



Pacific Linguistics

Electronic Publication E-4

SEALS XII

**Papers from the 12th
annual meeting of the
Southeast Asian
Linguistics Society
(2002)**

edited by
**Ratree Wayland
John Hartmann
and Paul Sidwell**

Pacific Linguistics
Research School of Pacific and Asian Studies



The Australian National University

Pacific Linguistics E-4

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, David Nash, Andrew Pawley, Paul Sidwell, Jane Simpson

EDITORIAL ADVISORY BOARD:

Karen Adams, *Arizona State University*
Alexander Adelaar, *University of Melbourne*
Peter Austin, *School of Oriental and African Studies*

Byron Bender, *University of Hawai'i*
Walter Bisang, *Johannes Gutenberg-Universität Mainz*

Robert Blust, *University of Hawai'i*
David Bradley, *La Trobe University*
Lyle Campbell, *University of Utah*
James Collins, *Universiti Kebangsaan Malaysia*

Bernard Comrie, *Max Planck Institute for Evolutionary Anthropology*
Soenjono Dardjowidjojo, *Universitas Atma Jaya*

Matthew Dryer, *State University of New York at Buffalo*

Jerold A. Edmondson, *University of Texas at Arlington*

Nicholas Evans, *University of Melbourne*

Margaret Florey, *Monash University*

William Foley, *University of Sydney*

Karl Franklin, *Summer Institute of Linguistics*

Charles Grimes, *Universitas Kristen Artha Wacana Kupang*

Nikolaus Himmelmann, *Ruhr-Universität Bochum*

Lillian Huang, *National Taiwan Normal University*

Bambang Kaswanti Purwo, *Universitas Atma Jaya*

Marian Klamer, *Universiteit Leiden*

Harold Koch, *The Australian National University*

Frantisek Lichtenberk, *University of Auckland*

John Lynch, *University of the South Pacific*

Patrick McConvell, *Australian Institute of Aboriginal and Torres Strait Islander Studies*

William McGregor, *Aarhus Universitet*

Ulrike Mosel, *Christian-Albrechts-Universität zu Kiel*

Claire Moyse-Faurie, *Centre National de la Recherche Scientifique*

Bernd Nothofer, *Johann Wolfgang Goethe-Universität Frankfurt am Main*

Ger Reesink, *Universiteit Leiden*

Lawrence Reid, *University of Hawai'i*

Jean-Claude Rivierre, *Centre National de la Recherche Scientifique*

Melenaite Taumoefolau, *University of Auckland*

Tasaku Tsunoda, *University of Tokyo*

John Wolff, *Cornell University*

Elizabeth Zeitoun, *Academica Sinica*

SEALS XII

**Papers from the 12th annual meeting of
the Southeast Asian Linguistics Society
2002**

edited by
Ratree Wayland
John Hartmann
and Paul Sidwell



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

Pacific Linguistics E-4

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.

Copyright in this edition is vested with Pacific Linguistics
First published 2007

National Library of Australia Cataloguing-in-Publication entry:

Southeast Asian Linguistics Society Meeting (12th : 2002 :

De Kalb, Ill.).

Seals XII : papers from the 12th annual meeting of the
Southeast Asian Linguistics Society, 2002.

ISBN 9780858835788 (pbk.).

1. Southeast Asia - Languages - Congresses. I. Wayland,
Ratree. II. Hartmann, John F. III. Australian National
University. Research School of Pacific and Asian Studies.
Pacific Linguistics. IV. Southeast Asian Linguistics
Society. V. Title.

495

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

<http://pacling.com>

Table of contents

Preface	vii
Note	ix
A look at North-Central Vietnamese <i>Mark J. Alves</i>	1
Genitive constructions in Lai <i>George Bedell</i>	9
Four-word phrases in Lao discourse: <i>yuu⁴ di² mi³ heej³</i> <i>Carol J. Compton</i>	23
Austronesian ergativity traced through two cycles <i>Joseph C. Finney</i>	37
Pragmatic influence on pronouns in Lai (Hakha) Chin, with especial reference to focus and contrast <i>Cue Hlun</i>	79
Sinification of the Zhuang people, culture, and their language <i>Pingwen Huang</i>	89
Orientation origins: the source of Jru' cardinals <i>Pascale Jacq and Paul Sidwell</i>	101
On the so-called abstract nominaliser <i>naak</i> in Lai (Hakha) Chin with remarks upon its other functions in Chin languages and its etymology <i>F.K. Lehman (Chit Hlaing) and Ceu Hlun</i>	107
The Mon-Khmer substrate in Chamic: Chamic, Bahnaric and Katuic contact <i>Paul Sidwell</i>	113
From Malay to Sinitic: the restructuring of Tsat under intense language contact <i>Graham Thurgood and Fengxiang Li</i>	129
Perceptual discrimination of Thai tones by naïve and experienced learners of Thai <i>Ratree Wayland and Susan Guion</i>	137
Tonogenesis in Khmer <i>Ratree Wayland and Susan Guion</i>	145
A comparative study of rice culture words in the Ge-Yang and Kam-Tai language <i>Zhou Guoyan</i>	153

SEALS papers not appearing in this volume:

Heritage language and culture: loss and maintenance
Rhodalyne Crail

Intergeneration transmission of the mother tongue in a minority language community:
a case study of Black Tai in Thailand
Natthida Chakshuraksha

The Laha language and its position in Proto-Kra
Jerry Edmondson

An optimality theoretic approach to Mandarin passive
Nissa Huishan Lin

Peeling back the past: GIS study of the emergence of Tai irrigated rice engineering and
culture
Wei Luo and John Hartmann

Borrowed animal terms in Hmong-Mien languages
David Mortensen

Orientalism and the standardization of spoken Vietnamese into Quoc Ngu
Hanh Nguyen

Historical genetic studies of Southeast Asian languages
Jim Placzek

Crossing the klong: learning inside and outside the classroom SEAsite staff -
www.SEAsite.niu.edu as a tool for teaching Southeast Asian languages and cultures
Herbert C. Purnell

Refinements in fieldwork methodology: case study - comparative linguistic work on Tai
irrigated rice agriculture in Southern China
Vinya Sysamouth

Subgrouping in Kuki-Chin
Kenneth Van Bik

Preface

The Southeast Asian Linguistics Society

History and Goals

The Southeast Asian Linguistics Society (SEALS) was conceived by Martha Ratliff and Eric Schiller in 1990 as a needed forum for the linguists who have the languages of mainland and Pacific Southeast Asia as their primary research focus. It is our hope that the activities of the Society will lead to:

1. greater communication within this group of scholars, especially across the gap which has heretofore divided researchers of mainland Southeast Asian languages and the Austronesian languages of the Pacific;
2. needed publication of descriptive, theoretical and historical accounts of these languages, in the first instance in the form of these proceedings volumes; and
3. greater awareness of these languages by non-specialist linguists, many of whom attempt to make universal and typological generalizations about the human language faculty without the important corrective which knowledge of Southeast Asian languages provides.

To these ends the Society hosts an annual international meeting as the primary means to support these goals. Specific projects, publications, and services beyond those of an annual meeting and the publication of the meeting proceedings will be at the discretion of the members of the Society.

Scope

The Southeast Asian Linguistics Society was founded with the idea of giving language researchers with a 'non-northern' Asian focus a place to share their findings and ideas. In terms of genetic affiliation, investigation into any aspect of Austroasiatic, Austronesian, Hmong-Mien, Tai-Kadai, or Tibeto-Burman languages may be relevant to our members. Although the common thread we recognize in the first instance is geographical, the boundaries of the Southeast Asian area are not clear, and we would not like to be responsible for trying to draw them rigidly. For example, students of languages which have a historical connection to the languages of the area but which are geographically outside and/or typologically unlike those in the Southeast Asian group would be welcome to participate in our meetings and publications as would students of the typologically similar Chinese languages of southern China.

The Twelfth Annual Meeting

The twelfth meeting of the Southeast Asian Linguistic Society took place at Northern Illinois University in De Kalb, Illinois on May 15-17, 2002. There were twenty-eight presentations in all. The keynote speakers were Professor Theraphan L. Thongkham of Chulalongkorn University in Bangkok and Professor Graham Thurgood of California State University-Chico, respectively. Dr. Thongkham spoke on “Old Tai Dam (Black Tai) and the Meanings of Ambiguous Words in Modern Thai Elaborate Expressions.” Dr. Thurgood’s topic was “From Malayic to Sinitic: The Restructuring of Tsat.” His paper is included in this volume.

Acknowledgements

We would like to express our appreciation to the Henry R. Luce foundation for support of two scholars from The Central University for Nationalities in Beijing: Dr. Huang Pingwen and Dr. Zhou Guoyan, respectively. Both of them were participants in a three-year fieldwork project studying Tai origins in southern China as part of the Luce Foundation cooperative research program. Support also came from the Center for Southeast Asian Studies at Northern Illinois University. Julia Lamb, Center Outreach Coordinator did a superlative job in setting up the arrangements for the program. Most importantly, we also would like to extend our gratitude and appreciation to all authors who contributed to the volume for their hard work on the manuscripts, their cooperation and their patience throughout the editing process. We would also like to thank Bradley McDonnell and Dr. Karen Adams at Arizona State University for their work in creating the language list and preparing other introductory material.

