ/ and /z/ are symbolized with *s* and *z* respectively, and /ç/ is symbolized with *xy*. All other spellings are easily related to their phonetic values.

The rime inventory is simpler. Open syllables are the rule; the only coda consonant is /-ŋ/ and it has a limited distribution. In the RPA, rimes with a nasal coda and nasalized vowels are marked by doubling the vowel, /ɨ/ is symbolized with *w* and all other spellings are easily related to their phonetic values.

---

<sup>3</sup> What are given here as preglottalized voiced initials in White Hmong have only recently developed from clusters with medial /l/ (Jarkey 1987). They correspond to the Green Hmong clusters in the column to the immediate left.

**Table 1:** Onset inventory (those limited to White Hm in bold, to Green Hm in italics)

p	pl	t	<i>tl</i>	?d	ts	t	tʃ	c	k	q	ʔ
p <sup>h</sup>	p <sup>h</sup> l	t <sup>h</sup>	<i>t<sup>h</sup>l</i>	?d <sup>h</sup>	ts <sup>h</sup>	t <sup>h</sup>	tʃ <sup>h</sup>	c <sup>h</sup>	k <sup>h</sup>	q <sup>h</sup>	
mp	mpl	nt	<i>ntl</i>		nts	nt	ntʃ	ɲc	ŋk	ŋq	
mp <sup>h</sup>	mp <sup>h</sup> l	nt <sup>h</sup>	<i>nt<sup>h</sup>l</i>		nts <sup>h</sup>	nt <sup>h</sup>	ntʃ <sup>h</sup>	ɲc <sup>h</sup>	ŋk <sup>h</sup>	ŋq <sup>h</sup>	
<b>hm</b>	<b>hml</b>	<b>hn</b>						<b>hɲ</b>			
m	ml	n						ɲ	ŋ		
f		s	hl			ɛ		ɕ			h
v			l			ʒ		j			

**Table 2:** Rime inventory (those limited to Green Hmong in italics)

Simple rimes	i	ɨ	u
	e	a	ɔ
Nasal rimes	ɛŋ	ãŋ	ɔŋ
Diphthongs	ia		ua
	ai	aɨ	au

There are seven primary tones and an eighth minor falling–rising tone which is a conditioned variant of the seventh tone. The starting and ending points of the tones are indicated by the pair of numbers following each word (on a scale where 5 is the highest pitch and 1 is the lowest pitch). In the RPA, the tones are marked by final consonants, as given in the middle column below.

**Table 3:** Tone inventory (common to both dialects, examples from Green Hmong)

[au 55	-b	to scratch the ground
[au 52	-j	a hammer
[au 24	-v	to light a fire
[au 33	-ø	six
[au 21	-s	to be full (of food)
[au 42 (breathy)	-g	to suffer (something)
[au 21? (creaky)	-m	kidney

### 3 Types of secret languages: *lug ntxeev* and *lug rov*

I collected nine secret languages in all. Although more extensive exemplification could have been presented, very little is actually needed in order to demonstrate the rules of each language. According to speakers, the simplest kind of secret language is *lug ntxeev* 'turned-over language'. In this type of language, there are no added nonsense syllables. Two secret languages of this type were described. In both of these languages and in the first four examples of *lug rov* (languages 1 through 6), the tone of the real word as indicated by the final consonant symbol always stays with the rime of the real word.

- 1) In this language, pairs of words exchange rimes while the onsets remain in place. This language was described by Pang Foua Her.

*koj le* 'you poss. = yours' > *ke loj*  
*Paaj Fuab* (her name) > *Puab Faaj*

- 2) Here sets of three words are involved. The middle word remains unchanged and the first and third words exchange rimes. This language was described by Pang Foua Her, Kalia Vang Vue, and Chue Vang Yang. Pang Foua Her described this as 'putting first word last and switching first letter'.

*mus ua ntej* 'go first' > *mej ua ntus*  
*kuv yog hmoob* 'I am Hmong' > *koob yog hmuv*

The second type of secret language, *lug rov* 'turned-back language', is more complicated. All examples of *lug rov* involve the addition of one or more full or partial nonsense syllables. They involve permutation of onsets, rimes, and/or tones as well. In this and all subsequent examples, the elements of the disguised real words have been underlined to make the pattern easier to see. The first four *lug rov* involve the addition of one or more full nonsense syllables.

- 3) The simplest of these involves one full nonsense syllable, here *leem*, *lab*, or *loob*, added to each individual word to make a pair of 'words' which disguise the one. The original onset is replaced by the nonsense onset, followed by the original onset with nonsense rhyme. These languages (which presumably can take as many forms as possible nonsense syllables) was described by Pang Foua Her and by Yee Moua's cousin.

with *leem*: *kuv* 'I' > *luv keem*  
 with *lab*: *kuv* > *luv kab*  
 with *loob*: *kuv* > *luv koob*

- 4) An extension of 3, here two full nonsense syllables may also be added in a three-syllable pattern. Here *lab* and *leem* are the nonsense syllables. The original onset is replaced with the nonsense onset, followed by the original onset paired with each of the two nonsense rhymes. This *lug rov* was described by Mao Chang Her.

*kuv* 'I' > *luv kab keem*

Yee Moua's cousin described the same language using *loob lia* as the nonsense words:

*kuv* 'I' > *luv koob kia*

- 5) Two full nonsense syllables, here *xab* and *xwj*, may also be added to one real word to create a three-syllable pattern: the original rhyme is replaced by the nonsense rhyme *-wj*, followed by entire unbroken nonsense syllable *xab*, followed by original word with the onset replaced with nonsense onset *x-*. This language is actually not very difficult to decipher, because all the hearer needs to do is pay attention to the beginning and ending of each set of three syllables. This language was described by Kalia Vang Vue and Chue Vang Yang.

*kuv* 'I' > *kwj xab xuv*

- 6) The next *lug rov* is the most remarkable of those I have heard. It was given to us by Yee Moua's cousin. Again, two full nonsense syllables, here *lab* and *leem*, are added in a two-syllable pattern. But what differentiates this from the others is that the nonsense syllable chosen is dependent upon syntactic constituent boundaries: *lab* is used phrase-medially, and *leem* is used phrase-finally. A longer sample is necessary to illustrate this pattern.

*Koj cov mov qab qab.* > [*Loj kab lov cab lov meem*] [*lab qeem*]  
 You CLF rice sweet sweet NP VP  
 'Your rice is good.'

*Kuv nyiam koj.* > [*Luv keem*] [*liam nyab loj keem*]  
 I like you NP VP  
 'I like you.'

*Kuv tau yuav ib tug npua.*  
 I get buy one CLF pig  
 'I bought a pig.'  
 > [*Luv keem*] [*tau tab luav yeem*] [*lib ab lug tab lua npeem*]  
 NP V NP

Notice that the entire VP is marked as a phrase with a pronominal object in the second example, but when the object is a full noun phrase as in the third example, the verbal complex is marked as one phrase and the object is marked as a separate phrase.

The last three *lug rov* involve the addition of one or more partial nonsense syllables. The last language also involves the wholesale substitution of a nonsense tone pattern over the entire utterance.

- 7) In this language, only a new onset (*y-*) and a new tone (*-m*) are added to the word to be disguised. In a two-syllable pattern, the nonsense tone replaces the actual tone of the word to be disguised, and then the nonsense onset replaces the onset of a copy of the same word. As language 5 above, this language is not very difficult to decipher, because all the hearer needs to do is pay attention to the beginning and ending of each set of two syllables. This language was described by Pang Foua Her, Za Her, and Mao Her Chang.

<i>kuv</i>	'T'	>	<i>kum yuv</i>
<i>puas</i>	Q particle	>	<i>puam yuas</i>
<i>zoo</i>	'good'	>	<i>zoom yoo</i>

- 8) In this language, two partial nonsense syllables are added in a three-syllable pattern. The first partial nonsense syllable is the toneless *ncaa* and the second partial nonsense syllable is just the onset *l-*. The onset of the disguised word appears in the first syllable only; the rime and tone of the disguised word appear in the second two syllables. The rime of *ncaaa* appears in the first syllable and the onset of *ncaaa* appears in the second. The onset *l-* appears in the third syllable. All syllables have the tone of the disguised word. This language is clearly easier to comprehend from the example than to describe! It was given to us by Za Her.

<i>kuv</i>	'T'	>	<i>kaav ncu<sub>v</sub> lu<sub>v</sub></i>
------------	-----	---	--------------------------------------------

- 9) The final *lug rov* combines the addition of a partial nonsense syllable and the substitution of an alternating nonsense tone pattern for the lexical tones. We recorded this language on two different occasions with three different speakers. The first, given to us by Kalia Vang Vue and Chue Vang Yang, involves a nonsense syllable made up of the onset *y-*, a copy of the vowel of the preceding disguised word, and the *-v* (rising) tone. These added syllables are pronounced between the real words, which are pronounced with the *-s* (low level) tone in place of the lexical tone. The 'melody' of this language is thus low-rise-low-rise-low-rise. This is clearly a deeper disguise than some of the other languages because the lexical tones have been eliminated, and are only recoverable from context.



*Koj yuav mus dab tsi* > *Kos yov mus yuv das yav tsis yiv?*  
 you will go what (yuav is not encoded)  
 'What are you going to do?'

Shoua Lao's mother-in-law gave us the same pattern with different variables: a nonsense syllable made up of the onset *tx-*, a copy of the vowel of the preceding disguised word, and the *-g* (breathy) tone, alternating with the real words pronounced with the *-j* (high falling) tone. The melody of this language is thus fall-breathy-fall-breathy-fall-breathy.

*Kuv tsis nyam koj* > *Kuj txug tsij txig nyaj txag koj txog*  
 I not like you  
 'I don't like you.'

These nine languages only represent the tip of an iceberg; this is clear from an examination of the other sources on Hmong secret languages listed below. Some of the languages described in these sources are reminiscent of those described above, but others present new twists: Catlin (1997) describes a sung *lug rov*, in which the real words are sung quickly on changing eighth notes between quarter notes. A series of distracting (rhyming) nonsense syllables are sung on the longer, and hence more salient quarter notes, which are always sung on the same pitch. In one of the *lug rov* published by Bliia Tcheu Vang (1994), there are two layers of disguise: once you have unscrambled the flipped onsets and rimes and have removed the nonsense elements, you find that the vowels of some or all of the real words have been replaced, so that *koj mus dab tsi* 'what are you going to do?' decoded once, is *koj mas dab tsa*, in need of yet more decoding.

#### 4 A window into linguistic structure

These secret languages, taken together, give clear support for two aspects of phonological structure which have now become well-established: the hierarchical structure of the syllable, and the independence of tone as captured in autosegmental models of phonology.

All of these secret languages move, copy, and/or replace the onsets and rimes of syllables, as opposed to whole syllables or individual phonemes. And although the first six of these languages might lead one to think that the tone was necessarily fused with the rime—since in these languages wherever the rime goes, the tone goes, too—the last three languages show that tone may be split off and changed independently from the rime. Secret languages may thus be added to the following types of indirect evidence for these units of linguistic structure:

- 1) the native Pahawh Hmong writing system, in which onsets and rimes are

symbolized and in which, over several versions, tone gradually came to be symbolized separately from the rime (Smalley et al. 1990), and;

- 2) those forms of Hmong instrumental music which directly represent language through tone-to-musical pitch associations (Catlin 1997).

Secret language 6 may give evidence for syntactic constituent structure as well as the significance of noun phrase 'weight': in this language that marks the ends of phrases with a different nonsense syllable, an object consisting of several words with a common noun as its head is treated as a separate constituent whereas a pronoun object is treated as part of the verb phrase. I would not like press this too far, however: upon reviewing the secret languages in Derrick-Mescua (1982) that alternated nonsense syllables, it was much harder to see a link with phrase structure. It may be that some speakers mark phrase structure and others don't, or considerations of euphony may govern the alternation of the nonsense syllables, and the change of to a different nonsense syllable phrase-finally only occurred in my sample by accident.

These secret languages also demonstrate that the syllable is the most salient unit of linguistic structure in Hmong. The rule seems to be that, with minor exceptions, every syllable gets its own disguise, even meaningless syllables in opaque compounds and disyllabic borrowings. For example, both syllables in the opaque compound *dabtsi* 'what' were disguised in secret language 9: *das yav tsis yiv*. And the borrowing *fabkis* 'français' in secret language 6 was pronounced *lab fab lis keem*. This contrasts with secret languages like English Pig Latin, in which structural disguise operates at the level of the word: 'Latin' is 'atin-lay', not 'a-lay in-tay'.

There are two exceptions to the rule that every syllable gets its own disguise. The occasional function word can simply be skipped (see language 9, where the future marker *yuav* is not encoded). More interestingly, reduplicated morphemes only get encoded once, or only once-and-a-half. In secret language 6, reduplicated *qab* 'sweet, delicious'—*qab qab* 'very delicious'—is only encoded once: *lab qeem*. And in secret language 5, the reduplicated verb *nco* 'miss (someone)'—*nco nco* 'miss (someone) greatly'—is not simply an encoding of a single *nco*, which would have been *ncwj xab xo*, nor is it an encoding of both syllables, which would have been *ncwj xab xo ncwj xab xo*, but comes out *ncwj ncwj xab xo*. The onset is paired with the nonsense rime twice, but the rime is only encoded once. This echoes the pronunciation of reduplicated words in Hmong, in which the first syllable is very short and has a reduced vowel, and suggests that reduplication in Hmong is a derivational rather than a syntactic process.

Secret languages also potentially provide a test for synchronic co-occurrence restrictions. Presumably, some co-occurrence restrictions are so strong that even secret language speakers would quickly adapt in order to avoid a highly illegal form, whereas other co-occurrence restrictions are not as strong, and forms illegal in the 'regular'

language would be allowed in this special context, much as they are in the exceptional phonology of expressives (Ratliff 1992, Chapter 4). The co-occurrence restriction between [+spread glottis] initials—which includes aspirated stops and affricates, voiceless fricatives, and voiceless sonorants—and the breathy (-g) tone is apparently not inviolable. One speaker using secret language 5, in disguising the name *Ntxawg*, produced *Ntxwj xab xawg*. *Xawg* [saɪ̯ 42, breathy tone] contains the combination of a voiceless fricative and the breathy tone, which normally do not co-occur.

Finally, these secret languages also give us information about which consonants Hmong treats as default, or neutral consonants. In secret language 4, I observed a dissimilation effect when strictly following the rules would result in a string of like onsets. The words to be disguised in each case began with either *hl-* or *l-*. The onsets of the two nonsense syllables in this language are also *l-*. To effect the dissimilation, one of the consonants *k-*, *n-*, or *y-* was chosen to replace not the nonsense onset, but the real onset *hl-*, while *r-* was chosen to replace the real onset *l-*. This must have made these words much harder to decipher. For example, *hlub* 'to care for' did not become \**lub hloob hlia*, but either *lub koob kia* or *lub noob nia*; *hlob* 'elder' became not \**lob hloob hlia* but *lob yoob yia*; and *lus* 'language' did not become \**lus loob lia*, but rather *lus roob ria*. It is probably not accidental that one of these replacement consonants, the *k-*, is the only consonant not symbolized in the native Pahawh Hmong orthography. In this onset-rime writing system, if the reader sees no onset symbol for a given word, he or she is to supply a *k-* onset. This is remarkable in that this writing system has not only a symbol for a glottal stop onset, but also has a symbol for the *absence* of an onset (Smalley et al. 1990, Chapter 4).

## 5 Phonological awareness

Equally significant to their implications for linguistic analysis are the psycholinguistic questions raised by these highly complex secret languages. Simply put, how do speakers do it?<sup>4</sup>

Prerequisites for speaking *lug rov* must be both the understanding that units smaller than the syllable exist and the ability to segment the speech stream into these units. Secret languages may thus pose a challenge to those psycholinguists who have claimed that 'phonological awareness', or awareness of linguistic units smaller than the syllable, can

<sup>4</sup> Mao Her Chang reported that speaking *lug rov* is a matter of talent; not everyone is able to do it, or do it well. Talent played an acknowledged role in the phonological awareness experiments with Chinese subjects in Read et al. (1986), and was also mentioned in connection with the Miao secret language discussed by Chu T'ing (1972).

only develop through learning an alphabetic writing system (Morais et al. 1979, Read et al. 1986):

'[S]egmentation skill' ... does not develop with cognitive maturation, non-alphabetic literacy [as the ability to read Chinese], or exposure to a language rich in rhymes and other segmental contrasts. It does develop in the process of learning to read and write alphabetically. (Read et al. 1986:31-32)

To test this limited (and Western-biased) notion of phonological awareness, we attempted to set up and record a situation in which two speakers of *lug rov* would work out the rules of a new secret language on-the-spot. We were not able to do this, but we did interview two speakers and asked them to teach us how to speak a secret language. The first session was held with Mao Her Chang, fifty-five years old at the time, and literate in Hmong (translation Ia Kue).

Mao Her Chang: I will teach the simplest type of *lug rov* first [language 7]. All you have to do with this type of *lug rov* is add the sound 'yos'. For example, the word *koj* ['you'] can be said *kom yoj* ... One simply adds the sound 'yos' to the end of the word

Xing Zer Kue: Do you know it's the 'yos' you're adding?

Mao Her Chang: We don't know that we are adding the 'yos', but we can hear the sound 'yos yos' in it.

Xing Zer Kue: So you can hear the sound, is that correct? Back then, did you know how to read and write? Did you know about the 'yos' or letter 'y'?

Mao Her Chang: We did not know, but now that I do know I've figured out that we are just adding a sound to it. When you say it, the sound, whether it is a high or low pitch depends on what vowel and tone you add to make the sound. When we would just speak it we would not know what it meant, but after we learned it then we knew.

The second session was held with Za Her, ninety years old at the time, and illiterate in Hmong (translation Ia Kue).

Za Her: We don't know the way you know things now. I don't know how to read or write, so I will just say this and you can add whatever you want to it.

Ia Kue: Can you teach me so that I can create my own?

Za Her: You just listen to the way I say it and then you just know. If you can say what I've been teaching you then you can create your own ... right away. It doesn't take a long time!

These exchanges suggest that literacy, and perhaps also analytic habits of mind that come with education, are necessary in order for speakers to be able to articulate what is

happening when they use secret languages (Koopmans 1987). But psycholinguists do not measure phonological awareness by the speaker's ability to describe tasks, but rather by their ability to perform tasks. Notice that Read et al. (1986) link phonological awareness to 'segmentation skill', not the ability to reflect upon that skill. I would think that the simple laboratory tasks that have been used by psycholinguists to measure phonological awareness—for example, upon hearing the stimulus 'pa', the subject must suppress 'p-' and produce '-a'—would be child's play for the Hmong women interviewed in this study, whether literate or illiterate.

## References

- Ai, Ch'ing, 1972, A Miao secret language. In Herbert R. Purnell, ed., 235–236. (First published 1957 in *Zhongguo Yuwen* 59:15.)
- Catlin, Amy, 1997, Puzzling the text: thought-songs, secret languages, and archaic tones in Hmong music. *The World of Music* 39/2:69–81.
- Chu, T'ing, 1972, A comment on 'A Miao secret language'. In Herbert R. Purnell ed., 237–238. (First published 1957 in *Zhongguo Yuwen* 62:36.)
- Derrick-Mescua, Maria, Judith Berman and Mary Beth Carlson, 1982, Some secret languages of the Hmong. In Bruce T. Downing and Douglas P. Olney eds *The Hmong in the West: observations and reports*, 142–159. Minneapolis: University of Minnesota Southeast Asia Refugee Studies Program.
- Her, Pang Foua, 1994, Hmong language play. Class paper, Linguistics 272, Wayne State University.
- Jarkey, Nerida, 1987, An investigation of two alveolar stop consonants in White Hmong. *Linguistics of the Tibeto-Burman Area* 10/2:57–70.
- Koopmans, Matthijs, 1987, Formal schooling and task familiarity: a reply to Morais et al. *Cognition* 27:109–111.
- Morais, José, Luz Cary, Jesús Alegria and Paul Bertelson, 1979, Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* 7:323–331.
- Mottin, Jean, 1979, *Fêtes du nouvel an chez les Hmong Blanc de Thaïlande*. Bangkok: Don Bosco Press.
- 1980, *55 Chants d'amour Hmong Blanc*. Bangkok: Don Bosco Press.
- Purnell, Herbert C. Jr., ed., 1972, *Miao and Yao linguistic studies: selected articles in Chinese* (translated by Chang Yü-hung and Chu Kwo-ray). Data Paper 88, Linguistics Series V. Ithaca: Cornell University Southeast Asia Program.
- Ratliff, Martha, 1992, *Meaningful tone: a study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong*. Monograph Series on Southeast

Asia, Special Report No. 27. DeKalb, IL: Center for Southeast Asian Studies, Northern Illinois University.

Read, Charles, Zhang Yun-fei, Nie Hong-yin and Ding Bao-qing, 1986. The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24:31-44.

Smalley, William, Chia Koua Vang and Gnia Yee Yang, 1990, *Mother of writing: the origin and development of a Hmong messianic script*. Chicago: University of Chicago Press.

Vang, Bliat Tcheu, 1994, Lus rov. *Liaj Luv Chaw Tsaws Tsab* 3/2.



# 3 *Variegated tonal developments in Tibetan*<sup>1</sup>

---

JACKSON T.-S. SUN

## 1 Introduction

Is the presence of analogous tone systems indicative of a genetic relationship among languages, or a more intimate genetic relationship among dialects of the same language? In a classic article on the subject of tonogenesis, we find definitive answers to both questions in the negative:

Tonal similarities—even regular tonal correspondences—are not to be taken uncritically as evidence for genetic relationship among languages. Indeed, tonal criteria are not even sufficient to establish genetic subgroupings for languages that are already known to be genetically related. (Matisoff 1973:89)

In this pioneering work Professor Matisoff portrayed, with his characteristic humor and insight, the fundamental mechanisms of tonogenesis in monosyllabic languages in terms of transphonologisation of inherent pitch perturbations of the surrounding consonants of a syllable into phonemic tone on the vocalic nuclei. He argued convincingly that these mechanisms may bring about typologically similar tone systems in unrelated languages provided that the languages in question are already drifting toward monosyllabic morpheme structure. Indeed, the tonogenetic scenarios he depicted apply not only to languages with predominant monosyllabic morphemes (Hmong-Mien, Tai-Kadai), but also tonal languages which, at an earlier stage, were incontestably sesquisyllabic, such as the

---

<sup>1</sup> An earlier version of this paper was presented at the 34th International Conference on Sino-Tibetan Languages and Linguistics (Kunming, Oct. 24-27, 2001). The helpful comments I received from Randy J. LaPolla, Brigitte Huber, and from the ICSTLL-34 audience are highly appreciated.

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds., *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 35–51. Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.



Mon-Khmer languages Vietnamese and Va,<sup>2</sup> or polysyllabic, such as the Austronesian languages Jabêm (Bradshaw 1979) and Tsat (Thurgood 1996).

Owing in part to Professor Matisoff's influential work, most Sino-Tibetanists today will not accept Hmong-Mien and Tai-Kadai as members of Sino-Tibetan on the basis of tonal parallels alone. An as yet unsettled area concerns modern Tibeto-Burman languages like Tibetan and Qiang,<sup>3</sup> which have both tonal and atonal dialects. With such languages, the question is whether the tonal dialects are more closely related to each other vis-à-vis the atonal ones. Indeed, the presence of Old Tibetan (henceforth OT) preradicals and phonemic tones were the main criteria in Róna-Tas' tentative subclassification of Tibetan dialects into 'archaic' and 'non-archaic' subgroups (1966:21). Likewise, Hu (1991:177) treats phonemic tones as the highest-order subgrouping criterion, classifying modern Tibetan thereby into tonal and non-tonal dialects. For these claims to make sense for genetic (rather than typological) classification, tonogenesis would have to be a unique set of phonological innovations occurring in a putative common ancestor of all the tonal dialects as it branched off from the atonal dialect(s). Qu (1988:321) explicitly denies the monogenesis possibility of Tibetan tone and suggests that tone arose first in the dBus-gTsang region during the eighth and ninth centuries and gradually diffused eastward to originally non-tonal Tibetan areas, without giving concrete evidence for the alleged spread of tone across dialect boundaries.

It will be my objective to advocate herein the alternative view that, tonogenesis being such a prevalent sound change areally and typologically, the possibility of tone emerging independently in different Tibetan-speaking areas cannot be lightly dismissed, particularly since different tonal dialects seem to have traveled divergent tonogenetic paths, as I wish to illustrate in the present study. The subsequent sections examine in turn the evolution of (distinctive as well as allophonic) tonal registers (§2), tonal contours (§3), as well as neutralisation of tone (§4) in various dialects reflecting different stages on the Tibetan tonality cline (J. Sun 1997:487). In §4, I summarise the heterogeneous paths of tonogenesis examined in this study and discuss their direct bearing on the subclassification of modern Tibetan.

---

<sup>2</sup> Va, a Waic language of Yunnan (personal research), has developed a predominant monosyllabic morpheme structure and a system of three phonemic tones: high, mid, and low, probably under strong areal influences from Tai. Interestingly, its sesquisyllabic Waic neighbors are still non-tonal: Paraok distinguishes two registers (tense versus slack), while the more conservative Lavɿa (personal research) has neither distinctive tone nor register.

<sup>3</sup> Qiang is customarily subdivided into a northern and a southern dialect; the latter is internally diverse with several mutually unintelligible 'subdialects'. The various subdialects of the southern dialect, despite being mostly tonal and more segmentally degenerate than the generally atonal northern dialect, are yet to be rigorously demonstrated to comprise a valid subgroup.

## 2 Evolution of tonal registers

It is generally agreed that the fundamental opposition of tone in (tonal dialects of) Tibetan is one between high and low registers. Old Tibetan did not seem to have tone, even at the phonetic level (i.e. the so-called 'habitual tone'). In this assumed purely atonal stage, still attested in some varieties of Amdo Tibetan, syllables that are not inflectional suffixes are generally high-registered. Tibetan tonal history can be characterised by the genesis of the low register, which has steadily invaded the former territory of the high register (J. Sun 1997:§3.1).<sup>4</sup> Three separate factors that gave rise to the innovative low register will be discussed: onset voicing, rhyme length, and aspiration.

### 2.1 Low register induced by onset voicing

This is crosslinguistically the best-known tonogenetic path; and can be partially explained in terms of articulatory and auditory phonetic mechanisms as due to the tendency for prevocalic voiced obstruents to depress pitch height (Hombert et al. 1979). Yet, as pointed out by Huang (1995:54), the dictum 'voiceless onsets yielded high tone and voiced onsets yielded low tone' does not fit all tonal Tibetan dialects. Rather, it seems that as the original OT onset system eroded through obstruent devoicing and cluster reduction, different modern dialects took the same onset-voicing-based tonogenetic course but responded differently to the complicated interplay between the preradicals and the root initials, resulting in different tone rules.

#### 2.1.1 Lhasa

The tonal developments of the Lhasa dialect are representative of a large number of tonal dialects. In Lhasa, register lowering applied across the board to all (originally) voiced obstruent initials. OT unprefixated sonorant initials also became low-toned, whereas prefixed sonorant initials became high-toned;<sup>5</sup> e.g.

ka <sup>H</sup> <bka> 'order'	ŋa <sup>L</sup> <nga> 'I'
ka <sup>L</sup> <sga> 'saddle'	ŋa <sup>H</sup> <mga> 'drum'
k <sup>ho</sup> L <go> 'to hear; to comprehend'	la <sup>L</sup> <la> 'mountain pass'
ko <sup>L</sup> <mgo> 'head'	la <sup>H</sup> <gla> 'wages'

<sup>4</sup> The low register has also gained increasingly independence from the phonation feature of breathiness concomitant with devoicing of the OT voiced obstruent initials.

<sup>5</sup> Exceptions include OT \*db- (and \*dbr-, \*dby-) and OT \*zl-, which consistently yielded high-toned and low-toned reflexes, respectively. One plausible account (Jiang 1997:40) is that prior to the time Lhasa Tibetan acquired distinctive tones, these anomalous clusters had already turned into very different consonants (OT db- > \*ʔw-ʔ-; OT zl- > \*<sup>n</sup>d-).

2.1.2 *Zaduo*

*Zaduo* (*rDza.rdo*), a distinct dialect of Yushu Prefecture in Qinghai, has richly diversified 'habitual tones' or allophonically predictable pitch patterns depending on initial phonation type (voiceless, voiced, voiceless breathy) as well as rhyme structure (short, long, checked) as shown in the table below (adapted from Huang et al. 1994:115). It suffices for our purposes here to note that, first, plain OT sonorants became low-registered in *Zaduo*, while OT prefixed sonorants became either devoiced (with OT *s-*) or preglottalised (with all other OT preradicals) in redundant high register. This phonetically transparent situation represents a likely precursor to the Lhasa sonorant tone split seen above, and shows that distantly related dialects may independently develop to varying extent the same tonegenetic seeds. Second, unlike in Lhasa, *Zaduo* reflexes of OT voiced obstruent initials did not indiscriminately carry the low register. Instead, the dialect underwent an important split whereby simplex OT voiced obstruents became devoiced, breathy, and low-registered, whereas OT voiced obstruents with preradicals remained voiced and took on *mid* register.

Table 1: *Zaduo* Monosyllabic Citation Pitch Patterns

	LONG RHYME	SHORT RHYME	CHECKED RHYME
HIGH	41 ta: <sup>41</sup> <gtar> 'to bleed [VT]' ʔna: <sup>41</sup> <gna'.ba> 'wild goat' 'mī: <sup>41</sup> <sman> 'medicine'	53 ta <sup>53</sup> <rtā> 'horse' ʔna <sup>53</sup> <rna> 'ear' 'na <sup>53</sup> <snga> 'early'	44 ta <sup>44</sup> <stag> 'tiger' ʔna <sup>44</sup> <mag> 'pus' 'nap <sup>44</sup> <snabs> 'snot'
MID	31 da: <sup>31</sup> <bdar> 'to grind'	32 da <sup>32</sup> <gda'> 'to be present'	23 da <sup>23</sup> <bdag> 'self'
LOW	121 t <sup>fi</sup> a: <sup>121</sup> <da> 'now' na: <sup>121</sup> <na.ba> 'marsh'	21 t <sup>fi</sup> a <sup>12</sup> -ta <sup>53</sup> <da.lta> 'now' na <sup>31</sup> <na> 'to be ill'	12 t <sup>fi</sup> a <sup>12</sup> <dag> 'to be pure' na <sup>12</sup> <nags> 'forest'

2.1.3 *Dege*

In *Dege* (*sDe.dge*), as in *Zaduo*, register lowering did not occur in syllables with voiced complex obstruent initials. The relevant *Dege* data, drawn from Qu (1988:323) and Qu (1991): *passim*, are all in the high register, e.g.

- gu<sup>H</sup> <dgu> 'nine'
- zi<sup>ʔH</sup> <gzig> 'leopard'
- dē:<sup>H</sup> <gdan> 'cushion; bolster'
- gā:<sup>H</sup> <sgam> 'trunk; box'

However, Gesang and Gesang (2002:109) expressly state for this dialect that ‘tone in (modern) syllables with voiced obstruent initials is unstable—one can pronounced them either in the high tone or in the low tone’.<sup>6</sup> On the basis of her instrumental study of Dege tones, Katrin Häslér notes that such voiced obstruents show a tendency toward devoicing, and ‘if a decrease of voicedness occurs the tendency towards low register tone is more pronounced’ (citing Haller 1999:88; fn 13). The production of the low register in Dege, therefore, seems to have fallen into step with obstruent devoicing—both processes occurred by stages, affecting the simplex obstruents first before extending their effects to prefixed obstruents. The tonality state of Dege is true of many other dialects of the Khams area, such as Diqing (*bDe.chen*; Lu 1990, Hongladarom 1996), Changdu (*Chab.mdo*; Liu 1984), and Batang (Gesang 1989, Haller 1999).<sup>7</sup>

#### 2.1.4 Baima

Baima, a divergent form of Tibetan spoken in Pingwu, Jiuzhaigou, and Wenxian counties across the Sichuan-Gansu border (Nishida and Sun 1990, Huang and Zhang 1995),<sup>8</sup> has minimally distinguished tone register on some of its syllables.<sup>9</sup> The history of tone development in this dialect is markedly idiosyncratic.

Among the short syllables (from OT syllables with zero, *-r*,<sup>10</sup> or stop codas), only those carrying prenasalised onsets (from OT stops/affricates with nasal preradicals *m-* and *n<sup>h</sup>*)

<sup>6</sup> Although her recordings show different tone values from those reported by previous authors on Dege, Huang (1995:§2.5) agrees with Gesang and Gesang (2002) on the free variability of tone in such syllables.

<sup>7</sup> The development of the low register seems to have acquired a firmer footing in Batang, where syllables with voiced obstruent can potentially contrast in tone register (see Gesang 1989:§4.2.5).

<sup>8</sup> Huang and Zhang (1995) recognises as many as five tones in the speech of a native Baima from Yaze Village of Baima Township in Pingwu County (Sichuan). This more inclusive tone inventory seems to incorporate pitch patterns that are strictly speaking allotones conditioned by stress (as is the case of the low falling contour occurring on syllables of reduced stress) or rhyme length. My reinterpretation of the Baima tonal history is made possible by the skillful primary analysis of these authors, who have painstakingly sorted out the major tonal correspondences which are, as in all dialects, plagued by exceptions. I have not availed myself of the Baima material in Nishida and Sun 1990 (representing the speech of a different village in Baima Township) since the latter source contains factual and analytical differences from Huang and Zhang (1995) I cannot resolve.

<sup>9</sup> Short syllables with non-prenasalised onsets automatically carry a lower register (phonetically <sup>42</sup>) if the onsets are voiced obstruents (from OT voiced oral obstruent clusters; e.g. *dʒa<sup>L</sup>* <ljags> ‘tongue’; *jæ<sup>L</sup>* <brgya> ‘hundred’), otherwise the associated register is high (phonetically <sup>53</sup>) irrespective of voicing of the proto-onset (e.g. *ʂ<sup>h</sup>a<sup>H</sup>* <sha> ‘meat’; *sa<sup>H</sup>* <za> ‘to eat’; *na<sup>H</sup>* <nags> ‘forest’). Syllables with breathy sonorant onsets (from certain OT sonorant cluster onsets) are redundantly long in the low-register (phonetically <sup>341</sup>; e.g. *lha<sup>L</sup>* <lha> ‘god, Buddha’; *mfiæ<sup>L</sup>* <sman> ‘medicine’).

<sup>10</sup> Huang and Zhang (1995:99) correctly attributes this distribution pattern to the early loss of OT *-r*.

show a tone register distinction. OT voiced and voiceless (aspirated) consonants with nasal preradicals merged into modern voiced prenasalised stops/affricates, causing a compensatory split in tone. Surprisingly, OT voiced and voiceless consonants in this case yielded respectively high and low registers, the reverse of the universal pattern of voicing-based tone bifurcation:

- $\text{ndz}\text{ə}^{\text{H}}$  <'dred> 'to trip up' [sic]<sup>12</sup>  
 $\text{ndzue}^{\text{H}}$  <mdzo> 'crossbreed between a yak and a cow'  
 $\text{ndue}^{\text{L}}$  <'thog> 'to pluck; to pick'  
 $\text{ndzue}^{\text{L}}$  <mtsho> 'lake'

Baima long syllables deriving from OT contracted disyllables display another tone split at variance with the expected voicing-based tonogenetic pattern, producing a low register if the OT disyllable began with voiceless consonants and a high register if the OT disyllable began with voiced consonants. Incidentally, since OT syllables with continuant codas except *-r*, the other main source of Baima long syllables, gave high-registered (phonetically <sup>35 ~ 354 ~ 55</sup>) reflexes regardless of onset voicing, tone register is also phonologically significant in long syllables with voiceless onsets. Consider these examples:

- |                                                                     |                                                             |
|---------------------------------------------------------------------|-------------------------------------------------------------|
| $\text{ka}^{\text{L}}$ <ka.ba> 'pillar'                             | $\text{ny}\text{æ}^{\text{H}}$ <nyi.ma> 'sun'               |
| $\text{sh}\text{æ}^{\text{L}}$ <sol.ba> 'charcoal'                  | $\text{wa}^{\text{H}}$ <ba.ba> 'goiter'                     |
| $\text{sh}\text{e}^{\text{H}}$ <shes> 'to know how'                 | $\text{ndz}\text{æ}^{\text{H}}$ <mtshan> 'night'            |
| $\text{c}\text{ç}\text{o}^{\text{H}}$ <kyang> 'to extend; to reach' | $\text{ndz}\text{ø}^{\text{H}}$ <'dzul> 'to enter (a hole)' |
| $\text{c}\text{ç}\text{o}^{\text{H}}$ <gyang> 'wall'                | $\text{j}\text{æ}^{\text{H}}$ <nyal> 'to go to bed'         |

## 2.2 Low register induced by rhyme length

Syllable duration is seldom included as a contributing factor in the historical development of tone; however, several cases of long rhymes fostering the rise of low register have been noted from various dialects. The association between long (bimoraic) rhymes and low register can be seen in the Lhasa low-tone spreading rule, whereby in disyllable phonological words the low tone in the initial syllable spreads to the following (tonally neutral) *long* syllable regardless of the original tone register of the latter (J. Sun 1997:505–509), and the Sherpa rule of tone neutralisation whereby non-initial long syllables become predictably low-registered (see further section §4.1). Rhyme length

<sup>11</sup> Written with the *achung* <'> in the Tibetan script.

<sup>12</sup> The verb means 'to slip; to slide' in standard written Tibetan.

either serves as a phonological environment for a subphonemic low register as in Zhangla and Qiuji, or underlies primary tonal splits as the case seems to be in Zhuoni.<sup>13</sup>

### 2.2.1 Zhangla<sup>14</sup>

The Tibetan spoken in Zhangla (*lCang.la*) District of Songpan County is a form of the Amdo dialect. In Zhangla, the low register correlates directly with rhyme length at the phonetic level, such that monosyllables containing long rhymes tend to be realised in a redundant low (phonetically low rising) tone, especially when the onset is voiced or voiceless aspirated:

ka: <sup>L</sup> <ka.ba> 'pillar'	sɛ: <sup>L</sup> <sprin> 'cloud'
ŋtʰõ: <sup>L</sup> <mthõng> 'to see'	ɣĩ: <sup>L</sup> <zhing> 'field'
dza: <sup>L</sup> <zla.ba> 'moon'	lɔ: <sup>L</sup> <lang> 'to stand'
	lʰõ: <sup>L</sup> <lhung> 'to fall'

### 2.2.2 Qiuji<sup>15</sup>

Qiuji (*Chos.rje*) is a little-explored dialect of eastern Ruo'ergai County and neighboring areas of Jiuzhaigou County. Among the unusual phonological traits of this dialect, OT voiced unprefixated obstruents were kept as such (e.g. *go* <go> 'to hear'), OT voiceless unprefixated obstruents became *voiced* (e.g. *ga:* <ka.ba> 'pillar'), whereas OT voiced obstruents with preradicals (nasal or non-nasal) became distinctively breathy (e.g. *kʰo* <sgo> 'door'; *ŋʰo* <sngo> 'blue/green').<sup>16</sup> There is no phonemic tone in Qiuji,<sup>17</sup> but

<sup>13</sup> In Zhouchu (*'Brug.chu*), long syllables with originally voiceless onsets also take the low register (e.g. *ka:<sup>21</sup>* <ka.ba> 'pillar'; *sʰa:<sup>22</sup>* <sems> 'mind'), at least in the limited data given in Huang (1995:49-53).

<sup>14</sup> Personal research. My Zhangla consultant comes from Mayi (*ntʰɛwɛ*) Village in Zhangla District of Songpan County.

<sup>15</sup> The Qiuji Tibetan data were collected in my recent fieldwork with the help of a native from Mazang Village in Qiuji Township (J. Sun, in preparation). Qiuji Tibetan is strikingly different from the surrounding dialects: Zhouqu (*'Brug.chu*) to the east, Amdo to the west, and Zhongu (*Zho.ngu.khog*) and Baima to the south. It is said to be mutually intelligible with the markedly different Thewo dialect of Ruo'ergai and Diebu counties, probably due to prolonged contact.

<sup>16</sup> As the OT voiced stop/affricates with oral preradicals became devoiced and distinctively low-toned in Lende Tibetan, they also became concomitantly breathy (e.g. *tʰo:<sup>L</sup>* <rdo> 'stone'; Huber (forthcoming). Qiuji has carried the development of breathy phonation much further than Lende, however. Incidentally, Qiuji prenasalised consonants and /r/ (< OT r-) are redundantly breathy (e.g. *ngɔ:<sup>L</sup>* <mgo> 'head'; *rʰi:<sup>L</sup>* <ru> 'to rot').

<sup>17</sup> Qiuji is the same dialect as what is called 'Ruo'ergai' in Huang (1995), for which as many as four tonemes are described. Our different tonal analyses stem from the different segmental inventories we

phonetic-level 'habitual tone' is more deeply entrenched in this dialect. Monosyllables (even with voiced or breathy onsets) are generally spoken on a high register, except that syllables containing long rhymes and aspirated<sup>18</sup> or breathy initials are *invariably* accompanied by low register (phonetically low-rising); compare:

k <sup>h</sup> ɔ: <sup>L</sup> <khang.ba> 'house'	mɔ: <sup>H</sup> <mang> 'to be many; much'
ʃe: <sup>L</sup> <sems> 'mind; soul'	gɔ: <sup>H</sup> <gangs> 'snow-capped mountain'
k <sup>h</sup> ɔ: <sup>L</sup> <sgam> 'trunk'	k <sup>h</sup> o <sup>H</sup> <kho> 's/he'
	k <sup>h</sup> o <sup>H</sup> <dgos> 'to want; to be necessary'

### 2.2.3 Zhuoni

Zhuoni (*Co.ne*) Tibetan is spoken at Lintan County of Gannan Prefecture in Gansu Province (Qu 1962). Zhuoni appears to distinguish only two tonemes, high and low.<sup>19</sup> Modern short syllables, which originated from OT open and checked syllables, underwent the usual register split caused by proto-voicing:

k <sup>h</sup> ə <sup>H</sup> <kha> 'mouth'	gu <sup>L</sup> <dgu> 'nine'
tu <sup>H</sup> <gtub> 'to cut'	tu <sup>L</sup> <dug> 'poison'
lo <sup>H</sup> <glo> 'girth'	lo <sup>L</sup> <lo> 'age'

Initial voicing is irrelevant to tonal development in long syllables, which come from OT monosyllables with sonorant codas and contracted disyllables.<sup>20</sup> Long syllables are generally low-toned, except that those deriving from OT syllables containing *non-nasal* preradicals are high-toned:

tʃə: <sup>L</sup> <tshil> 'grease'	ka: <sup>H</sup> <skal> 'share [N]'
ka: <sup>L</sup> <gang> 'where'	ba: <sup>H</sup> <sbang> 'to soak'
nam: <sup>L</sup> <nyan> 'to listen'	to: <sup>H</sup> <stong> 'to be empty'
tʃo: <sup>L</sup> <mthong> 'to see'	du: <sup>H</sup> <sdong.po> 'tree'
do: <sup>L</sup> <mdung> 'spear'	

Tone splits are therefore governed by entirely different factors in Zhuoni short and long syllables. The familiar tone splitting caused by onset voicing only affected short syllables, and most long syllables, irrespective of original onset voicing, took on the low register.

---

posit. I consider aspiration in Qiuji voiceless fricatives and breathy phonation as distinctive, whereas Huang does not.

<sup>18</sup> Including voiceless aspirated spirants.

<sup>19</sup> Zhuoni is a Tibetan dialect spoken at Lintan County of Gannan Prefecture in Gansu Province (Qu 1962). The four tones posited by Qu, short high (<sup>53</sup>), short low (<sup>21</sup>), long high (<sup>55</sup>), and long low (<sup>22</sup>), can be reasonably reduced to just high (<sup>H</sup>) and low (<sup>L</sup>) if one relegates rhyme length to the segmental or moraic tier.

<sup>20</sup> There are some exceptional cases of OT liquid codas *-r* and *-l* yielding short instead of long Zhuoni vowels.

Zhuoni thus provides another example in modern Tibetan of rhyme length being critically involved in the genesis and propagation of the distinctive low register.<sup>21</sup>

### 2.3 Low register induced by onset aspiration

Aspirated onset consonants have been observed to lead to tone lowering in various Chinese dialects and Southeast Asian languages (see Ho 1990). Despite its unclear underlying phonetic mechanisms and the existence of completely contrary developments,<sup>22</sup> the phenomenon is attested in at least two Tibetan dialects we have worked on: Qiuji and Tiebu.

As shown in §2.2.2, a Qiuji monosyllable takes on a subphonemic low register if it satisfies two conditions: (i) that the rhyme is long, and (ii) that the onset consonant is either breathy or aspirated.

The role of aspiration as a tone depressor works somewhat differently in Tiebu Tibetan.<sup>23</sup> In this dialect, modern syllables containing sonorant or voiceless unaspirated obstruent onsets have acquired a distinction in (high versus low) tone, in all other syllables tone register is predictable from onset consonant types and syllable structure. As a striking phonological trait of this dialect, aspirated onset consonants (i.e. voiceless aspirated stops, affricates, and spirants) are associated with the high register in checked syllables, but with the low register in non-checked ones, as in:

ph <sup>a</sup> ʔ <sup>H</sup> <phag> 'pig'	sh <sup>a</sup> L <sa> 'earth'
ts <sup>h</sup> aʔ <sup>H</sup> <khrag> 'blood'	x <sup>h</sup> e <sup>r</sup> L <shel> 'glass'

## 3 Evolution of tonal contours

On rare occasions, one finds Tibetan dialects showing further evolved tone systems with a minimally distinctive falling contour superimposed on the basic high-low register contrast.

<sup>21</sup> Qu (1988:323) suggests rather complicated tone shifts to account for the Zhuoni scenario.

<sup>22</sup> In the Nakhonsithammarat (Haas 1958) and Songkhla (Henderson 1959) dialects of Thai and in the Waic (Mon-Khmer) language Va (personal research), aspirated consonants conditioned high rather than low tone register.

<sup>23</sup> Tiebu (*The.bo*), like Qiuji, is an under-researched form of Tibetan spoken by the agricultural Tibetans of Chong'er, Re'er, and Donglie Townships of Tiebu District in Ruo'ergai County, as well as of the abutting Diebu District on the Gansu side of the Sichuan-Gansu border. I worked briefly with my first two Tiebu consultants in 1994. The Tiebu data cited herein, from another male speaker Mr. Tshe'bum, were recorded by my student Ms. Lin Youjing under my supervision during fieldwork in Sichuan in the fall of 2001.



### 3.1 *Falling contour induced by the glottal stop coda*

The widely attested tendency for the syllable-final glottal stop ? to lead historically to rising tone has been experimentally verified and accounted for by Hombert et al. (1979, §2.2.2), who showed that the postvocalic glottal stop produces a significant and perceptible rise in F<sub>0</sub> in the preceding vowel. As exemplified by the following varieties of Central Tibetan, the contrary effect seems to be the norm in Tibetan, attributable probably to creaky voice as a variant realisation of the glottal stop (according to Peter Ladefoged, cited in Hombert et al. 1979:51). Indeed, a variant realisation of the glottal stop coda in Lhasa Tibetan is a fall in pitch accompanied by some degree of glottal stricture (Chang and Chang 1978:xix–xx).

#### 3.1.1 *Lhasa*

In Lhasa, we find falling allotones in syllables taking the glottal stop coda, as in (Hu et al. 1982):

- ka<sup>L</sup> [ka<sup>12</sup>] <sga> 'saddle'
- ka<sup>?L</sup> [ka<sup>?132</sup>] <'gag> 'to be clogged'
- kam<sup>H</sup> [kam<sup>55</sup>] <skam> 'to be dry'
- kam<sup>?H</sup> [kam<sup>?52</sup>] <bskams> 'make dry [PF]'

#### 3.1.2 *Rikeze*

In some varieties of Central Tibetan the glottal stop completely elided, leaving in its wake a distinctive falling tone.<sup>24</sup> This is best exemplified by Rikeze (*gZhis.ka.rtse*; Qu 1981a:187, Qu 1988:324–326, Haller 1999) where the lexically random drop of the glottal coda triggered a secondary tone split, transforming the original two-tone system into a four-tone system:

- ke<sup>H</sup> [ke<sup>53</sup>] <ske> 'neck'
- tṣhà<sup>H</sup> [tṣha<sup>51</sup>] <khrag> 'blood'
- ro<sup>L</sup> [ro<sup>12</sup>] <ro> 'corpse'
- rè<sup>L</sup> [rè<sup>131</sup>] <ras> 'cloth'

<sup>24</sup> Marked below with the grave accent. Since the high register tone may also fall slightly at the phonetic level, especially when the syllable is short, the innovative falling tone is realised by a distinctively steep fall.

### 3.1.3 Zhibo

In Zhibo (Qu 1988:324), the split generating falling tone took effect only in the high register, resulting in two distinctive high-register tones. In the low register, the loss of the glottal stop coda is compensated for by vowel length only, e.g.

so<sup>H</sup> [so<sup>53</sup>] <so> 'tooth'  
 sò<sup>H</sup> [so<sup>51</sup>] <srog> 'life'  
 ma:<sup>L</sup> [ma:<sup>113</sup>] <mar> 'butter'  
 si:<sup>L</sup> [si:<sup>113</sup>] <gzig> 'leopard'

### 3.1.4 Langkazi

The tone split in question was narrowly confined to one phonological environment in Langkazi (*sNa.dkar.rtse*; Qu 1988:324–325), namely low-registered syllables closed by a nasal coda, where the OT secondary coda -s conditioned falling pitch, presumably compensating for the loss of an earlier glottal-stop reflex of OT -s. Notably, the falling contour in this dialect is allotonically predictable as it is always correlated with lack of vowel length:

ka:m<sup>L</sup> [kɑm<sup>113</sup>] <sgam> 'trunk'  
 kɑm<sup>L</sup> [kɑm<sup>131</sup>] <bgams> 'eat powdery food [PF]'

## 3.2 Falling contour induced by sonorant codas and syllable coalescence

In some varieties of Dzongkha (rDzong.kha) Tibetan, a basic high-versus-low register distinction is operational in all syllable types, but some syllable types (long open syllables and short syllables closed by bilabial codas) distinguish an additional level-versus-falling contour (Mazaudon and Michailovsky 1989).<sup>25</sup> The origins of the Dzongkha innovative contour tone are now clear, thanks to the skilled historical comparison provided by these authors. The loss of the OT liquid codas -r, and -l produced distinctive falling contour,<sup>26</sup> as shown by:

pa:<sup>H</sup> <spags> 'cut of meat'      bja:<sup>L</sup> <'bras> 'paddy (in the field)'  
 pɑ:<sup>H</sup> <dpar> 'picture'              bjɑ:<sup>L</sup> <dbyar> 'summer'

<sup>25</sup> In a valuable recent study (Watters 1996), a radically different phonological analysis of Dzongkha prosody is offered. Watters recognises more consonant types including preglottalised sonorants and prenasalised stops, and finds the variety of Dzongkha he investigates to be only incipiently tonal. Unfortunately, Watters does not specify exactly which Dzongkha dialect (down to the level of village) his data represent.

<sup>26</sup> The OT velar nasal coda was also reflected by (non-distinctive) falling tone on long nasalised vowels in Dzongkha.

Likewise, contracted disyllables containing the suffixes *-pa/-po*, *-ba/-bo*, *-ma/-mo* became Dzongkha monosyllables in falling tone:

sum <sup>H</sup> <gsum> 'three'	lam <sup>L</sup> <lam> 'road'
sùm <sup>H</sup> <srung.ba> 'locket'	bjàm <sup>L</sup> <sbrang.ma> 'bee'
çop <sup>H</sup> <shob> 'lie'	
çòp <sup>H</sup> <shog.pa> 'wings'	

#### 4 Tone neutralisation

The Tibetan tone system is a kind of *template word-tone* characterised by initial-dominance (J. Sun 1997:§4) whereby contrastive tone register is borne solely by the stressed initial syllable in a phonological word while non-initial syllables are redundantly high-registered.<sup>27</sup> Discussed in the following are two exceptions to the foregoing generalisation.

##### 4.1 Sherpa

In Sherpa (Shar.pa), the predominant register in tonally neutral non-initial syllables appears to be low rather than high (Tan 1987). Disyllabic words show two patterns of neutralisation depending on the morphological makeup. The minor pattern, motivated probably by trochaic stress, is found with words consisting of a main root not checked by a glottal-stop coda plus a suffix. The main root keeps its monosyllabic citation tone shape,<sup>28</sup> whereas the suffix is spoken invariably in low register (phonetically low falling<sup>31</sup>), as in:

a <sup>H</sup> -po <sup>L</sup> <gla.bo> 'day-laborer'	phe <sup>L</sup> -pa <sup>L</sup> <bod.pa> 'Tibetan'
ŋa <sup>H</sup> -mu <sup>L</sup> <mngar.mo> 'to be sweet'	tha <sup>L</sup> -mo <sup>L</sup> <dwangs.mo> 'to be sunny'

In the majority pattern occurring elsewhere, non-initial syllables show a quantity-related register split—short rhymes are high-registered (phonetically<sup>33</sup>) and long rhymes are low-registered (phonetically<sup>331</sup>); e.g.

phu <sup>L</sup> + tshap <sup>H</sup> <bu.tshab>	→	phu <sup>L</sup> -tshap <sup>H</sup> 'adopted son'
ça <sup>L</sup> + ne <sup>L</sup> <zha.ne>	→	ça <sup>L</sup> -ne <sup>H</sup> 'lead (metal)'
me <sup>L</sup> + khe:m <sup>H</sup> <me.khyem>	→	me <sup>L</sup> -khe:m <sup>L</sup> 'fire-shovel'
nup <sup>L</sup> + ŋim <sup>L</sup> <nub.nyin>	→	nup <sup>L</sup> -ŋim <sup>L</sup> 'day and night'

<sup>27</sup> Barring unstressed clitic syllables and occasional local assimilation to preceding low register (see J. Sun 1997:§2.2, 3.2).

<sup>28</sup> Except for a subphonemic detail: the pitch shape in high-registered long rhymes changes from (citation) high-falling to high level.

#### 4.2 Baima

As shown above, Baima is embryonically tonal with a register opposition restricted to certain syllable types, namely short syllables with prenasalised onsets and long syllables with voiceless onsets. Tonal distinction seems drastically reduced in Baima polysyllabic phonological words, as elsewhere in modern Tibetan. It is a token of the high aberrancy of this dialect that, unlike any other Tibetan dialect known to us, the target of Baima tonal neutralisation is the *initial* syllable. Thus, most disyllabic words cited in Huang and Zhang (1995) show an indistinct low register (phonetically a low fall <sup>21</sup>; unmarked herein) in the initial syllable:<sup>29</sup>

$jjæ^L [jjæ^{42}] + \text{ʂa}^H [\text{ʂa}^{53}] <\text{rgya.zhwa}> \rightarrow jjæ\text{-}\text{ʂa}^H [jjæ^{21}\text{-}\text{ʂa}^{53}]$  'Chinese-style hat'  
 $jy\text{ə}^H [jy\text{ə}^{53}] + \text{tsh}\text{ɔ}^H [\text{tsh}\text{ɔ}^{35}] <\text{lug.tshang}> \rightarrow jy\text{ə}\text{-}\text{tsh}\text{ɔ}^H [jy\text{ə}^{21}\text{-}\text{tsh}\text{ɔ}^{35}]$  'sheep-pen'

In fact, this low-high register pattern characterises most disyllabic words in Baima (Huang and Zhang 1995:84). Unfortunately, the data provided in this source are insufficient for determining the degree to which the Baima non-initial syllables are also subject to tonal neutralisation.

### 5 Summary and conclusions

Tibetan tonal typology is a subject attracting increasing scholarly attention (Mazaudon 1977, Qu 1981a, Qu 1981b, Qu 1988, Tan 1984, Yip 1993, Huang 1995, J. Sun 1997, Huber to appear). The present study makes a small contribution to this growing literature by demonstrating that (i) the same phonetic motivations underlying the evolution of the high-versus-low register contrast in Tibetan produced different developments in different dialects (§2.1); (ii) in addition to onset voicing and laryngeal codas, rhyme length and open glottal states (breathy voice and aspiration) also turn out to be important but hitherto underrated pathways leading to distinctive low register (§2.2); (iii) the emergence of the secondary level-versus-falling contour contrast may also be heterogeneously actuated in different dialects (§3), and (iv) though all known tonal dialects have template (i.e. non-spreading) word-tone systems, tone neutralisation strategies in polysyllabic words are not uniform across the tonal dialects.

The diverse diachronic tone rules inspected herein suggest that as spoken Tibetan was subject ubiquitously to the drift toward reduced consonantism and syllable structure, a

<sup>29</sup> There are quite a few apparent exceptions, some of which seem explainable; for example, O-V compounds are generally not treated as unitary phonological words in Baima, e.g.  $n\text{æ}^H g\text{ɔ}^L <\text{me.sgron}>$  'to light a fire'; this is also true of Amdo Ruo'ergai Xiaman (*mDzod.dge Byams.me*; J. Sun 1986), where phonological words are defined among other things by productive vocalic dissimilation. Furthermore, Baima monosyllables that derive from original disyllables did not seem to undergo tonal reduction either, e.g.  $na^H\text{-}d\text{ɔ}^H <na; <\text{ma.ba}> + d\text{ɔ}^H <\text{brdol}>$  'ear-hole'.

small number of dialects managed to bear the strain of segmental attrition without exploiting pitch modulations even at the allophonic level,<sup>30</sup> while most other dialects responded by setting their course for tonogenesis. Different dialects may well have independently developed to varying extents the same tonogenetic potentials or explored divergent tonogenetic paths, producing the observed scale of tonality types ranging from transparent allophonic pitch patterns (e.g. Zaduo, Qiuji) to stable and synchronically opaque phonemic tones comprising both distinctive registers and contours (Rikeze).

The moral of this study for the classification of the Tibetan dialects is clear. Rigorous methodology demands that only shared innovations unlikely to be independent developments are diagnostic of a period of common history, hence admissible as reliable subgrouping criteria. Given the strong possibility for the seeds of tone to sprout at different times and in different Tibetan-speaking areas, the mere presence of analogous tone systems in two forms of Tibetan tells us little about the degree of their genetic affinity. This is of course not to deny the value of *diachronic tone rules*. Quite the contrary, I believe that they are a largely untapped resource of phonological innovations that should properly figure in any comprehensive classification of modern Tibetan.<sup>31</sup> The important task at hand is to rank the observed tone rules in terms of their usability for the purpose of dialect subclassification. Phonetically well-motivated, recurrent global similarities of tone are probably due to parallel phonological development and are to be sifted out. On the other hand, idiosyncratic rules of tone split and neutralisation, some of which are presented in this paper, should be sought out in earnest for their potential value in the armory of Tibetan dialectologists.

In conclusion, the sagacity of the quote at the beginning of this paper holds true also for the classification of Tibetan dialects, and I only hope Prof. Matisoff will smile indulgently at my belaboring a point he already made so plain almost three decades ago.

## References

- Bielmeier, Roland, 1982, Problems of Tibetan dialectology and language history with special reference to the *sKyid.gron* dialect. *Zentralasiatische Studien des Seminars für Sprach- und Kulturwissenschaft Zentralasiens der Universität Bonn* 16:405–425.
- Bradshaw, Joel, 1979, Obstruent harmony and tonogenesis in Jabêm. *Lingua* 49:189–205.
- Chang, Kun and Betty Shefts Chang, 1978, *Spoken Tibetan Texts* vol.1 (Institute of History and Philology, Academia Sinica Special Publication 74). Taipei: Academia Sinica.

<sup>30</sup> An extreme case in point being Zhongu (J. Sun to appear), which remains completely atonal despite the fact that only as few as nine monophthongal rhymes remain in its inventory of contrastive rhymes.

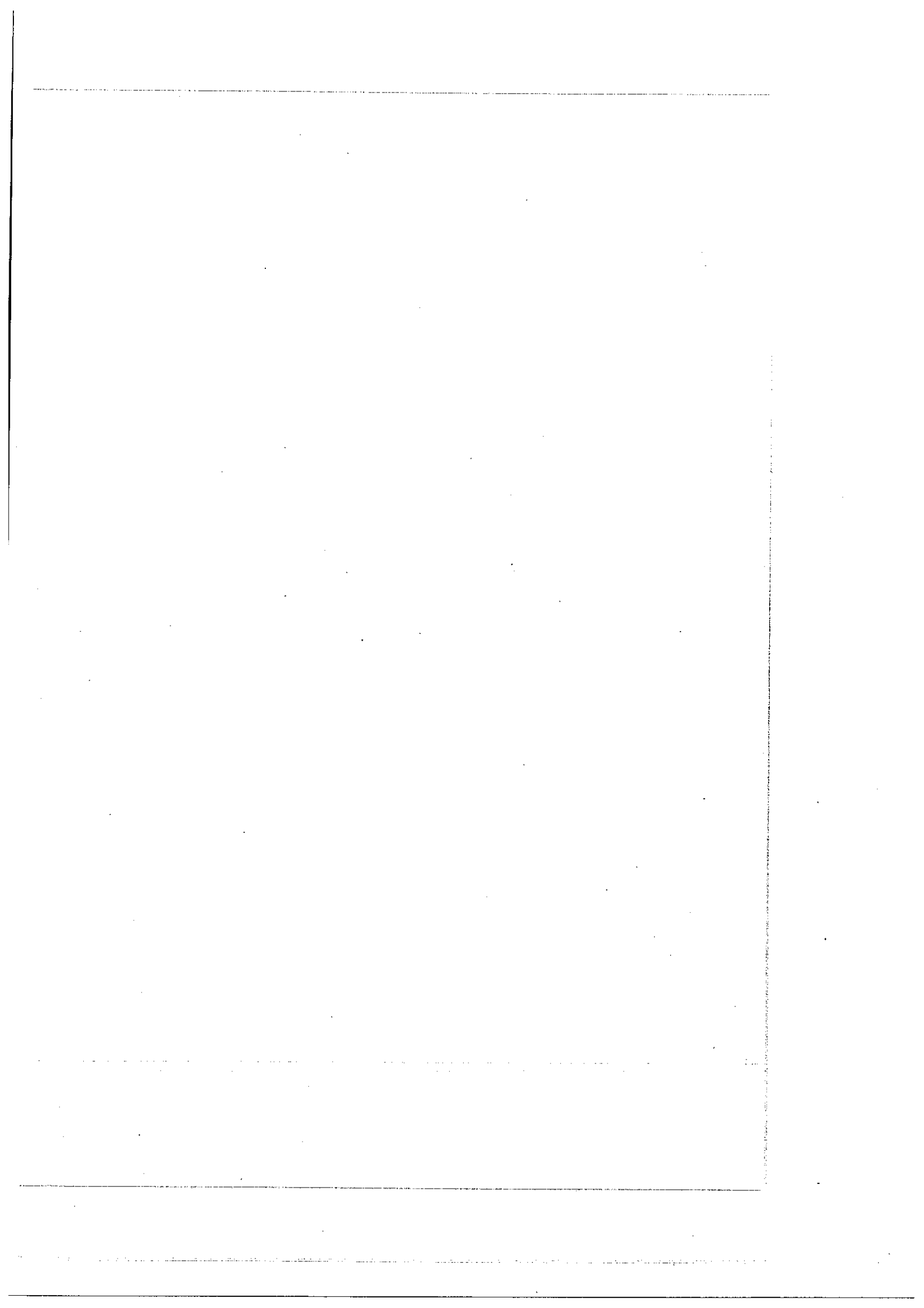
<sup>31</sup> Phonological innovations have been utilised in several recent papers bearing on Tibetan subclassification (e.g. Bielmeier 1982, Nishi 1986, Zhang 1993). But these deal mainly with segmental sound changes.

- Gesang Jumian, 1989, Phonological analysis of Bathang Tibetan. *Acta Orientalia Academiae Scientiarum Hungaricae* 43/2-3:331-358.
- Gesang Jumian and Gesang Yangjing, 2002, *Zangyu fangyan gailun* [Synopsis of Tibetan Dialects]. Beijing: Nationalities Press.
- Haas, Mary R., 1958, The tones of four Tai dialects. *Bulletin of the Institute of History and Philology* 29.2:817-826.
- Haller, Felix, 1999, A brief comparison of register tone in Central Tibetan and Kham Tibetan. *Linguistics of the Tibeto-Burman Area* 22.2:77-97.
- Henderson, Eugénie J.A., 1959, The tone of the Tai dialect of Songkhla. *Bulletin of the Institute of History and Philology* 30.1:233-235.
- Ho, Dah-an, 1990, Songqi fendiao ji xiangguan wenti (Tone splitting conditioned by aspiration and related issues). *Bulletin of the Institute of History and Philology* 60.3:765-778.
- Hombert, Jean-Marie, William G. Ewan and John J. Ohala, 1979, Phonetic explanations for the development of tones. *Language* 55/1:37-58.
- Hongladarom, Krisadawan, 1996, Rgyalthag Tibetan of Yunnan: A preliminary report. *Linguistics of the Tibeto-Burman Area* 19/2:69-92.
- Hu Tan, 1991, Zangyu [The Tibetan Language]. In Ma Xueliang et al. eds *Han-Zangyu gailun* [A general introduction to Sino-Tibetan languages], 109-207. Beijing: Beijing University Press.
- Hu Tan, Qu Aitang and Lin Lianhe, 1982, Zangyu Lasahua shengdiao shiyan [Experimental studies on Lhasa Tibetan tone]. *Yuyan Yanjiu* 2:18-38.
- Hua Kan and Duo Zangta, 1997, Zangyu Songpanhua de yinxi he yuyin de lishi yanbian [Sound system of Songpan Tibetan and its historical development]. *Zhongguo Zangxue* 29.2:134-144.
- Huang Bufan, 1991, Qiangyuzhi (Qiangic branch). In Ma Xueliang et al. eds *Han-Zangyu gailun* [A general introduction to Sino-Tibetan languages], 208-369. Beijing: Beijing University Press.
- 1995, Conditions for tonogenesis and tone split in Tibetan dialects. *Linguistics of the Tibeto-Burman Area* 18/1:43-62.
- Huang Bufan, Suonan Jiangcai and Zhang Minghui, 1994, Yushu Zangyu de yuyin tedian he lishi yanbian guiyu [Characteristics in Yushu Tibetan phonology and its rules of historical change]. *Zhongguo Zangxue* 26/2:111-134.
- Huang Bufan and Zhang Minghui, 1995, Baimahua zhishu wenti yanjiu [On the linguistic position of Baima]. *Zhongguo Zangxue* 2:79-118.

- Huber, Brigitte, to appear, Tone in Kyirong Tibetan: a conservative phonological development. In Tej Ratna Kansakar and Mark Turin eds *Themes in Himalayan languages and linguistics* (Proceedings of the 5th Himalayan Languages symposium, Kathmandu 1990).
- Jiang Di, 1997, Zangyu db- yinlei de yanhua guocheng ji shijian cengci [The evolutionary paths of Tibetan db- and their temporal ordering]. *Minzu Yuwen* 5:31-40.
- Liu Feng-hsi, 1984, Tones and tone sandhi in eastern Tibetan. Paper presented at the 17th International Conference on Sino-Tibetan Languages and Linguistics. Eugene, University of Oregon.
- Lu Shaozun, 1990, Zhangyu Zhongdianhua de yuyin tedian [Phonological characteristics of rGyal.thang Tibetan]. *Yuyan Yanjiu* 2:147-159.
- Matisoff, James A., 1973, Tonogenesis in Southeast Asia. In Larry M. Hyman ed. *Consonant types and tone* (Southern California Occasional Papers in Linguistics No. 1), 73-95. Los Angeles: Linguistics Program, UCLA.
- Mazaudon, Martine, 1977, Tibeto-Burman tonogenetics. *Linguistics of the Tibeto-Burman Area* 3/2:1-123.
- Mazaudon, Martine and Boyd Michailovsky, 1989, Lost syllables and tonal contour in Dzongkha. In David Bradley, Eugénie J.A. Henderson and Martine Mazaudon eds *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*, 115-136. Canberra: Pacific Linguistics.
- Nishi Yoshio, 1986, A classification of Tibetan dialects [in Japanese]. *National Museum of Ethnology Research Reports* 11.4:837-900.
- Nishida Tatsuo and Sun Hongkai, 1990, *A Study of the Baima-Chinese Vocabulary Baima Yiyu: the Structure and Lineage of the Baima Language* [in Japanese and Chinese]. Kyoto: Shokado.
- Qu Aitang, 1962, Zhuoni Zangyu de shengdiao yu shengyunmu de guanxi [The relationship between tone, onsets, and rhymes in Zhuoni Tibetan]. *Zhongguo Yuwen* 7:334-339.
- 1981a, Zangyu de shengdiao ji qi fazhan [Tibetan tone and its historical development]. *Yuyan Yanjiu* 1:177-194.
- 1981b, Zangyu de biandiao [Tone sandhi in Tibetan]. *Minzu Yuwen* 4:20-7.
- 1988, Zangyu gudiao zhi gouni [A reconstruction of ancient tone values in Tibetan]. *Zhongguo Yuyan Xuebao* 3:317-338.
- 1991, *Zangyu yunmu yanjiu* [A study on Tibetan rhymes]. Xining: Qinghai Minzu Chubanshe.
- Qu Aitang and Jin Xiaojing, 1981, Zhangyu fangyan de yanjiu fangfa [Methods of study the Tibetan dialects]. *Xinan Minyuan Xuebao* 3:76-83.

- Qu Aitang and Tan Kerang, 1983, *Ali Zangyu [mNga.ris Tibetan]*. Beijing: Shehui Kexue Chubanshe.
- Róna-Tas, András, 1966, *Tibeto-Mongolica: the Tibetan loanwords of Monguor and the development of the archaic Tibetan dialects*. The Hague: Mouton.
- Sun, Jackson T.S., 1986, *Aspects of the phonology of Amdo Tibetan: Ndzorge Shæme Xæra dialect (Monumental Serindica No. 16)*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- 1997, Typology of tone in Tibetan. *Chinese languages and linguistics IV: Typological studies of languages in China (Symposium series of the Institute of History and Philology, Academia Sinica; Number 2)*, 485–521. Taipei: Academia Sinica.
- To appear (in 2003), Phonological profile of Zhongu: a new Tibetan dialect of northern Sichuan. *Language and Linguistics 4/3* (Taipei: Institute of Linguistics, Academia Sinica).
- In preparation, Phonological profile of Qiuji Tibetan.
- Tan Kerang, 1984, Zangyu shuangyinjieci biandiao leixing [Tone sandhi patterns of disyllabic words in Tibetan]. In *Proceedings of the Lhasa seminar on Tibetology*, 626–39. Lhasa: Tibet People's Press.
- 1987, Xia'erba zangyu de shengdiao xitong [The tone system of Sherpa Tibetan]. *Minzu Yuwen 2*:22–28.
- Thurgood, Graham, 1996, Language contact and the directionality of internal drift: the development of tones and registers in Chamic. *Language 72/1*:1–31.
- Watters, Stephen A., 1996, A preliminary study of prosody in Dzongkha. Ph.D. Dissertation, University of Texas at Arlington.
- Yip, Moira J., 1993, Tonal register in East Asian languages. In Harry van der Hulst and Keith Snider eds *The phonology of Tone: the representation of tonal register*, 245–68. Berlin: Mouton de Gruyter.
- Zhang Jichuan, 1993, Zangyu fangyan fenlei guanjian [Opinions on the classification of Tibetan dialects]. In Dai Qingxia et al. eds *Minzu yuwen lunwenji [Papers on minority languages]*, 297–309. Beijing: Central Institute of Nationalities Press.





## 4 *Some case studies on linguistic variation and their implications*

---

CAROL GENETTI

### **Introduction**

One of the joys of reading the work of James A. Matisoff is his own clear delight in the messiness and infinite variation in language. Rather than trying to whittle a language down in order to achieve a single monolithic description, he fleshes it out, deftly exploring the subtlety and variation, and revealing the power of creativity of native speakers. In my own experience of conducting fieldwork on languages of the Himalayas, I have found myself repeatedly confronted by linguistic variation of a number of types and with a number of motivations. In my earlier days, I was eager to attribute such variation exclusively to differences of dialect (geographically or socially defined) or register. But eventually I had to admit that I was confronting variation that was neither, but was variation at the level of the idiolect. This confounded my ability to create the monolithic description that I thought grammars were supposed to be, and eventually, under the influence of Matisoff's work, I learned to work with the variation, indeed to give it a central role in my understanding of language.

For this volume, which honors Jim Matisoff and his tremendous accomplishments, I have decided to bring together several case studies of variation taken from my own work. The first study is on verb agreement in Nepāli, and presents a classic case of variation based on register. It shows that variation can persist across generations, and that the register-based system is the result of competing pressures on the system. The second study examines differences in syntactic constructions between two speakers of Kāthmāndu Newar who otherwise show very little difference in their speech patterns. While it is certainly possible that one could attribute these differences to dialect, probably socially defined, it is also possible that the differences are idiolectal. The third study examines

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 53–63. Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

idiolectal variation in more detail, comparing differences in the speech of first cousins raised as young children in the same household. Idiolectal differences in their phonological systems occur in environments where the differences carry no functional load. I also discuss differences in their use of two paradigmatically related suffixes, and suggest that the variation is the result of a decayed, relatively unmotivated distribution.

Throughout the paper, I will be referring to an expectation that is widely held in linguistics: the expectation of inter-speaker consistency. It is generally expected that when two or more speakers speak the same dialect, their phonological and grammatical systems will be the same. Thus one assumes that working with one speaker as an informant will produce the same result as working with another speaker, although there may be differences between them in their skill as informants, in the size of their vocabulary, etc. When variation is encountered, one can attribute it to dialect or register, and still happily feel that inter-speaker consistency is maintained. Part of the motivation behind the expectation of inter-speaker consistency comes from our tradition as grammarians to come up with a single description of each component of a language, and to present these descriptions as invariable, and as representing 'the way' a language works. It also derives in part from our tradition of conducting elicitation with a single, primary consultant, which naturally precludes any evidence of inter-speaker variation. And of course it is also due to the fact that there is ample evidence that social factors do indeed correlate with variation.

While the first study discussed below comfortably allows us to maintain the expectation of inter-speaker consistency, as variation is attributed to register differences, the second study raises doubts, and the third study, on idiolectal variation, runs counter to the expectation. One of the goals of this paper, then, is to make this expectation explicit, which is often implicit and assumed, and to demonstrate that it does not necessarily hold true. The second goal is to illustrate the richness of analysis which comes from the exploration of variation in detail. The historical sources and synchronic implications of linguistic variation, even at the level of the idiolect, are themselves a fascinating and enriching field of study.

## **1 Verb agreement in Nepāli: register-based variation persisting across generations**

The first case study concerns variation in verb agreement in Nepāli, an Indo-Aryan language which is the national language of Nepal. A full discussion can be found in Genetti (1999); here I will only summarise that study and briefly discuss the results.

Nepāli is a language with many varieties and a rich literature. There is a strong prescriptive tradition, and written Nepāli, which is taught to children in schools, is taken to be the 'correct' form of the language. In this prescriptive variety, the verb agrees with the subject in person, number, gender, and honorific status, all of which combine to create a

complex verbal paradigm (Genetti 1999:543–544). However, in casual spoken Nepāli, native speakers will frequently simplify the agreement, in particular the gender and number inflection. This variation is quite noticeable. In Genetti (1999), I presented clear evidence that amount and type of agreement is dependent on genre. In written Nepali there are very high percentages of agreement in gender and number (97.3 per cent, combined), in conversational Nepāli, percentages of agreement are quite low (9.3 per cent), and in spoken Nepāli narratives, the verb agreed in gender and number about half the time (51.9 per cent). Interestingly, the most inter-speaker variation found within genre was attested in the narrative data, the register intermediate in its formality. Here the percentages of different speakers varied significantly, but this variation could be attributed to different interpretations of formality that different speakers brought to the task of telling a story to a linguist with a tape recorder. Thus, in this case one can still maintain the expectation of inter-speaker consistency. Variation is attributed to register, and there is the expectation that if the speakers were consistent in their interpretation of level of formality, we would also find greater consistencies in their percentages of verb agreement.

One cannot help but wonder about the motivations for this register-based variation in the grammatical system. Where does it come from? And, if Nepāli speakers use verb agreement so sparingly in conversation, could it be that it is being lost, and that the current stage is a step away from the elaborate finite paradigm? There are strong internal and external pressures that would favor such a change. Internal pressures include the marked status of the feminine and plural categories in the structure of the finite and non-finite paradigms, as well as the low frequency with which these categories appear. The external pressure is substratum interference. Large numbers of speakers of Tibeto-Burman languages learn Nepāli as a second language, and use it as a lingua franca. Gender is not marked on the verb in any of the Tibeto-Burman languages of Nepal, and number is marked in only some of them. A large number of these speakers will learn the verb agreement system imperfectly, and simplify the paradigm, especially the gender and number agreement. This simplification is then an additional motivation for the simplification of the system by native speakers of Nepāli, who are thus frequently exposed to verb forms where the verb does not fully conjugate. This type of substratum interference is not the classic case of speakers shifting from one language to another (see, for example, Thomason & Kaufman 1988), but comes instead from the continued imperfect bilingualism of second-language learners.

While one might be inclined to think that the agreement system in the process of simplification in modern Nepāli, and that this variation represents historical change in progress, there is evidence that this is not a recent phenomenon. The earliest set of spoken Nepāli texts was published by Sir Ralph Lily Turner just after World War I (1921, 1922). They are personal narratives of Gurkha soldiers and their experiences fighting in the war. Due to the nature of the subject matter, there are no feminine referents, hence no data on gender agreement. However, there is evidence for number agreement: in these narratives

the verb agrees in number with inanimate plural subjects only 20 per cent of the time, while it agrees with animate plural subjects about 75 per cent of the time. We can see that lack of perfect agreement in the verb was a feature of Nepāli almost a century ago.

What we see from this study is that the Nepāli system of register-based variation in agreement has persisted over generations, and most likely will continue to persist into the future. The persistency of the variation is due to competing motivations. There is a strong motivation for a simplification of the system, with both internal and external pressures at work. As a counter-balance, there is the strong prescriptive tradition: the complex paradigm is taught in schools, enforced in published Nepāli writings, used in much of the Nepāli media, and used in formal contexts, such as academic lectures. These competing motivations are the source both of the register-based nature of the system, and the persistence of the variation over time.

## 2 Variation within the Kāthmāndu Newār community

In my early work on Kāthmāndu Newār, before my first trip to Nepal, I worked with two Newārs who were students at the University of Oregon. In most respects, the phonological and grammatical systems of the two speakers were the same; I only began to discover significant differences between them when exploring syntactic constructions in depth. Both speakers exhibited a difference between a 'long participle' and a 'short participle' form. The long participle form is used by both speakers for narrative chaining, and is often accompanied by a distinctive intonation contour. The short participle is used for the incorporation of auxiliaries (all of which are versatile verbs) into the clause. Thus, both speakers differentiate the following sentences:<sup>1</sup>

- (1) *w-āā*    *nay-āā*    *con-a.*  
 3s-ERG    eat-L.PART    stay-PST.DISJUNCT  
 'He ate and stayed.'
- (2) *w-ā*    *nay-ā*    *con-a.*  
 3S-ERG    eat-S.PART    stay-PST.DISJUNCT  
 'He was eating.'

For one speaker, Rajendra, the short participle is only used for the incorporation of versatile verbs as auxiliaries. The other speaker, Manoj, however, also uses the short participle in a very restricted construction with motion verbs. In this construction, the V2 of the sequence must be either *wan-e* 'go' or *wa-ye* 'come', and the motion verb must be the only member of its clause; intervening arguments, locations, or adverbials are

<sup>1</sup> The following abbreviations are used in this paper: ERG ergative; LOC locative; L.PART long participle; NEG negative; NR nominalizer/relativizer; PART participle; PL plural; PST past; S.PART short participle; STAT stative.

prohibited. In this construction the focus is on the first verb, but the speaker insists that the meaning of the motion verb is also distinctly conveyed:

- (3) *jī-ī*     *nay-ā*     *wayā*  
 1S-ERG eat-S.PART come-PST.CONJUNCT  
 'I already ate (before I came)'

It is interesting that the distinction between the presence and absence of the third, intermediate, construction correlates with another distinction between these two speakers, that of the interpretation of the scope of negation in clause chains. Rajendra, the speaker who lacks the intermediate construction, is quite free in his interpretation of the scope of negation in clause chains, allowing negation on the final verb to be applied to non-final clauses. Thus compare the non-negated sentence in (4) with the negated sentence in (5):

- (4) *jī-ī*     *bārcā*     *kurk-āā*     *tachyān-ā*  
 1S-ERG bowl drop-L.PART break-PST.CONJUNCT  
 'I dropped the bowl and broke it.'
- (5) *jī-ī*     *bārcā*     *kurk-āā*     *ma-tachyānā*  
 1S-ERG bowl drop-L.PART NEG-break-PST.CONJUNCT  
 'I dropped the bowl but didn't break it.' OR 'I broke the bowl without dropping it.'

Not any combination of clauses may have backward spreading of negation in this way, rather the clauses must indicate events that are thematically continuous, and both clauses must have the same subject referent.

In contrast, Manoj does not allow an interpretation of example (5) with the focus of negation on the non-final clause. The only conditions under which Manoj will allow backward scope of negation are the same conditions that hold for his intermediate chaining construction: the final verb must be a motion verb and it must directly follow the preceding verb. In this construction, the vowel of the participle may be either long or short, with no noticeable difference in meaning:

- (6) *wā-ā*     *nay-ā(ā)*     *ma-wā-ā*  
 3S-ERG eat-(L.)PART NEG-go-STAT  
 'He ate and didn't leave.' OR 'He didn't eat and left' (i.e., 'he left without eating')

There is more to say about these constructions and the differences between these speakers (see Genetti 1986 for a fuller description), but this much is sufficient to make the following point: speakers may appear to have very similar speech on initial observation, but prove to have significant and systematic differences at a quite deep grammatical level.

To what should one attribute the differences between these speakers? One obvious approach, and the one I took in my original analysis, is to assume that the two Newārs speak different dialects. This analysis is in accordance with the expectation of inter-speaker consistency. Kāthmāndu Newār society is highly complex and stratified. Members of the community are differentiated by caste, religion, occupation, gender,

socioeconomic and educational level, and neighborhood. The two speakers in question are from different castes and different parts of the city, such that an analysis of different dialects is certainly plausible. However, there are not other many obvious differences in the speech of these two consultants.

Another possible factor to which one might attribute this variation is the data elicitation technique. These data were collected in the artificial setting of a university office far removed from the speech community, out of context, and as translations of possible English sentences. So perhaps the difference in interpretation was due to one or both speakers losing track of their native intuitions in their role as informants.

A third possibility is that the variation is idiolectal, and that each speaker has independently constructed different grammatical systems of clause combining even though they had similar input as children. This analysis goes against the expectation of inter-speaker consistency, however, given my subsequent experience of working with significant idiolectal variation in Dolakhā Newār (discussed below), it has become clear to me that idiolectal variation is pervasive, and one cannot assume dialectal distinctions without clear independent evidence that the dialects exist.

### 3 Idiolectal variation in Dolakhā Newār

The expectation of inter-speaker consistency dominated my view of language until my second field trip to Nepal to study Dolakhā Newār. During my first field trip I had worked with one primary consultant named Kalpanā Shrestha, a young woman who had lived in Dolakhā until the age of twelve, then moved to Kāthmāndu with her family, where she was actively involved in the Dolakhā Newār community there and used the language regularly. Working with her, I collected vocabulary, conducted elicitation to determine the basic outlines of the grammar, and transcribed and translated a number of narratives produced by a variety of native speakers. When I returned to continue my work the following year, Kalpanā had taken a job and had little time to spare for work with me. She introduced me to Rama, her cousin two years older. With her, I also collected vocabulary, conducted elicitation and transcribed recorded texts, although I continued to work with Kalpanā when possible.

Kalpanā and Rama are first cousins. Their fathers are brothers, and, following the traditional Newār pattern, when they married they each brought their new wives into their shared house, and raised their children together. Thus, the two girls grew up in exactly the same linguistic community until Rama was seven and Kalpanā was five. At that point, Rama's father took a job with the government and was posted to other parts of Nepal. However, the family continued to consistently speak Dolakhā Newār in their home. Thus there are neither substantial regional or socio-demographic differences between the two women. The only possible difference between their linguistic backgrounds is that Rama

left the village at the age of seven at which point she interacted primarily with her parents for several years, and Kalpanā left the village at twelve. However, they both continued to be active members of the Dolakhā Newār community and continued to use the language consistently. It thus came as a surprise to me when I found variation in their speech at the phonological and morphosyntactic levels. I will present here a few, rather simple, examples of the variation in the speech of these two first cousins. There are many other areas in which their speech varies.

### 3.1 Variation in phonology

To begin with a simple example, Kalpanā pronounces the first person exclusive pronoun as /chiji/ whereas Rama pronounces it as /thiji/, although both women clearly have phonemic systems which differentiate between the aspirated alveolar stop /th/ and the aspirated alveopalatal affricate /ch/. This is a clear, if perhaps minor, violation of the expectation of inter-speaker consistency. When I pointed this difference out to the women one time when we were all together, they were both quite surprised and each laughingly insisted that her pronunciation was correct. In describing the pronominal system of the language then, the linguist is faced with a choice. Either choose one of the pronunciations as 'correct' and representing the pronoun, or list both pronunciations, thereby allowing for variation in the linguistic description. The problem with the first path is how to choose the 'correct' or 'basic' version. Perhaps one could do a survey of a representative sample of speakers, and choose the form that occurred most frequently, or one could choose /ch/ for systemic reasons as it corresponds to the second person pronoun *chi*, from which *chiji* is transparently formed (and for which there is no alternative *thi* to my knowledge). Or, one could choose *thi*, since it is more likely that this is historically prior, as the vowel /i/ together with the aspiration creates the ideal environment for palatalisation. But, if our linguistic descriptions are meant to be close and detailed portraits of the state of the language as we find it, then the most accurate description is the one that admits and highlights the variation. We may lose something in descriptive elegance, but we gain a more accurate and realistic portrait.

It is interesting to consider the *thiji/chiji* variation in light of the pronominal paradigms of which the forms are part. Although /th/ and /ch/ are clearly distinct phonemes and differentiate a number of nouns and verbs, they actually do not contrast forms within the narrow lexical class of personal pronouns. Assuming that speakers are aware when pronouns are being used in natural discourse—from their syntactic positioning, inflection, and discourse functions—they then do not have to maintain a strict phonemic distinction between similar phonemes. In this environment, the phonemic distinction can be relaxed, and variation easily tolerated without problems of intelligibility.

More extensive phonological variation between the two consultants is found in their vowel harmony systems. Dolakhā Newār has three verbal prefixes, the negative *ma-*, the



prohibitive *da-*, and the optative *tha-*. The vowels in all three prefixes are subject to harmony depending on the vowel of the stem. Rama has the simplest system of vowel harmony, as only the vowel /ā/ triggers harmony, for example *mā-y ā* 'didn't come'. Harmony is blocked when the stem has a glide in C2 position, for example, *ma-syāt* 'didn't kill', *ma-mwāl* 'didn't search'. Kalpanā's system is more complicated, as /ā/, /o/, and the sequence /wā/ all trigger harmony, thus, *dā-dāu* 'don't beat', *mo-sou* 'didn't see', *mwā-mwāl* 'didn't search'. A third consultant, a much younger cousin of the first two, has a more extensive system yet, with the feature [round] harmonising from /u/: *mo-pul* 'didn't pay'.

It is clear that all three speakers have formed distinct vowel harmony systems, but, as with the different pronunciations of the pronoun, none of them indicated awareness of any differences between their own speech and that of their cousins. Although the different vowels that appear in the prefixal allomorphs are distinct phonemes, they actually have no phonemic value in the prefixes themselves; these are the only three prefixes in the entire language and they are differentiated by the initial consonant. Thus the vowel carries no functional load and its precise quality is unimportant. As with the case of the pronoun, it thus appears that inter-speaker variation is especially tolerated in environments where it does not matter.

It is interesting to consider variation such as this in historical terms. Could this variation imply that vowel harmony in Dolakhā Newār is undergoing a change not yet completed, perhaps becoming more elaborate? If so, then we have an explanation for this counter-example to the expectation of inter-speaker consistency. It is simply a system in change, and over a matter of time inter-speaker consistency will once again be achieved. Of course this is a real possibility, and only time will tell, but there is no reason to assume that this is the case. It is just as likely that the reason for the synchronic variation is that the input that these speakers had as children was also variable, and that variation itself is a stable property of the language that persists over time.

### 3.2 Variation in morphosyntax

There are two nominalising suffixes in Dolakhā Newār, which are used in the formation of relative clauses, in complements of perception verbs, in complements of cognition verbs, in complements of *ju-en con-a* (roughly a mirative expression), in emphatic constructions, and in some types of questions. The analytical difficulty comes in determining under what conditions each of the two suffixes appear in these various environments. A full discussion is beyond the scope of the current paper (see Genetti 1994:154–170). Here I will begin by discussing the inflection of the verb in relative clauses. Note that since there are no simple functional terms which can be used to label these suffixes, I have resorted to calling them NR1 and NR2, for 'nominaliser/relativiser' 1 and 2 respectively.

When I first began working with Kalpanā on the distribution of these forms in relative clauses, I found her system to be quite straight-forward. NR1 was used in subject relative clauses (7), while NR2 was used in object relative clauses (8).

(7) *chē=ku ye-u mi-pen*  
 house=LOC come-NR1 person-PL  
 'people who came to the house'

(8) *jīn khoṅ-a keṭi*  
 1S.ERG see-NR2 girl  
 'the girl whom I saw'

This distribution was borne out in text counts using texts from a number of speakers; out of one hundred and twenty subject and object relative clauses examined, there was only one counter-example to this pattern, and that from a speaker who was openly criticised by others in the room for mixing up the story as she told it. Given this clean, motivated pattern in my data, it was thus surprising to find that Rama's opinions on the possible distribution of the suffixes differed from Kalpanā's. For subject relatives, Rama preferred to use NR1 consistently, but she also allowed the possibility of NR2 in examples where the aspect of the clause was imperfective, as in (9).

(9) *ām āmp kha-en coṅ-a / co-gu mucā*  
 that mango pick-PART stay-NR2 / stay-NR1 child  
 'the child picking mangos'

For object relatives, Rama again preferred NR2, but she said that NR1 was also possible in examples when the subject of the relative clause is third person, as in (10):

(10) *āmun khoṅ-a / khoṅ-gu mi-pen*  
 3S.ERG see-NR2 / see-NR1 person-PL  
 'the people that he saw'

Kalpanā, however, did not accept either of these optional patterns, and insisted that NR1 was the only form possible for the example in (9), and that NR2 was the only form possible for (10).

My first inclination upon seeing this variation was to attribute it to the elicitation setting. One must admit it is a rather strange task to translate series of sentences such as 'I saw the man who gave Sita the money' and 'I saw the money that the man gave Sita', and maybe it would be easy for a consultant to get mixed up. In short, I assumed that Rama's seemingly unmotivated system was probably due to error on her part. However, her opinions remained consistent over time. I then began to elicit oblique relative clauses from both consultants, and found that while both 'preferred' NR2 consistently in these examples, they both admitted that it was possible to use NR1 with oblique relatives if the subject of the clause is third person, thus:

(11) *ām misāmi ye-e / ye-u gāū*  
 that woman come-NR2 / come-NR1 village  
 'the village that the woman came from'

Text counts of oblique relative clauses with third-person subjects did indeed show such variation; while twenty-two examples had NR2 (judged 'preferred' by both consultants), five examples had NR1. Looking back to Rama's earlier judgments on object relatives, I noticed that the pattern she had given me for the inflection of the verb in object relative clauses was the same pattern that she and Kalpanā had both given me for the inflection of the verb in oblique relatives, and for which there was evidence in the texts. Thus I began to consider the possibility that, rather than being confused about the inflection of the verb in object relatives, Rama's judgments could result from a more finely tuned awareness of variation in the use of these forms.

In examining the distribution of NR1 and NR2 in other environments, I found that different factors condition their appearance in different environments: With complements of perception verbs, verbal transitivity is important; with complements of cognition verbs, perfectivity is the conditioning feature; in questions and emphatic constructions, person of the subject is relevant; in complements of *juen cona*, both transitivity and perfectivity are the factors speakers attend to. What emerges is a complex set of variables which underlies the distribution of the two paradigmatic suffixes, but no clearly motivated pattern. While the distinction between the suffixes can probably be traced to an old historical distinction based on transitivity (see Genetti 1994:169–171), there is no evidence for this synchronically. The attested variation appears, then, to be due to a decayed system which lacks a coherent functional motivation for the two suffixes, and whose use has thus become idiosyncratic and variable. Faced with this type of situation, speakers may choose to reanalyze and regularise the system. This seems to be happening in the case of relative clauses, the most frequent environment where the suffixes occur, and which appear for some speakers to have moved into a motivated system based on grammatical relations.