Editorial Comment

We have attempted in this volume to standardize the format and appearance of all of the papers. The task was not an easy one in that authors of individual papers work in different platforms (IBM vs. Macintosh), with a variety of IPA fonts (SIL-IPA vs. Mac-dependent STEDT-IPA). Two of the papers employed Southeast Asian language fonts (Vietnamese and Lao). Thanks to the tutelage of Peter Ross and the fastidious editing and formatting skills of Kip Thammavongsa (M.A. Computer Science, NIU), we have produced, we believe, a collection of papers where the publication standards and aesthetics are a source of deep pride and pleasure.

Ratree Wayland
John Hartmann
May 2004

Note

Starting with SEALS VIII, XII & XIII, *Pacific Linguistics* (with generous support of the *Centre for Research in Computational Linguistics*) will be the publisher and the distributor of future volumes of the conference proceedings. This welcome development redresses something of a hiatus in the publication of proceedings volumes, which reflected the general squeezing of resources for Southeast Asian Studies which has affected many programs in recent years.

It was a pleasure for the former Program for Southeast Asian Studies at Arizona State University to have initiated the publication series and to have provided an outlet for the important work of so many linguistic scholars. We would like to thank all past participants in this exciting venture, especially those volunteer efforts that are so essential for conference organization and publishing.

We believe *Pacific Linguistics* will do an exemplary job in the future, and we look forward to the continuation of the series. We are especially pleased to note that under the new publication arrangements SEALS will be distributed free of charge electronically, as well as being available for purchase in print. This initiative is in line with the Society's stated aim of promoting greater communication and awareness of Southeast Asian linguistics. It should be especially helpful to independent scholars, and to researchers located at institutions that do not subscribe to the series.

Karen Adams
Paul Sidwell

A LOOK AT NORTH-CENTRAL VIETNAMESE

Mark J. Alves

Montgomery College, Maryland
<Mark.Alves@montgomerycollege.edu>

1 Introduction

The purpose of this paper is twofold.¹ First, this paper presents phonetic and lexical field data on regional varieties of Vietnamese in North-Central Vietnam, a set of data that is not widely accessible from other sources. Second, it shows how the data, along with other information about other Vietic languages (such highly conservative languages as Rục (Nguyễn V. L. 1993, Nguyễn T. C. 1995), Thavung (Ferlus 1974 and 1979, Hayes 1984 and 1992), and Arem (Trần 1990)), underscore the archaic nature of the speech in this region and the important data for historical linguistic studies in this region.

Vietnamese is often divided into three main regional variants: Northern, Central, and Southern Vietnamese.² The differences between these variants are mainly phonological, but notable differences are evident among non-basic content vocabulary and, though less commonly, even more basic vocabulary and the functional part of the lexicon. In basic and grammatical vocabulary, it is in Central Vietnamese, in contrast with both Northern and Southern Vietnamese, that the differences are most striking. North-Central Vietnamese--generally thought of as the regional variety spoken in the provinces of Thanh Hoá, Nghệ An, and Quảng Bình--has also drawn attention (Emeneau 1951, Ferlus 1998, Guignard 1911, Maspero 1912, Nguyễn T. C. 1995, Thompson 1985), but besides the four general regional divisions, relatively little has been published about other local varieties.³ This paper presents phonological and lexical data on several local varieties of North-Central Vietnamese.

In 1997, I took speech recordings of several varieties of Vietnamese from the North-Central province of Nghệ An, in addition to representative standard speech from Hà Nội, Huế, and Hồ Chí Minh Cities. The data were collected by the use of (1) wordlists aimed primarily at identifying a full range of the phonetic realizations of tones and (2) sentence lists that provided ways to find distinctive etymologies.⁴ The data show that, although North-Central Vietnamese (hereafter, NCV) cannot be easily grouped with either Northern or Central Vietnamese and shares certain phonological and lexical characteristics with both, there are a few details that suggest a closer connection to Central Vietnamese. In addition, this paper deals with the language/dialect/accent continuum in Vietnam and the historical implications of linguistic data with respect to regional diversity and linguistic affiliation. NCV shows both conservative (segmental and lexical) characteristics and innovations (notable variety in realizations of tones). The sections in this paper include discussion on (1) dialect versus regional accent and vocabulary, (2) the genealogical position of North-Central Vietnamese, (3) some significant field data on NCV, and (4) the implications of that data. Table 6 following this article presents a tentative summary of tone systems amongst the varieties of Vietnamese studied.

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society* (2002). Canberra, Pacific Linguistics, 2007, pp.1-7.

2 Dialect versus Regional Accent/Vocabulary

The three traditional regional dialects, namely Northern Vietnamese (NV), Central Vietnamese (CV), and Southern Vietnamese (SV), have as their generally accepted regional centers Hà Nội, Vinh, Huế, and Hồ Chí Minh City (cf. Nguyễn Đình Hoà's Vietnamese-English dictionary (1966) and Thompson's Vietnamese grammar (1985)). Dialect neutralization has taken place along the well-traveled coastline,⁵ with some convergence towards regional center standards, while the restricted inner highlands have, as seen in data for this paper, maintained much smaller regional linguistic distinctions.

As the distinction between language, dialect, and accent is often not a clear one, it is necessary to consider the difference in Vietnamese. The difference between a dialect and simply a regional accent is the difference of more basic vocabulary, morphology, and syntax, though not so different as to cause significant interference in communication. Central Vietnamese has a significantly lower degree of comprehensibility to both speakers of Northern and Southern Vietnamese, a fact both linguistically predicted and anecdotally supported. In Vietnam, such is also the case for NCV, which has not only notable phonological distinctions (especially of tones), but also distinctions in basic vocabulary (such as 'knee' and 'head') and even distinctions in the etymologies for various grammatical vocabulary. This supports the claim that NCV is a distinct dialect with various local varieties.

It should also be noted that NCV is regarded as a kind of curiosity. It has been posited an archaic pocket by Maspero (1912) and later by Nguyễn T. C. (1995). In addition to the archaic linguistic features of rural NCV, there are also speakers of the closely related but more conservative Mường and highly conservative Minor Vietic languages. While Vietnamese has no complex initial consonant clusters, Mường does of obstruents plus [l]. More significantly, the Minor Vietic languages, with their more basic or even incipient tone systems, have complete presyllables with Mon-Khmer morphology. While this raises the issue of language contact as a source of mutual influence, in fact, put together, this scenario makes this region of Vietnam unique with regards to the historical background of Vietnamese, having all three subbranches of Vietic in one small region.

3 The Linguistic Position of North-Central Vietnamese

North-Central Vietnamese, as represented by speech in the city of Vinh, can be historically grouped with Central Vietnamese, as represented by the speech of Hue, due to their primary similarities in the unmarked tones and both grammatical and content vocabulary. The 'even' tones (*ngang* and *huyền*) generally have the same contour, as in Table 1. Also, they share several notable grammatical lexical items, as in Table 2.

Table 1: *Phonetic realizations of 'level' tones*

	NV	SV	CV	NCV
Ngang tone	33	33	35	35
Huyền tone	21	21	33	33

Table 2: Vietnamese lexical variants

Gloss	NV	SV	CV	NCV
‘thus’	vậy	vậy	rũa	rũa
‘how’	sao	sao	răng	răng
‘where’	đâu	đâu	mô	mô
‘this’	này	này	ni	ni

Nonetheless, NCV maintains a modern dialectal distinction from CV since the segmental phonological distinctions are great enough for interference in interregional communication (as discussed in section 4). There are, in addition, some notable lexical distinctions, for example, differences within the region for the word expressing negation (*nó* ‘no/not’, as opposed to several other mainstream alternatives *không*, *chẳng*, and *chả*) and the reflexive (*chắc* ‘each other’ versus standard *nhau*), spoken in Thành Chương Vietnamese in Nghệ An province (Alves and Nguyễn forthcoming).

4 Segmental Characteristics

NCV is phonologically conservative, maintaining more segmental phonemic distinctions than any of the other three main regional varieties of Vietnamese. In Table 3, the regional variations are shown with respect to Vietnamese Quốc Ngữ orthography (QN hereafter), which represents the original maximal number of phonemic distinctions in Vietnamese.

Table 3: Realizations of Quốc Ngữ in Regional Varieties

QN	NV	NCV	CV	SV
s	s	ʃ	ʃ	ʃ
x	s	s	s	s
tr	c	ʈ	ʈ	ʈ
ch	c	c	c	c
r	z	ɹ	ɹ	ɹ
d	z	ʒ	j	j
gi	z	z	j	j
v	v	v	j	j
-nh	ɲ	ɲ	n	n
-n	n	n	n/ɲ	n/ɲ
-ng	ŋ	ŋ	ŋ	ŋ
-ch	c	c	t	t
-t	t	t	t/k	t/k
-c	k	k	k	k

In geographic terms, the number of phonemic distinctions decreases either north or south of the region where NCV is spoken. Table 3 shows a geographic range of distinctions among varieties of Vietnamese, with NCV in the center. Whereas NCV has distinct phonemes for all the QN sounds, NV, CV, and SV all show mergers as indicated by the lack of lines between certain cells in the table.

5 Field data on NCV

Linguistic data were collected through acoustic phonetic recordings using SIL's Winccil.⁶ Speakers were given word lists (monosyllabic and bisyllabic words and short phrases and sentences). NCV shows a significant amount of both segmental and tonal variation, mostly restricted to specific lexical items, though some represented differences in phonemic systems. Some lexical differences were found between the rural and metropolitan varieties of central Vietnamese, as shown in Table 4. The tone values are indicated by the numbers 1 to 5, which show the starting and finishing points of the tones, while 'g' indicates glottalization accompanying tones. In Tables 4, the regional variants from Vinh downward are rural NCV variants. Major cities are named as cities, while the remainder of the examples show small townships (*xã*) preceded by their subregional districts (*huyện*).