#### 4 Implications of the studies

In the discussion of the case studies, I have made a number of points about the nature of linguistic variation, which may be summarised as follows:

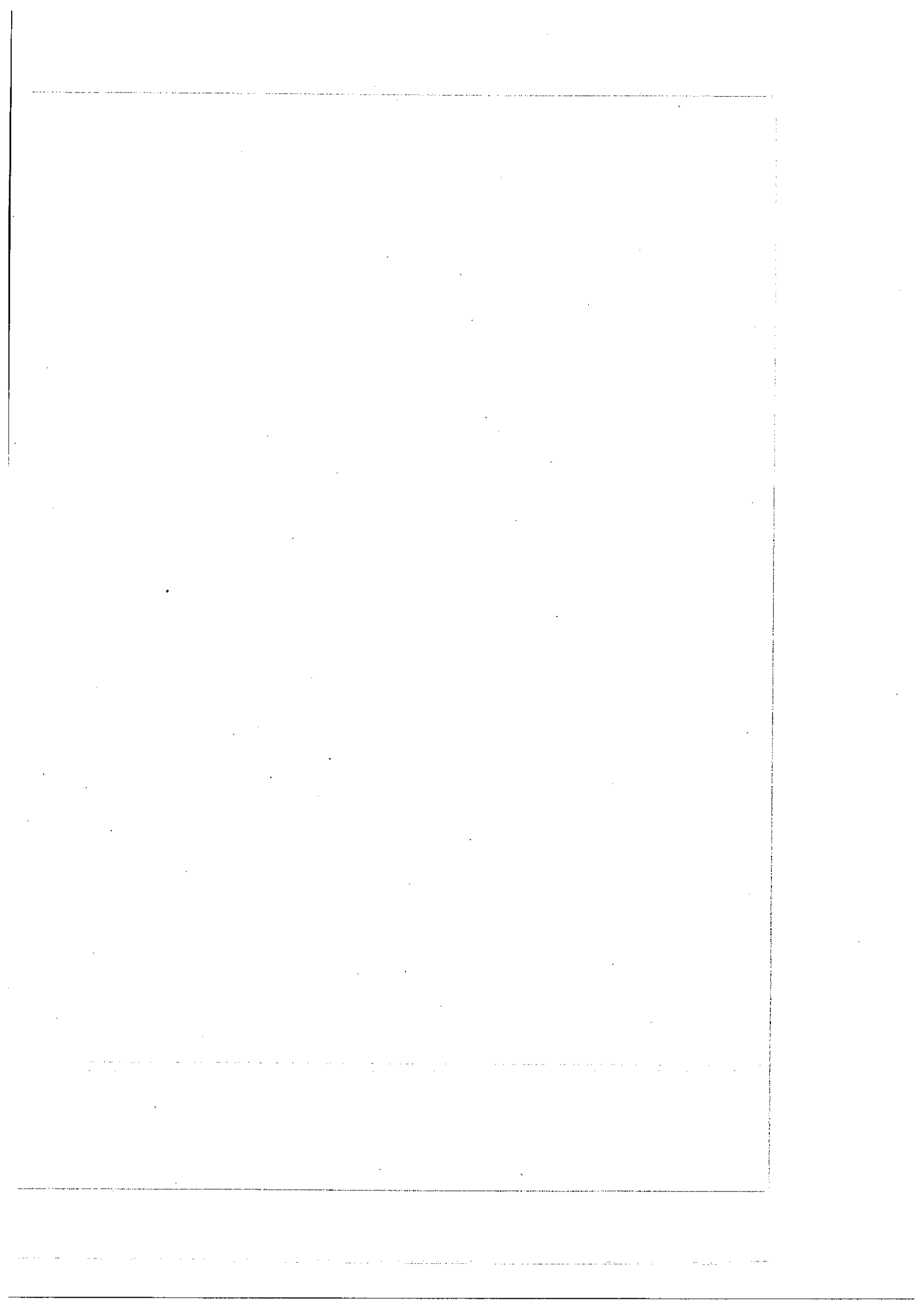
- Variation may persist across generations
- Variation may result from competing motivations, which are resolved differently by different speakers, or result in register-based variation
- Speakers may appear on the surface to have very similar linguistic systems, but differences may emerge at quite deep levels of the grammar
- Significant idiolectal variation may exist in the phonology, morphology and syntax, even among speakers with identical linguistic backgrounds
- One should not attribute variation to dialect without independently proving the existence of the dialects; in the absence of this evidence, the variation may be idiolectal as opposed to dialectal

- Idiolectal variation may be more likely to arise in cases where it makes no difference, that is, in cases where there is no functional loss resulting from the variation
- Variation may reflect systems in the process of historical change, but doesn't necessarily have to.

It is clear that one may draw many deep insights into the nature of language by exploring linguistic variation, and that incorporating variation into linguistic description allows for a richer understanding of language as well as a more accurate portrayal. Our tradition in linguistic theory is to expect inter-speaker consistency, and to assume that we can produce a single accurate statement about linguistic systems and subsystems. Variation is often attributed to dialect or register only, and so 'accounted for', often with little in-depth exploration or justification. With this approach we risk obscuring the richness and diversity of language in our search for generalisations. We are fortunate to have the work of Jim Matisoff as a model of an alternative approach, which inspires us not just to work with unkempt variation, but to revel in it.

## References

- Genetti, Carol, 1986, Scope of negation in Newari clause chains. Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics, Ohio State University, 11–14 September 1986.
- 1994, *A descriptive and historical account of the Dolakha Newari dialect* (Monumenta Serindica). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- 1999, Variation in agreement in the Nepali finite verb. In Warren Glover and Yogendra Yadava eds *Topics in Nepalese linguistics*, 542–555. Kathmandu: Royal Nepal Academy. (First published 1993, *South Asian Language Review* 3/2:90–104.)
- Thomason, Sarah Grey and Terrence Kaufman, 1988, *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Turner, Ralph Lily, 1921, Specimens of Nepali. *The Indian Antiquary* 50:84–92.
- 1922, Further specimens of Nepali. *The Indian Antiquary* 51:41–67.



# 5 *Recent language contact in the Nepal Himalaya*

---

MICHAEL NOONAN

## **Introduction**

The Nepal Himalayas have been the scene of extensive linguistic contact over a considerable period.<sup>1</sup> Languages of different genetic phyla, in particular Indo-European and Tibeto-Burman, have been involved, but so have languages within the Tibeto-Burman phylum representing different stocks with differing typological characteristics. Indeed, the long periods of contact between speakers of Tibeto-Burman languages of different stocks have resulted in considerable lexical and grammatical borrowing, which has tended to obscure genetic relationships. As a result, there is still a good deal of uncertainty as to how even major groupings of languages should be positioned within the family tree.

In recent times in the Nepal Himalaya, large-scale population movements, both from outside Nepal into the country and within Nepal itself, have resulted in a wide variety of contact situations involving at least the following groups:

- 1) speakers of Tibeto-Burman languages which have been in Nepal for long periods (that is, languages which are usually grouped together as 'Himalayish');
- 2) speakers of Tibeto-Burman languages which have moved south of the Himalayas within the last two millennia (that is, speakers of Bodish languages, especially Tamangic, Ghale, and languages of the Tibetan Complex, such as Baragaunle and Sherpa), and;

---

<sup>1</sup> The work reported on in this paper has been supported by the following grants from the National Science Foundation: DBC-9121114, SBR-9600717 and SBR-9728369. I would like to thank Edith Moravcsik and Colin Masica for helpful comments. I would also like to thank Karen Grunow-Hærsta for help with the Syanja Magar data.

- 3) speakers of Indo-European Nepali, which was well established in Western Nepal in ancient times, but has moved into central and eastern Nepal more recently.

Contact among speakers of these languages has resulted in considerable borrowing of lexical and grammatical material. In this paper, I will present data on sixteen Tibeto-Burman languages of Nepal with the aim of showing what these languages have borrowed, and from what source. I will concentrate here on grammatical borrowing as opposed to lexical borrowing, though in practice the two are sometimes difficult to disentangle.

The method that I am using to chart grammatical borrowing is a fairly simple but a rather crude one, subject to known errors of a variety of sorts. The method can be described as follows: for each of the major genetic groupings (Bodish, Himalayish, and Nepali—and for some subclassifications with the first two), I establish values for a set of typological parameters. For each language, grammatical features are compared to the typological profile for its group and, if the values are different, it is assumed that change has taken place. This sort of approach can only succeed if a large number of features and languages are considered, and I have tried to do just that.

The structural parameters chosen for the typological profiles are those which will yield differences among the three groups. So, for example, there would be little point in including a parameter for the order of major clausal constituents since all three groups are strongly SOV. On the other hand, the presence of distinctive tone, the order of adjectival modifiers and their heads, the existence of split ergativity of certain types, and so on represent parameters which will yield different values for these groups and thus can be used to map possible instances of borrowings or convergence. The structural features used in this survey are summarized in Table 1 and are discussed individually later in the paper.<sup>2</sup>

The language groupings themselves require some comment. Of the three groupings, Nepali is the least problematic since it consists of only one language. However, one finds a number of differences, even with regard to the parameters considered in this study, between standard literary Nepali and colloquial regional Nepali. For example, in the Baglung and Myagdi districts of Nepal, I have heard Nepali which is consistently ergative—that is, does not exhibit the aspect-based split ergativity of Standard Nepali—and in which the Standard Nepali distinction between the dental and retroflex series of consonants is neutralized in an apical alveolar series. The typological profile used here is based on the standard dialect, however, since this dialect, for the most part, represents a more conservative version of the language (i.e. more like the one speakers of TB languages would have encountered in times past) and is the one held out as a normative model in the schools and in the mass media. And, it is also worth noting that Indo-Aryan influence on the TB languages of Nepal predates the arrival of Nepali and its predecessors in the central and eastern portions of the country. Newari, for example, has been influenced by Indo-Aryan for a very long period.

---

<sup>2</sup> The tables can be found at the end of the paper.

For our purposes here, the Bodish group consists of the Tamangic languages, Ghale, and local representatives of the Tibetan Complex, such as Baragaunle (a variety of Loba) and Sherpa. The typological profile that can be assigned these three subgroups prior to contact with Nepali is pretty similar, though there are some differences as will be noted in the text. Though there are some uncertainties as to how the languages should be grouped—Ghale in particular, this grouping is clearly a genetic grouping, and a reasonably close one at that. The presumed genetic relationships among these languages are presented in Figure 1.

The Himalayish group is much more problematic. It isn't clear, for example, that this represents a genetic grouping at all as opposed to a geographic assemblage of TB languages that have been in contact in the sub-Himalayan region of Nepal for a long period. The exact relation of Newari to the rest is particularly problematic. Nonetheless, from a typological point of view, these languages can be presumed to have shared, prior to contact with Nepali and other Indo-European languages, a rather distinctive typology. And there is still a remarkable similarity in typological profile of the languages at the western and eastern extremes of their distribution—Kham and the Kiranti languages, languages which have experienced the least amount of interference from Indo-European and are thus presumed to have preserved better the earlier typological profile. We therefore assume—and this is a very strong and possibly incorrect assumption—that at an earlier period all the languages in this grouping had the typological profile still shared by Kham and the more conservative Kiranti languages.<sup>3</sup>

So, in sum, we assume that the languages classified as Himalayish have been in Nepal for a long period and had shared a distinctive typology. The Bodish languages, represented by the Tamangic group, had entered the sub-Himalayan region of Nepal perhaps 1500 years ago, with Ghale perhaps entering a bit later. The entry of members of the Tibetan Complex into this zone is relatively recent. Indo-Aryan Nepali, though well established in the west of Nepal a thousand years ago, has become important in central and eastern Nepal, where the languages in this survey are spoken, much more recently.

Before going on to discuss the structural features used to make the typological profiles, we should say something about the sample of languages discussed in this survey. The sixteen languages discussed here, although distributed across the genetic groupings, constitute a 'convenience sample' based on limitations of available data. Data from Sherpa and any of a number of additional Kiranti languages were available and would have increased the value of the study. In later versions of this paper, these data will be added. The languages used in the study and the sources of data for them are given in (1); their

---

<sup>3</sup> This is especially problematic for Newari, which, as noted above, has been under varying degrees of Indo-Aryan influence for a long period and whose exact relationship to the rest of the group is problematic.



genetic classification is provided in Figure 1 and their locations within Nepal can be found on the map in Figure 2:

- (1) Athpare: Ebert (1997a)  
 Baragaunle: Mary Brehm fieldnotes; Kretschmar (1995)  
 Camling: Ebert (1997b), Winter (1985)  
 Chantyal: Michael Noonan fieldnotes; Noonan et al, (1999), Noonan (2003a)  
 Chepang: Caughley (1982), Thompson (1990)  
 Ghale: Holly Smith fieldnotes; Smith (1999)  
 Gurung: Glover (1974)  
 Hayu: Michailovsky (1988b)  
 Kham: Watters (1998)  
 Limbu: van Driem (1987)  
 Magar [Syangja]: Karen Grunow-Harsta fieldnotes  
 Magar [Tanahu]: Karen Grunow-Harsta fieldnotes  
 Nar-Phu: Michael Noonan fieldnotes; Noonan (2003b)  
 Dolakha Newari: Genetti (1994)  
 Kathmandu Newari: Genetti (1994), Malla (1985), Michael Noonan fieldnotes  
 Thakali: Georg (1996), Michael Noonan fieldnotes

Nepali is not one of the languages in the survey, but it figures prominently in the discussion: almost all young adults in Nepal now speak Nepali fluently and so bilingualism is an every day fact of life. As the rate of lexical borrowing increases, so does the rate of structural influence, as we will see later in this paper. The values assigned to the languages in the survey for each of the structural features are summarized in Table 2. We will take up each in turn.

**Phonemic voicing contrasts:** I refer here to contrastive voicing in stops and fricatives. The typological profile of the Bodish languages of Nepal does not include contrastive voicing, whereas voicing is contrastive for the Himalayish group and Nepali.

Of the Bodish languages in our sample, all the languages are consistent with the profile save Chantyal and Gurung. Chantyal, with its massive borrowing of Nepali vocabulary (71 per cent) of the items listed in Noonan et al. (1999), has assimilated to the Nepali type. In Gurung, contrastive voicing is limited to certain tones. In the Himalayish group, all the languages are consistent with the profile except for the Kiranti languages. Ebert (1997a) reports that in Athpare voiced consonants are rare in initial position; for Camling, Ebert (1997b) reports voicing contrasts for labials and dentals only. In Limbu, in native vocabulary, a voicing contrast is found only in bilabials, though borrowings from Nepali are establishing the contrast for other points of articulation.

**Tone:** The Tamangic languages have been described as having a basic four-tone system (Mazaudon 1973, 1978a, 1978b, 1993-94). The members of the Tibetan Complex in Nepal mostly exhibit a two-tone system, which, as in Tamangic, is related to voicing of the initial consonant. Nepali is, of course, non-tonal, and we assume that tone in the

Himalayish group is of recent origin, as, for example, Michailovsky (1975) has shown for Khaling.

Chantyal has lost its tone system under Nepali—and perhaps Magar—influence. It retains an opposition of murmured versus plain syllabic nuclei, which resembles that found in some Magar dialects.

Of the Himalayish languages in our sample, only Kham has acquired a tone system, almost certainly under Bodish influence. A number of Kiranti languages have acquired tone systems, in particular those spoken in the northern reaches of the Kiranti-speaking area: these languages can be presumed to have had the most contact with Bodish languages.

**Murmur:** The presence of murmur and its role in the phonological system clearly distinguishes the three groups. For the Bodish languages, at least those in Nepal, murmur is a concomitant of tone, typically associated with low tone. In this way, murmur can be found in syllables with voiceless as well as voiced initials, at least in some languages. In Nepali, murmur is phonologically a feature of consonants. And in the Himalayish group, we can assume that the presence of murmur is an innovation deriving from contact with either Nepali or Bodish.

While in the main the Bodish languages in Nepal preserve the original role of murmur as a concomitant of tone, the large and ever-increasing number of borrowings from Nepali has resulted in disturbances in the traditional relation between tone and murmur with many words now having phonemic murmur on the Nepali pattern. Predictably, this affects Chantyal the most, and the most recent borrowings are taken in with murmur preserved according to the Nepali pattern where murmur can occur independently on consonants in syllabic onset or coda: earlier borrowings did not follow this pattern.

The situation in Himalayish is complex. Kham has murmur as a concomitant of tone in the manner of the Bodish languages. The Newari dialects are split: Dolakha lacks murmur altogether, but Kathmandu has it in the Nepali fashion. (Genetti 1994 reconstructs murmur for Proto-Newari.) The two Magar dialects have murmur in more-or-less the Nepali fashion, though Tanahu shows signs of a 'register' system similar to that which Chantyal must have had after losing its tone system. Kiranti languages have assimilated murmur in varying degrees: Limbu has murmured stops only in a few loan words, while in Athpare and Camling, murmured stops occur in (presumably) native words, but they are not numerous. In Chepang, Caughley finds phonetic murmur, but analyzes it phonemically as a sequence of voiced consonant and /h/. Murmur in these languages is generally in the Nepali fashion.

**Voicing opposition in liquids and/or nasals:** The Bodish languages in Nepal prototypically have a series of voiceless liquids and, occasionally, voiceless nasals, voiceless *m* being the most common. These are lacking in Nepali and are assumed to be lacking in earlier stages of Himalayish.

In Bodish, Chantyal and Ghale now lack these sounds, though both have murmured nasals and liquids, unlike Nepali.

In Himalayish, the Hayu-Chepong group have voiceless liquids and Chepong has voiceless nasals as well. For Athpare Ebert reports one word with /rh/, which she refers to only as an 'aspirated r'. For Camling Ebert reports /lh/, /rh/, /mh/ and /nh/, none of which are reported by Winter (1985), though Ebert provides minimal pairs with plain liquids and nasals. Kathmandu Newari has murmured liquids and nasals, but lacks a voiceless series.

**Retroflex series:** We refer here specifically to either of two sorts of oppositions among stops: dental versus true retroflex and dental vs alveolar, with the latter being affricated with a rhotacized off-glide such as [ɾ]. The first is characteristic of Nepali and languages influenced by it, the second characteristic of the Bodish group. See Michailovsky (1988a) for discussion.

Once again, Chantyal differs from the other Bodish languages, here in lacking a retroflex series of any sort. In this too, it may have moved to be more in accord with Magar: Magar natively has an alveolar series (perceived by Nepali speakers as retroflex) only and is (in Tanahu) acquiring an opposition with a dental series through borrowings from Nepali. Of the Himalayish languages in our sample, only Dolakha Newari has a retroflex series that is not obviously the product of recent borrowing from Nepali. Michailovsky reports that Dolakha Newari has retroflex consonants of the Nepali type.

**Fricatives and affricates:** The Bodish group is characterized by a phonemic opposition between an alveolar and an alveopalatal series of fricatives and affricates. Spoken Nepali lacks such an opposition—though a few purists pronounce written ऋ as an alveopalatal; the Himalayish languages also lack this opposition. In our sample, Chantyal once again assimilates to the Nepali type, having phonetic [s] and [ʃ] in complementary distribution. Marphatan Thakali, as described by Georg, has two affricates as separate phonemes, but not a corresponding pair of fricatives.

**Phonemic nasalized vowels:** Nepali has phonemically nasalized vowels, and distinctive nasality is assumed for Tamangic (Mazaudon 1993–94) and is widely found in Nepal Bodish. Distinctive nasal vowels are assumed here not to characterize the Himalayish group, though a feature like this may arise spontaneously in any linguistic grouping.

Having said that, we note that distinctive nasal vowels seem to be associated with geographical groupings. In our sample, the two Bodish languages spoken in the Kali Gandaki Valley, Thakali and Baragaunle, lack nasal vowels. The other Tamangic languages have phonemic nasal vowels, and so do the other members of the Tibetan Complex in Nepal for which we have data: Jirel and Sherpa. For Ghale, nasal vowels exist, but may have a low functional load.

Among the Himalayish languages, Kham has nasal vowels, as do both dialects of Newari and Hayu. The remaining Himalayish languages our sample lack them, save for

Camling, for which distinct nasalization is reported by Ebert only for /o/ and /a/. Winter reports nasalized counterparts for /e/ and /u/ also.

**Λ ~ ɒ allophony:** This refers to a characteristic of Nepali which has been passed on to a number of other Nepalese languages. In Nepali, the mid-central phoneme /ə/ has two allophones, a mid-central vowel and a low back rounded vowel in more-or-less free variation. This feature has been borrowed in Thakali, Chantyal, Ghale, Syangja Magar, and Kathmandu Newari, all of which have evolved 6-member vowel systems like Nepali's.

**Word initial /ŋ/:** The Bodic<sup>4</sup> languages are characterized by allowing the velar nasal to appear in word initial position; Nepali does not. In our sample, Chantyal and Kathmandu Newari have converged with Nepali in not allowing /ŋə/ to appear word initially. In the latter, as in Nepali, the velar nasal is present only allophonically, by assimilation. Michailovsky reports initial /ŋ/ to be rare in Hayu.

**Stress:** In Standard Nepali, stress is phonemic, though it is largely predictable from the orthography, which writes distinctions in vowel length that are no longer pronounced. In the Bodic groupings in our sample, stress was predictable and was generally fixed on the root. Where languages have borrowed large amounts of Nepali vocabulary, they have accommodated to the Nepali stress pattern. Of the languages in our sample, this is most evidently true of Chantyal, though it is true to lesser degrees for most of the other languages.

**Prefixes:** The Himalayish languages, as a group, are prefixing. The Bodish languages and Nepali, with a few exceptions (such as the negative prefix in Bodish), do not use prefixes.

All of the Bodish languages are consistent with their traditional typology and with Nepali in not allowing prefixes. The Himalayish languages continue to use prefixes, save for the Newari dialects and the Hayu-Chepeng group, which resemble the Bodish languages in having no prefixes but the negative.<sup>5</sup> In the Magar dialects, the number of prefixes is very small, however, in comparison to the number of suffixes.

**Person/number inflection on verbs:** The three groupings present different typologies: in Bodish there is no person/number agreement on verbs, in Nepali there is agreement only with the subject, and in the Himalayish languages there is agreement, potentially, with two arguments.

The Bodish languages all lack person/number agreement morphology, though Baragaunle has innovated an evidential system that resembles the conjunct/disjunct system found in Kathmandu Newari. In the Himalayish group, the Hayu-Chepeng group, Kham, and the Kiranti languages have person/number agreement with multiple arguments. Syangja Magar has subject agreement, but the Tanahu Magar dialect lacks argument/verb

---

<sup>4</sup> Note that 'Bodic' is not the same as 'Bodish': see Figure 1.

<sup>5</sup> In Chepeng, the negatives are expressed as suffixes.

agreement altogether. Kathmandu Newari also lacks agreement (though it has a conjunct/disjunct distinction for which person is relevant—see Genetti 1994); Dolakha Newari has an agreement system which references subjects only.

**Reflexive:** This parameter patterns exactly like the previous one: in accordance with their complex argument/verb agreement patterns, Himalayish typology would express reflexives as part of their verbal word. Nepali and the Bodish languages express reflexives analytically, with either a special reflexive form or an ordinary personal pronoun.

The Bodish languages all have analytic reflexives, with Chantyal and Nar-Phu using only ordinary personal pronouns. In the Himalayish group, the Hayu-Chepong group, Kham, and the Kiranti languages express reflexives as part of the verbal word, but the Magar dialects and the Newari dialects have analytic reflexives on the Nepali model.

**Adjectival word order:** All the groupings allow AN order, but Bodish allows also NA and some Bodish languages favor this order.

In our sample, all the Bodish languages allow NA order—this is really the preferred order in Nar-Phu—except for Chantyal, which has only the AN order. All the Himalayish languages in our sample have only AN.

**Demonstrative word order:** The Tibetan Complex has N Dem; all other groupings in our sample have Dem N. All the languages in our sample are consistent with their typological profile for this feature.

**Numeral word order:** Bodish languages have N Num, whereas Nepali and the Himalayish group have Num N. All the languages in the sample are consistent with their typological profiles save Chantyal, which has borrowed all its numbers from Nepali along with the order that goes with them, and Hayu and Camling, both of which allow N Num order along with Num N.

**Ergative syntax:** The Bodish languages in Nepal are fairly consistently ergative, but where they deviate from a strict ergative case assignment, the deviation is conditioned by pragmatic factors like topicality, volitionality, and so on. The Himalayish languages have a similar profile, though split ergativity often follows the animacy hierarchy, with speech act participants following an accusative syntactic pattern. Standard Nepali, on the other hand, has an aspect based split ergativity.

All of the languages in our sample are consistent with their grouping's typological profile save Tanahu Magar, which appears to have a split ergative syntax along the Nepali model, and Camling. Conservative varieties of the latter are consistent with the Himalayish typological profile, having a consistently ergative syntax for 3rd person subjects, but no case marking on 1st and 2nd person transitive subjects; however, Ebert reports that some speakers now show Nepali-type split ergativity.

**Antidative syntax:** By 'antidative', we mean the use of the case marker associated with indirect objects on certain classes of direct objects, especially animate patients (see Dryer

1986 and Noonan 1991). This syntagm is associated with Nepali and is absent from the traditional typological profiles of all the Bodic groups.

In our sample, only Ghale, Hayu, and the Kiranti languages Athpare and Limbu do not exhibit the antidative use of the indirect object case form. And in Ghale, there is evidence indicating that the phenomenon is taking root there too, but it is still marginal to the system at this point. (Note that Hayu lacks a dative case marker.) In a few cases (for example, Gurung and Camling), the Nepali dative case morpheme has been borrowed along with its syntax.

**Dative subjects:** The 'dative subject' construction is one in which the most animate argument is rendered in the case ordinarily assigned to indirect objects and, moreover, acquires many of the characteristics of subjects in the language. Semantically, dative subjects are typically non-volitional experiencers. See Masica (1991) for an extended discussion.

The dative subject construction is a prominent feature of Nepali syntax, but is not characteristic of the Bodic languages of Nepal.

Acquisition of this feature among the Bodic languages seems to be an index of Nepali influence. Chantyal, the Magar dialects, and the Newari dialects clearly have acquired it. Examples of the construction can be found in other Bodic languages in our sample, but in those languages, so far as we can tell, the construction has not yet been fully integrated into the syntax.

**Compound case:** 'Compound case' refers to the compounding of case clitics creating complex expressions, e.g. Chantyal:

*dəu-phyarā-mar-gomsə*  
tree-SUBSESSIVE-CIRCUMULATIVE-ABLATIVE<sup>6</sup>  
from down around the base of the tree'

This pattern, which also serves as a source of case clitic renewal, is characteristic of the Bodic languages. Nepali does not use this pattern and augments its case clitics from other sources.

All of the languages in our sample, even Chantyal, the one most affected by contact with Nepali, employ case compounding.

**Vertical case and vertical verbs:** 'Vertical case' refers to locative, ablative, and allative case forms whose meanings include the vertical directional senses 'up', 'down', and 'level'. 'Vertical verbs' refer to verbs with a sense like 'come' and 'bring' which includes also a vertical dimension, i.e. 'come from above', 'bring from below', etc.

These phenomena are characteristic of and probably restricted to the Kiranti languages (see Ebert 1994), though other Bodic languages, for example, Chantyal, may habitually

---

<sup>6</sup> Note that all of the case clitics in the example are in origin compound cases.

specify the vertical dimension by other means, for example, adverbials. Of the languages in our sample, all the Kiranti languages have vertical verbs, but only Camling has vertical case. Hayu may have vertical case (though this is not altogether clear from Michailovsky's description), but seems to lack vertical verbs; Chepang appears to lack both.

**Morphological valence increasing strategies:** We refer here to derivational processes which increase valence [applicative or causative]; all three groups have periphrastic causative constructions, and we are not concerned with them here.

Nepali has such derivational morphology as do all the Bodic groupings except Tamangic. In many of the Bodic languages, however, such strategies are nonproductive, though pairs resulting from these strategies are numerous. All the languages in the sample are consistent with their typological profiles except Chantyal, which has borrowed valence increasing morphology from Nepali. The morphology, which, interestingly, is not identical to that found in Nepali, is restricted to [the very numerous] Nepali borrowings.

**Morphological valence decreasing strategies:** Here too we are concerned with morphological, as opposed to syntactic, devices for decreasing valence such as passive. Nepali has such a device, but such strategies, we assume, are not part of the typological profiles of the Bodic languages. The Bodish languages do not have any such strategy, even Chantyal. Among the Himalayish group, the Newari dialects and Chepang do not have such a strategy either, though the other Himalayish languages in our sample have various kinds of detransitivizing morphology. For example, Tanahu Magar seems to have a middle construction, as well as a non-productive detransitivizing suffix. Athpare [122-4] has strategies for detransitivizing clauses, including an 'agent demotion' strategy, whereby the agent does not agree with the verb and is expressed in the ablative; these seem marginal to the system at this time.

**Evidentiality expressed in the verb complex, i.e. by verbals, not sentence particles:** This mode of expressing evidentiality is characteristic of the Tibetan Complex, but not of other groupings in our sample. Note that we are using the term 'evidential' in a fairly restricted way here, referring only to the opposition 'witnessed/highly reliable' versus 'not witnessed' or 'hearsay/not highly reliable', which involves a linking of speaker's certainty of the veracity of reported information with the source of the information, that is, whether witnessed or not.

All of the languages in our sample are consistent with their typological profiles except Nar-Phu and Kathmandu Newari. Nar-Phu, though a Tamangic language, is within the Tibetan cultural sphere and traditionally many speakers were bilingual in Nar-Phu and Tibetan dialects. This mode of evidentiality is central to the Nar-Phu verb system.

Kathmandu Newari evidences a 'conjunct/disjunct' system which has its origins in an evidential distinction of the sort described above. This system is not found in contemporary Dolakha Newari.

**Honorific noun and verb stems:** Honorific noun and verb stems are characteristic of the Bodish grouping within Nepal, but are not found in the other groupings.

Honorific nouns and verbs seem to be characteristic only of those Bodish-speaking groups in Nepal that adhere to the Tibetan Buddhist faith. So, among the Bodish languages in our sample, Baragaunle, Nar-Phu, and Thakali preserve honorific forms (though Thakali has only honorific verbs), while Chantyal, Gurung, and Ghale have lost them, at least in the varieties of these languages that we have investigated or have reliable information on.

**Numeral classifiers:** Numeral classifiers are entirely absent from the Bodish typology, but are present both in Himalayish and in Nepali. The Nepali classifier system is quite simple, consisting only of a human/non-human distinction; those in Himalayish languages can be considerably more complex.

The presence of a classifier system in the Bodish languages is directly connected with the preservation of native numerals: when these have been lost and replaced by Nepali numerals, as in Chantyal, the Nepali system of classifiers is usually imported along with the numerals. In Chantyal, however, the human classifier is seldom used except in very formal speech.

In Himalayish, native classifier systems are preserved in some cases and lost in others, usually matching the fate of native numerals. Limbu, however, has no classifier system and yet preserves native numerals.

**Verbal with nominal and adjectival functions:** The Bodic languages generally make use of a verbal with nominal and adjectival functions (Noonan 1997); Nepali lacks such a form. In our sample, only Tanahu Magar deviates from the Bodic norm in having innovated a specifically attributive deverbal form.

**Finite subordinate clauses:** The Bodish languages and all of Himalayish save Kiranti and Hayu do not allow finite subordinate clauses except as complements to 'say'. (Note that Chepang differs from Hayu here.) Himalayish Kiranti and Hayu and Standard Nepali allow finite subordinate clauses, though their typologies are somewhat different. (Note, however, that colloquial Nepali tends not to favor finite subordination.) All of the languages in our sample are consistent with their group's typological profiles.

**Correlative constructions:** The correlative construction that concerns us here is a complex construction formed with a relative pronoun in the first clause and a demonstrative in the second: *who believes my argument, that person will be enlightened*. The Bodish languages natively lacked this construction; it is characteristic of Nepali.

Reliable data on this construction are available for only eight languages in the sample. Of these eight, six languages have borrowed the syntagm; three—Chantyal, Athpare (Ebert 1997a:154) and Tanahu Magar—have borrowed both the construction and the *j*-class pronouns that go with it from Nepali. Hayu, Camling, and Syangja Magar use native words, though the construction otherwise follows the Nepali pattern.



Having examined all the features with regard to each of the languages in the sample, we can now provide a sort of profile for the languages. Excluding the last structural feature, for which we have incomplete data, we have the following, summarized in Table 3.

The following can be inferred from the Table 3 and the discussion that preceded it:

- 1) The Bodish languages were more different typologically from Nepali than were the Himalayish languages prior to contact. This can be seen, for example, in the relative values for +T/-N and +T/+N.
- 2) The table shows that Chantyal is the most deviant relative to the typological profile of its group. It is the only language in the sample where the number of deviations for the typological norm exceeds the number of instances of typological consistency.

Given that the values for the other Bodish languages are generally quite similar, the profile of Chantyal is rather striking. The history of Chantyal is quite special (see Noonan 1996 for discussion) and accounts for its unusual degree of divergence from the Tamangic typological norm.

- 3) The Newari dialects and the Magar dialects show the greatest number of deviations after Chantyal, and, again like Chantyal, these are primarily in the direction of convergence with Nepali—or, at least, with Indic. Most likely the influence of Indic on these languages has been considerable over a long period. The Newar dialects, particularly those in the Kathmandu Valley, have been in close contact with Nepali for more than two centuries and have had other Indic influences before that. Magar has had great exposure to Nepali over at least the last three centuries and many ethnic Magars, in particular those in the westernmost reaches of the ethnic Magar area (for example in Baglung and Myagdi), have been speaking Nepali for many generations.
- 4) Among the other Bodish languages, except for Chantyal Gurung has been most affected by Nepali. This is not surprising given the long contact between Gurungs and Nepali speakers.
- 5) Baragaunle and Nar-Phu, both spoken north of the great Himalayas in ecological (though not political) Tibet, show the least influence from Nepali. Until fairly recently, contact with Nepali speakers was not especially frequent. Chantyal apart, the Bodish languages in our sample have been affected structurally relatively little by Nepali in comparison to the Himalayish languages.
- 6) Kham and the Kiranti languages have very similar typological profiles despite the great physical distance separating them.
- 7) There are few instances overall of -T/-N, i.e. instances where these languages have changed so as to converge with a typological profile other than that of Nepali. Chantyal has likely borrowed from Magar in a few instances, and Kham has long

been in contact with Bodish languages and has borrowed from them. Hayu and Chepang, with the greatest number of -T/-N values have been in contact with Bodish (specifically Tamang) over a long period.

- 8) So, where change has occurred, it has generally been in the direction of convergence with Nepali, as seen by comparing the figures under -T/-N and -T/+N. This is hardly a surprising result, given the political and cultural situation in Nepal.

Finally, we need to evaluate the parameters themselves. In Table 4 are listed the instances of -T, that is, deviations from the typological profile for each grouping, of the thirty parameters used in our survey (again the final one is not given here due to incomplete data). From the data in Table 4 the following can be said:

- 1) Twelve out of the sixteen languages in our sample showed -T/+N values for antidative, that is, twelve of the languages acquired this feature from Nepali. Masica (1991) notes that antidative (not his term for the construction) is a comparatively new phenomenon in South Asia and has spread rapidly among Indo-Aryan languages. Note also that five languages have acquired the dative subject construction from Nepali. It is worth noting that not only have these constructions involving the dative spread from Nepali to the TB languages of Nepal, but also that they are often accompanied by the Nepali *-laai* dative, which is often borrowed along with the constructions in which it is used, for instance by Gurung and some Kiranti languages.<sup>7</sup>
- 2) Of the phonological features, some (for example, the presence of phonemic nasalized vowels) can easily arise spontaneously, so we must be careful in attributing its spread to areal influence. Other features are less likely to arise spontaneously and therefore when one finds them one has a better case for areal influence: murmur is such a feature. The  $\Lambda \sim \text{D}$  allophony is distinctive and idiosyncratic enough to be attributed to external influence.
- 3) All the Himalayish languages in our sample, save the Newari dialects, Chepang, and Camling<sup>8</sup>, have developed valence decreasing strategies, though these devices are a diverse set syntactically, resembling neither each other nor Nepali very much.
- 4) The largest number of -T/-N values for any structural feature is found with the Numeral Classifiers and is associated with Himalayish languages which have, presumably, lost their classifier systems: the Magar dialects, Chepang, and Limbu. Only Chantyal has a -T/+N value for this feature, having borrowed Nepali's minimalist classifier system along with Nepali numerals.

---

<sup>7</sup> The Bodish languages frequently have a dative in *\*la/ra*, accidentally similar to the Nepali dative. Prior to Nepali contact, all [or, at least, most] of the Kiranti languages lacked a dedicated dative altogether.

<sup>8</sup> Both Chepang and Camling have been analyzed as having inverse constructions.

I will close with two observations: first, the amount of borrowing from Nepali does not correlate directly with endangerment. Some languages, for example, Kathmandu Newari, have been in close contact with Nepali for a long period and have borrowed many features without being in grave danger of extinction. Chantyal also has survived for a long period in close contact with Nepali and, as we have seen, being massively influenced by it without succumbing to it, though some recent changes in social conditions in the Chantyal speaking villages may well result in the extinction of the language within the next generation (Noonan 1996). By contrast, the Kiranti languages have been in close contact with Nepali for a much shorter period and have borrowed much less, though many of these languages are in grave danger of extinction within the next few decades.

Second, many of the reference grammars consulted for this study tend to under-report borrowings from Nepali as these features are less interesting to Tibeto-Burmanists than native features, and linguists often strive in their grammars to describe only 'pure', uncorrupted structures where alternatives between native and borrowed structures still exist. A truer picture of the actual spoken languages would likely show greater convergence with Nepali than this study has shown.

## References

- Caughley, Ross C., 1982, *The syntax and morphology of the verb in Chepang*. Canberra: Pacific Linguistics.
- Driem, George van, 1987, *A grammar of Limbu*. Berlin: Mouton de Gruyter.
- Dryer, Matthew, 1986, Primary objects, secondary objects, and antipassive. *Language* 62/4:808-845.
- Ebert, Karen, 1994, *The structure of Kiranti languages*. Arbeiten des Seminars für Allgemeine Sprachwissenschaft, 13. Zürich: Universität Zürich.
- 1997a, *A grammar of Athpare*. München & Newcastle: Lincom Europa.
- 1997b, *Camling (Chamling)*. München & Newcastle: Lincom Europa.
- Genetti, Carol, 1994, *A descriptive and historical account of the Dolakha Newari dialect*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Georg, Stefan, 1996, *Marphatan Thakali*. München & Newcastle: Lincom Europa.
- Glover, Warren, 1974, *Sememic and grammatical structures in Gurung (Nepal)*. Norman OK: Summer Institute of Linguistics.
- Kretschmar, Monika, 1995, *Erzählungen und Dialekt am Südmustang*, vols 1-4. Bonn: VGH Wissenschaftsverlag.
- Malla, Kamal P., 1985, *The Newari language: a working outline*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

- Masica, Colin P., 1991, *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Mazaudon, Martine, 1973, *Phonologie Tamang*. Paris: Société d'études linguistiques et anthropologiques de France.
- 1978a, Consonantal mutation and tonal split in the Tamang subfamily of Tibeto-Burman. *Kailash: A Journal of Himalayan Studies* 6/3:157-179.
- 1978b, *Tibeto-Burman tonogenetics*. *Linguistics of the Tibeto-Burman Area* 3.2 (Monograph issue).
- 1993-94, Problèmes de comparatisme et de reconstruction dans quelques langues de la famille tibéto-birmane. PhD dissertation, Université Paris III.
- Michailovsky, Boyd, 1975, Notes on the Kiranti verb, 2: the origin of tone in Khaling. *Linguistics of the Tibeto-Burman Area* 2/2:205-218.
- 1988a, Phonological typology of Nepal languages. *Linguistics of the Tibeto-Burman Area* 11/2:25-50.
- 1988b, *La langue hayu*. Paris: Editions du CNRS.
- Noonan, Michael, 1991, Anti-dative shift. *Milwaukee Studies on Language* 5:50-58.
- 1996, The fall and rise and fall of the Chantyal language. *Southwest Journal of Linguistics* 15/1-2:121-136.
- 1997, Versatile nominalizations. In Joan Bybee, John Haiman and Sandra Thompson eds *Essays on language function and language type in honor of Talmy Givón*, 373-394. Amsterdam & Philadelphia: John Benjamins.
- 2003a, Chantyal. In Graham Thurgood and Randy LaPolla eds *The Sino-Tibetan languages*, 315-335. Richmond, Surrey: Routledge Curzon.
- 2003b, Nar-Phu. In Graham Thurgood and Randy LaPolla eds *The Sino-Tibetan languages*, 336-352. Richmond, Surrey: Routledge Curzon.
- Noonan, Michael, with Ram Prasad Bhulanja, Jag Man Chhantyal and William Pagliuca, 1999, *Chantyal dictionary and texts*. Berlin: Mouton de Gruyter.
- Smith, Holly, 1999, Ghale grammar. MA dissertation, University of Wisconsin-Milwaukee.
- Thompson, Chad, 1990, On the treatment of topical objects in Chepang. *Studies in Language* 14/2:405-427.
- Watters, David, 1998, The Kham language of West-Central Nepal (Takale Dialect). PhD dissertation, University of Oregon.
- Winter, Werner, 1985, *Materials toward a dictionary of Chamling: I. Chamling-English; II. English-Chamling*. Based on data collected by Dhan Prasad Rai. Preliminary version. Kiel: Linguistic Survey of Nepal.

**Table 1:** Structural features inventory by grouping

phonological feature	Bodish [Bodic]	Himalayish [Bodic]	Nepali [I-A]
phonemic voicing contrasts	no	yes	yes
tone	2- or 4-way tonal contrast	no tone	no tone
murmur	murmur concomitant with tone	absent	murmur contrastive with stops
voicing opposition in liquids and/or nasals	present	absent	absent
retroflex series	present	absent	present
fricatives	distinct alveolar & palato-alveolar series	one fricative [alveolar or palato-alveolar]	one fricative [alveolar or palato-alveolar]
affricates	distinct alveolar & palato-alveolar series	palato-alveolar series only	palato-alveolar series only
phonemic nasalized vowels	present <sup>9</sup>	absent	present
Λ ~ ɒ allophony	absent	absent	present
word-initial /ŋ/	present	present	absent
stress	relatively weak, word boundary stress: on root; may be pitch accent type	relatively weak, word boundary stress: on root	relatively weak phonemic stress

morpho-syntactic feature	Bodish [Bodic]	Himalayish [Bodic]	Nepali [I-A]
prefixes	absent [save for NEG]	present	absent
person/number marking	absent	complex [i.e. multiple arguments]	simple [i.e. subject (or absolutive) only]
reflexive	analytic; special refl word or pers pro	inflectional	analytic; special refl word
adjectival w/o	NA [AN possible]	AN	AN
demonstrative w/o	N Dem in Tibetan Complex, Dem N elsewhere	Dem N	Dem N
numeral w/o	N Num	Num N	Num N
ergative syntax	consistently	consistently	aspectually split

<sup>9</sup> Perhaps a split with the Tamangic languages having nasalised vowels and the rest not.

morpho-syntactic feature	Bodish [Bodic]	Himalayish [Bodic]	Nepali [I-A]
	ergative	ergative, or split on animacy hierarchy	ergativity
antidative syntax	absent	absent	present
dative subjects	absent	absent	present
compound case <sup>10</sup>	present	present	absent [?]
'vertical' case <sup>11</sup>	absent	present	absent
'vertical' verbs <sup>12</sup>	absent	present	absent
morphological valence increasing strategies <sup>13</sup>	absent in Tamangic; present elsewhere	present	present
morphological valence decreasing strategies	absent	absent	present
evidentiality expressed in VC by verbals	present in the Tibetan Complex, but not elsewhere	absent	absent
honorific verb & noun stems	present	absent	absent
numeral classifiers	absent	present	marginally present
verbal with nominal and adjectival functions	present	present	absent
finite subordinate clauses	absent	Kiranti & Hayu-Chepang; absent elsewhere	present
correlative constructions <sup>14</sup>	absent	absent	present

<sup>10</sup> This refers specifically to the compounding of locative case clitics creating complex expressions, e.g. Chantyal *dhiuŋ-p<sup>h</sup> yaraŋ-mar-gamsə* [tree-SUB-CIRC-ABL] 'from down around the base of the tree'.

<sup>11</sup> 'Vertical' case refers to LOC, ABL, and ALL cases which include vertical directional senses: 'up', 'down', and 'level'.

<sup>12</sup> 'Vertical' verb refers to a verb with a sense like 'come' and 'bring' which includes also a vertical dimension, i.e. 'come from above', 'bring from below', etc.

<sup>13</sup> By morphological, we mean by derivation; in all three groups there are periphrastic causative constructions.

<sup>14</sup> Complex constructions formed with a question word in the first clause and a demonstrative in the second: *who believes my argument, that person will be enlightened.*

**Table 2: Structural features inventory by language – Part 1**

1 = Bodish: 1a = Tibetan Complex, 1b = Tamangic, 1c = Ghale; 2 = Himalayish: 2a = Kham-Magar, 2b = Newari, 2c = Hayu-Chepeng, 2d = Kiranti

+T/-N = consistent with the typological profile of the grouping and not consonant with Nepali [i.e. feature has not converged with Nepali type]

+T/+N = consistent with the typological profile of the grouping and consonant with Nepali

-T/-N = inconsistent with the typological profile of the grouping but not consonant with Nepali [i.e. converging on a typological sort other than Nepali]

-T/+N = inconsistent with the typological profile of the grouping and converging with Nepali [i.e. the feature may have been borrowed from Nepali]

	Bara- gaunle: 1a	Nar- Phu: 1b	Gurung: 1b	Thakali: 1b	Chan- tyal: 1b	Ghale: 1c
<b>PHONOLOGICAL FEATURES</b>						
phonemic voicing	+T/-N	+T/-N	-T/+N	+T/-N	-T/+N	+T/-N
tone	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
murmur	+T/-N	+T/-N	+T/-N	+T/-N	-T/-N	+T/-N
nasal/liquid opp.	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	-T/+N
retroflex series	+T/+N	+T/+N	+T/+N	+T/+N	-T/-N	+T/+N
fricatives	+T/-N	+T/-N	-T/+N	-T/+N	-T/+N	+T/-N
affricates	+T/-N	+T/-N	-T/+N	+T/-N	-T/+N	+T/-N
nasalised vowels	-T/-N	+T/+N	+T/+N	-T/-N	+T/+N	+T/+N
ʌ ~ ɒ allophony	+T/-N	+T/-N	+T/-N	-T/+N	-T/+N	-T/+N
word initial /ŋ/	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
stress	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
<b>MORPHO-SYNTACTIC FEATURES</b>						
prefixes	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
person/number	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
reflexive	+T/+N	+T/-N	+T/+N	+T/+N	+T/-N	+T/+N
adjectival w/o	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
demonstrative w/o	+T/-N	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
numeral w/o	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
ergative syntax	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
antidative syntax	-T/+N	-T/+N	-T/+N	-T/+N	-T/+N	+T/-N
dative subjects	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
compound case	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
vertical case	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
vertical verbs	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
morph. val. increasing	+T/+N	+T/-N	+T/-N	+T/-N	-T/+N	+T/+N
morph. val. decreasing	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
evid. expressed in VC	+T/-N	-T/-N	+T/+N	+T/+N	+T/+N	+T/+N
honorific N&V stems	+T/-N	+T/-N	-T/+N	+T/-N	-T/+N	-T/+N
numeral classifiers	+T/-N	+T/-N	+T/-N	+T/-N	-T/+N	+T/-N
verbal as Nom & Adj	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
finite subordinate cl's	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
correlative const's		+T/-N			-T/+N	

Table 2: Structural features inventory by language – Part 2

1 = Bodish: 1a = Tibetan Complex, 1b = Tamangic, 1c = Ghale; 2 = Himalayish: 2a = Kham-Magar, 2b = Newari, 2c = Hayu-Chepang, 2d = Kiranti

+T/-N = consistent with the typological profile of the grouping and not consonant with Nepali [i.e. feature has not converged with Nepali type]

+T/+N = consistent with the typological profile of the grouping and consonant with Nepali

-T/-N = inconsistent with the typological profile of the grouping but not consonant with Nepali [i.e. converging on a typological sort other than Nepali]

-T/+N = inconsistent with the typological profile of the grouping and converging with Nepali [i.e. the feature may have been borrowed from Nepali]

	Kham: 2a	T.Magar : 2a	S.Magar : 2a	K.Newa- ri: 2b	D.Newa- ri: 2b
<b>PHONOLOGICAL FEATURES</b>					
phonemic voicing	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
tone	-T/-N	+T/+N	+T/+N	+T/+N	+T/+N
murmur	-T/-N	-T/+N	-T/+N	-T/+N	+T/-N
nasal/liquid opp.	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
retroflex series +T/- N +T/-N +T/-N +T/- N - T/+N fricatives +T/+N +T/+N +T/+N +T/+N +T/+N affricates +T/+ N +T/+N +T/+N +T/+N +T/+N nasalised vowels -T/+N +T/- N +T/-N -T/+N - T/+N $\Lambda$ ~ $\text{D}$ allophony +T/-N +T/- N -T/+N -T/+N - T/+N word initial /ŋ/ +T/-N +T/-N +T/-N	-T/+N	+T/-N		-T/+N	+T/-N
stress	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
<b>MORPHO-SYNTACTIC FEATURES</b>					
prefixes	+T/-N	+T/-N	+T/-N	-T/+N	-T/+N
person/number	+T/-N	-T/-N	-T/+N	-T/-N	-T/+N
reflexive	+T/-N	-T/+N	-T/+N	-T/+N	-T/+N
adjectival w/o	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
demonstrative w/o	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
numeral w/o	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
ergative syntax	+T/-N	-T/+N	+T/-N	+T/-N	+T/-N
antidative syntax	-T/+N	-T/+N	-T/+N	-T/+N	-T/+N
dative subjects	+T/-N	-T/+N	-T/+N	-T/+N	-T/+N
compound case	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
vertical case	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
vertical verbs	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
morph. val. increasing	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N



morph. val decreasing	-T/+N	-T/+N	-T/+N	+T/-N	+T/-N
evid. expressed in VC	+T/+N	+T/+N	+T/+N	-T/-N	+T/+N
honorific N&V stems	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
numeral classifiers	+T/+N	-T/-N	-T/-N	+T/+N	+T/+N
verbal as Nom & Adj1	+T/-N	-T/+N	-T/+N	+T/-N	+T/-N
finite subordinate cl's	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
correlative const's	+T/-N	-T/+N	-T/+N		

**Table 2: Structural features inventory by language – Part 3**

1 = Bodish: 1a = Tibetan Complex, 1b = Tamangic, 1c = Ghale; 2 = Himalayish: 2a = Kham-Magar, 2b = Newari, 2c = Hayu-Chepeng, 2d = Kiranti

+T/-N = consistent with the typological profile of the grouping and not consonant with Nepali [i.e. feature has not converged with Nepali type]

+T/+N = consistent with the typological profile of the grouping and consonant with Nepali

-T/-N = inconsistent with the typological profile of the grouping but not consonant with Nepali [i.e. converging on a typological sort other than Nepali]

-T/+N = inconsistent with the typological profile of the grouping and converging with Nepali [i.e. the feature may have been borrowed from Nepali]

	Che-pang: 2c	Hayu: 2c	Athpare : 2d	Camling : 2d	Limbu: 2d
<b>PHONOLOGICAL FEATURES</b>					
phonemic voicing	+T/+N	+T/+N	+T/+N	+T/+N	-T/-N
tone	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
murmur	-T/+N	+T/-N	-T/+N	-T/+N	+T/-N
nasal/liquid opp.	-T/-N	-T/-N	+T/+N	-T/-N	+T/+N
retroflex series	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
fricatives	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
affricates	+T/+N	-T/-N	+T/+N	+T/+N	+T/+N
nasalised vowels	+T/-N	-T/+N	+T/-N	-T/+N	+T/-N
λ ~ ɒ allophony	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
word initial /ŋ/	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
stress	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
<b>MORPHO-SYNTACTIC FEATURES</b>					
prefixes	-T/+N	-T/+N	+T/-N	+T/-N	+T/-N
person/number	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
reflexive	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
adjectival w/o	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
demonstrative w/o	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
numeral w/o	+T/+N	-T/-N	+T/+N	-T/-N	+T/+N
ergative syntax	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
antidative syntax	-T/+N	+T/-N	+T/-N	-T/+N	+T/-N
dative subjects	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
compound case	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
vertical case	-T/+N	+T/-N	-T/+N	+T/-N	-T/+N
vertical verbs	-T/+N	-T/+N	+T/-N	+T/-N	+T/-N

morph. val. increasing	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
morph. val. decreasing	+T/-N	-T/+N	-T/+N	+T/-N	-T/+N
evid. expressed in VC	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
honorific N&V stems	+T/+N	+T/+N	+T/+N	+T/+N	+T/+N
numeral classifiers	-T/-N	+T/+N	+T/+N	+T/+N	-T/-N
verbal as Nom & Adj	+T/-N	+T/-N	+T/-N	+T/-N	+T/-N
finite subordinate cl's	-T/-N	+T/+N	+T/+N	+T/+N	+T/+N
correlative const's		-T/+N	-T/+N	-T/+N	

Table 3: Feature values by language

1 = Bodish: 1a = Tibetan Complex, 1b = Tamangic, 1c = Ghale; 2 = Himalayish: 2a = Kham-Magar, 2b = Newari, 2c = Hayu-Chepeng, 2d = Kiranti

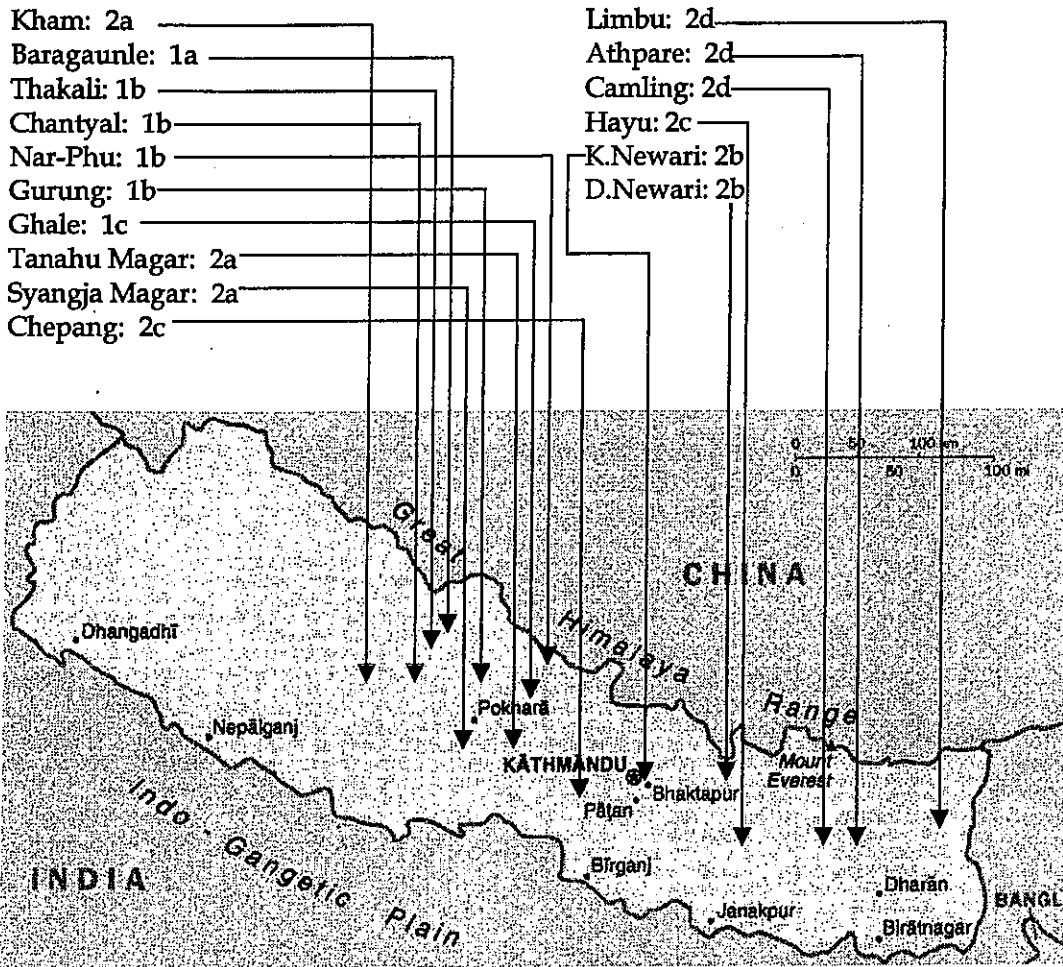
		+T/-N	+T/+N	-T/-N	-T/+N
Baragaunle	1a	22	6	1	1
Nar-Phu	1b	22	6	1	1
Gurung	1b	17	8	0	5
Thakali	1b	19	7	1	3
Chantyal	1b	7	6	2	15
Ghale	1c	18	9	0	3
Kham	2a	12	13	2	3
Tanahu Magar	2a	8	13	2	7
Syangja Magar	2a	8	13	1	8
Kathmandu Newari	2b	7	13	2	8
Dolakha Newari	2b	8	14	0	8
Chepeng	2c	12	10	3	5
Hayu	2c	13	9	3	5
Athpare	2d	14	13	0	3
Camling	2d	13	11	2	4
Limbu	2d	15	11	2	2

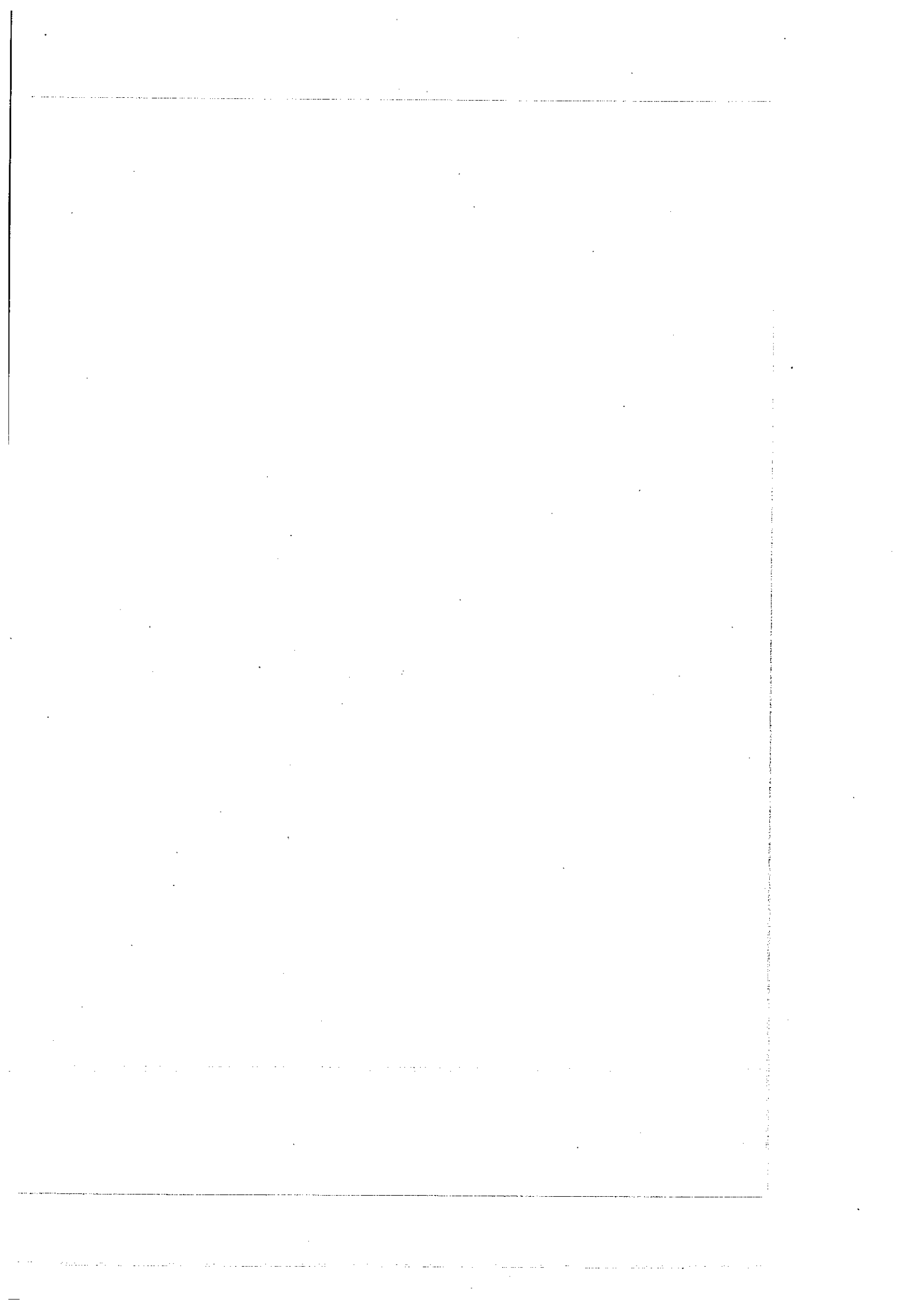
Table 4: Deviations from typological norm, by feature

	-T/-N values	-T/+N values	Total -T
<b>PHONOLOGICAL FEATURES</b>			
phonemic voicing	1	2	3
tone	1	1	2
murmur	2	6	8
nasal/liquid opp.	3	2	5
retroflex series	1	1	2
fricatives	0	3	3
affricates	1	2	3
nasalised vowels	2	5	7
ʌ ~ ɒ allophony	0	6	6
word initial /ŋ/	0	2	2
stress	0	1	1
<b>MORPHO-SYNTACTIC FEATURES</b>			
prefixes	0	4	4
person/number	2	2	4
reflexive	0	4	4
adjectival w/o	0	1	1
demonstrative w/o	0	0	0
numeral w/o	1	2	3
ergative syntax	0	1	1
antidative syntax	0	12	12
dative subjects	0	5	5
compound case	0	0	0
vertical case	0	3	3
vertical verbs	0	2	2
morph. val. increasing	0	1	1
morph. val. decreasing	0	6	6
evid. expressed in VC	2	0	2
honorific N&V stems	0	3	3
numeral classifiers	4	1	5
verbal as Nom & Adj	0	2	2
finite subordinate cl's	1	0	1

Figure 2: Map Showing Locations of Languages in Sample

1 = Bodish: 1a = Tibetan Complex, 1b = Tamangic, 1c = Ghale; 2 = Himalayish: 2a = Kham-Magar, 2b = Newari, 2c = Hayu-Chepang, 2d = Kiranti





# 6 *Prosodic tautomorphemicity in Sino-Tibetan*

BALTHASAR BICKEL

## 1 Introduction<sup>1</sup>

Sino-Tibetan is a prime example of how strongly a language family can typologically diversify under the pressure of areal spread features (Matisoff 1991, 1999). One of the manifestation of this is the average length of prosodic words. In Southeast Asia, prosodic words tend to average on one or one-and-a-half syllables. In the Himalayas, by contrast, it is not uncommon to encounter prosodic words containing five to ten syllables. The following pair of examples illustrates this.<sup>2</sup>

- (1) Hakha Lai (Kuki-Chin; W. Burma)  
( $\omega$  *na-tuk*)                      ( $\omega$  *ṅaa*)                      ( $\omega$  *l̄aay*).  
2SG.A-hit.with.stick 3PL.O                      FUTURE  
'You will hit them.'

- (2) Belhare (Kiranti; E. Nepal)  
( $\omega$  *mi-ṅṅ-u-ukg-att-u-n-chi-nn-h ak=ch a*).  
3NSG.A-NEG-roast-bring.down-PAST-3O-NEG-NSG.O-NEG-N=ADD  
'They didn't even roast it for them down here.'

<sup>1</sup> The research presented here was made possible by Grant No. 8210-053455 from the Swiss National Science Foundation. I am greatly indebted to Martha Paa (a.k.a. Kenneth Van Bik) and Bimala Pa (a.k.a. Lekh Bahadur Rai) for their help with Lai and Belhare, respectively. Parts of this paper were presented at the 29th International Conference on Sino-Tibetan Languages and Linguistics, Leiden, October 1996, and the 23rd Annual Meeting of the German Society for Linguistics, Leipzig, March 2001.

<sup>2</sup> Abbreviations are: A 'transitive subject', ADD 'additive focus', EXCL 'exclusive (of addressee)', O 'object', N 'nominalizer/focalizer', NEG 'negative', NSG 'nonsingular', PL 'plural', Q 'interrogative', SG 'singular'.

In the Southeast Asian language Lai, all but a few monomoraic prefixes license their own prosodic word (annotated here by 'ω'). The result of this in (1) is a sesquisyllabic word *natuk* 'you hit' followed by two prosodically and phonologically autonomous monosyllabic words *naa* 'them' and *lāay* 'FUTURE'. In example (2) from the Himalayan language Belhare, by contrast, no less than eight syllables are strung together into one polysynthetic word. There is but one main stress (on the first stem-containing syllable, here *ɲu*), and most morphemes undergo phonological adjustment when concatenated (Bickel 1996, 2003).

The monosyllabicity of Southeast Asian languages does not hold of **grammatical** or **lexical words**, but of **prosodic words**. Southeast Asian languages abound in bisyllabic or sometimes even multisyllabic compounds (often lexically frozen), and there is solid evidence that a sequence of morphemes such as the one in the Lai example (1), is just one grammatical word: the ordering of the object agreement marker *nhaa* and the future morpheme *lāay* is absolutely rigid, their occurrence is strictly tied to the simultaneous occurrence of a verb, and no phrasal constituent can ever intervene between morphemes. These are all properties of bound morphemes, and this contrasts systematically with the behavior of autonomous grammatical words: Lai word order is relatively unconstrained and clause constituents allow re-ordering, interruption and independent use. Note that the same is true of words such as *will* and *them* in the English translation, although they are very similar in function to their Lai equivalents: *will* and *them* can be re-ordered (cf. *will you hit them?*), interrupted (cf. *you will definitely hit all of them*), and used independently (cf. *will you?*).

However, in languages such as Lai, grammatical and prosodic structure are not independent of each other: while one grammatical word may contain several prosodic constituents (feet or syllables), there is strong bias against the opposite, that is, prosodic constituents (feet or syllables) containing more than one morpheme. I call this the 'Tautomorphemicity Principle', and the remainder of this paper is devoted to demonstrating that the effects of this principle are not only transparent in Southeast Asian Sino-Tibetan languages, but that they are also detectable in the polysynthetic and polysyllabic words of Sino-Tibetan languages in the Himalayas. There is evidence, then, that tautomorphemicity is not only an areally robust feature of Southeast Asia, but that at the same time, it is a genetically robust feature of Sino-Tibetan.

## 2 Prosodic Tautomorphemicity

The Sino-Tibetan Tautomorphemicity Principle is defined and further specified as follows:

- (3) The Tautomorphemicity Principle:  
Avoid feet or syllables containing concatenative morpheme boundaries.

The relevant prosodic constituents range over feet ('f') and syllables ('s'). The choice between these varies from language to language, but the claim here is that at least one of these two constituents disfavors internal morpheme boundaries in most Sino-Tibetan languages. The relevant boundaries are specified in (3) as concatenative (or 'agglutinative'). The principle does not apply to abstract morphemes that one might posit in accounting for tone patterns or stem alternations. Such patterns and alternations violate tautomorphemicity both necessarily and uninterestingly. Furthermore, the principle as set out in (3) is meant to capture a tendency, not a law. The precise degree and nature of its impact in any one language depends on a number of prosodic and morphophonological factors. Discussion of these is beyond the scope of this paper, but they can probably best be captured by Optimality Theoretic modelling (and see Bickel 1998 for a preliminary OT analysis of Belhare tautomorphemicity). In the following, I will put forward evidence for tautomorphemicity first on the syllable, then on the foot level.

### 2.1 Tautomorphemic syllables

The most obvious piece of evidence for tautomorphemicity on the syllable level comes from the lexical fact that most morphemes in Sino-Tibetan languages are at least one syllable long. This is generally true in Southeast Asia (Matisoff 1999), but it also holds for Belhare and other polysyllabic Sino-Tibetan languages. Lexical morphemes in Belhare all minimally contain one syllable, and the same is true for over 90 per cent of the grammatical morphemes.<sup>3</sup>

One systematic exception to this is found scattered all over Tibeto-Burman and consists of the well-reconstructed coronal augments *\*s-*, *\*-t* and *\*-s*. Interestingly, however, more often than not these subsyllabic augments have lost their concatenative character and have instead developed into processes of phonation, tonation, or aspiration. Prefix *\*s-* is well-known to have often developed into aspiration (for example, Lai *pɪt* 'be blocked' ~ *ḥpɪt* 'block', Belhare *pok-* 'rise' ~ *pʰok-* 'raise'), and the suffixes *\*-t* and *\*-s* have often developed into phonation (for example, Lai *fiaŋ* 'be clear' ~ *fiaŋ̃* 'make clear'; Peterson 1998).<sup>4</sup> A similar development is found with the Written Tibetan ergative *-s* which has developed into an umlaut and phonation alternation (for example, WT *kho-s* 's/he-ERGATIVE' = /kʰø̃/, cf. absolutive /kʰõ/; Tournadre 1996). Through such developments, morphemes lose their concatenativity and become hidden inside the stem. There, they are shielded off against the effects of the Tautomorphemicity Principle.

<sup>3</sup> Among about 80 such morphemes, there are only 3 clearcut exceptions: *-ŋ* '1SG.A', *-m* '1/2 PL.A', and *-t* or *-ʔ* 'NONPAST'. Another exception is limited to one allomorph of a morpheme, viz. the postvocalic allomorph *-n* of the negation suffix *-ni*.

<sup>4</sup> On *\*-t* in Belhare and other Kiranti languages, see below.



Another systematic threat to tautomorphemicity arises from concatenating CVC with V-initial syllables. Simple syllabification of this would result in strings like  $(\sigma CV)(\sigma C-V)$ , where the second syllable violates the Tautomorphemicity Principle. There seem to be two responses to this: onset-free syllabification and onset prothesis.

Onset-free syllabification can be exemplified by Garo (NE India), where the habitual tense suffix *-a* is always parsed as syllable-initial. Thus, *caʔ-a* 'eat-HABITUAL' is syllabified as  $(\sigma caʔ)-(\sigma a)$ , not as  $*(\sigma ca)(\sigma ʔ-a)$ . Likewise, *kat-a* 'go-HABITUAL' is syllabified as  $(\sigma kat)-(\sigma a)$ , not as  $*(\sigma ka)(\sigma t-a)$ . Structural evidence for this syllabification is (i) that there are no glottal syllable onsets in Garo, and (ii) that the /t/ in  $(\sigma kat)-(\sigma a)$  does not undergo aspiration, a process limited to syllable initials in Garo (Burling 1961:5).

Onset-free syllabification is also found in the form of bisyllabic parsing of underlying vowel sequences. This is well attested in Himalayan languages. For example, in Dolakha Newar, a string like *ye-e* 'come-N' is realised bisyllabically, rather than as one syllable with a long vowel (Genetti 1994:30). Another instance is Belhare, where a sequence like *yu-a* 'go.down-IMPERATIVE' is realised by two distinct syllables rather than by one (raising) diphthong. The same is exemplified by the sequence *u-uk* 'roast-bring.down' in (2) above, which is syllabified into two nuclear peaks (on initial *ŋ*, see below).

The other response to CVC-V strings is onset prothesis: an underspecified onset (notated in the following by bold font) is inserted at the morpheme boundary. This segment is then realised together with the preceding consonant as an ambisyllabic geminate. Thus, CVC-V turns into  $(\sigma CVC)-(\sigma CV)$ , with an ambisyllabic CC geminate.

The result of this is that the underlying morpheme boundary comes to lie between rather than within syllables, as required by the Tautomorphemicity Principle. This process can be exemplified by Meithei (also known as Manipuri; NE India), where onset prothesis turns forms like *tʰəm-u* 'keep-IMPERATIVE' into  $(\sigma tʰəm)-(\sigma mu)$ , or *təw-e* 'do-ASSERTIVE' into  $(\sigma təw)-(\sigma we)$  (Chelliah 1997:23, 67). Onset prothesis is also sometimes recruited as a solution for hiatus (that is, V-V) concatenations, which, too, are a potential threat to tautomorphemicity. In this case, the prosthetic onset is typically a glide. An example of this are Belhare forms like  $(\sigma so)-(\sigma yu)$  'wait-3SG.O' (that is., 'wait for him/her!'), rather than  $*(\sigma sou)$ , or  $(\sigma tu)-(\sigma yu)$  'dig-3SG.O' (i.e., 'dig it!'), rather than  $*(\sigma tuu)$  or  $*(\sigma tu:)$ . In Belhare, the environments triggering glide insertion are lexically restricted. The default response to hiatus in this language is onset-free syllabification, as observed above with examples like  $(\sigma yu)-(\sigma a)$  'go down!'.

## 2.2 Tautomorphemic feet

In many Sino-Tibetan languages, tautomorphemicity holds of the basic stress unit, the foot (either in addition to or in stead of tautomorphemicity of syllables). In Southeast Asia, foot tautomorphemicity is often trivially satisfied by having a one-to-one

correspondence between syllables, feet, and prosodic words. In languages such as Lahu, each morpheme is realised by a syllable that is a self-contained and equally stressed foot and indeed, word (Matisoff 1999). Foot tautomorphemicity is much less obvious in the Himalayas, but, as I argue in the following, there is one perspicuous morphophonological alternation that is best explained as an effect of just this principle.

We observed above that syllabic tautomorphemicity is sometimes ensured by means of onset prothesis. In Belhare this is found, as we saw, with CV-V sequences. Apart from this environment, onset prothesis also occurs in Belhare with CVC-VC sequences. This is illustrated by expressions such as *lap-uk* 'catch-bring.down', which is realised with a prothetic voiced consonant as  $(\sigma \text{ lap})-(\sigma \text{ buk})$ . Unlike unrepaired  $*(\sigma \text{ la})(\sigma \text{ b-uk})$ , the form with onset prothesis complies with the Tautomorphemicity Principle.<sup>5</sup> An example from the Maivā-Mevā dialect of Limbu (Michailovsky 1986) is *huk-en* 'hand-ARTICLE', which is realised as  $(\sigma \text{ huk})-(\sigma \text{ ken})$  'the hand', rather than as  $*(\sigma \text{ hu})(\sigma \text{ k-en})$ . Again, onset prothesis makes the syllables tautomorphemic.

Onset prothesis is not found, however, in CVC-V sequences, although they are as much a threat to syllabic tautomorphemicity as CV-V and CVC-VC sequences; for instance, Belhare *lap-u* 'catch-30' (that is, 'catch it!') is realised as  $(\sigma \text{ la})(\sigma \text{ b-u})$ , not as  $*(\sigma \text{ lap})-(\sigma \text{ bu})$ . In Belhare, only -VC morphemes trigger prothesis; heteromorphemic syllables resulting from suffixing -V remain unrepaired. Unlike Belhare, Limbu has a phonemic vowel length opposition, and in this language, long-vowelled suffixes trigger prothesis as well: *ke-sira-thaŋ-i* '2-pleasure-come.up-Q' is realised as  $(\sigma \text{ kɛ})-(\sigma \text{ si})(\sigma \text{ ra})-(\sigma \text{ dhaŋ})-(\sigma \text{ ŋi})$  'do you like it?' (Phedāppe dialect; van Driem 1987:144). This contrasts with short-vowelled suffixes: *ke-ni-siŋ-i* '2-see-REFLEXIVE-2PL' is realised as  $(\sigma \text{ kɛ})-(\sigma \text{ ni})-(\sigma \text{ si})-(\sigma \text{ ŋ-i})$  'you saw each other' (Phedāppe dialect; van Driem 1987:384).

Thus, onset prothesis is generally restricted to -VC or -VŪ morphemes.<sup>6</sup> These morphemes carry both two moras, and this in turn is criterial for feet in the Belhare and Limbu stress system. Stress in these languages follows a progressive trochaic pattern starting from the first stem mora. That the relevant feet are defined by bimoraicity, rather than bisyllabicity is evidenced by Belhare examples like *labhokarik* 'kind of small bird' which has secondary stress on  $(\phi \text{ rik})$ :  $(\phi \text{ 'labho})ka(\phi \text{ rik})$ . If the stress system were syllable-counting rather than mora-counting, we would expect secondary stress to fall on

<sup>5</sup> I have no explanation for the voicing, except perhaps that impressionistically it yields a more pronounced demarcation of the second syllable, and as we will shortly see, stress unit. Place and manner features spread from the left.

<sup>6</sup> There are three apparent exceptions in Maivā-Mevā Limbu: the negative imperative suffix -*z*, the interrogative -*i* and the vocative -*e*. Although short-vowelled, these suffixes trigger prothesis. A likely explanation of this is that the behavior of these suffixes reflects an earlier VC or VŪ structure, as is evidenced by their Phedāppe cognates -*z*?, -*i*? and -*e*?, respectively (*e*? is redundantly long; there is no short /e/ in Phedāppe).

\*( $\phi$  karik): \*( $\phi$  'lab<sup>h</sup>o)( $\phi$  karik). As the foot bracketing suggests, Belhare bans degenerate feet: the light syllable *ka* in this example remains unfooted and thereby unstressed (and indeed typically has a reduced realisation). The only exception to the ban on degenerate feet is monomoraic syllables under word-initial (that is, main) stress: ( $\phi$  'mi) 'fire', or ( $\phi$  'sa)(met). The Limbu stress system seems to be similar in these regards.

What this suggests is that the driving force behind onset prothesis in Limbu and Belhare is not tautomorphemicity of syllables but rather tautomorphemicity of the basic foot unit that underlies the stress system: ( $\sigma$  CVC)-( $\sigma$  CVC)- and ( $\sigma$  CVC)-( $\sigma$  CV:)- define sequences of two tautomorphemic bimoraic feet, that is, ( $\phi$  CVC)-( $\phi$  CVC)- and ( $\phi$  CVC)-( $\phi$  CV:)-, respectively. By contrast, onset prothesis in CVC-V would create a monomoraic CV syllable unable to sustain a foot of its own. In brief, **onset prothesis occurs just in case a tautomorphemic foot can be obtained.**

As a consequence of this, onset prothesis is absent not only from CVC-V sequences but also from many other patterns. Consider, for example, a CVC-V-C sequence. Such a sequence has the same number of skeletal positions as the sequence that triggers onset prothesis (that is, CVC-VC). Yet, in a CVC-V-C sequence, onset prothesis cannot possibly create tautomorphemic feet. Thus, strings like *lap-u-k=cha* 'catch-3O-2A=ADD' (that is, 'although you caught it') remain unrepaired, for \*( $\phi$  'lap)( $\phi$  b-u-k)=cha violates foot tautomorphemicity as much as \*( $\phi$  'la) ( $\phi$  b-u-k)=cha. In the absence of segmental proliferation techniques other than onset prothesis, the only way to comply with foot tautomorphemicity in the case of CVC-V and CVC-V-C sequences is prosodic underparsing of syllables. Thus, forms like *lap-u* 'catch-3O' and *lap-u-k=cha* 'catch-3O-2A=ADD' are parsed prosodically as ( $\phi$  'la)b-u and ( $\phi$  'la)b-u-k=cha, respectively. ( $\phi$  'la) is a degenerate foot, but it occurs in word-initial position, that is, the one position where Belhare tolerates degenerate feet (cf. above).

The result of underparsing is that there is no secondary stress in these examples. Suffix sequences show secondary stress only where foot structure is tautomorphemic. This is possible only with bimoraic suffixes, that is, with -CVC (4a) or -VC (4b, repeated from 2 above) suffixes. Note that as long as the suffix is bimoraic it does not matter where in the suffix string the secondary stress comes to lie:

(4) Belhare

- a. ( $\phi$  'ta)r-u-chi-k-( $\phi$  k<sup>h</sup>ak)=cha  
*tar-u-chi-k-k<sup>h</sup>ak=cha*  
 bring-3O-NSG.O-2A-N=ADD  
 'that you brought it to them as well'
- b. mi- $\eta$ ( $\phi$  'ju)-( $\phi$  uk)( $\phi$  ,gat)t-u-n-chi-n( $\phi$  n<sup>h</sup>ak)=cha  
*mi- $\eta$ -u-uk-att-u-n-chi-n-hak=cha*  
 3NSG.A-NEG-roast-bring.down-PAST-3O-NEG-NEG.O-NEG-N=ADD  
 'They didn't even roast it for them down here.'

Examples like these confirm the observation that (apart from the anti-hiatus cases mentioned in Section 2.1) onset prothesis in Belhare occurs in all and only those environments where the result is a tautomorphemic foot. This includes not only suffixal environments, but, as (4b) shows, also prefixal environments: in (4b), onset prothesis dissociates the stem-initial (and thus exceptionally degenerate) foot ( $\eta u$ ) from the negative prefix  $\eta$ .

Table 1 summarises the possible morpheme sequences of Belhare after CV and CVC stems. I leave out sequences with -CV and -CVC morphemes: -CVC morphemes license tautomorphemic feet to begin with; and -CV morphemes can easily be reconciled with the Tautomorphemic foot Principle by prosodic underparsing (non-stressing) of the kind noted just above. Note that in Belhare single C suffixes never follow a stem alone, without further affixes.

Table 1: Overview of morpheme combinations in Belhare

	-V(-C)	-VC	-C-V	-C-VC
CV	$\sigma$ : prothesis or onset-free syllabification $\phi$ : $\sigma$ -underparsing $(\phi (\sigma CV))-(\sigma CV)$ <i>or</i> $(\phi (\sigma CV))-(\sigma V)$	$\sigma$ : onset-free syllabification $\phi$ : well-formed $(\phi (\sigma CV))-(\phi (\sigma VC))$	<i>s</i> : unrepaired <i>f</i> : <i>s</i> -underparsing $(\phi (\sigma CV))-(\sigma C-V)$	$\sigma$ : C-deletion $\phi$ : C-deletion $(\phi (\sigma CV))-C-(\phi (\sigma VC))$
CVC	$\sigma$ : unrepaired $\phi$ $\sigma$ -underparsing $(\phi (\sigma CV))(\sigma C-V)$	$\sigma$ : prothesis $\phi$ $\sigma$ -prothesis $(\phi (\sigma CVC))-(\phi (\sigma CVC))$	<i>s</i> : unrepaired <i>f</i> : <i>s</i> -underparsing $(\phi (\sigma CVC))(\sigma C-V)$	<i>s</i> : C-deletion <i>f</i> : prothesis $(\phi (\sigma CVC))-C-(\phi (\sigma CVC))$

**Key:**

- first row in each cell: processes triggered by syllabic tautomorphemicity ('s')
- second row in each cell: processes triggered by foot tautomorphemicity ('f')
- third row in each cell: result

The first two columns in Table 1 contrast the two patterns discussed earlier: CV-V sequences are brought in compliance with syllabic tautomorphemicity through either onset prothesis or onset-free syllabification, depending on what is lexically licensed by a given morpheme; by contrast, syllabic tautomorphemicity is not enforced with CVC-V sequences. But like CV-V, CVC-V is brought into agreement with foot tautomorphemicity. This is achieved through prosodic underparsing, that is, non-stressing.

The second column illustrates V-initial bimoraic suffixes, which have the potential of projecting their own tautomorphemic foot. After CV morphemes, syllabic tautomorphemicity is achieved through onset-free syllabification (in the case of CV-VC). Foot tautomorphemicity is automatically ensured in this case. By contrast, CVC-VC sequences need onset prothesis in order to result in tautomorphemic feet. The resulting

structure, CVC-CVC also satisfies syllabic tautomorphemicity but this is an epiphenomenon rather than the underlying reason because, as the first column of Table 1 shows, violations of syllabic tautomorphemicity alone, as in the case of CVC-V sequences, is no reason for onset prothesis.

The next two columns illustrate single C suffixes. Like before, the phonological result depends on what follows (and there is always a segment following single C suffixes in Belhare). If a single C suffix is followed by a monomoraic suffix, syllabic tautomorphemicity is left violated. Thus, like (s CV)(s C-V) in the first column, both (σ CV)-(σ C-V) and (σ CVC)-(σ C-V) in the third column contain heteromorphemic syllables. Thus, just like before, violations of syllabic tautomorphemicity alone are not a sufficient reason for phonological repair. Examples of this involve the coronal augments -t and -s mentioned in §2.1: a string *hi-t-e* 'be.able-AUGMENT-PAST' is realised as (σ hi)(σ r-e) 's/he was able' (/t/ is regularly voiced and flapped between vowels). Foot tautomorphemicity is ensured, however, just like before, by underparsing: (ϕ (σ 'hi))(σ r-e). The same pattern is found with -C-V sequences after CVC morphemes: *mat-t-u* 'narrate-T.AUGMENT-3O' is realised as (ϕ (σ mat))(σ t-u) 'tell the story!', with a heteromorphemic syllable but only a strictly tautomorphemic foot.

As for single C suffixes followed by -VC (last column in Table 1), the Tautomorphemicity Principle is complied with by means of C-deletion. The only environment for this again involves the coronal augment. (No other single C morpheme is ever followed by -VC.) In examples like the following, deleting the augment -t ensures tautomorphemicity of both syllables and feet:

- (5) Belhare  
 n-(ϕ 'lu)-(ϕ at)t-u-n  
*n-lu-t-att-u-n*  
 NEG-tell-T.AUGMENT-PAST-3O-NEG  
 's/he didn't tell him/her.'

Incidentally, augment-deletion is also found before C-initial suffixes; the reason is again tautomorphemicity; cf. (ϕ 'hi)-ma 'to be able to' from *hi-t-ma* 'be.able-T.AUGMENT-INFINITIVE'.<sup>7</sup>

If the augment-VC sequence follows a CVC stem, as in *m-mat-t-att-u-n* 'NEG-narrate-T.AUGMENT-PAST-3O-NEG', C-deletion ensures tautomorphemicity of syllables, but at the same time it creates a sequence \**mat-att*, which in turn corresponds precisely to the triggering environment of onset prothesis. And onset prothesis is indeed what one finds here:

<sup>7</sup> The classical analysis in Kiranti studies is that augments delete before C. This does not account for (5).

## (6) Belhare

m-(*'mat*)-(dat)t-u-n*m-mat-t-att-u-n*

NEG-narrate-T.AUGMENT-PAST-3O-NEG

's/he didn't narrate it'

The result again confirms the observation that onset prothesis is found in all and only instances where a tautomorphemic foot can be created.

Other accounts of onset prothesis are conceivable—for example, by positing empty initial C-slots in -VC and -V: morphemes that are realised under certain lexical conditions; or by assuming that geminate-producing prothesis is triggered by morphemes that are prespecified as ending in syllable codas (Michailovsky 1986)—but all these alternatives miss the generalisation that the prothesis effect correlates with the distribution of stress and are therefore sensitive to foot structure.<sup>8</sup>

As noted above, in many Sino-Tibetan languages foot tautomorphemicity is satisfied trivially by having a one-to-one correspondence between morpheme, syllable, foot and word; the principle does not require in this case any of the complex mechanisms found in polysynthetic languages like Belhare and Limbu. However, under one analysis, Southeast Asian languages recruit syllable underparsing as a means to comply with foot tautomorphemicity. Like in other Southeast Asian languages, the only deviation from strict monosyllabism in Lai arises from prefixes like *na-*, as in *na-tuk* '2SG.A-hit.with.stick' in (1) above. Sequences like *natuk* are potential threats to foot tautomorphemicity since they can be analyzed in principle as unitary iambic feet. However, the fact that these prefixes are never stressed nor tone-bearing, and that they often have reduced vocalism (Matisoff 1999; cf. Chelliah 1997:38 on Meithei) suggests that these prefixes are not footed at all, that is, that the foot structure is *na-( $\phi$  tuk)* rather than ( $\phi$  *na-tuk*). If this is correct, these languages would indeed recruit the same underparsing technique as found in Belhare and Limbu above.

### 3 Conclusions

The analyses presented in this chapter suggest that a seemingly disparate number of morphophonological alternations, stress rules, and historical developments receives a unified explanation once prosodic tautomorphemicity is acknowledged as an underlying principle of many Sino-Tibetan languages. The principle turns out to be not only an areally robust feature of Southeast Asia, but that at the same time, it is also found in a branch of Sino-Tibetan that is typologically and geographically far away from Southeast Asia: the Kiranti languages of the Himalayas. This suggests that tautomorphemicity is not only an areally but perhaps also a genetically robust feature of Sino-Tibetan: it not only

<sup>8</sup> For a more detailed critique of alternative accounts, see Bickel (1998).

spreads during language contact and language shift, but it also appears to survive language splits and family branching.

However, while prosodic tautomorphemicity shows strong areality in Southeast Asia, the principle seems to resist any convergence pressure on those Sino-Tibetan languages that are in intense contact with Indo-European in the Himalayas. Indo-European languages do not show any trace of tautomorphemicity; morphemes appear in any size, from single consonants to bisyllabic feet, and there does not seem to be any systematic technique for aligning syllable or feet boundaries with morpheme boundaries of the kind found in Limbu or Belhare. In turn, languages like Limbu and Belhare keep prosodic tautomorphemicity despite very intense languages contact in the form of large-scale bilingualism and code-switching with Nepali.

A reason for this diminished diffusibility in the Himalayas is perhaps that unlike in Southeast Asia, the effects of tautomorphemicity in Limbu and Belhare are very indirect: unlike in the case of the one-to-one correspondence of morphemes, syllables and feet in Southeast Asia, Himalayan tautomorphemicity manifests itself only indirectly through its effects on onset prothesis, phonation and stress rules. Unlike in Southeast Asia, there is no single and unitary tautomorphemic surface pattern as it were. It has sometimes been argued that the absence of such direct surface patterns makes linguistic features less transparent, and that this in turn may lower their areal diffusion potential (Heath 1978, Johanson 1992, Bickel 1999). Prosodic tautomorphemicity could be an instance of just such a feature.

If further research confirms this, the prosodic Tautomorphemicity Principle may prove to be a genetically highly stable feature of Sino-Tibetan — indeed, a genetic signature of the whole family. High diffusibility would then be expected only in the presence of monosyllabism because this considerably increases the surface transparency of the principle.

## References

- Bickel, Balthasar, 1996, *Aspect, mood, and time in Belhare*. Zürich: ASAS Press.
- 1998, Rhythm and feet in Belhare morphology. Working Paper No.287, Rutgers. Optimality Archive; <http://roa.rutgers.edu>.
- 1999, Grammatical relations, agreement, and genetic stability. Ms., University of California, Berkeley; <http://www.spw.unizh.ch/mitarbeiter/bickel/papers>.
- 2003, Belhare. In Graham Thurgood and Randy J. LaPolla eds *The Sino-Tibetan languages*, 546-570. London: Routledge.
- Burling, Robbins, 1961, *A Garo grammar*. Poona: Linguistic Society of India.
- Chelliah, Shobhana L., 1997, *A grammar of Meithei*. Berlin: Mouton de Gruyter.

- Driem, George van, 1987, *A grammar of Limbu*. Berlin: Mouton de Gruyter.
- Genetti, Carol, 1994, *A descriptive and historical account of the Dolakha Newari dialect*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Heath, Jeffrey, 1978, *Linguistic diffusion in Arnhem Land*. Canberra: Australian Institute of Aboriginal Studies.
- Johanson, Lars, 1992, *Strukturelle Faktoren in türkischen Sprachkontakten*. Stuttgart: Steiner.
- Matisoff, James A., 1991, Sino-Tibetan linguistics: present state and future prospects. *Annual Review of Anthropology* 20:469–504.
- 1999, Genetic vs. areal linguistics in Southeast Asia: prosodic diffusibility in Southeast Asian languages. Ms., University of California, Berkeley.
- Michailovsky, Boyd, 1986, Structure syllabique et variation combinatoire: voisement et gémination en limbu. *Cahiers de Linguistique Asie Orientale* 15:193–204.
- Peterson, David A., 1998, The morphosyntax of transitivity: in Lai (Haka Chin). *Linguistics of the Tibeto-Burman Area* 21:87–153.
- Tournadre, Nicolas, 1996, *L'érgativité en tibétain*. Louvain: Peeters.



-----

-----

-----

# 7 *Ergativity in Thulung Rai: a shift in the position of pronominal split*<sup>1</sup>

AIMÉE LAHAUSOIS

Thulung Rai is a Tibeto-Burman language spoken in eastern Nepal, in the Solu Khumbu district. It is subgrouped with the Kiranti languages, all of which are spoken roughly in the same geographical area, based on its complex verbal morphology, marking both subject and object agreement with verbal suffixes. Thulung is endangered (my estimate is 1000 speakers in the year 2000), with no fluent speakers younger than twenty, and has been subject to the inroads of Nepali, the Indo-Aryan national language of Nepal.

I first became interested in Thulung because of an attempt of Jim's to have it as the target language for a field methods class at Berkeley. He had found, to everyone's amazement, a Thulung speaker in the Bay Area, and one who was willing to be our consultant at that. It quickly became obvious, in the very first session, that the consultant, while Thulung, did not fit the definition of speaker, and he struggled even with basic vocabulary. I had been studying Nepali at the time, and recognised some of the words he was giving as Nepali forms. We eventually switched to a different field methods language, but I remained intrigued by this elusive language, which was so clearly losing out to Nepali. As a result of this brief exposure, I ended up in Nepal working on Thulung with probably the last generation of native speakers, and encountering throughout the language

<sup>1</sup> Glosses used are the following: DAT=dative; ERG=ergative; NOM.rel=relativiser; NOM.inf=infinitiviser; OBL=obligation; PURP=purposive. Other abbreviations used are as follows: for all persons, s=singular, D=dual, P=plural, I=inclusive, E=exclusive. POL. preceding person/number indicates a polite form, where relevant; the absence of POL. indicates a default casual form. Transitive verb agreement suffixes show the agent first, separated from the patient by /. Thus hit-1s/3P indicates a 1s agent acting on a 3P patient. PST following the person combinations in the verbal suffixes indicates a past form of the verb. The non-past is considered the default and is unmarked in the gloss.

evidence of the intense influence of Nepali. The issue described in this paper is the result of such influence, and it was thus through Jim, in a roundabout way, that I came across this linguistic puzzle.

The issue at hand is an unusually positioned split conditioning morphological ergative marking in Thulung. This pattern became apparent when I did research in Nepal<sup>2</sup>, but when I looked at the work of Allen on Thulung in the 1970s, I found that a different distribution of case markers was present at that time. I will first describe transitivity patterns in the language, in so far as they relate to ergative marking or the lack thereof, first laying out the present pattern and then contrasting it with that described by Allen (1975). Because of the pronominal nature of this ergative split, as is the case with most Kiranti languages, I then discuss the pronominal systems of modern and older<sup>3</sup> Thulung, comparing these systems to that of co-territorial Nepali (which is Indo-Aryan). This appears to provide the key to the puzzle, as the Nepali pronoun system reveals itself to be the model for change in the Thulung system, with morphosyntactic repercussions as far as ergative marking is concerned. I also discuss the typology of pronominal ergative splits in so far as they relate to person and animacy hierarchies, and show that the shift seen in Thulung, although motivated by a contact phenomenon, results in a typologically consistent pattern.

## 1 Transitivity marking in modern Thulung

### 1.1 A marking<sup>4</sup>

Thulung exhibits a split system of morphological ergativity, with the split conditioned by person. In other words, the actor in a transitive clause will, for the appropriate persons,

---

<sup>2</sup> My data comes from ten months of field research in Nepal in 1999-2000, under the auspices of the Fulbright Commission, with additional funding from a UC Humanities and Social Sciences Research grant. My main informant was a 25-year old female, and data from her is from elicitation sessions. The bulk of my data is on the Mukhli dialect, and comes from analysis of narrative, collected in Mukli from 12 different speakers, aged 25-67. I first detected the ergative pattern I describe in the paper in the speech of the main informant, and noticed that in the stories I collected from speakers of a fairly wide age-range, the ergative was used consistently with the 'new' pattern. Allen seems to have worked with three main informants, two of whom were schoolmasters (1975:7). It could be that the data these informants provided was more conservative and perhaps even prescriptive, rather than reflecting the speech of most speakers at the time. This might explain how the change happened over the course of only 30 years.

<sup>3</sup> I will henceforth use these two labels to refer to the data I collected, in 1999-2000, and to Allen's data, from the 1970s, respectively.

<sup>4</sup> In discussing the assignment of grammatical markers, I use the semantico-syntactic primitives S, A and O, where S is the only argument in an intransitive clause, A is the most agent-like argument in a transitive clause, and O is the least agent-like argument in a transitive clause. These are the terms used by Dixon (1994).

receive ergative case marking, whereas the single argument of an intransitive clause will receive (zero) nominative marking. In my research on modern Thulung<sup>5</sup>, I found the ergative case marker (*-ka*) to be applied to agents when these were a second plural<sup>6</sup> or 'lower' (the notion of position in a hierarchy is taken from Silverstein 1976). The appearance of the ergative marker is entirely based on the person/number of the agent, and has nothing to do with tense. The following transitive sentences illustrate the distribution of the ergative marker for a range of pronouns:

No ergative marker *-ka* on agent:

1S AGENT, 3S PATIENT

- (1) *go mag qou-to*  
 1S mug drop-1S/3S.PST  
 'I dropped the mug.'

1PI AGENT, 3S PATIENT

- (2) *gui pe-pa fhal sul-mu basi*  
 1PL eat-NPST.PRT dish wash-NOM.INF OBL  
 'We must wash the dishes.'

2D AGENT, 3S PATIENT

- (3) *gatsi mam-lai krum-da lo-mu basi*  
 2D mother-DAT visit-PURP go-NOM.INF OBL  
 'You two must go visit mother.'

Ergative marker *-ka* present on agent:

2P AGENT, 3S PATIENT

- (4) *ganimim-ka go-lai jal-nini*  
 2P-ERG 1S-DAT hit-2P/1S  
 'You hit me.'

3S AGENT, 3S PATIENT

- (5) *gu-ka thulu-lwa si-mu basi*  
 3S-ERG Thulung-language learn-NOM.INF OBL  
 'She must learn Thulung.'

3P AGENT, 3S PATIENT

- (6) *gumimim-ka kam be-mri*  
 3P-ERG work do-3P/3S.PST  
 'They do work.'

<sup>5</sup> This is the same dialect which Allen used for his data in the 1970s. Mukli is considered the homeland of the Thulung people (and is where the king, when there was one, had his residence). It has the largest proportion of Thulung in its population, compared to other villages.

<sup>6</sup> But not second person dual, as we see from the examples. I use plural and dual as exclusive throughout this paper, as Thulung always distinguishes the two.

The above examples show that the pronominal split determining the presence of the ergative marker occurs between the second person dual and the second person plural. This is a surprising position for a pronominal split, as most other languages reveal a pattern whereby the entire person patterns as a unit, rather than showing divisions within a person, based on number.

## 1.2 O marking

In addition to nominative (marked  $\emptyset$ ) and ergative (marked *-ka*), Thulung also has a category called 'primary object'. This is the syntactic function proposed by Dryer (1986) to cover the situation where the indirect object of a ditransitive and the direct object of a monotransitive receive the same morphological marking. In the case of Thulung, the markers in such situations are *-lai*, most probably borrowed from Nepali (Allen 1975:92), and verb agreement.

The primary object marker is seen in monotransitive sentences:

- (7) *go mam-lai tsum dwak-pu*  
 1s mother-DAT much like-1s/3s  
 'I like my mother a lot.'
- (8) *i-lwak-ka i-mam-lai khhui*  
 2POSS-y.sibling-ERG 2POSS-mother-DAT help.3s/3s  
 'Your younger sibling helps your mother.'

For ditransitive sentences, it is the recipient which is marked with *-lai*, while the patient does not get overtly marked.

- (9) *mam-ka u-tsu-tsi-lai pomuʃhok gwak-ty*  
 mother-ERG 3POSS-children-DUAL-DAT food give-3S/3S.PST  
 'The mother gave her two children food.'
- (10) *go a-mam-lai tsutsu gwak-tomi*  
 1s 1POSS-mother-DAT child give-1S/POL.3S.PST  
 'I gave the child to my mother.'

The absence of marking on the patient in ditransitives can be explained in terms of the avoidance of potential ambiguity. There is a much greater possibility of the (animate, usually human) recipient being an agent-like participant than the (usually inanimate) patient. In other words, the marking on the recipient serves to diminish the chances of confusing it with the agent, by tagging it as clearly un-agentive.

It is also important to note that there is an animacy constraint on the PO marking in Thulung, as exemplified below.

O is marked with *-lai* when human:

(11) *gu-ka*      *gana-lai*      *jal-na*  
 3S-ERG      2s-DAT<sup>7</sup>      hit-3S/2S  
 'He hits you.'

(12) \**gu-ka* *gana* *jal-na*

A non-human animate can be either marked or unmarked:

(13) *gu-ka*      *khlea(-lai)*      *jal-y*  
 3S-ERG      dog(-DAT)      hit-3S/3S  
 'He hits the dog.'

Inanimate objects are unmarked :

(14) *gu-ka*      *gari(-\*lai)*      *thuur-y*  
 3S-ERG      car      drive-3S/3S  
 'He drives the car.'

La Polla (1992, 1994) argues that for Tibeto-Burman languages, a more appropriate term to cover the similar marking of recipient and patient is 'anti-ergative'. While Dryer's 'primary object'/'secondary object' pair describe syntactic functions, LaPolla (1992:5) believes the marking to be based in semantics and 'on an actor vs non-actor contrast, not on an object vs. non-object contrast.' It is indeed significant that the marker has an animacy constraint in Thulung, as in other Tibeto-Burman languages, as this is an indication that semantics are involved in the assignment of the markers. Dryer (1986:842) suggests that 'something along the lines of a human/non-human distinction is a likely diachronic source for primary objectivity'. LaPolla points out (1992:7) that 'primary object' marking is much more grammaticalised in languages of Nepal than others in the Tibeto-Burman family, as a result presumably of contact with Nepali (which is PO marking), from which some of the languages have borrowed the actual marker.<sup>8</sup>

## 2 Transitivity marking in older Thulung

Allen too notes a suffix *-ka*, marking agents of transitive sentences, but the distribution is different from that which I found. His data largely points to a clear pronominal split, with first and second person being in the nominative case, and third person and common NPs following an ergative pattern. He reports a few counterexamples to the pattern

<sup>7</sup> The choice of the term 'dative' follows the suggestion of Balthasar Bickel, as per his and Nichols' description of the term as 'sometimes used for primary objects' (2001:93).

<sup>8</sup> This is the case for Thulung, but also for Kham (*lay*) among others.

though, where (in Mukli dialect) ergative marking occurs on first and second persons (1975:93).<sup>9</sup>

We have seen how the marker *-ka* functions in older Thulung. The other transitive marker Allen mentions is *-la*<sup>10</sup> (1975:94). He says it is a Nepali suffix, used to mark direct or indirect objects when these are animate, and even then the use of the suffix is optional. At the time of his research, comparative data on Kiranti languages was virtually unavailable<sup>11</sup>, so he probably was not aware of the presence of *-lai* in other related languages, or of possible TB cognates. Furthermore, he mentions that 'there can be no doubt at all that traditionally both the direct and indirect objects have been unmarked.' It may be that Allen compared the speech of the oldest speakers he encountered with that of the younger generation, noticed a difference in transitive markers, and concluded that the change was due to Nepali influence. Allen points out that if the object of a transitive verb is marked with *-lai*, then there is no longer a correlation with the subject of intransitive verbs (which were, and still are, unmarked). To him, this is indicative of the breakdown of the ergative structure in the language.

### 3 Pronominal systems of Thulung and Nepali

As we have seen from looking at the transitivity patterns for modern and older Thulung, the ergative system functions according to a pronominal split, which has shifted over this time period. The second person plural used to be marked in the same way as the other second person pronouns, whereas now it is paired with third persons and common NPs. In order to see how this came about, we must look at the pronoun systems of the language and see where the change operated.

Modern Thulung shows a fairly complex pronoun system:

**Table 1:** Modern Thulung pronouns

<sup>9</sup> The counterexamples to Allen's neat ergative marking pattern are as follows (transliterated into my transcription system, with the ergative marker in bold; I leave Allen's gloss as in his text):

(1) *gutsi n**ip**hi-ka*  
'we both'

(2) *gana-ka derakhom ben-na*  
you lodging have-made

(3) ...*rwak-saka rwam basi guku-ka*  
...saying to-say it-is-necessary for-us (EXCL)  
'This is what we Thulung have to say.'

<sup>10</sup> Which he writes with a long /a:/. His data shows that he found contrastive vowel length at the time of his research, which is no longer present.

<sup>11</sup> Allen's grammar was printed in 1975 (his unpublished dissertation is from 1976), and even up to 1984 it remained the only existing grammar of a Kiranti language (Ebert 1994:8).

Person	SINGULAR		DUAL		PLURAL	
1	go		exclusive	inclusive	exclusive	inclusive
			gutsuku	gutsi	guku	gui
2	casual	polite	gatsi		ganimim	
	gana	gani				
3	casual	polite	gutsi		gumimim	
	gu	gumi				

The system laid out in Table 1 makes dual-plural as well as inclusive-exclusive distinctions, both characteristic of other related languages in the same general region of Nepal. The presence of honorifics is attested in Burmese and Tibetan, both languages with a long literary tradition, and in Newar, which has long been in contact with Nepali as well as being the language of the first kings of the Kathmandu Valley. So while the existence of honorific pronouns is documented for a few TB languages, these tend to be languages spoken in more explicitly hierarchical, urban societies than the smaller, rural ones that make up most of the language family. In other words, the pronoun system of modern Thulung is typical of its Kiranti heritage, while containing an unusual element in the existence of politeness distinctions for the second and third person singular pronouns.

The pronoun system of older Thulung is presented in Table 2.

**Table 2:** Older Thulung pronouns

Person	SINGULAR		DUAL		PLURAL	
1	go		exclusive	inclusive	exclusive	inclusive
			gutsuku	gutsi	guku	gui
2	gana		gatsi		gani	
3	gu		gutsi		gumi	

This shows a prototypical Kiranti pronoun system. We can see from a comparison of these two pronoun charts that the change rests in the addition of the honorific pronouns for the modern version of the dialect. The creation of a new set of polite pronouns for the second and third singular forms resulted in a rearrangement of the system: the second singular *gana* shifted into being used as an casual second singular, and the second plural *gani* became the polite second singular. Interestingly this is the same pattern as in many Indo-European languages, where the second plural has become the polite equivalent of the second singular pronoun. In Thulung, the shift of second plural into polite second singular resulted in a gap in the second person plurals. A new form was needed to replace the



missing plural, and this was created with the help of the nominal pluralising suffix, *-mim*,<sup>12</sup> resulting in a new second plural *gani mim*. The same situation occurred in the third person, resulting in a parallel shift in the pronouns and creation of a new plural form.

More schematically, where the old second person pronoun system was 2S *gana* / 2D *gatsi* / 2P *gani*, it shifted to 2S casual *gana* / 2S polite *gani* / 2D *gatsi* / 2P *gani-mim*. Similarly for the third person, 3S *gu* / 3D *gutsi* / 3P *gumi* changed to 3S casual *gu* / 3S polite *gumi* / 3D *gutsi* / 3P *gumi-mim*.

The pronoun system of modern Thulung is much more similar to those of Nepali and other Indo-Aryan languages. Nepali makes politeness distinctions, in fact making a three-way distinction<sup>13</sup>, for both second and third persons. Thulung, then, has followed the concept without achieving exactly the same result. Nepali pronouns are listed in Table 3.<sup>14</sup>

The Nepali plural pronouns for second and third person are formed by adding a nominal pluralising suffix *-haru* to the middle or high singular pronouns. This suffix is productive for pluralising animate objects: *manche*, 'person', for example, becomes *manche-haru*, 'people.' It is possible to use the pluralising suffix with inanimate objects, such as in *kalam*, 'pen' becoming *kalam-haru*, 'pens', but this results in a fairly marked form.<sup>15</sup>

Table 3: Nepali pronouns

Person/Number	SINGULAR			PLURAL	
1	<b>ma</b>			<b>hāmī</b>	
2	low	middle	high	Low/middle	high
	<b>tā</b>	<b>timī</b>	<b>tapāī</b>	<b>timīharū</b>	<b>tapāīharū</b>
3	low	middle	high	low/middle	high
	<b>u</b>	<b>unī</b>	<b>wahā</b>	<b>unīharū</b>	<b>wahāharū</b>

<sup>12</sup> This plural marker is used to pluralise nouns, and its use with pronouns appears to be a new phenomenon, calqued on Nepali. The distribution of the pluraliser *-mim* is as follows: it is an optional plural marker, but there is a scale in the frequency of appearance with certain noun classes. It appears most frequently with kin terms, where its use is fairly consistent, although examples are found where it is omitted. It is optional but generally used with non-kin humans, and considerably less frequent with non-human animates, such as herd animals. It tends not to appear very frequently with inanimate nouns (but this depends on the speaker).

<sup>13</sup> The levels of politeness are often called 'low', 'middle' and 'high' in grammars of Nepali: 'low' is used for animals, untouchable castes, and anyone to whom one wishes to express scorn or superiority, and additionally for wives and children sometimes. 'Middle' is most commonly used to refer to intimates, such as wives and children (although a wife will never refer to her husband by the same form, but rather use 'high'). 'High' is used in addressing strangers, parents and other elders.

<sup>14</sup> These pronouns are a simplified set of the pronouns given in Mathews (1998:39-44), omitting alternative pronouns as well as the less-commonly used feminine forms.

<sup>15</sup> Plural is often conveyed via context, particularly for inanimate objects.

In Thulung, pluralised nouns seem to pattern similarly to the way they do in Nepali, with a restriction based on animacy (as briefly described in footnote 13), except for situations when a marked plural form is desirable. The Thulung suffix for formation of plural forms of nominals is *-mim*, and we see how the creation of the new second and third plural pronouns, *gani-mim* and *gumi-mim*, is perfectly analogous to the corresponding plural forms in Nepali.

We have seen how both the ergative marking split and the pronoun system in Thulung have shifted over the last thirty years, and the next issue is to correlate the two. In older Thulung, the ergative splits along a clear first and second person vs third person pattern<sup>16</sup>, whereas it now splits within the second person. This is exemplified by the following representation, where // stands for the position of the split in the system:

Older Thulung system:

<i>go</i>	<i>gutsuku</i>	<i>gutsi</i>	<i>guku</i>	<i>gui</i>	<i>gana</i>	<i>gatsi</i>	<i>gani</i>	//	<i>gu</i>	<i>gutsi</i>	<i>gumi</i>
1S	1DE	1DI	1PE	1PI	2S	2D	2P	//	3S	3D	3P

Modern Thulung system:

<i>go</i>	<i>gutsuku</i>	<i>gutsi</i>	<i>guku</i>	<i>gui</i>	<i>gana</i>	<i>gani</i>	<i>gatsi</i>	//	<i>gani-mim</i>	<i>gu</i>
1S	1DE	1DI	1PE	1PI	2S	POL.2S	2D	//	2P	3S
<i>gumi</i>	<i>gutsi</i>	<i>gumi-mim</i>								
POL.3s	3D	3P								

If we look at the pronouns which are to the left of the // split mark, they are identical in form, even if their referents are slightly rearranged.<sup>17</sup> According to this representation, the shift in the split looks quite natural, and its explanation lies in the behavior of the pluralising suffix, *-mim*. By virtue of being exclusively a nominal pluraliser, before its relatively recent appearance on pronouns, *-mim* was limited to common nouns, which are by nature treated as third person. Looking at the representation of the new split system above, we see that all pronouns on the right side of the // mark (and therefore receiving ergative marking) are either third person or marked with the pluraliser associated with common NPs. As a result of the formation of the new second person plural pronoun by suffixation of the nominal pluraliser, *gani-mim* is treated as part of the third person-common NP subclass, and its ergative marking patterned accordingly.

<sup>16</sup> If we ignore the few examples where first and second person take ergative, *-ka* (these cases are not well explained, and more complete data would probably clarify their presence), and consider that the Ribdung dialect is outside the scope of this discussion.

<sup>17</sup> In other words, the second person pronouns to the left of the // are still *gana gatsi* and *gani*, even if the *gani* now represents a polite second singular instead of the plural it used to be.

#### 4 Typology of pronominal splits

Now that I have shown the likely explanation for the shift in pronominal split in Thulung's ergative marking, I will place the Thulung split in the typological context of other splits. Of the four major kinds of ergative split which occur in the world's languages (Dixon 1994:70–110), a split conditioned by the semantic nature of the NPs is a prominent possibility, along with conditioning based on tense/aspect, main versus subordinate clause, and the semantic nature of the verb. Within the category of split where the conditioning is based on the nature of the NP, as is the case for Thulung, a classic example is the Australian language Dyirbal, where the first two persons use a nominative-accusative system, and the third person and all other nouns an ergative-absolutive system. Many different patterns of pronominal split have been attested (Silverstein 1976) but the Thulung split, coming within the second person, is to my knowledge unattested. Silverstein's theory of person hierarchies states that first and second persons are always higher ranked than third persons, but says that which of first or second person is the highest ranked is difficult to determine. Dixon maps out the Nominal Hierarchy (1994:85), which I reproduce here as Table 4.

**Table 4:** Nominal hierarchy

1st person pronouns	2nd person pronouns	Demonstratives, 3rd person pronouns	Proper nouns	Common nouns: human, animate, inanimate
---------------------	---------------------	-------------------------------------	--------------	-----------------------------------------

← more likely to be in A than in O function ←  
(and therefore less likely to take ergative marking)

Dixon places first person pronouns at the top of the hierarchy, and says that in all but a very small number of languages (the exceptions are Algonquian languages) first always outranks second. For Dixon the issue of where a split occurs, in languages which do have split systems between nominative and ergative, is very much a semantic one, related to issues of control over certain types of action. The speaker is always considered to think of himself as having more control over an action, and therefore gets ranked at the left-most side of his hierarchy chart, followed by second and then third person, and eventually common nouns on the rightmost side of the chart. The main prediction of this chart is that if there is a split within the range of pronouns, proper nouns, and common NPs, ergative marking will always apply to the right side of the chart before the left side.<sup>18</sup> In other words, if a system has any ergative marking at all, there will be no instance of proper

<sup>18</sup> This is because the likelihood of ergative marking is inversely proportional to the likelihood of an NP being in A rather than O role.

nouns being marked when common nouns are not, or of third person pronouns marked when proper nouns are not, and so on up the hierarchy.

In the case of Thulung, the conditioning of the split is based on the association of *-mim* with third person/common NP behavior, following a contact-induced shift in the pronoun system. The original split in Thulung was based on the person hierarchy, with first and second person aligning differently from third person and other NPs, but the development of honorific forms, under the influence of Nepali, caused a shift to a different pattern. Yet it is interesting to note that Thulung, even after the shift in the system, still patterns along typologically predictable lines. This raises the question whether the new ergative pattern would have been possible had it violated the constraints of the hierarchies seen to control issues such as pronominal splits. This cannot of course be answered, as the data fits a typologically sound pattern, but raises the issue whether change, even through analogy, conforms to universal tendencies.

It is interesting that the influence of Nepali has been limited, as far as the ergative goes, to a reorganisation of the pronoun system and the resulting shift in position of the ergative split. The reason this is surprising is that Nepali also has a split ergative system, similar to that of Hindi, where the split is along the lines of tense: only perfective transitive sentences are marked for ergative. Considering the influence of Nepali in so many areas of the grammar, it is interesting that Thulung's ergative system is strong enough to fend off a possible Nepali-like tense split in favour of a much more Rai-like person-based split, all the while rearranging its pronouns to look more like those of Nepali.

## 5 Conclusion

In this paper, I have attempted to account for the position of the split in Thulung's ergative marking. The split occurs between the second person dual and plural, with the second person dual and all 'higher' pronouns (first person and second singular) being unmarked, and all 'lower' pronouns (second plural, third person, and all common NPs) receiving ergative marking. The position of the split is typologically sound, both before and after the change, in respecting the person hierarchy, and the explanation for the shift therefore lies elsewhere.

In having data from both the 1970s and 2000, it was possible to identify a change in the pronoun system, which proved to be at the root of the shift. Through contact, Thulung's pronoun system was expanded to be more similar to that of Nepali, resulting in two-way politeness distinction for second and third person singular. This shift then led to a new, compounded version of second and third person plural pronouns, which were created by the suffixation of the nominal pluraliser. Because of this pluraliser's previously exclusive association with common NPs, the new second person plural pronoun has been reinterpreted by analogy as being similar in nature to such NPs. The result is that it is now

marked as ergative in appropriate contexts, with a concomitant shift in the position of the ergative split.

Thulung and the other Tibeto-Burman languages of Nepal are in an intense contact situation with Indo-Aryan Nepali, as we saw in this case, with Nepali creating the pressure for a shift in the pronoun system which then has consequences in other areas, such as case marking. It is interesting to speculate on further contact-induced changes in Thulung, as the number of fluent speakers dwindles and Nepali makes further inroads into the language.

### References

- Allen, Nicholas J., 1975, *Sketch of Thulung grammar*. *Cornell University East Asia Papers* No.6. Ithaca: Cornell University.
- Bickel, Balthasar and Johanna Nichols, 2001, *Inflectional morphology*. Unpublished ms., U.C. Berkeley
- Comrie, Bernard, 1981, *Language universals and linguistic typology*. Chicago: University of Chicago Press. (2nd edition 1989. Oxford: Basil Blackwell.)
- Dixon, Robert Malcolm Ward, 1994, *Ergativity*. Cambridge: Cambridge University Press
- Dryer, Matthew S., 1986, Primary objects, secondary objects, and antitativity. *Language* 62/4:808–845.
- Ebert, Karen, 1994, *The structure of Kiranti languages*. *Arbeiten des Seminars für Allgemeine Sprachwissenschaft*, 13. Zürich: Universität Zürich.
- LaPolla, Randy, 1992, 'Anti-ergative' marking in Tibeto-Burman. *Linguistics of the Tibeto-Burman Area* 15/1:1–9
- 1994, Parallel grammaticalizations in Tibeto-Burman languages: evidence of Sapir's 'Drift'. *Linguistics of the Tibeto-Burman Area* 17/1:61–78.
- Matisoff, James A., 1991, Universal and areal dimensions of grammaticalization in Lahu. In Elizabeth Traugott and Bernd Heine eds *Approaches to grammaticalization*, vol.2, 383–453. Amsterdam: John Benjamins.
- Matthews, David, 1998, *A course in Nepali*. Richmond, Surrey: Curzon Press.
- Silverstein, Michael, 1976, Hierarchy of features and ergativity. In Robert Malcolm Ward Dixon ed. *Grammatical categories in Australian languages*, 112–71. Canberra: Australian Institute of Aboriginal Studies.

# 8 *Why languages differ: variation in the conventionalisation of constraints on inference\**

RANDY J. LAPOLLA

Language is the frail bridge that we fling across the chasm of the inexpressible and the incommunicable.

Matisoff 1979[2000]:2

## 1 Introduction

Sperber and Wilson (1996) and Wilson and Sperber (1993) have argued that communication involves two processes, ostension and inference, but they also assume there is a coding-decoding stage of communication and a functional distinction between lexical items and grammatical marking (what they call 'conceptual' vs. 'procedural' information). Sperber and Wilson have accepted a basically Chomskyan view of the innateness of language structure and Universal Grammar. In this paper I will also assume

---

This paper presents a view of language and communication that developed as a synthesis of what had originally been separate interests in grammaticalization, pragmatics, typology, and Sino-Tibetan linguistics, informed overall by the fact that language use is an aspect of human interaction. These are of course the very areas and approach of Prof. James A. Matisoff, and his influence on this paper and all of the work I have produced over the years cannot be overestimated. It is with great respect and affection that I submit this paper in his honor. Earlier versions of this paper were presented at City University of Hong Kong, the University of Melbourne, and the Australian National University. I would like to thank all those who participated in the discussions at those times, and I would also like to thank Sasha Aikenvald, Nick Enfield, Ruth Kempson, Steven Nicolle, Michael Pickering, Dan Sperber, and Elizabeth Traugott for sending me written comments.

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 113–144.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

that communication involves ostension and inference, but propose the hypothesis that there is no innate language structure, and no deterministic coding-decoding process, as all aspects of interpretation involve inference. The inferential process involved in communication is the creating of a context in which the ostensive act achieves relevance (makes sense). I also reject the idea of a functional difference between conceptual and procedural information, and argue that the role of all aspects of language use in communication is to constrain the inferential process, to help the addressee construct a context in which the communicator's ostensive act can be seen to be relevant. Contrary to most work in pragmatics which assumes that context disambiguates language, I am proposing that it is not context that disambiguates language, but language that disambiguates the context of interpretation. That is, rather than assume that the form of the utterance is given and the context is the manipulable variable, we should recognise that in real-world communicative situations, speakers have no choice in terms of the context they are communicating in, but as communicators have choices in terms of the form of the utterance they use. The more explicit the linguistic form, the more constrained the addressee is in constructing the context of interpretation. On this view a language and the rules for its use in a particular society are a set of social conventions which have evolved in the particular way they have in that society in response to the need to constrain the inferential process involved in communication in particular ways thought to be important in that society.<sup>1</sup> These conventions (or some subset of them) become habits of the individual speakers of the language. These conventions and habits are no different in nature from other types of conventions and habits that have developed in the society for performing particular actions, such as the conventions in the U.S. of eating with a fork, or of men wearing ties on formal occasions, or of driving on the right-hand side of the road. As each society develops (evolves) its own particular sets of conventions (linguistic and non-linguistic), each set (in totality) is unique to that society and so manifests the way that society construes and deals with the world. That is, the habits and conventions for carrying out actions (including communication) of a society reflect habits and conventions of thought. The significance of this view of language to typology is that there is no assumption of a universal genetically-encoded grammar, and each language is seen as a unique entity reflecting a unique society, and so in doing linguistic work, we must look at

---

<sup>1</sup> I am purposefully avoiding use of the word 'culture', as it is a problematic term for many in the anthropological tradition. What I am interested in here are the conventionalized methods and tools that a society develops for carrying out certain actions. For example, in a restaurant serving Italian food in the U.S., the table will by convention be set with at least one fork and a plate, while in a restaurant serving Chinese food in Beijing the table will by convention be set with chopsticks. Some conventions may be written into law, such as driving on the right-hand side of the street in the U.S., to ensure that everyone in the society follows the convention, but they are still conventions. As people follow the conventions on a regular basis, they become habits of thought and action, so that not following a particular convention will feel and often be considered 'wrong', and will often be difficult to change.

each language on its own terms and determine which functional domains it obligatorily constrains the interpretation of, to what extent it constrains them (if it does), and what formal mechanisms it uses to constrain the creation of the context of interpretation. It is in these ways that languages differ.

## 2 Ostensive-inferential communication

In human communication, one person (the communicator) does something (an ostensive act) with the intention to cause another (the addressee) to become aware of some phatic or factual information. The process by which the addressee becomes aware of the information is an inferential one. The hearer uses inference to recognise the communicative intention of the speaker, and, given the recognition of that intention and the particular form of the ostensive act, the addressee can (usually) infer the intended information. The ostensive act can be linguistic, but it need not be communication can (and often does) occur without language.<sup>2</sup> What is necessary for communication is not the exchange of symbolic expressions, but the successful determination of the reason for the communicator making the particular ostensive act that he or she made. In terms of a linguistic ostensive act, what is communicated is not what the communicator says, but what the addressee infers to be the intention behind the communicator making that particular ostensive act, that is, saying those particular words. Language is not the basis of communication, but simply an instrument used to help the interpreter more easily infer the speaker's communicative intention, as it constrains that inferential process by reducing the number of assumptions that could potentially be part of the context of interpretation. Even when the ostensive act is linguistic, there are often a great number of degrees of explicitness possible, depending on the speaker's estimation of the hearer's inferential abilities and current knowledge state; the more explicit the utterance, the more constrained the interpretation, as in the six different possible answers to the question given in (1) (all of which have the same 'meaning'; of these, the first is attested).

---

<sup>2</sup> Cf. Keller (1994: 25): A language facilitates communication, but it is not the condition of its possibility. To communicate with the help of conventional instruments such as linguistic ones is a special kind of communication, although this is for us the normal and prevailing way to communicate. We are so used to it that many think that the common possession of a stock of signs together with syntax is logically required (the condition of possibility) in order to communicate at all. If this were the case, we could neither meaningfully pose the question of how we as a species acquired language phylogenetically, nor how small children can learn their mother tongue ontogenetically. The reason is that the construction of rule hypotheses presupposes (among other things) successful communication.



- (1) Q: Do you want something to drink?  
 A1: (points to soup bowl)  
 A2: I have soup.  
 A3: No. I have soup.  
 A4: No, because I have soup.  
 A5: No, since I have soup, I don't need anything to drink.  
 A6: No, I don't want anything to drink. Since I have soup, I don't need anything else to drink right now.

A1 does not constrain the interpretation very much, and so the addressee must (a) infer that the communicator is pointing at the soup bowl and not something else, (b) must notice that the bowl is full, and (c) must infer that the fullness of the bowl is somehow relevant to the communicator's communicative intention, and (d) then infer that the relevance is that soup is something to drink, and then (e) infer that since the communicator has something to drink, she might not need something else to drink, and then (f) conclude that the communicator's communicative intention in pointing at the soup bowl was to alert the addressee to the fact that she does not require anything to drink, as she has soup, and that is enough. A2 constrains the interpretation somewhat more than A1, in that it narrows the context of interpretation by explicitly mentioning the soup (thereby eliminating steps (a-c) above), and so the addressee can then start with the assumption that having soup is relevant. The other inferences (d-f) still must be worked out, though. A3 constrains the context of interpretation even more by supplying the negative answer to the question; A4 makes explicit the cause and effect relation between the negative answer and the fact of having soup. A5 constrains the process of interpretation more than A4, and A6 constrains it more than A5. It is important to notice two things here: (i) there is no difference in communicative function between the linguistic and the non-linguistic responses; (ii) the additional words, such as *No* in A3, and the additional marking of semantic and grammatical relations, such as *because* in A4, both have the same function, to constrain the context of interpretation to a greater degree than not using those words.

The difference between non-linguistic communication and linguistic communication, or brief speech and highly explicit speech, is like the difference between ripping bread into pieces with your hands and cutting it carefully with a knife, a difference of tool or mode, with resulting differences in precision. What is important is the separating of the two parts of the bread; we should not mistake the tool used (the knife) for the process of separating the two parts. We often communicate with our hands or other body parts, such as pointing at the wrist to ask the time, or nodding the head, or wagging the finger with pouted lips to show disapproval (there are both conventionalised and non-conventionalised gestures). The entire process of interpretation involves inference, whether the ostensive act is linguistic or not. Interpretation of a linguistic ostensive act involves identification of the ostensive act as a linguistic act and recovery of its form. We are not usually conscious of this aspect of communication as involving inference, yet psycholinguistic studies (for

example, Warren 1970) show that we do use inference to construct the form of the utterance we are hearing. In a multilingual situation, such as where I work in Hong Kong, a place where three different languages (Mandarin, Cantonese, and English) are regularly used by the same people, and you don't know when they will use which language, the inference necessary to determine which language the person is using when they start to talk to you often becomes a conscious process.

Inference is also involved in identifying referents and delimiting the possible intended senses of words and structures, and then there is inference involved in deriving any implicatures that must be created in the processing of the resulting proposition. Interpretation then is not simply decoding a signal. Even interpreting something as code-like as  $1 + 1 = 2$  involves inference of a non-binary number system based on the appearance of the number 2 rather than 10. All linguists would agree that the pragmatic aspects of meaning, such as resolving ambiguities, correcting mistakes, identifying referents, identifying illocutionary force, recognising irony and humor, and completing incomplete utterances, must be interpreted by inference, but I would argue that all aspects of interpretation involve inference. The inference involved in interpretation is essentially guesses at what the communicator's intended message might be. These guesses are possible because of the unconscious assumption of the principle of relevance, given in (2).

(2) The principle of relevance (Sperber and Wilson 1996:260, 270):

1. Human cognition tends to be geared to the maximisation of relevance.
2. Every act of ostensive communication communicates a presumption of its own optimal relevance, such that;
  - a. The ostensive stimulus is relevant enough for it to be worth the addressee's effort to process it.
  - b. The ostensive stimulus is the most relevant one compatible with the communicator's abilities and preferences.

In order for the addressee to be able to infer the communicator's intention, the communicator must choose and tailor the utterance, in the case of linguistic communication, in such a way that the hearer will not have to expend unnecessary effort to create a context that will allow him/her to achieve relevance (arrive at the intended interpretation). In doing this, the speaker takes into consideration guesses as to what information is available to the hearer at the time of utterance for use in interpreting the utterance. A speaker must decide what to make explicit and what to make implicit (and also, among implicatures, what to make stronger or weaker implicatures), and this is done on the basis of the speaker's estimation of the hearer's processing abilities and contextual resources, but also partly on politeness considerations and what we think of as 'style'<sup>3</sup> The

---

<sup>3</sup> An example of the use of a particular utterance form in order to convey weak implicatures for the sake of politeness is the following:

more information that the speaker assumes the hearer is able to access in the processing of an utterance, the less explicit the utterance can be. Answer A1 in (1) above was interpretable in the situation in which it occurred, an interaction between a husband and wife at the dinner table, but would not be interpretable in a situation such as an interaction between a waiter and a customer in a restaurant. The wife could assume the husband could assemble a context of interpretation in which such a minimally explicit ostensive act would be relevant, but she could not make such an assumption in dealing with a waiter; here a form which constrains the context to a much greater degree, possibly as much as A5 or A6, would be necessary to communicate the same intention.

The degree to which the hearer is forced to deduce a particular interpretation depends on the degree to which the form of the utterance constrains the hearer in choosing the contextual assumptions necessary to achieve relevance in interpreting the utterance. In (1) we saw that having more words or more grammatical marking in the utterance can more greatly constrain the interpretation. The order of elements that the speaker chooses also influences the hearer's interpretation, as the hearer begins to assemble the context of interpretation as soon as the first word is uttered (or possibly earlier), and this initial set will influence the eventual set used for the overall interpretation. This is true both at the clause level and at the phrase level. Halliday (1994:197), for example, shows how the order of elements in the noun phrase in English is related to the degree to which the element helps the hearer identify a particular referent. In fact all aspects of language can be shown to constrain the interpretation, and that is in fact their *raison d'être*.

As mentioned above, work in Relevance Theory, while recognising the importance of constraining the context of interpretation, assumes a distinction between conceptual and procedural information (for example, Blakemore 1987, 1988a,b, 1990; Wilson and Sperber 1993; Nicolle 1997). For example, it is argued that one way the speaker can constrain the interpretation of implicature is to use discourse connectives such as *so* and *after all*, which are said to contain procedural information (procedures for manipulating conceptual representations), that is, information on how to interpret the proposition, to alert the hearer to the fact that one part of the utterance has a particular relationship to another part of the utterance, such as providing additional evidence or an explanation. In (3a-b) is an

A: Would you like to go see a movie tonight?

B: Thanks, but I have an important test tomorrow morning.

The implicated conclusion, a strong implicature, is that B cannot go to the movies that night, but there is a weak implicature that were it not for having that test the next day B WOULD go with A to the movies, and it is in order to convey this weak implicature that this particular form of utterance is chosen. (The proposition conveyed by the weak implicature need not be true; it may be that the speaker is just trying to be polite (save A's 'face')—the strength of an implicature is directly proportional to the degree to which the speaker takes responsibility for the hearer making that particular interpretation. In this case, if B wanted to be sure A made that interpretation, B could add *How about next week*, or some such expression.)

example with two possible interpretations (from Wilson and Sperber 1993:11). In one interpretation the statement in (3a) provides evidence for the conclusion in (3b); in the other the conclusion in (3a) is confirmed by the evidence presented in the statement in (3b). In this case it would be possible for the speaker to constrain the hearer's choice of one or the other of these two interpretations of the conceptual information by adding procedural information (either *so* or *after all*) to the beginning of the second clause, as in (4a) and (4b) respectively.

- (3) a. Peter's not stupid.
- b. He can find his own way home.
- (4) a. Peter's not stupid, so he can find his own way home.
- b. Peter's not stupid, after all, he can find his own way home.

Discourse connectives such as these are said by Wilson and Sperber to not encode concepts (that is, they do not contribute to truth conditions); they just constrain the inferential phase of the comprehension, narrowing down the search for relevance, and thereby make the search easier, and make the interpretation selected more determinate.

Gumperz (for example, 1977, 1982, 1989, 1992a, 1992b) has also argued that hearers interpret the meaning of an utterance based on inferences about the speaker's underlying strategies and intentions, and that these inferences are drawn on the basis of interpretive frames (contexts) evoked by certain linguistic or non-linguistic contextualisation cues produced by the speaker. But as with Relevance Theory, Gumperz sees a distinction between lexical content and contextualisation cues.

Work by both Gumperz and those working in the Relevance Theory framework assume that language involves two types of elements, those that express conceptual information, and those that simply constrain the interpretation of the elements which express conceptual information. I would argue that in fact there is no functional difference between the two types, as both types of elements constrain the creation of the context of interpretation for inferring the speaker's intention. In Gumperz's terms, I would say *all* of language is a contextualisation cue. The goal of communication is not to decode the meaning in words (a view which is implied in the conceptual-procedural distinction), but to understand the speaker's communicative intention. This can happen with or without language, and the only purpose of language in this process is to constrain the inferential deduction of the communicative intention. Therefore the so-called 'conceptual' items are also constraining interpretation.<sup>4</sup> If I hand an assistant a piece of paper and wave my hand towards someone else (who may be in a crowd of people) with the intention that the assistant should give the paper to him, the interpretation of the action and the person it is to be given to is quite

---

<sup>4</sup> The difference between lexical and grammatical items is the generalness of use; lexicalization (idiomization) and grammaticalization are the same process (conventionalization), but differ in terms of generalness. See below for discussion.

unrestrained (though may be unproblematic in that context). If, instead, I say *Give it to him*, the interpretation of the person it is to be given to is still relatively unconstrained, but if I say *Give this paper to the tall man with the red hat by the back door*, or *Give this paper to the teacher*, then I have constrained the interpretation of the referents involved considerably, and the constraining is mainly done by the extra lexical items. (In this example there are both grammatical and non-grammatical elements, but this would not be true in all languages). Given an expression like *the teacher*, do we want to say that the 'procedural' marking (the definite marking) helps us interpret the 'conceptual' item *teacher*, or do we want to say that the phrase *the teacher* (rather than pointing or using a more general noun phrase) helps us identify the relevant referent? One might argue that *the* is helping to constrain the identification of the teacher by alerting the hearer to the cognitive accessibility of the referent of *teacher*, and this is true, but the use of the word *teacher* itself, as opposed to a less specific term, is also helping to constrain the context of interpretation. That is, both items are helping the hearer to identify a particular referent. Communication does not necessarily involve language, but the use of any amount of language constrains the interpretation more than not having language involved, and generally the more explicit the language involved, the more constrained the interpretation. In this case, *teacher* would constrain the interpretation of a particular referent more than, for example, *him*, or *person*. The function of both lexical and grammatical means in constraining interpretation is the same.<sup>5</sup> For example, those making a distinction between so called 'conceptual' information and 'procedural' information might say that adding the expression *I guess* to an English declarative clause such as in *I guess he's coming* would be adding conceptual information, while adding an evidential particle marking a guess to a similar clause in some other language that has grammaticalised evidential marking would be considered as adding only procedural information, yet the function/information of both is the same. **It is precisely because they have this function that lexical items can grammaticalise into grammatical marking.**

### 3 The development of language structure

Givón (1979a, Ch. 5; 1979b) has argued that language develops from pragmatic, loosely structured linguistic modes to tighter, more structured modes, and that these modes can be seen in the differences between child language and adult language, between pidgin languages and standard languages, between spoken and written registers, between informal and formal registers, and between unplanned and planned discourse. Communication in the pragmatic mode depends largely on word order and the lexicon alone, while

---

<sup>5</sup> Langacker (1987) argues that there is no difference between 'conceptual' and 'procedural' from the point of view that all linguistic structures are meaningful. This is similar to what I am saying, but only if we interpret 'meaningful' as 'having a role in constraining the creation of the context of interpretation'.

communication in the more syntacticised mode depends much more on conventionalised constructions and morphology in tightly structured relationships. Grammar develops as the originally free collocations of lexical items become fixed in a particular structure. Paul Hopper (1987, 1988) has developed this idea into the conception of grammar as 'emergent' from discourse. He has argued that rather than taking grammar as a given (what he called '*a priori* grammar'), and then possibly seeing how discourse can affect grammar after it is established, linguists should see discourse as prior to grammar, and giving rise to grammar, as repeated patterns of discourse develop into what we think of as grammar. Grammar is then not seen as fixed structure, but something that is constantly evolving (see also Langacker 1987, Ono and Thompson 1995).

The emergent grammar view of the development of language structure is a natural corollary of the theory of ostensive-inferential communication just presented. The discourse patterns that lead to the development of grammar are those that are repeatedly used for constraining the interpretation of utterances in a particular way. For example, in Old English the word *lic* 'like' plus the instrumental suffix *-e* were used so often after an adjective to make explicit an adverbial relation to a verb that it became conventionalised and developed into the adverb-forming suffix *-ly* used obligatorily in many contexts in English today (Lass 1992). The frequent use of a demonstrative adjective to show that a referent was cognitively accessible conventionalised into definite marking in English (Pyles and Algeo 1982). In Mandarin Chinese the frequent use of a preverbal locative phrase where there was an implicature of an on-going event led to the development of a progressive marker from the locative verb *zài*. What begins as a conversational implicature over time becomes conventionalised, so it is then a conventional implicature, and then can become further conventionalised until it is simply a part of the grammar that forces a particular interpretation. The differences between these three (conversational implicature, conventional implicature, and obligatory marking forcing a particular interpretation) is the degree to which speakers are free to use or not use them to constrain the hearer's inferential process, and also the degree to which the form forces a particular interpretation. We can think about grammaticalisation and the fixing of particular patterns in language use like the creation of a path through a field (see Keller 1994). One can cross a field any number of ways, and there was originally no difference between the part that eventually became the path and the rest of the field, except that people found it expedient to all go the same way through the field, and so the grass was worn away, creating the path. Eventually people start using the path just because it is there, without thinking about whether it is the best way to go through the field. At some point, either out of simple conventionalisation or because of some social factor (for example, attitudes towards preserving the grass that is left), it may become recognised as the 'unmarked' way to go through the field and crossing any other way would be considered 'marked'. This conventionalisation is the same whether it is the fixing of a particular word order or construction, the fixing of a lexical item in a particular context such that it becomes

grammatical marking, or involves the extension of the use of already existing morphology. What we think of as a grammatical construction (or 'constructional schema'—Langacker 1987; Ono and Thompson 1995; Barlow and Kemmer 1994) is also simply a pattern of usage that was used often enough by enough people to constrain the hearer's interpretation in a particular way that that usage became conventionalised.

The fixing of repeated patterns into grammar is nothing more than the development of conventionalised forms that restrict interpretation, and Givón's cline of forms from more pragmatically to less pragmatically based types correlates with the degree to which interpretation is constrained grammatically rather than lexically. The development of grammar out of repeated discourse patterns then can be seen as the fixing of constraints on the search for relevance during the process of interpretation.<sup>6</sup>

There are at least four types of conventionalisation that affect language structure. Morphological means for constraining the interpretation of particular functional domains develop out of the repeated use of particular lexical items for constraining interpretation in a particular context. An example from English is the development of definite marking from the frequent use of a demonstrative pronoun for constraining the identification of the referent of an expression to a contextually accessible referent.

Already available morphological marking may be extended in new ways and become conventionalised in that new use. The extension of the use of the reflexive marker from direct reflexive situations to middle situations is a good example of this. In this extension a marker that originally was used only to mark direct reflexives comes to be used in some middle situations optionally with an emphatic sense to narrow the range of possible interpretations (for example, the use of *myself* in *I stood myself up* constrains the interpretation of the purposefulness of the action), and later comes to be used so often that it becomes obligatory for many verbs. This happened in the Romance languages (see Kemmer 1993), and also in the Tibeto-Burman language Dulong (LaPolla 1995b; see example below).<sup>7</sup>

---

<sup>6</sup> Though I am presenting this from the point of view of constraints on interpretation, I do not assume that linguistic change is hearer-driven. From one point of view we can say it is speaker-driven, as the patterns can only become conventionalized if speakers choose to use the patterns over and over again. From another point of view the conventionalization process takes time, and involves the same people as speakers and hearers. That is, a speaker uses a particular pattern and other people pick up on that (we are creatures of habit and imitation), and repeated use of that pattern by a number of people causes it to become grammaticalized (such as the same-subject interpretation of English clause-coordination). The same is true of lexicalization. Language development is an 'invisible hand' phenomenon; it is an epiphenomenon which results from the actions of many individuals (Keller 1994), and so we cannot say it is speaker or hearer driven.

<sup>7</sup> Once this happens, there is then no formal distinction between reflexives and middles, and so some languages then reinforce or renew the direct reflexive marking, again being driven by the desire to constrain the interpretation. This has happened, for example, in Dutch (Kemmer 1993).

A third type of conventionalisation is the fixing of syntactic (rather than morphological) constraints on interpretation, such as many of those associated with the concept of 'subject' in English, for example the cross-clause same-subject constraint in coordination with a reduced second clause (see discussion below). This development is simply the fossilisation of a frequent pattern of coreference. Like in the path analogy given above, a particular coreference pattern between the two clauses became so common it became the unmarked and assumed pattern through conventionalisation.

A fourth type of conventionalisation is a type of secondary grammaticalisation where a form that has grammaticalised from a lexical item and at first only constrains the interpretation of the external described situation later further grammaticalises in the direction of constraining the interpretation of subjective (speaker-oriented, expressive) aspects of the interpretation, with a stage in between of marking textual cohesion (that is, the path of development is 'propositional (> textual) > (expressive)'; Traugott 1990:497; see also Traugott 1982, 1988, 1989, 1990; Traugott and König 1991). An example of the full set of changes is English *since*, which developed from the propositional sense 'after, from the time that' to a marker of temporal relation, and from inference from the temporal relation to a marker of a causal relation (Traugott 1990:497).

#### 4 Why languages differ

Language is a tool which aids in the process of inference and so is shaped by the demands of that process, just as a hammer is shaped the way it is because the main use it is put to is hammering nails. The development of particular types of linguistic structure is not teleological, any more than the evolutionary development of species is. It is in fact a type of evolution, though an aspect of socio-cultural evolution rather than biological evolution. This applies equally to the development of the lexicon and the development of morphosyntax (which are actually not two separate things—see below). Language is what Keller (1994) calls 'a phenomenon of the third kind'. That is, it is not a natural phenomenon, and it is not an intentionally created artifact of humans. Language is the cumulative result of the actions of many individual humans, but their actions are not with the intention to create language; language is the unintended byproduct of their attempts to communicate effectively (constrain the addressee's inferential process effectively) on an individual level. It forms as if guided by some invisible hand, much the way economies and paths in fields develop (see above).

Just as the evolution of species is related to particular environments, many of the conventions of a people are responses or adaptations to particular environmental factors, such as building houses on stilts where there is frequent flooding. We find conventionalisations of language also related to particular environments. For example, it is no coincidence that the Qiang people of Sichuan, China, who live on the sides of mountains along river valleys, have conventionalised in their language (Qiang; Tibeto-



Burman) a complex system of direction prefixes including prefixes marking 'up-river' versus 'down-river' and 'up the mountain' versus 'down the mountain' (LaPolla to appear), as in (5).

(5) Qiang directional prefixes (*ɣue* 'throw')

<i>təʂu</i> 'throw up (the mountain)'	<i>zəʂu</i> 'throw towards the speaker'
<i>ɦəʂu</i> 'throw down (the mountain)'	<i>dəʂu</i> 'throw away from the speaker'
<i>səʂu</i> 'throw down-river'	<i>əʂu</i> 'throw inside'
<i>nəʂu</i> 'throw up-river'	<i>ɦəʂu</i> 'throw outside'

There has been work (e.g. Bernstein 1971, Perkins 1980, Trudgill 1996, 1997) showing that the size and complexity of the speech community will influence the use patterns of the language spoken, and this in turn will influence the form that the language takes. For example, complex inflectional deictic systems such as the one we find in Qiang are said to be more often found in small homogeneous communities rather than large complex communities (Perkins 1980).

We often find the same types of structures appearing in totally unrelated languages, but again, in parallel with evolution, where both sharks and dolphins have similar body shapes even though they are unrelated creatures, and bats, birds, and butterflies all have wings, similarities among conventions are not due to some predetermined or universal genetic imprint, but due to similar adaptations to similar environments. For example, many societies have independently developed bowl-like implements in response to the need to drink liquids and the nature of liquids.

The particular patterns we find used to constrain the interpretation reflect particular ways of construing and representing the world. As language structure is formed from repeated discourse patterns that constrain the hearer's interpretation in particular ways, it necessarily must be the case that those aspects that were being constrained were salient to the speaker and also assumed by the speaker to be salient or relevant to the hearer, at least in the contexts where the pattern was used. For example, Pawley and Lane (1998) argue that to understand the grammaticalisation of serial verb constructions in Kalam (Papuan, New Guinea), it is necessary to understand that in reporting an event, a speaker of Kalam is expected to make reference to a sequence of associated actions that express whether the actor was at the scene of the event or moved to the scene; what the actor did; whether the actor then left the scene, and if so whether the actor took the affected object along or not; and what the final outcome was. That is, where in English we would usually mention a single action to represent a series of related actions, for example, *I cut firewood*, in Kalam the individual associated (prerequisite and consequent) actions would be made explicit. The interpretation of these aspects of the action are then generally more constrained in Kalam than in English. The expression of the various aspects of the overall multi-scene event or action in Kalam can be elaborately spread over many clauses, spread over just a few clauses, or, in the case of relatively familiar multi-scene events, can be done with a

serial verb construction. It is the fact of the salience of mentioning all these different aspects of an event, plus the fact that some types of action sequences are performed regularly, that caused these serial verb constructions to become conventionalised (grammaticalised). In this case it is easy to see the 'smoking gun' of the demand on event narration that led to this grammaticalisation, but we do not need to find the 'smoking gun' in all cases to know that the grammaticalised patterns reflect the salience of the type of information being constrained.

Another example of how construal of the world affects the sort of conventionalisations that develop is discussed by Heine (1994, see also Heine 1997a, 1997b). He argues that there are four main basic event schemas (conceptual source structures) that give rise to the different types of comparative constructions found in the world's languages, and that how a particular group of people construe the comparative relation determines the type and structure of the comparative construction used by those people (X = the comparee (the thing being compared), Y = standard of comparison, Z = quality; this is necessarily brief—see Heine 1997a, Ch. 6 for more detailed discussion):

The Location Schema: the relation is construed in terms of relative location: X is Y at Z; e.g. Rawang *ədiu nū əpūŋ mədām yāŋz̄* [Adeu TOP Apung above/on tall-INTRANS.NPAST] 'Adeu is taller than Apung'.

The Action Schema: the comparee is seen as a kind of agent which surpasses, defeats, exceeds, etc. the standard of comparison in some way: X surpasses Z with regard to Y; e.g. Cantonese *ŋɔ<sup>33</sup> kɔ<sup>55</sup>-kwɔ<sup>33</sup> lei<sup>33</sup>* [1SG tall-surpass you] 'I am taller than you'.

The Polarity Schema: the relation is construed as antithetic juxtaposition of two antonymical qualities: X is Y, Z is -Y; e.g. Hixkaryana *kaw-ohra naha waraka, kaw naha kaywerye* [tall-not he.is Waraka tall he.is Kaywerye]. 'Kaywerye is taller than Waraka' (Stassen 1985:184, cited in Heine 1997a:117).

The Temporal or Sequence Schema: what comes earlier is seen as having more of the quality than what comes later: X is Y, then Z; e.g. Javanese *enak daging karo iwak* [is.good meat than fish] 'Meat is better than fish' (Stassen 1985:60, cited in Heine 1997a:118; *karo* is also used as a consecutive conjunction)

One subset of conventions often influences another subset of conventions. For example, in Australia and the US open-plan kitchens are popular, but in Hong Kong, as many people have maids, open plan kitchens are not popular. One set of conventions (related to housework) has influenced another set of conventions (related to housebuilding). In terms of the set of conventions related to communication (language and language use), we can also find that it is influenced by sets of conventionalisations involved in other (non-linguistic) sets of conventions. For example, the conventions of language use in Javanese are very much influenced by the conventions of the caste system and social status in the society in general (Errington 1988). The Jinghpaw people of China and Myanmar do not have a word for 'toilet' (i.e. a place to defecate); as they don't fertilise their fields, they do not save human waste and so do not construct toilets, they just

go out in the woods. Here the conventions of farming have influenced the conventions of architecture, and the latter have influenced the conventions of language.

There are also sometimes competing motivations (DuBois 1985) for one pattern or another, but the process of a particular form becoming conventionalised is the same. For example, English lost the distinction between singular and plural second person pronouns because of a repeated pattern of using the plural pronoun when referring to a singular referent out of politeness considerations (that is, constraining the context of interpretation to the individual was seen as less polite than not constraining it in that way, and then the form used to constrain the interpretation to the singular individual disappeared from the set of conventions), but some Southern (U.S.) dialects have conventionalised a second person plural form *y'all* from repeated use of *all* after *you* to constrain the interpretation of singular vs. plural referents.

To say that the original development of a particular pattern is motivated does not imply that the motivation will always be transparent. In many discussions of ethnosyntax, the opaqueness of certain structures is taken to be evidence that it is not possible to show a link between language and other sets of conventions. Yet in many aspects of our lives, once a particular way of doing something is conventionalised, the original motivation may be lost, while the conventionalised behaviour continues, simply because it is already a convention. For example, when British people first began drinking tea, they were forced to put the milk in the cup before the tea because the ceramics produced in Britain at that time were not able to withstand the heat of the tea directly without cracking, and so the milk was put in first to protect the teacup. Later they had access to better ceramics, and so the motivation for putting the milk in before the tea was gone, but by that time the practice had become conventionalised, and is still continued by many people to this day. Another example is that early clocks had chains with weights descending out the bottom of the clock, as the weights ran the clock. Modern clocks are now largely electronic, but are often still designed to have the weight chains (or stylised representations of them) because that particular conception of a clock had become so conventionalised. In language, we have many expressions that are no longer motivated by their semantics, but reflect earlier lexical uses or ways of construing the world. For example, we often use fixed expressions in English such as *pig in a poke*, *pass the buck*, *put it in the hopper*, or *the stars in the firmament*, yet few speakers of English know what a poke is, or what a buck is, or what a hopper is (why it is called a hopper), or that the word *firmament* derives from a view of the heavens as a fixed dome. We often say *dial a phone*, even though our phones now have push-buttons rather than dials. In Chinese the motivation for the old word for 'crow', *wū*, is not at all transparent, but if we reconstruct the original form of the word we can see that it was onomatopoeic (\*ʔa). In Chinese also the words for 'cash money', *xiànjīn* [current-gold], and 'bank', *yínháng* [silver company], were originally motivated by the fact that gold and silver were the conventional currencies. This is no longer the case, but the names continue to be used. The same is true of grammatical patterns/morphology: the original

motivation may no longer be transparent, but that does not mean there never was any motivation for the pattern, such as the *-r-* in *children*, a remnant of an old plural marker, now redundant because of the *-en* plural marker, but retained nevertheless.

I mentioned above Heine's work showing that how speakers of a language construe a particular situation, such as a comparative relation, determines the type and structure of the linguistic construction used by those people in talking about that situation. Heine (1994) also shows that there are clear areal distribution patterns (that cross genetic lines) for the different event schemas behind the different comparative construction types. Heine's conclusion is that 'areal distribution plays a major role in the cognitive patterning underlying the development of comparative constructions in the languages of the world, and areal distribution is suggestive of massive linguistic and cultural communication' (Heine 1994:66). That is, because of massive contact, the speakers of the languages of an area come to construe an aspect of the world in the same way, that is, share the same event schema, and this leads them to have similar linguistic constructions for representing that schema. The influence of language contact on language development is then not always directly linguistic. Learning another language means learning to think in a different way, or to construe the world in a different way, and this may then affect our native language. This is often what calquing is. Calquing is not necessarily direct linguistic influence, the way loan words are. It is often the result of influence in the way people construe events or situations. Substratum effects can also be of this type, that is, the effect of a way of thinking or the effect of deep-seated habits of language use. That is, if our native language obligatorily constrains the interpretation of some functional domain, when we learn a second language, we will tend to want to constrain the interpretation of that domain in the new language. For example, in Taiwan Mandarin we regularly find a complementiser *suɔʔ<sup>55</sup>* (= 'to say'), which is due to the fact that the majority of the speakers of Taiwan Mandarin speak Southern Min Chinese as their first language, and this language has a complementiser *kɔŋ<sup>53</sup>* (= 'to say') which helps to constrain the interpretation of complements. When speaking Mandarin, the Min speakers felt the need for such a complementiser because of their habit of constraining the interpretation in this way in their own language, and so created a comparable one based on the Mandarin word for 'say'. This is filling a perceived gap. The same sort of thing happens when English speakers learn Chinese. I mention below that in Chinese no genitive phrase is necessary in an expression that would translate as 'I washed my hair', but English speakers learning Chinese often will add a genitive phrase in that context when speaking Chinese because they feel it is needed to constrain the interpretation. In a similar way, due to their habit of marking tense in every finite clause, English speakers learning Chinese will overuse the perfective aspect marker in Chinese, essentially using it in any situation they would normally use a past tense in English. This is because they feel the need to constrain the interpretation of the utterance by marking it as past tense, but as Chinese does not have tense marking, they use the closest thing they can find (the perfective marker) to fill the

perceived gap. The feeling of needing to constrain the interpretation in a particular way may also come from outside one's own language. For example, the third person pronoun in Chinese does not inflect for animacy or gender, but in the early 20th century many Chinese intellectuals learned English, French, or German, and came to feel the need to constrain, at least in writing, the interpretation of the referent of the third person pronoun, and so developed different ways of writing the third person pronoun in Chinese for male, female, inanimate, and godly referents.

The spread of borrowed words and borrowed patterns is the same process as that for native words and patterns. Borrowed words and patterns may introduce new concepts/tools, but the ultimate meaning of the word or pattern will be determined by the use to which it is put. For example, the English word *gungho* is a loan of Chinese *kun<sup>55</sup> xy<sup>35</sup>* 'industrial cooperative' in Chinese, but in English it means 'to have great spirit or enthusiasm' (derived from the enthusiastic spirit which American soldiers felt characterised the workers in China's early industrial cooperatives). This is the same with other tools. In the Philippines, a fork and a spoon are often used when eating, due to Western influence, but instead of the fork being used to move the food to the mouth, the fork is used only to push the food onto the spoon, and then the spoon is used to carry the food to the mouth. That is, the 'use/meaning' of the fork in that system is different. Also, a set of conventions is a system, and sometimes if you try to change one aspect of a system, you have to change others because of a mismatch. For example, in Taiwan now it is common to use large plates to hold the rice when eating (instead of the traditional bowls), due to Western influence, but they still use chopsticks to eat, and as these two tools are not very compatible, the Chinese spoon is used to take the food off the plate, where it can then be better accessed by the chopsticks. That is, they use the spoon to pick up the food off the plate, and then use the chopsticks to eat the food off the spoon.

Our language use is a set of habits we form, and these habits are very hard to change. We are very much creatures of habit, and once we have a habit, it is hard to change, including habits of language and even thought. The most simple example is the habit we form in learning our first language: we learn to categorise certain sounds together as allophones of a single phoneme, and to distinguish among other sounds our language treats as distinct phonemes. This is entirely a habit, but as anyone who has learned a second language (or taken a class in phonetics) knows, it is difficult to break the habit and make distinctions we are not used to making. The habit even influences our perception, as (for example) a native English speaker will really 'hear' a voiceless unaspirated stop as if it is the same sound as a voiced stop (for example, hear the initial sound in *pei<sup>214</sup> tciŋ<sup>55</sup>* 'Beijing' as /b/). Another good example is phonotactics. There is a set of permissible syllable types in English, and the habit of speaking those types and only those types is so strong that when a writer makes up a new syllable, it will invariably conform to that template (Whorf 1940[1956]). This is also what is involved in second language learner

accents. The point is not that you cannot learn another set of phonotactic constraints, just that it is difficult because it is an ingrained habit.

## 5 How languages differ

It has been said that languages differ not so much in what they can say, but in what they must say. This is looking at it from the speaker's point of view. From the hearer's point of view, we can say that languages differ not so much in what can be understood, but in what must be understood. All languages can constrain the interpretation of just about any functional domain, but most languages have developed obligatory grammatical marking that obligatorily constrains the interpretation in certain functional domains to some extent. Which domains the speakers of a language will choose to constrain, and how they constrain the interpretation, are the two major ways languages differ from each other. Just as societies differ as to what tools they use for a particular activity, for example using chopsticks as opposed to using the hands or a fork for eating, and these tools can vary in terms of specificity (for example, Chinese people traditionally use fewer specialised tools for eating than Westerners), the tool we think of as language can differ between cultures in terms of how specialised its structures are. I would like to turn now to some examples of the ways that languages can differ in terms of specialisation, and show how this relates to interpretation.

For a number of years I have been arguing that Chinese and most other Sino-Tibetan languages do not work the same way, in terms of pivots and grammatical relations, as either languages with largely nominative-accusative structure, such as English, or those that have largely ergative structure, such as Dyirbal (LaPolla 1988, 1990, 1993, 1995a, 1996, 2002; Van Valin and LaPolla 1997, Ch. 6). For example, in a language with an [S, A] pivot for coordination (the accusative pattern), such as English, an argument shared by two conjoined non-passive clauses can be represented by a zero pronoun in the second clause only if it is in the A or S role in both clauses, as in (6a).

- (6)a. The man went downhill and Ø saw the dog.
- b. \*The dog went downhill and the man saw Ø.
- c. The dog went downhill and Ø was seen by the man.

It is not possible to have the representation of the actor of the first clause coreferring with a zero pronoun representing the undergoer (O role argument) of the second clause without using a passive construction, as shown in (6b). It is not possible to say \**The dog went downhill and the man saw*. If the argument the two clauses have in common is the undergoer of the second clause, in order for the two clauses to be conjoined, the representation of the argument (here the zero pronoun) must appear as the single direct argument of a passive construction, as in (6c).

In a language with an [S,O] pivot for coordination (the ergative pattern), such as Dyirbal (Dixon 1980:461ff), a shared argument which appears as a zero pronoun in the second of two conjoined clauses must be in the S or O role in each clause, as in (7a). If the argument in the second clause is instead in the A role, in order for the two clauses to be conjoined and for the argument to be represented by a zero pronoun in the second clause, the shared argument must appear as the single direct argument of an antipassive construction, as in (7b). It is not possible to say the equivalent of *The man went downhill and saw the dog* with a transitive second verb and a zero anaphor referring to an A argument, as in (7c) (from Dixon 1980:461-2).<sup>8</sup>

- (7) a. *Balan guda buŋa-n bangul yara-ŋgu bura-n.*  
 she+ABS dog+ABS descend-PAST he+ERG man-ERG see-PAST  
 'The dog went downhill and was seen by the man.'  
 (Lit.: The dog went downhill and the man saw  $\emptyset$ .)
- b. *Bayi yara buŋa-n bulralŋanyu bagun gudagu.*  
 he+ABS man+ABS descend-PAST see+PAST+ANTI he+ABS dog+DAT  
 'The man went downhill and saw the dog.'  
 (with antipassive indicator  $\geq a$ -y on the second verb).
- c. \**Bayi yara buŋa-n bura-n bangul guda.*  
 he+ABS man+ABS descend-PAST see-PAST he+ERG dog+ABS  
 'The man went downhill and saw the dog.'  
 (with transitive verb and A argument (*yara $\geq$ gu*) unexpressed).

In Chinese we don't find either the English or the Dyirbal type of restriction on cross-clause coreference. In Chinese it is possible for the shared argument of a conjoined structure to be deleted regardless of whether it is in the A or O role, as we can see from the examples in (8):

- (8) a. *Xiǎo-gǒu; zǒu dào shān-dìxià, nèi-ge-rén jiù kànjiàn-le  $\emptyset$ ;*  
 little-dog walk to mountain-bottom that-CL-person then see-PFV  
 'The little dog went downhill and was seen by the man.'  
 (Lit.: 'The little dog went downhill and the man saw  $\emptyset$ .)'
- b. *Nèi-ge-rén; zǒu dào shān-dìxià, jiù  $\emptyset$  kànjiàn-le xiǎo-gǒu.*  
 that-CL-person walk to mountain-bottom then see-PFV little-dog  
 'The man went downhill and saw the little dog.'

<sup>8</sup> Abbreviations used in the examples: 1, 2, 3 1st, 2nd, 3rd person; ABS absolutive; AGT agentive; ANTI antipassive; CL classifier; CSM change of state; DAT dative; DIR directional; ERG ergative; LOC locative; INF inferentially derived conclusion; INTRANS.PAST third person intransitive past; NPAST non-past declarative; PFV perfective; PL plural; PROG progressive; PS predicate sequence; R/M reflexive/middle; SG singular; TMDYS past tense, 1 day-1 year ago; TMHRS past tense, within today; TMYRS past tense, years ago; TRANS.PAST 3rd person transitive past.

The result of this situation is that in languages with grammatical constraints on the control of anaphor like those we've just looked at, those constraints force a particular interpretation of an utterance. For example, if the words 'The man saw the dog and went downhill' were said in English, the interpretation would have to be that the MAN went downhill; but if the equivalent words were used in Dyirbal (*balan guda bangul yaranngu buran bujan*), the meaning would have to be that the DOG went downhill. This forcing of the interpretation holds even if the resulting interpretation would be nonsensical given what English speakers normally assume about the world, as in (9), from Comrie (1988:191):

- (9) The man dropped the melon and burst.

Because of the grammatical constraint on conjunction reduction in English, this sentence has to be interpreted as saying that the man burst after dropping the melon. That is, when there is a coordinate structure such as this, the rules of English syntax force the interpretation that the zero pronoun is coreferential with the S or A role argument of the first clause, and block the inclusion in the context of interpretation of the assumption that when someone drops something, it is more likely that the thing dropped bursts rather than the person doing the dropping. In a language such as Chinese, though, where there is no such grammatical constraint on interpretation, the equivalent sentence would not force such an interpretation, even with the man being the topic of the utterance, as assumptions from real world experience would be included in the context of interpretation and influence the interpretation more than the syntactic structure. Over the years I have asked well over a hundred native speakers of Chinese to translate this sentence into Chinese and then tell me who or what burst. The answer is invariably 'Of course the melon burst.' They are generally quite surprised when I tell them that the English sentence MUST mean that the man burst.

In Rawang, a Tibeto-Burman language spoken in Northern Burma, we have the same lack of constraints on the interpretation of clause coordination, as evidenced by the pair of sentences in (10):

- (10) a. *Vpūngí Vdḥs̀̀ng vdip bḥà nḥ ngḥa:pmì*  
*Vpūng-í Vdḥ-s̀̀ng vdip bḥ-à nḥ ngḥ-ap-ì*  
 Apung-AGT Adeu-LOC hit PFV-TRANS.PASTPS cry-TMDYS-INTRANS.PAST  
 'Apung hit Adeu and cried.' (Adeu cried)
- b. *Vpūngí Vdḥs̀̀ng vdip bḥà nḥ vḥḥshì a:pmì*  
*Vpūng-í Vdḥ-s̀̀ng vdip bḥ-à nḥ vḥḥ-shì*  
 Apung-AGT Adeu-LOC hit PFV-TRANS.PAST PS laugh-R/M  
*ap-ì*  
 TMDYS-INTRANS.PAST  
 'Apung hit Adeu and laughed.' (Apung laughed).



Here the structures are exactly the same, though the actor of the second clause is interpreted differently due to real world expectations of who would be more likely to cry or laugh after an act of hitting. In fact the interpretation is quite unrestrained; although I've written 'Adeu cried' and 'Apung laughed' after the free translations, actually the interpretation could be that the one who cried or laughed was either one of these two people, or even a third person, such as someone standing nearby watching what was happening between Adeu and Apung. Most Sino-Tibetan languages are similar to Chinese and Rawang in not having syntactic constraints that force particular interpretations of cross-clause coreference.

Let's look at some other ways that the grammar of English constrains interpretation. One way is with verb agreement. Aside from the obvious effect that verb agreement has on the identification of particular arguments, it can also constrain the interpretation of the syntactic structure. To borrow an example from Green (1996:144), the use of singular versus plural agreement in (11a) and (11b) forces two different analyses of the structures. In (11a) pickles and ice-cream must be interpreted as two different items about which the same predication is made, while in (11b) they must be interpreted as one item (a dish with two things combined) about which a predication is made.

- (11) a. Pickles and ice cream are really great.  
 b. Pickles and ice cream is really great.

In Chinese it is not possible to constrain the interpretation in this way, as there is no agreement marking, so there would be only one form for both these meanings in Chinese; the inferential process involved in deciding on the proper structure (and therefore the proper interpretation) would not be constrained by the linguistic form in the way that it is in English.

In terms of whether a language constrains the interpretation of the relations between elements of a complex clause structure or not, we can give the example of verb juxtaposition in Lahu. Matisoff (1991:403) gives an example with the verb *qô* 'hoe' in simple juxtaposition with 12 other verbs, and contrasts the use of this one syntactic form (simple juxtaposition) in Lahu with the use of six different types of construction for expressing the same relations in English (see (12)). There is nothing in the grammar of this Lahu construction that constrains the interpretation of the relationship between the two verbs, while in English the interpretation is constrained to a greater degree by the different constructions used.

(12) complementary infinitives	<i>qô ša</i> 'easy to hoe'	<i>ga qô</i> 'help to hoe'
-ing complements	<i>qô kî</i> 'busy hoeing'	<i>tà qô</i> 'start hoeing'
modal auxiliaries	<i>qô cã</i> 'should hoe'	<i>ğa qô</i> 'must hoe'
adverbs	<i>qô bà</i> 'hoe away'	<i>qô? qô</i> 'hoe again'
prepositional phrases	<i>qô pî</i> 'hoe for smn'	<i>phô? qô</i> 'hoe in a group'
subordinate clauses	<i>qô ni</i> 'hoe and see'	<i>ca qô</i> 'go and hoe'

Tense marking also restricts the search for the relevant interpretation. For example, to interpret the proper time frame for the situation expressed by the Chinese sentence in (13a), the hearer must depend on inference based on the context, whether overall what is being talked about is something that happened in the past or a current situation. In English, though, as English has grammaticalised obligatory tense marking, the equivalent of (13a) would be (13b), (13c), or (13d), all of which constrain the interpretation of the time frame. (As can be seen from this example, the identification of the gender of the referent (and therefore the identification of the referent) of some pronouns is also constrained by the form of the pronoun, and this too in Chinese is unconstrained.)

- (13) a. *Tā qù xuéxiào.*  
 3SG go school  
 b. She/He went to school.  
 c. She/He is going to school  
 d. She goes to school./He goes to school.

We can see that compared to Chinese, English obligatorily constrains the interpretation of the time frame, limiting the identification to either a past or non-past situation, but within those broad categories, say, for example given a past tense form, to determine how far in the past the action was the interpreter of the utterance must rely on linguistically unconstrained inference. That is, if I say *I have had lunch*, then you will probably draw the inference that I ate within the last hour or two, or at least within today; if I say *I have been to the doctor*, then you may make the inference that it was within the last few days; if I say *I have been to Tibet*, then you will not make the inference that it was within the last one or two hours, or even within the last few days, as it could have been quite some time ago, but the differences among these three interpretations are not due to anything in the grammatical structure, they are due purely to inferences based on the real-world understanding of the actions involved. The search for the proper interpretation of the length of time from an overtly marked past action to the time of the speech act is not further constrained grammatically in English. If we then compare English to Rawang, we can see that in Rawang there is a four-way past tense system which marks whether the action took place an hour or two ago, a few hours ago but within this day, sometime from yesterday up to a year ago, or more than a year ago. The examples in (14) all are of the verb *dī* 'to go'.

- (14) a. *àng dī á:m-ì*  
 3SG go DIR-INTRANS.PAST  
 'S/he left, went away (within the last 2 hours).'  
 b. *àng dī dár-ì.*  
 3SG go TMHRS-INTRANS.PAST  
 'S/he went (within today, but more than two hours ago).'

- c. àng dī ap-mì.  
3SG go TMDYS-INTRANS.PAST  
'S/he went (within the last year).'
- d. àng dì yàng-ì.  
3SG go TMYRS-INTRANS.PAST  
'S/he went (some time a year or more ago).'

The point is that languages differ quite a lot in how much they constrain the search for the most relevant interpretation, and in what aspects they choose to constrain. As can be seen from these examples, while Rawang constrains the interpretation of the time frame more than English, it does not constrain the search for the referent of a pronoun as much as English does (and we saw Rawang does not have the cross-clause coreference constraint that English has). From this we can see that we can not talk about **languages** as being more or less grammaticalised or their interpretation more or less constrained, only particular **functional domains** being more or less grammaticalised or their interpretation more or less constrained in a certain language.

An interesting three-way contrast of what is or is not left to inference in different languages can be seen from a comparison of Chinese, Tagalog and English. The normal way of saying 'Let's go' in Chinese involves just a verb and a particle, as in (15a), and only the use of the hortative particle constrains the interpretation of the actor referent (so it could be 'you go' or 'we go', but not 'he goes'); in Tagalog, as in (15b), it is normal to just say *Tayo na*, which is the 1st person plural inclusive pronoun plus a change of state marker, with no verb, and leave the interpretation of the action suggested unconstrained (it could mean 'Let's go' or 'It's our turn'), while in English both the pronoun and the verb must be specified, so the interpretation of the actor and the action are both obligatorily constrained.

- (15) a. Zǒu ba!  
go HORTATIVE.PARTICLE  
'Let's go.' or '(Why don't) you go.'
- b. *Tayo na!*  
1PL.INCL CSM  
'Let's go.' or 'It's our turn.'

Languages can also differ in terms of the type of grammaticalisation used to constrain the interpretation of a particular functional domain. For example, in the Chinese sentence in (16a), there is no marking to constrain the interpretation of whose hair is being washed, and so the determination of this relationship is purely a matter of inference; in most contexts it would mean the person is washing his or her own hair, but given the right circumstances (such as a professional hair-washer in a barber shop) it could mean the person is washing someone else's hair. In both English and Rawang, on the other hand, the interpretation of whose hair is being washed is obligatorily constrained, but in different

ways. In English, example (16b), the NP which has *hair* as its head must include a possessive pronoun, but in Rawang, example (16c), there is no marking on the NP; the interpretation is constrained by the obligatory use of the reflexive/middle marking (see LaPolla 2000).

- (16) a. *Tā zài xǐ tóufǎ.*  
 3SG PROG wash hair  
 'S/he is washing (her/his) hair.' (Lit.: 'S/he is washing hair.')
- b. He is washing his hair.
- c. *àng nī zǐl-shì-ě.*  
 3SG hair wash-R/M-NPAST  
 'S/he is washing her/his hair.'

In both English and Rawang the interpretation of the relationship between the actor and the undergoer is constrained, but by very different grammatical categories.

## 6 Summary and implications of this view of language

A speaker (communicator) performs an ostensive act in order to communicate. This gets the attention of the hearer (interpreter), and the hearer must first infer that the speaker has a communicative intention and that it is directed at the hearer. Then the hearer must infer the reason for the communicator to make that particular ostensive act in the context of the communicative activity. All of this is done using inference; all aspects of interpretation involve the creation of a set of assumptions, a context, which can be added to whatever part of the signal or message has been recovered up to that point (it is a dynamic process) to deduce the most likely form and possible motivation for its production. This inference is possible because of the assumption that an ostensive act involves a guarantee of relevance, and that the communicator will choose the form for the ostensive act that will most likely lead to the intended interpretation. Because of this assumption, the speaker must tailor the ostensive act in such a way that the hearer will not have to expend unnecessary effort to create a context that will allow him/her to arrive at the intended interpretation. In doing this, the speaker takes into consideration guesses as to what information is available to the hearer at the time of utterance for use in interpreting the utterance. The speaker may constrain the hearer's construction of the context of interpretation in many ways. The most straightforward reflection of this constraining process is the amount of lexical content that the communicator includes in the utterance. Grammaticalised marking (including intonation) can also be used to help the hearer process the utterance by constraining the search for relevant assumptions to include in the context of interpretation. The grammatical marking performs the same role in constraining or guiding the interpretation of the utterance that an increase in the number of lexical items can have. Lexicalisation and grammaticalisation is the conventionalisation of repeated

patterns of language used for constraining the context of interpretation in a particular way. As each society of language users determines what functional domains will constrain and how it will do it, languages come to differ in terms of what will be constrained and what will not, in terms of the degree to which interpretation of a particular functional domain is constrained, and how particular functional domains are constrained. This is why we get the variety we do among the languages of the world. As individual speakers also have some choice in what they constrain and how they do it (and develop their own habits, parallel to societal conventions), we also get variety from speaker to speaker within a particular society of language users.

What I am arguing for involves a different way of viewing structure. Rather than assuming that language structures are the building blocks of relatively effortless deterministic interpretation and treating ambiguous expressions as aberrant, we should assume that forms used in communication are inherently indeterminate (Reddy 1979; Grace 1987), and look at structure from the point of view of how it constrains interpretation, that is, how interpretation is made more determinate by, for example, the grammaticalisation of subject or other grammatical categories. Most linguistic studies, even many of those that use natural language data rather than made-up sentences, still take the grammar as given, and only look for the 'interface' between semantics and syntax or pragmatics and syntax. For many, such as Susumo Kuno (for example 1987) and Ellen Prince (for example 1988), pragmatics is simply another module of the package, and not the foundation of communication and therefore of grammar.<sup>9</sup> The view I am presenting here is that the fundamental aspect of communication is not the linguistic structure, but the interaction of the speaker and hearer in performing a communicative activity. The role of the context in the performance of this activity involving the interpretation of utterances is not to simply supplement semantic meaning; the context is the base on which all communicative activity depends. That is, rather than saying that the context constrains the interpretation of the linguistic form, I argue that it is the linguistic form that constrains the context (that is, constrains the creation of the context of interpretation).

To take one example of what I mean by looking at grammar in a different way, we can look at Ekkard König's (1995) excellent study of the meaning of converb constructions. This paper focuses on how the converb constructions are vague and so need to be enriched by contextual factors. That is, König takes the form of the converb construction as something basic and then tries to see how contextual factors help us to interpret the meaning of the converb construction. He says that general background assumptions and contextual information and general principles of language use 'make an important contribution to an interpretive enrichment of the nonspecific basic meaning of converbs.' (p. 83). An alternative possibility is to look at the utterance and try to interpret the speaker's communicative intention, and see how the use of a particular structure, such as a

---

<sup>9</sup> For arguments against the modular view of pragmatics, see Wilson and Sperber (1986).

converb construction, constrains our search for the proper interpretation of the speaker's intention, that is, how the use of a particular grammatical form constrains our search through general background assumptions and contextual information and general principles of language use in order to help us create the context of interpretation that will lead to the intended interpretation. Rather than taking the grammatical form as basic and trying to interpret its meaning in different contexts, we should see inference as the basis of communication, and try to determine how a particular grammatical form develops to aid the hearer in constructing the proper context in which the ostensive act achieves relevance.

The view of grammar I am presenting here means not trying to define what, for example, a 'subject' is, the way Keenan (1976) did, assuming it is some sort of 'thing', but seeing what we call 'subject' in English as a set of constraints on the interpretation (identification) of referents in certain syntactic constructions such as clause coordination, etc., and seeing which constraints individual languages have or have not grammaticalised as part of their grammatical system (see Van Valin and LaPolla 1997, Ch. 6).<sup>10</sup> It also means not inventing covert movements and structures to try to explain all differences of interpretation as differences in syntactic structure.

One consequence of this view is that there is no difference in quality or type between lexicalisation and grammaticalisation. Both are processes of conventionalisation, and differ only in the generalness of application. That is, lexicalisation affects only a single specific item (whether long or short), whereas grammaticalisation applies more generally to a class of items. This forms something more like a continuum, rather than discrete categories. This view also implies that much of language use involves recall of complete forms, including sentences, from memory rather than pure generation of totally new forms, as these remembered forms are what become fixed syntactic patterns (constructional schemata). As with so many other things, Bolinger (1961, 1976; see also Pawley 1985, Grace 1987) was ahead of his time when he argued for something like schemata, what he

---

<sup>10</sup> In earlier papers (LaPolla 1990, 1993, 1996) I have compared patterns of syntactic behavior in Chinese with those in accusative, ergative, active, and Philippine type languages, and have shown that Chinese does not pattern like any of those systems. I have argued this is because Chinese has not grammaticalized a syntactic pivot for any of its constructions. Chinese therefore should not be considered accusative, ergative, active, or of the Philippine type, but is it another type, possibly called a 'neutral' type, or is it a non-type? Given the facts mentioned above, and others of a similar nature, the tendency has been to see Chinese as another syntactic type, to try to make a syntactic relation out of topic or topic chain (Huang 1989, Shi 1989, Her 1991), or to see 'topic prominence' as a syntactic type in opposition to 'subject prominence' (as many have done based on Li and Thomson's (1976) original proposal of these concepts). I would like to argue instead that a lack of evidence of constraints such as we find in Chinese is precisely that, a lack of constraints. When we say 'type', we mean a set of constraints of a certain type, and if a type is a set of constraints, then the lack of evidence of constraints in Chinese is evidence of the lack of a type, not a separate type. There are ways that Chinese has grammaticalized constraints that English has not, such as numeral classifiers, but in terms of the constraints associated with subject in English, Chinese simply has not grammaticalized them, and so what we get is less constrained inference.

called 'idioms', and combinations of schemata, what he called 'syntactic blends' to form new syntactic structures, and attempted to show 'the permeation of the entire grammatical structure by threads of idiom' (1961:366). He argued against a purely generative view of grammar, suggesting that our use of grammar was partly creative and partly a matter of memory:

At present we have no way of telling the extent to which a sentence like *I went home* is a result of invention, and the extent to which it is a result of repetition, countless speakers before us having already said it and transmitted it to us in toto. Is grammar something where speakers 'produce' (i.e. originate) constructions, or where they 'reach for' them, from a preestablished inventory, when the occasion presents itself? ... Probably grammar is both of these things ... (Bolinger 1961:381).

As Matisoff (1979[2000]:xv) says in talking about the prepatterned, collocational nature of language use,

Entire conversations can be made up of formulaic expressions so naturally that the interlocutors are not disturbed by their lack of "generative originality"—indeed, quite to the contrary: there is great comfort and security to be derived from fitting into a well-worn communicative groove.<sup>11</sup>

The book in which this quote appears is a collection of hundreds of Yiddish psychostensive expressions, showing just how prepatterned and collocational language use can be. A corollary of this view of language is that there are then no clear lines between lexicon, morphology and syntax, as they form continua of generalness and rigidity (the degree to which they are fixed) (see Bolinger 1976:3; cf. also Langacker 1987).

Looking at language this way makes possible explanations not only of why a particular type of marking develops, but also of why the use of marking that has already developed becomes extended in predictable ways, such as the development of agentive marking from ablative marking or the extension of reflexive marking to middle situations (see LaPolla 1995b). The development is in the direction of greater specificity and a more constrained set of possible interpretations, utilising resources already present in the language when possible.

This view of language development also has a number of other important implications for linguistic theory. I will mention three here:

- (a) As languages differ in terms of constraining interpretation, both in terms of constraining or not constraining a particular type of interpretation (functional domain), and also in the degree to which the interpretation is constrained and how it is constrained, the differences between languages are gradient differences, not simple parameters.

---

<sup>11</sup> See also Aijmer (1996) on preset conversational routines.

- (b) As these constraints are the result of grammaticalisation, they are therefore not genetically hard-wired.
- (c) (a second order conclusion) The human language ability then can not be an autonomous and genetically programmed module; language development and use must be based on general cognitive structures. In short, from the point of view presented in this paper, saying that there are genetically determined parameters for language features, such as [ $\pm$  configurational], makes no more sense than saying that there are genetically determined parameters for other conventionalised behaviors, such as [ $\pm$  necktie-wearing].

## References

- Aijmer, Karin, 1996, *Conversational routines in English: convention and creativity*. London & New York: Longman.
- Aikhenvald, Alexandra Y. and Robert Malcolm Ward Dixon, 1998, Dependencies between grammatical systems. *Language* 74/1:56–80.
- Barlow, Michael and Suzanne Kemmer, 1994, A schema-based approach to grammatical description. In Susan D. Lima, Robert L. Corrigan and Gregory K. Iverson, eds *The reality of linguistic rules*, 19–42. Amsterdam & Philadelphia: John Benjamins.
- Blakemore, Diane, 1987, *Semantic constraints on relevance*. Oxford: Basil Blackwell.
- 1988a, 'So' as a constraint on relevance. In Ruth M. Kempson ed. *Mental representations: the interface between language and reality*, 183–195. Cambridge: Cambridge University Press.
- 1988b, The organization of discourse. In Newmeyer ed. 1988, vol.II, 229–250.
- 1990, Constraints on interpretations. *Berkeley Linguistics Society* 16:363–370.
- Bernstein, Basil J., 1971, *Class, codes and control*. London: Routledge and Kegan Paul.
- Bolinger, Dwight L., 1961, Syntactic blends and other matters. *Language* 37:366–381.
- 1976, Meaning and memory. *Forum Linguisticum* 1.1:1–14.
- Comrie, Bernard, 1988, Coreference and conjunction reduction in grammar and discourse. In John Hawkins ed. *Explaining language universals*, 186–208. Oxford: Basil Blackwell.
- Dixon, Robert Malcolm Ward, 1980, *The languages of Australia*. Cambridge, London, & New York: Cambridge University Press.
- Du Bois, John W., 1985, Competing motivations. In John Haiman ed. *Iconicity in syntax (Typological Studies in Language 6)*, 343–365. Amsterdam & Philadelphia: John Benjamins.
- Errington, J. Joseph, 1988, *Structure and style in Javanese*. Philadelphia: University of Pennsylvania Press.



- Givón, Talmy, 1979a, *On understanding grammar*. New York, San Francisco & London: Academic Press.
- 1979b, From discourse to syntax: grammar as a processing strategy. In Talmy Givón, ed., 81–112.
- Grace, George C., 1987, *The linguistic construction of reality*. London, New York & Sydney: Croom Helm.
- Green, Georgia, 1996, *Pragmatics and natural language understanding* 2nd edition. Mahwah NJ: Lawrence Erlbaum Associates.
- Gumperz, John J., 1977, Sociocultural knowledge in conversational inference. In Muriel Saville-Troike ed. *Linguistics and anthropology, Georgetown University round table on languages and linguistics 1977*, 191–211. Washington DC: Georgetown University Press.
- 1982, *Discourse strategies*. Cambridge: Cambridge University Press.
- 1989, Contextualization cues and metapragmatics: The retrieval of cultural knowledge. *Parasession on Language in Context. Chicago Linguistic Society 25/2:77–88.*
- 1992a, Contextualization and understanding. In Alessandro Duranti and Charles Goodwin eds *Rethinking context*, 229–252. Cambridge: Cambridge University Press.
- 1992b, Contextualization Revisited. In Peter Auer and Aldo di Luzio eds *The contextualization of language*, 39–53. Amsterdam & Philadelphia: John Benjamins.
- Gumperz, John J., Gurider Aulakh, and Hanna Kaltman, 1982, Thematic structure and progression in discourse. In John J. Gumperz ed. *Language and social identity*, 22–56. Cambridge: Cambridge University Press.
- Halliday, Michael A.K., 1994, *An introduction to functional grammar* 2nd edition. London & New York: Edward Arnold.
- Heine, Bernd, 1994, Areal influence on grammaticalization. In Martin Pütz ed. *Language contact language conflict*, 55–68. Amsterdam & Philadelphia: Benjamins.
- 1997a, *Cognitive foundations of grammar*. New York & Oxford: Oxford University Press.
- 1997b, *Possession: cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.
- Her, One-Soon, 1991, Topic as a grammatical function in Chinese. *Lingua* 84:1–23.
- Hopper, Paul, 1987, Emergent grammar. *Berkeley Linguistics Society* 13:139–55.
- 1988, Emergent grammar and the A Priori Grammar Postulate. In Deborah Tannen ed. *Linguistics in context: connecting observation and understanding*, 117–134. Norwood NJ: Ablex.
- Huang, C.-R., 1989, Subcategorized topics in Mandarin Chinese. Paper presented at the 1989 CLTA Annual Meeting, Boston MA, Nov. 17–19.

- Huang, C.-T. James, 1984, On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15/4:531–574.
- Keenan, Edward L., 1976, Towards a universal definition of 'subject'. In C. N. Li ed., 305–333. New York: Academic Press.
- Keller, Rudi, 1994, *On language change: the invisible hand in language*. Translated by Brigitte Nerlich. London: Routledge.
- Keimiger, Suzanne, 1993, *The middle voice. Typological Studies in Language* 23. Amsterdam & Philadelphia: John Benjamins.
- König, Ekkehard, 1995, The meaning of converb constructions. In Kartin Haspelmath and Ekkehard König eds *Converbs in cross-linguistic perspective*, 57–95. Berlin & New York: Mouton de Gruyter.
- Kuno, Susumu, 1987, *Functional syntax: anaphora, discourse, and empathy*. Chicago: University of Chicago Press.
- Langacker, Ronald W., 1987, *Foundations of cognitive grammar, vol.I: Theoretical prerequisites*. Stanford: Stanford University Press.
- LaPolla, Randy J., 1988, 'Subject' and referent tracking: Arguments for a discourse-based grammar of Chinese. In Joseph Emonds, P.J. Mistry, Vida Samiian and Linda Thornburg eds *Proceedings of the west coast conference on linguistics I*:160–173. Dept. of Linguistics, California State University, Fresno.
- 1990, Grammatical relations in Chinese: Synchronic and diachronic considerations. PhD dissertation, University of California, Berkeley.
- 1993, Arguments against 'subject' and 'direct object' as viable concepts in Chinese. *Bulletin of the Institute of History and Philology* 63/4:759–813.
- 1995a, Pragmatic relations and word order in Chinese. In Pamela Downing and Michael Noonan eds *Word order in discourse*, 297–329. Amsterdam & Philadelphia: John Benjamins.
- 1995b, On the utility of the concepts of markedness and prototypes in understanding the development of morphological systems. *Bulletin of the Institute of History and Philology, Academia Sinica* 66/4:1149–1185.
- 1995c, review of Diana Blakemore, *Understanding utterances: an introduction to pragmatics*. *Language* 71/1:173–176.
- 1996, On the place of Chinese in the typology of grammatical relations systems. Paper presented to the Linguistics Seminar of the Dept. of Chinese, Translation, and Linguistics, City University of Hong Kong Oct. 31, 1996.
- 1997, Grammatical sketch of the Qiang language, with texts and annotated glossary. ms., City University of Hong Kong.