Table 4: NCV lexical variants for 'lighter'

'lighter' bật lửa			
Location	QN	IPA	Tone Value
Hà Nội City	bật lửa	bət liə	22-31
Huế City	bật lửa	bək liə	22-31g
Nha Trang City	bật lửa	bək liə	22-35
Vinh City	bật lửa	bət liə	22-31g
Thành Tướng, Thành Tướng	bật lò	bət lə	22g-31g
Nam Đàn, Nam Trung	mậy lả	me la	22-31g
Nghi Lộc, Nghi Ân	bật lo	bət lə	22-31
Nghi Lộc, Nghi Hưng	mậy lửa	me lə	35-31
Nghi Lộc, Nghi Khanh	bật lửa	bət liə	22-31g
Nghi Lộc, Nghi Lâm	mậy lửa	məj liə	31g-31g

As another indication of the high degree of dialectal variation, Table 5 shows diphthongization and additional vocalic mutation. Nghi Ân township is especially conservative; it leaves unrounded the pronunciation of words with back vowels followed by final velars, such as the Sino-Vietnamese word *không* "not/nebulus" pronounced in Nghi Ân as /k^hoŋ/ instead of the standard /xauŋ^m/.

Table 5: *Nghi Lộc, Nghi Ân Speech*

Gloss	QN	NV	Nghi Ân	Category
‘dishonest’	gian	zan	juən	a:ə
‘series’	loạt	lwat	luət	
‘outside’	ngoại	ŋwaj	ŋuəj	
‘toss’	quáng	kwaŋ	kwəŋ	
‘uncle’	bác	bak	bək	a:ɔ
‘gold’	vàng	vaŋ	vɔŋ	
‘gait’	dáng	zaŋ	zɔŋ	
‘exit’	ra	ra	rɔ	
‘must’	phải	faj	fɔj	
‘left’	trái	ʈaj	ʈɔj	
‘ghost’	ma	ma	mɔ	
‘to pass’	quá	kwa	kɔ	
‘river (Sino-Vietnamese)’	hà	ha	hɔ	

Another lexical difference was strictly urban versus rural. The word for ‘older sister’ is in the data consistently *chị* in the major cities, while the rural areas showed variously *ả* and *o*.

Finally, in field data, the tones in this region show a very high amount of phonetic variation according to township, as can be seen in Table 6 (at the end of this article). Despite the small geographic region studied, the varieties of NCV show substantial linguistic differences. Most tone variation is among the phonetically non-level *trắc* tones, though even the level *bằng* tones show some variation.⁷ There is a phonetic distinction in some varieties between the *sắc* tone in open and syllables with final voiceless stops.

6 Implications

This paper has presented various phonetic and lexical data on North Central Vietnamese, which is, though closer typologically to Central than to Northern Vietnamese, a distinct main dialect of Vietnamese with local variation even at the township level. NCV shows the largest number of segmental phonemic distinctions and amount of regional phonetic variation, particularly tones. Moreover, a high amount of archaic vocabulary and phonemic preservations further highlight the conservative aspects of NCV and fill in some of the blanks that link Vietic with other Mon-Khmer languages.

While not directly stated, these dialectal distinctions are reflective of the historical spread and differentiation of modern Vietnamese. Historical records show that during the Chinese Han dynasty, there existed in modern North and North-Central Vietnam the historical regions of Giao Chi and Cửu Chân respectively (Taylor 1983), corresponding to the north versus central dialect division in Vietnam, as noted by Ferlus 1999. Thus, NCV maintains a significant geographic position between these two historically significant regions and near both Mường and Minor Vietic languages. The amount of regional variation is interesting as it identifies the known time depth of the speech in the area for diversification to have occurred.

Table 6: *Tone Contour in Varieties of Vietnamese*⁸

Area	City/Town	Ngang	Huyền	Sắc (Open)	Sắc (Closed)	Nặng	Hỏi	Ngã	
North	Hà Nội City	33	21	24	45	22g	31	35g	
South	Hồ Chí Minh city	33	21	45		23	24		
Central	Huế City	35	33	13g	45	22	31g		
	Vinh City	35	33	11	11/55	22	31	13g	
	Nam Đàn, Nam Trung	35	33	13g	13g/45	22	31		
	Thành Chương, Thành Tróng	35	33	11g/13g		22g	31		
	Nghi Lộc, Nghi Khanh	35	33	55			53		44
	Nghi Lộc, Nghi Hưng	44	33	45			22	31	31/13g
	Nghi Lộc, Nghi Lâm	35	33	13g	45	22	31	31/13g	

Notes

1. The data for this paper was collected with the invaluable assistance of Mr. Nguyễn Duy Hương of the Institute of Linguistics in Hà Nội. However, all mistakes in this paper of content, presentation, and ideas are mine alone.
2. See Thompson 1985 and Friberg 1973. The French colonial period saw the division of Vietnam into the three regions recognized as dialectal boundaries though administrative boundaries have existed in North Vietnam since the Han dynasty, two millennia ago (see Taylor 1983 for historical details and Ferlus 1999 on the relationship between those ancient divisions and modern dialectal differences).
3. Two such works are Vương 1981 and Hoàng 1989.
4. Thanh Chương Vietnamese data (Alves and Nguyễn) were presented at the Eighth SEALS conference held in Kuala Lumpur in 1998.
5. Acknowledged in personal communication with Vietnamese linguists.
6. The phonetic software was downloaded at www.sil.org.
7. The *bằng* (meaning 'level') category refers to the mid-level/*ngang* and mid-falling tones/*huyền* tones, while the *trắc* (meaning 'uneven') category refers to all other Vietnamese tones, which have more distinctive contours and sometimes glottalization.
8. This table is certainly tentative as the data collected were based on very few subjects. Moreover, the issues of describing phonemic representations of Vietnamese tones are complicated by other factors (cf. Alves 1995 and Pham 2003).

References

- Alves, Mark. 1995. Tonal Features and the Development of Vietnamese Tones. Working Papers in Linguistics, Vol. 27, University of Hawai'i.
- Alves, Mark and Nguyen Duy Huong. forthcoming. Notes on Thanh Chuong Vietnamese in Nghe An Province. Proceedings of the Eighth Annual Meeting of the Southeast Asian Linguistics Society, Mark Alves ed.
- Emeneau, Murray B. 1951. Studies in Vietnamese (Annamese) grammar. Berkeley.
- Ferlus, Michel. 1974. Problèmes de mutations consonantiques en Thavung. *Bulletin de la Société de Linguistique de Paris* 69:311-323.
- _____. 1979. Lexique Thavung-Français. *Cahiers de Linguistique Asie Orientale* 5:71-94.
- _____. 1998. Les systèmes de ton dans les langues Viet-Muong. *Diachronica* XV.1:1-27.
- _____. 1999. Les disharmonies tonales en Viet-Muong et leurs implications historiques. *Cahiers de Linguistique - Asie Orientale* 28.1:83-99.
- Friberg, Barbara. 1973. Generative phonology as applied to Vietnamese dialects: a study based on middle Vietnamese, comparing the three major dialects of modern Vietnamese. Saigon University MA thesis.
- Guignard, Péro T. 1911. Note sur une peuplade des montagnes du Quảng-Bình: Les Tác-Cũ. *Bulletin de L'École Française d'Extrême Orient* 11:201-205.
- Hayes, La Vaughn H. 1984. The register systems of Thavung. *Mon-Khmer Studies* XII:91-122.
- _____. 1992. Vietic and Việt-Mường: a new subgrouping in Mon-Khmer. *Mon-Khmer Studies* 21:211-228.
- Hoàng, Th, Châu. 1989. Tiếng Việt trên các miền đất nước (phương ngữ học) (Vietnamese in regions of our land (dialect studies)). Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Maspero, Henri. 1912. Études sur la phonétique historique de la langue Annamite: les initiales. *Bulletin de l'École Française d'Extrême-Orient* 12:1-127.
- Nguy-n, Đình Hoà. 1966. Vietnamese-English Dictionary. Rutland, Vermon: Charles E. Tuttle Co.
- Nguy-n, Phú Phong. 1988. Lexique Vietnamien-Rục-Français. Université de Paris VII.
- Nguy-n, Văn Lị. 1993. Tiếng Rục (The Ruc language). Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Nguy-n, Tài Cn. 1995. Giáo trình lịch sử ngữ âm tiếng Việt (Textbook of Vietnamese historical phonology). Hà Nội: Nhà Xuất Bản Giáo Dục.
- Pham, Andrea Hoa. 2003. Vietnamese Tone A New Analysis. New York: Routledge.
- Taylor, Keith W. 1983. The Birth of Vietnam. Berkeley: University of California Press.
- Thompson, Laurence. 1965. A Vietnamese grammar. Seattle: University of Washington Press.
- Trần, Trí Dối. 1990. Nhận xét về thanh điệu trong thổ ngữ Arem (Some notes of the tones of the Arem dialect). *Tạp Chí Khoa Học* 1990.20:37-40.
- Vương, Hữu Lễ. 1981. Vài nhận xét về đặc điểm của vần trong thổ âm Quảng Nam ở Hội An (Some notes on special qualities of the rhyme in local Quang nam speech in Hoi An). *Một Số Vấn Đề Ngôn Ngữ Học Việt Nam (Some Linguistics Issues in Vietnam)*: 311-320. Hà Nội: Nhà Xuất Bản Đại Học và Trung Học Chuyên Nghiệp.

GENITIVE CONSTRUCTIONS IN LAI¹

George Bedell

International Christian University
<gdbedell@yahoo.com>

Lai has a set of pronominal particles, shown in (1), which accompany nouns and mark agreement with a preceding noun phrase.