- 2000, Valency-changing derivations in Dulong/Rawang. In Robert Malcolm Ward Dixon and Alexandra A. Aikhenvald eds *Changing valency: case studies in transitivity*, 282–311. Cambridge: Cambridge University Press.
- 2002, A problem with word order typology: Differing uses of word order among languages. Talk presented at the Research Centre for Linguistic Typology, Institute for Advanced Study, La Trobe University, Dec. 4, 2002.
- To appear, *A grammar of Qiang*. Berlin: Mouton de Gruyter.
- Lass, Roger ed., 1992, *The Cambridge history of the English language*, vol.III 1476–1776. Cambridge: Cambridge University Press.
- Li, Charles N. ed., 1976, *Subject and topic*. New York: Academic Press
- Li, Charles N. and Sandra A. Thompson, 1976, Subject and topic: A new typology of language. In Charles N. Li ed., 459–489. New York: Academic Press.
- Matisoff, James A., 1979, *Blessings, curses, hopes, and fears: psycho-ostensive expressions in Yiddish*. Philadelphia: ISHL. Second edition 2000 Stanford, California: Stanford University Press.
- 1991, Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth Traugott and Bernd Heine eds *Approaches to grammaticalization*, vol.II, 383–453. Amsterdam & Philadelphia: John Benjamins.
- Newmeyer, Frederick K. ed., 1988, *Linguistics: the Cambridge survey*. Cambridge: Cambridge University Press.
- Nicolle, Steve, 1997, Conceptual and procedural encoding: criteria for the identification of linguistically encoded procedural information. In Marjolein Groefmsma ed. *Proceedings of the University of Hertfordshire relevance theory workshop, 27–29 Oct. 1995*, 47–56. Chelmsford: Peter Thomas & Assoc.
- Ono, Tsuyoshi and Sandra A. Thompson, 1995, What can conversation tell us about syntax? In Philip W. Davis ed. *Alternative linguistics*, 213–272. Amsterdam & Philadelphia: Benjamins.
- Pan, Haihua, 1998, Thematic hierarchy, causative hierarchy, and Chinese resultative verb compounds. Paper presented at the annual research forum of the Linguistic Society of Hong Kong, the Polytechnic University of Hong Kong, 5–6 December.
- Pawley, Andrew, 1985, Lexicalization. In Deborah Tannen and James E. Alatis eds *Language and linguistics: the interdependence of theory, data, and application*, 98–120. Washington DC: Georgetown University Press.
- Pawley, Andrew and Jonathan Lane, 1998, From event sequence to grammar: Serial verb constructions in Kalam. In Anna Siewierska and Jae Jung Song eds *Case, typology, and grammar*, 201–227. Amsterdam & Philadelphia: John Benjamins.

- Perkins, Revere D., 1980, The evolution of culture and grammar. PhD dissertation, State University of New York at Buffalo.
- Prince, Ellen F., 1988, Discourse analysis: A part of the study of linguistic competence. In Newmeyer ed. vol.II:164–182.
- Pyles, Thomas and John Algeo, 1982, *The origins and development of the English language* 3rd edition. New York: Harcourt, Brace, Jovanovich.
- Reddy, M. J., 1979, The conduit metaphor—a case of frame conflict in our language about language. In A. Ortony ed. *Metaphor and thought*, 284–324. Cambridge: Cambridge University Press.
- Shi, Dingxu, 1989, Topic chain as a syntactic category in Chinese. *Journal of Chinese Linguistics* 17/2:223–262.
- Sperber, Dan and Deirdre Wilson, 1986, *Relevance: communication and cognition*. Oxford: Basil Blackwell. (2nd edition 1996 Cambridge MA: Harvard University Press).
- Stassen, Leon, 1985, *Comparison and universal grammar*. Oxford: Blackwell.
- Traugott, Elizabeth C., 1982, From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Winifred P. Lehmann and Yakov Malkiel eds *Perspectives on historical linguistics*, 245–271. Amsterdam & Philadelphia: John Benjamins.
- 1988, Pragmatic strengthening and grammaticalization. *Berkeley Linguistics Society* 14:406–416.
- 1990, From more to less situated in language: the unidirectionality of semantic change. In S. Adamson, V. Law, N. Vincent and S. Wright eds *Papers from the fifth international conference on English linguistics*, 497–517. Amsterdam & Philadelphia: John Benjamins.
- Traugott, Elizabeth C. and Ekkehard König, 1991, The semantics-pragmatics of grammaticalization revisited. In Elizabeth C. Traugott and Bernd Heine eds *Approaches to grammaticalization*, vol.1:189–219. Amsterdam & Philadelphia: John Benjamins.
- Trudgill, Peter, 1996, Dialect typology: isolation, social network and phonological structure. In Gregory A. Guy et al. eds *Towards a social science of language: papers in honour of William Labov*, 3–21. Amsterdam & Philadelphia: John Benjamins.
- 1997, Typology and sociolinguistics: linguistic structure, social structure and explanatory comparative dialectology. *Folia Linguistica* 23/3–4:349–360.
- Van Valin, Robert D., Jr. and Randy J. LaPolla, 1997, *Syntax: structure, meaning, and function*. Cambridge: Cambridge University Press.
- Warren, R., 1970, Perceptual restoration of missing speech sounds. *Science* 167:392–393.

Whorf, Benjamin Lee, 1956, Linguistics as an exact science. In John B. Carroll ed. *Language, thought, and reality: selected writings of Benjamin Lee Whorf*, 220–232. Cambridge, MA: MIT Press.

Wilson, Deirdre and Dan Sperber, 1986, Pragmatics and modularity. Papers from the Parasession on Pragmatics and Grammatical Theory. *Chicago Linguistic Society* 22/2:67–84.

Wilson, Deirdre and Dan Sperber, 1993, Linguistic form and relevance. *Lingua* 90:1–25.

# 9 *From discourse to grammar in Tamang: topic, focus, intensifiers and subordination*

MARTINE MAZAUDON

## 1 Introduction

Although the general classification of Tibeto-Burman languages remains uncertain, the Tamang language of Nepal belongs clearly in the broad group which Robert Shafer (Shafer 1955) called the 'Bodish Section' of the 'Bodic division' of Tibeto-Burman, of which Tibetan is the best known member. The Bodish section consists of two main branches, the Bodish Branch, containing classical Tibetan and all the so-called 'Tibetan dialects', which share the innovation of *bdun* for 'seven', and the 'Tamang Branch', earlier named 'Gurung Branch' by Shafer on account of the fact that the Gurungs, being soldiers in the British army, were the most visible members of this group. For the earlier scholars, the Tamang branch consisted of Tamang (also called 'Murmi'), Gurung and Thakali (also called 'Thaksya'), to which have been more recently added Manangke, Nar-Phu and the Seke dialects (from the group of villages known as Panchgaon, the 'five villages' in Nepal), plus Chantyal, which is in a very bad state of repair. All of these have dialectal variants or sub-dialects. All languages of the Tamang Branch share a four-tone tonal system resulting from a recent two-way tonal split of what can be reconstructed as an earlier two-tone system, not shared by Tibetan in any of its dialects. If not reconstructible to Proto-Tibeto-Burman, as I believe it is not (but this remains an open question), this proto two-tone system is a common innovation of the Tamang Branch.

Except for the northernmost members of the group, the Seke and Manang dialects, which underwent some recent Tibetan influence on their grammar, languages of the Tamang branch have a very plain verbal morphology, exhibiting none of the developments

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 145–157.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

of auxiliaries indicating evidentiality and orientation *vis-à-vis* the speaker, which are general in Tibetan dialects. In the languages of the Tamang branch, verb roots are invariable. In main clauses, they carry tense-aspect and mood suffixes, in subordinate clauses they carry a marker indicating the relationship of the subordinate clause to the main clause. The way verbal morphology starts to complexify, in this group, is not a matter of combining main verbs with auxiliaries, as in the Bodish branch, but rather of adding to the subordinate verb phrase what Matisoff (1973) has called 'unrestricted particles'—that is particles that mark both noun and verb phrases. In Tamang proper, exemplified here by its Eastern dialect as spoken in the village of Risiangku, we can observe the use of morphemes which are basically discourse particles on all kinds of grammatical structures, and their progressive grammaticisation as they become obligatory on some kinds of subordinate clauses.

## 2 Descriptive and analytic framework

The proper use of discourse particles in a foreign language is among the most difficult things to master. Along with tones, it is one of those aspects of grammar which leads the traditional grammarian to advise the learner to get hold of a native speaker and imitate him; always good advice in any case!

Since a quarter of a century, much progress has been made in the analysis of information structure. There remains nevertheless some imprecision and incompleteness, and all definitions are open to criticism as Dryer remarks (1996). Here we have elected to use the definitions proposed by Lambrecht for topic and focus (Lambrecht 1994), and the framework he provides for the analysis of different types of referents in view of their potential selection as topics, and for the classification of different focus structures. Lambrecht calls 'topic' 'a matter of standing interest or concern' about which relevant information is added in the sentence. 'The topic of a sentence is the thing which the proposition expressed by the sentence is about,' (author's emphasis) (Lambrecht 1994:118). The focus, in many ways, is the complement of the topic, but not necessarily so. A sentence always has a focus, but not necessarily an expressed topic, for instance in answer to the question 'What is going on?' where the whole sentence is the focus. Lambrecht's definition of focus is thus independent of that of topic: 'The focus of the proposition expressed by a sentence in a given utterance context is seen as the element of information whereby the presupposition and the assertion differ from each other,' (Lambrecht 1994:207). This new information is not necessarily a new *denotatum*, but can be a new relation between a *denotatum* and the proposition (Lambrecht 1994:217). The third notion we use, 'intensifiers', borrowed and extended from König (König 1997), refers to particular kinds of focus markers: words like 'even, only, also, self (non-reflexive)'.

### 3 Information structure in the Tamang simple sentence

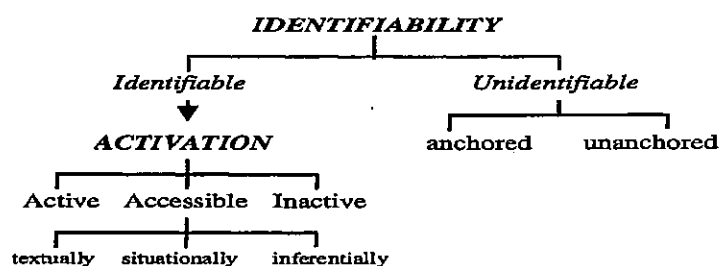
Tamang has two topic markers, *-mi/-m* for simple topics, and *'ca/-ca* for contrastive topics; one focus marker *-ka//ka*; and several intensifiers *-i/-e/-ja* 'also, even', *-n/-nun* 'self (non-reflexive), really', *'ce* 'only'. These markers are used to express the information structure of the sentence without changing its grammatical structure: case marking and word order can remain unchanged. In the following examples<sup>1</sup> the adverb 'now' is in the same position in the sentence, while the place of the topic marker shifts.

- (1) <sup>3</sup>*mi-la*      <sup>4</sup>*tamo-m*      <sup>3</sup>*rap-si*      <sup>3</sup>*cin-ci*  
 person-GEN    now-TOP    play-having    finish-PFV  
 'Hers [i.e. her tape-recorder] has now stopped playing.'
- (2) <sup>1</sup>*ŋa-ta-m*    <sup>1</sup>*tamo*      <sup>2</sup>*aru-la*      <sup>2</sup>*ph a:*      <sup>1</sup>*to:-ci*  
 I-DAT-TOP    now      aunt-GEN    repay      need-PFV  
 'Now I have to repay my aunt's [loan].'

The focus marker can occur on any element of the clause including the verb, without change in the syntactic structure, and its domain can be any element or the whole sentence.

#### 3.1 Topic and topicalisation in simple sentences

Lambrecht proposes an analysis of the cognitive states of referents in discourse in order to predict the likelihood of a given referent's being picked as topic in a sentence. We will follow the broad lines of that classification to characterise the different kinds of topics in Tamang. The main distinction to be made according to Lambrecht is the degree of identifiability of the referent for the hearer. Second is the degree of 'activation' of the referent in the consciousness of both speaker and hearer at the time of speech.



After Lambrecht (1994:109)

<sup>1</sup> The transcription follows the recommendations of the IPA, except that /c/ is used instead of IPA /ts/ and tones are transcribed by numbers from 1 to 4 from highest to lowest. Loan words from Nepali are in bold. Abbreviations used: DIR directional; IMP imperative; INT intensifier; IPFV imperfective; IRR irrealis; NMZR nominaliser; PFV perfective; RS reported speech.



The most likely candidate for topic is along the left branch of the tree, an identifiable, 'active'<sup>2</sup> referent; the least likely, along the right branch, an unidentifiable unanchored referent (that is one which cannot even be put in relation with a known element of the situation). With this definition, the topic referent can be a proposition as well as a cognitive entity corresponding to an NP (Dryer 1996:483).

### 3.1.1 *Active referent as topic*

The most common expression of an active topic in Tamang, whether in conversation or in running text, is zero: previously known elements are absent from the sentence (or pronominalised by zero), as the agent and patient of sentence (3).

- (3) <sup>2</sup>*kh aĩ*     <sup>4</sup>*por-ci* ?  
 where     bring-PFV  
 'Where did [she] take [them]?'

If present, an active topic is most commonly rejected after the verb, with lowered intonation, as an afterthought. Several active topics can occur in this position, with no particle attached.

- (4) <sup>2</sup>*kh aĩ*     <sup>4</sup>*por-ci*,     *kanchi-se*     <sup>4</sup>*me*  
 where     bring-PFV     Kanchi-ERG     cow  
 'Where did she take them, Kanchi, the cows?'

### 3.1.2 *Accessible referent as topic*

A topic which is accessible but not active (present in the slightly more distant context) is activated by being mentioned with the topic marker *-m(i)*. This is the case in (5) for the demonstrative *'oca* summing up what was said before (textually accessible), and for the pronoun *'ha* referring to the speaker (situationally accessible)<sup>3</sup>.

- (5) <sup>1</sup>*oca-m*     <sup>1</sup>*ha-i-mi*     *tʰa:*     <sup>3</sup>*are,*     *mai*  
 that-TOP     I-ERG-TOP     knowledge     is not     Mai  
 'That, as far as I am concerned, I don't know, Mai.'

<sup>2</sup> Remembering that 'active' here means 'in the foreground of the speech participants consciousness (i.e. 'activated'), not active vs passive.

<sup>3</sup> Although these two arguments could also be considered as 'active', following Lambrecht's examples p. 110, in which case we are in the construction described below in (3.1.2). This may be acceptable for *'oca*, but I don't believe it is right for the first person pronoun, in the sense that the speaker was not under discussion in the preceding conversation.

## 3.1.3 'Re-activated' active topic

If an active topic has to be mentioned in its proper grammatical place in the sentence, it is marked with the topic marker *-m(i)*. This structure occurs in story telling where the cohesion of discourse is ensured by the formal repetition of the topic from one sentence to the next (6).

- (6) <sup>4</sup>*kle* <sup>4</sup>*kik* *rani* <sup>4</sup>*kik* <sup>1</sup>*mu-pa* <sup>2</sup>*cim*  
king one queen one be-PAST EVID

<sup>1</sup>*oca* <sup>4</sup>*kle-then* *rani-mi* <sup>1</sup>*phjukpo-phjukpo* <sup>1</sup>*mu-pa<sup>2</sup>cim*  
that king-and queen-TOP rich-rich be-PAST.EVID  
'[Once upon a time] there was a king and a queen. That king and queen were extremely rich.'

## 3.1.4 Contrastive or selective topic

A contrastive topic is marked with the tonal particle <sup>4</sup>*ca*, sometimes reduced to *-ca*, possibly a loan word from Nep. *cahī*. The contrastive topic marker is often found in balanced sentences, like (7), where two topics are opposed.

- (7) <sup>2</sup>*sipai* <sup>4</sup>*ca* <sup>4</sup>*ηach<sup>h</sup>a* <sup>4</sup>*ηach<sup>h</sup>a*, <sup>2</sup>*oca* <sup>3</sup>*caca* <sup>1</sup>*lich<sup>a</sup>* <sup>1</sup>*lich<sup>a</sup>*  
soldier TOP ahead ahead that small behind behind  
'The soldiers were far in front, and the boy far behind.'

Clauses too can be topicalised in this way:

- (8) <sup>1</sup>*si-pa* <sup>4</sup>*ca* <sup>1</sup>*ηa-i* *patta* <sup>3</sup>*are*  
die-NMZR TOP I-ERG knowledge is.not  
'That she had died, I did not know.'

## 3.2 Focus and focalisation in the simple sentence

Lambrecht distinguishes three types of focus structures: argument-, predicate- and sentence-focus. In Tamang, predicate and sentence focus are not always easily distinguishable and can be grouped under the concept of broad focus, opposed to narrow, argument focus.

Plain focus<sup>4</sup>, whether narrow or broad, is left unmarked in Tamang. Given the propensity of Tamang to delete all known arguments from the spoken sentence, the focal element is frequently the only constituent present.

<sup>4</sup> Plain focus is to be distinguished from what Dryer (1996) calls 'simple focus', which is a strong focus marked by intonation only (rather than e.g. a cleft construction in English). Plain focus is the minimal level of focus that justifies an utterance.

- (9) *'pin-ci*  
give-PFV  
'[I/he/you] gave [it] [to me/you/him]'

Strong focus, whether narrow or broad, is marked by the particle *-ka*.

### 3.2.1 *Strong focus on an argument*

- (10) *'ai-la    'mar-ka    'ni:nun    'cuŋ-o    'pi-pa    'ŋa-i-mi*  
you-GEN gold-FOC two-INT sell-IMP say-IPFV I-ERG-TOP  
'It is your gold [earrings] that I said to sell both of (I did).'

Example (10) is uttered as a reprimand to the trickster boy of the story who has sold as slaves the two soldiers who were accompanying him, instead of his two earrings, as instructed by the soldiers. He mutters (11) in reply.

- (11) *'ŋa-i-e-n    'mar-ka    'pi-pa    na    'osem,    'ni:nun*  
I-ERG-also-INT gold-FOC say-IPFV uh then both  
'Me too, it is my golds which I say, uh, then, both.'

### 3.2.2 (*Strong*) *broad focus*

A strong focus on the predicate (12) is sometimes easy to contrast (semantically) with a sentence focus (13), but not always.

- (12) *'the-la    'mriŋ-ka    'cim*  
he-GEN wife-FOC EVID  
'That's his *wife*.'

- (13) *'ai-la    'mriŋ    'cim-ka*  
you-GEN wife EVID-FOC  
'You have a wife (already) [so why should I go out with you?].'

Note that the constituent which carries the marker *-ka*, in case of broad focus, is not necessarily directly contrasted to another possible item. The marker *-ka* often indicates that the assertion is globally contrary to expectancy, or to the wish of the speaker or the hearer.

- (14) *'ŋa-i-ia    'nipa-ka,    'apa*  
I-ERG-also go-FOC father  
'I want to go too, Daddy'

- (15) *'the-m    'lich-a-m    'jaŋ-ta-n    'jaŋ-pa-ka*  
it-TOP later-TOP us-DAT-INT find-IPFV-FOC  
'Those [riches], later, we will get them back (for sure/don't worry).'

- (16) *e,    'ŋa-i-mi    'a-<sup>3</sup>pa-lai-ka*  
eh I-ERG-TOP not-bring-IRR-FOC  
'Eh, me, I did not bring [it/any].'

The focus marker which semantically applies to the whole predicate can be affixed to the object of the verb as in (17) or to its subject in case of sentence focus (18).

- (17) <sup>2</sup>*ai-se-mi*    <sup>1</sup>*ŋa-la*    <sup>1</sup>*h o-ri*    <sup>1</sup>*ch amo-ka*    <sup>3</sup>*pa-ci*  
 you-ERG-TOP    I-GEN    top-LOC    cowife-FOC    bring-PFV  
 'You brought home a **co-wife** on me (of all things you could have done).'
- (18) <sup>2</sup>*nam-ka*    <sup>1</sup>*kh a-pa-ri*    <sup>4</sup>*mai-ci*  
 rain-TOP    come-NMZR-LOC    try-PFV  
 'It's going to rain (contrary to expectation)!'

In (17) the husband did not bring a second wife instead of a cow, in a substitution parallel to (10). Rather, the speaker underlines the nastiness of the whole action of bringing in a second wife.

### 3.2.3 Intensifiers

Several intensifiers are used for special focus. The meaning and use of *-i/-el-ja* 'also, even', and *ce* 'only' are clear enough; that of *-n/-nun* 'self (non-reflexive), really' is more subtle.

Although *-n* and *-nun* are variants in a number of contexts, they are distinct in others. When they are not morphophonemic variants of each other (conditioned by the phonological context) *-nun* seems to indicate the unique identification of the referent, while *-n* indicates rather the plenitude of the quality attributed to the referent.

- (19) <sup>2</sup>*li-mi*    <sup>2</sup>*sei-ci*    <sup>4</sup>*kle-nun,*    <sup>3</sup>*ko-ri* ...  
 face-TOP    know-PFV    king-INT    body-LOC ...  
 'The face, she recognised—it was the king all right, but on the body ...'
- (20) *pap*    *dharma*    <sup>4</sup>*kle-ta-n*  
 sin    merit    king-DAT-INT  
 'Sin or merit, it accrues fully to the king!'

The suffix *-nun* is used on numbers: <sup>3</sup>*mun* <sup>4</sup>*kik-nun* [night one-INT] 'the whole night', <sup>4</sup>*ni-nun* [two-INT] 'both'; while *-n* is often used on scalar quantifiers: <sup>2</sup>*ot-te-n* [this-amount-INT] 'that's all! (i.e. my story is finished)', <sup>3</sup>*lana-n* [much-INT] 'much (really much)', <sup>3</sup>*mokko-n* [all-INT] 'all (without exception)', <sup>3</sup>*sjo:-ri-n* [morning-in-INT] 'early in the morning'. The suffix *-n* is often used to strengthen the marker *-i/-el-ja* on NP's (see ex. 11, *-en* 'also' <*-e-n*).

## 4 The use of topic and focus markers in the construction of complex sentences

The same markers which indicate information structure in simple sentences are found in complex sentences, where they help to shift the basic descriptive reading of subordinating

suffixes towards a logical reading. Their use is optional with time and manner clauses, but with conditionals it has become obligatory. We will first look at the basic meanings of the subordinating suffixes.

#### 4.1 *Basic subordination patterns*

The relationships between clauses in Tamang are indicated by a set of suffixes affixed to the verbal root of the subordinate clause. The most important are the following:

- ma* temporal simultaneity
- si* temporal succession
- na* manner or aim
- sa-* condition

- (21) *ʎi:cha* <sup>2</sup>*kola* *gothe-ri* <sup>1</sup>*ni-ma* <sup>1</sup>*apa* <sup>4</sup>*tim-ri* <sup>ʎh</sup>*a-cim*  
 later child stable-LOC go-while father house-LOC arrive-PFV  
 'Then, while the child was going to the stables, the father arrived at the house.'
- (22) <sup>3</sup>*mokkon* *dikki-ri* <sup>4</sup>*por-si* <sup>2</sup>*to-ci*  
 all mill-LOC carry-after mill-PFV  
 'I took it all to the mill and milled it.'
- (23) <sup>2</sup>*mren-na* <sup>1</sup>*ca-ci*  
 satiate-MANNER eat-PFV  
 'He ate his full.'

These basic meanings can vary pragmatically. Thus an anteriority marker can indicate manner (24) or cause (25), and a simultaneity marker can express a condition (26).

- (24) <sup>1</sup>*sja-si* <sup>1</sup>*sja-si* <sup>ʎh</sup>*a-ci*  
 dance-ing dance-ing come-PFV  
 'She came dancing all the way.'
- (25) <sup>2</sup>*kola* <sup>1</sup>*si-si* *pir* <sup>ʎa</sup>*pa-ro*  
 child die-having pain happen-INF-RS  
 'He says he is sad because his child died.'
- (26) <sup>1</sup>*pʰut-si-n* <sup>2</sup>*ci* <sup>ʎo:</sup>*pa,* <sup>3</sup>*a-ʎpʰut-ma* <sup>1</sup>*si-pa* <sup>1</sup>*pʰut-ma* <sup>2</sup>*ʎuŋ-pa*  
 blow-ing-INT stay must-IPFV NEG-blow-when die-IPFV blow-when burn-IPFV  
 'You have to keep blowing, if (when) you don't blow on it, it [the fire] dies, if (when) you blow on it, it burns.'

#### 4.2 *Topic on subordinate clauses*

##### 4.2.1 *Simultaneity clause with topic marker*

Added to the suffix *-ma*, which indicates simultaneity in time, the topic marker *-m(i)* explicitly sets the clause as background, facilitating a causal reading.

- (27) <sup>3</sup>a-<sup>l</sup>kh a-ma-m,      <sup>4</sup>mai      <sup>1</sup>ni-ci  
neg-come-when-TOP search go-PFV  
'As he did not arrive, they went to look for him.'
- (28) <sup>1</sup>ca-ma-m      <sup>2</sup>itt<sup>h</sup>epa-n      <sup>2</sup>pho      <sup>3</sup>plo-pa  
eat-when-TOP this much-INT belly boil-IPFV  
'When I eat, my belly gurgles so much.'

In (29) the presence of the topic marker prepares the dramatic effect expressed in the main clause. Compare with the unemotional reporting in (30).

- (29) <sup>3</sup>kuŋ-te-ri      <sup>3</sup>to:ka-ma-m      <sup>1</sup>apa      <sup>1</sup>sjon-se      <sup>4</sup>lap      <sup>4</sup>por-ci-ro  
middle-about-LOC reach-DIR-when-TOP father river-ERG ADV carry-PFV-RS  
'But as he was reaching the middle, the father was carried off by the river.'
- (30) <sup>1</sup>sjon      <sup>3</sup>kuŋ-ri      <sup>3</sup>to:ka-ma      <sup>1</sup>sjon-se      <sup>4</sup>lap      <sup>4</sup>por-ci  
river middle-LOC reach-DIR-when river-ERG ADV carry-PFV  
'As he was reaching the middle (of the river), the river carried him off.'

Example (29) is said when first telling the story, (30) is used shortly after (29), in a factual description summing up the situation of the mother and her two children, when the father, who was helping them cross, disappears, leaving the mother on one side of the river and the children on the other.

#### 4.2.2 Sequential clause with topic marker

The most common extension of the reading of the sequence marker *-si* is causal, as in (25) above. Adding the topic marker *-mi*<sup>5</sup> allows a concessive reading as in (31).

- (31) <sup>3</sup>tante      suk<sup>h</sup>a      <sup>1</sup>iaŋ-si-m      <sup>1</sup>tik      <sup>1</sup>la-sai,      <sup>1</sup>lich<sup>h</sup>a      tuk<sup>h</sup>a  
now happiness find-ing-TOP what do-COND later unhappiness
- <sup>1</sup>iaŋ-ci      <sup>3</sup>pi-sam  
get-PFV say-if  
'Even though we may have happiness now, what good is it, if we get unhappiness later?' (lit. Having happiness now, what can we do?)

The common expression of human hopelessness <sup>1</sup>tik <sup>1</sup>la-sai 'what can we do?' is normally appropriate with the expression of sorrow, not of happiness. The logical expectation is thus broken, and the topic marker underlines this fact. The derived value of the subordinate clause is reversed from the usual causal reading of the sequential clause to a concessive reading.

<sup>5</sup> In Western Tamang, the *-m* post-suffix seems to have become stuck onto the sequential marker, yielding an unanalysable marker *-cim*. See Everitt (1973) and Taylor (1973). Western Tamang *-cim*, cognate to Eastern Tamang *-si*, a subordinate clause marker, should not be confused with Eastern Tamang *-cim*/<sup>2</sup>*cim*, translated here as 'evidential' or 'evidential perfective' in the interlinear gloss, where the final *-m* comes from the copula *'mu* 'to be', and not from the topic marker *-m(i)*.

### 4.3 Focus on subordinate clauses

When used on a subordinate clause, the narrow focus marker *-ka* has a contrastive value comparable to clefting in English.

- (32) <sup>1</sup>*ti:la*      <sup>1</sup>*ŋa*      <sup>1</sup>*oraŋ*      <sup>3</sup>*ŋjot-la*      <sup>3</sup>*pi-si-ka*      <sup>4</sup>*paŋ*  
 yesterday    I            that way    drunk-FUT    say-ing-FOC    strength

<sup>3</sup>*are-pa-ri*                      <sup>2</sup>*kʰ ana-i*              <sup>3</sup>*a-<sup>2</sup>waŋ-pa*  
 not be-NMZR-LOC    where-also    NEG-enter-IPFV

'Yesterday it was for fear of getting drunk, like that, in my weak condition, that I did not enter anywhere.'

Example (32) indicates that it is 'for this reason and no other' that the speaker has contacted no one, (and not because she did not wish to see her family).

### 4.4 The intensifiers *-n* and *-nun* on subordinate clauses

On a subordinated VP, *-n* in effect shifts the focus from its unmarked place on the main clause onto the subordinate clause (but without the contrastive value of the narrow focus marker *-ka*, (see 32).

- (33) <sup>1</sup>*sol*              <sup>4</sup>*cw-la*,              <sup>1</sup>*ch joi-si-n*              <sup>1</sup>*pʰ ep-o*  
 food (HON)    cook-FUT    eat(HON)-ing-INT    go(HON)-IMP  
 'I will cook lunch; please eat before you leave.' (lit. Please leave AFTER eating)

- (34) <sup>1</sup>*it-se*              <sup>1</sup>*klaŋ-si-n*              <sup>1</sup>*kan*              <sup>1</sup>*ca-ci-ka*  
 this-ERG    play-ing-INT    rice    eat-PFV-FOC  
 'This one gambles even while eating!'

In (34) *-ka* emphasises the whole proposition (the whole complex sentence), while *-n* indicates the focus inside that structure.

### 4.5 Conditionals

The conditional marker *-sa-* cannot be used nowadays without a following suffix, which is either the topic *-m(i)*, the strong focus *-ka*, or the intensifier *-i* (sometimes reinforced to *-i-nun*). The discourse particles, which are still used for stylistic effects on other subordinate clauses, have thus been grammaticised on conditional subordinates.

#### 4.5.1 Topic

The adjunction to *-sa-* of the topic marker yields a plain conditional, corresponding to the now well accepted idea that 'conditionals are topics' (Haiman 1978).

- (35) <sup>2</sup>*nam* <sup>1</sup>*kh a-sa-m* <sup>1</sup>*ŋa* <sup>3</sup>*a* <sup>1</sup>*kh a*  
 rain come-if-TOP I NEG come  
 'If it rains, I won't come.'

The verb of the main clause can be in the indicative mood as in (35). It can also be in the conditional, expressing greater doubt as to the certainty of the event.

- (36) <sup>1</sup>*iampu-ri* <sup>1</sup>*mar* <sup>1</sup>*ni-pa* <sup>3</sup>*mi* <sup>1</sup>*mu-sam,* <sup>1</sup>*ŋa-m* <sup>3</sup>*cakir*  
 Kathm.-LOC down go-NMZR man be-if I-TOP salary  
<sup>1</sup>*ca-pa-r'* <sup>1</sup>*ni-sai*  
 eat-NMZR-LOC go-COND  
 'If there were someone going down to Kathmandu, I would go [with him] to earn a salary.'

#### 4.5.2 Focus

The use of a focus marker on a conditional clause is less readily understandable. In Tamang, it is used to underline the fact that the condition is not fulfilled, at least as yet, and that the speaker wishes it were.

- (37) <sup>2</sup>*ai* <sup>1</sup>*ni-sa-ka* <sup>4</sup>*por-sai*  
 you go-if-FOC take along-COND  
 'If only you would come, I would take you along.' (in the most common reading = 'I would marry you')

It is thus most often used with a past main clause as a counterfactual marker.

- (38) <sup>1</sup>*t<sup>h</sup>aa* <sup>1</sup>*mu-sa-ka* <sup>4</sup>*pa-si* <sup>1</sup>*kh a-sai*  
 knowledge be-if-FOC bring-ing come-COND  
 'If only I had known, I would have brought some.'

#### 4.5.3 Intensifiers

The intensifier *-i* added to the conditional marker *-sa*, emphasises the condition (39) and in most cases yields a concessive reading 'even if' (40). In that meaning it is often strengthened by a second intensifier *-nun*.

- (39) <sup>2</sup>*chjai* <sup>1</sup>*ta-sa-i,* <sup>2</sup>*ai-se* <sup>2</sup>*p<sup>h</sup>a:* <sup>1</sup>*to:-pa* <sup>3</sup>*ahin*  
 debt occur-if-INT you-ERG pay must-IPFV isn't it?  
 'And if ever he gets into debt, you will have to pay it, right?'
- (40) <sup>1</sup>*the* <sup>3</sup>*a-kh a-sai(-nun)* <sup>1</sup>*ŋa* <sup>1</sup>*ni-la*  
 he not-come-if+INT(-INT) I go-FUT  
 'Even if he won't come, I will go.'



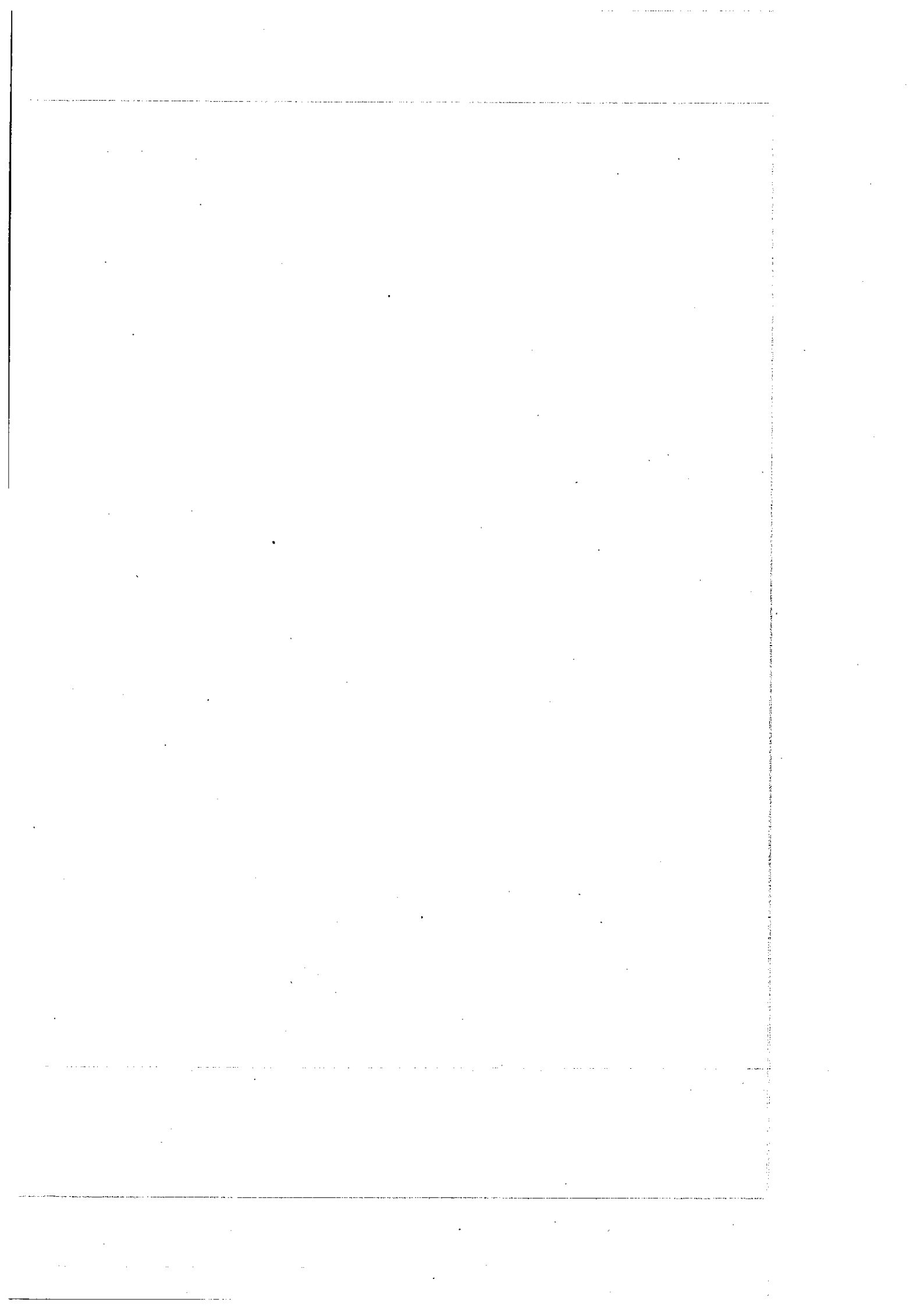
## 5 Conclusion

In the present situation of Tamang, we find the same particles used in different grammatical roles, with meanings that can be rather transparently derived from some basic meaning. Are they the same particles synchronically, and is the present state stable? The obligatoriness of these particles on conditional sentences tends to suggest that grammaticisation has occurred or is occurring there, and that in their use on subordinate clauses, these particles are slowly losing the information structure value which we can construe as their original value, in favor of a more grammatical value, where they blend with the subordination suffixes.

## References

- Dryer, Matthew S., 1996, Focus, pragmatic presupposition, and activated propositions. *Journal of Pragmatics* 26:475–523.
- Grierson, George A., 1967 [1908], *Linguistic survey of India*, vol.3, Tibeto-Burman family. Delhi: Motilal Banarsidass.
- Everitt, Fay, 1973, Sentence patterns in Tamang. In Ronald L. Trail ed. *Patterns in clause, sentence, and discourse in selected languages of India and Nepal*. Norman, Oklahoma: SIL. Part 1 Sentence and Discourse:197–234.
- Haiman, John, 1978, Conditionals are topics. *Language* 54/3:564–589.
- Haspelmath, Martin & Ekkehard König, 1998, Concessive conditionals in the languages of Europe, in Johann van der Auwera *et al.* eds *Adverbial constructions in the languages of Europe*, 563–640. Berlin: Mouton de Gruyter.
- König, Ekkehard, 1997, Towards a typology of intensifiers (emphatic reflexives). *16th International Conference of Linguists, Paris*. CD-Rom. Oxford: Pergamon.
- Lambrecht, Knud, 1994, *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Matisoff, James A., 1973, *The grammar of Lahu*. Berkeley: University of California Press.
- Mazaudon, Martine, 1988, Temps aspect et négation en tamang. In Nicole Tersis & Alain Kihm eds *Temps et aspects*, 179–185. Paris: Péeters/SELAF.
- 2003, Tamang. In Graham Thurgood & Randy LaPolla eds *Sino-Tibetan Languages*, 291–314. London: Routledge.
- n.d., Tamang texts. Web publication. <http://lacito.archivage.vjf.cnrs.fr/>.
- Shafer, Robert, 1955, Classification of the Sino-Tibetan languages. *Word* 11/1:94–111.

- Taylor, Doreen, 1973, Clause patterns in Tamang. In Austin Hale & David Watters eds *Clause, sentence and discourse patterns in selected languages of Nepal*. Norman, Oklahoma: SIL. Part 2 Clauses:81-174.
- Yoncan, Amrit, 2054 B.S., [1997], *Tamang vyakaran* [Tamang grammar]. Kathmandu: Royal Nepal Academy. [in Nepali].



# 10 Preliminary remarks on Gyarong negation particles<sup>1</sup>

---

YASUHIKO NAGANO

## 1 Introduction

Gyarong (*rGyal rong* in Written Tibetan (WT)) is a Tibeto-Burman (TB) language spoken in the northwestern part of Sichuan Province, China. This language has long attracted the attention of scholars, because of the striking similarity of some of its lexical items to those of WT as well as its complicated system of affixation, which could be regarded as reflexes of Proto-TB morphology.

The author has written two studies on the language (Nagano 1984, 2003), and this small paper is intended to supplement their discussion of the negation system. The negation particles that I point out in this paper have not been described in any previous works on Gyarong.

The majority of the Gyarong people inhabit the Aba (WT *mga ba*) Tibetan and Qiang Autonomous Prefecture and the Gantse (WT *dkar mdzes*) Tibetan Autonomous Prefecture of Sichuan Province. The exact population is unknown, since they are not officially recognised as an independent national minority, but are classified as part of the Tibetan nationality. The number of speakers is at least 150,000.

It is traditionally said that there are eighteen dialects. This classification came from the historical division of the area into eighteen administrative zones in the 12<sup>th</sup> century, with a separate dialect said to be spoken in each zone. On the basis of modern data, however, the

---

<sup>1</sup> The following abbreviations are used in this paper: 1 first person, 2 second person, 3 third person, AUX auxiliary verb, CAUS causative, ERG ergative, HON honorific, IMP imperfect, INF infinitive, INTERR interrogative, LOC locative, N noun, NEG negative, NOM nominaliser, NP noun phrase, PFT perfect, PL plural, PROG progressive, PROH prohibitive, PTT patient, SG singular, VP verb phrase.

language seems to have three dialects: northern, eastern and western. There are about 10,000 speakers of the northern dialect; Datsang (WT *da tshang*) is in the center of the northern area. There are about 50,000 speakers of the western dialect, in the Dzatang (WT '*dzam thang*) and Tampa (WT *bstan pa* or *dam pa*) areas. The eastern area covers a relatively large region including Cogtse (WT *lcog rtse*), Barkham (WT '*bar khams*), Suomo (WT *so mang*), Tsakunao or Tshako (WT *bkra shi gling*), Lishan (WT *lis rdzong*), Shaojin (WT *btsan lha*), Jinchuan (WT *rab brtan*), Heishui (WT *khro chu rdzong*) and Mawo (WT *bha dbo*). There are around 80,000 Gyarong people in this area.

These dialects are classified according to their initial consonant clusters and by the behavior of pronominal affixes in VPs. However, not all the descriptions of dialects are complete, so the classification remains tentative. Among these dialects, the conservative Cogtse (WT *lcog rtse*) dialect, which retains a set of affixes, is considered the standard. The following description is based on that dialect, unless otherwise noted.

The consonant phonemes of Gyarong are here transcribed as follows: *p ph b; py phy by; t th d; tr thr dr; k kh g; ky khy gy; ' ; ts tsh dz; c ch j; s z; sh zh; h, fi; m n ny ng; l hl r; w y*. As to the phonetic values, *ph phy th thr kh khy tsh ch* are aspirated. *Py phy by* and *ky khy gy* are palatalised. The palatalised velars also have palatal stop allophones [c j]; Chinese linguists often transcribe them as [cç cç' j]. *Tr thr dr* are retroflex; *c ch j* are palatoalveolar [tʃ tʃʰ dʒ], and *sh zh* are alveopalatal fricatives [ç ʒ]. The *fi* represents a voiced cavity fricative. *Ny ng* are palatal and velar nasals [ɲ ŋ], and *hl* is a voiceless alveolar lateral fricative [ɬ].

In addition to the above, there is /N/ as a phoneme of nasalisation, which occurs as a homorganic nasal before the stops and affricates, and in final consonant position, nasalising the preceding vowel. The vowels are /a i u e o ə/. The usual allophone of /i/ is [i]; /u/ has allophones [u] and [u̠]. Tone is not distinctive. Some say there are several minimal pairs, but these were in fact homonymous in the speech of my informants.

The syllable structure is (C)C<sub>i</sub>(G)V(C<sub>f</sub>)(s), where the parenthesised portions are optional. The (C) prefix position can only be occupied by *p t k r l s sh m* or *N*. All the voiceless consonants in this position become voiced when followed by a voiced initial consonant. All consonants except for *fi* can occur as initial consonant C<sub>i</sub>. G stands for medial glides *r l w y*. C<sub>f</sub> can either be *p t k ' c s fi m n ny ng l r w y* or *N*. There are syllables that could be interpreted as vowel-initial, but these are interpreted as 'V here.

## 2 Negation particles in previous works

The only negation marker previously described in all previous studies on the language has been the *ma-* particle. For instance, the exhaustive grammar of Cogtse dialect by Lin describes it as follows: 'In Gyarong, the adverb of negation mainly modifies verbs and/or adjectives, specifying negation and prohibition' (Lin 1993:312–313, my translation). The

adverbs he describes are *ma* [ma, mə] and *mə*. The former is the normal negation, while the latter means 'not yet finished' or prohibition. He lists the following examples.

- |     |                      |                            |
|-----|----------------------|----------------------------|
| (1) | <i>ŋa ma ki-ŋ</i>    | 'I don't/won't buy.'       |
|     | <i>mə mə zə-u</i>    | 'He doesn't/won't eat it.' |
|     | <i>no mə tə-pə-u</i> | 'You don't do it.'         |
|     | <i>ta-pu ma mfor</i> | 'The child is not cute.'   |
| (2) | <i>ŋa mə pəŋ</i>     | 'I haven't done it.'       |
|     | <i>wəjo mə za-u</i>  | 'He hasn't eaten.'         |
|     | <i>ŋa mə to-m</i>    | 'I haven't beaten him.'    |
|     | <i>ŋa mə mfor</i>    | 'I wasn't beautiful.'      |
| (3) | <i>no mə tə-zə-u</i> | 'Don't eat.'               |
|     | <i>no mə tə-let</i>  | 'Don't beat.'              |

My second informant, who passed away at the age of 65 in 1985, had the same negation system. As I wrote (Nagano 2003), 'Negation is always expressed by *ma* which is followed by VP-final, VP-nonfinal or auxiliary verb'. However, a more recent informant, belonging to a much younger generation, has different negation particles, *ja-* and *ji-*.

### 3 Negation particles in recent data

As a general tendency, *ja-/ji-* seem to appear in perfect aspect, while *ma-* appears in the imperfect. The distribution of *ja-* and *ji-* will be discussed at the end of this chapter. Prohibition is expressed exclusively by *ji-*.

#### 3.1 Negation particles in intransitive structures

The following examples show a typical contrast between *ja-/ji-* and *ma-*.

- |     |                                                               |             |                |             |             |                   |               |
|-----|---------------------------------------------------------------|-------------|----------------|-------------|-------------|-------------------|---------------|
| (4) | <i>wugyo</i>                                                  | <i>tsay</i> | <i>ta-ki-w</i> | <i>reN,</i> | <i>məza</i> | <i>tshongkang</i> | <i>ja-che</i> |
|     | he                                                            | vegetable   | PFT-buy-3PTT   | because     | she         | store             | not-go        |
|     | 'Because he bought vegetables, she didn't go to the store.'   |             |                |             |             |                   |               |
| (5) | <i>wugyo</i>                                                  | <i>tsay</i> | <i>ta-ki-w</i> | <i>reN,</i> | <i>məza</i> | <i>tshongkang</i> | <i>ma-che</i> |
|     | he                                                            | vegetable   | PFT-buy-3PTT   | because     | she         | store             | not-go        |
|     | 'Because he bought vegetables, she will not go to the store.' |             |                |             |             |                   |               |

In negative sentences, no direction prefix occurs in the final VP. Sentence (4) means that she judged she did not need to go to the store. The form *ji-che*, instead of *ja-che*, is only possible under the condition that his purchase of vegetables forced her not to go to the store.

- (6) *wugyo wa-rgyap nə-sar wu-Nkuy, magyu magyu che sa-mu-y*  
 he marriage PFT-get of-since often wine place-drink-LOC  
*ja-che*  
 not-go  
 'Since he got married, he did not go to the tavern often.'
- (7) *wugyo wa-rgyap nə-sar wu-Nkuy, che sa-mu-y ja-ta-che*  
 he marriage PFT-get of-since wine place-drink-LOC not- PFT-go  
 'Since he got married, he has never been to the tavern.' (he quit the habit of going to the tavern)
- (8) *wugyo wa-rgyap nə-sar wu-Nkuy, che sa-mu-y ma-nə-che*  
 he marriage PFT-get of-since wine place-drink-LOC not-PROG-go  
 'Since he got married, he is not going to the tavern.'

Example (6) means that he refrained from going to the tavern so often, whereas (7) means that he quit his habit of going to the tavern. The perfect marker *ta-* in (7) was originally a direction affix meaning 'upward', which now connotes completion. In these three examples, his not going to the tavern is based on his own will, and the form *ji-che* does not occur.

Another verb which means 'go', *thar*, requires a general movement prefix *yi-*. We may be able to say that *yi-* has been lexicalised. Therefore, *jikthar* in (9) should be analysed as *ja-yi-kə-thar* instead of *ji-kə-thar*.

- (9) *wugyo kuru zinka-y jikthar (<ja-yi-kə-thar)-je thakchot*  
 he Tibet area-LOC not-general.movement-1SG-go-NOM certain  
  
*nə-ngo*  
 down-AUX  
 'It's certain that he did not go to Tibet.'

In both (10) and (11), (*tə*-)*mnyak gyu* behaves as an intransitive verb meaning 'to sleep', although it literally is 'to close eyes'. When this literal meaning needs to be expressed, a causative prefix must precede *gyu*.

- (10) *wugyo wu-mnyak ja-gyu wuches, tə-mnyo no-pa-w*  
 he his-eyes not-sleep since show PFT-watch-3PTT  
 'Because he didn't fall asleep, he watched the show.'
- (11) *wugyo wu-mnyak ma-gyu wuches, tə-mnyo pa-w*  
 he his-eyes not-sleep since show Ø-watch-3PTT  
 'Because he does not fall asleep, he will watch the show.'

Natural phenomena seem to require *ji-* as in (12). The form *jupa* is a contraction of *ja-wu-pa*.

- (12) *wugyo-yo ta-mnya-ma wastot jupa(<ja-wu-pa)-nəma. pewa ta-mu*  
 they cultivation well not-3PL>3-AUX(NEG) this.year rain  
*kəmca ji-lat*  
 many not-fall  
 'It isn't that they didn't plant well. It didn't rain a lot this year.'

### 3.2 Negation particles in transitive structures

Basically, the differentiation of *ma-* and *ja-/ji-* seems the same as in the intransitive structure, as seen in the following examples. In parallel examples (13) to (15), (16) to (19) and (20) to (22), the *ja-* versus *ma-* distinction is rather straightforward. The informant did not accept any sentences with *ji-*.

- (13) *wugyo tama ku-pa ci, chitre ja-let*  
 he work PFT-do when car not-drive  
 'When he worked, he didn't drive a car.'
- (14) *wugyo tama pa-w ci, chitre ma-let*  
 he work do-3PTT when car not-drive  
 'When he works, he doesn't drive a car.'
- (15) *wugyo tama pa-w ci, chitre ma-nə-let*  
 he work do-3PTT when car not-PROG-drive  
 'When he works, he usually doesn't drive a car.'
- (16) *wugyo tama ja-pa-w reN, wu-Ngra munədet (<ma-wu-nə-det)*  
 he work not-do-3PTT because salary not-3PL>3-PROG-pay  
 'Because he didn't work, they are not paying him his salary.'
- (17) *wugyo tama ja-pa-w reN, wu-Ngra mudet (<ma-wu-det)*  
 he work not-do-3PTT because salary not-3PL>3-pay  
 'Because he didn't work, they will not pay him his salary.'
- (18) *wugyo tama ja-pa-w reN, wu-Ngra judet (<ja-wu-det)*  
 he work not-do-3PTT because salary not-3PL>3-pay  
 'Because he didn't work, they didn't pay him his salary.'
- (19) *wugyo tama ma-pa-w reN, wu-Ngra mudet (<ma-wu-det)*  
 he work not-do-3PTT because salary not-3PL>3-pay  
 'Because he doesn't work, they will not pay him salary.'
- (20) *wugyo-yo stə wu-gyim nu(<no-wu)-Npar ju(<ja-wu)-Npar*  
 they this of-house PFT-3-sell not-3-sell  
*nutho(<nə-wu-tho)*  
 PROG-3-ask  
 'They are asking whether or not they sold the house.'
- (21) *wugyo-yo stə wu-gyim ka-Npar ma-Npar nutho (<nə-wu-tho)*  
 they this of-house INF-sell not-sell PROG-3-ask  
 'They are asking whether or not they sell the house.'



- (22) *wugyo-yo stə wu-gyim nu(<no-wu)-Npar ju(<ja-wu)-Npar*  
 they this of-house PFT-3-sell not-3-sell

*tutho(<ta-wu-tho)*

PFT-3-ask

'They asked whether or not they sold the house.'

However, depending on the verb, the following alternatives occur:

- (23) *wugyo gyaga-y na-we nongo-y, nga ja-mto-ng*  
 he India-LOC high.to.low-come AUX-though I not-CAUS-see-1SG  
 'Even though he has arrived from India, I haven't seen/met him.'

- (24) *wugyo gyaga-y na-we nongo-y, nga ma-wa-rdo-ng*  
 he India-LOC high.to.low-come AUX-though I not-CAUS-see-1SG  
 'Even though he has arrived from India, I won't meet him.'

For (23), *ma-mto-ng* is theoretically grammatical. However, the informant rejected it, because the subordinate sentence is based on the direct information and therefore *ma-mto-ng* meaning 'I don't see him' constitutes a contradiction.

Unlike *ja-* in (23), *ji-* may appear for the same verb, as in:

- (25) *wugyo wu-gyim guy mak-nyi ji-mto-ng*  
 he of-house in NEG.AUX(EXISTENCE)-NOM not-see-1SG  
 'I didn't see him when he wasn't at home (he was always at home).'

- (26) *shtə wu-rmi kəce nongo-y, ji-mto-ng*  
 this of-man where AUX-though, not-see-1SG  
 'I didn't see this man anywhere.'

- (27) *məza thə ke-tsi ja-mto-w*  
 she what also not-see-3PTT  
 'She hasn't seen/observed anything.'

Comparing (23) and (27) to (24) and (25), volitionality may be hypothesised as a key for the distribution of *ja-* and *ji-*.

- (28) *wugyo su wu-gyim-guy no-nyis jikshing(<ji-kə-mshi-ng)*  
 he who of-house-in PFT-stay not-1SG-know-1SG  
 'I didn't know whose house he stayed in.'

- (29) *wugyo su wu-gyim-guy kə-nyis-tə makshing(<ma-kə-mshi-ng)*  
 he who of-house-LOC IMP-stay-NOM not-1SG-know-1SG  
 'I don't know whose house he stays in.'

- (30) *wu-tong mə-nə-mshi-w jikshing(ji-kə-mshi-ng)*  
 its-meaning INTERR-PFT-know-3PTT not-1SG-know-1SG  
 'Have you understood the meaning?' 'I haven't.'

- (31) *wugyo kupa wu-skat ma-mshi-w-tə nga jikshing(<ji-kə-mshi-ng)*  
 he China of-language not-know-3PTT-NOM I not-1SG-know-1SG  
 'I didn't know that he doesn't know Chinese.'

As far as the verb 'to know' is concerned, it requires *ji-* as its negation marker in the perfect. It is true that 'to know' is a transitive verb, but only the first sentence in (30) shows the typical transitive structure, judging from the pronominal affix pattern. The other sentences have the *kə-mshi-ng* (1SG/2SG-ROOT-1SG) pattern, which is the structure for intransitives. This dialect has a euphemiser *yi-* but it is not appropriate to analyze *ji-* as *ja-yi-* because the euphemiser cannot occur before a pronominal prefix.

- (32) *wugyo-yo kupa zinka-y kə-che-ny kəngos kəma ji-ches*  
 they China country-LOC INF-go-3PL AUX AUX(NEG) not-say  
 'They didn't say whether or not they were going to China.'
- (33) *wugyo-yo kupa zinka-y kə-che-ny kəngos kəma ma-ches*  
 they China country-LOC INF-go-3PL AUX AUX(NEG) not-say  
 'They won't say whether or not they are going to China.'
- (34) *stə thə kəngos kəma ji-ches*  
 this what AUX AUX(NEG) not-tell  
 'They didn't tell what this is (or is not).'
- (35) *stə thə kəngos kəma ma-nə-ches*  
 this what AUX AUX(NEG) not-PROG-tell  
 'They don't tell what this is (or is not).'

From the examples above, we may be able to conclude that 'to tell' also requires *ji-* for negation in the imperfect. The following examples show the *ja-/ji-/ma-* contrast.

- (36) *wugyo skyi ma-we-tə ji-suso-ng*  
 he here not-come-NOM not-remember-1SG  
 'I didn't remember that he isn't coming here.'
- (37) *wugyo skyi ma-we-tə ma-nə-suso-ng*  
 he here not-come-NOM not-PROG-remember-1SG  
 'I don't remember that he isn't coming here.'
- (38) *wugyo-yo kwor mupay(<ma-wu-pa-y) ji-suso-ng*  
 they help not-3PL>1PL-do-3PL>1PL not-remember-1SG  
 'I didn't think they won't help us.'
- (39) *wugyo-yo kwor mupay(<ma-wu-pa-y) ma-nə-suso-ng*  
 they help not-3PL>1PL-do-3PL>1PL not-remember-1SG  
 'I don't think they won't help us.'
- (40) *shtə wa-ma kə-ra kuma-tə ja-suso-w*  
 this of-work INF-needed AUX(NEG)-NOM not-regard-3PTT  
 'He didn't regard that this work is not important.'
- (41) *shtə wa-ma kə-ra kuma-tə ma-nə-suso-w*  
 this of-work INF-needed AUX(NEG)-NOM not-PROG-regard-3PTT  
 'He is not regarding that this work is not important'

- (42) *wugyo wa-rgyap kə-sar ji-suso nongo-y, wu-pa-ma jis*  
 he his-marriage INF-marry not-think AUX-though his-parents two  
*wu-su-sar nəngo*  
 3PL>3SG-CAUS-marry AUX  
 'Even though he did not think about marriage, his parents will make him.'
- (43) *wugyo wa-rgyap kə-sar ja-suso nongo-y, wu-pa-ma jis*  
 he his-marriage INF-marry not-want AUX-though his-parents two  
*wu-su-sar nəngo*  
 3PL>3SG-CAUS-marry AUX  
 'Even though he did not want to get married, his parents will make him.'
- (44) *wugyo wa-rgyap kə-sar ma-suso nongo-y, wu-pa-ma jis*  
 he his-marriage INF-marry not-want AUX-though his-parents two  
*wu-su-sar nəngo*  
 3PL>3SG-CAUS-marry AUX  
 'Even though he does not want to get married, his parents will make him.'

For the same verb root, both *ja-* and *ji-* occur. The root *suso* includes different grades of volitionality, that is, 'to remember' > 'to think' > 'to regard' > 'to dream' > 'to want', and it seems that the more volitional the meaning is, the more frequently *ja-* occurs. Note that the pronominal affix pattern is transitive only in (40) through (44).

### 3.3 Negation particles with adjectives

There are not enough examples to draw any conclusion concerning the distribution of *ja-* and *ji-* with adjectives.

- (45) *tham zhimpa-yo nyi-loti ja-sna*  
 nowadays farmer-PL of-production not-good  
 'These days farmers' production was no good.'
- (46) *tham zhimpa-yo nyi-loti ma-nə-sna*  
 nowadays farmer-PL of-production not-PROG-good  
 'These days farmers' production is no good.'
- (47) *yinyi tama ji-kə-skoy(<skos-y)-tə chung-gə yitrhul*  
 we work not-diligent-1PL-NOM government-ERG punishment  
*nə-pu-w*  
 PFT-charge-3PTT  
 'The government punished us for our non-diligent work.'

- (48) *yinyi tama ma-kə-skoy(<skos-y)-tə zhung-gə yitrhul*  
 we work not-diligent-1PL-NOM government-ERG punishment

*pu-w*  
 charge-3PTT

'The government will punish us for our non-diligent work.'

### 3.4 Negation particles with auxiliary verbs

The distribution of *ja-/ji-* and *ma-* is the same as the examples above. The verb *ra* 'to need' takes both *ja-* and *ji-* in (49) to (53), depending upon the gradation of volitionality.

- (49) *domor təmu kumca na-lat wuches, təmnya-y teji ka-lat*  
 Last.year rain much PFT-fall because field-LOC water INF-do

*ja-ra*  
 not-need

'Because it rained a lot last year, they didn't have to irrigate the fields.'

- (50) *domor təmu kumca na-lat wuches, təmnya-y teji ka-lat*  
 Last.year rain much PFT-fall because field-LOC water INF-do

*ma-ra*  
 not-need

'Because it rained a lot last year, they won't have to irrigate the fields.'

- (51) *domor təmu kumca na-lat wuches, təmnya-y teji ka-lat*  
 Last.year rain much PFT-fall because field-LOC water INF-do

*ma-n-ra*  
 not-PROG-need

'Because it rained a lot last year, they don't have to irrigate the fields.'

- (52) *zhung nyi-nbey thrə kumca ka-det ji-ra nongo-y,*  
 government of-toward tax many INF-give not-need AUX-though

*wu-zi-nkam yargyes kə-kte ta-che*  
 its-country development big PFT-go

'Though they didn't have to give much tax to the government, the country has developed a lot.'

- (53) *wugyo-yo nyi-shamdu ji-ra nongo-y, tuki(<ta-wu-ki)*  
 they their-gun not-need AUX-though PFT-3PL>3-buy

'Even though they don't need a gun, they bought one.'

The verb *cha* 'to be able to' also takes *ja-* and *ji-*.

- (54) *kuntren loto ka-sə-sna ja-cha*  
 commune production INF-CAUS-good not-able

'The commune was not able to produce well.'

- (55) *kuntren loto ka-sə-sna ma-cha*  
 commune production INF-CAUS-good not-able  
 'The commune will not be able to produce well.'
- (56) *kuntren loto ka-sə-sna ma-nə-cha*  
 commune production INF-CAUS-good not-PROG-able  
 'The commune is not able to produce well.'
- (57) *nga loptrey(<loptra-y) ka-che ji-cha-ng*  
 I school-LOC INF-go not-able-1SG  
 'I was not able to go to school.'

The verb *tso* 'to have time to' takes only *ji-* as in:

- (58) *nga loptrey(<loptra-y) ka-che ji-tso-ng*  
 I school-LOC INF-go not-have.time.to-1SG  
 'I did not have time to go to school.'

For *yo* 'to be allowed', there is only one example:

- (59) *nga tagyim wu-nkuy ka-ngo mə-yo ka-we ma-yo*  
 I house of-in INF-enter INTERR-allowed INF-enter not-allowed  
 'May I come in?' 'Not allowed to come in.'

The auxiliary verb *lo* 'to be about to' requires *ja-* and *ma-*. I have no examples with *ji-*.

- (60) *wugyo jis gyimguy(<gyim-nguy) kə-cwat ja-lo*  
 he two house-toward INF-return not-be.about.to  
 'Those two weren't about to return home.'
- (61) *wugyo jis gyimguy(<gyim-nguy) kə-cwat ma-lo*  
 he two house-toward INF-return not-be.about.to  
 'Those two aren't about to return home.'

The form *myo* 'to have the experience of' can be prefixed by *ja-*, *ji-* or *ma-*. But in most cases *ma* is used, as in (62) and (63).

- (62) *nga gyaga-y ka-che nə-myo-ng nongo-y, nəci*  
 I India-LOC INF-go PFT-have.experience-1SG AUX-though there  
*kumca ka-nyi ma-myo-ng*  
 long INF-stay not-have.experience-1SG  
 'Even though I've gone to India, I haven't stayed there for a long time.'
- (63) *wugyo kuru zinka-y ka-che ma-myo-w*  
 he Tibet area-LOC INF-go not-have.experience-3SG  
 'He has never been to Tibet.'

Conversely, (64) and (65) are grammatical. In these examples, the meaning is rather close to 'had not stayed' and 'had not been'. In these latter examples, informants say that *ji-* is used with first person and *ja-* with third person.

- (64) *nga gyaga-y ka-che nə-rnyo-ng nongo-y, naci kumca*  
 I India-LOC INF-go PFT-have.experience-1SG AUX-though there long  
*ka-nyi ji-rnyo-ng*  
 INF-stay not-have.experience-1SG  
 'Even though I've gone to India, I had never stayed there long.'
- (65) *wugyo kuru zinka-y ka-che ja-rnyo-w*  
 he Tibet area-LOC INF-go not-have.experience-3SG  
 'He had never been to Tibet.'

### 3.5 Negation particles in optative/desiderative structures

Optative is expressed by 'a-*ji-* + ROOT. The form 'a seems to be a cognate with the nominaliser 'a. Elsewhere, this affix is primarily prefixed to the direction markers to nominalise them: 'a-*tha* 'the above, upper place', versus *ta-* 'upward', 'a-*na* 'down, the lower place' versus *na-* 'downward'.

- (66) *wugyo mə-nə-go zə, tama 'a-ji-pa-w*  
 he uncontrollable.act-PROG-sick if work NOM-not-do-3PTT  
 'If he is sick, I hope he doesn't work.'
- (67) *wugyo mə-nə-go zə, tama ji-sə-pa-w*  
 he uncontrollable.act-PROG-sick if work NOM-not-do-3PTT  
 'If he is sick, don't make him work.'
- (68) *wugyo lhasa-y 'a-ji-che*  
 he Lhasa-LOC NOM-not-go  
 'I hope he doesn't go to Lhasa.'
- (69) *tamo 'a-ji-lat*  
 rain NOM-not-fall  
 'I hope it doesn't rain.'

### 3.6 Prohibition

Prohibition is consistently marked by *ji-*.

- (70) *shci ro-we-n*  
 here towards.speaker-come-2SG  
 'Come here!'
- (71) *shci ji-we-n*  
 here not-come-2SG  
 'Don't come here!'
- (72) *shci ji-nbyi-n*  
 here not-come(HON)-2SG  
 'Please don't come here.'

- (73) *ka-pshi ji-pa-w*  
 INF-sing not-do-2SG  
 'Don't sing!'

### 3.7 Irregularities

The contrast between *ja-*, *ji-* and *ma-* has now been described. However, the following sentences show an irregularity. (75) and (76) are normal structures; the reason for (74) having two negation particles is unknown.

- (74) *wugyo tama ma-ji-pa-w zə, wu-Ngra ka-bja me*  
 he work not-not-do-3PTT if his-salary INF-draw AUX(NEG)  
 'If he is not going to work, there is no salary to draw/get.'
- (75) *wugyo tama ma-pa-w zə, wu-Ngra ka-bja me*  
 he work not-not-do-3PTT if his-salary INF-draw AUX(NEG)  
 'If he is not going to work, there is no salary to draw/get.'
- (76) *wugyo tama ja-pa-w reN, wu-Ngra ka-bja no-me*  
 he work not-not-do-3PTT because his-salary INF-draw PFT-AUX(NEG)  
 'Because he didn't work, there was no salary to draw/get.'

### 3.8 Distribution of *ja-* and *ji-*

Sentences (4) through (65) show that volitionality of verbs lies behind the behavior of *ja-* and *ji-*. The following examples may further support this.

- (77) *wugyo ka-we makcha(<ma-kə-cha) mak, ma-we nongos*  
 he INF-come not-3SG-can AUX(NEG) not-come AUX  
 'It is not that he cannot come, he doesn't come.'
- (78) *wugyo ka-we jikcha(<ji-kə-cha) mak, ja-we nongos*  
 he INF-come not-3SG-can AUX(NEG) not-come AUX  
 'It is not that he couldn't come, he didn't come.'
- (79) *wugyo ji-we nəma, ja-we nongos*  
 he not-come AUX(NEG) not-come AUX  
 'It is not that he couldn't come, he didn't come.'

Sentences (78) and (79) are similar in meaning. In (79) the meaning of the first clause is that it is not that some external factor prevented him from coming, while the second clause in (78) and (79) means that he did not come of his own will. The following two examples show a parallel to the preceding.

- (80) *wugyo khri zə kəma ji-Ndza-w*  
 he rice other.than anything not-eat-3PTT  
 'Except for rice, he didn't eat anything. (he couldn't eat anything else)'

- (81) *wugyo khri zə . kama ja-ndza-w*  
 he rice other.than anything not-eat-3PTT  
 'Except for rice, he didn't eat anything.'

Example (81) means that because of his own taste, he ate only rice; while (80) means that because of some external reason, the only thing he could eat was rice.

#### 4 Summary and speculations

Summarising the above, we can tentatively conclude the following:

- a) The Cogtse dialect of Gyarong has three negation particles, *ja-*, *ji-* and *ma-*.
- b) The form *ma-* is used in the imperfect while *ja-* and *ji-* are used in the perfect.
- c) The form [mə] of Lin (1993) corresponds to *ja-* and *ji-* here.
- d) The differentiation of *ja-* and *ji-* is related to volitionality; the more volitional the main verb is, the more frequently *ja-* appears.
- e) Prohibition is always marked by *ji-*.

A similar system of negation is observed in the Kyomkyo dialect of Gyarong. Prins reports that in Kyomkyo, there is a contrast between *nga masyiong* 'I don't know' and *nga djasyiong* 'I didn't know', adding that the vowel change has to do with mood and so on. This contrasts with the volitionality meaning of *ja-* and *ji-* in Cogtse.

What then is the origin of *ja-* and *ji-*? These forms have never been described in previous studies of Gyarong. But, according to my informant, these were common at least among his grandparents' generation. If so, we can speculate that, although *ja-/ji-* have existed for a long time, *mə-* was dominant in the standard Gyarong while *ja-* and *ji-* were used as colloquial forms. It is probable that, since *mə-* is identical with the interrogative marker, the negation particle split into *ma-* and *ja-/ji-* in order to avoid confusion. As for dialects of Gyarong other than Cogtse and Kyomkyo, there is no report of the coexistence of *ja-/ji-* and *ma-*.

The historical origin of *ja-/ji-* may perhaps be in Proto-TB prohibitive \*ta. It seems plausible that this old verbal negative prefix was palatalised in Gyarong, becoming *ja-/ji-*.

#### References

- Bauman, James J., 1975, Pronouns and Pronominal Morphology in Tibeto-Burman. PhD dissertation, U. C. Berkeley.
- Chang Kun and Betty S. Chang, 1975, Gyarong historical phonology. *Bulletin of the Institute of History and Philology* 46/3:391–524.
- DeLancey, Scott, 1981, An interpretation of split ergativity and related patterns. *Language* 57/3:626–657.



- Driem, George van, 1993, The Proto-Tibeto-Burman verbal agreement system. *Bulletin of the School of Oriental and African Studies* 56/2:292–334.
- Jin Peng, 1949, Etude sur le Jyarung. *Han Hiue* 3:211–310.
- et al. 1957/58, Jiarongyu Suomohua de yuyin he xingtai (Phonology and morphology of the Suomo dialect of Gyarong) [in Chinese]. *Yuyan Yanjiu* 2:123–151, 3:71–108.
- LaPolla, Randy, 1992, On the dating and nature of verb agreement in Tibeto-Burman. *BSOAS* 55/2:298–315.
- Lin Xiangrong, 1993, *Jiarongyu yanjiu* (Gyarong language research) [in Chinese]. Chengdu: Sichuan Nationalities Press.
- Nagano, Yasuhiko, 1984, *A historical study of the rGyarong verb system*. Tokyo: Shiseido.
- 2003, Cogtse Gyarong. In Graham Thurgood and Randy LaPolla eds *The Sino-Tibetan languages*, 469–489. London: Routledge.
- Qu Aitang, 1984, Jiarongyu gaikuang (An outline of the Gyarong language) [in Chinese]. *Minzu Yuwen* 2:67–80.
- Wolfenden, Stuart N., 1929, *Outlines of Tibeto-Burman linguistic morphology*. London: Royal Asiatic Society.
- 1936, Notes on the Jyarong dialect of eastern Tibet. *T'oung Pao* 32:167–204.

# 11 *Agreement and grammatical relations in Hyow\**

---

DAVID A. PETERSON

## 1 Introduction

This paper is intended as a primarily descriptive contribution on some basic morphosyntactic characteristics of Hyow, one of two Kuki-Chin languages I have been working on in the Chittagong Hill Tracts of southeastern Bangladesh. Hyow is spoken by about two thousand people to the north and east of the town of Bandarban. The data was collected during two field trips (November 1999–April 2000 and October 2000), in addition to a short visit to verify the accuracy of the examples and paradigms included here (April 2001). The main field site was Gongru village (cf. Bernot and Bernot 1958).

First, I will demonstrate the principles of case marking for core grammatical relations, including an infrequently attested number-based ergativity split. Second, I will illustrate the language's relatively complicated system of verbal pronominal agreement. The system of agreement is not complex in terms of the number of forms that it involves (as is the case for K'cho as reported by Bedell 2000), but as we will see, formal indeterminacy gives the system considerable complications that the systems of many Kuki-Chin languages do not exhibit.

---

\* This research was made possible by a Fulbright fellowship and by a postdoctoral fellowship from the Max Planck Institute for Evolutionary Anthropology. I would also like to thank audiences at the Max Planck Institute for Evolutionary Anthropology, SEALS XI, Rice University, and Dartmouth College for helpful comments. I have worked with a number of Hyow speakers, but I am most indebted to Kewsamong Khyang and Mongtoay Khyang of Gongrupara for their tireless assistance. Lastly, and most importantly, although I remember he was initially a bit hesitant to encourage me to work in Bangladesh, this work would never have even begun if not for Jim's influence.

## 2 Nominal case marking

We begin with a discussion of nominal case marking. The examples in this section of course contain agreement morphology, but little will be said about agreement until §3.

### 2.1 Case marking of S, A, and O

At first blush, case relations are marked by a standard-looking, person-based split-ergative pattern. As seen in (1), Ss (the single argument associated with intransitive predicates) have no case marking.<sup>1</sup>

- (1) *samuyco lodo=wa (?e-)cet-aʔy-so*  
 boy Bandarban=LOC 3SGS-go-FUT-CONCL  
 'The boy is going to Bandarban.'

In (2), first and second person As (the argument associated with transitive predicates which has greater control of the action) and Os (the argument associated with transitive predicates which has less control of the action) likewise show no case marking.

- (2) a. *key vok-so ke-hleʔy*  
 1SG pig-meat 1SGA-buy  
 'I bought some pork.'
- b. *naʔ vok-so ne-hleʔy*  
 2SG pig-meat 2SGA-buy  
 'You bought some pork.'

However, (3) shows that third person As are marked by a case clitic, =*la*, which is thus glossed as an ergative marker.

- (3) *yontuʔa uy=la key ʔo-ʔoʔwey-so*  
 yesterday dog=ERG 1SG 1O-bite-CONCL  
 'Yesterday a dog bit me.'

Third person As always have this marker, and Os are always unmarked, regardless of their animacy.

Table 1 gives the full array of independent pronouns. Note that there are three numbers, and in the dual, there is an inclusive/exclusive distinction (the inclusive dual pronoun is *hnihniʔ*, but some speakers also produce and accept the formally more transparent *naʔkey*). There is no inclusive/exclusive distinction in the plural. Third person forms are built on demonstrative elements.

<sup>1</sup> Transcription in examples is for the most part standard; however, voiceless sonorants are indicated by a preceding *h*, and the vowel represented by *u* is actually somewhat more centralised than this symbol usually indicates.

Table 1: Hyow independent pronouns

	singular	dual inclusive	dual	plural
1	<i>key</i>	<i>hnihni?</i>	<i>key-hni?</i>	<i>key-ni?</i>
2	<i>naŋ</i>		<i>naŋ-hni?</i>	<i>naŋ-ni?</i>
3	<i>cu(ʔuy, ni)</i>		<i>cu-huʔy (ʔuy-, ni-)</i>	<i>cu-khol (ʔuy-, ni-)</i>

It is acceptable and not uncommon in Hyow to include independent pronominals in sentences which one might expect technically do not require them because of the extensive system of verbal pronominal morphology; thus, it is not difficult to examine the case marking of independent pronominals. When we consider various combinations of these pronominals in A and O function, we observe a wider distribution for the =*la* ergative case marker than we did initially. The examples here are all in the future tense and they are all affirmative, though as far as I have been able to determine, these case marking principles operate in any tense and also in negatives.

As shown in (4)–(7), first and second person dual and plural As are marked with =*la* if the O is a non-singular third person.

- (4) a. *keyhni?=-la cuhuʔy kihni-hyul-aʔy*  
 1D=ERG 3D 1D.A-follow-FUT  
 'We two are going to follow those two.'
- b. *keyhni?=-la cukhol kihni-hyul-aʔy*  
 1D=ERG 3PL 1D.A-follow-FUT  
 'We two are going to follow them.'
- c. *keyhniʔ(\*=-la) cu kihni-hyul-aʔy*  
 1D=ERG 3SG 1D.A-follow-FUT  
 'We two are going to follow her.'
- (5) a. *keyni?=-la cuhuʔy kini-hyul-aʔy*  
 1PL=ERG 3D 1PL.A-follow-FUT  
 'We're going to follow those two.'
- b. *keyni?=-la cukhol kini-hyul-aʔy*  
 1PL=ERG 3PL 1PL.A-follow-FUT  
 'We're going to follow them.'
- c. *keyniʔ(\*=-la) cu kini-hyul-aʔy*  
 1PL=ERG 3SG 1PL.A-follow-FUT  
 'We're going to follow her.'
- (6) a. *naŋhni?=-la cuhuʔy hnihni-hyul-aʔy*  
 2D=ERG 3D 2D.A-follow-FUT  
 'You two are going to follow those two.'
- b. *naŋhni?=-la cukhol hnihni-hyul-aʔy*  
 2D=ERG 3PL 2D.A-follow-FUT  
 'You two are going to follow them.'

- c. *naḡhni?(\*=la) cu hnihni-hyul-aʔy*  
 2D=ERG 3SG 2D.A-follow-FUT  
 'You two are going to follow her.'
- (7) a. *naḡni?=la cuhuʔy nini-hyul-aʔy*  
 2PL=ERG 3D 2PL.A-follow-FUT  
 'Y'all are going to follow those two.'
- b. *naḡni?=la cukhol nini-hyul-aʔy*  
 2PL=ERG 3PL 2PL.A-follow-FUT  
 'Y'all are going to follow them.'
- c. *naḡni?(\*=la) cu nini-hyul-aʔy*  
 2PL=ERG 3SG 2PL.A-follow-FUT  
 'Y'all are going to follow her.'

However, this is not the case if a first or second person singular A acts on a third person O, as shown in (8) and (9):

- (8) a. *key(\*=la) cu ku-hyul-aʔy*  
 1SG(=ERG) 3SG 1SG.A-follow-FUT  
 'I'm going to follow her.'
- b. *key(\*=la) cuhuʔy ku-hyul-aʔy*  
 1SG(=ERG) 3D 1SG.A-follow-FUT  
 'I'm going to follow those two.'
- c. *key(\*=la) cukhol ku-hyul-aʔy*  
 1SG(=ERG) 3PL 1SG.A-follow-FUT  
 'I'm going to follow them.'
- (9) a. *naḡ(\*=la) cu nu-hyul-aʔy*  
 2s=ERG 3SG 2SG.A-follow-FUT  
 'You're going to follow her.'
- b. *naḡ(\*=la) cuhuʔy nu-hyul-aʔy*  
 2s=ERG 3D 2SG.A-follow-FUT  
 'You're going to follow those two.'
- c. *naḡ(\*=la) cukhol nu-hyul-aʔy*  
 2s=ERG 3PL 2SG.A-follow-FUT  
 'You're going to follow them.'

The only way for first or second person singular As to be expressed in these cases would be without =*la*.

Next, first person dual and plural As are marked by =*la* if the O is a non-singular second person, as seen in (10) and (11):

- (10) a. *keyhni?=la naḡhni? kini-hyul-aʔy*  
 1D=ERG 2D 1A.2O-follow-FUT  
 'We two are going to follow you two.'

- b. *keyhni?=la naŋni? kini-hyul-aʔy*  
 1D=ERG 2PL 1.A.2O-follow-FUT  
 'We two are going to follow y'all.'
- c. *keyhniʔ(\*=la) naŋ kini-hyul-aʔy*  
 1D=ERG 2SG 1.A.2O-follow-FUT  
 'We two are going to follow you.'
- (11)a. *keyni?=la naŋhni? kini-hyul-aʔy*  
 1PL=ERG 2D 1.A.2O-follow-FUT  
 'We're going to follow you two.'
- b. *keyni?=la naŋni? kini-hyul-aʔy*  
 1PL=ERG 2PL 1.A.2O-follow-FUT  
 'We're going to follow y'all.'
- c. *keyniʔ(\*=la) naŋ kini-hyul-aʔy*  
 1PL=ERG 2SG 1.A.2O-follow-FUT  
 'We're going to follow you.'

But a first person singular A acting on a second person O may not be marked by *=la*, as seen in (12):

- (12)a. *key(\*=la) naŋ kini-hyul-aʔy*  
 1SG(=ERG) 2SG 1.A.2O-follow-FUT  
 'I'm going to follow you.'
- b. *key(\*=la) naŋhni? kini-hyul-aʔy*  
 1SG(=ERG) 2D 1.A.2O-follow-FUT  
 'I'm going to follow you two.'
- c. *key(\*=la) naŋni? kini-hyul-aʔy*  
 1SG(=ERG) 2PL 1.A.2O-follow-FUT  
 'I'm going to follow y'all.'

Similarly, second person dual and plural As are marked by *=la* if the O is a non-singular first person (13–14):

- (13)a. *naŋhni?=la keyhni? khrɔŋ-hnihni-hyul-aʔy*  
 2D=ERG 1D 1O-2D.A-follow-FUT  
 'You two are going to follow us two.'
- b. *naŋhni?=la keyni? khrɔŋ-hnihni-hyul-aʔy*  
 2D=ERG 1PL 1O-2D.A-follow-FUT  
 'You two are going to follow us.'
- c. *naŋhniʔ(\*=la) key khrɔŋ-hnihni-hyul-aʔy*  
 2D=ERG 1SG 1O-2D.A-follow-FUT  
 'You two are going to follow me.'
- (14)a. *naŋni?=la keyhni? khrɔŋ-nini-hyul-aʔy*  
 2PL=ERG 1D 1O-2PL.A-follow-FUT  
 'Y'all are going to follow us two.'

- b. *naŋmi?=*la**     *keyni?*     *khɾɔŋ-nini-hyul-aʔy*  
 2PL=*ERG*     1PL     1O-2PL.A-follow-FUT  
 'Y'all are going to follow us.'
- c. *naŋniʔ(\*=*la*)*     *key*     *khɾɔŋ-nini-hyul-aʔy*  
 2PL=*ERG*     1SG     1O-2PL.A-follow-FUT  
 'Y'all are going to follow me.'

But it is never possible to mark a second person singular A with =*la* if the O is a first person, as in (15):

- (15)a. *naŋ(\*=*la*)*     *key*     *khɾɔŋ-ni-hyul-aʔy*  
 2SG(\*=*ERG*)     1SG     1O-2SG.A-follow-FUT  
 'You are going to follow me.'
- b. *naŋ(\*=*la*)*     *keyhni?*     *khɾɔŋ-ni-hyul-aʔy*  
 2SG(\*=*ERG*)     1D     1O-2SG.A-follow-FUT  
 'You are going to follow us two.'
- c. *naŋ(\*=*la*)*     *keyni?*     *khɾɔŋ-ni-hyul-aʔy*  
 2SG(\*=*ERG*)     1PL     1O-2SG.A-follow-FUT  
 'You are going to follow us.'

Finally, for the sake of completeness, first person dual inclusive As are marked by =*la* if the O is a non-singular third person:

- (16)a. *hnihni?=*la**     *cuhuʔy*     *ni-hyul-aʔy*  
 1D.INCL=*ERG*     3D     1D.INCL.A-follow-FUT  
 'We two are going to follow those two.'
- b. *hnihni?=*la**     *cukhol*     *ni-hyul-aʔy*  
 1D.INCL=*ERG*     3PL     1D.INCL.A-follow-FUT  
 'We two are going to follow them.'
- c. *hnihniʔ(\*=*la*)*     *cu*     *ni-hyul-aʔy*  
 1D.INCL=*ERG*     3SG     1D.INCL.A-follow-FUT  
 'We two are going to follow him.'

Thus, in addition to the generalisation that all third person As are marked with =*la*, first and second person As are also marked with =*la* if they are dual or plural and if the O also is dual or plural.

This distribution of the ergative marker is noteworthy in two respects. First of all, ergativity splits which are based on number are relatively rare cross-linguistically, so it is good to have another example to generalise from. Second, the Hyow case also involves marking more distinctions in the non-singular categories than in the singular categories, which is in contradiction to a principle which Dixon (1994) claims often governs number-based splits (92).

However, for the split to occur where it does makes good functional sense. Singular As and Os make the clearest agents and patients, whereas dual and plural As and Os do not have as clear a concentration of agency and patienthood. Thus, if ergative splits are to be understood in terms of a marking of less-likely agents (following the general functional explanation for case-marking splits suggested by Comrie (1989)), agents which are non-

singular and which are acting on non-singular entities are less prototypical, and should be expected in some cases to be marked ergatively.

Dixon does mention that in Chukchi, plurals tend to make more distinctions than singulars, in direct contrast to the patterns frequently attested in Australian languages. Perhaps there are really two competing tendencies at work here: a tendency to make more formal distinctions in singulars than in plurals, and also the well-supported tendency to mark less prototypical As. In some languages the first tendency manifests itself clearly, whereas in other languages the second tendency takes precedence.

Finally, I should note that there are two things which are still not clear about the distribution of *=la*. For some members of the speech community, the use of the marker does not appear to extend beyond its use with third person subjects; these are mostly older speakers, so I would hypothesise that the spread of *=la* to first and second person subjects is a change in progress among the younger speakers of the language. Secondly, even for the speakers that have it, it is not always the case that they use it under direct elicitation or in spontaneous discourse. There are many possible explanations for these inconsistencies, including differences in the semantics of the predicates involved, degree of control of the agent, and so on, which I have not yet tested to my satisfaction. Resolution of these issues will have to be the subject of future inquiry.

## 2.2 Case marking of multiple objects

Verbal agreement in Hyow codes the O of monotransitives and the *recipient* of three place verbs, but recipients are marked with the locative case clitic, as can be seen in (17)–(20).

- (17) *cu=la key=a cɔ ʔe-pek*  
 3SG=ERG 1SG=LOC book 1O-give  
 'He gave me a book.'
- (18) *key naŋ=a cɔ kini-pek*  
 1SG=ERG 2SG=LOC book 1.A.2O-give  
 'I gave you a book.'
- (19) *key piya naŋ=a kini-dɔn-aʔy-sɔ*  
 1SG wife 2SG=LOC 1.A.2O-show-FUT-CONCL  
 'I'm going to show you my wife.'
- (20) *naŋ piya key=a khrɔŋ-nɔ-dɔn-aʔy-sɔ-wey*  
 2SG wife 1SG=LOC 1O-2SG.A-show-FUT-CONCL-INTERR  
 'Are you going to show me your wife?'

The treatment of objects is thus neither purely of the direct object type, nor of the primary object type. It exhibits a split between the two marking strategies.

## 3 Pronominal agreement

Moving on to verbal agreement morphology, the intransitive affirmative paradigm is given in Table 2 ( $\Sigma$  indicates the position of the stem with respect to agreement markers).



**Table 2:** Intransitive affirmative

1s	1d	1d incl	1p	2s	2d	2p	3s	3d	3p
<i>kV-Σ</i>	<i>kihni-Σ</i>	<i>ni-Σ</i>	<i>kini-Σ</i>	<i>nV-Σ</i>	<i>hnihni-Σ</i>	<i>nini-Σ</i>	<i>(ʔV-)Σ</i>	<i>(ʔi)hni-Σ</i>	<i>(ʔi)ni-Σ</i>

Concerning the phonological content of these agreement markers, a *V* in markers that contain it is an underspecified vowel which harmonises with the first stem vowel. Most of these markers consist of elements well known from elsewhere in Kuki-Chin (*k* in first person, *n* in second person; the *ni* plural element and the *hni* dual element are also clearly direct reflections of the independent pronominal morphology). The distribution of the third singular *ʔV-* marker is not clearly definable at present. There are some roots which it definitely cannot occur with, roots which it definitely must occur with, and others for which it may or may not be present, with no apparent semantic difference to condition its presence or absence. It would seem to be lexically conditioned.

Interestingly, negatives are not indicated by a separate negative marker, but are encoded instead by a distinct suffixal agreement pattern given in Table 3.

**Table 3:** Intransitive negative

1s	1d	1d incl	1p	2s	2d	2p	3s	3d	3p
<i>Σ-ŋa</i>	<i>Σ-hniʔŋa</i>	<i>Σ-pu</i>	<i>Σ-ʔuŋa</i>	<i>Σ-ti</i>	<i>Σ-hniʔti</i>	<i>Σ-cu</i>	<i>Σ-aʔ</i>	<i>Σ-huʔy</i>	<i>Σ-ʔu</i>

In the first and second dual of these markers, the first part (*hniʔ*), like the *hniʔ* of the independent pronouns, is etymologically from the number 'two' (< \**g-nis*); the second element in these markers is the same person marker seen in first person (*-ŋa*) and second person (*-ti*) in the other numbers. The first person plural marker also clearly consists of the general first person marker (*-ŋa*) and the same plural element that may be seen in the third person plural (*-ʔu*).

This separate set of agreement markers is noteworthy in light of full suffixal agreement paradigms attested elsewhere in the family, such as Tedim (otherwise known as Tiddim Chin) and Sizang, as seen in 21 and 22.

(21) Tedim (Henderson 1965:109–111):

<i>Σ-iŋ</i>	'1st singular'	<i>Σ-uŋ</i>	'1st plural excl'	<i>Σ-haŋ</i>	'1st plural incl'
<i>Σ-teʔ</i>	'2nd singular'	<i>Σ-uʔteʔ</i>	'2nd plural'		
<i>Σ-(iʔ)</i>	'3rd singular'	<i>Σ-uʔ</i>	'3rd plural'		

(22) Sizang (Stern 1963:264):

<i>Σ-bo-ŋ</i>	'I not'	<i>Σ-bua-u-ŋ</i>	'we not'
<i>Σ-bua te</i>	'you not'	<i>Σ-bua-u-te</i>	'y'all not'
<i>Σ-boʔ/bua</i>	'he/they not'		

In Tedim and Sizang these markers actually occur in both positive and negative contexts, but the forms in (21) and (22) are nonetheless in many cases obviously related to the ones seen in the Hyow negative paradigm.

For transitives, the fundamental generalisation is that there is no coding of O number. First, consider Table 4 which gives the paradigm for third person Os.

Table 4: Third person O (affirmative and negative)

	1sA	1dA	1d inclA	1pA
affirmative	<i>kV-Σ</i>	<i>kihni-Σ</i>	<i>ni-Σ</i>	<i>kini-Σ</i>
negative	<i>Σ-ŋa</i>	<i>Σ-hniŋa</i>	<i>Σ-pu</i>	<i>Σ-ŋuŋa</i>

	2sA	2dA	2pA
affirmative	<i>nV-Σ</i>	<i>hniŋni-Σ</i>	<i>nini-Σ</i>
negative	<i>Σ-ti</i>	<i>Σ-hniŋti</i>	<i>Σ-cu</i>

	3sA	3dA	3pA
affirmative	<i>Σ</i>	<i>(ŋ)hni-Σ</i>	<i>(ŋ)ni-Σ</i>
negative	<i>Σ-a?</i>	<i>Σ-huŋy</i>	<i>Σ-ŋu</i>

A consideration of these forms reveals that they are essentially the intransitive markers used with transitive stems—third person O is entirely unmarked. Note, however, that third singular As are obligatorily *unmarked* in the transitive affirmative paradigm as opposed to the intransitive paradigm, where third singular Ss are sometimes marked. We will see a possible motivation for this gap shortly, when we consider the first person O forms.

Somewhat more complicated are the second person O forms, as seen in Table 5. The basic pattern is similar to that in Table 4: there is a prefixal marker for the A argument in the affirmative which is absent in the negative; negative As are expressed by predictable suffixal morphology.

Table 5: Second person O (affirmative and negative)

	1s.A	1d.A	1p.A
affirmative	<i>kini-Σ</i>	<i>kini-Σ/kihni-Σ</i>	<i>kini-Σ</i>
negative	<i>kini-Σ-ŋa</i>	<i>kini-Σ-hniŋa</i>	<i>kini-Σ-ŋuŋa</i>

	3s.A	3d.A	3p.A
affirmative	<i>ni-Σ</i>	<i>ni-Σ/hni-Σ</i>	<i>ni-Σ</i>
negative	<i>ni-Σ-a?</i>	<i>ni-Σ-huŋy</i>	<i>ni-Σ-ŋu</i>

The variation in coding for first person dual and third person dual A may perhaps be explained in the following way: there is a second person O marker *ni-* which sometimes can be pushed out of its slot by a *hni-* dual A marker for dual first and third person As. On this analysis, the *-ni* of the *kini-* in the first singular A form would be this second person O marker, as would the *ni* in the third person singular A form; the *ni* in the first plural and third plural A forms could either be the second person O marker or a plural marker. In the first and third dual A forms, however, either the second person O *-ni* marker, or the *-hni* dual A marker might be present in the affirmative.

The persistence of *ki-* at the beginning of the first person A on second person O forms in the negative is anomalous, however, as indicators of A are otherwise invariably expressed postverbally in the negative. I do not believe, however, that synchronically there is anything to say about this discrepancy.

The most complicated agreement patterns involve first person Os. Table 6 gives the simplest parts of the paradigm, those for first persons acted on by second or third person duals and plurals. These forms consist simply of a first person O marker *khɔŋ* followed by the normal preverbal or postverbal S/A agreement markers.<sup>2</sup>

**Table 6:** First person O (including dual inclusive) with third dual and third plural A (affirmative and negative)

	2d.A	2p.A
affirmative	<i>khɔŋhni-Σ</i>	<i>khɔŋmi-Σ</i>
negative	<i>khɔŋ-Σ-hni?ti</i>	<i>khɔŋ-Σ-cu</i>
	3d.A	3p.A
affirmative	<i>khɔŋni-Σ</i>	<i>khɔŋni-Σ</i>
negative	<i>khɔŋ-Σ-huɔy</i>	<i>khɔŋ-Σ-?u</i>

Next, Table 7 shows that for third singular As, you get either the *khɔŋ*- first person object marker, or you get a different marker, a prefixal underspecified vowel formally identical to the one sometimes seen for third person singular Ss. This marking perhaps motivates the absence of third person A coding by such a marker in the affirmative of transitives, which we saw earlier.