(1)		singular	plural
	1st person	<i>ka</i>	<i>kan</i>
	2nd person	<i>na</i>	<i>nan</i>
	3rd person	<i>a</i>	<i>an</i>

These pronominal particles appear identical to the set of pro-nominal particles which accompany verbs and mark agreement with the subject.² The second and third person plural pronominal particles are illustrated in (2) and (3) accompanying a noun, and in (4) and (5) accompanying a verb.

(2) *an zal chung khan* ‘out of their bags’ (2:11)

(3) *nan zal ah* ‘in your bags’ (10: 9)

(4) *annah an kal hnuah* ‘after they went’ (2:13)

(5) *nan kal* ‘you go’ (23:15)

Such parallelism between noun phrases and clauses is not uncommon in the languages of the world; the question naturally arises as to what accounts for it.

One possibility often suggested is a parallelism in the overall syntactic organization of clauses and noun phrases. It has been argued in earlier work that subject-verb agreement markers in Lai are the syntactic heads of clauses, and the verb is attached to them, as illustrated in (v).³

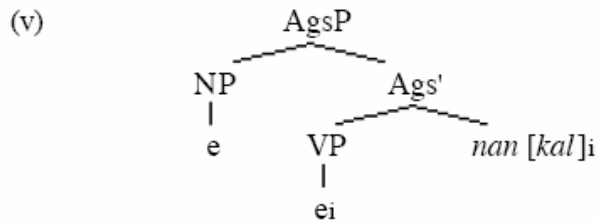
¹ Lai (often called (Hakha) Chin) is a Tibeto-Burman language spoken primarily in central Chin State, Myanmar. The examples accompanied by a chapter and verse notation are taken from the Gospel according to Matthew in the 1999 Lai Bible. I am grateful to Kenneth Van Bik and Stephen Hre Kio for assistance with some points of Lai grammar.

² See Bedell (1995) for a discussion of subject-verb agreement in Lai.

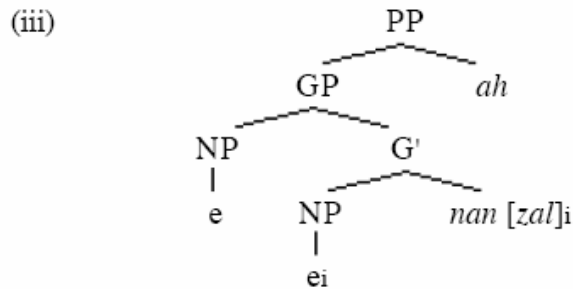
³ See the papers cited in Bedell (1999) and (2000).

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society* (2002). Canberra, Pacific Linguistics, 2007, pp.9-21.

© George Bedell



If phrases like (2) and (3) are parallel to those like (4) and (5), the structure of (3) might be as in (iii).



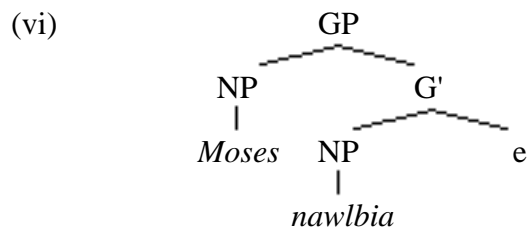
The empty NP in (v) occupies the subject position, and is understood as second person plural. The first empty NP in (iii) occupies the corresponding position; it too is understood as second person plural. *Nan* in (v) belongs to the category Ags (Subject Agreement), which is the head of this clause structure; Ags and AgsP appear as I and IP (where I represents the more general Inflection) in some versions of this terminology. *Nan* in (iii) belongs to the category G (Genitive); it too is the head of the given structure. In some versions of this terminology used for the discussion of English, G and GP appear as D and DP (where D represents the more general Determiner). In Lai, as will be argued below, the projection of G represented in (ii) is distinct from that headed by deictic elements. The meaning expressed by the pronominal particles in examples like (2) and (3) is possession, but the syntactic structure in (iii) is by no means restricted to possession. In its semantic variability also, that position resembles the subject position of a clause as in (v). The term ‘genitive’ will be used in general for the construction in (iii), to emphasize its syntactic nature, and the empty NP in (ii) will be said to occupy the genitive position.

In (6) and (7), the genitive position is filled with a noun phrase.

(6) *Moses nawlbia* ‘the law of Moses’ (7:12)

(7) *profet hna cawnpiaknak* ‘the teachings of the prophets’ (7:12)

That is, the structure of (6) is something like (vi).



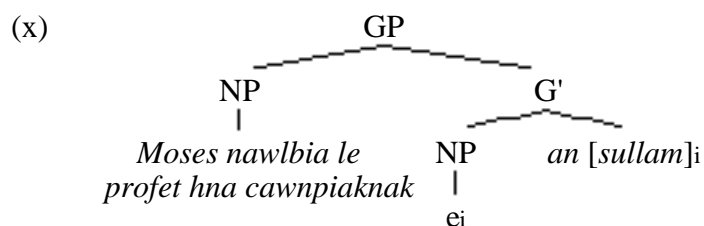
Thus (6) and (7) have a different structure than the semantically parallel (8) and (9).

(8) *a nawlbia* ‘his law’

(9) *an cawnpiaknak* ‘their teachings’

Particles like *a* and *an* have very different syntactic properties from noun phrases like *Moses* and *profet hna* ‘the prophets’. The difference is clear in examples where both a particle and a noun phrase co-occur, as in (10).

(10) *Moses nawlbia le profet hna cawnpiaknak an sullam* (7:12)
 ‘the meaning of the law of Moses and the teachings of the prophets’



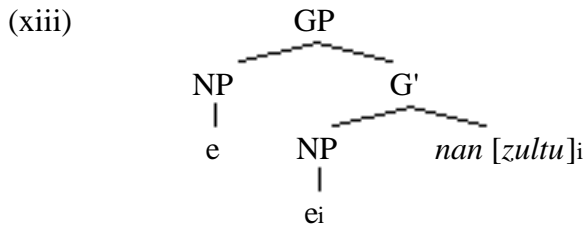
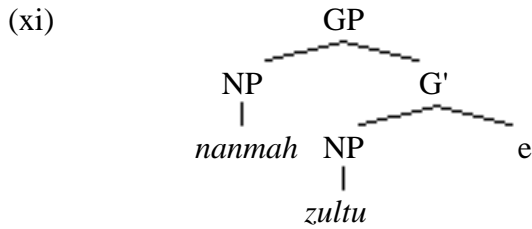
Pronouns represent a special case of noun phrases, and they may appear in the genitive position, as in (11) and (12).

(11) *nanmah zultu* ‘your disciple(s)’ (12:27)

(12) *anmah zultu tlawmpal hna* ‘a few of their disciples’ (22:16)

The structure of (11) is parallel to (vi), and quite different from that of (13) parallel to (iii), despite their similarity in meaning.

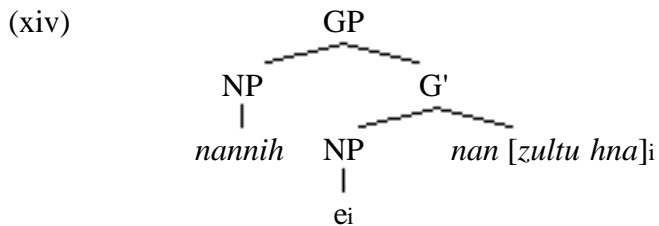
(13) *nan zultu* ‘your disciple(s)’



Again, this is clear from examples like (14) in which a pronoun in the genitive position and the corresponding pronominal particle co-occur.

(14) *nannah nan zultu hna* ‘your disciples’ (12:27)

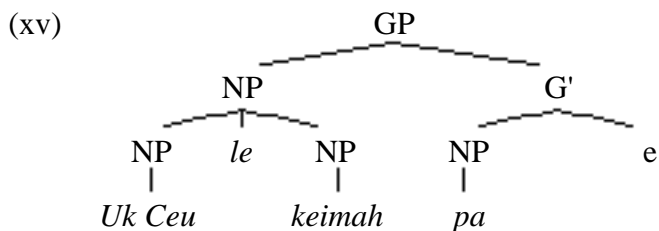
The structure of (14) is parallel to (x).



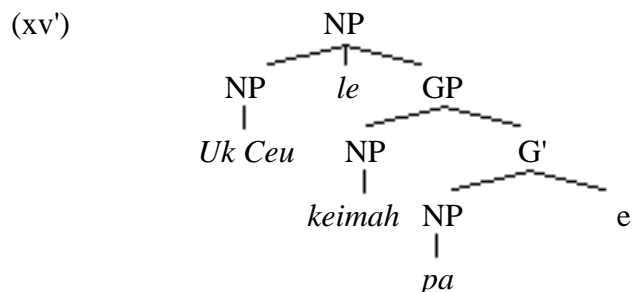
An additional argument showing clearly the distinct syntactic status of genitive particles and noun phrases, even though they are usually adjacent as in (10) or (14), comes from conjunction. Phrases like (15) are potentially ambiguous.

(15) *Uk Ceu le keimah pa* ‘Uk Ceu’s and my father’
 ‘Uk Ceu and my father’

In the first reading, the conjunction involves Uk Ceu and keimah ‘I’; the structure will be as in (xv).



In the second reading, the conjunction involves *Uk Ceu* and *keimah pa* ‘my father’; as in (xv’). The ambiguity arises because of the absence in Lai of a genitive marker like English ‘-’s’.



In an example like (17), only the second interpretation is possible; a noun phrase like *Uk Ceu* cannot conjoin with a particle like *ka*.

(17) *Uk Ceu le ka pa* ‘Uk Ceu and my father’

In fact, genitive particles cannot conjoin at all.

(18) **a le ka pa* ‘his and my father’

It may be worth observing that the meanings conveyed by the genitive constructions in examples (6) through (14) do not include possession. In the case of (7), the relation between the genitive noun phrase *profet hna* ‘the prophets’ and *cawnpiaknak* ‘teachings’ is often called the ‘agent’ relation. *Cawnpiaknak* is a nominalized form of the verb *cawnpiak* ‘teach’, and in a clause containing that verb it will be the subject which has that relation. In (11) on the other hand, the relation between *nanmah* ‘you’ and *zultu* ‘disciple’ is often called the ‘patient’ relation. *Zultu* is a nominalized form of the verb *zultu* ‘follow’, and in a clause with that verb it will be the object which has that relation. In (10), the noun *sullam* ‘meaning’ is not a nominalization, and the semantic relation between that noun and the genitive *Moses nawlbia le profet hna cawnpiaknak* ‘the law of Moses and the teachings of the prophets’ is different from either agent or patient. In the case of (6), the relation between *Moses* and *nawlbia* ‘the law’ may appear to resemble the agent relation as in (7), but it is rather vaguer; we do not know exactly what it is if we are not familiar with the Bible story of the Ten Commandments. The meaning which characterizes the genitive position in a genitive construction depends in large part on the meaning of the head noun, but some details may depend on the particular genitive noun phrase or the context. This situation is parallel to the interpretation of clause subjects, and found in all languages.

If the analysis just proposed is correct, then the Lai pronominal particles in (1) are agreement markers when they accompany nouns, just as they are when they accompany verbs. In the former case they mark agreement between the head noun and a noun phrase in the genitive position. Nonetheless there are differences in the two cases which it will not do to ignore. For one thing, a clause generally contains a subject in addition to a verb, so that finite verbs are always accompanied by an agreement particle. In the presence of

agreement, the subject position may often be empty, as in (iv). Noun phrases, however, do not require that the genitive position be present, so that most nouns are not accompanied by any agreement marker. Like the subject position, the genitive position may be empty in the presence of genitive agreement, as in (ii) or (xi). Secondly, an overt clause subject is echoed by the corresponding agreement marker on the verb; but as illustrated in examples like (6), (7), (11) and (12), the genitive position in a noun phrase can contain an overt noun phrase without any agreement particle accompanying the head noun. Examples like these are quite ordinary, with examples like (10) or (14) relatively unusual.

There is in addition a restriction on the use of pronouns as genitives. Lai has two sets of pronouns as in (19) and (20), which are more or less interchangeable in many contexts.

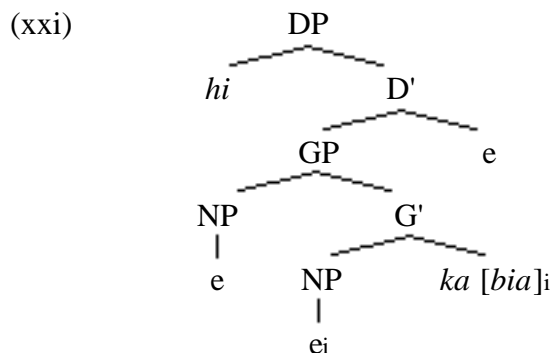
- (19) *kei* 'I' *kannah* 'we'
 nang 'you' *nannah* 'you'
 anih 'he/she' *annah* 'they'

- (20) *keimah* 'I' *kanmah* 'we'
 nangmah 'you' *nanmah* 'you'
 amah 'he/she' *anmah* 'they'

Genitive pronouns are usually from the second set, in which the element *-mah* originally meant 'self'. The more basic pronouns in (19) sound odd, though there are a few examples in the Bible; (14) is one such.

As mentioned above, Lai differs from English in treating deictics and genitives separately. A clear example appears as (21), with the structure shown in (xxi).

- (21) *hi ka bia* 'these words of mine' (7:24)



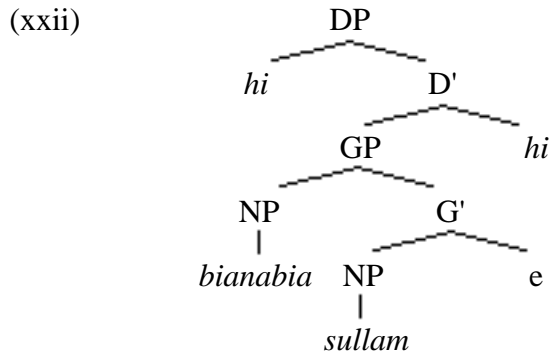
According to the analysis given in Bedell (1999), *hi* is located under DP as shown. In English a demonstrative like 'these' and the genitive pronoun 'my' appear to occupy the same position, since they cannot co-occur *'these my words', *'my these words', and (21) can be glossed in English only by using a post-nominal 'of' phrase.⁴ Since the genitive

⁴ Phrases like 'these my words' may be possible in some styles of English. If so, they do not fit the usual assumed structure.

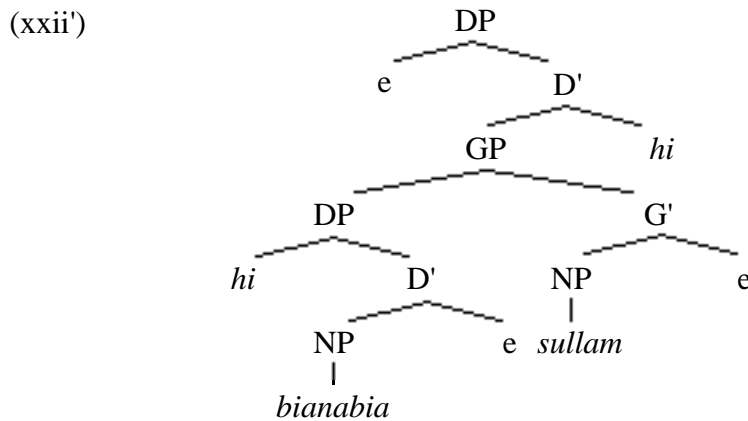
position and the head D position are empty in (21), it may appear possible to place *hi* in the former, obviating the need for distinct G and D projections.

In example (22), both the genitive position and the head D position are filled, and the structure will be as in (xxii).

(22) *hi bianabia sullam hi* ‘this meaning of the parable’ (13:18)

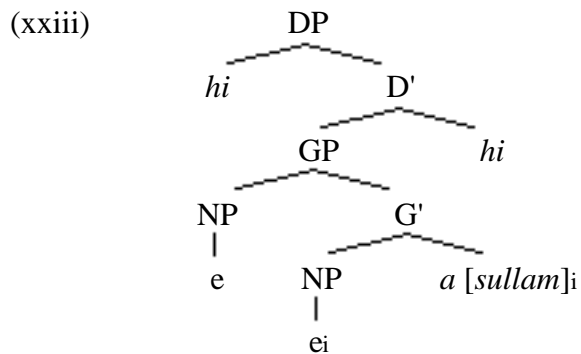


(22) may not be conclusive, since it would be possible to take the first *hi* to be accompanying the noun *bianabia* ‘parable’ rather than the noun *sullam* ‘meaning’. If so, the gloss ought to be ‘the meaning of this parable’. In that case the structure would be (xxii’) rather than (xxii), and the D and G projections might still be collapsed.



In examples like (23) to (25), however, it is clear that both the head position of G and the head position of D are filled and therefore there must be two distinct projections. The structure of (23) will be as in (xxiii).

(23) *hi a sullam hi* ‘this meaning of it’ (24: 15)



(24) *ka fapa hi* ‘this son of mine’ (17: 15)

(25) *nan sayapa hi* ‘this teacher of yours’ (9: 11)

This is clear also from genitive constructions in which the head noun is understood in context. This position cannot be left empty as in English, but must be occupied by the dummy noun *ta*. Such constructions may have a genitive particle as in (26), a genitive pronoun as in (27), or other genitive noun phrases as in (28) and (29).

(26) *na ta* ‘yours’ (25:25)

(27) *nanmah ta* ‘yours’ (5: 3)

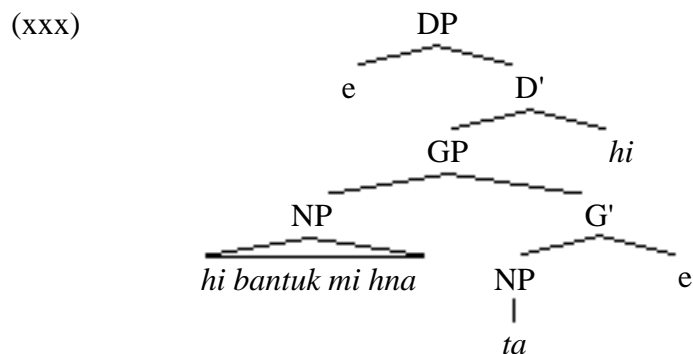
(28) *Pathian ta* ‘God’s’ (22:21)

(29) *siangpahrang ta* ‘the king’s’ (22:21)

In an example like (30), even though no agreement marker is present, the G projection must be distinct from the D projection, as shown in (xxx).