**Table 7:** First person O with third singular A (affirmative and negative)

	3s.A
affirmative	<i>ʔV-Σ</i> or <i>khɔŋ-Σ</i>
negative	<i>khɔŋ-Σ-a?</i> ( <i>ʔV-Σ-a?</i> )

Lastly, the second singular A on first person O forms seen in Table 8 either do or do not involve the *khɔŋ*- object marker, but always include one of the two second person participant markers that we have seen already. If the *khɔŋ*- object marker is not present, though, these forms have considerable potential for ambiguity. A form like *ni-hyul*, for instance, could indicate 'you follow me,' 'they follow him,' 'they follow you,' or 'we (inclusive) follow him'.

**Table 8:** First person O with second singular A (affirmative and negative)

	2sA
affirmative	<i>khɔŋnV-Σ, khɔŋni-Σ, nV-Σ, ni-Σ,</i>
negative	<i>khɔŋ-Σ-ti, Σ-ti</i>

<sup>2</sup> This element appears to have arisen from the independent lexical item *khɔŋ* 'man'. A comparable grammaticalisation of 'man' has been reported in Kiranti languages (Ebert 1991, Michailovsky 2001).

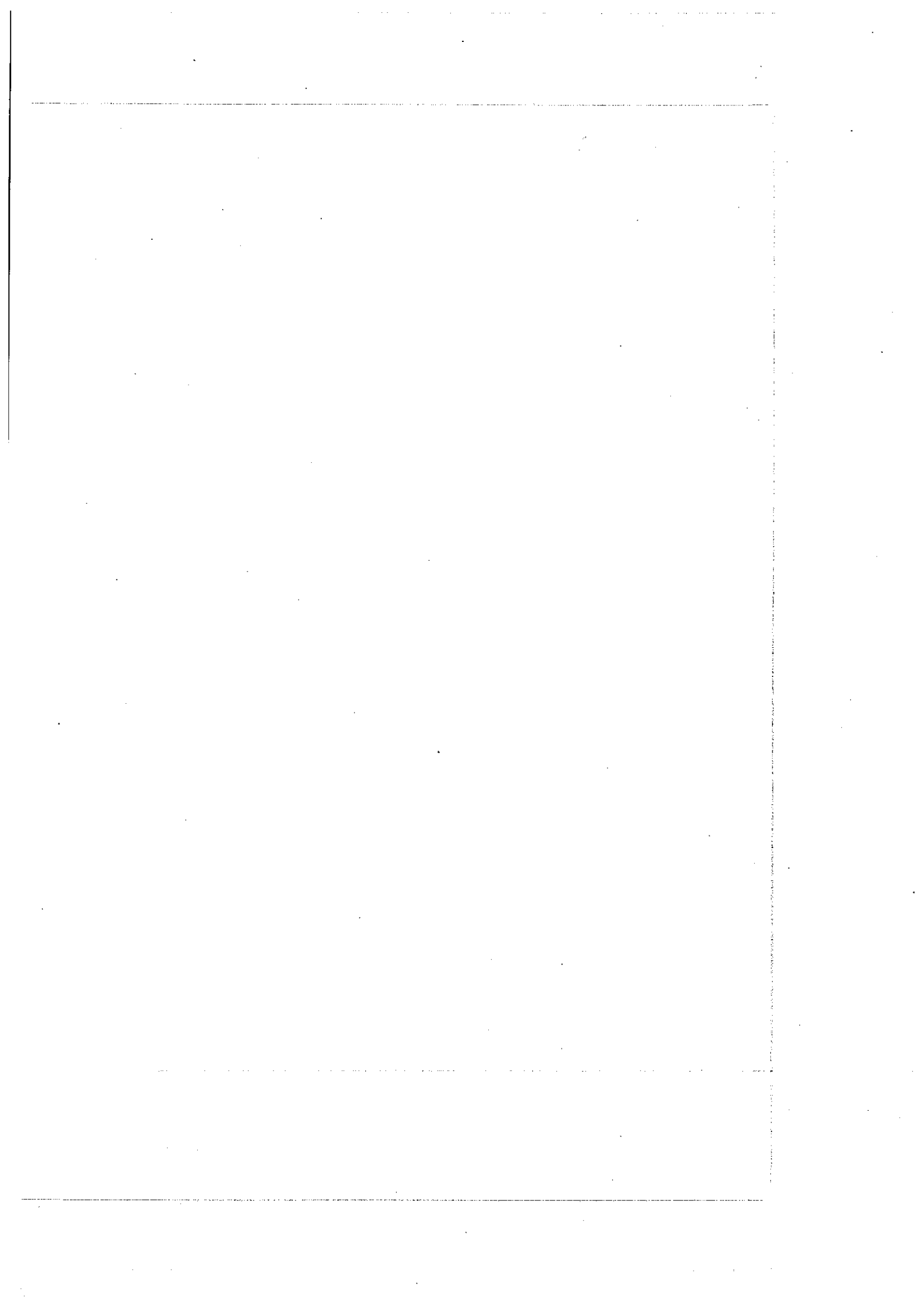
#### 4 Conclusion

To summarise what I have tried to demonstrate in this paper concerning Hyow morphosyntax: verbal pronominal agreement has a roughly nominative/accusative alignment, though extensive and idiosyncratic syncretism obscures this; in terms of verbal agreement, multiple objects show a primary object orientation; finally, nominal alignment is split between neutral and ergative/absolute for core case relations, and multiple objects are treated in a direct object fashion.

Future work will have to concentrate on third singular intransitive marking, to better account for its distribution, and on finding (semantic) correlates for the different options seen in many of the transitive paradigm slots, e.g. the *ŋV-* vs. the *khrɔŋ-* first person object markers in the third singular A on first person O portion of the paradigm. It seems likely that factors such as focus may play a role in the choice of one or another of these options, as suggested by Hartmann (2000) to account for a similar variability seen in Daai, but I have not yet had an opportunity to test this possibility. Also, some issues concerning the distribution of the number-based ergativity split in first and second persons were mentioned at the end of section 2. Further investigation will shed light on these problems.

#### References

- Bedell, George, 2000, Agreement in Cho. Paper presented at 33rd International Conference on Sino-Tibetan Languages and Linguistics, Bangkok/Trang.
- Bernot, Denise and Lucien Bernot, 1958, *Les khyang des collines de Chittagong (Pakistan oriental)*. Paris: Plon.
- Comrie, Bernard, 1989, *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Dixon, Robert Malcolm Ward, 1994, *Ergativity*. Cambridge: Cambridge University Press.
- Ebert, Karen, 1991, Inverse and psuedo-inverse prefixes in Kiranti languages: evidence from Belhare, Athpare, and Dungmali. *Linguistics of the Tibeto-Burman Area* 14/1:72–92.
- Hartmann, Helga, 2000, The Daai pronouns and pronominal agreement system: alternate sets and focus. Paper presented at the fifth international Pan-Asiatic linguistics symposium, Ho Chi Minh City.
- Henderson, Eugénie J.A., 1965, *Tiddim Chin: A descriptive analysis of two texts*. Oxford: Oxford University Press.
- Michailovsky, Boyd, 2001, Limbu *nous autres* and 1st person morphology. *Linguistics of the Tibeto-Burman Area* 24/1:145–156.
- Stem, Theodore, 1963, A provisional sketch of Sizang (Siyin) Chin. *Asia Major* 10:222–278.



# 12 *Contact induced variation and syntactic change in the Tsat of Hainan*<sup>1</sup>

---

GRAHAM THURGOOD AND FENGXIANG LI

## 1 Introduction

Tsat is an Austronesian language located on Hainan Island. The 1982 census lists 4131 Utsat people largely in the villages of Huihui and Huixin near Sanya on Hainan Island (which has recently been designated as a province), 3849 of whom still speak Tsat. Virtually all the Tsat speakers also speak one or more Chinese dialects, typically Fukienese or Cantonese, the languages of business, and Mandarin, the language of school.

Genetically the closest language to Tsat is the Northern Roglai of Vietnam, a Chamic language (Austronesian) which it split off from first around 982, with a second migration probably around 1471. Despite the genetic closeness, Tsat is now radically different both phonologically and syntactically from Northern Roglai. Phonologically, Northern Roglai is sesquisyllabic and atonal whereas Tsat is monosyllabic and fully tonal. Structurally, Northern Roglai is much, much more like the other Chamic languages of Vietnam which, in turn resemble the Mon-Khmer languages of the region, while Tsat, not surprisingly, is much like the Chinese dialects that surround it.

---

<sup>1</sup> This paper is a modified version of a paper presented first presented at the 34th International Conference on Sino-Tibetan Languages and Linguistics, held in Kunming, Yunnan, October 2001, at the Linguistics Society of America annual meeting, held in San Francisco, January 3-6, 2002, and at the Berkeley Linguistics Society meeting, February, 2002.

The abbreviations used here are: GEN genitive, NEG negative, SG singular, PERF perfective, PAST past, PRES present, CLF classifier.

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 185–200.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.

Increasingly, all that remains of Tsat is the vocabulary, with the structure being Chinese, albeit with Tsat lexical items. Thus Tsat provides some exceptionally clear examples of contact-induced syntactic variation and change. Work has been done on genetic affiliations of Tsat (Benedict 1941), the history of the Chamic languages including Tsat (for example Thurgood 1996, 1999), and on the description of Tsat itself, Ouyang and Zheng (1983), Zheng (1986, 1997), with the later work by Zheng including numerous valuable observations on the influence of Chinese on Tsat, both identifying Chinese borrowings and commenting on Chinese structural influence.

## 2 Tsat contact

Changes in Tsat resulting from contact with neighboring languages of Hainan are quite obvious. Phonologically, it has gone from sesquisyllabic and registral to monosyllabic and tonal. Lexically, it contains four layers of borrowings reflecting contact patterns since the Tsat arrival in Hainan: a Hlai (= Li) strata, an early Chinese level reflecting early contact with speakers of Min dialects such as Hainanese and various Cantonese dialects, a later layer of contact with the Mandarin spoken by the army and officials, and most recently the Mandarin of the schools. The intensity of the last layer of contact looks to have initiated rapid and through restructuring of the language.

Here we will restrict our examination to four constructions with extant variation, two involving word orders that do not correlate with VO order (Dryer 1992) and two involving word orders that do correlate: genitive constructions, demonstratives and head nouns, adjectives and head nouns, and comparative constructions. All show the structural influence of Chinese.

### 2.1 *Genitive (associative) constructions*

The genitive patterns have been divided into those with full noun phrases as the genitive and those involving pronouns, reflecting the differences in their historical paths of change.

#### 2.1.1 *Genitives with full noun phrases*

In Northern Ronglai full NPs are postposed. In Tsat, even in the most colloquial, non-Sinicised texts, full GenNPs already show Chinese influence: all full genitive NPs are preposed with the genitive construction marked by *sa*<sup>33</sup>. In the case of locative NPs, the preposed NPs look to be developing into prepositions. Elsewhere, the genitive marker is a marker of pre-head modification.

## Northern Ronglai: Nh GENNP

- (1) *gaʔ sa:k* (Lee 1966:65)  
 roof house  
 'the roof of the house'

Tsat (colloquial): GENNP *sa*<sup>33</sup> Nh

- (2) *?a<sup>44</sup>?ba<sup>44</sup> sa<sup>33</sup> sa:ŋ<sup>33</sup>* (Zheng 1997:70)  
 father's.older.brother GEN house  
*bófù de fángzi*  
 'father's eldest brother's house.'
- (3) *?ia<sup>33</sup>?be<sup>24</sup> sa<sup>33</sup> ka:n<sup>33</sup>* (Zheng 1997:71)  
 river GEN fish  
*hélǐ de yú*  
 'the fish in the river'
- (4) *...pyay<sup>33</sup> sa<sup>33</sup> za:ŋ<sup>22</sup>* (Zheng 1997:95)  
 ...village GEN person  
*...cūn de rén*  
 'people of the village...'

Tsat (Mandarinised with *ti*<sup>33</sup>): GENNP *ti*<sup>33</sup> HeadNP

The same pattern with the genitive borrowed from Mandarin.

- (5) *tan<sup>33</sup> khua<sup>55</sup> ti<sup>33</sup> si<sup>44</sup>ha:w<sup>21</sup>, la<sup>33</sup> pyay<sup>33</sup> sa<sup>33</sup> za:ŋ<sup>22</sup>* (Zheng 1997:4.1.4)  
 arrive daybreak GEN after under village GEN person  
*dào tiānliàng de shǐhòu xià cūn de rén*  
*dìer tiāntiānliàng shǐhòu, cūnlǐ de nánrén hé nǚrén...*  
 'Early the next morning, the villagers...'
- (6) *sui<sup>44</sup>tso<sup>21</sup> kay<sup>43</sup>kak<sup>24</sup> kh'a:y<sup>33</sup>ph'a:ŋ<sup>33</sup> ti<sup>33</sup> sin<sup>33</sup>zi<sup>24</sup>* (Zheng 1997:3.3.3)  
 along.with reform bloom GEN penetrate  
*suǐzhe gǎigé kāifàng de shēnrù*  
*suǐzhe gǎigé kāifàng de shēnrù*  
 'As the Reform and Open-door policies continue...'

Mandarin: Poss *de* NP

- (7) *duìzhǎng de érzi* (Zheng 1997:71)  
 captain GEN child  
 'the captain's son'
- (8) *lǎoshī de shū* (Zheng 1997:71)  
 teacher GEN book  
 'the teacher's books'

In the genitives, as with the other constructions, the more Mandarinised variants tend to co-occur with borrowed Mandarin grammatical markers (which in some cases seem to



mark the construction), tend to have more Mandarin borrowings in the sentence (marked in this paper through the underlining of both the borrowed Tsat term and its corresponding Mandarin (in Pinyin)), and, if one examines the texts, occur in the more Mandarinised texts. For instance, the text on the origin of the Tsat has less Mandarin influence than does the text describing the Japanese invasion of Hainan with its Mandarin-influenced political content.

**Table 1:** Noun modification: Genitives (full NPs)

	simple modification	prehead <i>sa</i> <sup>33</sup> pattern
Northern Roglai	Nh GENNP	—
Colloquial Tsat	—	GENNP <i>sa</i> <sup>33</sup> Nh
Mandarinised Tsat	—	GENNP <i>tj</i> <sup>33</sup> Nh
Mandarin	—	GENNP <i>de</i> Nh

In the case of full noun phrases, the Northern Roglai post-head genitives have been totally replaced by Chinese-influenced prehead genitive constructions. Nonetheless, contact has resulted in variation as the colloquial pattern uses a Tsat genitive marker while the Mandarinised pattern uses a borrowed genitive marker. Both patterns show the structural influence of Chinese.

### 2.1.2 Genitives with pronouns

Northern Roglai: Nh GENPr

- (9) *sa:k hã* (Lee 1966:65)  
 house you  
 'your house'

Tsat (colloquial): Nh GENPr

- (10) *ŋa:n<sup>33</sup> kaw<sup>33</sup> ki<sup>24</sup>* (Zheng 1997:97)  
 hand I painful  
*shǒu wǒ tòng*  
*wǒ de shǒu tòng.*  
 'My hand hurts.'

- (11) *ko<sup>24</sup>bu<sup>24</sup> naw<sup>33</sup> sa<sup>24</sup>* (Zheng 1997:92)  
 head.hair she messy  
*tóufa tā luàn*  
*tā de tóufa luàn*  
 'Her hair is messy.'

Tsat (Chinese influenced, with *sa*<sup>33</sup>):

- (12) *naw*<sup>33</sup> *sa*<sup>33</sup> *ko*<sup>24</sup>*bu*<sup>24</sup> *sa*<sup>24</sup>. (Zheng 1997:97)  
 she GEN head.hair messy  
*tā de tóufa luàn.*  
 'Her hair is messy.'

- (13) *ha*<sup>33</sup> *sa*<sup>33</sup> *ʔa*<sup>11</sup>*sa.w*<sup>11</sup> *sa*<sup>33</sup> *ha*<sup>33</sup>. (Zheng 1997:87)  
 you GEN elder.brother. seek you  
*nǐ de sǎo zhǎo nǐ*  
 'Elder brother's wife seeks you.'

Mandarin: GENPr Nh

- (14) *wǒ fùqin shì tā bófù* (Zheng 1997:77)  
 I father be he uncle  
 'My father is his uncle.'

GENPr *de* Nh

- (15) *wǒ de shǒu tòng* (Zheng 1997:97)  
 I GEN hand painful  
 'My hand hurts.'

Table 2: Noun modification: Genitives (pronouns)

	simple modification	prehead <i>sa</i> <sup>33</sup> pattern
Northern Roglai	Nh GENPr	—
Colloquial Tsat	Nh GENPr	—
Mandarinised Tsat	—	Pr <i>sa</i> <sup>33</sup> Nh
Mandarin	Pr Nh	Pr <i>de</i> Nh

### 2.1.3 Demonstratives and head nouns

Like genitive pronouns, demonstratives are postposed in Northern Roglai and the colloquial Tsat, but preposed in Chinese-influenced Tsat and Mandarin. In the Chinese-influenced Tsat, however, the demonstratives are often accompanied by a genitive marker, a pattern that matches the adjective plus genitive construction immediately below, a construction that reflects Mandarin influence.

Northern Roglai:

- (16) *sa:k ghen ʔunǐ* (Lee 1966:65)  
 house big this  
 'this big house'

- (17) *dua +a:k labu? ?anā? sia:p nū ?anī la sa:k* (Lee 1966:66)  
 two person plural child good he this in house  
 'these two children of his in the new house'

Tsat:

- (18) *?ay<sup>33</sup> ni<sup>33</sup> sat<sup>24</sup> ?an<sup>33</sup>.* (Zheng 1997:84)  
 water this truly cold  
*shuǐ zhè zhēn lěng*  
*zhè shuǐ zhēn lěng.*  
 'This water is very cold.'

Tsat (Chinese influenced): (this + GEN) + CLF

- (19) *ni<sup>33</sup> sa<sup>33</sup> ta<sup>11</sup> p<sup>h</sup>an<sup>32</sup> pi<sup>11</sup> kiaw<sup>33</sup> lu<sup>33</sup>.* (Zheng 1997:75)  
 this GEN one CLF CM much  
*zhè de yī fèn bǐjiào duō*  
*zhè yī fèn bǐjiào duō.*  
 'This portion is bigger.'

Mandarin: this + CLF

- (20) *zhè lù.* (Zheng 1997:75)  
 this road  
 'This road...'

- (21) *zhè shuǐ...* (Zheng 1997:84)  
 this water...  
 'This water...'

the classifier version

- (22) *zhè gè dà fángzi*  
 this CLF big house  
 'this big house'

Needless to say, the demonstrative-noun order is a result of Chinese contact. This word order change induced by extensive and prolonged contact with Chinese is quite systematic and pervasive throughout the grammatical system of Tsat. It is found in texts collected from the same speaker by Zheng Yiqing in the 1980s published in Zheng (1997). It is interesting to note that the borrowed patterns are found in texts that describe more recent phenomena, whereas the native patterns are used in texts of traditional stories.

**Table 3:** Noun modification: Demonstratives

	simple modification	prehead <i>sa<sup>33</sup></i> pattern
Northern Roglai	Nh Dem	—
Colloquial Tsat	Nh Dem	—
Mandarinised Tsat	—	Dem <i>sa<sup>33</sup></i> Nh
Mandarin	Dem Nh	Dem CLF Nh

## 3 Adjectives and head nouns

Northern Roglai has postposed adjectives as does the colloquial Tsat, while the Mandarinised Tsat and Mandarin itself have preposed adjectives.

Northern Roglai: postposed adjectives

- (23) *sa:k ghen ʔunt* (Lee 1966:65)  
 house big this  
 'this big house'

Tsat: postposed adjectives

- (24) *na<sup>11</sup>tsun<sup>33</sup> pyoŋ<sup>ʔ22</sup> poy<sup>24</sup>:* (Zheng 1997:1.1.9)  
 bird big say  
*niǎo dà shuō*  
*niǎo dà shuō:*  
 'The big bird said:...'

- (25) *th<sup>33</sup>un<sup>33</sup>zau<sup>33</sup> pyoŋ<sup>ʔ22</sup> siv<sup>ʔ43</sup>liang<sup>11</sup>,* (Zheng 1997:1.2.21)  
 tree big relax.in.cool.place  
*shù dà xiēliáng*  
*dà shù xià xiūxi de,*  
 relaxed under a big tree,

Tsat (Mandarin-influenced): preposed adjectives

- (26) *hu<sup>11</sup>tsa:m<sup>ʔ22</sup> mi<sup>33</sup> san<sup>11</sup> na:y<sup>ʔ22</sup> sin<sup>33</sup> na:y<sup>ʔ22</sup> pa:w<sup>33</sup>,* (Zheng 1997:4.2.1)  
 Tsat we believe good heart good reward  
*Huízú wǒmen xìn hǎo xīn hǎo bào*  
*wǒmen Huízú rén xiāngxìn hǎo xīn de rén yīdìng dédào bào...*  
 'We Tsat people believe that people with kind hearts will be rewarded...'

- (27) ... *kiu<sup>33</sup> san<sup>33</sup>,* (Zheng 1997:2.1.1)  
 ...old village  
 ...*jiù cūn*  
 ... *jiù cūn*  
 '... the old village'

preposed with *sa<sup>33</sup>*, a calque on Mandarin *de*

- (28) *na:y<sup>ʔ22</sup> sa<sup>33</sup> saŋ<sup>33</sup>hua<sup>24</sup>,* (Zheng 1997:2.1.10)  
 good GEN life  
*hǎo de shēnghuó*  
*hǎo de shēnghuó*  
 '(the) good life'

Mandarin

- (29) *dà niǎo shuō* (Zheng 1997:1.1.9)  
 big bird say  
 'The big bird said:'

preposed with *de*, a 'genitive' marker

(30) *hǎo de shēnghuó,*  
 good GEN life  
 '(the) good life'

(Zheng 1997:2.1.10)

Note that under the influence of Chinese, the preposed adjectives of Tsat are often accompanied by a genitive marker (or, as Li and Thompson (1981:113–116) term it, an 'associative' marker).

**Table 4:** Noun modification: Adjectives

	simple modification	prehead <i>sa</i> <sup>33</sup> pattern
Northern Roglai	Nh Adj	—
Colloquial Tsat	Nh Adj	—
Mandarinised Tsat	Adj Nh	Adj <i>sa</i> <sup>33</sup> Nh
Mandarin	Adj Nh	Adj <i>de</i> Nh

#### 4 The spread of the *sa*<sup>33</sup> construction

The *sa*<sup>33</sup> construction is a calque on the Mandarin *de* construction illustrated throughout this paper: *X sa*<sup>33</sup>/*de* NP, in which the first element (X) modifies the final NP. This construction, marked by , is expanding in use. Its initial use appears to have been with preposed genitive NPs involving full NPs and then expanded to other parts of the grammar. The three constructions already discussed show this movement from posthead to prehead using the *sa*<sup>33</sup> construction (see Table 5).

**Table 5:** Noun modification: Spread of the *sa*<sup>33</sup> construction

	Genitive NP	Genitive pr	Dem	Adj
Northern Roglai	Nh NP	Nh Pr	Nh Dem	Nh Adj
Colloquial Tsat	NP <i>sa</i> <sup>33</sup> Nh	Nh Pr	Nh Dem	Nh Adj
Mandarinised Tsat	NP <i>sa</i> <sup>33</sup> Nh	Pr <i>sa</i> <sup>33</sup> Nh	Dem <i>sa</i> <sup>33</sup> Nh	Adj Nh Adj <i>sa</i> <sup>33</sup> Nh
Mandarin	NP <i>de</i> Nh	Pr Nh Pr <i>de</i> Nh	Dem Nh Dem CLF Nh	Adj Nh GENNP <i>de</i> Nh

For full NP genitives, the older Northern Roglai postposed NPs have been completely replaced by preposed NPs using the *sa*<sup>33</sup> construction; for genitive pronouns, the older postposed genitive pronouns are still used in more colloquial contexts, but in more Mandarinised speech these are now preposed. For demonstratives, the situation is more

complex. All demonstratives were postposed in Northern Rglai and tend to retain this posthead position in the more colloquial Tsat. However, in the more Mandarinised speech demonstratives are variably preposed through the use of the *sa<sup>33</sup>* construction. The distribution between the postposed and the preposed with *sa<sup>33</sup>* variants appears to correlate albeit only loosely with both the register and the type of NP involved. Adjectives were postposed in Northern Rglai, are postposed in the more colloquial Tsat contexts, and even in the most Mandarinised Tsat texts still remain postposed some of the time, at other times being preposed using the *sa<sup>33</sup>* construction. The use of this pattern has expanded beyond the examples in this paper to include prehead relative clauses, a construction highly marked for an SVO language like Tsat.

## 5 Comparative constructions

The existence of contact-induced word order variation is obvious in the two distinct Tsat comparative patterns: the native pattern is inherited from Chamic; the other is borrowed from Chinese:

- X - Adj - CM/ST     native pattern  
 X - CM/ST - Adj     Chinese influenced pattern

In the native pattern, Zheng (1997:75) notes that the word order is quality-marker-standard (X - Adj - CM/ST), that is, the quality being compared, followed by the preposition *la:w<sup>32</sup>* 'CM; pass' (which serves as the comparative marker), followed by the standard of comparison, typically a pronoun. The extent or degree of the quality may also be marked, in which case it is through modification of the quality.

When relationships are compared, the comparative marker is the preposition *la:w<sup>32</sup>* 'CM; pass; exceed', derived from a verb. For example,

Tsat (colloquial):

- (31) *naw<sup>33</sup> ma<sup>243</sup> la:w<sup>32</sup> ha<sup>33</sup>.* (Zheng 1997:75)  
 he fat CM you  
*tā pàng bǐ nǐ*  
*tā bǐ nǐ pàng.*

'He is fatter than you.'

- (32) *lu<sup>33</sup> pyoŋ<sup>22</sup> la:w<sup>32</sup> ʔbo<sup>11</sup>koy<sup>24</sup>.* (Zheng 1997:89)  
 coconut.palm big CM pomelo  
*yēzi dà guò yòuzi*

*yēzi bǐ yòuzi dà.*

'The coconut palm is bigger than the pomelo.'

- (33) *ʔa<sup>11</sup>ko<sup>33</sup>    ph<sup>h</sup>ay<sup>ʔ43</sup>    k<sup>h</sup>a:n<sup>21</sup>    na:y<sup>ʔ22</sup>    la:w<sup>ʔ22</sup>    ʔa<sup>11</sup>th ay<sup>11</sup>* (Zheng 1997:75)  
 elder.brother read book good CM younger.brother  
*gēge    dú    shū    hǎo    guò    dìdì*  
*gēge xuéxí bǐ dìdì hǎo.*  
 'Elder brother studies more than younger brother.'

### 5.1 *Tsat influenced by Chinese*

However, as Zheng goes on to note, under the influence of Chinese, comparatives often follow a Chinese order, namely, comparative marker, standard, quality (X - CM/ST - Adj), using *pi<sup>11</sup>* 'CM; compare' borrowed from Chinese to mark the comparison. For example:

- (34) *kaw<sup>33</sup>    pi<sup>11</sup>    ha<sup>33</sup>    tsat<sup>24</sup>tso<sup>33</sup>    ki<sup>33</sup>    sun<sup>33</sup>.* (Zheng 1997:75)  
 I CM you short three inch  
*wǒ    bǐ    nǐ       ǎi    sān    cùn*  
*wǒ bǐ nǐ ǎi sān cùn.*  
 'I am three inches shorter than you.'

- (35) *mi<sup>33</sup>    sa<sup>33</sup>    saŋ<sup>33</sup>huat<sup>24</sup>,    ta<sup>11</sup>    zay<sup>33</sup>    pi<sup>43</sup>    ta<sup>11</sup>    zay<sup>33</sup>    pu<sup>33</sup>    na:y<sup>ʔ22</sup>,*  
 (Zheng 1997:2.1.4)  
 we GEN life, one day CM one day NEG good  
*wǒmen de shēnghuó yī tiān bǐ yī tiān bù hǎo*  
*wǒmen de shēnghuó yītiān bǐ yītiān chà...*  
 '...our life went downhill each day.'

- (36) *zin<sup>11</sup>min<sup>11</sup>    sa<sup>33</sup>    saŋ<sup>33</sup>huat<sup>24</sup>    ta<sup>11</sup>    zay<sup>33</sup>    pi<sup>43</sup>    ta<sup>11</sup>    zay<sup>33</sup>    na:y<sup>ʔ22</sup>    a<sup>0</sup>.*  
 (Zheng 1997:2.1.16)  
 people GEN life one day CM one day good PART  
*rénmǐn de shēnghuó yī tiān bǐ yī tiān hǎo a*  
*rénmǐn de shēnghuó cái yītiān bǐ yītiān hǎo yuè a.*  
 '...people's lives began to get better and better.'

Mandarin

- (37) *wǒ    bǐ    nǐ       ǎi    sān    cùn* (Zheng 1997:75)  
 I CM you short three inch  
 'I am three inches shorter than you.'

- (38) *tā    bǐ    nǐ    pàng* (Zheng 1997:75)  
 he CM you fat  
 'He is fatter than you.'

In these examples, both the word order and the comparative marker itself are Chinese. Instead of the native pattern of quality-standard-noun illustrated by the example in (1), we have noun-marker-standard-quality exemplified by examples in (2a) to (2c). In fact, this kind of almost wholesale borrowing from Chinese is not confined to a limited number of grammatical structures in Tsat. In other words, it is quite pervasive throughout the grammatical system of Tsat.

## 6 Adverbs and conjunctions from Chinese

Adverbs, prepositions, and conjunctions are all borrowed from Mandarin.

Adverbs: (the examples given here are intensifiers)

- (39) *tʰ a.y<sup>33</sup> ?diai<sup>24</sup>. pʰ ay<sup>33</sup> sian<sup>21</sup> na.y<sup>32</sup>.* (Zheng 1997:76)  
 very hot extremely good  
*tài rè fēicháng hǎo*  
*tài rè fēicháng hǎo*  
 'very hot', 'extremely good'

- (40) *na.y<sup>32</sup> ket<sup>43</sup>. sa<sup>24</sup> ti<sup>55</sup>*  
 good extremely really white  
*hǎo jí zhēn bái*  
*hǎo jí zhēn bái*  
 'extremely good' 'truly white'

Correlative conjunctions:

- (41) *ziu<sup>33</sup> pa<sup>33</sup> ziu<sup>33</sup> ha:i<sup>33</sup>.* (Zheng 1997:84)  
 both hungry and tired  
*yòu è yòu lèi*  
*yòu è yòu lèi.*  
 'Both hungry and tired.'

Both the Mandarin and the Tsat have exactly the same structure with the key morphemes borrowed from Mandarin.

Clausal conjunctions:

- (42) *zi<sup>11</sup>ko<sup>11</sup> kʰ iy<sup>43</sup> tʰ a.y<sup>33</sup> ?dai<sup>24</sup>, kaw<sup>33</sup> saw<sup>43</sup> pu<sup>33</sup> na.u<sup>32</sup> lə<sup>33</sup>.* (Zheng 1997:85)  
 if tomorrow very hot, I then NEG go PERF  
*rúguǒ míngtiān tài ré, wǒ jiù bù qù le.*  
*rúguǒ míngtiān tài ré, wǒ jiù bù qù le.*  
 'If tomorrow is very hot, I won't go.'

What makes these examples particularly interesting is that not only are they borrowed but that for the most part their syntax in Tsat matches their syntax in Mandarin. That is, what has been borrowed is a construction still marked by its characteristic lexical item.

## 7 Other Han influenced constructions

Not all Mandarin influence has resulted in patterns of synchronic variation. However, even when this sort of syntactic variation is no longer found, it is still fairly obvious that contact has been at work. Whenever Tsat word order patterns differ from those of the Chamic languages of Vietnam, they are either identical with or close to the patterns found in Chinese. And, of course, the fact that often times grammatical morphemes are borrowed



together with the syntactic constructions, even serving to define the construction, is noteworthy.

Examples abound. In (43) below are three separate constructions showing Chinese syntactic influence. The first, indicated by the initial double underlining, is the extension of the prehead modification of the *sa*<sup>33</sup> construction to produce a prehead relative clause. This type of typologically marked prehead relative clause has developed under Chinese influence in at least three independent but parallel cases, once in Karen, once in Bai, and once in Tsat, has been extended, under the influence of Mandarin, to include other constructions quite new to Tsat.

- (43) ʔdi<sup>55</sup>nan<sup>33</sup> sa<sup>33</sup> mo<sup>33</sup> si<sup>11</sup> may<sup>33</sup> sa<sup>33</sup> (Zheng 1997:73)  
 lie.down that GEN cow be female GEN  
 tǎng nà de huángniú shì mǔ de  
 tǎngzhè de nà huángniú shì mǔ de.  
 'The yellow cow lying down is female.'

The second is the use of the Mandarin borrowing *si*<sup>11</sup> to mark the equative construction; the Chamic languages seem to use simple juxtaposition. The third is the use of a postposed *sa*<sup>33</sup> as a nominalisation in the *mai*<sup>33</sup> *sa*<sup>33</sup> 'female'. All three reflect Mandarin influence.

## 8 Other languages

None of this is restricted to Tsat, of course. Strikingly parallel developments are in progress throughout the Chinese dominated area of Asia. Everywhere where intense Chinese influence is manifested languages are undergoing major restructuring, resulting in word order variation and change. Contrary to the general belief that a very long period of time of persistent contact is needed for structural borrowing to occur, the wholesale restructuring is taking place quite rapidly. Intensity of contact more than duration seems to be the crucial factor. The level of bilingual proficiency and the instability of the social structure of the borrowing language seem to have a major impact on the length of time needed and the extent of the structural shift in contact induced changes.

### 8.1 Oroqen

For instance, in the southeastern and central dialect regions of Oroqen (a Tungusic language of the northeast; Li and Whaley 2000, Whaley, Grenoble, and Li 1999; Sun and Li 2001; Whaley and Li 1998, 2000), Chinese contact did not occur until after the settlement in the late 1950s and early 1960s. Nonetheless, within a short span of only a few decades, we already see signs of strong contact influence on the grammatical structure of the language. A case in point is the fact that one of our informants from the central dialect region used the adverb *mafaŋ*, which is a Chinese borrowing meaning 'immediately' spontaneously without noticing it. When fed the Oroqen word *diyaldzi*

meaning the same thing, he accepted it, but strongly prefers the Chinese borrowing. Interestingly enough, our informants from the western and northeastern dialect regions adamantly rejected the Chinese form insisting that it is not an Oroqen word.

Even in the northeastern dialect region where Oroqen is preserved the best, we saw strong Chinese influence, which is shown in the examples in (44).

- (44) a) *fī*        *tīmana*        *ŋənə-ni*        *yafa*  
          you        tomorrow        go-2SG.PRES        Q.2SG.PRES  
          'Are you going tomorrow or not?'  
       b) *yabufa*        *hafi*        *yafa*  
          walk.PAST        still-be        Q.PAST  
          'Went or not?'

In (44), we have two examples of the A-not-A question formation in Oroqen. Notice that the informant produced the Chinese *haishi* 'still be: or not' in (44b). She did so without realising it at all until it was pointed out to her. This informant feels at ease with both languages. In fact, she possesses native proficiency in both Chinese and Oroqen.

This kind of phenomenon suggests that when a speaker reaches a certain level of bilingual proficiency, borrowing between the languages occurs much more readily than is generally assumed in the literature. Thus, it does not take a very long time for a language to shift to a completely different typological pattern in its grammatical structures. Central to the rate of such structural shifts are sociolinguistic factors, particularly, the so-called intensity of contact. Our work suggests that for both Tsat and Oroqen a crucial factor has been schooling in Mandarin.

## 8.2 Mulam

Mulam (Zheng 1988), a Kam-Sui language of the Guangxi area related ultimately to Thai, is undergoing many of the same changes Tsat is: borrowing of conjunctions, adverbs, and prepositions along with the introduction of constructions with new word orders under the influence of Chinese. In fact, Mulam even has its own equivalent of the *sa<sup>33</sup>* construction, built on a different genitive marker but, like its Tsat counterpart, resembling the *de* construction of Mandarin.

Zheng Guoqiao (1988:173), in discussing the Mulam borrowing of Chinese adverbs and conjunctions, notes that 'degree and quantity adverbs are all borrowed from Han' and that 'borrowed adverbs generally are subject to the same syntactic rules as Han'. That is, what is borrowed is not just a lexical item but a lexical item along with accompanying syntax—in short, a construction.

Mulam, despite being a large geographical distance from Tsat, has calqued the *de* construction of Mandarin very much as Tsat has, and Mulam is borrowing many of the same constructions along the representative morpheme, leading to convergence with

Mandarin. Sometimes this has produced variation with the native patterns competing with the borrowed patterns; in other cases, the native pattern has been completely replaced.

The Mulam data would benefit from a broader historical perspective. It is not just Mulam, but the whole Kam-Sui branch of Tai-Kadai that shows such developments. A particularly striking example can be found in Kam (Dong). The northern dialects often have more evidence of Mandarin influence than the southern dialects. For a comparison that largely supplements this paper, see Edmondson (2001.) and Long and Zheng (1998).

## 9 Observations

In short, under intense Chinese contact, similar, rapid restructuring is occurring in geographically distant languages belonging to distinct language families: in Tsat, an Austronesian language of Hainan; in Oroqen, a Tungusic language in northeastern China; and in Mulam, a Tai-Kadai language found south of the Yangtze.

Although this is not our major focus, it is obvious that sociolinguistic factors rather than structural factors provide the impetus for the word order changes. Although some linguists consider structural similarity and functional congruence as the most important factors in cases of grammatical borrowing, (for example, Weinreich (1953)), Tsat and Oroqen seem instead to support Thomason and Kaufman's contention (1988:35) that 'it is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact. Purely linguistic considerations are relevant but strictly secondary overall.' In the case of Tsat, a major impetus to wholesale restructuring has been the fluency brought about by schooling in Mandarin.

Most central to the data presented are the paths of diachronic change. Grammatically, one configurational grammatical structure is being replaced by another, construction-by-construction, with the older Chamic word order being replaced by its Mandarin equivalent. In some cases, the Mandarin-influenced construction, often marked by a transparently Mandarin grammatical morpheme, is simply a marked alternative, as with comparatives, but in other constructions the word order of the native Chamic construction has been completely superseded, as with the genitives. The word order changes have entered the language as borrowed constructions, marked by a characteristic often-borrowed grammatical morpheme. Still further influence is manifested in the overgeneralisation of calques, cf. the spread of the *sa*<sup>33</sup> construction in Tsat.

The complex set of conditions responsible for the restructuring are only partly explainable by internally and externally motivated principles proposed in the literature. A profound understanding of the situation must take account of the dynamic changes that take place in not only linguistic structures but in the social conditions as well. There is a range of social factors that contribute to borrowing and structural shifts. In the case of Tsat, encroaching bilingualism with a powerful dominant language (along with schooling

and social mobility), are among the most prominent factors that lead to the massive borrowing and drastic structural shift.

The rapid changes taking place in these languages make it imperative that the nature of the speech community be specified in far more detail than it is now and as quickly as possible. We can already see extensive language change in these particular languages whose speaker community is constituted by bi- or multi-lingual linguistic and ethnic minorities but the details need to be documented and the paths of change need to be examined. A lot more work needs to be done and done quickly before the languages in question cease to exist.

Finally, a comment on the obvious: It is the construction, rather than just its characterising grammatical morpheme, that is the typical unit of borrowing.

## References

- Benedict, Paul K., 1941, A Cham colony on the island of Hainan. *Harvard Journal of Asiatic Studies* 4:129–134.
- Dryer, Matthew S., 1992, The Greenbergian word order correlations. *Language* 68/1:81–138.
- Edmondson, Jerold A., 2001, Typology and language contact phenomena in Southeast Asian perspective. Paper presented at the southeast asian linguistic society conference, Bangkok.
- Lee, Ernest Wilson, 1966, Proto-Chamic phonologic word and vocabulary. PhD dissertation. Indiana University.
- Li, Charles N. and Sandra A. Thompson, 1981, *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- Li Fengxiang, 1996, A report on the study of the consonant systems of two Manchu-Tungusic languages in China: Evenki and Oroqen. Paper presented at the annual meeting of the Linguistic Society of America, San Diego.
- Li Fengxiang and Lindsay J., Whaley, 2000, Emphatic reduplication in Oroqen and its Altaic context. *Linguistics* 38/2:355–372.
- Long Yaohong and Zheng Guoqiao, 1998, *The Dong language in Guizhou province, China*. Translated by D. Norman Geary. Arlington TX: Summer Institute of Linguistics and University of Texas at Arlington.
- Ouyang Jueya and Zheng Yiqing, 1983, Hainan Yaxian Xian Huizu de Huihuihua [The Huihui language of the Hui nationality of Yaxian County, Hainan]. *Minzu Yuwen* 1:30–40.

- Sun Hongkai and Li Fengxiang, 2001, Contact, attrition, and structural shift: the case of Anong and Qiang. Paper presented at the Tibeto-Burman workshop at the University of California, Santa Barbara, July 27–29, 2001.
- Thomason, Sarah Grey and Terrence Kaufman, 1988, *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thurgood, Graham, 1996, Language contact and the directionality of internal 'drift': the development of tones and registers in Chamic. *Language* 71/1:1–31.
- 1999, *From ancient Cham to modern dialects: two thousand years of language contact and change*. Oceanic Linguistics Special Publication No.28. Honolulu: University of Hawai'i Press.
- Weinreich, Uriel, 1953, *Languages in contact*. The Hague: Mouton.
- Whaley, Lindsay J., Lenore A. Grenoble, and Fengxiang Li, 1999, Revisiting Tungusic classification from the bottom up: A comparison of Evenki and Oroqen. *Language* 75/2:286–321.
- Whaley, Lindsay J. and Fengxiang Li, 1998, The suffix *-kan* in Oroqen. *Studies in Language* 22:181–205.
- 2000, Oroqen dialects. *Central Asiatic Journal* 44/1:105–130.
- Zheng Guoqiao, 1988, The influences of Han on the Mulam language. In Jerold A. Edmondson and David B. Solnit eds *Comparative Kadai: linguistic studies beyond Tai*, 167–177. Arlington: Department of Linguistics, University of Texas at Arlington.
- Zheng Yiqing, 1986, Cai tan Huihuihua de diwei wenti [A further discussion of the position of Huihui speech and its genetic relationship]. *Minzu Yuwen* 6:37–43.
- 1997, *Huihui yu yanjiu* [A study of Tsat]. Shanghai Yuandong Press.

# 13 *Comparative notes on verb compounding in English and East Asian languages*<sup>1</sup>

---

BENJI WALD

## 1 Foreword

While the East Asian (EA) speech area is well known for a variety of verb compounding and associated grammatical processes, English verb compounding is of such recent emergence that its evolution and current status have only begun to be explored (as in Wald and Besserman 2002). For English, verb compounding refers to the word formation process by which a syntactic verb is formed by the fusion of two constituent verb roots into a single verb stem, allowing it to be the nucleus of a single clause, as in *drop-kick*, *freeze-dry*, *sleep-walk*, *spell-check*, *stir-fry*, *strip-search*, among many other examples. Essentially, the differing constraints on verb compounding in the individual EA languages is a reflection of how the process interacts with other grammatical processes in the individual languages. Nevertheless, it is possible to extract a number of recurrent

---

<sup>1</sup> It is my pleasure to acknowledge the debt I owe to Jim Matisoff for giving me my initial appreciation of the Sino-Tibetan languages and the larger speech area in which they function, enabling me to explore them further with my present concerns. Beyond that, I want to take this opportunity to express my appreciation of his vast linguistic knowledge, enthusiasm and interest in all languages, and above all, his unwavering friendship. With regard to this paper, I also want to express my gratitude to other contributors to this celebratory volume, and especially to Randy LaPolla for discussing with me many points which had puzzled me with respect to specific Mandarin verb compounds. There are clearly still some shortcomings in my discussion, but there would have been many more if not for his generous and patient help. Lastly, I am thankful to Masatomo Ukaji for replying on a few points concerning Japanese verb compounding, as indicated at the appropriate point in the text.

dimensions along which verb compounding varies across these languages, and these dimensions also have a relevance to English verb compounding. The following discussion explores some of these dimensions for points of comparison among various EA languages and English verb compounding.

## 2 Productivity

How freely do two verbs combine to form a verb compound in English and various EA languages? In approaching an answer to this question I have chosen to generalise from attested examples of English verb compounds rather than try to intuit a distinction between unattested possible and impossible verb compounds. At the current stage of development of English verb compounding, I do not think that constraints on its productivity are intuitively obvious. Therefore, my approach imposes a control on a possible (unconscious) tendency to over or underestimate the variety of contexts in which verbs may be compounded in English. A similar approach to EA verb compounding seems to distinguish between syntactically productive and lexicalised verb compounds. In general, the most commonly discussed syntactically productive type is labeled resultative, of which there are several sub-types, and the directly lexicalised type can be labeled coordinate, of which there are also several subtypes.

Only Korean, areally transitional between the core EA verb-medial area, dominated by Chinese, and the eastern peripheral EA verb-final area, represented by Japanese, seems to distinguish these two major types of compounding by surface construction (cf. Sohn 1999:254). The Korean "resultative" type is more generally a serialising construction represented by V1-*e/a*V2, in which V1 and V2 are ordered by temporal iconicity, for example, *tul-e ollita* (take-*e/a* lift < ascend.CAUS) 'hold up', and the coordinate type is represented by V1-V2, e.g., *olu-naylita* (ascend-descend) 'go up and down', generally expressing iterativity or simultaneity of events. In Japanese, both types are expressed by a single surface construction, V-i-V, formally most similar to the Korean serialiser, e.g., Japanese *kam-i-kiru* (bite-i-cut) 'bite and thus cut/cut by biting' and coordinate *nak-i-wameku* (cry-i-shout) 'cry and shout' (Shibatani 1990:246). In the core area, represented by Mandarin, both types are expressed by a single surface construction most similar to the Korean coordinate, i.e., V1-V2, e.g., resultative *tuī-dào* (push-fall.over) 'push down/over' and coordinate *tān-chàng* (play-sing) 'sing and play (an instrument) at the same time' (examples from deFrancis 1996). For well understood reasons involving avoidance of rampant homonymy, Chinese, and Mandarin in particular, has productively developed the coordinate subtype of two quasi-synonymous verbs, e.g., *wán-bì* (finish/use.up-finish/accomplish) 'finish/complete' vs. *wán-nòng* (play/trifle.[with]-play.with) 'play/juggle with'. Elsewhere in the core area, quasi-synonymous compounding is generally described as a literary artifice, influenced by written Mandarin, and the

productivity of the more general coordinate type is problematic. English verb compounding formally resembles the core East Asian type of V1-V2, but we will see that by its nature it does not engage in coordinate compounding (see discussion connected with examples 11 and 12 in §7 below). Its productivity resembles the resultative type, but occurs largely in distinct semantic contexts in which V1 most commonly resembles in function an adverbial qualifier of a head V2.

With regard to the relative productivity of EA compounds, Matisoff's comments on Lahu verb compounding are instructive:

Neither of the elements in a true compound are juxta-productive; each occurs in at most a few compound-combinations with verbs of compatible semantic nature. Compounds once established acquire the status of unitary lexical items. It is as difficult to invent a comprehensible and acceptable Lahu compound as it is to create any neologism. Binary versatile concatenations, on the other hand, are freely 'inventable'. (Matisoff 1973:209)

In context he is contrasting as true compounds, the apparent coordinate type of Lahu *nù-quâ* (stink-be.bitter) 'be acrid-smelling' with various syntactically produced types where either V1 or V2 may be freely chosen as the head of the construction, and the other constituent is chosen from a relatively limited number of verbs, labeled versatile, that can modify the head in such a construction, for example, *qô ša* (hoe easy) 'easy to hoe', where V2 represents the limited set, *tā qô* (begin hoe) 'begin to hoe', where V1 represents another limited set; thus, *tā ša* (begin easy) 'easy to begin' or 'begin to be easy', depending on which constituent verb is taken as the head (op. cit. 201).

When English VVs were highly limited in number they conformed to Matisoff's concept of (true) lexical compounds. However, many now consist of impressively juxta-productive verbs. To give one example, *kick*, as V1, is attested with the V2s *-block, -box, -break, -chop, -flip, -jump, -punch, -push, -save, -serve, -start, -stop, -stretch, -stroke, -turn*. As V2, it is attested with the V1s *block-, bounce-, chip-, chop-, crash-, dive-, flick-, flip-, flutter-, fly-, hitch-, hook-, jump-, punch-, return-, skip-, slap-, slice-, slide-, snap-, spin-, splash-, stab-, sweep-, tap-, touch-, trip-*. In such cases, neither of the constituent verbs appears to be a 'versatile', since their meanings are not specialised in context and they do not seem to form a closed set. However, in some cases, there are semantic parallels between productive English V1s and Lahu (and more general core EA) pre-head versatile, for example, English *sneak-* attachable to *-drink, -eat, -smoke*, etc., and Lahu *qhô* (lit. 'steal'), e.g., *qhô-na* (steal-listen) 'eavesdrop' (Mandarin *qiè-ting<sup>1</sup>* or *tou<sup>1</sup>-ting<sup>1</sup>*). While Lahu verb compounds may be either head-initial or -final, and various Mandarin verb compounds are analogous but controversial for headedness, we'll see that English verb compounds, like Japanese, are invariably head-final.

With regard to the criterion of comprehensibility, this seems to be a matter of context in discourse, as discussed further in §3 below. Acceptability is a more difficult matter to assess, as we will see in (15) below, and is no doubt subject to individual variation. The



issue of productivity will continue to be of concern throughout this paper. For the moment, note that I only knew a few of the above and other attested English VVs in advance of research, for example, *kick-start*. In essence, I coined the others and then did a web-site search to check their prior coinage. In a period of two months, starting in January 2000, I attested well over a thousand verb compounds. Thus, the question arises whether we are dealing with a new English syntactic pattern involving verb fusion. If so, is it competing with or replacing any prior English syntactic pattern performing a similar function? And what is that function?

### 3 Heterogeneity of sources

To what extent is verb compounding a homogeneous process in English? I have already noted distinct sources for EA verb compounding in the preceding discussion, with respect to coordinate compounds, which appear to be directly coined and lexicalised, and resultative compounds, which are syntactically derived but may be subject in particular cases to subsequent lexicalisation. One aspect of lexicalisation in core EA languages is the phenomenon of bound verbs, that is, verbs which enter into compounds, but do not occur as independent verbs, and whose meaning may even be obscure to current speakers. Matisoff (1973:198) gives a Lahu example in his initial discussion of verb compounds, *qa-mi* ('sing, pass air noisily in or out of mouth', cf. English *snore, pant*) 'sing'. He observes that V1 *qa* is currently used alone only poetically (archaism), and that V2 *mi* is only used in compounds. He infers a free origin for both verbs, noting that such fusions are a historically recurrent phenomenon among Sino-Tibetan languages. This phenomenon is indicative of the length of time in which verb compounding processes have been active in core EA languages. The phenomenon does not occur among English verb compounds, no doubt due to their recency; comparable phenomena are only found in English noun compounding, unproblematically a productive word formation process predating English verb compounding (and even English) by many millennia, e.g., *-hood* as in *neighbourhood*, and *-dom* as in *kingdom*.

The notion that English verb compounding is, or is becoming, a unified grammatical process is supported by the transparency with which the constituent verbs, as independent entities, contribute to the meaning of the compound as a whole. There are, however, some problems to be acknowledged. The difference between (1) and (2) below illustrates the issue.

- (1) ... Another measure would have provided \$10 million in bonds for a plant to freeze-dry coffee.  
 starbulletin.com/2001/05/15/business/ story1.html<sup>2</sup>

<sup>2</sup> For a website source of example, punctuation is always left as is; the search-word is given in bold type.

The compound *freeze-dry* is typical of a great many attested verb compounds. One need not be familiar with the particulars of the machinery involved in the process of freeze-drying coffee in order to recognise the contribution of the constituent verbs to the meaning of the compound verb.

The following example is more problematic.

- (2) ... At a recent conference in Austria, he **hang-glided** through the Grand Canyon without ever leaving the convention hall ...

[www.dr tomorrow.com/profiles.html](http://www.dr tomorrow.com/profiles.html)

In principle, the semantic contribution of the constituent verbs of *hang-glide* is no less transparent than in *freeze-dry*. However, understanding of the contextual meaning is enhanced by knowledge of the instrument *hang-glider*. Thus, the issue of heterogeneity of verb compounding in English emerges in whether the verb *hang-glide* should be recognised as a backformation from the noun *hang-glider*. If so, its internal structure may be irrelevant to interpretation, and *hang-glide* can be distinguished from various other apparent verb compounds such as *drop-kick*, which may be a conversion of the noun *drop-kick* (cf. Marchand 1969:58ff). For that matter, English VV compounds are often alternatively analysable as NV compounds, for example, *sleep-walk* as 'walk while sleeping' (VV, where the gerund paraphrase of V1 implies a verb, since English gerunds have verb properties that nouns lack, for example, adverbial rather than adjectival qualification), or as 'walk in sleep' (NV). My approach to this problem is to recognise that for individual lexical items different speakers may have different internal analyses, if any at all (cf. Wald and Besserman 2002). In general, the analysis matters, because a VV analysis reinforces the productivity of the VV pattern. In view of speaker variability in analysis, it is more precise to speak of individual examples susceptible to alternative analyses as possible (rather than indisputable) VVs, with the expectation that some, probably many, speakers interpret such compounds on first hearing them (and often thereafter) on the basis of their constituent verbs. To this extent, verb compounding can be considered a unified process in English. Subsequent discussion will provide further examples of English verb compounds whose sources may be historically diverse, but which can be unified by a general grammatical characterisation of English verb compounding discussed in §7 below.

#### 4 Degree of Fusion

English verb compounding is characterised by the fusion of the constituent verb roots so that nothing may intervene between them. The fusion is a consequence of the historical evolution of the pattern of English verb compounding from other forms of compounding, most notably nominal. EA languages differ in the degree of fusion exhibited by compound verbs. All EA languages seem to have some examples of inviolable fusion, but they vary

in the contexts and the extent to which this pattern is productive. In the core area, the productive resultative pattern is not fused. A limited number of additional elements may intervene between the constituent verbs, most notably a negative marker. Thus, Matisoff (1973:266) notes that Lahu resultatives have two meaningfully contrasting negatives; for example, *γò t̂ʔ* (pull come.out) 'pull out' (or 'extract by pulling') has the external negative *m̂ γò t̂ʔ* 'not (even try to) pull out' and the internal negative *γò m̂ t̂ʔ* (pull NEG come.out) 'pull but fail/not succeed in extracting'. Similar phenomena are noted for the same construction in various other EA languages, e.g., Mandarin (Li and Thompson 1981:427), Vietnamese (Nguyen 1979:178). Interestingly, Gorgoniyev (1966:76) states that internal negation is not common in Khmer resultatives, as if to say external negation may serve the same purpose, but notes that the verb *haj* 'cause' optionally but commonly intervenes in cause resultatives, e.g., *phat (haj) lyyam* (clean cause shine) 'polish'.

Alternative forms of syntactic bonding, a precondition for fusion, are also noted for some resultative compounds consisting of more than two constituent verbs. Thus, for example, Li and Thompson (1981:64) note that the notional object of the Mandarin verb *duān* 'serve (e.g., a bowl of soup)' may intervene in the directional (resultative) compound *duān-sháng-lái* (serve-ascend/up-come) 'serve up (toward speaker)' either between *duān* and *sháng-lái* or *duān-sháng* and *lái*. Similarly, with respect to alternative bonding, the complex *zhì-zào-chéng* (fabricate-create-transform) 'manufacture (e.g., goods from spare parts)' blends the simpler attested compounds of *zhì-zào* 'manufacture' and *zào-chéng* 'create'.

Historical shifts in bonding of verb constituents are evident in the difference between Mandarin *huí lái le* (return come Perfective) vs Cantonese *fa:an-jó làih* (return-Perfective come) 'come back' (Matthews and Yip 1994:46), where bonding seems more advanced in Mandarin than in Cantonese. Similar shifts in bonding are evident in negative placement in Burmese, where NV compounds, for example, *hnou' hse'* (mouth-join) 'greet', are usually negated immediately before the verb, for example, *hnou' mahse'*, but occasionally maintain the integrity of compound, e.g. *mahnou'-hse'* (Okell 1969:40-41).

Japanese and Korean are more similar to English in the fusion of verb compounds. However, there is one difference which reveals an important point about the semantic nature of English verb compounding. As noted above, the Japanese verb compound has an "infinitive" suffix *-i-* intervening between V1 and V2. This element is productively associated with adverbialisation of verbs in subordinate clauses, for example, *wara-i]-nagara mukaeru* (smile-*i*]-while greet) 'greet while smiling' (Shibatani 1990:313), cf. Classical Tibetan verb clips, for example, *n:u-bod* (< *n:u-ba-r bod*; weep-nominaliser-adverbialiser exclaim) 'exclaim while weeping < weepingly exclaim' (Beyer 1992:95). Thus, Japanese *-i* effectively subordinates V1 to V2 in compounds. English similarly subordinates V1 to V2 in compounding, with syntactic consequences, discussed further below. However, English verb compounding strictly forbids the V1 root to be overtly marked in any way. Thus, in contrast to other nouns, the English gerund, as an activity

nominal with verb syntactic properties, for example, adverbial rather than adjectival modification (cf. 'wrongly *spelling* the word', gerund, versus 'wrong *spelling* of the word', noun), does not occur in the formation of English compound verbs, and, indeed, would not contribute anything to the meaning of the compound other than to explicitly mark V1 as subordinate to V2 (as in Japanese). The restriction supports the productivity of English verb compounding at the expense of the NV pattern. Consider (3) below.

- (3) ... After you have typed your Tibetan text, it can be **spelling-checked** using the normal WordPerfect methods.  
[www.tibet.dk/tcc/Tibetan4b.htm](http://www.tibet.dk/tcc/Tibetan4b.htm)

Here *spelling* is not the gerund, but a product nominal of the root verb meaning of *spell*. Significantly, this particular NV has largely given way to the VV *spell-check*, in which the activity expressed by V1 is only related to the product nominal *spelling* by pragmatic inference.

Similarly, passivised V1, that is, **\*\*V1-en-V2**, is not allowed in English verb compounding. Thus, for example, the compound adjective *spun-dyed*, used in textile manufacture, has a corresponding compound verb *spin-dye*, not **\*\*spun-dye**. Voice is a property of the compound as a whole, or of V2 as the head. V1 cannot independently undergo any grammatical process.

## 5 Order of constituents

It is already evident from the preceding discussion that order of constituents plays a crucial role in the semantic interpretation of English verb compounds. The role of V2 as the head of the compound is confirmed by the argument structure of English compounds when the constituent verbs differ in their argument structure. The argument structure of the compound is invariably the same as the argument structure of V2. Thus, consider:

- (4) ... She got up and **jump kicked** Drusilla ...  
[www.slayerfanfic.com/R/RickyGarcia/wouldyou.html](http://www.slayerfanfic.com/R/RickyGarcia/wouldyou.html)

In context, the argument structure of transitive *jump-kick* is that of transitive V2 *kick*, but not of intransitive V1 *jump*. When the order of constituents is reversed, as in (5) below, the compound becomes intransitive in accordance with V2 *jump*:

- (5) ... Justice **kick-jumped** off the wall and fired his cutting beam from the air ...  
[www.projectmetaverse.org/stories/Winger/tempest/tempest2.html](http://www.projectmetaverse.org/stories/Winger/tempest/tempest2.html)

Currently, very few attested English verb compounds provide reverse pairs to test the headedness properties of V2, even though most verbs attested as V1 in some compounds are also attested as V2 in other compounds. In the broadest sense, this situation is similar to Mandarin, in which most verbs figure as V1 in some compounds and V2 in others,

though rarely with the same companion verbs, as noted further in §6 below. There are, however, a few such verbs in English, of which *test* is a highly productive example, with minimal, if any, effect on the meaning of the compound according to its position. Thus, consider the pair,

- (6) ... Don't write the rules down until you've **play tested** your game and ironed out the details ...  
[www.discovergames.com/skate5.html](http://www.discovergames.com/skate5.html)

- (7) ... the guitar is strung up, tuned and **play-tested** ...  
[www.tejagerken.com/Article\\_Folder/factory\\_tour.html](http://www.tejagerken.com/Article_Folder/factory_tour.html)

When the companion verb is transitive, as in the above examples, it shares its (logical) object with *test*. However, when the other verb is intransitive, V2 (whichever it is) determines the argument structure of the compound, as in:

- (8) ... After we've **sleep tested** the product, we find that it does not live up to our expectations ...  
[www.specialtybed.com/magazine/1998/fall/feature2.php3](http://www.specialtybed.com/magazine/1998/fall/feature2.php3) (contrast **test-sleep** > **-slept** the product')

- (9) ... Well I **test-listened** to these speakers couple of times before I bought them ...  
[www.audioreview.com/reviews/Spaker\\_product\\_6806.shtml](http://www.audioreview.com/reviews/Spaker_product_6806.shtml) (contrast **listen-tested** to these speakers')

There are a few verbs with the order versatility of *test*, for example, *check*, as in attested (police) *check-stop/stop-check* (vehicles). More commonly only one order is attested for any potential pair of compound verbs, so that even *\*test-sleep* and *\*listen-test* are not (yet) attested as verbs. A second language learner who said *\*dry-blow* or *\*fry-stir* would be readily understood but corrected.

## 6 Lexicalisation

So far only two temporal orderings of English verb compounds have been exemplified, one in which the referenced activities are temporally unordered, for example, *stir-fry*, *play-test*, *test-play*, *hang-glide*, and another in which they are iconically ordered, for example, *jump-kick*, *freeze-dry*, *spell-check*. These are indeed the most commonly observed possibilities, just as in EA languages. However, counter-iconic order is also attested, for example,

- (10) ... The remaining trees are then grown on without further thinning until age 25–35 years when they can be **clear-felled** as the final timber crop ...  
[www.dpi.qld.gov.au/hardwoods/d/1815.html](http://www.dpi.qld.gov.au/hardwoods/d/1815.html)

Here V1 *clear* (the forest) is the purpose of V2 *fell* (the trees). A resultative order would be unattested *\*fell-clear*. Note the different argument structures of V1 and V2, i.e., *clear* the forest of/\*\*from trees vs. *clear* trees from/\*\*of the forest. Another example is *drop-ship* (a package, etc.), where the package is *shipped* (V2) and then *dropped* (V1) [off] at its destination (as opposed to *drop-kick* or *drop-feed*, where the order is iconic).

I will leave as problematic whether temporal order should be lexically specified for English verb compounds, or considered a matter of pragmatic interpretation. Decisive in favour of pragmatic interpretation would be a single item which may receive either an iconic or counter-iconic interpretation according to context, but I have not found such examples. Nevertheless, it is clear that English compounding is not constrained by temporal order. As discussed further in §7 below, it seems likely to me that headedness alone is the determinant of order. Thus, *\*dry-blow* and *\*fry-stir*, for example, are corrected because the corrector perceives that they have been mistakenly headed, demoting the “more essential” (head) elements *dry* and *fry*.

It is worth noting that the problem of ordering to lexicalisation would be somewhat different for coordinate compounds. As we will see, English does not have coordinate verb compounds. If it did, the issue of headedness would be controversial, as it has been for Mandarin (cf. Packard 2000, 2001; Chung 2001). Li and Thompson (1981:54) consider the parallel (quasi-synonymous) compound to be as productive as the resultative compound. However, their only hint to the ordering of such compounds is that the simple verb *mǎi* alternates with *gòu-mǎi* (buy-buy) ‘buy’ according to the larger prosodic context, not a lexical matter. The few equivalent English quasi-synonymous coordinate expressions, largely peculiar to legal contexts, are not compounds, but are rhythmically ordered as monosyllabic verb-polysyllabic verb, for example, *aid and abet*, *cease and desist*, *keep and maintain* (cf. Mellinkoff 1963:121). A possible colloquial synonymous expression is *pick and choose*. However, the polysemy of *pick* with respect to picking up an object before choosing to buy it, as in fruit and vegetable shopping, allows an iconic temporal motivation for the order, cf. Mandarin *tiāo xuǎn* (pick/carry select) ‘pick/choose’. There are a few apparently reversible compounds in Mandarin, example, *jì-suàn* (count-calculate): *suàn-jì* ‘count/calculate’, *dòu-zhèng* (fight-struggle): *zhèng-dòu* ‘struggle/fight’, *wèn xún* (ask-ask) *xún-wèn* ‘ask/inquire about’. But in most cases reversal leads to meaningfully distinct compounds from distinct grammatical sources, for example, *qǐ-diào* (rise-hang) ‘lift (with a crane)’ versus *diào-qǐ* ‘hoist (with a rope)’, where the second is analysed as a directional resultative (DeFrancis 1996). The first seems to be indeed lexicalised, if the implied instrument is limited to a crane or similar device. Similarly, *zá-róu* (be mixed.up- knead) ‘blend’ but *róu-zá* ‘be jumbled together’ seem to imply lexicalisation, in the absence of an interpretation by headedness. Giving pause are pairs like *jī-pò* (attack-break) ‘destroy/defeat’ and *pò-jī* ‘sabotage/attack and destroy’. The first seems similar to a resultative, cf. *dǎ-pò* (hit-break) ‘break/smash’, while the second closely resembles the first in its more general sense. It might be coordinate, with

indifference to temporal order, but could not be resultative due to the reversed order. Meanwhile, the specialised sense 'sabotage' of *pò-jī* suggests lexicalisation of that order in that sense.

In contrast to Mandarin, reversibility, or indifference to order, is an explicitly recognised feature of a number of Vietnamese quasi-synonymous coordinate compounds, for example, *kính-trong* (admire-respect) or *trong-kính* 'admire/respect' (Thompson 1965:132, Nguyen 1979:xviii). Nguyen adds that for rhetorical effect the linking particle *với* 'and' can optionally intervene between the two constituents of a coordinate compound, for example, *buôn (với) bán* [buy.wholesale (and) sell] 'engage in commerce'. The Mandarin equivalent *mǎi-mài* (buy-sell) is strictly ordered. Thus, it is evident that many Vietnamese coordinate compounds are not fused, and are associated with their syntactic source. The Mandarin coordinate compound is well documented to have evolved from an explicitly coordinate expression, i.e., *V1-ér-V2* (Norman 1988:121), but in its current state of development it seems to establish nuanced meaning differences in internal order, eventually lexicalising coordinate compounds and removing them from their productive syntactic sources.

## 7 Semantic context

The issue of semantic context has been implicit in much of the preceding discussion. A useful approximation of the semantic structure of English verb compounding is the paraphrase *V2 PREP V1-ing*, where *V1-ing* is a gerund, not a more specialised *V-ing* nominal (such as the concrete product nominals *spelling*, *painting* or *building*). The paraphrase captures the semantic headedness of *V2*. *PREP* is key to the generality, and productivity, of English verb compounding. It represents a specific semantic relationship between *V1* and *V2* beyond head and subordinate (modifier, qualifier, specifier). It must be abstract because, as seen above, different semantic relationships obtain, depending on the sense of the particular compound. For example, *freeze-dry* 'dry (X) by freezing (X)', cf. *play-test*, but *clear-fell* 'fell (X) for clearing (X/Y)', cf. *test-play*. Sometimes more than one *PREP* paraphrase is arguably appropriate, for example, *strip-search*, 'search (X) by/while stripping (X)', cf. *stir-fry*. Paraphrase, of course, facilitates cross-language comparison for semantic context without necessarily ascribing the same syntactic (headedness) analysis to the compared language, e.g., Mandarin *shì-shè* (test-shoot) 'test-fire (a weapon)', cf. *lì-shè* (stand-shoot) 'fire (a weapon) while standing' (cf. Wu 1999; deFrancis lists these examples only as nominals). Meanwhile, the greatest virtue of paraphrase is the language-specific comparison of English verb compounding with alternative grammatical strategies in the same semantic context, as discussed further below. This is of particular interest to the future productivity of English verb compounding. Is it competing with any other grammatical strategies toward the same semantic ends? If so, what are its chances for success in a particular semantic context?

Effectively excluded from verb compounding by its semantic characterisation are coordinate compounds. There are a few apparent examples of English coordinate compound verbs, for example, *drink-drive*, *slash-burn*, *hit-run*. They are all ordered by temporal iconicity, V1 *before* V2, and thus are interpretable as V2 *after* V1-ing, e.g., *drink-drive* as 'drive *after* drinking'. However, they are transparently related to coordinate expressions, the latter usually modifying a noun, for example, '*slash-(and)-burn* agriculture', '*hit-(and)-run* accident'. The following example of *hit-run* demonstrates the effect of converting the coordinate expression into a compound verb:

- (11) ... I remember standing there, waiting for them to **hit run** us over, ...  
[www.lostdream.com/users/dmulhern/dreams/dmulhern\\_004.htm](http://www.lostdream.com/users/dmulhern/dreams/dmulhern_004.htm)

In the coordinate expression V2 *run* is intransitive. But in the compound verb it is transitivised in order to take the object properly belonging to V1 *hit*. The resulting V2 is a transitive phrasal verb *run over*. The semantics of *run* has been changed to conform to the requirement that the object of the compound verb must also be the object of V2. Consequently, only in reference to the expression *hit-and-run* is it possible to interpret the compound as meaning '*run* us over and then *run* (away)'. This parasitic relation between the meaning of the compound and the meaning of another expression removes it from the semantic structure and productive use of English verb compounding. Such verbs are best viewed as direct conversions from the attributive compounds upon which their meanings depend.

Apparent coordinates are sometimes revealed in other ways as not the product of verb compounding, for example:

- (12) ... if you drank drove you'd be picked up by the cops ...  
[www.kuro5hin.org/story/2001/2/7/21155/69149](http://www.kuro5hin.org/story/2001/2/7/21155/69149)

The inflection on V1 *drink* removes it from verb compounding; *\*drink-drove* is not attested. On the other hand, *slash-burned* is attested. Thus, unlike *hit-run* and *drink-drive*, it has been remodelled according to verb compounding, and is not a coordinate compound in that context despite its transparent historical origin, cf. *drop-kick* (kick after dropping), etc.

A very productive semantic context for EA verb compounding is reflected in directional compounds, for example, Mandarin *tiào-guò* (jump-cross), Japanese *tobi-kosu* (jump-surpass) 'jump over'. As translations of this semantic type invariably show, the English equivalent is a phrasal verb, not a compound verb. The English semantic pattern allows *jump-cross*, and other verbs equivalent to EA directionals, for example, *rise*, *lower*. However, in contrast to EA languages, English has and prefers adverbials like *up*, *down*, *over*, etc. Here the least that can be suggested is that the English adverbials and their use in forming phrasal verbs has a historical lead in development of almost a millennium over verb compounding in the directional context. There is little discernible motivation to replace them with verb compounds. In addition, as the head position, V2 of VV is



equivalent to the V(1) of directional phrasal verbs in its freedom of selection and determination of argument structure. Thus, when *cross* figures in compounding, as it frequently does, it patterns with the adverbials in assuming the status of V1 rather than V2, for example, attested *cross-jump* (alongside more common examples in which V2 is transitive, for example, *cross-cut*), cf. *down-load*, *back-pedal*. In sum, verb compounding is an under-utilised resource in directional contexts and likely to remain so. That is, there is little foreseeable probability that *\*jump-rise* (or even *\*rise-jump*) will emerge to compete with *jump up*.

A similar situation of under-utilisation occurs in quasi-aspectual contexts, for basically the same reasons, again in contrast to EA verb compounding. Thus, only one arguable case of verb compounding is so far attested in an aspectual context:

- (13) ... Click OK to **run start** the installation. Power Word requires minimum of 100 MB of free hard disk space ...  
home.freeuk.com/zian/

Paraphrase as *start* PREP (e.g. ?by) *running* is dubious. The sense is simply 'start running (the program)'. In addition, *run* is easily construable as the nominal object of V2 *start*, i.e., 'start a run (of the program)'. OV (Object-Verb) is a type of NV compound, not a VV. A compound like *strike-break* is more readily recognisable as an OV, not a VV. I have not found examples in which the latter takes an object (or is used passively), although V2 *break* is transitive. The manifest reason is that *strike* is already the object. Returning to the less obvious analysis of *run-start*, there are many examples of *-start* as V2, e.g. following *jump-*, *kick-*, *push-* etc. In all these cases, manner/means of 'starting' is the contribution of V1, facilitating use of *by* as the appropriate paraphrase PREP, e.g. *kick-start* as 'start by kicking', and all are ways of 'run-starting'.

Note that in aspectual contexts the core EA languages tend to iconically order aspectually related verbs, so that 'begin' figures as V1 with respect to its complement, while 'finish/complete' figures as V2, for example, Lahu *tā qô* (begin-hoe) 'start to hoe' but *và? pà* (put.on-finish) 'have already put on'. Similarly, Mandarin features a number of V1 *kāi* 'open > begin' compounds, for example, *kāi-dòng* (open = begin-move) 'set in motion', but V2-*wán* 'use up > finish/complete V2-ing' as the general completive marker, for example, *nòng-wán* (do-finish) 'finish doing'. The same is true of Thai and Vietnamese (cf. Noss 1964:118,128; Thompson 1965:209). In contrast, in Japanese and Korean, as in English, 'begin' and 'finish' are organised as members of the same auxiliary set (cf. Takahashi and Takahashi 1984, Lee 2000). In English the immense historical priority of aspectual auxiliaries preceding their complements, and their general integration into a larger set of syntactically similar auxiliaries and modals, leaves little prospect for verb compounding to become active in this semantic context.

The situation is different with English cause resultatives. English has a number of compound verbs that can be construed as resultative; *freeze-dry* has already been

mentioned several times. The alternative resultative construction is *V1 ADJ2*, for example, 'freeze (something) dry'. cf. *rub raw, slap silly, tickle pink*, etc. In the core EA area verb compounding is not distinct from the *V1 ADJ2* pattern, but English sometimes overtly derives verbs from adjectives, as in the following of example, where the ADJ *flat* must be converted to the verb *flatten* according to the pattern:

- (14) ... The first operation *roll forms* the metal to produce the five thicknesses or folds.  
The second operation **roll flattens** these to produce the tight seam ...  
[www.all-pak.com/t\\_glossary\\_metal.asp](http://www.all-pak.com/t_glossary_metal.asp)

So far, English verb compounding is relatively limited in resultative contexts, but it is difficult to predict what the future may have in store, cf.

- (15) ... Hagan's eyes **open widened** at the sight of the huge creature ...  
[www.annexia-rpg.com/PBEM/past/story.cfm?Story\\_ID=7](http://www.annexia-rpg.com/PBEM/past/story.cfm?Story_ID=7)

Many readers may reject the example on aesthetic grounds, but the writer shows an impeccable grasp of the semantic structure of verb compounding.

In sum, the semantic characterisation of English verb compounding facilitates exploration of alternative English grammatical strategies used in the same semantic contexts. English verb compounding seems to be a currently underused grammatical resource. It is particularly underused where other syntactic resources are highly organised and actively used, for example, the system of phrasal verbs and prepositional adverbs which control expression of path and direction, and the (auxiliary) verb-complement system which, among other things, controls quasi-aspectual expressions, like 'start/begin' and 'finish/end'. In resultative contexts verb compounding may be gaining ground. The major use of English verb compounding, frequently exemplified above, is to qualify V2 with a V1 activity verb describing the manner/means by which the V2 activity is accomplished (or achieved). This use of verb compounding has been recognised for some EA languages (for example, for Japanese by Matsumoto 1996), and is evident in others, cf. the Mandarin examples given at the beginning of this section, that is, 'test-fire', 'stand-shoot', and also various areally more widespread examples resembling Lahu versatile-prehead compounds, for example, 'steal-listen' (that is, 'eavesdrop' or '(electronically) bug'), mentioned in §2 above. A complicating factor in Mandarin analysis is the possibility of 0-derivation of verbs to adverbs in V1 position, reminiscent of the analytical ambiguity of 0-derivation from verbs to nouns in English, for example, *sleep* in *sleep-walk*. Thus, in contrast to Matisoff's maintenance of the category verb for Lahu *qhô* 'steal' as a pre-head versatile, (p.212), deFrancis analyses the Mandarin parallel *qiè* (and *tou*) 'steal' as an adverb 'secretly' in the same contexts. Matisoff simply notes for Lahu that *qhô* functions in context like the English adverb *stealthily*, which overtly runs the gamut of derivational processes from verb to adverb, that is, *steal*[v-th][n-i]adj-ly]adv. This form of Mandarin compounding remains to be sorted out for the most revealing analysis. So, it

remains to be seen if Mandarin, like English, has a general grammatical resource of "manner/means" verb compounding, available for use when convenient.

## 8 Argument structure

The preceding characterisation of the semantic structure of verb compounding as *V2 PRÉP V1-ing*, is incomplete. It is clear that the process is not inhibited by object sharing, since it does not require shared transitivity. But what about subject sharing? Does the subject of V2 necessarily have to be the subject of V1? This issue also comes up in comparing different EA verb compounding systems. Although Li's (1973) comparison of Mandarin and Japanese resultatives, and Gamerschlag's (2000) more general study of the argument structure of Japanese compound verbs, focused on object sharing of the constituent verbs, their discussions indicate that these languages also differ with respect to subject sharing.

The paradigmatic example for the EA difference is the cause resultative 'beat to death', Mandarin *dǎ-sǐ* (beat-die) but Japanese *naguri-korosu* (beat-kill). The Japanese resultative requires the subject to be shared by V1 and V2. The Chinese cause resultative, in principle, does not require any argument sharing between V1 and V2; the semantic context simply requires that V1 and its arguments be construed as the cause and V2 and its arguments be the outcome. Thus, for example, X *dǎ-sǐ* Y could mean 'X beat Y until Y died' (without subject sharing) or 'X beat Y until X died' (with subject sharing). Similarly, Matthews and Yip (1994:154–155) provide a Cantonese example without any shared arguments; X *haam séng* Y (cry wake.up) 'X cried so that Y woke up'.

Japanese resultative compounding is much more restricted. Typical of Japanese is the contrast between transitive *yaki-korosu* (burn-kill) 'kill Y by burning Y' and intransitive *yake-shinu* (burn-die) 'die by burning', where the V1s *yaki* and *yake* (< *yake-i*) are based on the transitive/intransitive verb pair *yaku* and *yakeru* respectively (examples from Takahashi and Takahashi 1984). Required selection of the V1 of appropriate transitivity status insures the shared subject. Gamerschlag (2000:6) exemplifies one exception to Japanese subject-sharing, *mai-ageru* (dance-raise) '[wind] whirls up [dry leaves]', the transitive counterpart of the typical intransitive *mai-agaru* (dance-rise) '(leaves) whirl up'. He notes the more general rarity of Japanese intransitive-transitive verb compounds (as opposed to transitive-intransitive compounds where V2 is motional, for example, *moti-aruku* (hold-walk) 'walk (while) holding [something]'). However, in other cases of the rare mixed transitivity order, subject-sharing is preserved (as elsewhere), for example, *warai-tobasu* (laugh-let.fly) 'laugh [something] off' (p.c. Masatomo Ukaji). Thus, *mai-ageru* is quite exceptional.

The semantic structure of English verb compounding makes no reference to object sharing, cf. *jump-kick*, *test-listen*, but it does seem to indicate subject sharing. Thus, no

examples like Japanese *mai-ageru* are attested in English. In contexts where they are apparent, V1 is invariably attested as “unaccusative” in interpretation. Thus, for example, for attested ‘*flow-coat* something with paint’, V1 *flow* is also attested independently as transitive, i.e. ‘*flow* paint over something’. Note that, V2 *coat* alone determines the grammatical object of *flow-coat*, as expected by its status as the head.

A more irreducible exception to subject-sharing occurs with a few compounds like *spell-check*, *copy-protect*, etc. In such cases the subject of V1 is non-specific, and may or not include the subject of V2 as a possible referent, depending on the pragmatic context. For example, in *spell-check* (‘check for spelling’), the subject of V1 *spell* is unspecified, and is whoever spelled in the first place, whether the subject of V2 *check* or not. In *copy-protect* (‘protect from copying’), it is pragmatically unlikely that the subject of V2 *protect* is the subject of V1 *copy*, since under ordinary circumstances the subject of V2 ‘protect (a document)’ is opposed to unspecified other subjects of V1 who might want to ‘copy (the document)’, cf.

- (16) ... a newly developed twist-push-pull cap that is **tamper-protected** by a perforated shrink band ...

[www.petpla.net/petplanet/insider/2001/04-05/articles/inthemarket.shtml](http://www.petpla.net/petplanet/insider/2001/04-05/articles/inthemarket.shtml)

With respect to objecthood, the case of *tamper-protect* is the same as *flow-coat*. That is, V2 alone determines the logical object of the compound. The example shows that grammatical subject sharing can be restored, thanks to the long established ability of English to passivise the object of a prepositional phrase, so that *cap* as grammatical subject of passive V1 *tampered with* shares that status with passive V2 *protected*. The difference between *flow-coat* and *tamper-protect* is simply that in the latter the oblique (*with*) object of V1 is the direct object of V2, rather than the other way around. Thus, *with* does not occur in the grammatical context of (16) above.

The generalisation seems to be an additional condition to the effect that the subject of V1 can be unspecified, and, subject to pragmatic interpretation, it may include the subject of V2. Specifically disallowed by this formulation is that the subject of V1 has some other argument relation to V2 rather than subjecthood. Thus, the presumed type of Japanese *mai-ageru* is not possible in English verb compounding. It also seems that when V1 in English verb compounding is unspecified, it necessarily shares a non-subject argument with V2, but not necessarily the same grammatical non-subject type of argument, so that sharing of identical subjects can be restored by passivisation, as in (16) above. The simplest statement of the additional condition is that for purposes of verb compounding English verbs have only two kinds of arguments, subject and non-subject, and that referential identity is required of either a subject or non-subject argument of both verbs.

Currently, the additional condition to the formula *V2 PREP V1-ing* is only used with a few V2s, for example, *-check*, *-protect*. In most cases specific subjects are shared by verb compounds, and where V1 is transitive it most commonly shares the direct object with V2.

The additional condition anticipates further productivity in the less usual as well as the more usual contexts.

## 9 Concluding remarks

With an eye toward the future development of what is still a relatively new grammatical strategy in English, the preceding notes have generalised a productivity for English verb compounding beyond its most active current uses. In some cases where it is very productive in most EA languages, it is rarely used in English, for example, in directional semantic contexts, where it does not have any obvious advantage over currently productive English grammatical strategies, especially, adverbial prepositions and phrasal verbs. Its strength lies elsewhere. Its most salient strength is the economy of expression it provides for binding two (or more) activities into a single process, especially when its headedness highlights the culminating activity (accomplishment/achievement) of most interest in the context of use, for example, *freeze-dry*, *stir-fry*, *drop-kick*, (trained dogs) *sniff-search* (luggage), etc. Thus, it is not surprising to find this economy most exploited in technical domains, where recurrent complex processes are frequently referred to. There are now a sufficiently large number of generally familiar examples to promote current speakers' general awareness of verb compounding. Its utility lies in solving syntactic problems for which alternative strategies are awkward or much less economical, for example, involve wordy prepositional phrases. The following example illustrates:

(17) ... She then **drip-paints** the figures in enamel and finishes by stenciling on the words.

... [www.biddingtons.com/os/category/FLXcur167\\_23.shtml](http://www.biddingtons.com/os/category/FLXcur167_23.shtml)

What is the alternative? Maybe leaving V2 as a simple verb *paint* and then inserting a prepositional phrase 'by *dripping* enamel (paint)', most likely further necessitating an additional goal argument for *drip*, e.g. 'on it (= the posterboard, or whatever)'.

It is not clear to me that the core EA languages would profit much from adopting a pattern of verb compounding as general as that described above for English, since their syntax seems already quite economical in allowing juxtaposition of verbs without fusion (and omitting understood arguments). Thus, the productivity of semantic contexts for verb compounding beyond those generally discussed in the literature on these EA languages, but most favoured in English, remains at issue.

The situation is different for the peripheral more highly inflected EA languages such as Japanese. Here more severe constraints on the argument structure of verb compounding may play a role in preventing greater productivity. Thus, for example, Shibatani (1990:239) describes a process by which verbs can be compounded into VV<sub>n</sub> nominals, including intransitive-transitive order, for example, *tati-yomi* (stand-read) '(do) reading while stranding'. The compound economises on using a suffix like *-nagara* 'while' with

V1, but requires the use of the verb *suru* 'do' as an auxiliary to allow use as a verb. Interestingly, Beyer (1992:110 fn10) describes a similar situation for Lhasa Tibetan on the western periphery of the EA verb compounding area, for example, *ce-t.en ts.he* (< *skvel-Ndren byed*; *accompany-lead*]n do) 'ship (e.g. by courier)'. Many English VV-ing nominal compounds that do not (yet) have verb compounded counterparts are similarly attested, e.g., *bite-fighting*, *bounce-floating*, *cling-holding*, *freeze-cutting*, *hook-shooting*, *rush-cooking*, *straddle-riding*, and innumerable others. However, according to the semantic structure of English verb compounding, nothing prevents the grammatical remodeling of these nominal compounds as verbs by whoever finds it convenient to do so. In the event they are remodelled, the semantic structure of English verb compounding is sufficient to interpret the compounds without reference to the nominal compounds, to the same extent that the nominal compounds are interpretable without further context. It remains to future observation to see how soon any or all such VV-ing]n nominals are replaced by VV]v compounds. It similarly remains to the future to see how long grammatical constraints stand in the way of greater productivity of EA verb compounding. For the present it remains to be determined what those constraints are, and especially for the core EA languages, to what extent apparent limits on verb compounding are determined by pragmatics rather than by grammar. With regard to pragmatic limits, Matisoff's statement for Lahu compounds, cited more fully in section 2 above, is worth repeating: "It is as difficult to invent a comprehensible and acceptable Lahu compound as it is to create any neologism." My web-searching experience with English compound verbs suggests that speakers are currently creating neologisms faster than they can be compiled by the most ambitious dictionary.

## References

- Beyer, Stephan V., 1992, *The Classical Tibetan language*. Albany: State University of New York Press.
- Chung, Karen Steffen, 2001, Review of Packard (2000). National Taiwan University. LINGUIST List 12.844.
- DeFrancis, John ed., 1996, *ABC Chinese-English dictionary: alphabetically based computerized*. Honolulu: University of Hawaii Press.
- Gamerschlag, Thomas, 2000, *Deriving argument structure in Japanese V-V compounds*. Arbeiten des SFB 282 'Theorie des Lexikons' 114. Duesseldorf: Faculty of Philology, University of Duesseldorf.
- Gorgoniyev, Iuri A., 1966, *The Khmer language*. Moscow: Nauka.
- Lee Eun-Hee, 2000, *Dynamic and stative information in temporal reasoning Korean tense and aspect in discourse*. [Online Resource] Groningen: University Library Groningen.

- Li Yafei, 1973, Structural head and aspectuality. *Language* 69/3:408–504.
- Li, Charles N. and Sandra A. Thompson, 1981, *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press
- Marchand, Hans, 1960, *The categories and types of present-day English word-formation*. (2nd edition 1969.) München: Beck.
- Matthews, Stephen and Virginia Yip, 1994, *Cantonese: a comprehensive grammar*. London and New York: Routledge.
- Matisoff, James A., 1973, *The grammar of Lahu*. University of California Publications in Linguistics, 75. Berkeley and Los Angeles: University of California Press. Reprinted 1982.
- Matsumoto, Yo, 1996, *Complex predicates in Japanese*. Stanford: CSLI Publications.
- Mellinkoff, David, 1963, *The language of the law*. Boston: Little, Brown & Co.
- Nguyen Dinh Hoa, 1979, *201 Vietnamese verbs: compounds and phrases for everyday usage*. Hauppauge, NY: Barrons Educational Series.
- Norman, Jerry, 1988, *Chinese*. Cambridge: Cambridge University Press.
- Noss, Richard B., 1964, *Thai reference grammar*. Washington, DC: Foreign Service Institute, Department of State.
- Okell, John, 1969, *A reference grammar of colloquial Burmese*. London: Oxford University Press.
- Packard, Jerome L., 2000, *The morphology of Chinese: a linguistic and cognitive approach*. Cambridge: Cambridge University Press
- 2001, Author's response to Chung (2001). LINGUIST List 12.891.
- Shibatani, Masayoshi, 1990, *The languages of Japan*. Cambridge: Cambridge University Press.
- Sohn, Ho-Min, 1999, *The Korean language*. Cambridge: Cambridge University Press.
- Takahashi, Hiroshi and Kyoko Takahashi, 1984, *Pocket Romanized Japanese-English dictionary*. Tokyo: Taiseido.
- Thompson, Laurence, 1965, *A Vietnamese reference grammar*. Seattle: University of Washington Press.
- Wald, Benji and Lawrence Besserman, 2002, The emergence of the verb-verb compound in twentieth century English and twentieth century linguistics. In Donna Minkova and Robert Stockwell eds *Studies in the history of English: millennial perspective*, 417–448. Berlin: Mouton de Gruyter.
- Wu Ching-jung, 1999 [1976], *Pinyin Chinese-English dictionary*. Beijing: Commercial Press.

---

# 14 *Deictic patterns in Lisu and Southeastern Tibeto-Burman*<sup>1</sup>

---

DAVID BRADLEY

## 1 Introduction

As a result of the excellent work of Matisoff, Lahu is the best-described Tibeto-Burman (TB) language, which sets a very high standard for descriptions of related languages, and permits and inspires us to investigate the structure of closely related languages in a more insightful way. When we do so, due to Matisoff's work we can also more readily document any structural differences. This is particularly interesting when the forms are mainly cognate, reflecting different patterns of structural change in closely related languages.

The demonstratives of Lisu form a syntactic class, unlike the corresponding forms in Lahu. Furthermore, the syntax of the demonstratives in Lisu differs greatly from that of Lahu: they have a completely fixed order, one which is not possible in Lahu, and have a different structure from the genitive pattern found by Matisoff for the Lahu spatial demonstrative class and some occurrences of the syntactically distinct Lahu proximal form *chi*.<sup>2</sup> In Lisu, the demonstratives operate in a nominalisation or relative clause structure. In this paper, I will first outline the structure of the demonstratives in Lisu, and then make comparisons with the demonstratives of Lahu and other very closely related TB languages. In conclusion, I will draw some conclusions about the nature of deixis in these languages and the implications for Proto-TB.

---

<sup>1</sup> I am pleased to acknowledge the support of the Australian Research Council (A59701122, A59803475, A00001357) and the UNESCO endangered languages programme, and the assistance of many Lisu colleagues over a number of years. Naturally all errors are solely my responsibility.

<sup>2</sup> Forms in Lahu are cited in Matisoff's transcription.



Lisu, like Lahu, is a member of the Central subbranch of the Ni<sup>3</sup> branch of Burmese-Lolo. Lisu was originally spoken in an area immediately to the north of the traditional Lahu area in China, but has now spread far and wide: to the north into Nujiang and subsequently westwards into northern Burma and northeastern India; to the southwest into the Shan State of Burma, and thence into northern Thailand; and to the northeast into north central Yunnan and southern Sichuan. The Lisu themselves divide the core speech community into three main varieties: Northern, which is called /lo<sup>33</sup> wu<sup>55</sup>/ 'upriver', covering northwestern Yunnan, northern Burma and India; Central, called /ɕa<sup>44</sup> ɕa<sup>44</sup>/ (meaning unclear), spoken in southwestern Yunnan and along the eastern border of Burma; and Southern, called /lo<sup>33</sup> ɣl<sup>44</sup>/ 'long valley', spoken in parts of the Shan State in Burma and in Thailand. A fourth, more distinct, variety is spoken in north central Yunnan by the /li<sup>33</sup> p<sup>h</sup>o<sup>21</sup>/ 'Li men' group; most Lisu do not include this within the Lisu, but some Lipo are classified as members of the Lisu nationality in China; others are included within the much more diverse Yi nationality.

## 2 Deixis in Lisu

The Southern variety of Lisu has a maximal system of nine distinct bound demonstrative markers, including five with spatial reference, three which combine spatial and anaphoric reference, and one which may have either meaning. The forms are given in (1) below.

- |     |                                      |                                                       |
|-----|--------------------------------------|-------------------------------------------------------|
| (1) | /thø <sup>33</sup> /                 | 'this (by speaker)'                                   |
|     | /a <sup>55</sup> thø <sup>33</sup> / | 'that (by addressee)'/ 'this (by speaker, anaphoric)' |
|     | /go <sup>33</sup> /                  | 'that (by third person, on same level)'               |
|     | /ko <sup>55</sup> /                  | 'that (visible, fairly near)'                         |
|     | /nø <sup>33</sup> /                  | 'that (on higher level)'                              |
|     | /dʒø <sup>33</sup> /                 | 'that (on lower level)'                               |
|     | /a <sup>55</sup> go <sup>33</sup> /  | 'that (by third person, on same level, anaphoric)'    |
|     | /a <sup>44</sup> nø <sup>33</sup> /  | 'that (on higher level, anaphoric)'                   |
|     | /a <sup>55</sup> dʒø <sup>33</sup> / | 'that (on lower level, anaphoric)'                    |

The anaphoric forms in the Southern dialect are productively derived by the addition of an /a<sup>55</sup>/ prefix to all demonstratives except /ko<sup>55</sup>/. The /ko<sup>55</sup>/ form is somewhat isolated in Southern Lisu, but forms part of a system of three high-tone demonstratives in Central Lisu, /ko<sup>55</sup>/ 'that (further away than /go<sup>33</sup>/, on same level)', /nø<sup>55</sup>/ 'that (further away or up than /nø<sup>33</sup>/)', and /tɕø<sup>55</sup>/ 'that (further away or down than /dʒø<sup>33</sup>/)'. Note the parallel

<sup>3</sup> This subgroup of TB was formerly known as Loloish, a term now viewed as pejorative in China. There it is known as the Yi Branch, based on the new post-1950 name for the largest nationality within this subgroup, the Yi. Yi is an exonym, not used by any group as their specific autonym. I have argued (Bradley 1995:1 footnote 2) that a better name for this subgroup of TB would be Ni, based on a very widespread autonym. This also avoids confusion between the Yi nationality of China and the Ni branch of TB.

devoicing of initials where possible, and the expressive use of high tone to express greater distance, and the semantic difference in the meaning of /ko<sup>55</sup>/ in Southern Lisu and Central Lisu.

The /a<sup>55</sup> thø<sup>33</sup>/ form is both the by-speaker anaphoric form and the by-addressee demonstrative in the Southern and some Central dialects, and is also used in this way in the composite literary Lisu language. The /a<sup>55</sup>/ prefix is not used in the Northern dialect and is restricted to the /a<sup>55</sup> thø<sup>33</sup>/ form in most Central dialects. In Southern Lisu there is a partial merger of /ø/ into /e/ underway, which has affected the by-speaker and by-addressee forms which are usually [the<sup>33</sup>] and [a<sup>55</sup> the<sup>33</sup>], but not the distal higher and lower forms.

The semantics of /go<sup>33</sup>/ and forms derived from it is somewhat multifarious. In Southern and Central Lisu its basic meaning is that something is not close to the speaker or addressee, and on the same level. Sometimes it just has a general distal meaning; it can also mean next to a third person. In contrast with /ko<sup>55</sup>/ in Southern Lisu it means not visible, but is also used to refer to things which are visible but somewhat more distant; more than twenty metres or so away. In contrast with /nø<sup>33</sup>/ and /dzø<sup>33</sup>/, it refers to things on the same level; but in context it may also be used to refer to things which are known to be higher or lower and have previously been specified as such by the use of /nø<sup>33</sup>/ or /dzø<sup>33</sup>/; or it can be used to refer to distant non-visible things whose relative height is unknown; so it is the unmarked form and the form used when relative height is not specified.

In Northern Lisu, /go<sup>33</sup>/ is used to refer to things which are close to the addressee but not the speaker; the /a<sup>55</sup> thø<sup>33</sup>/ form is absent, though known to speakers who have contact with Central or Southern Lisu or speakers who are literate, since the literary language uses this form. The /ko<sup>55</sup>/ form is rare in Northern Lisu, so there /go<sup>33</sup>/ can also refer to things which are near and visible. It is thus much more frequent in Northern Lisu. There are also some allomorphs of the /go<sup>33</sup>/ form in Northern Lisu, depending on which nominal marker follows: /o<sup>55</sup> kwɑ<sup>44</sup>/ 'there', /yo<sup>35</sup> lø<sup>33</sup>/ 'that way' and so on.

Northern Lisu has some internal phonological diversity in these forms. Some speakers use /o/ or /u/ in all three distal forms, and thus have /go<sup>33</sup>/, /no<sup>33</sup>/ and /dzo<sup>33</sup>/ or /gu<sup>33</sup>/, /nu<sup>33</sup>/ and /dzu<sup>33</sup>/. Some Central speakers use /gwe<sup>33</sup>/ corresponding to Southern Lisu anaphoric /a<sup>55</sup> go<sup>33</sup>/, and some Northern speakers who normally use /go<sup>33</sup>/ use /gu<sup>33</sup>/ instead in this anaphoric context.

The demonstrative /ko<sup>55</sup>/ and similar high tone forms in Central Lisu are used only rather infrequently in Central and Southern Lisu, and they are also syntactically more restricted in various ways. There is no Southern anaphoric \*/a<sup>55</sup> ko<sup>55</sup>/, nor a single-syllable locative form (see 2.1), but these forms do occur in two-syllable locative forms (see 1.2).

All demonstrative forms are bound in every variety of Lisu; they may never occur without something following them: a nominal postposition (locative, manner, time and so on), an extensive nominal, or in the absence of one of these, the nominaliser/relativiser /mɑ<sup>44</sup>/. If there is a head nominal present, the demonstrative must follow it, as seen in (2).

If there is also a quantifier construction (numeral plus classifier) present, the quantifier construction follows the demonstrative, and is usually then followed by the nominaliser/relativiser /ma<sup>44</sup>/, as seen in (3). Sometimes, especially when the classifier is the general classifier /ma<sup>33</sup>/, the second /ma<sup>44</sup>/ may be omitted by some speakers—though it is usually present. It is also very frequently the case that there is no overt head nominal; the demonstrative plus /ma<sup>44</sup>/ very frequently acts as the head, with or without an enclosed quantifier construction, as seen in (4) and (5) below. A quantifier construction alone may also act as a head.

- (2) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ma<sup>44</sup>  
 dog this NMZR  
 'this dog/these dogs'
- (3) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup>  
 dog this five CLF NMZR  
 'these five dogs'
- (4) tʰɸ<sup>33</sup> ma<sup>44</sup>  
 this NMZR  
 'this/these'
- (5) tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup>  
 this five CLF NMZR  
 'these five'

The examples of the type seen in (2) to (5) function like any other NP, and may thus be followed by a case-marking postposition, a topic postposition, or a sequence of case plus topic or topic plus topic postpositions. Examples based on (3) are seen in (6) to (9).

- (6) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup> tɛ<sup>55</sup>  
 dog this five CLF NMZR OBJ  
 'these five dogs (object)'
- (7) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup> ɲa<sup>44</sup>  
 dog this five CLF NMZR TOP  
 'these five dogs (topical)'
- (8) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup> tɛ<sup>55</sup> ɲa<sup>44</sup>  
 dog this five CLF NMZR OBJ TOP  
 'these five dogs (object, topical)'
- (9) a<sup>55</sup> na<sup>21</sup> tʰɸ<sup>33</sup> ŋwa<sup>21</sup> ma<sup>33</sup> ma<sup>44</sup> na<sup>21</sup> ɲa<sup>44</sup>  
 dog this five CLF NMZR TOP TOP  
 'these five dogs (highly topical)'

One factor influencing the omission of /ma<sup>44</sup>/ after a demonstrative is a fairly strong constraint favouring constituents with an even number of syllables. If the constituent would otherwise have an odd number of syllables, as in NPs like (6) or (7), then the omission of the /ma<sup>44</sup>/ is more likely, especially when the NP contains something following it. This may also be why many nouns have a one-syllable and a two-syllable

alternative, such as /tsho<sup>33</sup>/ or /la<sup>21</sup> tsho<sup>33</sup>/ 'person', which permits the number of syllables in an NP to be even.

### 2.1 Contracted locative demonstratives

All forms except /ko<sup>55</sup>/ have one-syllable contracted locatives combining the demonstrative plus the locative postposition /kwa<sup>44</sup>/ into a very frequently used one-syllable form, such as /thø<sup>33</sup> kwa<sup>44</sup>/ > /thə<sup>33</sup>/ 'here', all with the initial and tone of the demonstrative and the vowel of the locative, as seen in (10). In Southern Lisu and some varieties of Central Lisu, the contracted form /a<sup>55</sup> tə<sup>33</sup>/ has two meanings, by addressee or by speaker anaphoric. Southern Lisu has all the other forms in (11) with anaphoric meaning as well; these do not occur very frequently. All uncontracted locative forms also occur, especially in environments where they allow the NP constituent to have an even number of syllables, but are much less frequent. The demonstrative /ko<sup>55</sup>/ has a contracted locative form, but it occurs only in reduplicated forms, as discussed below.

- (10) thə<sup>33</sup>      a<sup>55</sup> thə<sup>33</sup>      gwa<sup>33</sup>      nwa<sup>33</sup>      dzwa<sup>33</sup>  
       'here'      'there (by you)'      'there'      'up there'      'down there'
- (11) a<sup>55</sup> thə<sup>33</sup>      a<sup>55</sup> gwa<sup>33</sup>      a<sup>55</sup> nwa<sup>33</sup>      a<sup>55</sup> dzwa<sup>33</sup>  
       'here'      'there'      'up there'      'down there'

The locative postposition /kwa<sup>44</sup>/ has a number of alternative forms including voiced-initial /gwa<sup>44</sup>/<sup>4</sup> and reduced-initial /wa<sup>44</sup>/ or /a<sup>44</sup>/. Postposed /a/ and /o/ (with whatever tone) are not counted as independent syllables in Lisu traditional literature, and are indeed of shorter duration; thus these locative contractions follow an existing pattern, once the initial of the postposition is eroded. However, a contraction of /<sup>33</sup>/ + /<sup>44</sup>/ usually produces /<sup>35</sup>/, but in these locatives the outcome is instead /<sup>33</sup>/, keeping the tone of the demonstrative.

It should be noted that the locative form for 'there (lower)' is phonologically problematic. In a few Central subdialects there is a contrast between retroflexes [tʂ] etc. and alveopalatals [tɕ] etc. before /a/, but this is allophonic elsewhere: alveopalatals occur before front vowels and medial /j/, and retroflexes occur before nonfront vowels and medial /w/. In most Central and all Northern dialects, there is no contrast, though many Central speakers attempt to maintain one before /a/ as it is distinguished in literary Lisu. In all Northern and most Central Lisu, palatoalveolars [tʃ] etc. occur before nonfront vowels and /w/ instead of retroflexes; alveopalatals occur before front vowels. Most words which have an alveopalatal plus /a/ in those subtypes of Central Lisu which do contrast retroflexes have an alveopalatal plus /e/ in Northern Lisu, maintaining the contrast by changing the vowel; but not in this word. In Southern Lisu the retroflexes are replaced by alveolars. This means that the 'up there' form is [dʒwa<sup>33</sup>] in a few Central Lisu varieties,

<sup>4</sup> This is the song language pair word for /kwa<sup>44</sup>/, which also occurs in some subdialects as the main locative marker.

[dʒwa<sup>33</sup>] in Northern and most of Central Lisu, and [dzwa<sup>33</sup>] in Southern Lisu as seen in (10) and (11).

## 2.2 High and superhigh tone demonstratives and locative reduplication

All the demonstratives (except /ko<sup>55</sup>/ which already has a high tone) have alternative expressive forms in Southern Lisu with high tone to express an augmentative meaning: /tʰø<sup>55</sup> ma<sup>44</sup>/ 'this one very close to me', /a<sup>55</sup> tʰø<sup>55</sup> ma<sup>44</sup>/ 'that one very close to you', /go<sup>55</sup> ma<sup>44</sup>/ 'that one far over there', /nø<sup>55</sup> ma<sup>44</sup>/ 'that one way up there' and /dʒø<sup>33</sup> ma<sup>44</sup>/ 'that one way down there'. For the distal demonstratives /go<sup>33</sup>/, /nø<sup>33</sup>/ and /dʒø<sup>33</sup>/, this can be further augmented into a superhigh tone well beyond the top of the speaker's normal pitch range, to indicate that the referent is really very distant. There are also parallel high and superhigh tone forms of the contracted locatives.

In Southern and some Central Lisu varieties all the contracted distal forms also have reduplicated forms. Reduplicated forms which have the first syllable with a high or superhigh tone and the second syllable with the usual /<sup>33</sup>/ tone refer to things further away than nonreduplicated /<sup>33</sup>/ tone forms, but less far away than a single high tone form. Reduplicated forms with two high tones as in (15) are usually further than nonreduplicated high tone forms as in (14); the forms in (15) are quite infrequent and not all speakers use them. A superhigh tone instead of normal high in the first or both syllables of forms such as (14) and (15) emphasises that the referent is even more distant. The normal single-syllable distal forms as seen in (10) and repeated in (12) are usually the least far, and the reduplicated forms in (15) but with superhigh tone are the furthest. Sometimes the reduplicated forms have an alternative diminutive meaning in Southern Lisu, and so there (13) can be closer than (12) and (15) can be closer than (14). Forms given for the 'there (lower)' form are the Southern Lisu ones.

- |      |                                          |                                          |                                            |
|------|------------------------------------------|------------------------------------------|--------------------------------------------|
| (12) | <i>gwa<sup>33</sup></i>                  | <i>nwa<sup>33</sup></i>                  | <i>dzwa<sup>33</sup></i>                   |
|      | 'there'                                  | 'up there'                               | 'down there'                               |
| (13) | <i>gwa<sup>55</sup> gwa<sup>33</sup></i> | <i>nwa<sup>55</sup> nwa<sup>33</sup></i> | <i>dzwa<sup>55</sup> dzwa<sup>33</sup></i> |
|      | 'there'                                  | 'up there'                               | 'down there'                               |
| (14) | <i>gwa<sup>55</sup></i>                  | <i>nwa<sup>55</sup></i>                  | <i>dzwa<sup>55</sup></i>                   |
|      | 'there'                                  | 'up there'                               | 'down there'                               |
| (15) | <i>gwa<sup>55</sup> gwa<sup>55</sup></i> | <i>nwa<sup>55</sup> nwa<sup>55</sup></i> | <i>dzwa<sup>55</sup> dzwa<sup>55</sup></i> |
|      | 'there'                                  | 'up there'                               | 'down there'                               |

In addition to the reduplicated locative forms like those seen in (13) and (15), Southern Lisu has two-syllable forms combining the distal non-visible same-level locative /go<sup>33</sup>/ and a contracted form of the distal visible locative /ko<sup>55</sup>/, which does not occur in a single-syllable form like those in (12). These exist in every possible combination: with the /g/-initial form first (with high tone) or second (with high or mid tone), and even with the /k/-

initial form in both positions, as seen in (16) and (17). The meanings of (16) and (17) differ from the meanings of (13) and (15) in that the referent must be visible, but is further away than something normally referred to by /ko<sup>55</sup> kwa<sup>44</sup>/ alone would be; note also the mid tone form of /kwa<sup>33</sup>/ in (16) – not the locative /kwa<sup>44</sup>/ but a contracted form of /ko<sup>55</sup> kwa<sup>44</sup>/. The forms in (16) and (17) are not just a sequence of demonstrative plus locative /kwa<sup>44</sup>/; the tone is not the same. Superhigh tone forms do occur for the forms in (16) and (17), but some speakers find them semantically anomalous and so reject them.

(16) kwa<sup>55</sup> gwa<sup>33</sup>      kwa<sup>55</sup> kwa<sup>33</sup>      gwa<sup>55</sup> kwa<sup>33</sup>  
 'there (fairly far but visible)'

(17) kwa<sup>55</sup> gwa<sup>55</sup>      kwa<sup>55</sup> kwa<sup>55</sup>      gwa<sup>55</sup> kwa<sup>55</sup>  
 'there (extremely far but visible)'

### 2.3 Demonstratives and extensives

Apart from occurring before /ma<sup>44</sup>/ and before a variety of nominal postpositions, some of the demonstratives also occur before the bound nominal extent forms of Lisu. This is a class of up to ten adjectival verbs of positive extent which grammaticalise in most Ni branch languages (Bradley 1995). This grammaticalisation is one of the distinctive morphosyntactic properties of this subgroup within TB.

(18) + mja <sup>44</sup>	'number'quantity'
+ yu <sup>44</sup>	'distance'
+ ɛl <sup>44</sup>	'length'
+ mu <sup>44</sup>	'duration'
+ vu <sup>44</sup>	'size'
+ mo <sup>44</sup>	'height'
+ ci <sup>44</sup> / + hja <sup>44</sup> / + hi <sup>44</sup>	'width' <sup>5</sup>
+ ne <sup>55</sup>	'depth'
+ i <sup>h</sup> u <sup>44</sup>	'thickness' <sup>6</sup>

These forms are mainly used in reply to questions containing a question extensive nominal, as in (19). As the various possible replies seen in (19) show, it is not only demonstratives which occur in this slot; but they are more frequent than nouns or pronouns.

<sup>5</sup> The first form is the Southern Lisu one, the second is found in some Central Lisu varieties, and the third is the usual Northern and literary form and also found in most Central varieties.

<sup>6</sup> This form 'thick' occurs as a verb in all varieties of Lisu, but is not in the class of grammaticalised extensives in Southern Lisu, unlike Central and Northern Lisu. Lipo has a further extensive nominal form /li<sup>33</sup>/ derived from the adjective verb /li<sup>21</sup>/ 'heavy' which is absent from Lisu proper. For more details on the extensives of Lisu, see Bradley (2003).

- (19)  $\alpha^{44}$   $mo^{44}$   $t\phi^{h}i^{44}$   $dzo^{33}$   $\underline{a}^{21}$   $th\phi^{33}$   $mo^{44}$   $t\phi^{h}i^{44}$   $dzo^{33}$   $a^{44}$   
 how.tall reach have ? this height reach have DEC  
 'How tall is it?' 'About this tall.'
- $th\phi^{33}$   $mo^{44}$   $dzo^{33}$   $a^{44}$   
 this height have DEC  
 'This tall (exactly).'
- $hi^{33}$   $mo^{44}$   $t\phi^{h}i^{44}$   $dzo^{33}$   $a^{44}$   
 house height reach have DEC  
 'About as tall as a house.'
- $hi^{33}$   $mo^{44}$   $dzo^{33}$   $a^{44}$   
 house height have DEC  
 'As tall as a (particular) house.'
- $\eta wa^{33}$   $mo^{44}$   $t\phi^{h}i^{44}$   $dzo^{33}$   $a^{44}$   
 I height reach have DEC  
 'About as tall as me.'
- $\eta wa^{33}$   $mo^{44}$   $dzo^{33}$   $a^{44}$   
 I height have DEC  
 'Exactly as tall as me.'

The demonstratives which most frequently precede an extensive nominal in such constructions are / $th\phi^{33}$ / or / $go^{33}$ /; others are unusual but possible in sufficient context. Again, there are differences in the syntax of these forms between varieties of Lisu. In Southern Lisu, the forms without / $t\phi^{h}i^{44}$ / have a more exact meaning, and those with it have an approximative meaning. In Central and Northern Lisu the forms without / $t\phi^{h}i^{44}$ / are ungrammatical, and those with / $t\phi^{h}i^{44}$ / lack the meaning of exactness.

#### 2.4 Metaphorical use of demonstratives

The two demonstratives / $th\phi^{33}$ / and / $go^{33}$ / have various metaphorical extensions. Their temporal extension means time before/time after, using the contracted locative form / $th\alpha^{33}$ / 'here' or the base demonstrative form / $go^{33}$ / 'that', both followed by the bound nominal directional / $po^{55}$ / 'side': literally, 'here side' and 'that side' as in (20); these two temporals are not syntactically parallel. They often occur with a preceding nominal. Note that \*/ $th\phi^{33}$   $po^{55}$ / and \*/ $gwa^{33}$   $po^{55}$ / are not possible.

- (20) / $th\alpha^{33}$   $po^{55}$ / 'up to now/before now' / $go^{33}$   $po^{55}$ / 'from now/after now'

The former form is also further lexicalised into the quantifier nominal phrase / $\eta\alpha^{33}$   $th\alpha^{33}$   $po^{55}$ / 'all', literally 'be here side'.

There is also a more fully lexicalised temporal form based on the 'up there' demonstrative, / $\alpha^{21}$   $n\phi^{33}$ / 'in ancient times', usually followed by the temporal postposition / $th\epsilon^{21}$ / . This also has a reduplicated form / $\alpha^{21}$   $n\phi^{33}$   $n\phi^{33}$ / with stronger meaning. Note that the first syllable has a different tone from the Southern Lisu anaphoric form, and is found

in all varieties of Lisu, including Northern Lisu which completely lacks the /a<sup>55</sup>/ prefix on demonstratives. No other demonstrative has an /a<sup>21</sup>/ prefixed form. This is the usual introductory word in a traditional Lisu story; it is also used in the Northern Lisu compound noun 'story' /a<sup>21</sup> nɔ̃<sup>33</sup> ɲo<sup>21</sup>/, literally 'ancient.time language'; in other varieties of Lisu there is another word, /ma<sup>55</sup> mjɔ̃<sup>21</sup>/, literally 'teach story'.

### 3 Demonstratives in other Ni languages

#### 3.1 Central Ni

Diachronically the most similar system to Lisu is found in Lahu, which is also genetically very close. As Matisoff (1982:51ff, 110ff) indicates, the inventory is as follows. The proximal form is syntactically distinct, and also has a distinct locative form. The other four are grouped in Matisoff's spatial demonstrative class.

(21) <i>chi</i>	[tɕh <sup>33</sup> ]	'this'	<i>chò</i>	[tɕh <sup>o21</sup> ]	'here'
<i>ô</i>	[o <sup>54</sup> ]	'that (same level)'			
<i>nô</i>	[no <sup>54</sup> ]	'that (higher level)'			
<i>mô</i>	[mo <sup>54</sup> ]	'that (lower level)'			
<i>cô</i>	[tɕo <sup>54</sup> ]	'that (far)'			

As indicated in Bradley (1979 etyma 455, 456, 458) the cognate status of the corresponding Lisu forms with three of these, the distal, distal higher, and distal far forms, is fairly clear. The noncognacy of the proximal forms lends support to Matisoff's suggestion (1982:113), based primarily on its syntactic differences from the others, that the *chi* form is recent in Lahu. Perhaps the Lahu locative form is the cognate of the Lisu proximal /tɕh<sup>33</sup>/, though the initial place and tonal correspondences are not perfect. The Lahu distal lower form is also noncognate. Since these forms operate within a system, it is not so surprising that correspondences are somewhat irregular; there is paradigm pressure toward the same tone and rhyme, [o<sup>54</sup>] in all Lahu spatial demonstratives and [o<sup>21</sup>] in the locative proximal; [ɔ̃<sup>33</sup>] in three Lisu demonstratives and [o] in the other two.

Unlike Lisu, the position of the demonstratives in Lahu is relatively free. The four spatial demonstratives, like the corresponding forms in Lisu, are bound: if not followed by some other nominal marker (locative, temporal, etc.) they must be followed by the genitive/relative/nominaliser *ve* Matisoff (1982) analyses these as genitive constructions. Genitives and relatives marked by *ve* in Lahu more often precede the possessed head nominal, but may also follow, and this is particularly true for Matisoff's class of spatial demonstratives which occur before or after the head. When after the head they may precede or follow the numeral plus classifier, also unlike Lisu.<sup>7</sup> The Lahu proximal form

<sup>7</sup> The more frequent order in Lahu is demonstrative plus quantifier, especially when there is no overt nominal head.



*chi* also occurs in a parallel construction, before or after the head nominal and followed by *ve*, but, unlike the Lahu spatial demonstratives, it may also occur without the *ve* after the head (before or after a numeral plus classifier) and sometimes before the head. The form without *ve* often has an anaphoric function (Matisoff 1982:112), parallel to the Southern and Central Lisu /a<sup>55</sup> tʰɔ<sup>33</sup>/ form. Matisoff indicates that *chi* is productive without *ve* in posthead position, but fossilised and literary when before the head (1988:521). As in Lisu, so also in Lahu, the demonstrative construction frequently occurs without a head nominal, alone or with a numeral plus classifier before or after it.

In Lisu there is normally no genitive marker where a head is present; the possessor simply precedes the possessed head. If no head is present, a dummy head /gu<sup>21</sup>/ (or /ɣu<sup>21</sup>/ in Southern Lisu) is used, as shown in (22) below. Some Southern Lisu speakers very occasionally use the genitive head marker before a possessed head. By contrast, Lahu genitive *ve* may sometimes be omitted between two nouns, but is very frequent in all varieties of Lahu (Matisoff 1982:56ff). Thus a genitive analysis of the demonstrative forms is possible in Lahu, but not in Lisu.

(22)	(N/C Lisu)	ɲwa <sup>33</sup>	a <sup>55</sup> na <sup>21</sup>	ɲwa <sup>33</sup>	gu <sup>21</sup>	*ɲwa <sup>33</sup>	gu <sup>21</sup>	a <sup>55</sup> na <sup>21</sup>
		I	dog	I	GEN	I	GEN	dog
		'my dog'		'mine'		*		
	(S Lisu)	ɲwa <sup>33</sup>	a <sup>55</sup> na <sup>21</sup>	ɲwa <sup>33</sup>	ɣu <sup>21</sup>	ɲwa <sup>33</sup>	ɣu <sup>21</sup>	a <sup>55</sup> na <sup>21</sup>
		I	dog	I	GEN	I	GEN	dog
		'my dog'		'mine'		'my dog'		
	(Lahu)	ɲa <sup>21</sup>	ph <sup>54</sup>	ɲa <sup>21</sup>	ve <sup>33</sup>	ɲa <sup>21</sup>	ve <sup>33</sup>	ph <sup>54</sup>
		I	dog	I	GEN	I	GEN	dog
		'my dog'		'mine'		'my dog'		

In Lisu the relative or nominalised construction of most stative verbs is marked by a clause with a final /a<sup>55</sup> ma<sup>44</sup>/; such a clause may occur preceding the head or (more frequently than in Lahu) following the head; note that this is distinct from the /ma<sup>44</sup>/ alone which follows the demonstratives. Agentive-headed relatives with active intransitive or transitive verbs are instead marked with /su<sup>33</sup>/ in Lisu, and instrument-headed relatives are marked with /du<sup>33</sup>/. The Lahu cognate of the former form is the Lahu third person remote pronoun, not used as a relativiser in Lahu. Cognates of the latter form also occurs in Lahu, but as the purposive nominaliser /tu<sup>21</sup>/, not a relativiser. Thus the Lisu evidence also does not support an analysis of the demonstrative plus /ma<sup>44</sup>/ as a relative, unlike Lahu.

Another closely related Central Ni language is Lalo. In Lalo, as in many other languages in other subbranches of Ni, the demonstratives occur following the head in constituency with a classifier, with or without an intervening numeral (Björverud (1998:117ff). The general classifier in Lalo is [ma<sup>55</sup>]. As is usually the case in Lahu and often in Lisu, Lalo relative clauses precede the head; the relative marker is [ma<sup>55</sup>]. There is also a very frequent topic marker [ma<sup>55</sup>] which occurs at the end of an NP, and thus often directly follows the classifier [ma<sup>55</sup>]. This topic marker also occurs as a nonfinal clause

topicaliser [ma<sup>55</sup>] (Björverud 1998:142ff), following verbs. In Lalo, possession is optionally marked by [ɣə<sup>21</sup>] between the possessor and the head; this appears cognate with Lisu [gu<sup>21</sup>]/[ɣu<sup>21</sup>], and is structurally parallel to Lahu and Southern Lisu.

The inventory of demonstratives is much more restricted in Lalo; there are proximal and distal forms, including fused pronominal, locative and manner forms as well as the bound forms which precede a (numeral plus) classifier (Björverud 1998:76).

(23)		bound	pronominal	locative	manner
	proximal	tɕə <sup>55</sup>	tɕa <sup>55</sup>	tɕu <sup>55</sup>	tɕe <sup>55</sup>
	distal	nə <sup>55</sup>	na <sup>55</sup>	nu <sup>55</sup>	ne <sup>55</sup>

The pronominal forms may be derived from a contraction of the demonstrative plus [ma<sup>55</sup>]; Björverud notes that several of the [ma<sup>55</sup>] forms also have an alternative [a<sup>55</sup>] form. The locative forms may be derived from a contraction of the demonstrative plus the [ku<sup>55</sup>] 'inside' form (Björverud 1998:130). The manner demonstratives are presumably of a similar source; the corresponding Lisu manner postposition, very frequently used after the proximal and distal same level forms, is [lɔ<sup>44</sup>] or Southern Lisu [le<sup>44</sup>], which may account for the Lalo vowel here. The Lalo proximal form appears to be cognate with the Lahu proximal, and the distal form with the Lisu and Lahu distal higher forms. The Lalo topic nominal or nonfinal clause topic marker, relative marker and general classifier [ma<sup>55</sup>] is cognate with the Lahu general classifier *mà* [ma<sup>21</sup>] and the Lahu nominaliser *ma* [ma<sup>33</sup>], and with the Lisu general classifier [mɔ<sup>33</sup>] and relative marker/nominaliser [mɔ<sup>44</sup>]. It is striking that it is the relative or nominaliser form in all three languages which follows the demonstratives, despite the fact that the Lahu *ve* form is not cognate with the Lisu [mɔ<sup>44</sup>] and Lalo [ma<sup>55</sup>].

It is also striking that the semantics, syntax and forms of demonstratives are so different in these three rather closely related Central Ni languages. The following three examples in (24) are parallel NP forms in Lalo (Björverud 1998:120), Lisu and Lahu, all meaning 'that one (topical) monkey'; of course the Lahu form has a number of other possibilities, with the spatial demonstrative plus *ve* preceding the head or following the numeral plus classifier.

(24)	(Lalo)	a <sup>55</sup> mu <sup>21</sup>	nə <sup>55</sup>	tɕhə <sup>21</sup>	ma <sup>55</sup>	ma <sup>55</sup>
		monkey	that	one	CLF/REL	TOP
	(Lisu)	tɕa <sup>35</sup> mjə <sup>21</sup>	go <sup>33</sup>	ih <sup>21</sup>	ma <sup>33</sup>	ma <sup>44</sup> ja <sup>44</sup>
		monkey	that	one	CLF	MNZR TOP
	(Lahu)	mɔ <sup>21</sup>	o <sup>54</sup>	ve <sup>33</sup>	te <sup>54</sup>	khə <sup>33</sup> lɛ <sup>21</sup>
		monkey	that	GEN/REL	one	CLF TOP

The striking differences in constituency are obvious; also the potential for reanalysis which may have led to them.

Briefly considering other languages in this subgroup, we can note Sani /t<sup>h</sup>e<sup>33</sup>/ 'this' and /k<sup>h</sup>e<sup>44</sup>/ 'that', Axi /tɕ<sup>h</sup>i<sup>33</sup>/ 'this' and /k<sup>h</sup>u<sup>33</sup>/ 'that', Lolo (Nanhua County) /i<sup>33</sup>/ 'this' and

/k<sup>h</sup>u<sup>33</sup>/ 'that', Western Lipo /t<sup>h</sup>e<sup>44</sup>/ 'this' and /g<sup>o</sup><sup>44</sup>/ 'that', Eastern Lipo /h<sup>e</sup><sup>55</sup>/ 'this' and /k<sup>h</sup>u<sup>55</sup>/ 'that': a great deal of diversity in forms and much less complex systems than Lisu or Lahu. Western Lipo has forms essentially identical to the Lisu proximal and distal forms. Among the others, there is considerable agreement here on the distal form; this is not relatable to the Lisu, Lahu or Lalo forms, but has direct cognates in some Northern Ni languages. The Sani proximal form, like the Western Lipo form, is similar to the Lisu form; the Axi proximal form is similar to the Lahu and Lalo forms. Allowing some leeway for palatalisation, these six proximal forms might be connected.

### 3.2 Southern Ni

The syntactic situation in some Southern Ni languages is parallel to that seen in Lalo: demonstratives occur followed by a classifier, or followed by a numeral plus a classifier. The forms in Akha and Hani, two very closely related languages, are as follows.

(25)	Akha	x <sup>r</sup> <sup>33</sup>	t <sup>r</sup> <sup>33</sup> [t <sup>h</sup> <sub>Y</sub> <sup>33</sup> ]	x <sup>ø</sup> <sup>55</sup>
	Hani	ɕi <sup>55</sup>	t <sup>ø</sup> <sup>55</sup> [t <sup>h</sup> <sub>ø</sub> <sup>55</sup> ]	ø <sup>55</sup>
		'this'	'that by you'	'that'

As in various other languages, an Akha or Hani demonstrative or numeral plus classifier, or a demonstrative plus numeral plus classifier, may serve as the head of an NP. There are no distinctive fused locative forms; these are just marked by a demonstrative plus one of various locative postpositions.

In Mpi, Sah-iam (1996) describes a much more complex system of seven degrees of distance with the inventory of demonstratives given in (26) (op. cit. 97). Only four of these demonstratives are also used pronominally, with exactly the same forms: the first, second, fourth and fifth.

(26)	hi <sup>55</sup>	h <sup>r</sup> <sup>55</sup>	th <sup>r</sup> <sup>55</sup>	nr <sup>55</sup>	na <sup>55</sup>	nu <sup>55</sup>	nar <sup>55</sup>
	'this'	'that'	'that'	'that'	'that'	'that'	'that'
		(close)	(by you)	(fairly close)		(far)	(very far)

All seven demonstratives have corresponding locative forms as given in (27) (op. cit. 116). Again, we can see a fused form for some of the locatives: the first and third contain the vowel /a/; the other five are identical to the non-locative forms.

(27)	ha <sup>55</sup>	ni <sup>55</sup>	th <sup>a</sup> <sup>55</sup>	nr <sup>55</sup>	na <sup>55</sup>	nu <sup>55</sup>	nar <sup>55</sup>
	'here'	'here'	'there'	'there'	'there'	'there'	'there'
		(close)	(by you)	(fairly close)		(far)	(very far)

In Mpi the demonstratives occur immediately following the head noun, and do not require anything else to follow: no classifier, nominaliser, genitive or other subsequent item as seen in the full NP examples in (28) (op. cit. 98-99).

(28)	<i>ʔi<sup>55</sup></i>	<i>hi<sup>55</sup></i>	<i>pa<sup>33</sup></i>	<i>hy<sup>55</sup></i>	<i>ʔi<sup>55</sup></i>	<i>naŋ<sup>55</sup></i>
	house	this	book	that (close)	house	that (very far)
	'this house'		'that (very nearby) book'		'that (very distant) house'	

As in Akha and Hani, we note the by-addressee forms with initial [tʰ]. All the distal forms have initial [n] so there is presumably some kind of internal process which has led to the development of the four distinct degrees of distal marker. A similar process may account for the two proximals, which also involve vowel differences with the same initial and tone. Thus this complex system appears to have developed from a three-term system rather like that of Akha and Hani. In terms of cognacy, the various [n]-initial Mpi distal forms may be related to the Lisu and Lahu distal (higher) form.

### 3.3 Northern Ni

The Nuosu or Northern Yi system of standard Shengza as spoken in much of southwestern Sichuan and some adjacent parts of northwestern Yunnan is /tsʰɿ<sup>44</sup>/ 'this', /a<sup>33</sup> dzɿ<sup>44</sup>/ 'that (by addressee)' and /a<sup>33</sup> di<sup>55</sup>/ 'that'. The proximal appears related to various proximals elsewhere. Apart from the anaphoric forms of Lisu, the two-syllable /a<sup>33</sup>/ prefixed forms have no other syntactic parallels. The second syllable of the former may be related to the Lisu 'that (lower)' and Lahu 'that (far)' forms, and the second syllable of the latter has a chance similarity to the modern spoken Burmese 'this' form in (29) below.

The Nisu or Southern Yi forms from various areas of south central Yunnan include (from northwest to southeast) Mojiang County /e<sup>55</sup>/ 'this' and /ku<sup>55</sup>/ 'that'; Shuangbai County /ŋe<sup>55</sup>/ 'this' and /yɔ<sup>21</sup>/ 'that'; Xiping County /tɕi<sup>55</sup>/ 'this' and /ky<sup>55</sup>/ 'that'; and Shiping and Honghe counties /dy<sup>33</sup>/ 'this' and /ky<sup>55</sup>/ 'that'. We can note some links with various Central Ni proximal and distal forms here; for example, the Nisu and some Central Ni distals (Sani, Axi, Lolo and Eastern Lipo, which are also the Central Ni languages geographically closest to Nisu).

There are various Eastern Yi varieties spread across northeastern Yunnan, western Guizhou and northwestern Guangxi. The best described of these is Neasu of Weining County in Guizhou (Gerner 2002). This has a three-term demonstrative system, /tʰa<sup>55</sup>/ 'this', /na<sup>55</sup>/ 'that (nearby)' and /ga<sup>55</sup>/ 'that'. It also has a two-term anaphoric demonstrative system /tʰɔ<sup>33</sup>/ 'this (aforementioned)' and /gɔ<sup>55</sup>/ 'that (aforementioned)'; like the Lisu forms these are clearly derived from some of the basic demonstratives, but here by suffixation. As in Lalo, Hani and Akha, the Neasu demonstrative normally occurs after the head noun and with a following classifier. Gerner (2002) also cites Ngopho of Longlin County in Guangxi as having a three-term demonstrative system, /tɕʰe<sup>33</sup>/ 'this', /yɔ<sup>33</sup>/ 'that (by addressee)' and /ku<sup>55</sup>/ 'that', but no separate anaphoric forms. Nasu of Dafang County in Guizhou has a similar system but again with different forms: /tsʰɔ<sup>13</sup>/ 'this' /na<sup>55</sup>/ 'that (by addressee)' and /ŋur<sup>55</sup>/ 'that'. Most other Eastern Yi varieties mainly having two-term systems, mostly with [yur<sup>55</sup>] 'that' and a variety of forms for 'this' including (from west to

east) [ts<sup>h</sup><sub>1</sub><sup>33</sup>] in Nasu of Wuding County in Yunnan and [th<sub>0</sub><sup>55</sup>] in Nasu of Luquan County in Yunnan, and [ts<sup>h</sup><sub>0</sub><sup>13</sup>] in Nasu of Dafang County in Guizhou. Again, we can note some possible cognates here: the initial [th] and [tɕ<sup>h</sup>] forms of the proximal and the initial [g] or [ɣ] forms of the distal form, also for the by-addressee form in Ngopho.

#### 4 Burmish languages and Jinghpaw

It is striking that some Burmish languages, in the other component of this branch of TB, have demonstratives preceding the head. This is true in Burmese, where, as in Mpi, the demonstrative requires no other marking (relative, nominalisation, genitive, classifier etc.). Furthermore, the forms of the Burmese demonstrative system has changed completely since the language started to be written nearly 900 years ago. Example f(29) shows the inscriptional (and modern literary) forms as compared to the modern spoken forms.

(29) Literary	<i>ʔi</i>	<i>sap</i>	<i>tho</i>
Spoken	<i>di</i>	<i>è di</i>	<i>ho</i>
	'this'	'this (anaphoric)'	'that'

There are pronominal forms of the spoken proximal demonstratives which fuse the nominal topic marker *ha* onto the end, giving the forms *da* and *è da*. The pronominal topic form *ho ha* does not get contracted in this way, as it would then just be homophonous with the topic marker alone.

Another related language, Gong, is spoken by a small and diminishing community in western central Thailand. The demonstrative system provides another interesting pattern.

(30) <i>heŋ<sup>21</sup></i>	<i>thəŋ<sup>21</sup></i>	<i>tuk<sup>55</sup> nɔŋ<sup>33</sup></i>	<i>mɔk<sup>55</sup> nɔŋ<sup>33</sup></i>
'this'	'that' (near/visible)	'that (far/to S or W)'	'that (far/to N or E)'

The final [ŋ] of the first two forms is a fused form reflecting a reduced version of the locative postposition /nɔŋ<sup>33</sup>/, and is sometimes syllabic. There are also directional motion forms /thi<sup>33</sup> ɲe<sup>21</sup>/ 'toward the south or west' and /thi<sup>33</sup> mɔ<sup>21</sup>/ 'toward the north or east'; the former has no obvious connection with the semantically related demonstrative form. These of course imply that there is a fused /-k<sup>55</sup>/ suffix on the demonstratives. The compass direction demonstratives have no clear significance in Gong history or culture. Syntactically, Gong is similar to Burmese in that the demonstrative comes alone before the head noun; it can of course be noted that the demonstratives are followed by a locative suffix, whether fused or independent. The Gong number plus classifier component must follow the head noun, again as in Burmese.

In some Burmish languages of the Kachin group and in closely related Ngoqchangh (Achang), the pattern is more like that seen in the Ni languages. The demonstrative follows the head noun and can occur alone as in Mpi, or it may be followed by a (numeral plus) classifier as in Lalo, Hani and Akha. Example (31) gives some Ngoqchangh examples from Dai and Cui (1985:31).

- (31)  $t\zeta o^{55}$   $xai^{55}$   $t\zeta o^{55}$   $xai^{55}$   $zu^{75}$   $t\zeta o^{55}$   $xai^{55}$   $sək^{55}$   $zu^{75}$   
 person this person this CLF person this two CLF  
 'this person/these people' 'these two people'
- $t\zeta o^{55}$   $th e^{55}$   $t\zeta o^{55}$   $th e^{55}$   $zu^{75}$   $t\zeta o^{55}$   $th e^{55}$   $sək^{55}$   $zu^{75}$   
 person that person that CLF person this two CLF  
 'that person/those people' 'those two people'

A five-term system with considerable syntactic complexity is found in Zaiwa; the system as described in Xu and Xu (1984:48) is given in (32).

- (32)  $xji^{51}$   $xau^{51}$   $xje^{51}$   $xu^{51}$   $mə^{21}$   
 'this' 'that (by you)' 'that' 'that (higher)' 'that (lower)'

As in Ngoqchangh, in Zaiwa there may be a classifier after the demonstrative, but it is also possible for there not to be, as in /pju<sup>51</sup> xji<sup>51</sup>/ 'person this' (Duoshi et al. 1984:928).

Jinghpaw, which is geographically adjacent to Burmish but in another rather distant branch of TB, shows a similar semantic pattern to Lisu, Lahu and Zaiwa: a three-way distinction between distal demonstratives on the same level, higher or lower, given in (33).

- (33)  $n^{33}$   $tai^{33}$   $wo^{51}$   $th o^{55}$   $le^{51}$   
 'this' 'that' 'that (higher)' 'that (lower)'

The partial semantic congruence between Jinghpaw and Zaiwa is perhaps due to close contact; Zaiwa, like several other Burmish languages, participates in the Kachin culture complex in China and northern Burma, and virtually all Zaiwa also speak Jinghpaw natively and use it as their literary language (Bradley 1996). Burling shows that there has been very substantial convergence of semantic categories between Jinghpaw and another similar Burmish language, Maru; the example he gives is of the kinship systems (Burling 1971:27–40), but it is very frequent for there to be close lexical equivalents in various semantic fields between Jinghpaw and the other languages included in the Kachin culture complex, including Zaiwa and Maru.

## 5 Conclusion

It is not possible to posit a single syntactic pattern for the demonstratives of Lisu and the various other Ni languages discussed here. The sole generalisation appears to be that the demonstrative comes after the head nominal in all Ni and some Burmish languages; Lahu is a partial exception, though the demonstratives are most frequently postnominal there too. A postverbal demonstrative is also found in some Burmish languages, such as Ngoqchangh and Zaiwa. However the demonstrative precedes the nominal in some other Burmish languages such as Burmese and Gong. One possible reason for this is that the

development of classifier constructions, which is relatively recent,<sup>8</sup> has taken different syntactic paths in the various Ni and Burmish languages, and the demonstratives have become involved in different ways if at all. We can also note the partial parallel between the Lalo, Hani and Akha pattern in which the demonstrative is in constituency with a following (numeral and) classifier and the modern Mandarin pattern; but in Mandarin, the demonstrative, numeral and classifier precede the head noun.

The cognate forms have rather diverse meanings. For example, the etymon *\*tam*<sup>2</sup> has cognates which mean 'this' in Central Ni Lisu, Western Lipo and Sani, and 'that (by addressee)' in some Southern Ni languages. It may well be that it is partly for this reason that the prefixed Lisu form /ɑ<sup>55</sup> tɰ<sup>33</sup>/ has the additional meaning of 'that (by addressee)' as well. This etymon is perhaps further relatable to the etymon *\*to*<sup>1</sup> meaning 'that' in literary Burmese, 'that (near or visible) in Gong, and 'that (higher)' in Jinghpaw.

The etymon *\*nam*<sup>2</sup> has cognates which mean 'that (higher)' in Central Ni Lisu and Lahu, 'that (by addressee)' in Northern Ni Neasu and some Nasu, and the unmarked distal 'that' in diverse languages including Central Ni Lalo and Southern Yi Mpi. A contrasting distal etymon *\*(g)o*<sup>2</sup> is reflected in the Lisu distal visible and non-visible, Lahu, Akha, Hani, Neasu, various Nasu and spoken Burmese 'that' forms, and Ngopho 'that (by addressee)'. Another related distal etymon *\*(?)ko*<sup>2</sup> is seen in various places including Central Ni Sani, Axi and Eastern Lipo, and Northern Ni Nisu and Ngopho; but in a few places this contrasts with the preceding etymon, as in Neasu.

The etymon *\*jam*<sup>2</sup> is reflected in the Lisu 'that (lower)' form, the Lahu 'that (far)' form, and the Nuosu 'that (by addressee)' form. The original semantic value of this etymon is unclear, though it is not proximal. The only forms that may be related to the Lahu 'that (lower)' are from Burmish: the Gong 'that (far, to the north or east)' form and the Zaiwa 'that (higher)' form. However the Lahu form implies an etymon *\*mam*<sup>2</sup> with a rhyme parallel to several other demonstratives, while the Burmish forms imply something like *\*mo(k)* with a different rhyme. There is a corresponding Lisu adverbial form /mo<sup>44</sup>/ 'up there', but with the opposite meaning; the Lisu form /ɰ<sup>33</sup>/ 'down there' does not correspond.

The most problematic of all the basic forms is the proximal. It seems that the Lisu, Western Lipo and Sani forms are related to each other and to the 'that (by addressee)' etymon, found elsewhere; possibly the Lahu, Lalo and Axi proximals are related forms derived by palatalisation. One wonders whether there could be any connection between the spoken Burmese [d] initial and these forms.<sup>9</sup> We also have proximal forms in Southern

<sup>8</sup> Classifiers are absent from many branches of TB and show rather different syntax in some branches of TB. For example, in Bodo-Garo languages, classifiers, which occur only with the lowest numbers, precede the number; this classifier plus number may precede or follow the head noun.

<sup>9</sup> Voiced initials are in all cases secondary in Burmese. They are a regular outcome of juncture voicing, but occur initially in a few very frequent grammatical words such as the question nominals and the spoken proximal demonstrative, as well as some nouns with possible juncture origins.

Ni and some Burmish languages with [x] or [h] initials (and [ç] in Hani); Northern Yi forms with [ts<sup>h</sup>] and [ʔ] initials; Lolo (Central Ni) and literary Burmese forms with [ʔ] initial; spoken Burmese [d]; and Jinghpaw [nd]. Most of the proximals have the vowel [i], reflecting the universal phonesthetic tendency for forms with this meaning. The tones are also quite diverse.

Overall, by far the most complex of the modern systems in terms of the inventory is that found in the Southern dialect of Lisu. This has a six-term basic system with four further prefixed anaphoric forms (one homophonous with one of the basic terms): a total of nine distinctive forms for these ten meanings. Lisu locative forms, including the basic six, high and superhigh tone expressive forms along with reduplication provide a distinction of up to 36 different degrees of locative distance, not including anaphoric forms. These 36 include the six basic forms; three distals with expressive high tone, three with expressive superhigh tone; three reduplicated distals, each having alternative forms with one or two high tones and two more with one or two superhigh tones; six combinations of the visible and nonvisible distal with one or two high tones, and six more with one or two superhigh tones. Syntactically, Lisu is also unusual in having the demonstrative and its associated nominalisation postposition as an ambifix around the numeral plus classifier construction.

## References

- Björverud, Susanna, 1998, *A grammar of Lalo*. Lund: Department of East Asian Languages, Lund University.
- Bradley, David, 1979, *Proto-Loloish*. Scandinavian Institute of Asian Studies Monograph Series No.39. London and Malmö: Curzon Press.
- 1995, Grammaticalisation of extent in Mran-Ni. *Linguistics of the Tibeto-Burman Area* 18/1:1–28.
- 1996, Kachin. In Stephen A. Wurm, Peter Mühlhäusler and Darrell Tryon eds *Atlas of languages for intercultural communication*, 749–751 and Map 90. Berlin: Mouton de Gruyter
- 2003, Lisu. In Graham Thurgood and Randy LaPolla eds *Sino-Tibetan languages*, 222–235. London: Routledge.
- Burling, Robbins, 1971, The historical place of Jinghpaw in Tibeto-Burman. *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics* II:1–24. Urbana: Department of Linguistics, University of Illinois.
- Dai Qingxia and Cui Zhuchao, 1985, *Achangyu jianzhi* [Achang language outline]. Beijing: Nationalities Press.
- Duoshi Yongtang, Xu Xijian and Xu Guizhen, 1984, *Ham-Azi ai dian* [Chinese-Zaiwa Dictionary]. Chengdu: Sichuan Nationalities Press.



Gerner, Matthias, 2002, Demonstratives, articles and topic markers in the Yi Group. Ms.

Matisoff, James A., 1973, *The grammar of Lahu*. University of California Publications in Linguistics vol. 75. Berkeley: University of California Press. (Reprinted 1982.)

1988, *The dictionary of Lahu*. University of California Publications in Linguistics vol. 111. Berkeley: University of California Press.

Sah-Iam, Sitthichai, 1996, Deixis in the Mpi Language [in Thai]. PhD dissertation, Chulalongkorn University.

Xu Xijian and Xu Guizhen, 1984, *Jingpoyu jianzhi (Zaiwayu)* [Jingpo language (Zaiwa) language outline]. Beijing: Nationalities Press.

---

# 15 *Suffix-runs and counters in Kiranti time-ordinals*

---

BOYD MICHAILOVSKY

## 1 Introduction

Kiranti languages are notable for rich sets of time-ordinals, that is, single words for, e.g. tomorrow, the day after tomorrow, this year, last year, etc.,<sup>1</sup> which allow speakers to put off resorting to phrasal expressions rather longer than in English. The record is held by words like Kulung *nokthum* '6 days hence' and Yamphu *cukniŋ* '6 years ago', although the latter contains independently attested morphemes '6' and 'year'.

Analysis of the Kiranti time-ordinals reveals that they and their component elements are more variable and difficult to reconstruct across languages than words in other parts of the vocabulary. There are several likely explanations for this: the words (at least those expressing times closest to the present) are very common, and, as might be expected in such a structured domain, they are under heavy paradigm pressure, a difficulty which Matisoff (1997) confronted in his exhaustive study of Tibeto-Burman numerals.

To facilitate analysis, I will first present the time-ordinals in tables that bring out relationships across Kiranti, and then in tables that bring out intra-language connections between time-ordinals and numerals, between day-ordinals and year-ordinals, and between past and future domains in each language. The main mechanisms that structure this semantic area are identified as 'runs' of affixes (Matisoff 1997)—and 'counters', some of them related to the numerals, which mark off successive ordinals at increasing distance from the present.

---

<sup>1</sup> The notations D+1, D+2, Y0, Y-1, etc., will be used in the tables below.

2 **Comparative tables****Table 1:** Kiranti day-ordinals<sup>2</sup>

WK		CK					
D:	Hayu	Sunwar	Bahing	Thulung	Wambuie	Dumi	Khaling
-4			<i>lekkhabol</i>	<i>blunem</i>	<i>thāthāccum</i>		<i>aalang Kane</i>
-3			<i>sukkhabol</i>	<i>sunem</i>	<i>thāccum</i>	<i>attha:mkuna</i>	<i>aasaam Kane</i>
-2	<i>nitibuk</i>	<i>sāber na:kti</i>	<i>nikhabol, sani</i>	<i>neonem</i>	<i>sāḍi</i>	<i>atina</i>	<i>aathaasne</i>
-1	<i>tīdzon</i>	<i>'sina:kti</i>	<i>sā:ti</i>	<i>basta</i>	<i>sāiso</i>	<i>asina</i>	<i>aamiski</i>
0	<i>turi</i>	<i>mulā:kti</i>	<i>ana</i>	<i>anep</i>	<i>tyāso</i>	<i>amni-</i>	<i>ānōl'</i>
+1	<i>nukun</i>	<i>'disā</i>	<i>dilla</i>	<i>dika</i>	<i>diskānā</i>	<i>a'selma</i>	<i>disā</i>
+2	<i>niha</i>	<i>'nit-na:kti</i>	<i>niti</i>	<i>nahaḍḍa</i>	<i>nusso</i>	<i>na:mni-</i>	<i>'nām</i>
+3	<i>tshumma</i>	<i>sā:bo-na:kti</i>		<i>suhaḍḍa</i>	<i>sukbu</i>	<i>simdinna</i>	<i>'samne</i>
+4	<i>blimma</i>			<i>bluhaḍḍa</i>	<i>plyāku</i>	<i>likdinna</i>	<i>'lanne</i>
+5							
+6							

**Table 2:** Kiranti year-ordinals

WK		CK					
Y:	Hayu	Sunwar	Bahing	Thulung	Wambule	Dumi	Khaling
-6							
-5							
-4				<i>bluna</i>			
-3	<i>blikthon</i>			<i>sunā</i>	<i>mumunthot</i>		<i>aalaakto</i>
-2	<i>tshukthon</i>	<i>sāb-thoce</i>	<i>niwa</i>	<i>nōna</i>	<i>munthot</i>		<i>aajhaakto</i>
-1	<i>niḡanon</i>	<i>'sāy-thoce</i>	<i>santho</i>	<i>mamtha</i>	<i>numthot</i>	<i>a'dzəŋkhini</i>	<i>ājhāmnu</i>
0	<i>tī:thon</i>	<i>mulā-thoce</i>	<i>anamoltse</i>	<i>o:thotse</i>	<i>tyāthot</i>	<i>tom kho'o</i>	<i>ātem</i>
+1	<i>niḡahe</i>	<i>māykta</i>	<i>thotse, mata</i>	<i>pali</i>	<i>nāthot</i>	<i>namma</i>	<i>naamaa</i>
+2	<i>tshukthonḡhe</i>	<i>ni:yu</i>	<i>niwa</i>	<i>nou</i>	<i>nussothot</i>	<i>tsimma</i>	<i>chumaa</i>
+3	<i>blikthonḡhe</i>			<i>su</i>	<i>sukbuthot</i>		<i>domaa</i>
+4				<i>blu</i>	<i>plyākkuthot</i>		

<sup>2</sup> Sources of data are as follows: Hayu, Bahing Michailovsky (n.d.); Sunwar, Hale (1973); Thulung, Lahaussis (2002, revised; see table 13); Wambule, Oppenort (2002); Dumi, van Driem (1993); Khaling, Toba and Toba (1975); Kulung, Tolsma (n.d.); Camling, Ebert (1997); Bantawa, Michailovsky (n.d.), Rai (1985); Athpare, Ebert (1997); Yamphu, Rutgers (1998); Belhare, Bickel (n.d., 1997); Limbu, Michailovsky (2002), Kainla (2002).

Transcription is as in the original sources: *y* represents IPA [j] in all languages; *c* and *j* represent prepalatal affricates distinguished from alveolar *ts*, *dz* only in Sunwar and Hayu; elsewhere they represent either prepalatal or alveolar affricates; apostrophe represents high tone in Sunwar, Thulung, and Khaling; it represents stress in Dumi; *ā* represents IPA [a] in Sunwar and Wambule; *ā* represents [æ], *ǝ* represents [ø], *aa* represents [a], *a* represents [ɔ] and *ng* represents [ŋ] in Khaling; *ñ* represents [ɲ] in Wambule; *ɽ* represents [ɽ] in Bantawa.

Table 1 (cont.)

SK		EK					
D:	Kulung	Camling	Bantawa	Athpare	Yamphu	Belhare	Limbu
-4							
-3			acchosa	aumbu	khombreʔŋa	aūbu	khemya:n
-2	keska		asen	achumbu	simbreʔŋa	a'chumbu	siŋya:n
-1	espa	ase(ma)	akhumaŋ	asen	aseʔŋa	asen	asen
0	ese	ale	ai	hatle	ayu	hamba	ain
+1	desa	sela(ma)	maŋkolen	handeŋ	wagiyasiŋʔaram	warembe	ta:ndik
+2	chindi	suspaŋəi	chintolen	chindeŋ	siŋʔa	chinumba	etchinda:n
+3	ɔkthum	yaspaləi	suyaŋkolen	khaŋdeŋ	khonsum	onumba	ekkhəmda:n
+4	khethum			atupdeŋ	roksum	khonumba	ε'uda:n
+5	watthum				əksum		
+6	nocthum				cuksum		

Table 2 (cont.)

SK		EK					
Y:	Kulung	Camling	Bantawa	Athpare	Yamphu	Belhare	Limbu
-6					cukniŋ		
-5					nakniŋ		
-4					rombaniŋ	khometniŋ	
-3				aumbuniŋ	khombaniŋ	ommetniŋ	khemliŋ
-2	metəŋka		acchibatniŋ	himiniŋ	simbaniŋ	chimetniŋ	sumliŋ
-1	temniŋka	namnuŋ	anemniŋ	namniŋ	nemniŋ	namniŋ	metliŋ
0	ini	alpanuŋ	aidoŋ	naniŋ	aniŋ	nania	aillamba enniŋ
+1	nammo	nammo, waruŋ	nammaŋ	haŋdemaŋ	namma	nemma	stnimma
+2	chimo		chinmaŋ	chindemaŋ	simma	chimma	etchimma
+3	ɔmo			khaŋdemaŋ	khomma	omma	
+4	khemo			atupdemaŋ	romma	khomma	

A tentative subgrouping is proposed at the outset for ease of reference. Starting at the western edge of the range, the languages from Bahing through Khaling will be referred to as Central and Western Kiranti (CWK), and from Kulung through Limbu as Southern (SK) and Eastern (EK) Kiranti. The languages are as follows:

- WK: Hayu, Bahing, Sunwar, Thulung, Wambule;
- CK: Dumi, Khaling;
- SK: Kulung, Camling, Bantawa;
- EK: Athpare, Belhare, Yamphu, Limbu.

Hayu, the westernmost language, is somewhat marginal; it was placed outside of the East Himalayish (our Kiranti) subgroup by Shafer (1955). Limbu data, unless otherwise specified, is from the Mewa Khola dialect (Michailovsky 2002).

### 3 Single-language tables

In the following tables, day and year ordinals, past and future, are listed in four parallel columns for each language. This facilitates the identification of counters, which appear across rows, and of affix runs, which appear in columns. The numerals (if not borrowed from Nepali) are shown for reference in each table. Prefixes identified as belonging to runs are separated from following elements by a closing square bracket. Suffixes appearing in suffix-runs are marked off by an opening square bracket and listed in the last row of the column.

Table 3 illustrates this arrangement for Limbu. In each domain in Limbu there is a suffix-run: the suffixes in question are listed in the last row of the table. The two suffixes used in day-ordinals and the past years suffix are semantically motivated (see §4). The prefix *et* forms a run in the future domains. This prefix also appears in locative postpositions, for example in *etthap* 'up at/on/in', *etyo* 'down(hill) at', etc. A suffix-run is also found in the numerals: all Limbu numerals from 2 to 9 bear the plural/collective suffix *si*.

Table 3: Limbu time-ordinals

i=	number	D+i	D-i	Y+i	Y-i
0		<i>ain</i>		<i>aillamba, en[niŋ]</i>	
1	<i>thik</i>	<i>ta:ndik</i>	<i>asen</i> <i>mi:pma<sup>a</sup></i>	<i>et]nim[ma</i>	<i>met[liŋ</i> <i>mi:t[liŋ<sup>a</sup></i>
2	<i>netchi</i>	<i>et]chin[da:n</i>	<i>siŋ]ya:n</i>	<i>et]chim[ma</i>	<i>sum[liŋ</i> <i>~sil[liŋ<sup>a</sup></i>
3	<i>sumsi</i>	<i>ek]khem[da:n</i>	<i>khem]ya:n</i>	—	<i>khem[liŋ</i>
4	<i>lisi</i>	<i>ε<sup>2</sup>]u[da:n</i>	—	—	—
suffixes:		<i>ta:n</i>	<i>ya:n</i>	<i>ma</i>	<i>liŋ</i>

<sup>a</sup> Panchthar dialect.

Looking across the rows in Table 3, *\*sin* and *\*khem* can be identified as counters, that is, as elements whose function is ordinal rather than representing any specifically temporal meaning such as 'day' or 'year'. The form and distribution of these elements will be discussed in §6 below.

Time-ordinals in the remaining EK and SK languages are constructed on similar principles to those in Limbu and share some of the same elements. The data are presented in entirety in Tables 4–16.

Table 4: Belhare time-ordinals

i=	number	D+i	D-i	Y+i	Y-i
0		ha[mba		nania	
1	i	ware[mba	a]sen	nem[ma	nam[niŋ
2	sik	chin[umba	a]chum[bu	chim[ma	chim[metniŋ
3	sum	on[umba	a]ū[bu	om[ma	om[metniŋ
4	—	khon[umba	—	khom[ma	khom[metniŋ
suffixes:		umba	bu	ma	metniŋ, niŋ

Table 5: Yamphu time-ordinals

i =	number	D+i	D-i	Y+i	Y-i
0		ayu		a[niŋ	
1	ikko	wagiyasiŋ?aram	ase[?ŋa	nam[ma	nem[niŋ
2	nitci	siŋ?a <sup>a</sup>	sim[bre?ŋa	sim[ma	sim[baniŋ
3	sumji	khon[sum	khom[bre?ŋa	khom[ma	khom[baniŋ
4	ri?um	rok[sum	—	rom[ma	rom[baniŋ
5	ŋa?um	æk[sum	—	—	nak[niŋ
6	cu?um	cuk[sum <sup>a</sup>	—	—	cuk[niŋ
suffixes:		sum	bre?ŋa, ?ŋa	ma	baniŋ, niŋ

<sup>a</sup> Tentatively corrected from Rutgers 1999 '6 days ago'.

Table 6: Athpare time-ordinals

i=	number	D+I	D-i	Y+i	Y-i
0		hatle		naniŋ	
1	thik	han[deŋ	a]sen	han[demaŋ	nam[niŋ
2	i-	chin[deŋ	a]chum[bu	chin[demaŋ	himi[niŋ
				~achim[maŋ	
3	sum-	khan[deŋ	a]jum[bu	khan[demaŋ	aumbu[niŋ
4	—	atup[deŋ	—	atup[demaŋ	—
suffixes:		deŋ	bu	demaŋ, maŋ	niŋ

Table 7: Bantawa time-ordinals

i=	number	D+I	D-i	Y+i	Y-i
0		ai		aidoŋ	
1	iktat	maŋ[kolen	a]khuməŋ	nam[maŋ	a]nem[niŋ
2	hīa	chin[tolen	a]sen	chin[maŋ	a]cchibat[niŋ
3	sumka	suyaŋ[kolen	a]cchosa	—	—
suffixes:		(C)olen	—	maŋ	niŋ

Table 8: Camling time-ordinals

i=	number	D+I	D-i	Y+i	Y-i
0			<i>ale</i>		<i>alpa[nuŋ]</i>
1	<i>i-</i>	<i>sela, selama</i>	<i>ase, asema</i>	<i>nam[mo]</i>	<i>nam[nuŋ]</i>
2	<i>haka-</i>	<i>sus[paŋi]</i>	—	—	—
3	<i>sum-/sim-</i>	<i>yas[paŋi]</i>	—	—	—
suffixes:		<i>paŋi</i>		<i>mo</i>	<i>nuŋ</i>

Table 9: Kulung time-ordinals

i=		D+I	D-i	Y+i	Y-i
0			<i>ese</i>		<i>ini</i>
1	<i>i-</i>	<i>desa</i>	<i>espa</i>	<i>nam[mo]</i>	<i>temniŋ[ka]</i>
2	<i>nitci</i>	<i>chindi</i>	<i>keska</i>	<i>chi[mo]</i>	<i>metoŋ[ka]</i>
3	<i>supci</i>	<i>dok[thum]</i>		<i>do[mo]</i>	
4	<i>li:ci</i>	<i>khet[thum]</i>		<i>khe[mo]</i>	
5	<i>ŋaci</i>	<i>wat[thum]</i>		—	
6	<i>tukci</i>	<i>nok[thum]</i>		—	
suffixes:		<i>thum</i>		<i>mo</i>	<i>ka</i>

Table 10: Khaling time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>ānōl</i>		<i>ātem</i>
1	<i>'tu, tak-</i>	<i>disä</i>	<i>aa]miski</i>	<i>naa[maa</i>	<i>ājhämu</i>
2	<i>'saak</i>	<i>'nām</i>	<i>aa]thaas[ne</i>	<i>chu[maa</i>	<i>aa]jhaak[to</i>
3	<i>'suk</i>	<i>'sam[ne</i>	<i>aa]saam[kane</i>	<i>do[maa</i>	<i>aa]laak[to</i>
4	<i>bhāl</i>	<i>'lan[ne</i>	<i>aa]lang[kane</i>	—	—
suffixes:		<i>ne</i>	<i>kane</i>	<i>maa</i>	<i>to</i>

Table 11: Dumi time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>amni-</i>		<i>tom kho?o</i>
1	<i>tik</i>	<i>a'selma</i>	<i>a]si[na</i>	<i>nam[ma</i>	<i>a'dzəŋkhini</i>
2	<i>sak</i>	<i>na:mni-</i>	<i>a]ti[na</i>	<i>tsim[ma</i>	—
3	<i>suk</i>	<i>sim[dinna</i>	<i>a]ttha:mku[na</i>	—	—
4	<i>balik</i>	<i>lik[dinna</i>	—	—	—
suffixes:		<i>dinna<sup>a</sup></i>	<i>na</i>	<i>ma</i>	

<sup>a</sup> ?Cf. Nepali *din* 'day'.

Table 12: Wambule time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>tyāso</i>		<i>tyā[thoce<sup>a</sup></i>
1	<i>kwāl</i>	<i>diskānā</i>	<i>sāiso</i>	<i>nā[thoce<sup>a</sup></i>	<i>num[thoce<sup>a</sup></i>
2	<i>nisi</i>	<i>nusso, nusswām</i>	<i>sāḍi</i>	<i>nusso[thoce<sup>a</sup></i>	<i>mun[thot</i>
3	—	<i>sukbu</i>	<i>thāccum</i>	<i>sukbu[thoce<sup>a</sup></i>	<i>mumun[thot</i>
4	—	<i>plyākku</i>	<i>thāthāccum</i>	<i>plyākku[thoce<sup>a</sup></i>	—
suffixes:				<i>thoce, thot</i>	<i>thoce, thot</i>

<sup>a</sup> Also *-thot*.

Table 13: Thulung time-ordinals (Lahaussis 2002<sup>a</sup>)

i =	number	D+I	D-i	Y+i	Y-i
0			<i>anep, ane</i>		<i>othotse</i>
1	<i>ko</i>	<i>dika</i>	<i>basta</i>	<i>pali<sup>b</sup></i>	<i>mamtha</i>
2	<i>no<sup>c</sup></i>	<i>na[haḍḍa</i>	<i>neo[nem</i>	<i>no</i>	<i>no[na</i>
3	— <sup>d</sup>	<i>su[haḍḍa</i>	<i>su[nem</i>	<i>su</i>	<i>su(na)</i>
4	— <sup>e</sup>	<i>blu[haḍḍa</i>	<i>blu[nem</i>	<i>blu</i>	<i>blu(na)</i>
suffix:		<i>haḍḍa</i>	<i>nem<sup>f</sup></i>		<i>na</i>

<sup>a</sup> The day-ordinals have been revised slightly by Lahaussis (p.c. 21.11.2002).

<sup>b</sup> Nepali *pāli* 'time, turn, year'.

<sup>c</sup> Allen 1975: *nə* '2'.

<sup>d</sup> Allen 1975: '*su ~ sium* '3'.

<sup>e</sup> Allen 1975: *blə* '4'.

<sup>f</sup> 'day'.

Table 14: Bahing time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>ana</i>		<i>anamoltse</i>
1	<i>koṅ</i>	<i>dilla</i>	<i>sā:ti</i>	<i>thotse, ma:ta</i>	<i>santho</i>
2	<i>niksi</i>	<i>niti</i>	<i>ni[khabəl, sani</i>	<i>niwa</i>	<i>niwa</i>
3	<i>sam</i>	—	<i>suk[khabəl</i>	—	—
4	<i>le</i>	—	<i>lek[khabəl</i>	—	—
suffix:			<i>khabəl</i>		



Table 15: Sunwar time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>mu[ɫā:kti</i>		<i>mulā-thoce</i>
1	<i>kā:</i>	<i>'disā</i>	<i>'si[na:kti</i>	<i>māykta</i>	<i>'sāy-[thoce</i>
2	<i>'niksyi</i>	<i>'nit-[na:kti</i>	<i>sāber [na:kti</i>	<i>niryu</i>	<i>sāb-[thoce</i>
3	<i>'sā:</i>	<i>sā:bo-[na:k</i>			
suffixes:		<i>na:kti</i>	<i>na:kti</i>		<i>thoce</i>

Table 16: Hayu time-ordinals

i =	number	D+I	D-i	Y+i	Y-i
0			<i>tiri</i>		<i>tī:[thoŋ</i>
1	<i>koŋ</i>	<i>nukun</i>	<i>tīdzoŋ</i>	<i>niŋa[he</i>	<i>niŋanoŋ<sup>a</sup></i>
2	<i>nak-</i>	<i>niha</i>	<i>nitibuk<sup>b</sup></i>	<i>tshukthoŋ[he</i>	<i>tshuk[thoŋ</i>
3	<i>tshuk</i>	<i>tshum[ma</i>	—	<i>blikthoŋ[he</i>	<i>blik[thoŋ</i>
4	<i>bli-</i>	<i>blim[ma</i>	—	—	—
suffixes:		<i>ma</i>		<i>he<sup>c</sup>, thoŋhe</i>	<i>thoŋ</i>

<sup>a</sup> The suffixed element *noŋ* is a locative suffix 'at'.<sup>b</sup> Hodgson (1880:232) cites *buk* as a quantifier for days. ?Cf. Athpare, Belhare suffixed *bu*.<sup>c</sup> Locative suffix 'in'.

#### 4 Lexical elements in time-ordinals

Lexical elements in time-ordinals are of two types, (1) independently attested elements with (usually) time-related meanings and (2) other elements occurring in time-ordinals in more than one of the languages and whose meaning can be inferred. Discussion of affixes appearing in single languages and of counters, whether or not related to numerals, is deferred to §5 and §6 below.

##### 4.1 Kiranti words for 'day', 'now', 'today'

The Proto-Kiranti (PK) etymon *\*len* 'day' is best attested in SK and EK: Thulung *lem* 'day [bound quantifier]'; Kulung *lei* 'day'; *le:pa* 'afternoon'; Camling *lēi, lei, lēi* 'day', *kholēi* 'day, daylight' (cf. *khosai* 'night'); Bantawa *len* 'day', *kholen* 'day'; Athpare *lemba* 'day, light'; Yamphu *lemda* 'afternoon'; Belhare *lemba* 'daytime'; Limbu *len* 'day [bound quantifier]', *lendik* '(by) day' (vs *sendik* '(by) night'). Van Driem (1987) suggests the presence of *\*len* in Dumi *lemma* 'daydream' (Cf. also STC #82 *maŋ* 'dream', PK *\*maŋ* 'spirit' and PK *\*senmaŋ* 'dream', the latter a compound with 'night'). In reflexes of PK *\*len* and of the etymon 'night' (below), final *n* is often dropped, particularly in SK and CK, leaving a *Vi* diphthong. Reflexes appear in Athpare *hatle* and Camling *ale* 'today', and in suffix-runs in SK future day ordinals: (C)olen in Bantawa ('D+1', 'D+2', 'D+3') and *palēi* in Camling ('D+2', 'D+3').

WK *\*ti* 'day, daytime' appears sporadically: in Hayu *tiri* 'today' and future days, in Bahing *niti* 'day after tomorrow ['second day']' and *sā:ti* 'yesterday ['(last) night-day']'

(vs *santho* 'last year'), in Sunwar, and probably in Wambule (*sāq̄i* 'day before yesterday' ?< \**san-ti* cf. Opgenort 2002:9) and Kulung (*chindī* 'day before yesterday'). It is opposed to an element meaning 'evening, night' in Sunwar: *nā:kti* 'daytime, afternoon' vs *nā:kdo* 'evening'; *'sinā:kti* 'yesterday' vs *'sinā:kdo* 'last night'. Cf. also Bahing *namti* 'daytime', a compound with PK \**nam* 'sun' (Benedict 1972:148n).

In looking for reflexes of PTB \**niy* (STC #81) we find candidates as suffixes in day-ordinals in Dumi (*n̄*), Khaling (*ne*) and Bahing (*nī*).

PK *han ~ an* 'now': Bahing *ana* 'today', Khaling *ān*, Kulung *hən*, Bantawa *han*, Athpare *hatle* 'today, now' (compound with 'day'), Yamphu *hago*, Limbu *allo* (phonologically \**an-lo*). Athpare *handeŋ* 'tomorrow' probably reflects this etymon.

Probably related are Limbu *ain*, Bantawa *ai* Belhare *hamba* and Yamphu *ayu* 'today'. 'Today' appears in Bantawa *aidoŋ* 'this year' (compound with 'year'), Limbu *aillamba* (\**ain-lamba*) and Bahing *anamoltse* 'this year'.

Khaling *ānōl* 'today' is composed of *ān* 'now' and *nōl* 'day, afternoon'. Camling *ale* 'today' is composed of \**len* 'day' with a prefixed *a* (cf. *aso* 'now'); *alpaŋuŋ* 'this year' can be analysed as *al(e)* 'today' + *pa* '[nominaliser/modifier]' + *nuŋ* 'year'. (See Ebert 1997:10 on *i ~ u* variation in Camling.)

The Wambule words for 'today' and 'this year' contain an element *tyā*, identified by Opgenort (2002) as a bound morpheme meaning 'now'; cf. *tyāño* 'from now'.

For Thulung *anep* 'today' and *othotse* 'this year', cf. Thulung *a-* 'this', *nem* 'day', *thotse* 'year [bound form]'.

#### 4.2 'Night', 'yesterday'

PK \**sen* 'night': Thulung *sintha*, Dumi *sina*, Khaling *sene*, Kulung *se:pa*; Camling *khosai*; Yamphu *senda*; Belhare *semba*; Limbu *sendik*. (Note the parallels with words for 'day' containing reflexes of \**len*.) The most widespread word for 'yesterday', PK \*(*a*)*sen* is clearly related to 'night', a widely observed semantic connection. This is best seen in EK: Limbu *assen* (Phedap *anchen*) 'yesterday', Belhare *asen*, Yamphu *aseŋa*, Athpare *asen*. Cf. also Dumi *asina*, Wambule *sāiso* [analysed as 'past'-'day' by Opgenort 2002]—this last, like some other WK forms seems to reflect \**san*. In Bahing, this etymon appears in both *sā:ti* 'yesterday' and *santho* 'last year' (cf. Sunwar *'sina:kti*, *'sāy-thoce*). Bantawa *asen* and Camling *ase* have been pushed back to 'D-3' and 'D-2'.

Limbu has a specific word *mit* 'yesterday', which appears in Panchthar dialect *mipma* 'yesterday' and *mitliŋ ~ mi?liŋ* (Mewa Kholā *metliŋ*) 'last year'.

#### 4.3 'Tomorrow'

CWK \**dis* ~ \**disa* appears to be a specific CWK etymon meaning 'tomorrow', attested in Sunwar, Bahing, Thulung, Wambule, Kaling and Kulung. It is probably borrowed into Kulung since (1) otherwise initial *t* would be expected and (2) it is not found elsewhere in SK or in EK. This element (or at least \**diC-*) also appears in Khaling *dicchä* and Kulung *detcha* 'morning' (?also borrowed), a common 'tomorrow' etymology. The element *ka* in Thulung *dika* is a general adverbial suffix whose use with other time-ordinals is optional.

In Dumi, we find a different etymon, *sel* 'tomorrow', shared with Camling.

Yamphu and Belhare have a common element beginning with *wa*. Cf. Chhingtangya *warangda*, Lohorong *wengda*, Lambichhong *waring* 'tomorrow' (Hodgson 1880:186,206).

Limbu *ta:ndik* 'tomorrow' contains the Limbu element *ta:n* 'morning, day'—used as a suffix in Limbu D+2 and following—and a suffix *dik*, which also appears in *sendik* '(by) night' and *lendik* '(by) day'; cf. Yamphu *sendak*, *lendak*. The suffix *dik* is related to the general Limbu adverbialising suffix *rik/lik*. Cf. Limbu *ta:nchoppa* 'morning star'.

#### 4.4 'Year'

The element *nij* 'year' (*lij* in Limbu, an irregular correspondence) (STC #368) appears as a bound form throughout SK and EK in words for 'this year' and for past years.

Athpare *nanij* and Belhare *nania* 'this year' contain a proximal demonstrative *na* 'this'; Limbu *enni* has the same structure.

Words for future years bear suffixes reconstructible as *\*maŋ* (SK, Athpare, also CK) or *\*ma* (EK), whose etymology is unknown.

PK *?\*toŋ* 'year'<sup>3</sup> has reflexes in all branches of Kiranti, usually as bound forms (this can not be determined in all of the sources): Hayu *thoŋ* 'year'; Sunwar *thoce* [?bound form], Bahing *tho* [bound form]; Thulung *tho*; Wambule *thot ~ thoce* [bound form], Khaling *tho* 'year', *to* [suffix in past-year-ordinals]; Kulung *doŋ* [bound quantifier]; ?Camling *camduŋ* 'season'; Bantawa *doŋ* 'year'; Limbu *toŋ* [bound quantifier], *toŋbe* 'year'. This element appears in many time-ordinals, in suffix-runs and in Bantawa *aidoŋ* 'this year ['today-year']' and in Kulung *metoŋka* 'year before last'. The initial *t* in Kulung *metoŋka* suggests *\*met-doŋ*, ?cf. Limbu *met* 'yesterday'.

#### 4.5 Next year, last year; past-future contacts

An etymon *nam ~ nem ~ nem ~ nim* appears in almost all CK, SK and EK words for 'next year', with a suffix *maŋ ~ ma ~ mak*, usually beginning a suffix-run. Limbu *etnimma* is prefixed, beginning a prefix-run, and the vowel has been influenced by the following item in the run, *etchimma* 'Y+2'. In many of the languages, elements similar to *nam* (etc.), appear in words for 'Y-2' as well. The original meaning of *nam* (etc.) is not clear. The fact that it usually permutes with *\*sim ~ chim* in the adjacent item meaning 'Y+2' or 'Y-2' might suggest that it is a counter. But it only occurs in the year domain. It may have originally meant 'next year' and become extended in some languages to 'Y±1', like the famous Hindi words *kal* 'D±1' and *parsō* 'D±2'. The languages in which it appears in both 'last year' and 'next year' are Yamphu, Belhare, Bantawa and Camling; in all there is a difference of suffix, and in all but the last a difference of vowel between the two forms.

Bahing *niwa* 'Y±2' is a past/future term, probably containing the numeral '2'.

Hayu *nija* appears in both *nijanoŋ* 'Y-1' and *nijahe* 'Y+1', with different locative suffixes. Similarly, Hayu has *tshukthoŋ* 'Y-2', *tshukthoŋhe* 'Y+2'; *blikthoŋ* 'Y-3',

<sup>3</sup> The reconstruction is questionable because the manner-series correspondence is irregular. The regular correspondences are CWK \*t ~ SK \*d ~ EK \*th and CWK \*d ~ SK and EK \*t (Michailovsky 1994). CWK aspiration is secondary. Note that this etymon is missing in EK except for Limbu, and that there is no evidence for a final in CWK except in Hayu.

*blikthoghe* 'Y+3'; the Y+ ordinals differentiated from the past by the addition of a locative suffix. The elements *ni*, *tshuk*, and *blik* suggest the numerals '2', '3' and '4', but *niŋa* may rather reflect SK and EK *niŋ* 'year'.

Bahing *ma:ta* and Sunwar *māykta* 'next year' suggest a common Bahing-Sunwar ancestor \**matta*, not attested elsewhere.

Athpare *handemaŋ* 'next year' is formed on *handeŋ* 'tomorrow'.

Table 17: Suffix-runs in day-ordinals

suffix	Language	in day-ordinals	note
<i>ta:n</i>	Limbu	D+2, D+3, D+4	'morning'
<i>ya:n</i>	Limbu	D-2, D-3	'day'
<i>mbu</i>	Athpare, Belhare	D-2, D-3	
(V) <i>mba</i>	Belhare	D0, D+1, D+2, D+3, D+4	
(bre?) <i>ŋa</i>	Yamphu	D-1, D-2, D-3	
<i>sum</i>	Yamphu	D+3, D+4, D+5	
<i>deŋ</i>	Athpare	D+1, D+2, D+3, D+4	?cf. CWK ?* <i>de</i> 'tomorrow' cf. <i>demaŋ</i> in Athpare, Y+ ordinals
(C) <i>olen</i>	Bantawa	D+1, D+2, D+3	cf. <i>len</i> 'day'
<i>palai</i>	Camling	D+2, D+3	cf. <i>lai</i> 'day'
<i>thum</i>	Kulung	D+3, D+4, D+5, D+6	
<i>kane</i>	Khaling	D-3, D-4	cf. Dumi <i>kuna</i> 'D-3'
<i>ne</i>	Khaling	(D-2), D+2, D+3	cf. Khaling <i>kane</i> , Thulung <i>nem</i>
<i>dinna</i>	Dumi	D+3, D+4	cf. Khaling <i>ne</i> ; ?Nepali <i>din</i> 'day'
<i>Cu</i>	Wambule	D+3, D+4	(?)
<i>so</i>	Wambule	D0, D-1, D+2	(not a continuous run)
<i>nem</i>	Thulung	D-2, D-3, D-4	cf. <i>nem</i> 'day'; cf. Khaling <i>ne</i>
<i>haqqa</i>	Thulung	D+2, D+3	
<i>khabol</i>	Bahing	D-2, D-3, D-4	
<i>ma</i>	Hayu	D+3, D+4	

## 5 Affixes and affix-runs

Time ordinals often have the form of compounds in which the second element is shared with neighboring ordinals belonging to the same series, whether of days or years. The sources of a few of these elements, with independently attested meanings 'day' or 'year', can be found in the previous section. But, as Matisoff has pointed out, affix-runs are partly or even wholly phonologically motivated, to facilitate rhythmic recitation, independently of the semantic origin of the affixed elements. In the case of time-ordinals, it may happen that the final element in the word for, say, D+i, is extended to D+(i+1) and following items and acquires an association with 'day' or 'year' that it did not have originally.

We have also encountered a few prefixes, usually prefixed *a*. Runs of prefixes in the past days domain often start from 'today'.

Year-ordinals often contain recognisable morphemes for 'year', precisely to distinguish them from related day ordinals. But suffix-runs in day-ordinals are highly variable and generally limited to single languages. These are listed in Table 17. Most have no independently attested meaning.

## 6 'Counters'

By 'counters' I mean elements whose meaning is primarily ordinal, even if limited to the temporal domain, rather than associated with specific temporal concepts like 'day', 'yesterday', etc. In practice, the distinction is not necessarily clear-cut.

The status of 'counter' is clearest in elements which are etymologically numerals and which are used for both days and years, past and present. Such elements are found in Western Kiranti. In Eastern and Southern Kiranti we find elements of similar distribution which are probably not etymologically numerals.

The element *nim* in Limbu 'next year' has been discussed in §4.5 above. Elements *nam/nem* appear for 'Y±1' throughout EK and SK except in Athpare, where 'next year' is based on 'tomorrow', in Limbu, where 'last year' is based on 'yesterday', and in Kulung 'Y-1'. They do not appear to be counters.

The Limbu counter *\*sin* was mentioned in presenting Table 3 above. Final *n* is reconstructed internally because the finals *m* and *ŋ* do not usually assimilate. Limbu *ch* only occurs as an allophone of *s* after *t* or *n* (here after the *t* of the prefix *εt*). Panchthar dialect *silliŋ* 'Y-2' can only reflect *\*sin+liŋ* phonologically. Probably *\*sin* is the original Limbu form, replaced by *sum* '3' in the Phadap and Mewa Khola dialect terms for 'Y-2'. It is possible that *\*sin* (in fact, EK *\*chin* — see below) somehow reflects *\*sum* '3', but this hypothesis is rejected here in view of the phonological difference and of the fact that the following counter, *khem* is not related to any known numeral '4'.

The counter *\*chin* is well represented elsewhere in EK and in SK at the i=2 level. It has the vowel *u* before the suffix *bu* in Athpare and Belhare, and the final *m* before suffix-initial bilabials. EK and SK initial *\*ch* (corresponding to PK *\*c*) regularly has the reflex *ch* in Limbu and Yamphu, merging with reflexes of PK *\*s*: compare, for example, the reflexes of PK *\*cap* 'to write': Dumi *tsəpt-*, Kulung *chap-*, Camling *chapd-*, Bantawa *chapt-*, Athpare *chept-*, Yamphu *sap-*, Belhare *chap-*, Limbu *sapt-*.

EK and Kulung have counters at higher levels (i>2), but these differ between languages. Still, at the i=3 or i=4 level all four EK languages have a counter *khVN*, and three of them also have a counter beginning with *o* or *u*. These counters, beginning with i=2 and ignoring some phonetic variation, are shown in Table 18. Separate counters for past (-) and future (+) are listed for Athpare. Kulung has counters only in future time-ordinals. Bantawa and Camling are omitted as lacking counter-series.

Table 18: Kulung and EK counters

i=	Kulung+	Athpare+	Athpare-	Yamphu	Belhare	Limbu
2	<i>chin</i>	<i>chin</i>	<i>chin</i>	<i>siŋ</i>	<i>chin</i>	<i>sin</i>
3	<i>dək</i>	<i>khaŋ</i>	<i>um</i>	<i>khəŋ</i>	<i>on</i>	<i>khəm</i>
4	<i>khet</i>	<i>atup</i>	—	<i>rok</i>	<i>khon</i>	<i>u</i>
5	<i>wat</i>	—	—	<i>(n)ak</i>	—	—
6	<i>nək</i>	—	—	<i>cuk</i>	—	—

The counters of Table 18 show few resemblances to Kiranti or TB numerals. An exception is Yamphu *cuksum* '6 days hence' and *cukniŋ* '6 years ago', which clearly contain Yamphu *cuk* '6', itself a reflex of EK \**tuk* '6' (cf. STC #411, completed by Matisoff 1997:81). Note that Yamphu regularly affricates EK \**t* before *i*, and sometimes before *u*, e.g. Yamphu *cupt-*, Athpare *tup-*, Limbu *tum-* 'to meet'. The preceding Yamphu counters, *ro(C)* and *nak* ~ *æk* are probably influenced by the Yamphu numbers *ri?* '4' (PK \**bli*, EK \**li*) and *ŋak-* '5' (cf. *nakpoŋ* '50'). Yamphu *r* is the regular reflex of PK initial \**l*; cf. Yamphu *ram* 'road', *ruŋma* 'liver', etc.

In the western languages (CWK), Khaling and Dumi resemble SK in the future years paradigm (e.g. Khaling *naamaa* 'Y+1', *chumaa* 'Y+2', *domaa* 'Y+3'; cf. Kulung *namə*, *chimə*, *dəmə*). The correspondences *ch* ~ *ch* and *d* ~ *d* between Kulung and CK point to borrowing, probably by CK.

Other counters in CWK are generally related to numerals. The Kiranti numerals 2, 3, and 4 often have *k* finals, an oddity from a Tibeto-Burman point of view, and these appear in some of the counters. In comparing reflexes of PK \**ni(C)* '2' and \**sum* '3' it should be noted that Khaling *a* and Thulung *ɔ* (ə in Allen 1975) often reflect PK close vowels.

Reflexes of PK \**ni(C)* '2' are found as counters in WK time-ordinals: *nus* appears in Wambule 'D+2' and 'Y+2', *na* ~ *neo* ~ *nə* in Thulung (all time-ordinals at the i=2 level), *ni* in Bahing, *nit* in Sunwar (future only), *ni* in Hayu (i=2 for days, ?i=1 for years—but Hayu *niga* may be related to SK and EK *niŋ* 'year'). The peculiar CK etymon \**sak* '2' does not seem to figure as a counter in CK or elsewhere. Dumi *namni-* and Khaling *hām* appear to be specialised words for 'D+2' rather than to reflect a counter '2'.

Khaling and Dumi *samne* and *sindinna* 'D+3' and Khaling *aasaamkane* 'D-3' reflect PK \**sam* ~ \**sum* '3', while Wambule *sukbu* 'D+3', Bahing *sukkhəbəl* 'D-3' and probably Thulung *su-* (in all time-ordinals for i=±3) reflect \**suk* (cf. also Hayu *tshuk-* '3'). It is interesting that some of the languages have final *k* in the numeral but final *m* in the counter (Khaling, Dumi, Hayu 'D+3'), while Bahing has the opposite.

CK 'D±4' reflects PK \**bli* '4', with variable final consonants. Reflexes with final *k* appear in Wambule (i=+4), Bahing ('D-4'), and Hayu ('Y±3'). Dumi *likdinna* 'D+4' clearly contains Dumi *balik* '4', minus the dimidiated *b* prefix, which appears in Wambule *plyākku* (in 'D+4', 'Y+4'). The vowel of Thulung *blu* (<'4') appears to show the influence of the preceding *su* (<'3'). It is interesting that PK '3' and '4' are represented in Thulung and Wambule time-ordinals, although the corresponding numerals have been lost.

Wambule uses reduplication rather than counters to advance from 'Y-2' to 'Y-3' and from 'D-3' to 'D-4'. Reduplication is not used elsewhere in Kiranti time-ordinals.

The CWK counters are summarised in Table 19. Sunwar is omitted, as the relevant words are simply numeral expressions.

**Table 19:** Numeral-related counters in CWK

i=	Hayu	Bahing	Thulung	Wambule (+)	Dumi (D+)	Khaling (D)
2	<i>ni</i>	<i>ni</i>	<i>no</i>	<i>nus</i>		
3	<i>tshum</i> (D+3) ~ <i>tshuk</i> (Y-2)	<i>suk</i>	<i>su</i>	<i>suk</i>	<i>sim</i>	' <i>sam</i>
4	<i>blik</i> (Y±3) ~ <i>bli</i> (D+4)	<i>lek</i>	<i>blu</i>	<i>plyāk</i>	<i>lik</i>	' <i>lan</i>

## 7 Conclusion

Time-ordinals are a well-defined semantic area, like numerals, with perhaps less extensibility, but with more dimensions, at least as compared to the positive integers. This multi-dimensionality gives scope for exchange of morphemic material along different axes—between past and future, or from days to years—in addition to the possibility, shared with the numbers, of ‘runs’ along a single axis. These multiple possibilities have given rise to wide variation among closely related languages.

Perhaps the most striking aspect of the Kiranti time-ordinals is the great variety of morphemes that have been either borrowed or created to fill out the domain. Of course, a certain number of etyma, like ‘day’ and ‘night’, appear repeatedly, but it is remarkable that over a dozen otherwise unattested morphemes, as far as we know, have been created and used to form suffix-runs, each in a single language. These are concentrated in the day-ordinals.

The ‘counters’ are remarkable both for their similarities to numerals and for their differences from them. In West Kiranti, the counters are clearly related to ordinary inherited numerals. In the East this does not appear to be the case, except at higher levels in Yamphu. It is interesting that inherited numeral morphemes can be preserved in time-ordinals when they have been eliminated by borrowings in the numeral domain (e.g. in Wambule and Thulung). This would seem to reflect the pressure of commerce on the numerals. Even Kiranti numerals composed of Tibeto-Burman raw material may show evidence of such pressure, perhaps from Tibetan; time ordinals may furnish clues to an earlier state, altered by the many analogic processes that are at work in this domain.