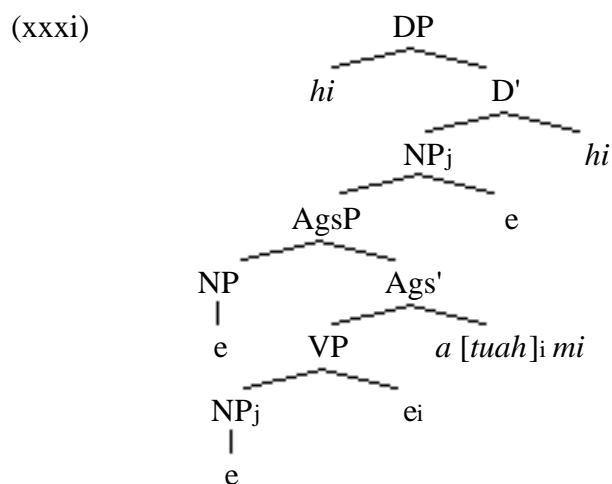
(30) *hi bantuk mi hna ta hi* ‘of people like these’ (19:14)

In this example the initial deictic *hi* probably has to be taken as attached to *bantuk* rather than to the entire phrase, since *bantuk* requires a deictic. Thus (xxx) is parallel to (xxii’) rather than to (xxii).



Example (31) resembles (23), except that the head noun *tuahmi* ‘action, deed’ is derived from the verb *tuah* ‘do’. The structure might be the same as (xxiii) with *tuahmi* replacing *sullam*. But it also might be analyzed as (xxxi), containing a kind of relative clause.

(31) *hi a tuahmi hi* ‘these doings of his’ (8:17)



In this case, the particle *a* accompanies a verb rather than a noun, and thus marks agreement of the verb *tuah* with its empty subject rather than (as in the analyses parallel to (xxiii)) of the noun *tuahmi* with its empty genitive, and the gloss should be ‘these things which he did’. It is difficult to decide which of these two analyses is preferable. Examples like (32) and (33) provide evidence in favor of the genitive analysis.

(32) *hi tleipi kung cung i ka tuahmi hi* ‘these doings of mine to the fig tree’ (21:21)

(33) *Cucu Bawipa tuahmi a si.* ‘That is the Lord’s doing.’ (21:42)

In (32), the postpositional phrase *tleipi kung cung i* ‘to the fig tree’ has the postposition *i* (used to modify nouns) rather than *ah* (used in adverbial modification). In (33), no agreement marker appears, a characteristic of genitive constructions and not of clauses. On the other hand, an example like (34) cannot be analyzed as a genitive construction.

- (34) *thratnak a tuahmi he thratlonak a tuahmi he* (5:45)
 ‘both those who do good and those who do evil’

Examples like (35) and (36) also favor a relative clause analysis.

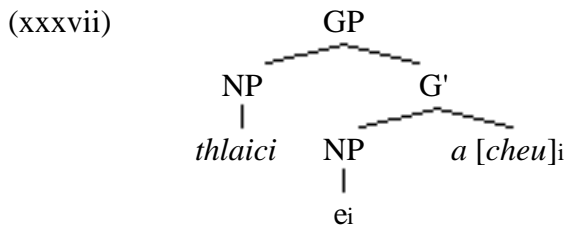
- (35) *nan chung i khuaruahharnak ka tuahmi vialte hi* (11:21)
 ‘all the miracles which I performed in you’
- (36) *na chungah ka tuahmi khuaruahharnak vialte hi* (11:23)
 ‘all the miracles which I performed in you’

In (36) the relative clause is *na chungah ka tuahmi* ‘which I performed in you’ modifying the head noun *khuaruahharnak* ‘miracle’, while in (35), the relative clause would seem to be *nan chung i khuaruahharnak ka tuahmi*, a head-internal structure.⁵

Genitive constructions are used in Lai with a partitive meaning, as in (37) or (38).

- (37) *thlaici a cheu* ‘some of the seed’ (13: 8)
- (38) *phungbia cawnpiaktu saya a cheukhat* ‘some of the teachers of the law’ (9: 3)

The predominant pattern with *cheu* ‘part’ is to have genitive agreement, whether the genitive noun phrase appears overtly or is interpreted in context. It is a noun which is restricted to the head position of a genitive construction; the structure of (37) is (xxxvii).

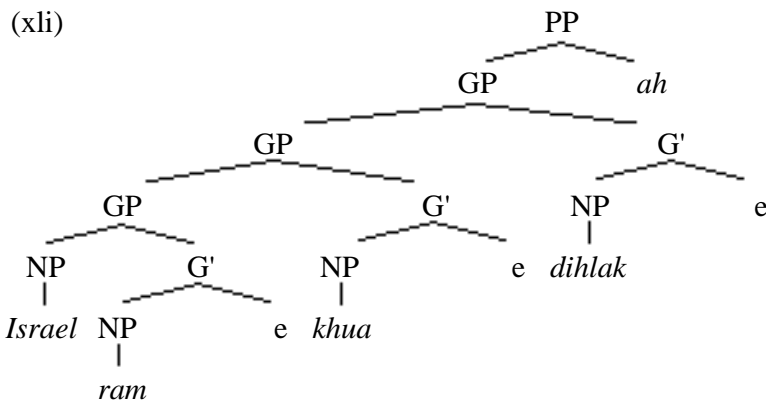


A similar situation obtains for *dihlak* ‘all’, as in (39) to (42).

- (39) *a dihlak in* ‘all of it’ (6:22)
- (40) *nan dihlak in* ‘all of you’ (18:18)
- (41) *Israel ram khua dihlak ah* ‘to all of the towns of Israel’ (10:23)
- (42) *na lungthin dihlak le na ruahnak dihlak le na nunnak dihlak in* (22:37)
 ‘with all your heart and all your thoughts and all your life’

⁵ See Kathol and Vanbik (1999) for a discussion of head-internal relatives in Lai.

Dihlak does not require genitive agreement, but it is restricted to a postpositional phrase which may modify another noun phrase, or be adverbial. The structure of (41) will be as in (xli).



The Lai word *dang* ‘other’ seems to belong with this group of nouns also. Like *cheu* it seems restricted to the head position of a genitive construction, though not clearly partitive in this case.

(43) *tahchunhnaak bia a dang* ‘another parable’ (13:33)

(44) *tahchunhnaak bia a dang pakhat* ‘one other parable’ (13:31)

(45) *Mari dang pakhat* ‘another Mary’ (27:61)

(46) *a dang Mari* ‘the other Mary’ (28: 1)

As analyzed in Bedell (2000), Lai has a large subclass of nouns whose meaning is expressed in English by different prepositions. English objects of such prepositions correspond to Lai genitives on those relational nouns. The full range of genitives appear, as illustrated in (47) through (58).

(47) *na sinah* ‘to you’ (14:28)

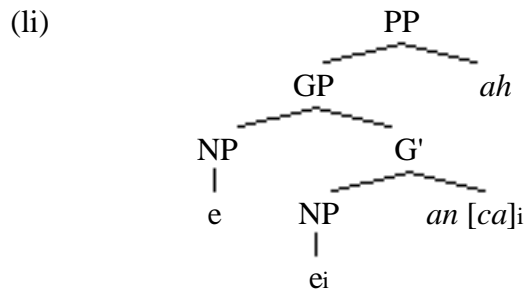
(48) *na sin i* ‘to you’ (26:35)

(49) *nan chung* ‘your insides’ (23:28)

(50) *nan chung i* ‘in you’ (6:23)

(51) *an caah* ‘for them’ (18: 6)

In (47) to (51), genitive agreement appears with no overt genitive noun phrase. The structure of (51) will be as in (li).



(52) *Josef sinah* ‘to Joseph’ (2:13)

(53) *amah sinah* ‘to him’ (4:24)

(54) *lawng chungah* ‘in the boat’ (4:21)

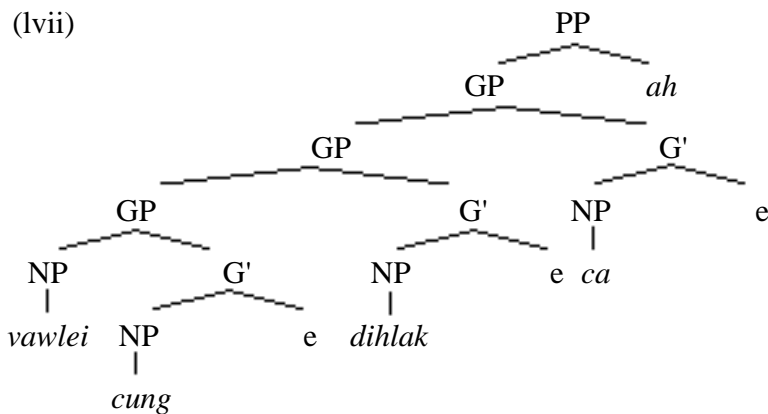
(55) *nangmah chung (in)* ‘out of you’ (2: 6)

(56) *miphun vialte hna caah* ‘for all people’ (5:13)

(57) *vawlei cung dihlak caah* ‘for all on earth’ (5:14)

(58) *nang caah pakhat, Moses caah pakhat, Elijah caah pakhat* (17: 4)
 ‘one for you, one for Moses, one for Elijah’

In (52) to (58), there is a genitive noun phrase with no agreement. In (53), (55) and one of the conjuncts in (58) those noun phrases are pronouns. In the last case the pronoun belongs to set (19) instead of the usual (20). The structure of (57) will be as in (lvii).



Finally, Lai has a large subclass of compound verbs of which the first member is a noun, and on which one argument is marked as genitive agreement. Many of these compounds are ‘psycho-collocations’ as described in Van Bik (1998).