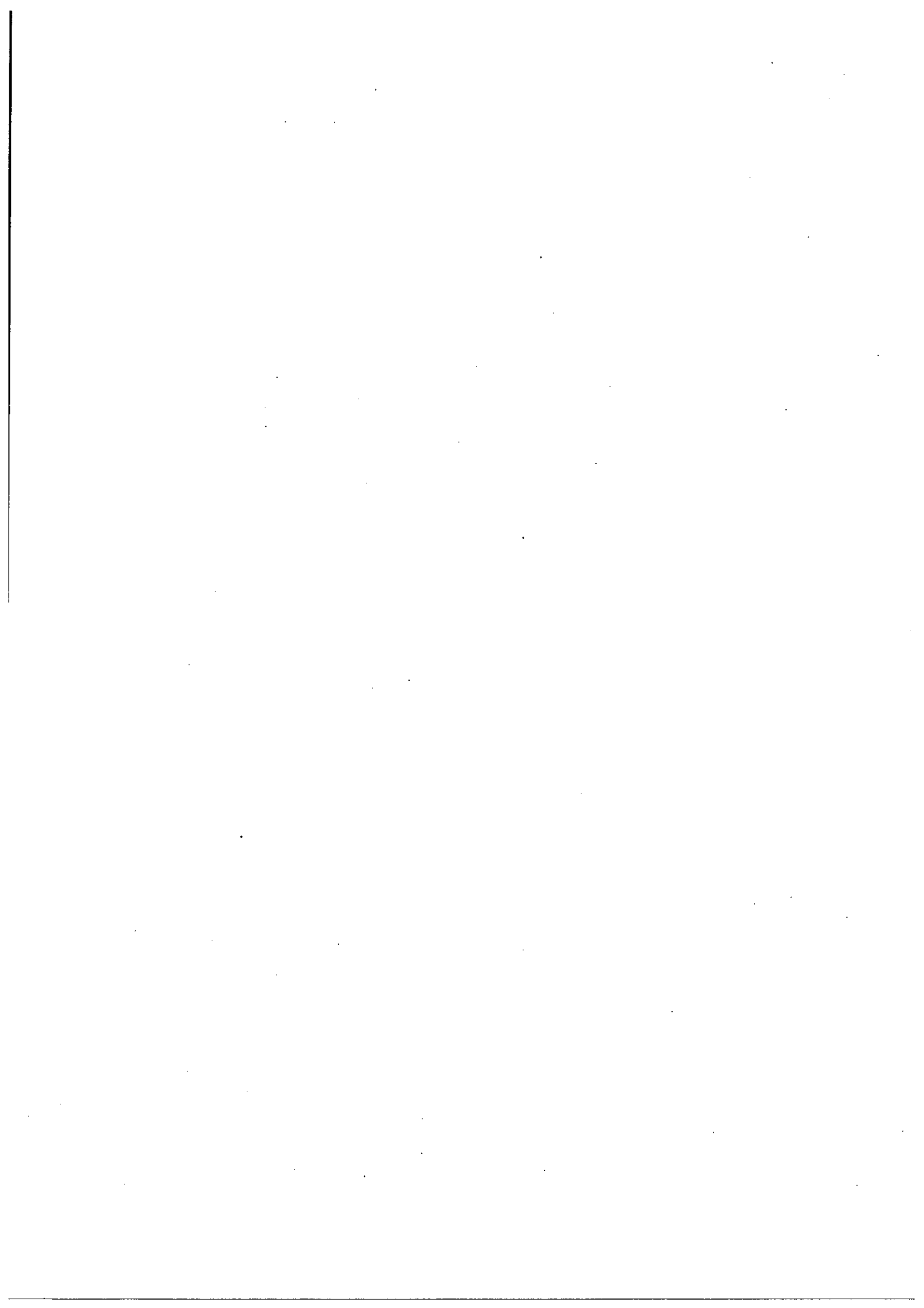
The study of Kiranti time-ordinals confirms what Matisoff (1977) has shown by his study of the Tibeto-Burman numerals: the richness and complexity of what might seem, *a priori*, to be a straightforward semantic domain.

## References

- Allen, Nicholas, 1975, *Sketch of Thulung Grammar*. Cornell University East Asian Papers No.6. Ithaca: Cornell University.
- Benedict, Paul K., 1972, *Sino-Tibetan: a conspectus*. (contributing editor James A. Matisoff). Cambridge: Cambridge University Press. [Proto Tibeto-Burman (PTB) reconstructions from this source are cited as ‘STC #xxx’ in the text.]
- Bickel, Balthasar, 1997, Time reckoning in Belhare. Ms.

- 1998, Belhare vocabulary. Ms.
- Driem, George van, 1987, *A grammar of Limbu*. Berlin: Mouton de Gruyter. [Phedap dialect]
- 1993, *A grammar of Dumi*. Berlin: Mouton de Gruyter.
- Ebert, Karen, 1997, *A grammar of Athpare*. München: Lincom Europa.
- 2000, *Camling texts and glossary*. München: Lincom Europa
- Hale, Austin, 1973, *Clause, sentence, and discourse patterns in selected languages of Nepal 4*. Word Lists. Norman, OK: S.I.L. [The Sunwar data cited is by Dora Bieri and Marlene Schulze.]
- Hodgson, Brian Houghton, 1880, *Miscellaneous essays relating to Indian subjects*. London: Trübner.
- Kainla, Bairagi, ed., 2059 V.S. [2002]. *Limbu-Nepali-English dictionary*. Kathmandu: Royal Nepal Academy.
- Lahaussais, Aimée, 2002, Aspects of the grammar of Thulung Rai: an endangered Himalayan language. PhD dissertation, University of California Berkeley.
- Matisoff, James A., 1997, *Sino-Tibetan numeral systems: prefixes, protoforms and problems*. *Pacific Linguistics* series B-114. Canberra: Australian National University
- Michailovsky, Boyd, 1975, Notes on the Kiranti verb. *Linguistics of the Tibeto-Burman Area* 2/2:183–218.
- 1994, Manner vs place of articulation in the Kiranti initial stops. In Kitamura, H. et al. eds *Current issues in Sino-Tibetan linguistics*, 766–772, 781. Osaka: National Museum of Ethnology.
- 2002, *Limbu-English dictionary of the Mewa Khola dialect*. Kathmandu. Mandala Book Point.
- n.d., Bahing, Bantawa, Hayu vocabularies. Mss.
- Opgenort, Jean-Robert, 2002, The Wāmbule language. PhD dissertation, University of Leiden
- Rai, Novel Kishor, 1985, A descriptive study of Bantawa. PhD dissertation, Deccan College, Pune.
- Rutgers, Roland, 1998, *Yamphu*. Leiden: CNWS.
- Shafer, Robert, 1955, Classification of the Sino-Tibetan languages. *Word* 11/1:94–111.
- Toba, Sueyoshi & Ingrid Toba, 1975, *A Khaling-English English-Khaling glossary*. Kirtipur, Nepal: S.I.L.
- Tolsma, Gerard, n.d. [?1999], Kulung dictionary. <http://www.starling.rinet.ru>.





# 16 *The impact of English loanwords on the Cantonese syllabary*

---

ROBERT S. BAUER

## 1 Introduction

No Chinese variety has had more intimate and longer contact with English than Cantonese. Their contact began just over 300 years ago when the early English-speaking traders arrived in Guangzhou to exchange silver for Chinese tea, porcelain, silk, and other goods. Today the effect of English influence on the Cantonese language shows up most clearly in the Hong Kong Cantonese lexicon in which hundreds of English loanwords occur (cf. Bauer and Benedict 1997:361–405). Some of these loanwords have been in use in Hong Kong Cantonese for such a long time that they have an official status and are written with Chinese characters, for example, *ba<sup>1</sup>-si<sup>6/3</sup>* ‘bus’ and *tek<sup>7A</sup>-si<sup>6/3</sup>* ‘taxi’;<sup>1</sup> further, many Cantonese speakers who do not speak English assume such words as these and others are ordinary Cantonese words.

In studying English loanwords in Hong Kong Cantonese we have several advantages over trying to identify the source languages of other kinds of loanwords (cf. Bauer 1996): first, the borrowing was not so long ago as to render the identity of the source words in English irrecoverable; second, the contact and borrowing are still ongoing with the objects

---

<sup>1</sup> The Cantonese lexical tone categories are indicated by raised numbers following the syllables as follows (the two-digit number following the tone contour name represents the start point and end point of the tone contour on the five-point scale): <sup>1</sup> High Level (55); <sup>2</sup> Mid-low Falling (21); <sup>3</sup> High Rising (25); <sup>4</sup> Mid-low Rising (23); <sup>5</sup> Mid Level (33); <sup>6</sup> Mid-low Level (22); <sup>7A</sup> High Stopped (55); <sup>7B</sup> Mid Stopped (33); <sup>8</sup> Mid-low Stopped (22). The morphological changed tone (known as *pinjam*) is marked by separating the basic tone from the changed tone which is usually tone 3 High Rising with the slash ‘/’; for example, *si<sup>6/3</sup>* indicates the basic tone is tone 6 Mid-low Level and the changed tone is tone 3 High Rising.

of our study in use right before our eyes (and ears) and so easily accessible; and third, the phonological systems of English and Cantonese are so markedly different and the loanwords so abundant that the phonetic adaptation of English loanwords into Cantonese can be described as a series of distinctive and principled correspondences.

One aspect of the borrowing of English words into Cantonese that has particularly interested me for some time has been the impact that the phonetic adaptation of English loanwords has had on the structure and development of the Cantonese syllabary. Structure refers to the phonetic shapes of syllables and the total number of them that occur in Cantonese; development recognises that the syllabary has been continuously evolving as new syllables are created and then added to it. The phrase 'new syllables' means just that: syllables that did not exist prior to the borrowing of the loanword have been constructed through the combination of existing initial consonants and rimes to form new syllables with which to represent the loanwords. Although there are constraints on the structure of syllables imposed by the phonotactics of the language, the Cantonese phonological system is still flexible and expansive enough to accommodate loanwords; it has no need to borrow any sounds from English it does not already have but has been able to handle loanwords through the recombination of its existing phonetic resources.

The value of studying the impact of English loanwords on the Cantonese syllabary is that it provides us with concrete evidence of language change that has resulted from language contact.

## 2 Cantonese syllabary

The first step in the identification of uniquely loanword syllables is to map out the Cantonese syllabary. The syllabary can be constructed by first listing the 20 initial consonants across the top of the page and the 56 rimes down the far left side. The resulting intersections of columns of initial consonants with rows of rimes form a matrix of syllables and yields a potential of 1,120 syllables (recognising the occurrence of the two nasal syllabics  $\_$  and  $\eta$  raises the possible number of syllables to 1,122). One can then check through the individual syllables to identify them as either occurring or non-occurring. Appendix 3 of Bauer and Benedict (1997:486–487) has been updated and reproduced in this paper as Appendix 2. The syllables in the syllabary are classified into four categories as follows:

- 1) Cantonese syllables that are associated with the standard Chinese characters (which are used to write the modern standard Chinese language) as their standard reading pronunciations. These syllables are unmarked in the syllabary.
- 2) Cantonese syllables that occur in the colloquial stratum of the lexicon and are not cognate with their semantic and functional equivalents in standard Chinese. These colloquial morphosyllables may not have standard Chinese characters associated

with them as their written forms; but if they do, the characters have been borrowed for their homophonous pronunciations. Many colloquial morphosyllables are represented by dialectal characters which have been especially created for this purpose. The colloquial syllables are marked with the superscript 'c'.

- 3) Syllables that only occur in the representation of English loanwords. These loanword syllables are marked with the superscript '+
- 4) The non-occurring syllables are marked with the superscript 'n

As far as I am aware, this syllabary is the only one that explicitly marks the non-occurring syllables.<sup>2</sup> The purpose in doing this is to try to determine if there is an underlying pattern of avoiding certain types of syllables. This idea is further considered in §4 below. Because syllables occurring in loanwords can be homophonous with syllables belonging to categories 1 and 2, the syllabary described here makes it possible to identify those syllables that uniquely occur in the representation of English loanwords.

### 3 English loanword syllables

In my first attempt (Bauer 1985) to produce a comprehensive Cantonese syllabary in which the loanword syllables were explicitly marked, a total of 30 loanword syllables were

---

<sup>2</sup> A number of Cantonese syllabaries have been published over the years, and the forms these syllabaries have taken have varied. One type of syllabary has simply listed the series of morphosyllables in phonetic transcription followed by standard Chinese characters which are pronounced with the morphosyllables, as in Yu 1982. A second type of syllabary has been based on the matrix of phonetically transcribed initial consonants either down or across one side of the page and the rimes in the corresponding direction; Chinese characters are written at the intersections of initials and rimes to mark the occurrence of morphosyllables, as in Kao (1971:177–184) and Yue-Hashimoto (1972:205–398). Rao et al (1981:296–302), however, indicated the occurrence of morphosyllables with a single or double slanted line. This second type of syllabary has typically recognised the standard and many colloquial morphosyllables but excluded those morphosyllables that occur only in English loanwords. In terms of its recognition of the first three kinds of occurring morphosyllables referred to above, Yue-Hashimoto's syllabary is by far the most comprehensive by listing all the standard and almost all the colloquial Chinese characters with which Cantonese morphosyllables are written; her syllabary also recognises some English loanwords. However, marking the occurrence of Cantonese morphosyllables in a syllabary with Chinese characters has necessarily limited the morphosyllables that can be recognised to those which are associated with Chinese characters: even if the morphosyllables do exist in the spoken language, they have no written forms and so are overlooked. The syllabary on the inside front cover of the index of standard Chinese characters in Cantonese romanisation compiled by the Linguistic Society of Hong Kong (Linguistic Society of Hong Kong 1997) has this limitation. In order to recognise the characterless morphosyllables, Kao (1971) marked them with the check mark. Yue-Hashimoto recorded English glosses enclosed within parentheses to indicate these syllables. In addition, the phrase 'taboo syllable' recognised morphosyllables regarded as vulgar and obscene (even though dialectal characters exist for writing these).

recognised. However, a recheck of these reduced the number to 26 because it was found that four syllables were actually homophonous with colloquial syllables. Twelve years later Bauer and Benedict (1997:410) identified 40 English loanword syllables. My most recent tally has now noted 49 such syllables and these have been listed and exemplified in Appendix 1. It is possible that my earlier research for identifying the loanword syllables may have undercounted some of them, and that the number of new syllables that entered Cantonese over the past two decades may be less than 23; nonetheless, I do believe the pattern of increase from 26 loanword syllables in 1985 to 40 in 1997 and to 49 in 2002 does point to a gradual expansion of the Cantonese syllabary as a result of the contact between Cantonese and English.

Represented among the set of 49 loanword syllables listed in Appendix 1 are 23 different rimes. When we look at these rimes within the framework of the Cantonese syllabary's four categories of syllables as described above, the most remarkable thing we discover is that there is nothing unusual about them. First, all these rimes belong to the set of 56 Cantonese rimes, so none are non-occurring. Second, they can be categorised into two sets: (1) rimes that occur in syllables that are associated with the standard Chinese characters: *-i, -æ, -u, -ɔ, -iw, -ow, -ej, -ɔj, -im, -in, -im, -it, -eŋ, -eŋ, -ek, -at, -ɔn, -ɔt*; and (2) rimes that occur in colloquial morphosyllables: *-ew, -em, -en, -ep, -et*. Admittedly, this is the only syllabary that has recognised all five of these colloquial rimes (the syllabary in Rao et al (1981:298) listed *-em, -ep, and -et*, but only syllable *kem* is marked as occurring). Appendix 4.1 in Bauer and Benedict (1997:488–496) listed lexical items from the colloquial lexicon (including both words and phrases) in which the five rimes occur.

#### 4 Conclusion

The phonetic structure of loanword syllables seems to be an extension of phonetic patterns observed in the occurrence of colloquial syllables. In looking at all four types of syllables, can we identify accidental and systematic gaps in the Cantonese syllabary? Do the non-occurring syllables establish a pattern that underlies Cantonese phonotactics in which certain types of syllables are disallowed? In examining the non-occurring syllables that are marked in Appendix 2, we observe that four labially articulated initial consonants have the following number of non-occurring syllables out of the 56 that are possible: *khw*-41, *kw*-35, *w*-27, *f*-23. These numbers may indicate a tendency to avoid syllables with labial initial consonants. Other non-occurring syllables seem to suggest a tendency to avoid syllables that have both labially-articulated initial and final consonants, yet we must also note the occurrence of *piw, phiw, miw, pɛw, phɛw, mɛw, paw, phaw, maw* among others as readings of standard Chinese characters, and *pɛm, pɛp, phɛp, mɛm, and wiw* in the colloquial lexicon. There may also be a tendency to avoid syllables in which all three elements, including the nuclear vowel, are labially-articulated; but again we find that *pow*,

*phow* and *mow* are readings of standard Chinese characters, while *wow* occurs in the colloquial lexicon and *fow* in an English loanword. For any of these patterns that appear to underlie the labial dissimilation constraint that has been posited for Cantonese (Yue-Hashimoto 1972:139; Light 1977:79), counterexamples occur among the syllables associated with standard Chinese characters, syllables from the colloquial lexicon, and those that occur in loanwords. In view of the complex distribution of syllables within the syllabary and the trend for the number of loanword syllables to increase, the matters of identifying the accidental and systematic gaps in the Cantonese syllabary and mapping out Cantonese phonotactics still await a satisfying resolution.

## References

- Bauer, Robert S., 1985, The expanding syllabary of Hong Kong Cantonese. *Cahiers de Linguistique Asie Orientale* XIV/1:99–111.
- 1994, Fieldnotes on Hong Kong Cantonese. University of Hong Kong.
- 1996, Identifying the Tai substratum in Cantonese. *Proceedings of the fourth international symposium on languages and linguistics, Pan-Asiatic linguistics V 1806–1844*. Bangkok: Institute of Language and Culture for Rural Development, Mahidol University, Salaya.
- 1997, Fieldnotes on Hong Kong Cantonese. Hong Kong Polytechnic University.
- 2001, Fieldnotes on Hong Kong Cantonese. Hong Kong Polytechnic University.
- Bauer, Robert S. and Paul K. Benedict, 1997, *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.
- Chan, Mimi & Helen Kwok, 1982, *A study of lexical borrowing from English in Hong Kong Chinese*. Hong Kong: Centre of Asian Studies, University of Hong Kong.
- Cheung Kwan-hin, 1986, The phonology of present-day Cantonese. Ph.D. dissertation, University College, London.
- Huang Xiling, 1986, *Yueyin yunhui, a Chinese syllabary pronounced according to the dialect of Canton*. Hong Kong: Zhonghua Shuju.
- Kao, Diana L., 1971, *Structure of the syllable in Cantonese*. The Hague: Mouton.
- Kiu, K.L., 1977, Tonal rules for English loan words in Cantonese. *Journal of the International Phonetic Association* 7/1:17–22.
- Light, Timothy, 1977, The Cantonese final: an exercise in indigenous analysis. *Journal of Chinese Linguistics* 5/1:75–102.
- Rao Bingcai, Ouyang Jueya and Zhou Wuji, 1981, *Guangzhouhua fangyan cidian* [dictionary of Cantonese dialect]. Hong Kong: Shangwu Yinshuguan.

- Xianggang Yuyanxue Xuehui, 1997, *Yueyu pinyin zibiao* [table of Chinese characters in Cantonese phonetic transcription]. Hong Kong: Xianggang Yuyanxue Xuehui.
- Yu Bingzhao, 1982, *Tongyin zihui* [syllabary of homophones (in Cantonese phonetic transcription)]. Hongkong: New Asia Press.
- Yue-Hashimoto and Ann Oi-kan, 1972, *Studies in Yue dialects 1: phonology of Cantonese*. Cambridge: Cambridge University Press.
- Zhang Hongnian, 1972, *Xianggang Yueyufa yanjiu* [study of Hong Kong Cantonese grammar]. Hong Kong: Chinese University of Hong Kong Press.
- Zhang Liyan and Ni Liehuai, 1999, *Gangshi Guangzhouhua cidian* [Hong Kong Cantonese dictionary]. Hong Kong: Wanli Jigou.
- Zhang Liyan, 1983, Guangzhouhua you yin wu zi yinjie zongji [tally of syllables with sound but no written form in Cantonese]. *Yuwen Zazhi* 10:28–35.
- Zhang Risheng, 1986, Xianggang Guangzhouhua Yingyu yiyin jieci de shengdiao guiyu [tonal rules for transliterating English loanwords in Hong Kong Cantonese]. *Zhongguo Yuwen* 1:42–50.

## Appendix 1: 49 "new" Cantonese syllables created by borrowing English words.

<i>pɔːŋ</i> 'boy'	(Bauer 1985: 101)
<i>pɛːn</i> 'band'	(Zhang 1972: 221)
<i>phæː</i> in <i>phæː-l-sɛːn</i> 'percent'	(Rao et al. 1981: 179)
<i>phaːt</i> in <i>phaːt-naː</i> 'partner'	(Zhang 1972: 221)
<i>phɔːt</i> in <i>phaː-l-siː-phɔːt</i> 'passport'	(Zhang 1972: 225)
<i>tɛːn</i> in <i>khɔːn-tɛːn-saː</i> 'condenser'	(Zhang 1972: 223)
<i>tɔːn</i> in <i>aː-tɔːn</i> 'Don'	(Bauer 1994)
<i>thiː</i> in <i>thiː-sət</i> 'T-shirt'	(Zhang 1972: 220)
<i>thej</i> in <i>thej-l-siː</i> 'taste'	(Bauer 2001)
<i>theːn</i> in <i>theːn-niː-siː</i> 'tennis'	(Zhang 1986: 48)
<i>kɛːm</i> 'game'	(Zhang 1972: 222)
<i>khɔː</i> 'call'	(Bauer 1985: 101)
<i>khɔːn</i> 'coin'	(Bauer 1985: 101)
<i>khiːp</i> in <i>khiːp-fiːt</i> 'keep fit'	(Bauer 1985: 101)
<i>kheːp</i> in <i>kheːp-thɔːn</i> 'captain'	(Zhang 1972: 221)
<i>khaː</i> 'card'	(Yue-Hashimoto 1972: 329)
<i>khɔːt</i> 'court'	(Bauer 1994)
<i>khwiːm</i> 'cream'	(Cathy Wong pc 4/12/97)
<i>khwiːn</i> 'queen'	(Chan & Kwok 1982: 114)
<i>khwiːt</i> 'quit'	(Bauer 1997)
<i>muː</i> in <i>muː-fiː</i> 'movie'	(Zhang 1972: 224)
<i>mɛːn</i> MAN 'manly'	<dung1 jing1 zai2> March 1997, #45:22 (adultcomic book)
<i>mɔːn</i> 'monitor (for computer)'	(Cheung Kwan-hin pc 1997)
<i>mɛːk</i> in <i>mɛːk-kheːn</i> 'mechanical'	(Kiu 1977: 19)
<i>fiːw</i> in <i>fiːw-l-siː</i> 'fuse'	(Zhang 1972: 217)
<i>fow</i> in <i>siː-fow</i> 'civil'	(Bauer & Benedict 1997: 365)
<i>feːn</i> 'friend'	(Zhang 1986: 44)
<i>foː</i> 'volt'	(Zhang 1972: 227)
<i>fɛːk</i> in <i>fɛːk-l-siː</i> 'fax'	(Zhang & Ni 1999: 393)
<i>sɛːw</i> in <i>sɛːw-l-siː</i> 'sales'	(Bauer 1997)
<i>sɛːm</i> 'semester'	(Bauer 1997)
<i>sɛːn</i> in <i>phæː-l-sɛːn</i> 'percent'	(Rao et al 1981: 179)
<i>heːp</i> in <i>heːp-phiː</i> 'happy'	(Zhang & Ni 1999: 393)
<i>tsej</i> in <i>tiː-tsej</i> 'DJ (disk jockey)'	(Zhang 1986: 45)
<i>tseːm</i> 'jam'	(Zhang 1972: 217)
<i>tɔːn</i> in <i>aː-tɔːn</i> 'John'	(Bauer 1994)
<i>wiːn</i> in <i>wiːn-naː</i> 'winner'	(Zhang 1972: 222)
<i>wɛːn</i> in <i>wɛːn-tsej</i> 'van'	(Cheung 1986: 33)
<i>wɔːn</i> in <i>wɔːn-lɛː-laː</i> 'vanilla'	(Zhang 1986: 49)
<i>leːp</i> in <i>leːp-pɔː</i> 'net-ball (in table tennis)'	(Cheung Kwan-hin pc 1997)
<i>jeːw</i> 'yell at'	(Cheung Pak-man pc 6/3/00)
<i>jeːn</i> '(Japanese) Yen'	(Cheung 1986: 33)
<i>juː</i> 'U' (short form for <i>university</i> )	(Zhang & Ni 1999: 396)
<i>iːn</i> in <i>iːn-tɛi</i> 'inch'	(Zhang 1986: 46)
<i>ɛŋ</i> in <i>ɛŋ-liː</i> 'Eng. Lit.'	(Zhang 1986: 48)
<i>ɛːm</i> in <i>ɛːm-l-siː</i> 'MC'	(Zhang 1986: 45)
<i>ɛːn</i> in <i>ɛːn-tɛi</i> 'engine'	(Zhang 1972: 224);
<i>ɛːn-l-khɔː</i> 'encore'	(Zhang & Ni 1999: 392);
<i>ɛːn-tɛi</i> N.G. 'no good'	(Bauer & Benedict 1997: 405)
<i>ɛːŋ</i> in <i>waːŋ-ɛŋ-kow</i> 'wide-angle'	(Zhang 1972: 225)
<i>ɛːk</i> in <i>ɛːk-l-siː-kwɔːŋ</i> 'X-ray'	(Zhang 1972: 213)



Appendix 2: Cantonese syllabary.

No mark = syllable associated with standard character; ° = colloquial; + = in English loanword; ° = non-occurring.

	P-	ph-	t-	th-	k-	kh-	kw-	khw-	m-	n-	ŋ-	f-	s-/ç-	h-	ts-/tç-	tsh-/tç-	w-	l-	j-	-
-i	°pi	°phi	°ti	°thi	°ki	°khi	°kwi	°khwī	°mi	°ni	°ŋi	°fi	°si	°hi	°tci	°tchi	°wi	°li	°ji	°i
-iw	°piw	°phiw	°tiw	°thiw	°kiw	°khiw	°kwiw	°khwīw	°miw	°niw	°ŋiw	+°fiw	°siw	°hiw	°tciw	°tchiw	°wiw	°liw	°jiw	°iw
-im	°pim	°phim	°tim	°thim	°kim	°khim	°kwim	+°khwim	°mim	°nim	°ŋim	°fim	°sim	°him	°tcim	°tchim	°wim	°lim	°jim	°im
-in	°pin	°phin	°tin	°thin	°kin	°khin	+°kwīn	+°khwīn	°min	°nin	°ŋin	°fin	°sin	°hin	°tcin	°tchin	+°win	°lin	°jin	+°in
-ip	°pip	°phip	°tip	°thip	°kip	°khip	°kwip	+°khwip	°mip	°nip	°ŋip	°fip	°sip	°hip	°tcip	°tchip	°wip	°lip	°jip	°ip
-it	°pit	°phit	°tit	°thit	°kit	°khit	°kwit	+°khwit	°mit	°nit	°ŋit	°fit	°sit	°hit	°tcit	°tchit	°wit	°lit	°jit	°it
-y	°py	°phy	°ty	°thy	°ky	°khy	°kwy	°khwīy	°my	°ny	°ŋy	°fy	°sy	°hy	°tcy	°tchy	°wy	°ly	°jy	°y
-yn	°pyn	°phyn	°tyn	°thyn	°kyn	°khyñ	°kwyn	°khwīyn	°myñ	°nyñ	°ŋyn	°fyn	°syn	°hyñ	°tcyn	°tchyn	°wyn	°lyn	°jyn	°yn
-yt	°pyt	°phyt	°tyt	°thyt	°kyt	°khyt	°kwyt	°khwīyt	°myt	°nyt	°ŋyt	°fyt	°syñ	°hyt	°tcyt	°tchyt	°wyt	°lyt	°jyt	°yt
-ej	°pej	°phej	°tej	°thej	°kej	°khej	°kwej	°khwēj	°mej	°nej	°ŋej	°fej	°sej	°hej	+°tsej	+°tshēj	°wej	°lej	°jej	°ej
-ej	°pej	°phej	°tej	°thej	°kej	°khej	°kwej	°khwēj	°mej	°nej	°ŋej	°fej	°sej	°hej	+°tsej	+°tshēj	°wej	°lej	°jej	°ej
-ek	°pek	°phek	°tek	°thek	°kek	°khek	°kwek	°khwēk	°mek	°nek	°ŋek	°fek	°sek	°hek	°tsek	°tshək	°wek	°lek	°jek	°ek
-ε	°pe	°phe	°te	°the	°ke	°khe	°kwe	°khwē	°me	°ne	°ŋe	°fe	°se	°he	°tse	°tshē	°we	°le	°je	°ε
-ew	°pew	°phew	°tew	°thew	°kew	°khew	°kwew	°khwēw	°mew	°new	°ŋew	+°few	+°sew	°hew	°tsew	°tshēw	°wew	°lew	°jew	°ew
-em	°pem	°phem	°tem	°them	°kem	°khem	°kwem	°khwēm	°mem	°nem	°ŋem	°fem	+°sem	°hem	+°tsem	+°tshēm	°wem	°lem	°jem	+°em
-en	+°pen	+°phen	+°ten	+°then	+°ken	+°khen	+°kwen	+°khwēn	+°men	+°nen	+°ŋen	+°fen	+°sen	+°hen	+°tseñ	+°tshēñ	+°wen	+°len	+°jen	+°en
-ej	°pej	°phej	°tej	°thej	°kej	°khej	°kwej	°khwēj	°mej	°nej	°ŋej	°fej	°sej	°hej	°tsej	°tshēj	°wej	°lej	°jej	°ej
-ep	°pep	°phep	°tep	°thep	°kep	°khep	°kwep	°khwēp	°mep	°nep	°ŋep	°fep	°sep	°hep	°tsep	°tshēp	°wep	°lep	°jep	°ep
-ei	°pei	°phei	°tei	°thei	°kei	°khei	°kwei	°khwēi	°mei	°nei	°ŋei	°fei	°sei	°hei	°tsei	°tshēi	°wei	°lei	°jei	°ei
-ek	°pek	°phek	°tek	°thek	°kek	°khek	°kwek	°khwēk	+°mek	°nek	°ŋek	+°fek	°sek	°hek	°tsek	°tshək	°wek	°lek	°jek	+°ek
-æ	°pæ	°phæ	°tæ	°thæ	°kæ	°khæ	°kwæ	°khwæ	°mæ	°næ	°ŋæ	°fæ	°sæ	°hæ	°tçæ	°tchæ	°wæ	°læ	°jæ	°æ
-æj	°pæj	°phæj	°tæj	°thæj	°kæj	°khæj	°kwæj	°khwæj	°mæj	°næj	°ŋæj	°fæj	°sæj	°hæj	°tçæj	°tchæj	°wæj	°læj	°jæj	°æj
-æk	°pæk	°phæk	°tæk	°thæk	°kæk	°khæk	°kwæk	°khwæk	°mæk	°næk	°ŋæk	°fæk	°sæk	°hæk	°tçæk	°tchæk	°wæk	°læk	°jæk	°æk
-æj	°pæj	°phæj	°tæj	°thæj	°kæj	°khæj	°kwæj	°khwæj	°mæj	°næj	°ŋæj	°fæj	°sæj	°hæj	°tçæj	°tchæj	°wæj	°læj	°jæj	°æj
-æn	°pen	°phen	°ten	°then	°ken	°khen	°kwen	°khwēn	°men	°nen	°ŋen	°fen	°sen	°hen	°tseñ	°tshēñ	°wen	°len	°jen	°en
-æt	°pet	°phei	°tet	°thei	°ket	°khei	°kwei	°khwēi	°met	°nei	°ŋei	°fei	°sei	°hei	°tsei	°tshēi	°wei	°lei	°jei	°æt
-ej	°pej	°phej	°tej	°thej	°kej	°khej	°kwej	°khwēj	°mej	°nej	°ŋej	°fej	°sej	°hej	°tsej	°tshēj	°wej	°lej	°jej	°ej
-ew	°pew	°phew	°tew	°thew	°kew	°khew	°kwew	°khwēw	°mew	°new	°ŋew	°few	°sew	°hew	°tsew	°tshēw	°wew	°lew	°jew	°ew

Appendix 2: Cantonese syllabary (continued).

	p-	ph-	t-	th-	k-	kh-	kw-	khw-	m-	n-	ŋ	f-	s-	h-	ts-	tsb-	w-	l-	j-	ø-
-am	<sup>0</sup> phem	<sup>0</sup> phem	<sup>0</sup> tam	<sup>0</sup> them	ken	khem	<sup>0</sup> kwem	<sup>0</sup> khwem	<sup>0</sup> mam	<sup>0</sup> nem	<sup>0</sup> ŋem	<sup>0</sup> fem	sem	hem	tsem	tshem	wem	lem	jem	em
-en	pen	phen	ten	then	ken	khen	<sup>0</sup> kwen	<sup>0</sup> khwen	men	nen	ŋen	fen	sen	hen	tzen	tshen	wen	len	jen	en
-ej	pej	phej	tej	thej	kej	khej	<sup>0</sup> kwaj	<sup>0</sup> khwej	mej	nej	ŋej	faj	sej	hej	tzej	tshaj	waj	lej	jej	ej
-ep	pep	phep	<sup>0</sup> tep	<sup>0</sup> thep	kep	kehp	<sup>0</sup> kwep	<sup>0</sup> khwep	<sup>0</sup> mep	<sup>0</sup> nep	<sup>0</sup> ŋep	<sup>0</sup> fep	sep	hep	tsep	tshap	<sup>0</sup> wep	lep	jep	ep
-et	pet	phet	ret	<sup>0</sup> thet	ket	kheth	<sup>0</sup> kwet	<sup>0</sup> khwet	met	net	ŋet	fet	set	het	tset	tshet	wet	let	jet	et
-ek	pek	phek	tek	<sup>0</sup> thek	<sup>0</sup> kək	khək	<sup>0</sup> kwək	<sup>0</sup> khwək	mek	nek	ŋək	fək	sek	hek	tsek	tshək	<sup>0</sup> wək	tək	jek	ək
-aj	pa	pha	ta	tha	ka	kha	<sup>0</sup> kwa	<sup>0</sup> khwa	ma	na	ŋa	fa	sa	ha	tza	tsha	wa	la	ja	a
-aj	paj	phaj	taj	thaj	kaj	khaij	<sup>0</sup> kwaj	<sup>0</sup> khwaij	maj	naj	ŋaj	faj	saj	haj	tzej	tshaj	waj	lej	fej	aj
-aw	paw	phaw	<sup>0</sup> taw	<sup>0</sup> thaw	kaw	khaw	<sup>0</sup> kwaw	<sup>0</sup> khwaw	maw	naw	ŋaw	<sup>0</sup> faw	saw	haw	tsw	tshaw	<sup>0</sup> waw	law	jaw	aw
-am	pam	pham	tam	tham	kam	khām	<sup>0</sup> kwaŋ	<sup>0</sup> khwaŋ	mam	nam	ŋam	fam	sam	ham	tsam	tsham	wam	lam	jam	am
-an	pan	phan	tan	than	kan	khan	<sup>0</sup> kwān	<sup>0</sup> khwān	man	nan	ŋan	fan	san	han	tsan	tshan	wan	lan	jan	an
-aj	paj	phaj	taj	thaj	kaj	khaj	<sup>0</sup> kwaj	<sup>0</sup> khwaj	maj	naj	ŋaj	faj	saj	haj	tsaj	tshaj	waj	laj	jaj	aj
-ap	pap	phap	tap	thap	kap	khap	<sup>0</sup> kwap	<sup>0</sup> khwap	map	nap	ŋap	fap	sap	hap	tsap	tshap	wap	lap	jap	ap
-at	pat	phat	tat	thāt	kat	khāt	<sup>0</sup> kwat	<sup>0</sup> khwat	mat	nat	ŋat	fat	sat	hat	tsat	tshat	wat	lat	jat	at
-ak	pak	phak	tak	thak	kak	khak	<sup>0</sup> kwak	<sup>0</sup> khwak	mak	nak	ŋak	fak	sak	hak	tsak	tshak	wak	lak	jak	ak
-u	pu	phu	tu	thu	ku	khu	<sup>0</sup> kwu	<sup>0</sup> khwu	+ mu	nu	ŋu	fu	su	hu	tsu	tshu	wu	lu	ju	u
-uj	puj	phuj	tuj	thuj	kuj	khuj	<sup>0</sup> kwuj	<sup>0</sup> khwuj	muj	naj	ŋuj	fuj	suj	huj	tsuj	tshuj	wuj	luj	juj	uj
-un	pun	phun	tun	thun	kun	khun	<sup>0</sup> kwun	<sup>0</sup> khwun	mun	nun	ŋun	fun	sun	hun	tsun	tshun	wun	lun	jun	un
-ut	put	phut	tut	thut	kut	khut	<sup>0</sup> kwut	<sup>0</sup> khwut	mut	nut	ŋut	fut	sut	hut	tsut	tshut	wut	lut	jut	ut
-ek	pek	phek	tek	thek	kek	khek	<sup>0</sup> kwək	<sup>0</sup> khwək	mek	nek	ŋək	fək	sək	hek	tsek	tshək	wək	lək	jək	ek
-ow	pow	phow	tow	thow	kow	khow	<sup>0</sup> kwow	<sup>0</sup> khwow	mow	now	ŋow	fow	sow	how	tsow	tshow	wow	low	jow	ow
-oj	poj	phoj	toj	thoj	koj	khoj	<sup>0</sup> kwōj	<sup>0</sup> khwoj	moj	noj	ŋoj	foj	soj	hoj	tsōj	tshōj	woj	loj	jōj	oj
-ok	pok	phok	tok	thok	kok	khok	<sup>0</sup> kwok	<sup>0</sup> khwok	mok	nok	ŋok	fok	sok	hok	tsok	tshok	wok	lok	jok	ok
-o	po	pho	to	tho	ko	kho	<sup>0</sup> kwō	<sup>0</sup> khwo	mo	no	ŋo	fo	so	ho	tso	tsho	wo	lo	jo	o
-oj	poj	phoj	toj	thoj	koj	khoj	<sup>0</sup> kwōj	<sup>0</sup> khwoj	moj	noj	ŋoj	foj	soj	hoj	tsōj	tshōj	woj	loj	jōj	oj
-on	pon	phon	ton	thon	kon	khon	<sup>0</sup> kwon	<sup>0</sup> khwon	mon	non	ŋon	fon	son	hon	tsōn	tshōn	won	lon	jon	on
-oj	poj	phoj	toj	thoj	koj	khoj	<sup>0</sup> kwōj	<sup>0</sup> khwoj	moj	noj	ŋoj	foj	soj	hoj	tsōj	tshōj	woj	loj	jōj	oj
-ot	pot	phot	tot	thot	kot	khot	<sup>0</sup> kwot	<sup>0</sup> khwot	mot	not	ŋot	fōt	sōt	hōt	tsot	tshot	wōt	lōt	jōt	ot
-ok	pok	phok	tok	thok	kok	khok	<sup>0</sup> kwək	<sup>0</sup> khwok	mok	nek	ŋək	fək	sək	hək	tsək	tshək	wək	lək	jək	ok



# 17 *On borrowing from Middle Chinese into Proto Tibetan: a new look at the problem of the relationship between Chinese and Tibetan*<sup>1</sup>

MICHEL FERLUS

## 1 Introduction

It is well known that Chinese and Tibetan are genetically related. But it has been asserted by some scholars that an important part of the supposed common vocabulary represents an ancient layer of borrowings from Middle Chinese into Ancient Tibetan. It is important to separate the loanwords from inherited vocabulary in order to evaluate the closeness of the genetic relationship.

The study of borrowings depends on the relationship between the languages in contact. When the languages are not genetically related, borrowing usually concerns whole words, and it is in general relatively simple to sort out loanwords. But when genetically related languages are in contact, with some degree of intercomprehension, or at least a certain awareness of correspondences, and in a hierarchical relationship of prestige, more complex types of influence can occur. These may affect only one segment in a word, one constituent (consonant, vowel or rhyme) in a syllable, or one syllable in a dissyllabic word. Such phenomena have been observed by the author in situations of linguistic contact between Vietnamese and related languages within the Vietic linguistic group (Ferlus 1991, 1995,

<sup>1</sup> The following abbreviations are used:

- MC Middle Chinese (Karlgren's Ancient Chinese): the stage of the *Qie<sup>4</sup> Yun<sup>4</sup>* reflected in the *Yun<sup>4</sup> Jing<sup>4</sup>*.
- OC Old Chinese (Karlgren's Archaic Chinese): the stage of the rhymes of the *Shi<sup>1</sup> Jing<sup>1</sup>*.
- WT Written Tibetan.
- PT Proto-Tibetan (The stage just before Middle Chinese influence).
- PST Proto-Sino-Tibetan (in a restricted sense).
- TB Tibeto-Burman.

2001). Meillet appealed to a phenomenon of this kind, under conditions of bilingualism, to explain the *h* in French *haut* (Germanic *hoch*, Latin *altus*) (Meillet 1936:99–103).

These phenomena will be illustrated here by borrowings, or partial borrowings, from Middle Chinese (MC) into Proto Tibetan (PT), the supposed stage of the language immediately preceding MC influence. The Proto Tibetan forms have been reconstructed by the author for the needs of the present article. This Proto Tibetan could just as well be called pre-Old Tibetan. The result of the Chinese influence is reflected in Old Tibetan and recorded in Written Tibetan (WT). It can be asserted that a part of PT vocabulary remained relatively close to Proto Sino-Tibetan (PST).

## 2 A theory of monosyllabisation from OC to MC

Before proceeding further, it is necessary to review the author's theory of the phenomenon of monosyllabisation that occurred between OC and MC (Ferlus 1998). This theory is used in the explanation of the influence of MC on Tibetan that follows.

Old Chinese was a disyllabic language, in the sense that while part of the vocabulary was monosyllabic, another part contained disyllabic words, more precisely of the sesquisyllabic type (as defined by Matisoff). This type is still widely represented in many Austroasiatic languages of Southeast Asia. A sesquisyllable is a type of disyllable composed of a main syllable preceded by a presyllable. The main syllable is similar to a monosyllabic word, while the presyllable is a reduced and unstressed syllable in which vocalic oppositions are neutralised. The presyllable can be a morphological prefix or a neutral element without any meaning.

monosyllable: CV(C)  
sesquisyllable: C-CV(C)

According to my theory, OC sesquisyllables developed phonetic tenseness (T) while monosyllables developed laxness (L). Then, when sesquisyllabic words became monosyllabic by the loss of the presyllable, the earlier contrast of syllabic type, between C-CV(C) and CV(C), was replaced by the new contrast of tense vs lax (T/L). This phenomenon was associated with a vocalic split, with vowel lowering in T syllables and vowel raising in L syllables. Later, in a second step after these changes, the lenition of medial *-r-* further blurred the situation. This is the stage of MC characterised by the well-known system of four divisions: the T syllables belong to Division I/IV (syllables without medial *-r-* in OC) or to Division II (medial *-r-* in OC), while the L syllables belong to Division III (with or without medial *-r-* in OC), characterised by the famous yod of Karlgren's (1957) reconstructions.

Old Chinese (OC)		Middle Chinese (MC)	divisions
C-CV(C) ( <i>tenseness</i> )	>	CV(C) / T	( <i>vowel lowering</i> ) I/IV ( <i>-r</i> ) or II ( <i>+r</i> )
CV(C) ( <i>laxness</i> )	>	CV(C) / L	( <i>vowel raising</i> ) III ( <i>±r</i> )

In my system, the symbol [°] is the mark of Division III. It indicates a raising and centralisation of the vowel associated with (what I suspect to be) breathy voice. The symbol ['] is the mark of Division II. It indicates the result of the lenition of OC medial *-r-*, probably a kind of velar spirant. Thus Division II is simply an offshoot of Division I (and IV, below); together these three divisions continue the old T category. In L syllables, the softened OC medial *-r-* became obscured by the breathiness of the vowel and was lost; there was no split

analogous to the one that gave rise to Division II. Thus the entire L category is continued intact by Division III. No special mark characterises Division I or IV. Division IV is in complementary distribution with Division I, apparently a device to represent the single MC front diphthong. This theory is summarised in Table 1.

**Table 1:** Proposed origins of MC divisions

syllable-type	OC medial (Baxter)	MC division
(without medial -r-)		
tense	∅	I/IV / T
lax	-j-	III / L
(with medial -r-)		
tense	-r-	II / T(r)
lax	-rj-	III / L(r)

The examples in Table 2 are taken from Baxter (1992). I have added my own phonetic interpretation between square brackets.

**Table 2:** Diachronic examples of syllable-types with phonetic interpretation

T/L	div.	Man.	MC	OC	
T	I	納	<i>na<sup>4</sup></i>	nop [nɒp]	*nup [T(C-)nup] 'send in' (695h)
L	III	入	<i>ru<sup>4</sup></i>	nyip [n <sup>ə</sup> ip]	*n-j-up [L-nup] 'enter' (695a)
T	IV	銘	<i>ming<sup>2</sup></i>	meng [mieŋ]	*meng [T(C-)meŋ] 'inscription' (826d)
L	III(>IV)	戍	<i>ming<sup>2</sup></i>	mjieng [m <sup>ə</sup> jeŋ]	*m-j-eng [L-meŋ] 'name' (826a)
T	I	股	<i>gu<sup>3</sup></i>	ku <sup>X</sup> [kɔ <sup>?</sup> ]	*ka <sup>?</sup> [T(C-)ka <sup>?</sup> ] 'thigh' (51a)
T(r)	II	假	<i>jiā<sup>3</sup></i>	kæ <sup>X</sup> [k <sup>ɛ</sup> æ <sup>?</sup> ]	*k-r-a <sup>?</sup> [T(C-)kra <sup>?</sup> ] 'false, simulate' (33c)
L(r)	III	莒	<i>ju<sup>3</sup></i>	kjo <sup>X</sup> [k <sup>ə</sup> ʌ <sup>?</sup> ]	*k-rj-a <sup>?</sup> [Lkra <sup>?</sup> ] 'round basket' (76j)

Before the complete, structural monosyllabisation that affected the whole sesquisyllabic vocabulary, there may have existed a slower process of random monosyllabisation affecting individual words.

### 3 The mode of borrowing from Middle Chinese into Proto Tibetan

Language A (here MC) is in a dominating position with a genetically related language B (here PT). Language A is regarded as prestigious by speakers of B who, by a kind of affectation, are led to imitate some characteristic sounds of A unknown in B. This results in a phonetic compromise, a segment of an A word being borrowed and substituted for the corresponding segment of the cognate B word. The sounds of MC that did not exist in PT are the segments (rhymes or main syllables) that characterise Division III (i.e. the presumed breathiness marked by [\$]) and Division II (i.e. the spirantised velar sound marked by [f]).

Speakers of PT tried, unconsciously or not, to imitate these unfamiliar sounds which were felt to carry with them the prestige of the dominating language. But in contrast to the usual process, in which whole words are borrowed, only the the characteristic MC segments of cognate words were borrowed by speakers of PT.

**Table 3:** Chinese and Tibeto-Burman numerals

	Karlgren 1957	Coblin 1986	Coblin 1986	Coblin 1986	Benedict 1972	Pulleyblank 1991	
	Ar C > An C	PST	OC > MC	TB	TB	EMC	
1	一 <i>zi</i> <sup>1</sup>	--- > tsjäk	gtyik	tjik > tsjäk	g-tyik	t(y)ik	tɕiajk
2	二 <i>er</i> <sup>4</sup>	ńjər > ńzi-	gnyis	njidh > ńzi	gnyis	g-nis	ɲi <sup>h</sup>
3	三 <i>san</i> <sup>1</sup>	səm > sâm	gsum	səm > sâm	g-sum	g-sum	sam
4	四 <i>si</i> <sup>4</sup>	sjəd > si-	btyid	sjidh > si-	blyiy	b-liy	si <sup>h</sup>
5	五 <i>wu</i> <sup>3</sup>	ngo > nguo:	lngay	ngagx > ngwo:	l-nga	l-ŋa	ŋɔ <sup>?</sup>
6	六 <i>liu</i> <sup>4</sup>	ljök > ljuk	dljəkw	ljəkw > ljuk	d-ruk	d-ruk	luwk
7	七 <i>qi</i> <sup>1</sup>	ts'jēt > ts'jēt	shnjis	tshjit > tshjet	s-nis	s-nis	tsh'it
8	八 <i>ba</i> <sup>1</sup>	pwät > pwät	priat	priat > pwät	pryat	b-r-gyat	pəət/pə:t
9	九 <i>jiu</i> <sup>3</sup>	kjüg > kjəu:	dkwjəyw	kjəgwɔx > kjəu:	d-kuw	d-kuw	kuw <sup>?</sup>
10	十 <i>shi</i> <sup>2</sup>	đjəp > źjəp	gřip	djəp > źjəp	gip	gip	dzip
100	百 <i>bai</i> <sup>3</sup>	päk > pok	pria <sup>?</sup>	prak > pek	prya	r-gya	pa:jk/pə:jk

**Table 3:** (continued)

Baxter 1992	Ferlus 1998	Sagart 1999	Matisoff 1997	WT
OC > MC	OC > MC	OC > MC	PST	
(tjek > tsyek)	Ltek > tɕ <sup>ə</sup> ek	-----	g-t(y)i-k ɤ tya-k	<i>gcig</i> 1
(njits > n yij <sup>H</sup> )	Lnits > ɲ <sup>ə</sup> ij <sup>h</sup>	bni[jt]-s > nyij <sup>H</sup>	g-ni-s/k	<i>gnyis</i> 2
sum > [sam]	Tksum > [sam]	a <sup>s</sup> -hlɪm > sam	g-sum	<i>gsum</i> 3
s(p)jij/ts > sij <sup>H</sup>	Lslits > s <sup>ə</sup> ij <sup>h</sup>	b <sup>s</sup> -hli[j]-s > sij <sup>H</sup>	b-liy = b-ləy	<i>bzhi</i> 4
nga <sup>?</sup> > ngu <sup>X</sup>	Tlŋa <sup>?</sup> > ŋɔ <sup>?</sup>	aŋa <sup>?</sup> > ngu <sup>X</sup>	l-ŋa ɤ b-ŋa	<i>lŋa</i> 5
C-rjuk > ljuwk	Lruk > l <sup>ə</sup> uwk	BCə-ruk > ljuwk	d-ruk / d-k-rok	<i>drug</i> 6
thsjit > tshit	Ltsh'it > tsh <sup>ə</sup> it	b <sup>s</sup> -hnit > tshit	s-nis	( <i>bdun</i> ) 7
(pret > pət)	Tpret > p <sup>ɪ</sup> ət	a <sup>pr</sup> [e]t > peat	b-r-gyat ɤ b-g-ryat	<i>brgyad</i> 8
k <sup>w</sup> ju <sup>?</sup> > kjuw <sup>X</sup>	Lk <sup>w</sup> u <sup>?</sup> > k <sup>ə</sup> uw <sup>?</sup>	bku <sup>?</sup> > kjuw <sup>X</sup>	d-kəw ɤ s-gəw ɤ d-gaw	<i>dgu</i> 9
gijp > dzyip	Lgip > dz <sup>ə</sup> ip	b <sup>g</sup> ip > dzyip	gip ɤ gyap	<i>bcu</i> 10
prak > pæk	Tprak > p <sup>ɪ</sup> æk	-----	b-r-gya ɤ b-g-rya	<i>brgya</i> 100

This process will be illustrated first in the numerals. Table 3 shows the principal reconstructions and interpretations of the set of numerals 'one' to 'ten' and 'hundred': Archaic Chinese and Ancient Chinese (Karlgren 1957), their equivalents Old Chinese and

Middle Chinese (Coblin 1986, Baxter 1992, Ferlus 1998, Sagart 1999), Early Middle Chinese (Pulleyblank 1991), Proto Sino-Tibetan (Coblin 1986, Matisoff 1997), and Tibeto-Burman (Benedict 1972, Coblin 1986).

#### 4 Comparison of Tibetan and Chinese numerals

Table 4 is the reference chart for the following discussion. The reconstructions used here are by the author. For OC and MC they are based on Baxter 1992. For PT and PST they have been elaborated for the needs of the present article.

Table 4: Tibetan numerals

			OC > MC	PT + MC	hypothetic	WT
1	雙	<i>zhi</i> <sup>1</sup>	Ltek > tɕ <sup>ə</sup> ek	k[tek] + tɕ <sup>ə</sup> ek	> ktɕ <sup>ə</sup> ek	<i>gcig</i>
2	二	<i>er</i> <sup>4</sup>	Lmits > n̥ <sup>ə</sup> ijh	k[nits] + n̥ <sup>ə</sup> ijh	> kn̥ <sup>ə</sup> ijh	<i>gnyis</i>
3	三	<i>san</i> <sup>1</sup>	Tksum > [sam]	ksum	> (unchanged)	<i>gsum</i>
4	四	<i>si</i> <sup>4</sup>	Lslits > s <sup>ə</sup> ijh	p[sits] + s <sup>ə</sup> ijh	> ps <sup>ə</sup> ijh	<i>bzhi</i>
5	五	<i>wu</i> <sup>3</sup>	Tlŋa? > ŋɔ <sup>?</sup>	lŋa	> (unchanged)	<i>lŋa</i>
6	六	<i>liu</i> <sup>4</sup>	Lruk > l <sup>ə</sup> uwk	truk	> (unchanged)	<i>drug</i>
7	七	<i>qi</i> <sup>1</sup>	Lts <sup>h</sup> it > ts <sup>h</sup> it			( <i>bdun</i> )
8	八	<i>ba</i> <sup>1</sup>	Tpret > p <sup>1</sup> et	pr[et + p] <sup>1</sup> et	> pr <sup>1</sup> et	<i>brgyad</i>
9	九	<i>jiu</i> <sup>3</sup>	Lk <sup>w</sup> u? > k <sup>ə</sup> uw <sup>?</sup>	t[ku] + k <sup>ə</sup> uw <sup>?</sup>	> tk <sup>ə</sup> uw <sup>?</sup>	<i>dgu</i>
				or tku	> (unchanged)	
10	十	<i>shi</i> <sup>2</sup>	Lgip > dz <sup>ə</sup> ip	p[gip] + dz <sup>ə</sup> ip	> dz <sup>ə</sup> i(p)	<i>bcu</i>
100	百	<i>bai</i> <sup>3</sup>	Tprak > p <sup>1</sup> æk	pr[ak + p] <sup>1</sup> æk	> pr <sup>1</sup> æk(k)	<i>brgya</i>

'One ~ alone':

PT 'one' \*ktek, WT *gcig*.

OC 'alone' (tjek) [Ltek] > MC (tsyek) [tɕ<sup>ə</sup>ek] > *zhi*<sup>1</sup> 雙 (1260c), not cited in

Baxter (1992). The current word for 'one' is *yi*<sup>1</sup> < MC ?jit [ʔ<sup>ə</sup>it] < OC \*ʔjit [Lʔit].

PST \*ktek.

The presyllable k-, reconstructed on the basis of WT g-, was lost in pre-OC times by random monosyllabisation. The division III of MC requires the reconstruction of a monosyllable in OC.

As part of the interference of MC forms with PT forms, the main syllable in PT \*ktek was replaced by the unfamiliar pronunciation for Tibetan speakers of MC [tɕ<sup>ə</sup>ek]. The combination \*k[tek] + tɕ<sup>ə</sup>ek gave rise to the hypothesised intermediate form \*ktɕ<sup>ə</sup>ek, well represented by WT *gcig*.

'Two':

PT \*knits, WT *gnyis*.

OC (\*njits) [Lmits] > MC (nyij<sup>h</sup>) [n̥<sup>ə</sup>ijh] > *er*<sup>4</sup> 二 (564a), not cited in Baxter.

PST \*knits.



The presyllable *k-*, reconstructed on the basis of WT *g-*, was lost in pre-OC times by random monosyllabisation. MC division III points to an OC monosyllable. The final *-ts* changed into *-js* > *-j<sup>h</sup>* by final cluster simplification (Baxter 1992:568–9).

The main syllable in PT \**knits* was replaced by the unfamiliar pronunciation of MC *ŋ<sup>o</sup>ij<sup>h</sup>*. The combination \**k[nits]* + *ŋ<sup>o</sup>ij<sup>h</sup>* (with the possibility of a pre-MC form *ŋ<sup>o</sup>is*) gave rise to the hypothesised intermediate form \**kn<sup>o</sup>ij<sup>h</sup>*, well represented by WT *gnyis*.

'Three':

PT \**ksum*, WT *gsum*.

OC \**sum* [*Tksum*] > MC *sam* [*sam*] (irregular rhyme) > *san<sup>l</sup>* 三 (648a). The regular MC rhyme is [*-am*] (Baxter -om).

PST \**ksum*.

The pronunciations of MC *sam* as well as any other MC forms in *-am* (see below), all belonging to the divisions I or II, were not exotic for Tibetan speakers. So no imitation occurred, and WT *gsum* derives directly from PT \**ksum* without interference from MC.

The problem raised by the reconstruction of 'three' and its word family is a very complex one. It has been treated in detail by Sagart (1999:14F152). A detailed discussion here would lead us too far from the present subject. I will only briefly outline my point of view.

Sagart proposed two forms for 'three': OC \**a<sup>s</sup>-hlɿm* > MC *sam* > *san<sup>l</sup>* 三 for the simple graph and OC \**a<sup>s</sup>-hlɿm* > MC *tshom* > *can<sup>l</sup>* 參 for the complex graph (or *da<sup>4</sup>xie<sup>3</sup>*). First of all, I consider that the rhymes *-um* / *-up* must be reconstructed, and that the changes *-um* > *-ɿm* and *-up* > *-ɿp* occurred after OC times. Aside from the basic form OC *Tksum* > *san<sup>l</sup>* 三 'three', the word family comprises the MC meaning of 驂 (read *can<sup>l</sup>*), MC *tshom* [*ts<sup>h</sup>am*] 'three horses in a team', and both MC meanings of 參 (read *can<sup>l</sup>*): MC *tsho* [*ts<sup>h</sup>am*] 'three, a triad' and (read *shen<sup>l</sup>*) MC *srim* [*ʂ<sup>o</sup>im*] 'the triad star of Orion'. The character 參 (read *san<sup>l</sup>*) is also used even today as a complex graph for 'three'. It must be noted that MC rhymes in *tshom* [*ts<sup>h</sup>am*] (division I) and in *srim* [*ʂ<sup>o</sup>im*] (division III) are regular in respect to the OC rhyme *-um* [*-um*]. For these two words I propose the reconstructions OC \**srum* [*Tksrum*] > MC *tshom* [*ts<sup>h</sup>am*] and OC \**srjum* [*Lsrum*] > MC *srim* [*ʂ<sup>o</sup>im*], which I consider as secondary forms of OC *Tksum*. To summarise (with my reconstructions only):

pre-OC *ksum* > OC *Tksum* > MC *sam* (irr.) > *san<sup>l</sup>* 三 'three'.

pre-OC *krsum* > (metathesis of *-r-*) OC *Tksrum* > MC *ts<sup>h</sup>am* 參 'three, a triad' > *can<sup>l</sup>*  
(also, incidentally, the reading with the meaning 'take part, visit'), also MC *ts<sup>h</sup>am* 驂  
'three horses in a team'.

pre-OC *krsum* > (loss of *k-* and metathesis of *-r-*) OC *Lsrum* 參 'the triad star of  
Orion' > MC *ʂ<sup>o</sup>im* > *shen<sup>l</sup>* (also the reading with the meaning 'ginseng').

The change of pre-OC *krsum* into OC *Tksrum* or *Lsrum* by metathesis of *-r-* from the presyllable to the main syllable is, of course, purely hypothetical. But the phenomenon of metathesis, although refused by some scholars (Handel 2002), can help us to understand the curious instability and the intrusive behaviour of some OC medial *-r-*, and the fact that items with or without this medial can occur in the same phonetic series. The idea of an ancient metathesis of *-r-* is supported by some lexical correspondences between WT and OC. Compare WT *rdul* 'dust' with *chen<sup>2</sup>* 塵 < MC *drin* [*d<sup>o</sup>in*] < OC \**drjɿn* [*Ldrɿn*] (< pre-OC \**drɿl*) 'id.' (example from Coblin 1986:68).

I propose that the archaic character for 參 previously had the meaning 'three horses in a team' because this notion was more familiar than 'triad of Orion'. The meaning of three horses is expressed by the upper part of the archaic character, which rather clearly shows the three horse's heads, contra some other scholars who prefer to see three stars in it (why should the three stars be tied on?). The lower part of the character has sometimes been interpreted as the phonetic element, but neither the element *zhen*<sup>3</sup> 珍 (OC rhyme -ɳ) nor *shan*<sup>1</sup> 杉 (OC rhyme -am) fits phonetically with *can*<sup>1</sup> 參 (OC rhyme -um). For myself, I prefer to see in the lower part of the archaic character for 參 the image of reins hung with ornaments.

The ancient pronunciation of 參 was used to derive numerous other characters that belong to the phonetic series GSR 647.

'Four':

PT \*psits, WT *bzhi*.

OC \*s(p)jij/ts [L-slits] > MC sij<sup>H</sup> (s<sup>o</sup>ij<sup>h</sup>) > si<sup>4</sup> 四 (518a).

PST \*plsits ~ \*pplits.

The presyllable p- (perhaps a prefix ?), reconstructed on the basis of WT *b-*, was lost in pre-OC times. The medial -l- is justified by occurrences in some Tibeto-Burman languages.

The main syllable in PT psits was replaced by a corrupted form of the unfamiliar pronunciation of MC s<sup>o</sup>ij<sup>h</sup>. The combination \*p[sits] + s<sup>o</sup>ij<sup>h</sup> gave rise to the hypothesised intermediate form \*ps<sup>o</sup>ij<sup>h</sup>, which is rather well represented by WT *bzhi*.

'Five':

PT \*lŋa, WT *lŋa*.

OC \*ŋaʔ [ʔlŋaʔ] > MC ŋu<sup>X</sup> [ŋɔʔ] > wu<sup>3</sup> 五 (58).

PST \*lŋaʔ.

Like 'three', WT *lŋa* derives directly from PT.

'Six':

PT \*truk, WT *drug*.

OC \*C-rjuk [L-ruk] > MC ljuwk [l<sup>o</sup>uwk] > liu<sup>4</sup> 六 (1032a).

PST \*truk ~ t-ruk.

The first element t- was lost during pre-OC times, suggesting that tr- must have been a kind of disjoined cluster. In Proto Thai proper, 'six' is reconstructed as \*<sup>h</sup>rok, the voicelessness being irrefutable proof of the presence of an old presyllabic element. As in the case of 'three' and 'five', the WT form derives directly from PT.

'Seven':

PT \*pdun, WT *bdun*.

OC \*thsjit [L<sup>h</sup>ts<sup>h</sup>it] > MC tshit [ts<sup>h</sup>it] > qi<sup>1</sup> 七 (400a).

Pre-OC (for PST) \*snit ~ s<sup>h</sup>nit. Note that the rhyme -it(s) exists in knits 'two', plits ~ pplits 'four (2+2)' and snit ~ s<sup>h</sup>nit 'seven (5+2)'.

The Tibetan and Chinese forms are not genetically related.

'Eight':

PT \*pret, WT *brgyad*.

OC (\*pret) [ʔpret] > MC (pɛt) [p<sup>h</sup>ɛt] > ba<sup>1</sup> 八 (281a).

PST \*pret.

The rhyme *-et* in PT pret was replaced by the unfamiliar pronunciation of the segment *-ʃet* of MC *pʃet*. The combination *\*pr[et + p]ʃet* gave rise to the hypothesised intermediate form *\*prʃet*, rather well represented by WT *brgyad*. The segment *-gyad* is the result of the interpretation of *-ʃet* in the phonetic system of Tibetan. Some scholars have interpreted *-g-* as an epenthetic element, but it must be remarked that epenthesis normally occurs at the junction of two syllables; it did not occur in the Tibetan word for 'eight'.

'Nine':

PT *\*tku*, WT *dgu*.

OC *\*kʷjuʔ* [*Lkʷuʔ* ~ *Lkuʔ*] > MC *kjuw<sup>X</sup>* [*kʷuwʔ*] > *jiu<sup>3</sup>* 九 (992a). Baxter's reconstruction of a labiovelar before a high rounded vowel is surprising, in spite of his solid argument, so I will propose an alternate form *Lkuʔ*.

PST *\*tkuʔ* ~ *\*tku*.

At first sight, WT *dgu* appears to derive directly from PT *\*tku*, but the voiced velar *-g-* does not fit perfectly with PST and could be a result of the laxness of the MC form. If so, the combination *\*t[ku] + kʷuwʔ* gave rise to the hypothesised intermediate form *\*tkʷuwʔ* represented by WT *dgu*.

'Ten':

PT *\*pgip* (?), WT *bcu*.

OC *\*gjip* [*Lgip*] > MC *dzyip* [*dzʰip*] > *shi<sup>2</sup>* 十 (686a)

PST *\*pgip*.

The presyllable *p-* (perhaps a prefix ?) is reconstructed on the basis of WT *b-* assuming that these forms are related. I suppose that the combination *\*p[gip] + dzʰip* gave rise to the hypothesised intermediate form *\*pdzʰip*. Could the WT rhyme *-cu* represent MC *dzʰip* after the loss of final *-p*? I must confess that I am not sure.

'Hundred':

PT *\*prak*, WT *brgya*.

OC *\*prak* [*Lprak*] > MC *pæk* [*pʰæk*] > *bai<sup>3</sup>* 百 (781a).

PST *\*prak*.

The demonstration for 'hundred' is parallel to that for 'eight'. The combination *\*pr[ak + p]ʰæk* gave rise to the hypothesised intermediate form *\*prʰæk* > *prʰæ*, rather well represented by WT *brgya*. The loss of final *-k* is unexplained but not unprecedented.

Among the eleven comparisons between Tibetan and Chinese numerals, nine can be considered as good correspondences. The words for 'seven' are not cognate, and the correspondence for 'ten' is not absolutely sure. In the correspondences for 'three', 'five' and 'six', the WT forms derive directly from PT without MC interference. They must be considered as pure inherited correspondences. In the correspondences for 'one', 'two', 'four', 'nine' (possibly) and 'ten' (if related), the MC monosyllable replaced the main syllable in the PT form; in the case of 'eight' and 'hundred' the segment replaced was the rhyme. These are what could be called corrupted (or modified) inherited correspondences, in which the Tibetan word is the result of a compromise between an inherited form and a borrowed segment while the Chinese term remains unchanged. These modified correspondences fall between pure inherited correspondences and full borrowing.

## 5 Additional examples

In this section, further examples are presented to illustrate both corrupted and regular correspondences between Tibetan and Chinese. They are taken from Coblin (1986) and Gong (1995). OC and MC forms are between square brackets; PT and PST forms are the author's. OC and MC forms follow the system of Baxter (1992); when not attested they are placed between brackets. We begin with examples where the inherited correspondence has been perturbed by direct influence.

'Weary, exhausted':

PT \*bral, WT 'fatigue, weariness' *o-brgyal*, 'to faint' *brgyal*.

OC (\*brjaj) [bral > Lbraj] > MC (bje) [b<sup>ə</sup>e] > pi<sup>2</sup> 疲 (25d), also pi<sup>2</sup> 罷 (26a).

PST \*bral.

It seems here that the influence extends only to the vowel of the PT form: \*br[a]l + [b]<sup>ə</sup>e gave rise to the hypothesised intermediate form \*br<sup>ə</sup>al, represented by WT *brgyal*.

'Dwell, establish':

PT \*bdoks, WT 'to sit, dwell' *bzhugs-pa*.

OC (\*djo?) [Ldo?] > MC dzyu<sup>X</sup> [dʒ<sup>ə</sup>u<sup>ʔ</sup>] > *shu*<sup>4</sup> 樹 (127j).

PST \*bdoks.

It seems that the segment -do- of PT \*bdoks was corrupted by the MC form according to the formula \*b[do]ks + dʒ<sup>ə</sup>u<sup>h</sup>, the result being represented by WT *bzhugs*.

'To flow, flowing':

PT \*run, WT *rgyun*.

OC (\*wrjɿn) [Lwrɿn] > MC (hwin) [ɣw<sup>ə</sup>in] > *yun*<sup>2</sup> 云 (see GSR 460 and 227).

The character 云 given by Coblin (1986) is not attested in Karlgren (1957),

Pulleyblank (1991), or Baxter (1992).

PST \*(C)rɿ/un (vowel reconstruction uncertain).

The rime of PT \*run was influenced by the rime of MC: r[un] + [ɣw<sup>ə</sup>in], obviously at a stage earlier than MC proper, before the OC medial -r- merged into the breathiness of the division III lax syllable. See 'eight', OC (\*pret) [Tpret] > MC (pɛt) [p<sup>ʰ</sup>ɛt] > *ba*<sup>1</sup> 八, WT *brgyad*. In any case the segment -gyun of WT cannot be an inherited form. This correspondance shows what Bodman (1980) called 'primary yod', considered today as indicating an acquired correspondence.

'Center, middle':

PT \*gruŋ, WT 'middle, midst' *gzhung*.

OC (\*k-ljuŋ) [Ltruŋ] > MC (trjuwŋ) [t<sup>ə</sup>uwŋ] > *zhong*<sup>1</sup> 中 (1007a).

PST \*kruŋ/truŋ.

The segment -ruŋ of PT \*gruŋ was replaced by MC [t<sup>ə</sup>uwŋ] according to the formula \*g[ruŋ] + [t<sup>ə</sup>uwŋ] with an approximative phonetic adjustment, the result being represented by WT *gzhuang*.

'Loyal, sincere':

PT \*gruŋ, WT 'to attend to, sincere' *gzhung*.

OC \*[Ltruŋ] > Lluŋ > MC [t<sup>ə</sup>uwŋ] > *zhong*<sup>1</sup> 忠 (1007k).

The OC and MC forms are set up on the model of *zhong*<sup>1</sup> §§ (1007a).

PST \*kruŋ/truŋ.

The process is the same as for 'center, middle'.

'Salt, salty':

PT \*ram, WT *rgyam-tshwa*.

OC (\*<sub>(C)</sub>-rjam) [(<sub>(C)</sub>ram > <sup>L</sup>ram] > MC (yem) [j<sup>ə</sup>em] > *yan*<sup>2</sup> 鹽 (609n).

The MC initial is irregular; according to the basic phonetic -ram of GSR 609, the regular initial would be l-.

PST \*<sub>(C)</sub>ram.

The rime of PT ram was replaced by an earlier form of MC: r[am] + j<sup>ə</sup>em, the result being reinterpreted by WT *rgyam*. For the interpretation of OC -r- by WT -gy- see 'eight' and 'hundred' above, 'weary, exhausted' and 'to blow'.

'To see':

PT \*mk<sup>h</sup>en, WT *mkhyen-pa*.

OC \*kens [T<sup>h</sup>kens] > MC ken<sup>H</sup> [kien<sup>h</sup>] > *jian*<sup>4</sup> 見 (241a). The MC vowel -ie- is characteristic of division IV.

PST \*<sub>(C)</sub>ken. A presyllable must be reconstructed to explain divisions I/IV.

The rime of PT \*mk<sup>h</sup>en was replaced by the rime -ien of the MC form according to the formula \*mk<sup>h</sup>[en] + [k]ien[s], the result being represented by WT *mkhyen*.

'Imitate, conform to':

PT \*sbaŋ, WT 'to learn, study, exercise' *sbyong(s)*, *sbyang(s)*.

OC 'imitate' \*pjang? [<sup>L</sup>paŋ?] > MC pjang<sup>X</sup> [p<sup>ə</sup>aŋ?] > *fang*<sup>3</sup> 放 (740i); also OC 'method, norm' \*pjang [<sup>L</sup>paŋ] > MC pjang [p<sup>ə</sup>aŋ] > *fang*<sup>1</sup> 放 (740a).

PST \*sbaŋ.

The rime of PT \*sbaŋ was replaced by the rime of the MC form: \*sb[aŋ] + [p]<sup>ə</sup>aŋ, the result being represented by WT *sbyong/sbyang*.

'Taste':

PT \*snep (?), WT *snyab-pa*.

OC (\*snep > \*thep) [snep > T<sup>h</sup>thep] > MC (thep) [thiep] > *tie*<sup>1</sup> 帖 (618p). The MC vowel -ie- is characteristic of division IV.

PST \*snep.

Coblin (1986) proposed OC hniap > MC thiep and reconstructed PST sniap, in fact on the basis of the WT form.

The rime of the supposed PT snep was replaced by the rime -iep of the MC form according to the formula \*sn[ep] + [th]iep, with palatalisation of the nasal, the result being represented by WT *snyab*. This case is similar to that of *mkhyen* 'to see'.

By way of contrast, we present below some correspondences involving OC medial -r- in which the two languages have evolved independently without interference (as in the cases of 'three', 'five' and 'six'). These are considered to be inherited correspondences, to be compared with the corrupted correspondences above.

'Add, apply':

PT \*pkral, WT 'to impose, to appoint to' *bkral*.

OC (\*kral >) \*kraj [kral > T<sup>h</sup>kraj] > MC kæ [k<sup>h</sup>æ] > *jia*<sup>1</sup> 加 (15a).

PST \*pkral.

'Bear, rear':

PT \*srel, WT 'bring up, rear' *srel*.

OC (\*srel >) \*srjen [\*srel > <sup>T</sup>sren] > MC sren [ʃ<sup>l</sup>ɛn] > *chan*<sup>3</sup> 產 (194a).

PST \*srel

'Busy, employed at':

PT \*brel, WT *brel*.

OC (\*brels) [<sup>T</sup>brels] > MC (ben<sup>H</sup>) [b<sup>l</sup>ɛn<sup>h</sup>] > *ban*<sup>4</sup> 辦 (219f).

PST *brel*.

'Hard, strong':

PT \*kraŋ, WT 'hard' *khraŋ*.

OC 'strong' \*kraŋ? [<sup>T</sup>kraŋ?] > kæng<sup>X</sup> [k<sup>l</sup>æŋ<sup>?</sup>] > *geng*<sup>3</sup> 梗 (745e).

PST \*kraŋ

'Shell, armor':

PT \*krap, WT 'shield, coat of mail' *khraβ*.

OC \*krap [<sup>T</sup>krap] > MC kæp [k<sup>l</sup>æp] > *jia*<sup>3</sup> 甲 (629a).

PST \*krap

'Weep':

PT \*krap, WT 'a weeper' *khraβ-khraβ*.

OC \*krjɿp [<sup>L</sup>krjɿp] > MC khip [k<sup>ɿ</sup>ip] > *qi*<sup>4</sup> 泣 (694h).

PST \*krap

## 6 Conclusion

We observe that the Tibetan language shows a two widely differing types of phonetic developments from PST and PT: one is straightforward and can be considered as regular while the other can be regarded as abnormal.

The regular changes, where the rimes are well preserved, can be represented by examples such as 'three' (\*ksum > *gsum*), 'add, apply' (\*pkral > *bkral*) and 'weep' (\*krap > *khraβ*), showing clear correspondences with OC. Other examples are: 'five' (\*lŋa > *lŋa*), 'six' (\*truk > *drug*), 'nine' (\*tku > *dgu*), 'bear' (\*srel > *srel*), 'busy' (\*brel > *brel*), 'hard' (\*kraŋ > *khraŋ*) and 'armor' (\*krap > *khraβ*). However, some of these may be pure borrowings of the whole words from MC into PT.

The abnormal changes, in which the rimes have been corrupted, can be represented by such examples as 'eight' (\*pret > *brgyad*), 'weary, exhausted' (\*bral > *brgyal*) and 'salt' (\*ram > *rgyam*), showing irregular correspondences with OC. Other examples are: 'one' (\*ktek > *gcig*), 'two' (\*knits > *gnyis*), 'four' (\*psits > *bzhi*), 'dwell' (\*bdoks > *bzhugs*), 'to flow' (\*run > *rgyun*), 'center' (\*gruŋ > *gzhung*), 'loyal' (\*gruŋ > *gzhung*), 'salt' (\*ram > *rgyam*), 'to see' (\*mk<sup>h</sup>en > *mkhyen*), 'imitate' (\*sbaŋ > *sbyong*) and 'taste' (\*snep > *snyab*). It is precisely to explain such correspondences that the author has proposed the borrowing of segments from Middle Chinese forms into Proto Tibetan as detailed above.

This mode of borrowing proposed, in which only a part of the word (main syllable or rhyme) is affected, is the consequence of a particular situation. The two languages in contact are genetically related, with a certain degree of intercomprehension and in a hierarchical relation of prestige. The Chinese language of MC times, being in a dominating position,

was regarded as prestigious by speakers of Tibetan, who were led to imitate, by a kind of affectation, the characteristic features of Divisions II and III which were unknown in Tibetan.

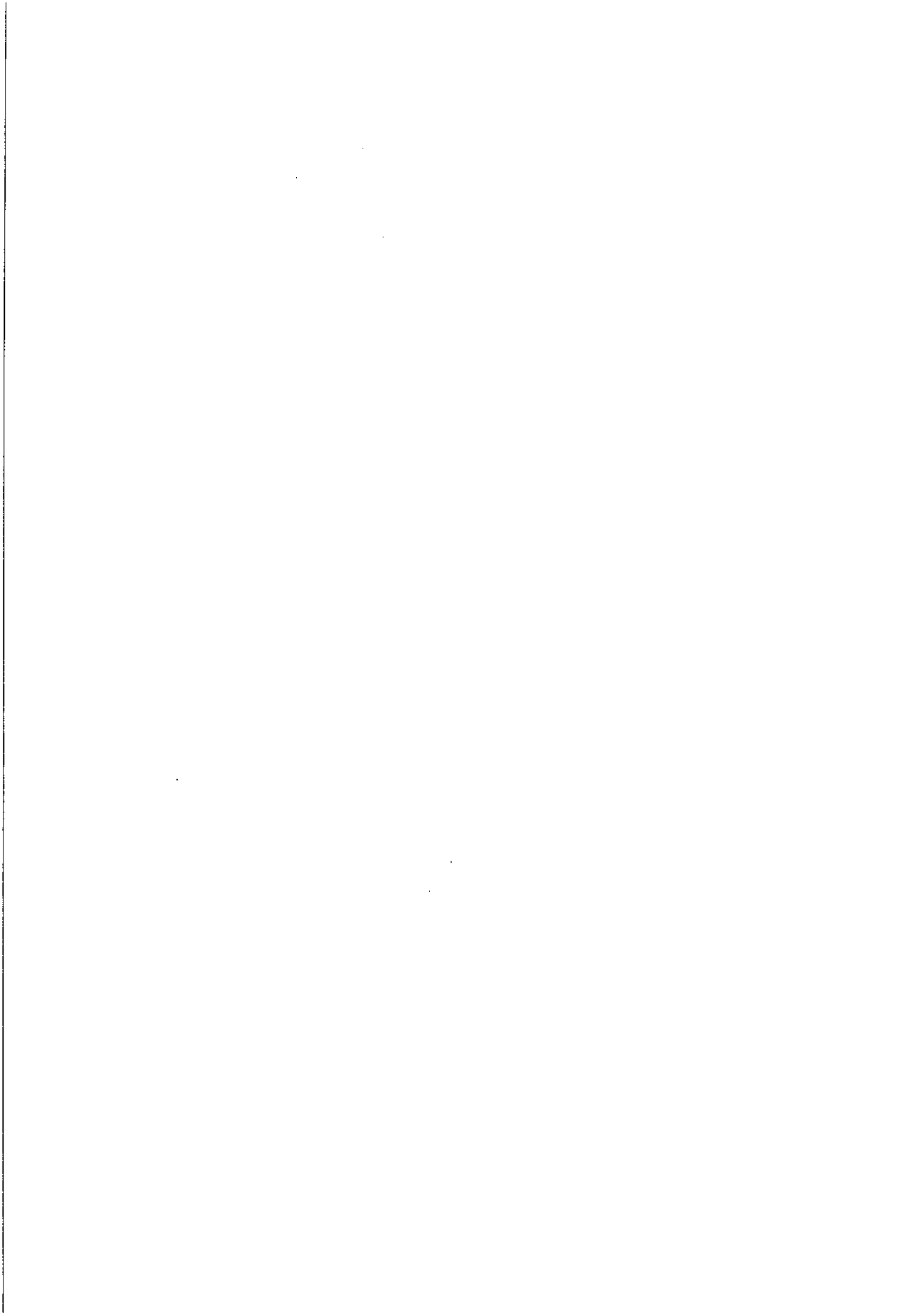
Consideration has been limited here to a restricted domain, mostly the numerals, which behave in general as a group, but the analysis could be extended with profit to most of the Tibetan vocabulary. I propose to call this special process of borrowing 'hypercorrection by affected imitation'. Such layers of borrowing have never been clearly identified in historical comparative studies.

## References

- Baxter, William H., 1992, *A handbook of Old Chinese phonology*. Berlin, New York: Mouton de Gruyter.
- 1993, Some phonological correspondences between Chinese and Tibeto-Burman. Paper presented to the 26th International Conference on Sino-Tibetan Languages and Linguistics, Osaka, Japan.
- Benedict, Paul K., 1972, *Sino-Tibetan: a conspectus*. Contributing editor, James A. Matisoff. Cambridge: Cambridge University Press.
- 1976, Sino-Tibetan: another look. *Journal of the American Oriental Society* 96:167–197.
- Bodman, N.C., 1980, Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. In Franz van Coetsem and Linda Waugh eds *Contributions to historical linguistics: issues and materials*, 34–199. Leiden: Brill.
- Coblin, W. South, 1986, *A Sinologist's handlist of Sino-Tibetan lexical comparisons*. Monumenta Serica Monograph Series XVIII. Nettetal: Steyler.
- Ferlus, Michel, 1991, Le dialecte vietnamien de Vinh. Paper presented at 24th International Conference on Sino-Tibetan Languages and Linguistics, Ramkhamhaeng University, Bangkok, Thailand.
- 1995, Particularités du dialecte vietnamien de Cao La Ha (Quang Binh, Vietnam). Paper presented at dixièmes journées de linguistique de l'Asie Orientale, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- 1998, Du chinois archaïque au chinois ancien: monosyllabisation et formation des syllabes *tendu / lâche* (Nouvelle théorie sur la phonétique historique du chinois). Paper presented at 31st International Conference on Sino-Tibetan Languages and Linguistics, University of Lund, Sweden.
- 2001, Les hypercorrections dans le thô de Lang Lo (Nghê An, Vietnam): les pièges du comparatisme. Paper presented at quinzièmes journées de linguistique de l'Asie Orientale, Ecole des Hautes Etudes en Sciences Sociales, Paris.
- Gong Hwang-cherng, 1980, A comparative study of the Chinese, Tibetan and Burmese vowel system. *Bulletin of the Institute of History and Philology, Academia Sinica* 51/3:455–490.
- 1995, The system of finals in Proto-Sino-Tibetan. In William S-Y. Wang ed. *The ancestry of the Chinese language*, Journal of Chinese Linguistics Monograph Series No. 8, 41–92.

- Handel, Zev., 2002, Rethinking the medials of Old Chinese: where are the *r*'s? *Cahiers de Linguistique Asie Orientale* 31/1:3–32.
- Karlgren, Bernhard, 1957, *Grammata Serica recensa*. Reprint from *Bulletin of the Museum of Far Eastern Antiquities* 29:1–332.
- Matisoff, James A., 1997, *Sino-Tibetan numeral systems: prefixes, protoforms and problems*. Canberra: Pacific Linguistics.
- Meillet, Antoine, 1936, Sur le bilinguisme. *Linguistique historique et linguistique générale*, vol. II. Collection Linguistique, Société Linguistique de Paris, 40:99–103. Paris: Klincksieck.
- Miller, Roy Andrew, 1988, The Sino-Tibetan hypothesis. *Bulletin of the Institute of History and Philology, Academia Sinica* 59/2:509–540.
- Norman, Jerry and W. South Coblin, 1995, A new approach to Chinese historical linguistics. *Journal of the American Oriental Society* 115/4:576–584.
- Peiros, Ilia and Sergei A. Starostin, 1996, *Comparative dictionary of five Sino-Tibetan languages*. Melbourne: University of Melbourne Department of Linguistics and Applied Linguistics.
- Pulleyblank, Edwin G., 1973, Some new hypotheses concerning word families in Chinese. *Journal of Chinese Linguistics* 1/1:111–125.
- 1991, *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese and Early Mandarin*. Vancouver: University of British Columbia.
- 1998, Old Chinese and Sino-Tibetan. Paper presented at 31st International Conference on Sino-Tibetan Languages and linguistics, University of Lund, Sweden.
- Sagart, Laurent, 1995a, Some Remarks on the Ancestry of Chinese. In William S-Y. Yang ed. *The ancestry of the Chinese language*. Journal of Chinese Linguistics Monograph Series No. 8, 195–223.
- 1995b, Questions of method in Chinese-Tibeto-Burman comparison. *Cahiers de Linguistique Asie Orientale* 24/2:245–255.
- 1999, *The roots of Old Chinese*. Amsterdam, Philadelphia: John Benjamins.
- Schuessler, Axel, 1974, *R and L in Archaic Chinese*. *Journal of Chinese Linguistics* 2/2:186–199.
- Starostin, Sergei A., 1996, Vowel correspondances in Sino-Tibetan. Paper presented to the 29th International Conference on Sino-Tibetan Languages and Linguistics. University of Leiden, the Netherlands.





# 18 *Phonetic fidelity vs. suggestive semantics: variations in Chinese character choice in the writing of loanwords*

---

MARK HANSELL

## 1 Introduction

The Chinese writing system has long been thought to be qualitatively different from other writing systems. In Western scholarship, early theories that described hanzi (Chinese characters) as ideographs have given way to theories that acknowledge the role of sound in the structure and use of hanzi.<sup>1</sup> Yet even these more informed and sophisticated approaches typically draw a sharp distinction between 'logographic' (or pleremic) Chinese and other 'phonographic' (or cenemic) writing systems. The former refers to a type of writing system in which individual graphs represent meaningful elements, and represent sound only secondarily if at all; the latter refers to a system in which individual graphs represent only sound, and represent meaningful elements such as morphemes or words only secondarily, as surrogates for spoken forms.

The one exceptional situation, in which almost everyone will agree that hanzi are used phonographically, is in the writing of foreign loanwords. Loanwords are typically described as using the same set of hanzi as native vocabulary, but with the hanzi 'emptied'<sup>2</sup> of their meaning (in the sense that the reader is intended to ignore a hanzi's usual meaning, and read it only for sound.) Some claim that there is a particular subset of hanzi used in the transcription of loanwords, constituting a sort of syllabary within the total set of hanzi; others maintain that the hanzi used in the writing of loanwords are selected arbitrarily from the set of all hanzi at the whim of whomever is the first to fix the loanword

---

<sup>1</sup> For in-depth discussion of the concept of ideography as applied to Chinese and Japanese see the various articles in Erbaugh 2001.

<sup>2</sup> See Haas (1976, 1983) for the origin of this term, which is also used in Hansell (1989b).

in a written form.<sup>3</sup> In either case, the writing of loanwords is taken to be a totally distinct, cenic process, as opposed to the normally plemic writing of Chinese native vocabulary.

Of the three possible types of lexical loans (semantic, graphic, and phonetic), the loanwords referred to above belong to only the phonetic. In semantic loans (loans in which the Chinese morphemes are chosen on the basis of their semantic similarity to the morphemes in the original), the connections between the individual hanzi and their usual Chinese meanings are unbroken. For example, in 熱狗 *règǒu* 'hot dog' (<English), the usual connections between the hanzi 熱 and the meaning 'hot' and between the hanzi 狗 and the meaning 'dog' respectively remain unchanged, the only innovation introduced by the semantic loan is the combining of the two morphemes into a new compound with a new meaning. In graphic loans, the choice of the hanzi is controlled by the source language original, and the resulting morphemes are those that are associated with those graphs in the recipient language (e.g. 社會 *shèhuì* 'society' <Japanese 社會 *shakai*.) Although the Chinese loan is created on the basis of graphic similarity to the Japanese, the connections between the particular morphemes and the hanzi are not any different from their usual connections in native Chinese lexical items. In a phonetic loan, however, no connection between meaning and graph is necessary. In 麥克風 *màikèfēng* 'microphone' (<English), any reference to the individual meanings of the constituent hanzi — 'wheat', 'conquer', and 'wind' respectively—is at best irrelevant and at worst misleading in interpreting the word. Choice of hanzi seems to be strictly on the basis of phonetic similarity to the English original.

A phonetic loan ideally represents the closest approximation of source-language pronunciation that recipient-language phonology can construct. Since there is no possible motivation for recipient-language speakers to deviate from source-language pronunciation, it is normally assumed that differences between model and replica are the result of discrepancies between the phonological structures of the two languages. How such differences come into play in the borrowing process can be formalised, for instance in Silverman's (1992) and Yip's (1993) treatments of English loans into Cantonese. Yip proposes a set of ordered constraints on the phonological form of loanwords, with 'FAITHFULNESS' (fidelity to the phonetic shape of the source-language model) ranking second, behind only the syllable structure constraints of the recipient language. Once the most faithful rendition of the model that also obeys the recipient-language syllable structure constraints is found, a new loanword is created, consisting of a string of Chinese syllables. Then, as stated or implied in Lou (1992), Hansell (1989b), French (1976), and Godwin (1979), it is given written form through the choice, for each syllable, of one Chinese character from the homophonous sets of characters that could possibly represent that syllable.

Perhaps the most familiar example (used in French 1976, Haas 1976 and 1983, Sampson 1985, among others) is the proper name *Marx*. Assuming a British r-less pronunciation, it can be adapted to Mandarin by first mapping English phones onto Chinese phonemes, then resyllabifying with epenthesis. The result is three Mandarin

<sup>3</sup> Kratochvil (1968), French (1976) and to a lesser degree Sampson (1985) support the idea of a small subset of characters being used for transliteration. Novotna (1968) finds much more variability, and Lou (1992) points out that Mainland renditions of phonetic loans use a much smaller and more regular set of hanzi than loans in Taiwan and Hong Kong.

syllables unspecified for tone: *ma*, *ke*, and *si*. For the first syllable, one hanzi (馬) is chosen from the set of homophones:

摩媽嗎癩麻馬嗎瑪碼嗎罵 etc.,

all pronounced *ma* (18 in all, according to Liang 1992). For the second syllable, 克 is chosen from among the 34 hanzi pronounced *ke*, and for the third syllable, 思 is chosen from among 37 syllables pronounced *si*. The result is the written form 馬克思 *mǎkèsī*, with the tones of the spoken form being dictated by the normal pronunciation of the hanzi used to represent each syllable.

Despite the clarity of the model presented above, and the clear distinctions drawn between the three types of loan, the distinction between semantic and phonetic loan does not hold up in the face of actual data. In the preceding example, the choice of 馬 is undoubtedly due to the fact that it is a common Chinese surname, and therefore contributes to the 'meaning' of the word by marking it as a proper name. Various scholars (Gelb 1963, Novotna 1968, Hansell 1989b, Tang 1989, Yao 1992, among others) have commented on a significant class of loans into Chinese which resemble the source language model in both sound and (though often to a lesser extent) meaning. These loans, called 'semanticised transcriptions' by Hansell (1989) and 音中有義 by Yao (1992), resemble folk etymologies. For example:

- 1) 維他命 *wéitāmìng* 'vitamin' ('support' + 'him/other' + 'life')
- 2) 聲納 *shēngnà* 'sonar' ('sound' + 'receive')
- 3) 香吉士 *xiāngjǐshì* 'Sunkist' ('fragrant' + 'lucky' + 'scholar')

In all three examples, the Chinese pronunciation is quite close to the English original, but there is also some relevant semantic content attached to the particular hanzi that have been chosen to represent those sounds. The meaning may be quite explicit (1), somewhat vague (2), or nothing more than positive connotations associated with the hanzi chosen (3), nevertheless it is difficult to attribute it to pure chance.<sup>4</sup>

The large number of homophones expressed by hanzi gives a possible explanation for this phenomenon. For example, the second syllable of (1) is *ta*. Since tone is not a feature of English, it can be safely ignored in the process of adapting English words into Chinese. There are about 30 different hanzi that are pronounced *ta* in Mandarin (as well as about 70 *wei* and 17 *ming*.) The choice of those particular hanzi can be attributed to the post-phonological-adaptation writing strategy described above: first adapt the source language model through a phonological sinicisation process; then take the output of that process (three toneless syllables), and choose, from among the homophonous hanzi with those pronunciations, the three whose morphemes can singly or in combination express the meaning of the original.

Attractive though it may be, this explanation only works if the details of phonological adaptation are not considered. None of these three examples is the most faithful adaptation of the English model into Mandarin syllables. In (1), given that Mandarin [w] is the usual

<sup>4</sup> Chinese is by no means unique in having different connotations attached to different graphic means of representing the same sound. Bolinger (1946) gives examples from English in which different connotations are attached to one spelling variant or another, from <grey> vs. <gray> to the agentive suffixes <-er> vs. <-or>, to <old shop> vs. <Olde Shoppe>. Such examples are relatively rare exceptions, however.

substitute for English [v] in loanwords, we would expect *wai* rather than *wei* in the first syllable<sup>5</sup> and *min* instead of *ming* in the third. In (2) the first syllable should be *song* rather than *sheng*, and in (3) the first syllable should be *seng* or *sun* rather than *xiang*. The phonetic distortions in (1)–(3) make it clear that the hanzi were not simply chosen from lists of homophonous characters after phonological adaptation was complete, but that the quest for meaning has influenced the ultimate phonological form.

Deviation from the ideal phonological adaptation is not based simply on a vague impulse towards meaning in general, but often is directed towards a very particular meaning. For example, the familiar (4) could just as easily have been rendered into Chinese as (5), which would be phonologically more faithful to the English, and equally meaningful:

4) 可口可樂 *kěkǒu kělè* 'Coca-Cola' ('tasty'+ 'cola')

5) \*口渴可樂 *kǒukě kělè* ('thirsty'+ 'cola')

Naturally, the marketers of the product who were responsible for coining the Chinese name preferred (4) over (5), despite its lower degree of phonological faithfulness. The choice of hanzi in semanticised loans can be manipulated at will to express the writer's attitude toward the referent, even when sales of a commercial product are not at stake. For example, (6) and (7) are two common written variants of the same English word.

6) 雅輩 *yǎbèi* 'yuppie' ('elegant'+ 'generation')

7) 雅痞 *yǎpǐ* 'yuppie' ('elegant'+ 'scoundrel')

Examples (1)–(7) demonstrate that there is a process by which phonetic loans into Chinese are assigned hanzi in a nonrandom, meaningful way. They also show that phonological faithfulness can be sacrificed in some way to enhance expression of meaning. The remainder of this paper will be devoted to exploring how, why, and by whom this alignment of sound and meaning is arrived at, based on the analysis of three different types of loanwords: commercial brand names, geographic place names, and general common vocabulary. Conclusions drawn from that analysis will be applied to larger questions about hanzi and loanwords in general.