- (59) *ka rawl a tram* ‘I was hungry’ (25:35)
- (60) *ka zang an fak* ‘I pity them’ (15:32)
- (61) *na khuk naa bil i* ‘you kneel’ (4: 9)
- (62) *nan thinphang hlah u* ‘do not be afraid’ (10:31)
- (63) *an khuaruah a har* ‘they were amazed’ (7:28)
- (64) *an thin a hunter ngaingai hna* ‘it made them very angry’ (15:12)

Unlike other genitive constructions, it is not possible to have a genitive noun phrase replacing the agreement as in (65) or (66). There can be a noun phrase, as in (67) or (68), but it is the subject of the verb and not a genitive.

- (65) **keimah rawl a tram*
- (66) **anmah khuaruah a har*
- (67) *keimah (cu) ka rawl a tram* ‘I was hungry’
- (68) *mi zapi cu ... an khuaruah a har* ‘the crowd was amazed’ (7:28)

Exactly what the structure of such compounds should be is not clear, but it is certainly their morphological nature which precludes full syntactic expression. Some speakers say that it is possible to have (65) or (66) when ‘pragmatically contrived’.

References

- G. Bedell, 1995, ‘Agreement in Lai’, *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society*, Tempe, Arizona: Program for Southeast Asian Studies, Arizona State University, pp. 21-32.
- G. Bedell, 1999, ‘The Syntax of Deixis in Lai’, presented to the 32nd ICSTLL, Urbana, to appear in *Linguistics of the Tibeto-Burman Area*.
- G. Bedell, 2000, ‘Postpositions and Relational Nouns in Lai’, presented to the Fifth International Symposium on Languages and Linguistics, Ho Chi Minh City, Vietnam.
- A. Kathol and K. Vanbik, 1999, ‘Morphology-Syntax Interface in Lai Relative Clauses’, *NELS* 29 pp.427-41.
- K. Van Bik, 1998, ‘Lai Psycho-collocations’, *LTBA* 21:1. pp. 201-33.
- (D. Van Bik), ‘Matthai sinin Thawngthra’ (the Gospel according to Matthew), *Lai Baibal Thiang* (The Holy Bible in Lai), Bible Society of Myanmar, centennial edition, 1999.

FOUR-WORD PHRASES IN LAO DISCOURSE:

yuu⁴ dii² mii³ hɛɛŋ³

Carol J. Compton

Madison, Wisconsin

<compton@facstaff.wisc.edu>

In his 1975 article, “Rhyme, Reduplication, etc. in Lao,” G. Edward Roffe presented and discussed an extensive collection of polysyllabic phrases in Lao which may occur either in oral or written discourse. Using material from Roffe’s article and some of his field notes, as well as material from my own field work, I focus on the types of four-word phrases that occur in the Lao data and discuss aspects of the patterns found in them. Such knowledge of four-word phrases and related patterns is seen as significant in developing a deeper understanding of Lao discourse and poetics.

1 Introduction

In this article I focus on four-word phrases in Lao in which each word in the phrase can function independently in some way, but when woven into a phrase produces a meaning which is different from or more intense than smaller sections of the phrase.¹ From my perspective, such four-word phrases are basic components of the discourse structure and poetics of many Tai languages, certainly so in Lao. These phrases occur with high frequency in oral language and are also found in written language. In Lao, they are found not only in everyday speech, written short stories, folk tales and epics, but also in newspapers and magazine articles. The modern short story writer and the traditional *mohlam* singer make use of these phrases because they are a basic part of the language. This weaving of words into four-word phrases is a process which is evident in most Tai languages; I believe all or nearly all of them use these phrases as intellectual building blocks, expressing old concepts with set four-word phrases and creating ways to reflect new ideas using the productive four-word phrases. Not only four-word, but five-word and six-word phrases fill similar positions in many of these languages. In this paper, I focus specifically on the four-word phrases.

2 Four-Word Phrases in Lao

For Lao, the first discussion of four-word phrases in the English literature appears to be in the 1975 article, “Rhyme, Reduplication, etc. in Lao,” by G. Edward Roffe. In that article, he presented and discussed an extensive collection of polysyllabic phrases in Lao which may occur either in oral or written discourse

In his article, Roffe presented four major categories of four-word expressions: Fixed expressions, adaptation of foreign terms, reduplication, and rhyming. The fixed expressions he refers to show a “semantic relationship” of the individual words but do not involve reduplication or rhyme. The criteria for a phrase to be considered a “fixed expression” appear to be that the four words are frequently found in a specific order in a phrase with a general meaning related to the meanings of the individual words. An example of a fixed expression is:

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society* (2002). Canberra, Pacific Linguistics, 2007, pp. 23-35

© Carol J. Compton

sap ⁴	sin ¹	ŋən ³	thɔɔŋ ³
treasure	wealth	silver	gold

with the meaning given by Roffe as ‘riches.’²

The adaptation of foreign terms is a rather small category and includes some repetition. An example of such an item is:

miɪ³ nit⁴ miɪ³ nat⁴ from the French ‘minister’; the phrase means ‘cabinet minister.’

Roffe’s third category of four-word phrases, reduplication, is very common and has many variations. Such phrases may contain combinations which include nonsense syllables in part of the phrase or uncommonly used expressions. For Roffe, reduplication may be of exact words or similar sounds. This form has many productive patterns.

The fourth category, rhyming, is a pattern which is quite frequent and also very productive. Usually in this pattern, syllables two and three of the four-word phrase rhyme.

In his paper, Roffe notes that “the basic structure of these expressions is a polysyllabic phrase with an equal number of syllables on either side of an imaginary vertical dividing line” (285). In the data on which my paper is based, I have chosen to focus my attention on the four-word phrases.

3 The Data

In this section, I will present some of the patterns and examples of four-word phrases from field notes and connected discourse in Lao; these are from the beginning stages of my current project on indigenous aspects of Lao grammar. Eventually I hope to be able to tell more about the use, distribution and frequency of occurrence of four-word phrases in Lao. First, we will look primarily at some of the variations and similarities we find with the Lao patterns; then, we will discuss briefly similar patterns found in a few other Tai languages. Finally, we will comment on what appear to be similar patterns in Vietnamese, Khmer and Hmong.

Examples of four-word phrases from G. Edward Roffe’s field notes were collected.³ Hundreds of such phrases occur in his material. From a continuous section of 2,100 lexical entries, 100 items were four-word phrases. I limited the selection of phrases to those having only four words, and I have categorized them in a somewhat different manner than did Roffe. Of these 100 four-word phrases, I found that 70 had a repetition pattern, 22 had a rhyming pattern and 8 involved semantic linking. The 70 phrases involving repetition can be further divided into five types.

1. First of all, there are those phrases in which words 1 and 3 are exactly the same and words 2 and 4 share a **clear semantic** link. Of the 70, 34 phrases fell into this category. For example: *khii⁵ loŋ¹ khii⁵ lʰim³* ‘absent-minded’.
2. Secondly, there are those phrases in which words 1 and 3 are exactly the same and 2 and 4 share a **vague semantic** link. Of the 70 phrases having repetition, 13 were of this type.

3. In this group, **repetition with compounds**, words 1 and 3 are again the same while 2 and 4 are words that are often used as a compound, but in the phrase are separated by the repeated words. There are 17 of these phrases.
4. In the next group, **repetition with opposites**, 1 and 3 are exactly the same, but 2 and 4 are opposites in meaning or opposed in meaning in some way. There are five such phrases.
5. Just one example was found of this final type of repetition, **double reduplication**. In this case words 1 and 2 are the same and represent one compound meaning ‘different’ while words 3 and 4 are the same and represent another compound meaning ‘various.’ Roffe glosses this phrase as ‘all sorts of things.’

The following table illustrates the types of repetition found in the 100 entries from Roffe’s field notes. In the first four types, words 1 and 3 are identical. In the fifth type, 1 and 2 are the same word and 3 and 4 are the same word.

Table 1: *Types of repetition in Lao four-word phrases*

1.	0	0	0	0	
	1	2 _{cs}	3	4 _{cs}	(where <i>cs</i> stands for <i>close semantic link</i>)
2.	0	0	0	0	
	1	2 _{vs}	3	4 _{vs}	(where <i>vs</i> stands for <i>vague semantic link</i>)
3.	0	0	0	0	
	1	2 _c	3	4 _c	(where <i>c</i> stands for an element of a <i>compound</i>)
4.	0	0	0	0	
	1	2 _{op}	3	4 _{op}	(where <i>op</i> stands for an <i>opposing</i> meaning)
5.	0	0	0	0	
	1	1 _r	2	2 _r	(where <i>r</i> stands for a <i>reduplicated</i> word)

In the next table, examples are presented of each type of repetition found among those hundred phrases. All of the examples in the table below are from Roffe’s field data.

Table 2: *Examples of repetition in Lao four-word phrases*

1.	Repetition with Close Semantic Link				
	thuun ³	hua ¹	thuun ³	kaw ⁶	
	lift up	head	lift up	head	‘adore’
	0	0	0	0	
	1	2 _{cs}	3	4 _{cs}	(close semantic link)

2. Repetition with Vague Semantic Link
 taam² mii³ taam² kəət
 as have as born ‘according to one’s means’
 0 0 0 0
 1 2vs 3 4vs (vague semantic link)
3. Repetition with a (Split) Compound
 pen² thaa⁴ pen² thaaj³
 be manner be way ‘(in the) accepted manner’
 0 0 0 0
 1 2c 3 4c (elements of a compound)
 thaa⁴ thaaj³ = ‘manner, character’
4. Repetition with Opposites
 kuay⁴ hua¹ kuay⁴ haaj¹
 put head put tail ‘alternate’
 crosswise crosswise (head to feet, as lying down)
 0 0 0 0
 1 2op 3 4op (opposing elements)
5. Double Reduplication
 taaj⁴ taaj⁴ naa³ naa³
 different various ‘all sorts of things’
 0 0 0 0
 1 1r 2 2r (r stands for a reduplicated word)

The following table illustrates the types of internal rhyme found in the four-word phrases from Roffe’s data. As I noted earlier, 22 of the 100 four-word phrases in his data showed a rhyming pattern. In 19 of these phrases standard internal rhyme, in which words two and three rhyme, was found. Three phrases were found to have variant rhyme. In two of these phrases, words two and four rhymed; one phrase had rhymes between words one and three.