## 2 Characteristics of the corpora

Semanticised phonetic loans are particularly numerous in the brand names of foreign products. In order to better understand the phenomenon by analyzing this particularly rich concentration, I collected a corpus of 537 foreign brand names used in Taiwan from advertising signs, print and electronic ads, and product labels. Brand names are particularly interesting as linguistic data because of the transparency of their origins and motivations. They are carefully designed by specialists with the goal of maximising the memorability of the name and the desirability of the product. Their etymologies are unlikely to get lost in the mists of the receding past, as happens to so many common lexical items, and their forms are standardised (by law!)

Of course, many of these advantages turn into decided disadvantages when one wishes to generalise from brand name data to other types of borrowing. Such clarity of

<sup>5</sup> Though if this word were originally borrowed into Cantonese, and later relayed by graphic loan into Mandarin, the first syllable would indeed be a regular adaptation.

motivation, careful calculation of creation, and standardisation of finished product are not usually found in other types of loanword. Brand names therefore cannot be used as a simple model for other loanwords, and phenomena observed in brand names cannot be directly generalised to all loanwords; they need to be seen more as artificially bred laboratory animals, whose behavior under carefully controlled conditions reveal the basic mechanics of the model, which once established can form the basis for observation of animals in the wild.

To complement these laboratory rats, I have assembled two other corpora which reflect different sets of motivations and constraints. One is a set of phonetically borrowed place names (toponyms) from a recent map of the world published in Taiwan. They are mostly the names of cities and physical features, chosen from all areas of the world at random. The toponyms were almost exclusively pure phonetic loans, and are very useful for comparison with the highly semanticised brand names.

The other corpus is a set of 163 phonetic loans culled from a much larger corpus of lexical borrowings of common nouns. All are either fully phonetic or semanticised phonetic loans, and all are clearly identified with a foreign source (usually English).

Of 537 brand names, 137 or 25.5 per cent were not phonetic loans (they were either semantic loans, graphic loans, or not loans at all.) Though no such thorough count was done on the other two corpora, it is clear that nearly all toponyms are phonetic loans, while less than 50 per cent of lexical loans in common vocabulary are phonetic loans.

### 3 Types of phonetic loans

The phonetically borrowed brand names can be divided into three categories based on the degree and type of semanticisation.

#### 3.1 Coherently semanticised loanwords (CSL)

These are loans in which the hanzi chosen to approximate the sound of the source language model represent morphemes that can be construed in a semantically compositional way to form a coherent phrase of some sort. For example:

- 8) 倍耐力 *bèinàilì* ('multiple durability') 'Pirelli' (tires)
- 9) 雅仕 *yǎshì* ('elegant official') 'Astor' (cigarettes)
- 10) 力多精 *lìduōjīng* ('essence of great strength') 'Lactogen' (baby formula)

CSLs most resemble folk etymologies, in that the source language sound sequence ends up encoded in a way that resembles as closely as possible a native word or phrase. Definitions of folk etymology differ, with some emphasising the unconscious and erroneous nature of the assumed derivation (Bolinger 1975:406–407, Crystal 1993), and others emphasising the reshaping of the word or phrase (Trask 1993:105). Whatever definition is used, there is always an element of seeking for meaning, and the idea that folk etymology transforms a lexical item into something that better satisfies a craving for meaningfulness. CSLs are consciously designed and therefore lack the unconscious, accidental origin that characterises folk etymology, but they pander to the same desire for transparent meaningfulness in the lexicon.

### 3.2 Randomly semanticised loanwords (RSL)

Here the hanzi chosen may be semantically related to the product referred to, or may be chosen simply on the basis of positive connotations or auspiciousness. The morphemes normally represented by the hanzi do not form a coherent or well-formed phrase in Chinese:

- 11) 沛綠雅 pèilǜyǎ ('copious' + 'green' + 'elegant') 'Perrier'
- 12) 豐力康 fēnglìkāng ('abundant' + 'strong' + 'healthy') 'Femcare' (formula)
- 13) 喜美 xǐměi ('happy' + 'beauty') (Honda) 'Civic'

In both (11) and (12), most of the hanzi refer directly or obliquely to the product. In (11), 沛 'copious' carries connotations of water, both because of the water radical in the hanzi itself, and because of the compound 沛沛 'copiously flowing, a great flow of water'. 綠 'green', besides its connotations of coolness, is the color of the Perrier bottle. In (12), abundance is associated with having plenty to eat (good nutrition), and strength and health are the qualities that parents want to nurture in their babies. Example (13), on the other hand, illustrates that hanzi may be chosen that have no direct relationship to the product, but that only bring pleasant associations.

While the examples in (11)–(13) are all fully semanticised, in that every hanzi is associated with the meaning or connotation that is to be conveyed, there are also RSLs that are only partially semanticised. Some of their hanzi are clearly semantically motivated, while others seem to be only phonetically motivated:

- 14) 雲絲頓 yúnsīdùn ('cloud' + 'silk' + 'pause') 'Winston' (cigarettes)
- 15) 蜜絲扛 mīsīkàng ('honey' + 'silk' + 'Buddha') 'Max Factor' (cosmetics)
- 16) 吉比 jībǐ ('auspicious' + 'compare') 'Skippy' (peanut butter)

In (14) both 'cloud' and 'silk' are associated with the smooth flavor that cigarettes like to boast of, while 'pause' is hardly relevant. 'Honey' and 'silk' in (15) resonate with desirable characteristics of cosmetics, but 'Buddha' is from a completely different arena. 吉 'auspicious' has no particular association with peanut butter, but provides overall positive connotations for (16) in a way that 比 'compare' does not.

### 3.3 Purely phonetic loanwords (PPL)

Loanwords that are rendered using hanzi that carry no special connection to the referent, and no special positive connotations, can be presumed to be chosen strictly on the basis of phonological fidelity. This is not to imply that RSLs must necessarily have been consciously designed to contain meaningful hanzi; it is certainly conceivable that some of the RSLs mentioned above might have acquired their meaningful hanzi purely by chance, or unconsciously. Since it is impractical to track down and interrogate the inventor of each brand name, and since it is always safer to err on the side of caution, it makes sense to simply restrict the category PPL to brand names with no semantically relevant hanzi at all, and avoid the sticky issue of chance vs. intent.

- 17) 米其林 mǐqíllín ('rice' + 3prs.pron + 'forest') 'Michelin' (tires)
- 18) 奧斯摩比 àosīmóbǐ ('mysterious' + 'this' + 'rub' + 'compare') 'Oldsmobile'

Out of a total of 400, the proportions of the different types of phonetic loan in the brand names corpus is shown in Table 1.

**Table 1:** Loan types (brand names)

Coherently Semanticised Loans	141		35.25%
Randomly Semanticised Loans	209		52.25%
Fully Semanticised		117	
Partially Semanticised		92	
Purely Phonetic Loans	50		12.5%

Compare the distribution in the 163 phonetic loans from common vocabulary shown in Table 2:

**Table 2:** Loan types (common loans)

Coherently Semanticised Loans	58		35.6%
Randomly Semanticised Loans	28		17.2%
Fully Semanticised		5	
Partially Semanticised		23	
Purely Phonetic Loans	77		47.2%

It is striking how similar the percentages of CSLs are in the two corpora. Equally striking is that the preponderance of RSLs over PPLs in the brand names is reversed in the common loans. Combining these two facts with the observation that in the third corpus, the toponyms, CSLs are virtually nonexistent, it becomes clear that different strategies of phonetic borrowing are used in these three areas of the lexicon.

The semanticisation of phonetic loans can be expected to have two different effects on the selection of hanzi, that of increasing phonetic distance from the original, and of increasing the frequency of certain hanzi. The former is the phenomenon seen in examples (1)–(5) above: that the written form of a phonetic loan may have a pronunciation that deviates from the ideal phonological adaptation of the source language model. If none of the homophonous hanzi corresponding to a given syllable have a desirable meaning or connotation, a non-homophone may be chosen on the basis of its more desirable semantics. It makes sense to hypothesise that the most semanticised words (CSLs and fully semanticised RSLs) have undergone this process to the greatest extent, and that they therefore would experience the most phonetic deviation from the source language originals.

As for frequency of hanzi, it also makes sense to assume that positive connotations make certain hanzi more likely to be chosen under a semanticising strategy. Since certain areas of the lexicon use the semanticising strategy more than others, the frequency distribution of hanzi in the different areas should reflect the difference: in brand names semantically potent hanzi should have the highest frequency; in toponyms semantically neutral ones should rank highest; and common loans should be in between. In the following section, both of these hypotheses will be tested.



#### 4 Quantitative analysis: phonetic fidelity

The phonetic fidelity of a given word can be estimated using a scoring system that measures the distance between an actual phonetic loan and the ideal phonological adaptation of the source language model. The basis for comparison is the ideal adaptation arrived at by replacing each of the source-language segments with the Chinese segment most closely matching it in place and manner of articulation, breaking up consonant clusters with an epenthetic vowel ([ə] or [ɔ], depending on the initial, except for sibilant initials, which use the apical vowel), and syllabifying non-nasal final consonants.<sup>6</sup> Tone is ignored.<sup>7</sup> The ideal adaptation is then compared to the existing Chinese loan, and discrepancies are rated according to the following scoring scheme:

- |        |                                                                                                            |
|--------|------------------------------------------------------------------------------------------------------------|
| 3 pts. | missing or extra CVC syllable: 歐舒康 <i>ōu shūkāng</i> 'Orthoxicol'                                          |
| 2 pts. | missing or extra smaller syllable: 愛馬 <i>àimǎ</i> 'Amana'                                                  |
|        | wrong front/backness of vowel: 固齡玉 <i>gùlíngyù</i> 'Kolynos'                                               |
| 1 pt.  | missing consonant: 蘭寇 — <i>lánkòu</i> 'Lancome'                                                            |
|        | wrong place or manner of articulation of C: 禮蘭 <i>lǐlán</i> 'Dearland'                                     |
|        | 1 pt. for each deviation in degree of vowel height, or in lip rounding:<br>賓士 <i>bīnshì</i> 'Benz' (1 pt.) |

##### Allowable exceptions:

- |                                         |                          |
|-----------------------------------------|--------------------------|
| missing post-vocalic /r/ or /l/         | /l/ for /r/              |
| retroflex-dental confusion in sibilants | confusion of /n/ and /ŋ/ |

##### Allowable conventional Chinese adaptations:

- |                  |                           |                                  |
|------------------|---------------------------|----------------------------------|
| wrong aspiration | <i>jia</i> for /ga/, /ka/ | <i>chu-, tsu-</i> for /dr/, /tr/ |
|------------------|---------------------------|----------------------------------|

The allowable exceptions and allowable conventional adaptations refer to frequently observed deviations from ideal adaptations. Allowable exceptions are those that are potentially explainable in terms of dialectal phonology (either Taiwan Mandarin or Cantonese).<sup>8</sup> The conventional adaptations are correspondences that have become so common that they are used regardless of phonetic dissimilarity. For instance, despite the fact that aspiration is an extremely important part of the phonological systems of both Chinese and English, aspiration in Chinese loanwords from English is 'consistently inconsistent'—it seems to be randomly assigned, with no attention to whether or not the English model is aspirated (as noted in Novotna 1968). Likewise, Mandarin *jia* is often used for English /ga/ or /ka/, because a large number of Cantonese-based loanwords from English use hanzi such as 加 (Mandarin *jia*, Cantonese *ka*) for English /ga/ or /ka/, and the

<sup>6</sup> Novotna (1968) gives a thorough account of the phonological adaptations that phonetic loans into Mandarin undergo.

<sup>7</sup> Lou (1992) points out that tone is assigned more or less randomly in Mandarin loanwords, due to the influence of transliteration practices. Because in Hong Kong borrowing takes place in a much more oral, face-to-face borrowing situation, Cantonese is the opposite: in loanwords from English, there are very regular correspondences between tone patterns and the stress patterns of the English models (Kiu 1977, Silverman 1992).

<sup>8</sup> For a systematic analysis of the differences between Taiwan and Mainland Mandarin, see Cheng (1985).

correspondence has become conventional despite the lack of strict phonetic similarity in Mandarin. The *chu-*, *tsu-* for /tʃ/ conventional adaptation may seem strange in terms of strict phoneme mapping, but is acoustically quite good.

One more adjustment to the coding scheme was the consideration of Cantonese pronunciation. A great many foreign consumer goods available in Taiwan first entered the Greater Chinese market through Hong Kong. Their brand names would therefore be sinicised according to Cantonese pronunciation, with the Mandarin version used in Taiwan created simply by graphic loan from Cantonese. For example:

19) 高露潔 'Colgate' Mand. [kaw lu tɕjɛ] Cant. [kow low kit]

20) 立頓 'Lipton' Mand. [li tuən] Cant. [lip tən]

In both (19) and (20), missing consonants in the Mandarin pronunciation (the /t/ in *Colgate* and the /p/ in *Lipton*) are present in the Cantonese pronunciation. The Cantonese is therefore much more phonetically faithful than the Mandarin, and has a lower phonetic distance score. In such cases, the lower of the two scores is used in this analysis.<sup>9</sup>

The full operation of phonetic distance scoring is shown in two examples below:

21) English original: 'Gatorade' Ideal adaptation: *gei te lei d(e)*

Actual Chinese: 開特力 *kāitèlì*

Scoring by syllable:

*gei* - kai 1 point (vowel height). (Aspiration irrelevant)

*te* - te 0 points (identical)

*lei d(e)* - li 2 points (vowel height, missing consonant)

Total score: 3

22) English original: 'Quaker' Ideal adaptation: *kui ke*

Actual Chinese: 桂格 *guìgé*

Scoring by syllable:

*kui* - gui 0 points (Aspiration irrelevant)

*ge* - ge 0 points (identical)

Total score: 0

Table 3 gives the average phonetic distance scores for the different types of loans (score per hanzi is given to eliminate the possibility of word-length effects):

<sup>9</sup> Robert Sanders (p.c.) suggests that Shanghai dialect was also a major source of loans, and should be included in the scoring. While I agree in principle, practical matters prevent it at this time.

**Table 3:** Phonetic distortion scores

		# of words	# of hanzi	word avg.	hanzi avg.
<b>Brand names</b>	CSLs	138	354	2.17	0.876
	RSLs	196	508	1.74	0.671
	PPLs	50	126	0.64	0.254
	<b>Total</b>	<b>384</b>	<b>988</b>	<b>1.78</b>	<b>0.691</b>
<b>Toponyms</b>		<b>100</b>	<b>316</b>	<b>0.48</b>	<b>0.152</b>
<b>Common loans</b>	CSLs	58	127	1.07	0.488
	RSLs	28	69	1.11	0.449
	PPLs	77	175	0.56	0.246
	<b>Total</b>	<b>163</b>	<b>371</b>	<b>0.83</b>	<b>0.367</b>

A brief look at the subcategories within brand names shows that there is certainly a trade-off between phonetic fidelity and semantics. Semanticised loans (CSLs and RSLs) show significantly greater phonetic distance from the foreign model than unsemanticised loans (PPLs). Major sacrifices of phonetic fidelity are being made to accommodate semanticisation. There is also a major discrepancy between fully semanticised and partially semanticised RSLs, as shown below:

#### **Brands RSL Breakdown**

Fully semanticised	113	279	2.23	0.903
Partially semanticised	83	229	1.07	0.389

Fully semanticised RSLs show the same high degree of phonetic distance as CSLs, while the partially semanticised are closer to PPLs.

The trade-off is also present in common loanwords, though not nearly to the same extent. PPL phonetic distance for brands and for common loans are virtually identical, while CSLs and RSLs are much lower for common loans than for brands. Toponyms have by far the lowest phonetic distance score, lower than any of the other PPLs.

This numerical analysis of phonetic fidelity in loans, though based on a rough and ready coding system, yields clear-cut results. The three corpora show sharp differences, differences that are easily correlated to their sociolinguistic status under the trade-off hypothesis. Brand names show the greatest phonetic distance, because they are created in a situation that puts a premium on meaningfulness, with phonetic accuracy being a secondary consideration. Common loans are much less market-driven, there is no such imperative to put the item being named in a positive light, though if a memorable and appropriate name should present itself, it will be used. Speakers will accept phonetic distortions compensated for by semantic suggestiveness, but they won't stretch too far for them. Toponyms represent the opposite end of the scale. The highest premium is placed on accuracy, since there is no accessible semantic content to most place names, and there is no motivation for injecting semantic content into them. The prime motivation is to render them in a way that will best approximate the sounds of the name of the place in a foreign language.

#### 4 Hanzi frequency

A comparison of the highest frequency hanzi in the three corpora should also show differences. The existence of a large number of RSLs in the brand names should increase the frequency of hanzi with positive connotations. On the other hand, the phonetic fidelity of toponyms should produce no such effect, and lack of attention to semantics would preclude use of many different homophones for the same syllable.

The high frequency hanzi for each corpus are listed below:

##### Brands: (out of 325 types, 988 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	23	達	dá	'arrive, attain'
2	22	樂	lè	'joy', 'happy'
	22	美	měi	'beautiful'
4	17	愛	ài	'love'
5	16	麗	lì	'beautiful'
	16	利	lì	'advantage'
	16	士	shì	'scholar, worthy'
8	15	力	lì	'strong'
9	13	雅	yǎ	'elegant'
10	12	百	bǎi	'hundred'
	12	保	bǎo	'treasure'
	12	斯	sī	'this'

The highest frequency hanzi in the brand name corpus almost all refer to the attributes of the good life that consumers (indeed, everyone) wish to attain—love, beauty, strength, plenty, etc. It is not until 10th place that a hanzi with no positive connotations (indeed, no connotative or denotative content at all), 斯 *sī* 'this', is seen.

##### Toponyms: (240 types, 851 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	40	斯	sī	'this'
	40	拉	lā	'pull'
3	39	亞	yǎ	'inferior, second'
4	27	克	kè	'to overcome'
5	26	加	jiā	'to add'
6	24	巴	bā	'to hope anxiously'
	24	利	lì	'advantage'
8	20	特	tè	'special'
9	19	馬	mǎ	'horse'
	19	爾	ěr	3prs. pron.

The toponym list is strikingly different from the brands list. Only two hanzi appear on both: 利 *lì* 'advantage, benefit' appears in roughly the same position in both; and 斯 *sī* 'this' is first in toponyms (tied for 10th in brand names). Otherwise, the toponym list is almost entirely devoid of positive connotations, and in some cases devoid of any

connotative or denotative meaning whatsoever (爾 ěr 3prs. pron., 斯 sī 'this'). The high frequency hanzi also make up a much higher percentage of the total tokens in toponyms as opposed to brands. The 10 highest frequency hanzi account for 32.67 per cent of all tokens in toponyms, but only 17.4 per cent in brands.

**Common loans:** (211 types, 371 tokens)

Rank	Number	Hanzi	Pronunciation	Meaning
1	18	克	kè	'to overcome'
2	10	卡	kǎ	'stuck'
3	8	沙	shā	'sand'
4	6	斯	sī	'this'
5	5	特	tè	'special'
	5	司	sī	'preside over'
	5	力	lì	'strong'
	5	達	dá	'arrive, attain'

(8 tied for 9th place)

Once again, the common loans are intermediate between brands and toponyms. Though semantically neutral hanzi predominate, there are a few with positive connotations lower down in the frequency list ('special', 'strong', 'arrive, attain'). The 10 most frequent hanzi account for 18.87 per cent of all tokens, a figure much more similar to brands than to toponyms.

## 6 Discussion

The above data can be brought to bear directly on questions about the process of hanzi choice in phonetic loans. They indicate clearly that hanzi are not chosen randomly from a set of homophones, since random choice would result in fairly equal distribution of hanzi in loanwords, and is inconsistent with the high frequencies that some hanzi achieve. However, any attempt to find a regular, conventional set of hanzi reserved syllabary-style for transliteration is equally doomed to failure. Such a model would predict that no two homophonous hanzi would appear at high frequencies in loanwords, a claim easily disproven by 利, 麗, and 力 (all pronounced *lì*) which are 5th, 5th (tied), and 8th respectively in frequency for the brand name corpus. This phenomenon is by no means uniform over the three different corpora, however: the fact that the ten highest frequency hanzi in the toponym corpus make up 32.67 per cent of all tokens, compared to just over half that amount for the other two corpora, show that the syllabary model works better for that particular type of loan than for others.

Just as clearly, the semanticisation of phonetic loans cannot be purely a product of sporadic folk etymology. In brand names, this is true because of the circumstances of their creation and dissemination: they are carefully crafted by specialists, and protected from any kind of change by trademark laws. It is true for the other categories of loanword for two reasons: first of all, the different distribution in the different corpora indicates that semanticisation cannot be the result of a random process, but is rather related to the sociolinguistic functions served by loanwords. Secondly, the degree of phonetic distortion

in semanticised loanwords, and especially in RSLs, indicates that in the trade-off between phonetic faithfulness and meaningfulness, speakers are willing to stretch quite a ways to accomplish some kind of meaning, even if it is not a meaning which would conventionally be associated with the referent.

The trade-off involved indicates that semanticisation is not a post-adaptation process, as suggested in Hansell (1989b) and Lou (1992), but is present throughout the borrowing process. Only such a model can explain the large degree of phonetic distortion in so many loans.

The implications of this finding go beyond the study of loanwords, to the conception of what hanzi are and how writing systems in general work. The common tendency to divide writing systems into abstract categories of 'pleremic' vs. 'cenemic' (Haas 1976, 1983) or 'logographic' vs. 'phonographic' (Sampson 1985) ignores the flexibility that they can display in actual use. Even appealing to the notion of cenemic use of an otherwise pleremic writing system (Coulmas 1989, Hansell 1989b) fails to do justice to the facts, since we have seen so many examples above of simultaneous consideration of sound and meaning in hanzi choice. These facts point up the need for a type of analysis of writing that does not depend on purely structural considerations, but on a functional consideration of the communicative goals of the writer and reader as well.

The data presented here also serve as a warning that the role of writing in lexical borrowing in general, and in Chinese loanword phonology in particular, cannot be ignored.<sup>10</sup> The study of writing, and of borrowing phenomena, need to be situated in a framework of broader understanding of the social meaning of linguistic forms, the communicative goals of language users, and the strategies that language users employ to achieve those goals.

## References

- Bolinger, Dwight, 1946, Visual morphemes. *Language* 22:333–340.  
1975, *Aspects of language*. New York: Harcourt, Brace Jovanovich.
- Cheng, Robert, 1985, A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language* 61:352–377.
- Coulmas, Florian, 1989, *The writing systems of the world*. Cambridge, MA: Basil Blackwell.
- Crystal, David, 1993, *An encyclopedic dictionary of language and languages*. London: Blackwell.
- French, M.A., 1976, Observations on the Chinese script and the classification of writing systems. In William Haas ed. 1976a:101–129.
- Erbaugh, Mary S., ed., 2002, *Difficult characters: interdisciplinary studies of Chinese and Japanese*. Columbus, OH: Ohio State University/National East Asian Languages Resource Center.
- Gelb, I.J., 1963, *A study of writing*. Chicago: University of Chicago Press.
- Godwin, Christopher D., 1979, Writing foreign terms in Chinese. *Journal of Chinese Linguistics* 7:246–267.

---

<sup>10</sup> As acknowledged in Yip (1993: 265).

- Haas, William, ed., 1976a, *Writing without letters*. Manchester, Eng.:Manchester University Press.
- 1976b, Writing: the basic options. In William Haas ed. 1976a:131-208.
- 1983, Determining the level of a script. In Florian Coulmas and Konrad Ehlich eds *Writing in focus*, 15-29. New York: Mouton.
- Hansell, Mark, 1989a, Lexical borrowing in Taiwan. Unpublished UC Berkeley PhD dissertation.
- 1989b. Non-logographic Chinese and the non-alphabetic alphabet. *Berkeley Linguistics Society* 15:102-113.
- 2002, Functional answers to structural problems in thinking about writing. In Mary Erbaugh ed., 124-176.
- Kiu, K.L., 1977, Tonal rules for English loan words in Cantonese. *Journal of the International Phonetics Association* 7:17-22.
- Kratochvil, Paul, 1968, *The Chinese language today*. London: Hutchinson & Co.
- Liang, Shiqiu, 1992, *Far East Chinese-English dictionary* (concise edition). Taipei: Far East Book Company.
- Lou Chengzhao, 1992, Transliterating non-Chinese proper nouns into Chinese: a comparative study of usage in mainland China, Taiwan, and Hong Kong. *International Journal of the Sociology of Language* 97:121-133.
- Novotna, Zdenka, 1968, Contributions to the study of loan-words and hybrid words in modern Chinese, part II. *Archiv Orientalni* 36:295-325.
- Sampson, Geoffrey, 1985, *Writing systems*. Stanford, CA: Stanford University Press.
- Silverman, Daniel, 1992, Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9:289-328.
- Tang Ting-Chi, 1989, Xincichuangzao yu Hanyucifa [The creation of new words and Chinese morphology]. In Tang Ting-Chi ed. *Hanyucifajufa xuji [Further writings on Chinese morphology and syntax]*, 93-146. Taipei: Student Book Company.
- Trask, Robert L., 1993, *A dictionary of grammatical terms in linguistics*. London: Routledge.
- Yao Rongsong, 1992, Taiwan xianxing wailaiyude wenti [Problems of current loan vocabulary in Taiwan]. *Bulletin of National Taiwan Normal University* 37:329-362.
- Yip, Moira, 1993, Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics* 2:261-291.

# 19 *Original and sophisticated features of the Lepcha and Limbu scripts*

---

RICHARD KEITH SPRIGG

I value this opportunity of making a contribution to the volume to honour James Matisoff especially because it brought to mind the introductory remarks to the article that he contributed, 'A new Sino-Tibetan root \*d-yu-k', for my festschrift, *Prosodic analysis and Sino-Tibetan linguistics: to honour R.K. Sprigg* (Bradley, Henderson & Mazaudon 1989).

For auld lang syne, my dear,  
For auld lang syne,  
We'll tak a cup of kindness yet  
For auld lang syne!

Only three Himalayan languages have scripts of their own, and all three of them are Nepalese languages: Newari, Lepcha, and Limbu, though the majority of Lepcha-speakers live outside Nepal. This article is confined to four remarkable features of the Lepcha and Limbu scripts. I have 4,826 for the number of Lepcha-speakers in Nepal, mostly in the Ilam area, and the number of Limbu-speakers, according to the 1991 Survey also, is 254,288 (Kansakar 1996:6, 8).

The Lepcha script is known these days in the form in which it appears in Mainwaring (1876:1–18; see Appendix 1); and the Limbu script is probably known only in the form in which it appears in Chemjong (1962:21–23; see Appendices 2 and 6); but there are earlier versions of both scripts to be found in handwritten books collected in Darjeeling in the 1840s by Hodgson (1847). These earlier books give an original, and strikingly different, picture of the two scripts from the later, and better known, versions, in which the influence

---

David Bradley, Randy LaPolla, Boyd Michailovsky and Graham Thurgood, eds, *Language variation: papers on variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, 291–304.

Canberra: Pacific Linguistics, 2003.

Copyright in this edition is vested with Pacific Linguistics.



of the Tibetan script on the Lepcha, and the Devanagari script on the Limbu is easy to detect.

## 1 The number of syllable-initial symbols

The syllable-initial symbols are customarily presented in the form of a syllabary in which a particular member of the syllabic-vowel system is symbolised jointly with an initial consonant or consonant cluster or a non-syllabic vowel (*y*-, *w*-, or *h*-), or, in the case of 'a, an initial glottal-plosive consonant in free variation with the syllabic vowel as itself syllable-initial (Appendices 1 and 2).

### 1.1 *Lepcha*

Mainwaring's Grammar gives the total number of symbols in the Lepcha syllabary, the 'á-mo 'mother', as 35; one of the two handwritten books in the Hodgson Collection, *róng-kup sa lháp-sho bo-sho-mú gum* (vol. 79, 1-50, 1903 VS, 1847 AD), now housed in the India Office Library, London, also gives the number of symbols in the syllabary as 35; but they are not quite the same as Mainwaring's symbols. For one thing the symbol *tsha* is missing from the syllabary in both of the Hodgson books, presumably through an oversight. This would, of course, reduce the syllabary's total from 35 to 34 (or 33 for Hodgson's second book, which also has *bla* missing); but they make up the total to 35 (or 34) by including 'a in the syllabary, as the first member of the series: 'a, ka, ga, etc. Minawaring, on the other hand, has treated 'a not as a member of the syllabary but as 'the basis of all vowels, eight in number: 'a, 'á, 'i, 'o, 'ó, 'u, 'ú, 'e' (3). In other words the Hodgson books have treated 'a on the model of the *a* symbol in the Tibetan syllabary, the *gsal-byed sum-cu* 'thirty radicals', in which *a* is placed thirtieth and last, except that 'a is placed first in the Lepcha syllabary of the Hodgson books.

Mainwaring, on the other hand, appears to have treated the 'a, 'á, 'o and 'ó of Lepcha on the model of the sub-sets of four Devanagari symbols shown at (i) below in contrast with the remaining sub-set of six symbols, making a total of ten *svaravaṛṇa* symbols (eleven if *ṛ* is included):

i.	'a	-'a								'o	'au
ii.			<i>i</i>	- <i>i</i>	<i>u</i>	- <i>u</i>	<i>e</i>	<i>ai</i>			( <i>r</i> )

However, Mainwaring's scheme goes beyond these two Devanagari sets of vowels; it groups the remaining four Lepcha symbols 'i, 'í, 'u, 'ú and 'e in a single set with 'a, 'á, 'o and 'ó, the initial symbol ' combined with the eight vowel symbols -, -i/ -í, -u, -ú and *e*, on the same principle as the Tibetan script.

Before leaving this controversial topic, it is useful to recall that while the symbols 'a, 'á, 'i, 'í, 'u etc. have an *optional* pronunciation with a syllable-initial consonant sound,

namely the glottal plosive (in addition to their pronunciation with a syllable-initial syllabic-vowel sound) mentioned earlier, the glottal plosive consonant is *obligatory* in the combination 'y- and must be pronounced before the non-syllabic vowel for the following y ([ʔj-], in such words as 'yok 'work' (Tibetan *gYog*), 'yu 'turquoise' (Tibetan *gYu*), and ta-'yu 'wife, female'.

## 1.2 Limbu

In Limbu too the position as regards 'a is very much like that of 'a in Lepcha. In the three Hodgson-Collection books vol. 84, pages 1–22 and 23–99 and vol. 87 (*lību haru ko kakhaharā*) has been included in the syllabary, but in the third place, after *ka* and *pa*, in a syllabary of 20 symbols, not in the first place as in Lepcha. Campbell (1855), however, has only 19 symbols in the syllabary because, like Mainwaring for Lepcha as discussed in section (1.1) above, he has removed 'a from the syllabary and associated it with a vowel series: 'a, 'e, ē, ĩ, ī, u, ō, ŏ, ai (210; cf. also van Driem 1987:548).

From the phonetic spelling in Devanagari script that has been added to each member of the syllabary in the Hodgson books it is clear that the vowel sound intended for *ka*, *pa*, 'a, etc. is the open back vowel [ɑ] (Devanagari *kā*, *pā*, -'a, etc., as in reciting that Tibetan syllabary, the *gsal-byed sum-cu*, *ka*, *kha*, *ga*, *nga*, etc., not the vowel sound used in reciting the Devanagari *ka*, *kha*, *ga*, *gha*, *nga*, etc. in Nepali and Hindi.

Chemjong (1962) agrees with Campbell (1855) in removing 'a from the syllabary and incorporating it in a vowel series, *svavararṇa*: 'a, -'a, 'i, 'u, 'e, 'ai, 'o, 'au, 'a:, 'e: (21–2; see Appendix 2); but he has gone further, and changed the vowel quality of 'a from the quality resembling the Tibetan, the open back quality, [ɑ], to the Limbu vowel quality nearest to the quality used in reciting *ka*, *kha*, *ga*, *gha*, etc. for Nepali, a half-open back rounded vowel sound ([ɔ]).

Chemjong has gone even further in changing the character of what he calls *śirijaṅgā lipi* (1962:20); he has adapted it to the needs of other Nepalese languages by adding eleven other symbols to the 20 given in the Hodgson Collection books. The extra symbols, 'nayaṅgā akṣara', are: *ga*, *gha*, *ja*, *jha*, *da*, *dha*, *ba*, *bha*, *tra*, *sa*, and *gya*. The symbols *ga*, *ja*, *da*, and *ba* are not needed in a Limbu script because the sounds that they symbolise in reading the Devanagari script, voiced sounds, are, with a very few exceptions, in complementary distribution with the corresponding voiceless sounds of *ka*, *ca*, *ta*, and *pa* (Sprigg 1966:452, n.7); so either of these two sets of symbols could represent these voiced/voiceless pairs of sounds, the other set being superfluous. The same relationship, complementary distribution, also applies to the Devanagari symbols *gha*, *jha*, *dha*, and *bha* as compared with *kha*, *cha*, *tha*, and *pha*. It shows remarkable acuity on the part of some linguistically minded Limbu, perhaps 'Siri-jungna, called also the Dorze Lama of Yangrup' (*Gazetteer* 1894:37; see also Sprigg 1959:591) more than two hundred years ago, in that he had detected this phonetic relationship and taken it into account while



no.1:

no.2: *ha ra kha ta tha da la kla pla fla va sa*

It is not clear to me why Thikung Men Salong, or whoever it was that devised the Lepcha script, should have chosen to put the symbols in either of these two orders; but, to some extent, phonetic principles can be seen to be at work:

- a. two of the symbols that have a velar-plosive initial sound, *ka* and *ga*, are grouped together, as the 2nd and 3rd members of the syllabary, in both books;
- b. four of the labial-initial symbols, *pa*, *fa*, *ba*, and *ma*, are grouped together, as the 4th, 5th, 6th and 7th symbols, in both books;
- c. three of the dental-initial symbols, *ta*, *tha*, and *da*, are grouped together, as the 11th, 12th, and 13th symbols of the syllabary as it has been given in book no. 1, and as the 26th, 27th and 28th symbols in book no. 2;
- d. the lateral-initial symbol *la* all six of the symbols that have a lateral cluster as their initial sounds, *kla*, *pla*, *fla*, *bla*, *gla*, and *m̥la*, and the voiceless lateral symbol *hla* are grouped together as symbols 14 to 19 and 21 to 22 in book no. 1 (*va* has been inserted between *gla* and *m̥la* as the 20th symbol); and, in book no. 2, *gla*, *m̥la*, and *hla* have been grouped together as the 8th, 9th, and 10th symbols, while *la*, *kla*, *pla*, and *fla* have been grouped as the 29th, 30th, 31st, and 32nd symbols (as has already been noted above, *bla* is missing from the syllabary in the form in which it appears in book no. 2; it should have come after *fla*);
- e. if the members of the syllabary are looked at from the point of view of blocks of symbols, then the seven symbols 'a to *ma* form the first block in both books; from that point onward the two books diverge, because the second block in book no. 1, comprising the eleven symbols *ha* to *bla*, forms the third block in book no. 2 (with the exception of *bla*, which is missing); the third block of book no. 1, comprising the fifteen symbols *gla* to *cha* (an extra symbol, *va*, has been inserted, in book no. 1, between *gla* and *m̥la*) is the second block of book no. 2:

book no.1: block 1, block 2(+ *va*), block 3

book no.2: block 1, block 3, block 2;

so in book no.2, the order of blocks 2 and 3 is the reverse of what is to be found in book no.1.

## 2.2 Limbu

### 2.2.1 Chemjong's version, on the model of Devanagari

Chemjong has followed the Devanagari script by beginning with the ten *svavarṇa* 'a to 'e:, except that *i*, *u*, and *ṛ* have been omitted and 'a: and 'e: have been added; he continues with the 30 *vyañjana-varṇa* from *ka* to *nya*, and then from *ta* to *ha*, with *tra* and *jnya* added (see Appendix 2).

### 2.2.2 The earlier order, from the Hodgson books

The Hodgson books give the following order:

<i>ka</i>	<i>pa</i>	<i>'a</i>	<i>ma</i>	<i>ta</i>	<i>ya</i>	<i>tha</i>	<i>na</i>	<i>sha</i>	<i>nga</i>	<i>sa</i>	<i>wa</i>
<i>ha</i>	<i>la</i>	<i>ca</i>	<i>pha</i>	<i>kha</i>	<i>ra</i>	<i>cha</i>	<i>nya</i>				

There are no clear indications of a phonetic analysis in this, the earlier, order of the symbols, though there is a slight resemblance to the two earlier orders of the Lepcha script shown in 2.1.2 above: *ka*, *pa*, *'a*, and *ma* all occur in the first four places in the Limbu order of symbols; and the same four occur in the first four places of the Lepcha too, if we ignore the Lepcha *ga*, *fa*, and *ba* symbols, which are not found in Limbu:

Lepcha:	<i>'a</i>	<i>ka</i>	<i>(ga)</i>	<i>pa</i>	<i>(fa</i>	<i>ba)</i>	<i>ma</i>
Limbu:		<i>ka</i>	<i>pa</i>	<i>'a</i>			<i>ma</i>

## 3 The syllable-final symbols as a separate system

### 3.1 Lepcha; the nine 'á-kup 'children'

It is mysterious, and quite remarkable, that the Lepcha script should have a set of diacritics to symbolise its (eight) syllable-final consonants: *-k*, *-m*, *-l*, *-n*, *-p*, *-r*, *-t*, *-ng* (*kàng*), this last of which comes near to being in complementary distribution with *-ang* when written with *nyín-dó* 'sun-moon'; all of these diacritics are superscript, that is to say they are written above the 'radical' symbol (*'á-mo*), except for two, the *kàng* and the *nyín-dó*, which are prescript (written before the (*'á-mo*)) (Tamsang 1982:5; see Appendix 3). Diringer refers to these diacritics in his authoritative study of writing systems of the world in the following terms:

peculiar features of the Lepcha character are the vowel signs and the final marks of eight consonants (*k, ng, t, n, p, m, r, l*) which consist of dashes, dots and small circles and are placed above and before the preceding letter (1948:280)

These nine superscript and prescript diacritics for the eight syllable-final consonants are almost unique in the writing systems of the world. It is only in one of the Tibetan styles of writing, the cursive style (*'khyug-yig*), that I can find a parallel: in *'khyug-yig* Tibetan writing too syllable-final *m*, as in *lam* 'road', is written above the radical (*gsal-byed*) in the form of a bar and a loop.

Some of the syllable-final consonant sounds, the sounds for *-k*, *-p*, and *-t*, are slightly different from the consonant sounds that are used in syllable-initial position, for *k-*, *p-*, and *t-*: the syllable-final sounds are stops — they have no audible release; but the syllable-initial sounds are plosives — they have plosion, an audible release into a following vowel (or, on occasions, a consonant, *l* or *r*); so, because of this phonetic difference, slight

though it is, one might agree that *-k*, *-p*, and *-t* should be written differently from the *k*-, *p*-, and *t*- of, for example, *ka(-)*, *pa(-)*, and *ta(-)*; but there is no such difference between the syllable-final consonant sounds for *-ng*, *-m*, *-n*, *-r*, and *-l* and the syllable-initial sounds to be heard when one pronounces, for example, the syllables *nga*, *ma*, *na*, *ra*, and *la*. Since these five consonant sounds are pronounced the same in both these two positions in the syllable, it might seem strange that whoever it was who devised the Lepcha script should have chosen not to write them with the same symbols.

In answer to this problem, some linguists at the present time would agree that it does seem strange, and would come to the conclusion that the inventor of the Lepcha script had, in the '*á-kup ka-kyót*', the nine syllable-final symbols, introduced nine unnecessary symbols into the script; but during the last sixty years another school of linguistics has come into being that would give the answer 'no' to this question, and would consider the inventor of the script to be correct in having devised a separate set of symbols for the eight final consonants.

This more recent school of linguistics, to which I myself belong, would point out that since only eight consonants are distinguished in Lepcha in syllable-final position while some thirty-five consonants and consonant clusters like *kl-* and *pl-* are distinguished in syllable-initial position, such as the consonant sounds at the beginning of the syllables *ka*, *ga*, *kla*, and *gla*, the distinctive value (or power of making a distinction in the meaning of words) of *-k*, *-ng*, etc. in a set of only eight possible consonants must be quite different from the distinctive value of the initial consonants symbolised in the syllables *ka*, *nga*, and all the other thirty-three consonants and consonant clusters that need to be distinguished in pronouncing the syllable-initial set. If I compare the Lepcha script to players in football teams, it is as though the Lepchas had invented two different kinds of football game, one game for teams of eight players and the other game for teams of thirty-five players. The value of a member of the eight-member team to his team is quite high, one to eight, one eighth of the total; so his value is quite different from the comparatively low value of a member of the thirty-five-member team to his team, one to thirty-five on average, or one thirty-fifth of the total. If a member of the eight-member team is sent off by the referee, or has to leave the field because of injury, it very much reduces his team's chances of winning; but the thirty-five member team might hardly notice losing one of its members.

This theory that distinguishes separate sets of sound units for different places in the syllable and the word, as I have just illustrated from Lepcha through the thirty-five '*a-mo*' in syllable-initial position and the nine '*á-kup*' in syllable-final position (though the number of final consonant units is only eight because the *kàng* and the *nyín-dó* are both used for *-ng*) was first put forward in 1935, by J.R. Firth, while writing about the Marathi language. Palmer has described how

for the nasals in Marathi he noted a two-term alternance initially, a three-term alternance finally, but, though phonetically there were eight different sounds, one

'unique' homorganic nasal before medial consonants, he comments 'I should not want to identify all those *n* sounds'. (1970:x-xi).

The term 'polysystemic approach' has been given to an analysis such as Firth's, in which there are a number of separate and independent systems; so the inventor of the Lepcha script, astonishing though it may seem, was applying the 'polysystemic approach' to devising the Lepcha script more than two hundred years, perhaps, before Firth had developed his theory and before the term 'polysystemic approach' had been introduced.

### 3.2 *Limbu*

The main differences between the Limbu syllable-final diacritics (Appendix 6) and the corresponding Lepcha symbols are that:

- (i) the Limbu symbols, *-k*, *-t*, *-m*, *-m*, *-ng*, *-n*, *-p*, and *-l*, are postscript, except for *-t* and *-ng*, which are subscript;
- (ii) the two final symbols *-k* and *-p* are almost the same in shape as the syllable-initial symbols *k-* and *p-*, and so, perhaps, hardly qualify to be classed as diacritics; and
- (iii) Limbu has clusters of syllable-final symbols, such as *-m-* and *-k*, *-ng-* and *-k*, and *-k-* and *-t*.

On the other hand, the Limbu syllable-final symbols resemble the Lepcha syllable-final symbols at 3.1 above in forming a separate set, or system, of mutually defining units; so Limbu can also claim that its script implies that the phonological analysis on which it is based should be regarded as polysystemic. Such an analysis would, of course, commend itself to the followers of J. R. Firth's theory, prosodic analysis.

## 4 The seven *la-thyu* as Lepcha diphonic symbols

The fourth remarkable feature of the Lepcha script, for which the credit, according to Tamsang, should be given to a Lepcha, Thikung Men Salong, is the set of seven symbols *kl-*, *pl-*, *fl-*, *ml-*, *gl-*, and *hl-* termed *la-thyu*; each symbolises a cluster of two consonant sounds the second of which is a lateral sound, except that *hl-* symbolises a single sound, a voiceless frictionless lateral consonant or voiceless 'l'sound (cf. Tamsang 1982:5-6). The six members of the *la-thyu* that symbolise consonant clusters, all the *la-thyu*, that is, except *hl-*, can be termed diphonic because for them a single symbol is used to symbolise not one but two consonant sounds linked together in a cluster. Diphonic symbols such as these are not to be found in the Tibetan script; indeed they are rarely to be found in any other of the world's scripts. The only other such symbols that readily come to mind are the letters *zeta*, *xi*, and *psi* of the Greek alphabet, which symbolise, respectively, the clusters of consonants [zd] (or perhaps, [dz]), [ks], and [ps], which Allen refers to as 'consonant groups represented by single symbols' (1968:53-57).

## 5 Conclusion

I think that the four original, and sophisticated, aspects of the Lepcha and Limbu scripts that I have mentioned above amply justify my claim that the scholars who are credited with devising these two scripts, Thikung Men Salong, perhaps, or Chador Namgyal, the third Chogyal of Sikkim, for the Lepcha and, for the Limbu, Sirijunga, 'the Dorze Lama of Yangrup', were outstanding and forward-looking linguists.

## References

- Allen, William S., 1968, *Vox Graeca*. Cambridge: Cambridge University Press.
- Bradley, David, Eugénie J.A. Henderson, and Martine Mazaudon, eds, 1989, *Prosodic analysis and Asian linguistics: to honour R.K. Sprigg*. Canberra: Pacific Linguistics.
- Campbell, Archibald, 1855, Note on the Limbo alphabet of the Sikkim Himalaya. *Journal of the Asiatic Society of Bengal* XXIV:202–203, Plate XIV.
- Chemjong, Iman Singh, 2018 NS [1962 AD], *Limbu-Nepali-English dictionary*. Kathmandu: Nepal Academy.
- Diringer, David, 1948, *The alphabet*. London: Hutchinson.
- Driem, George van, 1987, *A grammar of Limbu*. Berlin: Mouton de Gruyter.
- Firth, John Rupert, 1935, Phonological features of some Indian languages. *Proceedings of the second international congress of phonetic sciences* [reprinted 1957, in J.R. Firth, *Papers in Linguistics 1934–1951*, 47–53. London: Oxford University Press].
- Gazetteer of Sikkim*, 1894, (Calcutta) reprinted 1972. Delhi: Oriental Publishers.
- Hodgson, Brian H., 1847, The Hodgson Collection. India Office Library, vols. 79, 84, 87.
- Kansakar, Tej Ratna, 1996, Language planning and modernization in Nepal. *Nepalese Linguistics* 13:1–13.
- Mainwaring, George Byrn. 1876, *A grammar of the Róng (Lepcha) language*. Calcutta: Baptist Mission Press.
- Palmer, Frank R., ed., 1970, *Prosodic analysis*. London: Oxford University Press
- Sprigg, Richard Keith, 1959, Limbu books in the Kiranti script. *Akten des vierundzwanzigsten internationalen orientalisten Kongresses, München 1957*, 590–592. Wiesbaden: Morgenländische Gesellschaft.
- 1966, Phonological formulae for the verb in Limbu. In Charles E. Bazell, John C. Catford, Michael A.K. Halliday and Robert H. Robins eds, *In memory of J.R. Firth*, 431–453. London: Longman.
- Tamsang, Khárpú, 1982, *Róng chi-míng 'án lâ-zóng, The Lepcha-English encyclopedic dictionary*. Kalimpong: Mani Printing Works.



Appendix 1: Lepcha Consonants རྗེའི་ལྟོས་ ('ámo re)

k	kh	g	ng	
ཕ	ཁ	ཨ	ཀ	
c	ch	j	ny	
པ	ཅ	ཇ	ཉ	
t	th	d	n	
ཐ	ཌ	ཎ	ཏ	
p	ph	f	b	m
པ	ཕ	ཆ	པ	མ
ts	tsh	z	y	
ཐ	ཅ	ཇ	ཉ	
r	l	h	v	
ལ	ལ	ཨ	ཨ	
	s	sh	w	
	ཨ	ཇ	ཉ	
kl	gl	pl	fl	
ཕ	ཕ	ཕ	ཕ	
bl	ml	hl		
ཕ	ཕ	ཕ		

## Appendix 2: Limbu (Chemjong 1962:21–22)

## Svaravarṇā (vowels)

## स्वर-वर्णा

अ	आ	इ	उ	ए
ऐ	ओ	औ	अः	रः

## Vyañjanvarṇā (consonants)

## व्यञ्जन-वर्णा

क	ख	ग	घ	ङ
च	छ	ज	झ	ञ
त	थ	द	ध	न
प	फ	ब	भ	म
य	र	ल	व	श
ष	स	ह	ज्ञ	

Appendix 3: Limbu (Campbell 1855:Plate I)

<p>The Limboo, or Yakkhung ba Language consists of Twenty</p>	<p>- o āng - ० āng - ॐ ah - ॣ ap.</p>
<p>Eight Letters viz. nineteen Consonants, and nine vowels, which are as follow.</p>	<p>Similar with the Tibetan and Lepcha this language has also a "ya x ka" affixal</p>
<p>Consonants.</p>	<p>thus</p>
<p>Ka Ba Ma Ta Ya 𑄧 𑄨 𑄩 𑄪 𑄫</p>	<p>- २ ya - ॣ Ra The vowels, Finals and ya</p>
<p>Tha Na Sha Nga Sa 𑄬 𑄭 𑄮 𑄯 𑄰</p>	<p>&amp; Ra are thus affixed to the Letters.</p>
<p>Wa Ha Ia Ja Bha 𑄱 𑄲 𑄳 𑄴 𑄵</p>	<p>𑄧 Ka 𑄨 Ke 𑄩 Ke 𑄪 Ke 𑄫 Ke 𑄬 Ke 𑄭 Ke 𑄮 Ke 𑄯 Ke 𑄰 Ke</p>
<p>Kha Ra Chat Nya 𑄶 𑄷 𑄸 𑄹</p>	<p>𑄲 Kai</p>
<p>Vowels.</p>	<p>𑄧 Kāk 𑄨 Kēk 𑄩 Kōk 𑄬 Kīk 𑄭 Kūk 𑄮 Kōk 𑄯 Kōk 𑄰 Kāik</p>
<p>a ē ē ī u ō ō ai 𑄱 𑄲 𑄳 𑄴 𑄵 𑄶 𑄷 𑄸</p>	<p>𑄧 Kāk 𑄨 Kēk 𑄩 Kōk 𑄬 Kīk 𑄭 Kūk 𑄮 Kōk 𑄯 Kōk 𑄰 Kāik</p>
<p>There are also Seven</p>	<p>𑄱 Kām 𑄲 Kēm 𑄳 Kēm</p>
<p>Finals</p>	<p>𑄧 Kāk - 𑄨 Kēk - 𑄩 Kōk</p>





















Appendix 4: Lepcha Final Consonants

Uniting these with  $\mathcal{R}$  a, the basis of all the vowels, they, with their several names and pronunciations, stand thus—

<i>Finals.</i>	<i>Names.</i>	<i>Powers.</i>
$\mathcal{R}^{\acute{a}}$ ak	$\mathcal{W}\bar{\mathcal{E}}$ la kát	k
$\mathcal{R}^{\grave{a}}$ am	$\mathcal{W}\bar{\mathcal{E}}$ la nyat	m
$\mathcal{R}^{\hat{a}}$ al	$\mathcal{W}\bar{\mathcal{C}}\bar{\mathcal{U}}$ la sám	l
$\mathcal{R}^{\circ}$ an	$\mathcal{D}$ nun	n
$\mathcal{R}^{\text{ab or ap}}$	$\mathcal{O}\ \mathcal{E}$ ba kup	b or p <sup>t</sup>
$\mathcal{R}^{\text{ar}}$	$\mathcal{X}$ dar	r
$\mathcal{R}^{\text{at}}$	$\mathcal{E}$ kat	t
$\mathcal{R}^{\text{ang}}$	$\mathcal{E}$ kang	ng
$\mathcal{R}^{\text{ang}}$	$\mathcal{S}\bar{\mathcal{E}}\ (\mathcal{X})$ nyindó	ang

Appendix 5: Limbu Finals (Chemjong 1962:23)

मात्रा तथा संयुक्ताक्षर (Conjoint Letters)

				
t	f	l	r	sh
				
n	p	m	k	ng
				
d	th	n	p	m
				
y	r	l	v	lo

# 20 *Three Tibeto-Burman languages of Vietnam*

---

JEROLD A. EDMONDSON

## 1 Introducing the Tibeto-Burman peoples of Vietnam

In this modest study to honor our esteemed colleague, Dr. James A. Matisoff, I will be focusing on three Tibeto-Burman (TB) languages of Vietnam referred to in Vietnamese language sources as: Phù Lá, Xá Phó, and Lô Lô.<sup>1</sup> The most important contribution this essay makes, I feel, is to present **reliable data** for these languages. To the best of my knowledge there is very little published about any of these much beyond general information such as their names, approximate locations, and some reports of their customs and practices, cf. Lajonquière 1906, Abadie, 1924, DEMG 1975, Lục and Nòng 1975, 1995, Nguyễn 1975, EMPV 1978, Hemmet 1995 and Mai 1995, 1998. To aid in making linguistic data widely available, I have compiled as a companion piece to this paper a comparative word list of about 500 lexical items including the data elicited during my fieldwork sessions on these three as well as four other Tibeto-Burman languages of

---

<sup>1</sup> The research reported on here has been sponsored by a 1995 grant NEH RT-21754-95 from the National Endowment for the Humanities and by the grants SBR 9511285 and SBR9729043 from the National Science Foundation to the author and Dr. Kenneth J. Gregerson all entitled 'Languages of the Vietnam-China Borderlands'. I wish also to acknowledge the assistance of Professors Nguyễn Văn Lợi, Hoàng Văn Ma, and Ta Văn Thang, who arranged and accompanied me on the field trips that led to the data and analysis here. Many thanks as well are due Dr. Pete Unseth, who spent many hours digitising the data from my original tape recordings, and Trần Thuận for help with some of the Vietnamese data. Most of all I wish to thank Graham Thurgood who was able to unlock the system of tonal development in all of these languages. Some of the information about Phù Lá is taken from the MA thesis of my student Robb Fried.

Vietnam that can be accessed on my Web Site //ling.uta.edu/~jerry/. One other contribution I hope to make is to present **current information** about the locations, distribution, populations, and linguistic situation of these groups. All of them have small speaker numbers and inhabit areas of the country very near the China border. As their old people tell it, they immigrated to these places at some time in the near or remote past to escape harsh physical conditions or political turmoil and now live in small numbers in Vietnam. For some of these cases we have concrete evidence in support of their immigrant origins from very closely related groups in China who live nearby. Much about their history, culture, and languages remains unclear though and deserves further study, Nguyễn 1975 and Mai 1995, 1998. A third point I wish to make concerns the **position of these groups** within the larger spectrum of Lolo-Burmese languages, which is the subgroup of Tibeto-Burman to which all of them belong.<sup>2</sup> In pursuing these points I will provide some general information about each of these languages and then note some of their distinctive features.

It is useful to mention that the policy characterising relations between the majority society and small people groups within Vietnam has followed the practices of the Soviets and the Chinese. That is to say, selected groups are granted official recognition with certain rights and privileges; such decisions are often based on historical, cultural, linguistic, and political considerations. As a result, people groups, cultures, and languages do not completely overlap with such officially recognised minorities. So, for example, the TB peoples of Vietnam discussed here are classified into the official nationalities: Phù Lá and Lô Lô. From our own data as well as information gathered from Vietnamese sources, however, it is evident that there exists a greater number of language groups than official ethnicities: Lô Lô has two language groups (one discussed here) and Phù Lá also has at least two groups (both discussed here).

## 2 The Phù Lá nationality

The population of the Phù Lá was determined to be 6,424 in the last official census (April 1, 1989). People of the Phù Lá nationality live mostly in Lào Cai Province and are divided into several different subgroups but treated as one nationality (cf. Map). So, for example, EMPV 1978 distinguished six separate subgroups: (1) the Phù Lá Hán, Phu Kha, or China Phù Lá living at Bắc Hà District at Lùng Phình Commune and Tả Chu Phùng of Mường Khương District, Lào Cai Province, (2) the Phù Lá Lão or Xá Phó living on the southwest bank of the Red River in Sa Pa District and in Lai Châu Province to the west, (3) the Phù Lá Đen or Black Phù Lá living Bát Xát District in A Lù Commune and in Nàn Xi of Xín Mần District, Hà Giang Province, (4) the Phù Lá Hoa or Flowery Phù Lá also living in Bát Xát District in A Lù Commune and also some living with the Xá Phó of Bảo Thắng

---

<sup>2</sup> See Matisoff (1988) for a discussion of this choice of name.

District, Lào Cai, (5) Chù Lá Phù Lá in Lùng Phình of Bắc Hà District, Lào Cai, and (6) the Phù Lá Trắng or White Phù Lá living in Mường Khương District, Lào Cai Province. Twenty years later the description of these peoples and their linguistic and cultural situation had changed considerably. Mai (1995:63) states that in the early 1980's the Phù Lá Đen and the Phù Lá Hoa of Bát Xát District of A Lù Commune returned to China where they once lived and, moreover, the Ph<sup>h</sup> Læ Trắng and Chù Lá Phù Lá were now regarded as belonging to the Phù Lá Hán subgroup. Thus, Vietnamese sources say there are today only two linguistic groups living in Vietnam, the Phù Lá Hán to the north and east of the Red River and the Phù Lá Lão to the south and west of the Red River. Moreover, our informant, an official who was familiar with the linguistic situation in many Phù Lá villages, reported that language shift is highly advanced in many places; younger speakers of Phù Lá are opting to speak other regional languages and Vietnamese, so there is some urgency in getting more information on these groups quickly while there are still some completely fluent speakers available. In this paper we will report on the Phu Kha or Phu Khla language of Bảo Yên District, Lào Cai and also on the Xá Phó or Laghuu language of Sa Pa District, Lào Cai, which are quite different in phonology and lexicon and are certainly not comprehensible in running conversation.

Finally, it is worth noting that there are pictures of the Yi of Yunnan Province in traditional dress on the Web Site //travel.yn.cninfo.net/nyminfq/minzu26/yizu/yizu-e.htm/. Picture 28 entitled 'Yi of Wenshan Prefecture' shows women wearing clothing identical to the Phu Kha of Vietnam. That further confirms the presence of peoples north of the border likely to be culturally and linguistically close to the Phu Kha.

### 3 Phù Lá Hán or Phu Kha Language

This subgroup of the Phù Lá call themselves /p<sup>hu</sup><sup>31</sup> k<sup>ha</sup><sup>33</sup>/ or /p<sup>hu</sup><sup>31</sup> k<sup>h</sup>l<sup>a</sup><sup>33</sup>/ and are often called Phù Lá Hán by neighboring groups. They are said to have arrived in Vietnam in two waves, the first in the 15th century and a later one in the 18th century. The fact that some of the Phù Lá nationality has returned to China recently suggests that there still exist closely related groups north of the border. We were not told, however, that members of this group regularly visit China in search of spouses or to pay visits to relatives or to honor ancestors even though some of these people live very close to the Sino-Vietnamese border. Many Phu Kha speakers know a form of Guanhua (Southwestern Mandarin) Chinese and it is not surprising that we detected a number of modern Chinese loanwords in their lexicon.

As far as what the autonym /p<sup>hu</sup><sup>31</sup> k<sup>ha</sup><sup>33</sup>/ can tell us, we note that the designation [p<sup>hu</sup><sup>31</sup>] 'people' is a very common one for TB groups, especially those belonging to the Eastern Yi nationality in Yunnan. Some autonyms of groups in China are: /p<sup>hu</sup><sup>21</sup>/, /p<sup>ho</sup><sup>21</sup>/, /t<sup>h</sup>l<sup>u</sup><sup>21</sup> la<sup>21</sup> ba<sup>21</sup>/, /p<sup>hu</sup><sup>21</sup> ua<sup>33</sup>/, /a<sup>21</sup> t<sup>h</sup>u<sup>21</sup> p<sup>hu</sup><sup>55</sup>/ and /p<sup>hu</sup><sup>33</sup> va<sup>33</sup>/. There have not been many reports about the Phù Lá before, but one early mention is found in Abadie (1924:182f), who notes that



the Phù Lá are related to the Lô Lô but their vocabulary is clearly different. Another work on the Phù Lá is the report of Lục Bình Thủy and Nông Trung (1975).

### 3.1 *Distinctive features of Phu Kha*

Four tones, stops in four manners of articulation, plain, aspirated, voiced, pre-nasalised (for the labial series) and syllabic nasals suggest a northern Yi heritage for Phu Kha. Further support for this claim come from the fact that Phu Kha possesses two series of clustered initial consonants, a dental series /t̥ t̥ʰ d/ and a velar series /k̥ k̥ʰ gl ŋl/ as well as /p̥/ as in *p̥ca*<sup>31</sup> 'leaf' and *p̥su*<sup>31</sup> 'vomit'.

#### Initials.

p	p <sup>h</sup>	b	mb	f	v	m	w		
pj	p <sup>h</sup> j	bj		fj	vj	mj			
t	t <sup>h</sup>	d		s	z	n	ɬ	l	
t̥	t̥ <sup>h</sup>	d̥							
ts	ts <sup>h</sup>	dz							
t̥ɕ	t̥ɕ <sup>h</sup>	d̥ʒ		ɕ					
tj	t <sup>h</sup> j	dj				nj	ɬj	lj	j
kw	k <sup>h</sup> w	gw		xw					
k̥	k̥ <sup>h</sup>	gl							

Phu Kha has three tones with values 31, 35, 33, as well as a few words with the tone value 53. Several of these tone values can be accompanied with sharp glottal constriction at the end of the syllable. This feature is still easily perceived and has not as yet resulted in a tense/lax voice quality contrast as has happened in many Yi languages. Examples of some of these tonal contrasts are:

ha <sup>31</sup>	'dry plot'	ŋa <sup>31</sup>	'five'	vur <sup>31</sup>	'far'	ci <sup>31</sup>	'gold'	la <sup>31</sup>	'tiger'
k <sup>h</sup> o <sup>31</sup>	'inside'	ja <sup>31</sup>	'chicken'	mu <sup>31</sup>	'horse'	t̥ <sup>h</sup> tu <sup>31</sup>			'face'
ŋur <sup>31</sup>	'ox'	su <sup>31</sup>	'firewood', 'to die'						
sa <sup>35</sup>	'three'	t̥ <sup>h</sup> tu <sup>35</sup>	'four'	nia <sup>35</sup>	'cat'	ŋa <sup>35</sup>	'fish'	su <sup>35</sup>	'teeth'
t̥ <sup>h</sup> a <sup>33</sup>	'one'	ja <sup>33</sup>	'soul'	ŋ <sup>33</sup>	'two'				
ŋu <sup>53</sup>	'needle'								

Pitch plots for these tone shapes are provided below:

Figure 1: Tone plot for Phu Kha / $\eta^{33}$   $\eta^{33}$ / 'egg'

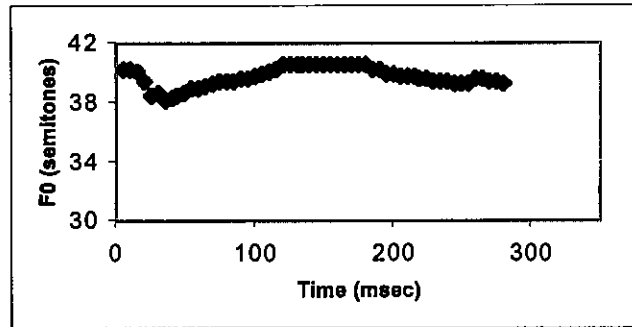


Figure 2: Tone plot for Phu Kha / $m^{31}$   $t\epsilon a^{53}$ / 'wind'

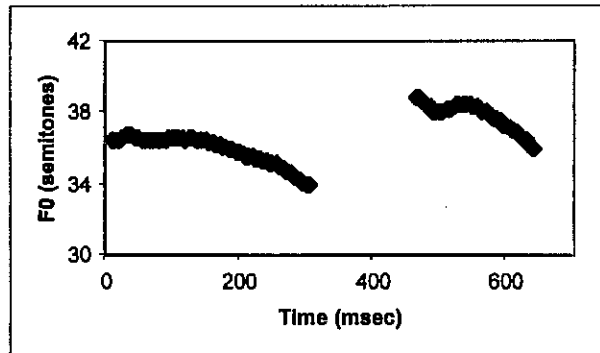
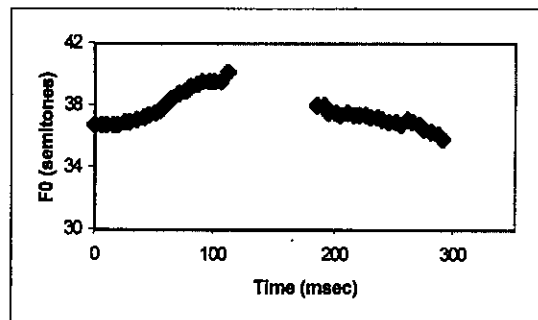


Figure 3: Tone plot for Phu Kha / $m i^{35}$   $k h u^{31}$ / 'ashes'



In regard to syllable rhymes and codas, Phu Kha possesses syllabic nasals:  $\eta^{31}$  'blow',  $\eta^{33}$  'two', and  $\eta^{35}$  'steam' and it preserves the coda /-p/, as well as /-n -ŋ/, which is quite unusual for Northern Yi. Most northern Yi languages demonstrate strong atrophy of the right side of the syllable. Examples of these phenomena are:

ap <sup>35</sup> k <sup>hu</sup> 53 ma <sup>31</sup>	'pumpkin'
tɕap <sup>31</sup> ɕi <sup>31</sup>	'pangolin'
ljap <sup>35</sup> k <sup>ha</sup> 33	'guest'
hai <sup>31</sup> ɕəp <sup>33</sup> k <sup>ha</sup> 33	'husband'

### 3.2 *Comparative notes*

Phu Kha reflects the proto tone category \*1 as 31 as is seen in: barking deer, bear, cat, sparrow, bat, foot, intestine, hammer, alcohol, star, cloud, rain, house, wide, spicy, drink, bark, roar, etc. and also some in 33. Proto tone \*2 is reflected as 35 (or sometimes 31) as is seen in: horse, buffalo, tiger, dog, belly, liver, salt, price. Proto tone \*3 is reflected as 33 as in: moon, dry, full, forget, and wait. The two dead tones are usually realised with tone values 31 and 33.

Dental clusters are an important feature of Phu Kha, because they are not widely attested in Yi languages except in Nisu, one of three main branches of Yi, cf Bradley 1999. According to Chen et al 1985 and Wu 1992, this kind of Yi, the Yiliang E Yi, is classified as Eastern Yi. In fact, Phu Kha appears to be situated close to a special subgroup of Nisu or Nyisu that includes the lectal continuum from Lunan County about 75 km SE of Kunming to the Vietnamese border.<sup>3</sup> The dental-lateral clusters represent a hallmark innovation that defines this subgroup showing a gradient pattern of variation (data from Wu 1992).

This diagram below shows that pre-nasalised and voiced dental clusters are lost and clusters simplified variably with the change having started in the south in the Wenshan/Xichou areas as well as in Sani areas, whereas Yanshan and Lushan Nyisu retain the full set of four manners of articulation. Phu Kha shows family resemblance in being very similar to Phuwa and Phulaba (which are found at locations near to the Phu Kha of Lào Cai Province, Vietnam) in its place in the process of ongoing change within the lectal continuum. This pattern of implicational variation accords well but not perfectly with geographic distribution; Lunan Sani is, for example, out of place, but this unexpected result may be the result of this location being a regional center of contact possibly promoting sound change.

<sup>3</sup> The Lunan Nyisu [ɲi<sup>55</sup> su<sup>33</sup> p<sup>hu</sup>55] are called locally Da Hei Yi (Great Black Yi). They number only 2,000 and live in a mountainous area far from the county seat. Their language is quite distinct from the much larger Sani of the same county. Local gazetteers from the Tang Dynasty onward report that they are descended from the Wuman.

1. Lunan Muzhu [ɲi <sup>55</sup> su <sup>33</sup> p <sup>h</sup> u <sup>55</sup> ] <sup>4</sup>	tʰ	thʰ	ɖɟ	ndɟɰ
2. Yanshan [p <sup>h</sup> o <sup>21</sup> ] and 3. Mengzi [p <sup>h</sup> u <sup>21</sup> ]	tʰ	thʰ	ɖɟ	ndɟ
4. Wenshan Phu Ua	tʰ	thʰ	ɖɟ	ɖɟ
5. Wenshan Phulaba	tʰ	thʰ	ɖɟ	ɖɟ
6. Phu Kha of Lào Cai Vietnam	tʰ	thʰ	dl	dl
7. Xichou [p <sup>h</sup> u <sup>33</sup> va <sup>33</sup> ] and 8. Lunan Sani	tʰ	t	ɖɟ	ɖɟ
9. Wenshan [a <sup>21</sup> tʰa <sup>21</sup> p <sup>h</sup> u <sup>55</sup> ]	tʰ	t	t	t

This entire system represents an innovative development in these lects from proto-Loloish prefixes or initials where the parent language usually had labial clusters or the prefix \*m- or \*ʔ-, cf. Thurgood 1982 and Wu 1992. As the chart below shows the cross lectal variation of wavelike change in the forms of Nyisu of Lunan County ɲi<sup>55</sup> su<sup>33</sup> p<sup>h</sup>u<sup>55</sup> (1), Yanshan p<sup>h</sup>o<sup>21</sup> (2), and Mengzi p<sup>h</sup>u<sup>21</sup> (3), which have a complete system of plain unaspirated, aspirated, voiced, and pre-nasalised voiced stops with lateral clusters; the speech of Wenshan Phuwa (4) and Wenshan Phulaba (5), which have only three of the four initials with stop + lateral sequence, a situation identical to the Phu Kha of Vietnam (6); finally, Xichou p<sup>h</sup>u<sup>22</sup> va<sup>22</sup> (7), Lunan Sani (8) as well as Wenshan a<sup>21</sup> tʰa<sup>21</sup> p<sup>h</sup>u<sup>55</sup> (9), which has simplified these clusters even more. Examples are:

Gloss	Loloish	1, 2, 3	4, 5	6	7, 8	9
'kick'	---	ndɟa <sup>33</sup>	ɖɟu <sup>21</sup>	---	ɖɟo <sup>33</sup>	tʰu <sup>33</sup>
'full'	*m-bliŋ3	ɖɟe <sup>21</sup>	ɖɟu <sup>33</sup>	dlɛ <sup>33</sup>	ɖɟe <sup>33</sup>	tʰɛ <sup>33</sup>
'bee'/'wasp'	*bya2	ɖɟu <sup>33</sup>	ɖɟa <sup>33</sup>	dla <sup>33</sup>	ɖɟa <sup>21</sup>	tʰɛ <sup>33</sup>
'fly'	*byam1	ɖɟɿ <sup>21</sup>		dlu <sup>31</sup>		
'silver'	*plu1	tʰtʰo <sup>33</sup>	tʰtʰo <sup>33</sup>	tʰshu <sup>31</sup>	tʰu <sup>33</sup>	tʰø <sup>21</sup>
'white'	*plu1	tʰtʰu <sup>21</sup>	tʰtʰo <sup>33</sup>	tʰshu <sup>31</sup>	tʰu <sup>33</sup>	tʰø <sup>21</sup>
'face'	*pyu2	tʰtʰo <sup>55</sup>	tʰtʰɿ <sup>33</sup>	tʰtʰur <sup>31</sup>	tʰo <sup>21</sup>	tʰu <sup>21</sup>
'lame'	---	tʰa <sup>33</sup>	tʰa <sup>33</sup>	---	tʰo <sup>33</sup>	tʰu <sup>55</sup>

This regularity was discussed in Thurgood 1982 as an innovation as well (he included all classes of voiced stops, not just the labials). The table above records a larger data set with the gradient devoicing and cluster simplification on display.

That Phu Kha reflects dental clusters in parallel to other Nyisu languages, proto-tones in a complex way, the presence of syllabic nasals, and the four manners of stop articulation (plain, aspirated, voiced, and pre-nasalised voiced) suggests that Phu Kha belongs to northern Loloish, Nisu of the Eastern subbranch and supports Bradley's claim.

<sup>4</sup> This 2000-strong Yi group of Lunan County—point 1—lives in a remote area, and preserves their traditional lifestyle much more than the much larger and better known Sani—point 9—of the same county.

#### 4 Xá Phó, Laghuu, or Phù Lá Lão

The Xá Phó are officially classified in Vietnam as a part of the Phù Lá nationality, but the Xá Phó language is not mutually intelligible with the Phù Lá at other places. Xá Phó speakers are found in Lào Cai Province at Văn Bàn, Bảo Thắng, and Sa Pa Districts as well as in Cam Đường very near to Lào Cai City. A few are also located in Yên Bái Province to the south at Văn Yên and in Sơn La Province near Thuận Giáo City, cf. Map. In this paper we rely largely on data collected from Mr. Mã A Suân of Nậm Sang Commune, Sa Pa District, Lào Cai Province, but we also have some data from a location in Yên Bái Province just across the border from Lào Cai where the phonological system is somewhat different.

##### 4.1 Distinctive features of Xá Phó

Xá Phó has the following set of initial consonants, including plain, aspirates, voiced stops, and pre-nasalised voiced stops.

p	t	ts	tç	k	ʔ
p <sup>h</sup>	t <sup>h</sup>	tsh	tç <sup>h</sup>	k <sup>h</sup>	
b	d	dz	dʒ	g	
mb	nd	ndz	ndʒ	ŋg	
m	n l		ɲ	ŋ	
f	s		ç	x	h
v	z		ʒ	ɣ	

It also has a series of velar plus lateral cluster, /kl k<sup>h</sup>l gl ŋk<sup>h</sup>l/

In comparison to the initials the rhymes of Xá Phó are comparatively simple.

i	ɨ	u	u	
	ə		o	
ɛ			ɔ	
			a	
ɨ	kh <sup>h</sup> ɨ <sup>33</sup>			'four'
i	mi <sup>21</sup> ne <sup>24</sup>			'cat'
u	mu <sup>33</sup> ɲe <sup>33</sup> k <sup>h</sup> u <sup>33</sup> ba <sup>33</sup>			'wind'
u	mu <sup>24</sup> pa <sup>21</sup>			'horse'
ə	ɲə <sup>44</sup> pa <sup>21</sup>			'buffalo'
o	mo <sup>33</sup> tho <sup>24</sup>			'sky'
ɛ	me <sup>21</sup> tç <sup>h</sup> o <sup>54</sup> ma <sup>33</sup>			'window'
ɔ	mo <sup>33</sup> li <sup>21</sup> va <sup>44</sup>			'rain'
a	ma <sup>33</sup> po <sup>33</sup>		ma <sup>44</sup> ma <sup>33</sup>	'bamboo'

Nasal codas are /-m -n -ŋ/. The labial and dental nasal codas occur with much less frequency than the velar.

m	mom <sup>33</sup>	'fishnet'
n	tin <sup>33</sup> ɲaŋ <sup>33</sup>	'field', 'wet'
ŋ	moŋ <sup>31</sup> tɕhu <sup>54</sup> tɕhu <sup>21</sup> pa <sup>44</sup>	'fog'

Xá Phó has only one rising tone with the value 24. The other tones have the shapes 54, 44, 33, and 21

#### 4.2 Comparative notes

We were able to obtain a small amount of data from a second location in Sơn La, which confirms the Lào Cai speech forms, whose autonym is [la21 ɔ44]. The tonal system appears to be identical to Lào Cai, as are many of the initials and finals.

Gloss	Lào Cai /la <sup>21</sup> ɣu <sup>44</sup> /	Sơn La /la <sup>21</sup> ɔ <sup>44</sup> /
'star'	ne <sup>33</sup> ɣu <sup>33</sup> khɿ <sup>33</sup> ba <sup>33</sup>	ma <sup>33</sup> mu <sup>22</sup> kli <sup>22</sup>
'rain'	mɔ <sup>33</sup> li <sup>21</sup> va <sup>44</sup>	mɔ <sup>44</sup> e <sup>55</sup> a <sup>22</sup>
'river'	i <sup>33</sup> dzu <sup>44</sup>	i <sup>21</sup> tɕhu <sup>44</sup>
'stone'	lu <sup>53</sup> khɿ <sup>33</sup>	lu <sup>54</sup> khɿ <sup>44</sup>
'tree'	se <sup>44</sup> ma <sup>33</sup>	se <sup>54</sup> khɿ <sup>54</sup>
'leaf'	se <sup>44</sup> kha <sup>44</sup>	se <sup>22</sup> khɿ <sup>44</sup>

Xá Phó speech forms differ from Phu Kha significantly by not undergoing the developments that have led to the dental cluster series /tʰ thʰ dʰ/, illustrated above in Phu Kha. Xá Phó appears instead to have undergone an innovative change of its own, as the labials, which are the source of /tʰ thʰ dʰ/, have developed into velars, including velar clusters.<sup>5</sup> Therefore, we can conclude that it is not likely that Xá Phó is a member of the Yiliang Eastern Yi subgroup.

While helping us to exclude Xá Phó from the Yiliang branch of Eastern Yi, these developments in Xá Phó can also help us locate where it belongs in the wider picture of Yi languages. Many lexical items of the parent language with original labials including many with additional /l j/ have become velars often with clustered lateral, e.g. \*b-le2 'four' is khɿ<sup>33</sup> or khɿ<sup>44</sup>, \*plu1 'white, silver' is khɿ<sup>33</sup> or khɿ<sup>21</sup>, \*pyu2 'face' is ki<sup>44</sup>, and \*bun1 'shirt' is khɿ<sup>21</sup>. Wu (1992) reports that Wenshan Zuoke shows this same kind of development:

<sup>5</sup> The labial initials are not the only source of velar + lateral clusters in Xá Phó.

Gloss	Xá Phó	Wenshan Zuoke
'excrement'	khj <sup>54</sup>	khj <sup>33</sup>
'silver'	khju <sup>33</sup>	khju <sup>33</sup>
'white'	khju <sup>21</sup>	khju <sup>21</sup>
'fly'	ŋgo <sup>33</sup>	ŋu <sup>21</sup>
'face'	ki <sup>44</sup>	khui <sup>44</sup> du <sup>44</sup>
'bee'	ŋga <sup>33</sup>	ge <sup>21</sup>

Wenshan Zuoke belongs to Southeastern Yi, Wenxi subtype, and is found in a location close to the place where Xá Phó speakers are found today. While the evidence unearthed to date remains somewhat meager, it appears that Xá Phó belongs to the Southeastern Yi subgroup. What is more certain, however, is that Phu Kha and Xá Phó, though belonging to one nationality in Vietnam, come from two differing and relatively distant subgroupings of Yi that have undergone rather different innovations with regard to the prefixed/clustered labials.

## 5 Lolo

The Lolo nationality of Vietnam is divisible into two mutually unintelligible linguistic groups—the Flowery and Red Lolo, on the one hand, who live in Hà Giang Province at Đồng Văn and Mèo Vạc Districts, and the Black Lolo, who live in Cao Bằng Province at Bảo Lạc District. The Flowery Lolo of Hà Giang reside in Xín Cái (Mèo Vạc) and at Lũng Cù (Đồng Văn) within 5 km of the Sino-Vietnam border. The Black Lolo live in the villages Hồng Tri and Đức Hạnh (Bảo Lạc) not far from the Bảo Lạc District capital on the China border. They had a combined population of 3,100 in the 1989 census.<sup>6</sup> We report here on the speech of Ms. Lò Thị Mi of Mèo Vạc, who was 28 years of age at the time we elicited the data. The Red Lolo call themselves [mã<sup>53</sup> tsi<sup>53</sup>]. There is another group in Wenshan Prefecture, Funing County, Muyang Jieshe, Daping Village Cluster, Musang village who speak a very similar language and who call themselves [mo<sup>31</sup> ndzi<sup>31</sup>]. The [mo<sup>31</sup> ndzi<sup>31</sup>] live about 25 km to the north of the Vietnam border and are thought to belong to the Southeastern Yi. The linguistic situation in Wenshan Prefecture is very complex; however, as very many Yi groups reside there, possessing many different groups of Loloish speakers, cf. //yi.peoples.org/. On the basis of the geographic proximity and linguistic resemblance, it is supposed that the Lolo of Vietnam are genetically close to the Lolo of Wenshan prefecture, //yipeople.org/southeastern/s\_lolo.htm/.

<sup>6</sup> Nguyễn (1985) reports that there are a few living in Mường Khương in Lào Cai as well as at Phóng Thổ in Lai Châu Province.

5.1 Special features of Vietnam Lolo

Lolo has the most complex initial system of the three languages reported on, possessing three sibilant series: plain [s ts tsh], retroflex [ʂ tʂ tʂh], and palatal [ç tç tçh] as well as having four kinds of stops plain /p t k q/, aspirated /ph th kh qh/, voiced /b d g/, and prenasalised before a voiceless stop /mp nt ŋk nq/.<sup>7</sup> The complete set of initials is:

p	ph	b	mp	m	f	v	
ts	tsh		nts				
t	th	d	nt	n	s	z	l
tʂ	tʂh	dʂ	ntʂ		ʂ	ʐ	
tç	tçh	dç	ntç	ɲ	ç	ʝ	
k	kh	g	ŋk	ŋ	x	ɣ	
q	qh		nq				

Red Lolo [mã<sup>53</sup> tsi<sup>53</sup>] has five tones with values 53, 31, 312, 44, and 35, which we have not plotted here.

Rhymes are:

i	ε	e	ɔ	o	u	a
		ẽ	õ	õ		ã
ia	ua	iã	uã			
i	tɕi <sup>31</sup>					'teeth'
ε	qhe <sup>44</sup>					'excrement'
e	qhe <sup>312</sup>					'eye'
ɔ	ŋɔ <sup>44</sup>					'duck'
u	ɲu <sup>53</sup>					'cow'
a	tɕa <sup>35</sup>					'boil'
ẽ	qẽ <sup>53</sup>					'ear'
õ	çõ <sup>53</sup>					'nose'
õ	bõ <sup>53</sup>					'belly'
ã	bã <sup>44</sup>					'to fly'
ia	phia <sup>35</sup>					'leaf'
ua	vua <sup>44</sup>					'pig'
iã	piã <sup>35</sup>					'left'
uã	qhuã <sup>44</sup>					'room'

<sup>7</sup> This complex initial corresponds elsewhere in Yi to pre-nasalised voiced stops. Phonetically, however, in Viet Lolo the stop or affricate component is voiceless.



## 5.2 *Comparative comments*

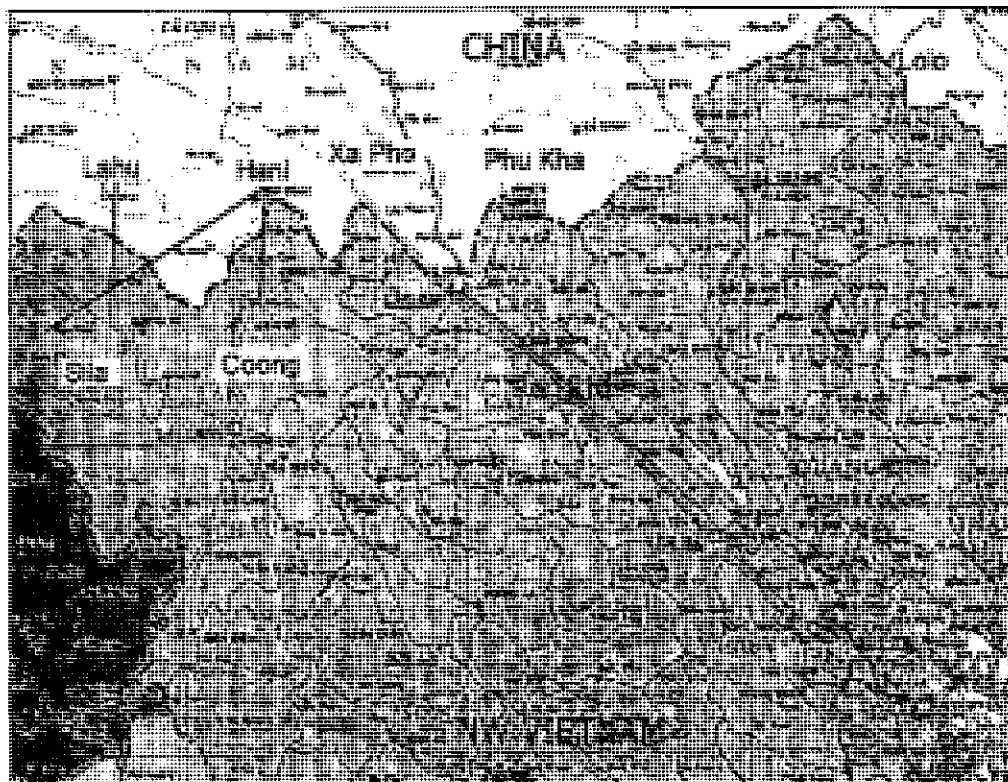
From our data sets it is evident that Lolo is phonologically and lexically quite unlike either Phu Kha or Xá Phó. We believe that Vietnam Lolo (LL) is, in fact, a form of Mo'ang or [muaŋ<sup>51</sup>], which is found in the extreme east of the Yi (Loloish) settlement area in Funing County on the border of Yunnan and Guangxi. This language is called White Lolo (WL) by the local people. Wu 1993, 1994 reports on this language at length, concluding that it differ quite significantly from other forms of Yi of the area. Viet Lolo possesses features more like Mo'ang, which we will focus on here. Mr. Lama Ziwo of our department has recently returned from a field work trip to this area where he found two languages called Lolo in Xinhua Township of Funing County autonyn [ma<sup>33</sup> ŋa<sup>33</sup>], and exonym Long Haired Lolo and in Bailun Township of Funing County autonyn [maŋ<sup>33</sup>] and exonyms White Lolo or Short Trouser Lolo.

The resemblance of Wu's Mo'ang data and Viet Lolo is clear and unquestionable. Some important common features are: uvular consonants /q q<sup>h</sup> ŋq/, diphthongs (Wu treats them as secondary articulations /pj p<sup>h</sup>j mbj tj t<sup>h</sup>j/, where I have transcribed them as diphthongs. There remain considerable differences; notably the former has a pre-glottalised series of initials /ʔm ʔmj ʔn ʔnj ʔl ʔlj ʔʔ/ (LL has none), retains final /-p -t -k -m -n -ŋ/ (LL has none), and has a full set of tense/lax contrasts, which LL may also have. There are also many uniquely similar forms (WL/LL): p<sup>h</sup>jaŋ<sup>33</sup>/p<sup>h</sup>iã<sup>44</sup> 'face', pja<sup>33</sup>/pia<sup>53</sup> 'bee', ʔa<sup>33</sup>/ʔa<sup>44</sup> 'son', kau<sup>33</sup>/kua<sup>31</sup> 'nine', ma<sup>31</sup>/mɔ<sup>44</sup> 'dream', nei<sup>33</sup>/nie<sup>44</sup> 'today'.

Thus, the local exonym, the autonyn, and the geographic location all make it very probable that the Lolo of Vietnam are related to the White Lolo of Funing and neighboring areas. Beyond this evidence is the similarity of phonological inventory (esp. uvular consonant initials), which establish the link between these two. Most importantly, there are examples of unique vocabulary not found at other Yi locations. Despite similarities, it is also evident that White Lolo of Funing possesses even more archaic features than the Red Lolo language of Vietnam in the strong retention of final consonants.

The number of Yi speakers in Vietnam is not great, yet it is evident that they and their brethren across the border in China are all peripheral to the main body of Yi speaker further to the north and west. It is not to be forgotten as well that the Black Lolo of Cao Bằng Province remain to be studied. But even without material on this subgroup, it is clear that through isolation Viet Lolo appears to have preserved some important characteristics that will, one hopes, give us a better understanding of this language group. The dictum that small groups are often more informative about history than larger ones seems to apply here.

**Map:** Tibeto-Burman groups of northern Vietnam.



## References

- Abadie, M., 1924, *Les races de Haut-Tonkin de Phong-Tho à Lang-Son*. Paris: Société d'éditions géographiques, maritimes et coloniales.
- Bradley, David, 1977, Phunoi or Cồng. *Papers in South-east Asian linguistics* no.5, 67-98. Canberra: Pacific Linguistics.
- 1979, *Proto-Loloish*. Scandinavian Institute of Asian Studies Monograph Series no.39. London: Curzon Press.
- 1996, Tibeto-Burman languages of PDR Lao. *Linguistics of the Tibeto-Burman Area* 19/1:19-27.
- 1997, Tibeto-Burman languages and classification in *Papers in Southeast Asian linguistics*, no.14. In David Bradley, ed. *Tibeto-Burman languages of the Himalayas*, 1-72. Canberra: Pacific Linguistics.
- 1999, Language endangerment in China: The Yi of Kunming. *La Trobe Papers in Linguistics* 10:1-15.
- Chen, Shilin, Bian Shiming, Li Xiuqing, 1985, *Yiyu Jianzhi* [Yi language sketch]. Beijing: Nationalities Press.
- DEMG, 1975, *Vấn Đề Xác Định Thành Phần Các Dân Tộc Thiểu Số ở Miền Bắc Việt Nam* [On determining the groupings of ethnic minorities of northern Vietnam]. Hanoi: Ethnic Cultures Publishing House.
- Edmondson, Jerold A. and Lama Ziwo, 1999, Laghuu or Xá Phó, a new language of the Yi group. *Linguistics of the Tibeto-Burman Area* 22/1:1-10.
- EMPV, 1978, *Các dân tộc ít người ở Việt Nam (Các tỉnh phía Bắc)*. [Ethnic minority peoples of Vietnam (the northern provinces)]. Abbreviated reference: EMPV.
- Đỗ Phương, 1996, *Việt Nam Hình Ảnh Cộng Đồng 54 Dân Tộc*. [Vietnam images of the community of 54 ethnic groups]. Hanoi: Ethnic Cultures Publishing House.
- Fried, Robert Wayne, 2000, A preliminary phonological sketch of Phu Kha, a Tibeto-Burman language spoken in Northern Vietnam. MA dissertation, University of Texas at Arlington.
- Hemmet, Christine, 1995, *Montagnards des pays d'Indochine*. Paris: Éditions Sépia.
- Hùng Đình Quy, 1994, *Văn Hóa Truyền thống các dân tộc Hà Giang*. [Traditional culture among the ethnicities of Hà Giang Province]. Hà Giang.
- Lajonquière, Lunet de, 1906, *Ethnographie du Tonkin septentrional*. Paris: E. Leroux.
- Lê Quý Dôn, 1962, *Kiến văn Tiểu Lục*. [Things seen and heard]. Hanoi: first published in the mid 1700s.
- Lò Giàng Páo and Hoàng Nam, 1994, *Truyện cô Lô Lô*. Hanoi: Nhà Xuất Bản Văn Hóa dân tộc hội văn học nghệ thuật Hà Giang.

- Lục Bình Thủy and Nông Trung, 1975, Người Phù Lá ở Lào Cai [The Phù Lá people of Lào Cai]. *Tạp chí Dân tộc học* 1:80.
- Mai Thanh Sơn, 1995, Một vài tư liệu về lịch sử người Phù Lá qua việc tìm hiểu tộc danh [Some materials about the history of the Phù Lá people (work based on the perspective of their names.)]. *Tạp chí Dân tộc học* 3:58–64.
- 1998, Đôi nét về tập quán ăn uống của người Phù Lá [Two features in the customs of eating and drinking of the Phù Lá people]. *Tạp chí Dân tộc học* 1:61–65.
- Matisoff, James A, 1972, *The Loloish tonal split revisited*. Research monograph no.7. Berkeley: Center for South and Southeast Asia Studies.
- 1991, Sino-Tibetan Linguistics: Present state and future prospects. *Annual Review of Anthropology* 20:469–504. Palo Alto: Annual Reviews Inc.
- Nguyễn Văn Huy, 1985, *Văn hóa và nếp sống Hà Nhi, Lô Lô*. [Culture and life style of the Hani and Lolo]. Hanoi: Nhà Xuất Bản Văn Hòa.
- 1975, Bước Đầu Tìm hiểu mối Quan Hệ Tộc Người Giữa Hai Nhóm Phù Lá và <Xã Phó>. [A first step to understanding the differences between the Xa Pho and other Phu La groups]. In *DEMG*, 415–428.
- PV, 1998. Peoples of Vietnam. Darawan Printing: Chiang Mai, Thailand.
- Qiu, Fuyuan, 1998, A phonetic and phonological overview of the Yi (Lolo) language. MA dissertation, The University of Texas at Arlington.
- SDSRV, 1991, Tổng Cục Thống Kê [General Office of Statistics]. *Số liệu thống kê Cộng hòa XHCN Việt Nam* [Statistical data of the Socialist Republic of Vietnam 1976–1990]. Hanoi: Statistical Publishing House. Abbreviated reference: SDSRV 1991.
- Shintani, Tadahiko, et al., 1999, *Basic vocabularies of the languages spoken in Phongsaly, Lao P.D.R.* Tokyo: Institute for the Studies of Languages and Cultures of Asia and Africa.
- Sun Hong Kai et al. eds, 1991, *Zangmianyu yuyin he cihui* [Tibeto-Burman phonetics and vocabulary]. Beijing: Social Sciences Press.
- Thurgood, Graham, 1982, Subgrouping on the basis of shared phonological innovations: a Lolo-Burmese case study. *Berkeley Linguistics Society* 8:251–260.
- VNA, 1996, Thông Tấn xã Việt Nam. [Vietnamese News Agency]. *Việt Nam: Hình Ảnh Cộng Đồng 54 Dân Tộc* [Vietnam: Image of the community of 54 ethnic groups]. Hanoi: Ethnic Cultures Publishing House. Abbreviated reference: VNA.
- Wu Zili, 1992, Nisu Yiyu Sebianyin zai fangyan he qingshu yuyan zhong de duiying [Dental-lateral stops in Yi-Nisu: their correspondences in other Yi dialects and related languages]. In *Yi-Mianyu yanyiu/Lolo-Burmese language research*, 171–183. Chengdu: Sichuan Nationalities Press.

1993, Yunnan Funing Mo'anghua chutan [A preliminary study of Mo'ang speech, Funing County, Yunnan Province]. *Minzu Yuwen* 2:53-63.

1994, Yunnan Sheng Guangnan Xian Kasuhua chutan [A preliminary study of the Kasu language of Guangnan County, Yunnan Province]. *Minzu Yuwen* 2:39-49.