Table3: *Internal rhyme in Lao four-word phrases*

1. *Standard internal rhyme two-three rhyme*

0	0	0	0	
1	2ry	3ry	4	(ry stands for a <i>rhyming</i> word)
viak ⁶	baan ⁶	kaan ²	hian ³	
work	house;	work	house	‘domestic affairs’
	village			
thon ³	khom ¹	ʔom ²	vaan ¹	
endure	bitter	keep	sweet	‘take the bitter with the sweet’
		in mouth		

2. Variant internal rhyme two-four rhyme

0	0	0	0	
1	2ry	3	4ry	(ry stands for a <i>rhyming</i> word)
pen ²	kaan ²	pen ²	ŋaan ³	
be	work	be	work	‘be businesslike’
tham ³	haay ⁶	haaŋ ⁴	kaay ²	
do	bad	form	body	‘assault’

One-Three Rhyme

0	0	0	0	
1ry	2	3ry	4	(ry stands for a <i>rhyming</i> word)
hen ¹	phit ³	pen ²	səp ⁶	
see	error	be	to like	‘mistake bad for good’

Table 4 illustrates the semantic linking found in four-word phrases that do not have specific repetition of words. These are phrases which are linked primarily by the meanings of the individual words; alliteration and rhyme may be found, but it is the semantics that dominates. Eight such phrases were found among the 100 four-word phrases from Roffe’s material.

Table 4: *Semantic Linking in Lao Four-Word Phrases*

0	0	0	0	
1	2	3	4	(<i>semantically linked</i> words)
hua ¹	nəɯn ³	paay ²	tin ²	
head	sleep	tip	feet	‘background of a person’
vaw ⁶	nam ⁶	thuum ⁵	thoŋ ⁴	
speak	water	flood	field	‘talk a blue streak’
səp ⁶	saa ⁶	ŋoo ⁴	ŋaaw ⁴	(with <i>alliteration</i>)
stupid	stupid	ignorant	foolish	‘blundering’
puu ⁴	ŋaa ⁴	taa ²	ŋaay ³	(with <i>internal rhyme</i>)
paternal		maternal		
grandfather,		grandfather,		
grandmother		grandmother		‘ancestors’

Note that though there are three main patterns of linkage discussed above for four-word phrases in Lao, there can be overlapping of these patterns. Many four-word phrases may be found to have at least two of these linkage characteristics (repetition, rhyme, or semantic links); some may have all three. As Roffe notes in his article on this topic, he long sought “some underlying principle governing the construction of expressions of this kind”; though he failed to find it, he did suggest that “it may lie buried deep in the subconscious” (1975: 285). Though one can find patterns of linguistic linkage weaving such phrases together, clear criteria for what constitutes a four-word phrase need to be

developed, taking into account information on how native speakers distinguish and interpret such phrases.

4 Similar Phrases in Lao Discourse Data

Having compiled the list of 100 examples of four-word phrases from Roffe's field data, I sought four-word phrases in connected discourse in a number of Southern Lao *mohlam* performances which I had translated (Compton 1975; 1979; 1992) to see how they fit with the three major types of four-word phrase patterns resulting from the above analysis. Examples of the patterns found in that material are presented in the following table.

Table 5: *Repetition examples from Lao discourse*

Pattern: Repetition with close semantic link

0	0	0	0	
1	2cs	3	4cs	
hɔɔt ⁶	hian ³	hɔɔt ⁶	law ⁶	(1992: 242)
reach	house	reach	granary	'reach...your house and granary'
con ²	hɔɔŋ ⁶	con ²	hay ⁵	(1992: 243)
until	cry	until	cry	'until (one) weeps'
tɛɛ ⁴	kii ⁶	tɛɛ ⁴	kɔɔn ⁴	(1979: 206)
since	first	since	before	'from the beginning'

Pattern: Repetition (with alliteration)

0	0	0	0	
1	2	3	4	
con ²	cuan ²	con ²	cɛɛŋ ⁶	(1979: 234)
until	almost	until	bright	'nearly dawn'

Table 6: *Rhyme examples from Lao discourse*

Pattern: Internal rhyme

0	0	0	0	
1	2ry	3ry	4	
ʔaan ⁴	hen ¹	pen ²	day ⁶	(1975: 71)
read	see	be	able	(we) 'can understand (it)'

Table 7: *Semantic link examples from Lao discourse*

Pattern: Semantic link (with variant rhyme)

0	0	0	0	
1	2ry	3	4ry	
din ²	daaw ⁶	faay ¹	laaw ³	(1992: 242)
earth	forest;	side	Lao	'Laos'
	region			

In my own data, the four-word phrases with the repetition pattern were found most often, probably because many of them are productive and thus give the *mohlam* flexibility. As I commented in an earlier paper, “These flexible four-word expressions are thus useful to the *mohlam*, who can manipulate them creatively to fit the variety of topics and performance situations they face” (Compton 1992:233). Examples with internal rhyme or semantic links as the dominant pattern, however, were found less frequently in this material.

5 Similar Phrases in Other Tai Languages

Four-word phrases showing the same or similar patterns to those discussed for Lao can be found in many Tai languages. Below I present examples of such patterns found in the work of those who have been looking at this phenomena over the past century. These examples of four-word phrases are from Shan, Thai, Black Tai, and Lue; examples from other Tai languages can also be found, but these will suffice to illustrate this point. I will begin with examples from Shan material that was originally published in 1914.

Shan Examples:

In his “Introduction” to the phonetic version (2000) of Cushing’s 1914 *Shan-English Dictionary*, Hudak notes that Cushing mentions two types of elaborate expressions, which he calls “double phonetic couplets.”

In the first, a four-syllable expression is created with the following: 1 - a word, 2 - an open syllable with the initial consonant of the word plus the vowel [i], 3 - a repetition of the word, and 4 - the usual phonetic couplet of the word....The open syllable always has the third tone regardless of the one on the original word.... The second type derives from compound nouns and verbs. With this type, the phonetic couplet is separated from its principal word by the repetition of the first syllable of the compound (xvii-xviii).

Table 8: *Repetition in Shan four-word phrases or “double phonetic couplets”*

Pattern: Repetition with alliteration

shuk ⁵ shi ³	shuk ⁵ shak ⁵	(xvii)	
confused C+I	confused (usual phonetic couplet)		‘confused, disordered’

Pattern: Repetition with compound

maak ² keŋ ⁴	maak ² kaŋ ⁴	(xviii)	
fruit tamarind	fruit (usual phonetic couplet)		‘tamarind’
maak ² keŋ ⁴ = ‘tamarind’			

The examples from Shan in the table above differ from the Lao examples presented earlier in that the Shan examples use syllables which may or may not have individual meanings; four-word phrases with “double phonetic couplets” found in the Lao data were not analyzed for this paper. There do appear to be some examples in Hudak’s phonetic version of Cushing’s dictionary which have the double reduplication found in the Lao phrases.⁴

Thai Examples:

In the introduction to her *Thai-English Dictionary* (1964), Haas says that elaborate expressions, her term for phrases similar to the Lao four-word phrases discussed in this paper,

are usually colloquial but a few are considered elegant. They are frequently based on compounds.... and are expanded by repeating a part of the compound and adding a new part, by inserting a syllable for the sake of rhyme, or by inserting a syllable which has some vague semantic relation to one of the original parts” (xvii).

Haas notes that most such expressions “are made up of four parts.” She identifies three basic types of these expressions: semi-repeated expressions, expressions characterized by rhyme, and those that have a semantic linkage.

Table 9: *Thai patterns and examples from Haas (1964)*

Semi-repeated expressions: (those having the first and third word the same)

0	0	0	0	
1	2	1	3	
rûab	hũa	rûab	hãaŋ	(xvii)
to gather	head	to gather	tail	
together		together		‘to gather everything together; to sum everything up’

(or those having the second and fourth word the same)⁵

0	0	0	0
1	2	3	2

Rhymed expressions: (those in which, usually, the second and third words rhyme)

0	0	0	0	
1	2r	3r	4	
mũu	hèd	pèd	kàj	(xviii)
pig	mush-	duck	chicken	‘meats of various kinds’
	room			

(Note that in this pattern, 2r may have no meaning, inconsistent meaning or more or less consistent meaning with whole.)

Vague semantic relation: (the semantically related words often occur in the second and fourth position)

0	0	0	0	
1	2s	3	4s	
lyym	hũu	lyym	taa	(xviii)
open	ear	open	eye	‘to open one’s eyes to what is going on around one’ ⁶

Tai Dam Examples:

Jay Fippinger in his article on Black Tai sentence types (1975) notes that “Syntactic structure is basically a form - a device specific to a given language - by which its speakers express underlying semantic concepts, many of which are common to human beings throughout the world” (130). Some of the sentences he presents as data in his article (although not intended to do so) provide examples from yet another of the Tai languages of patterns of four-word phrases in discourse. All but one of the examples which occur in that data happen to be of the type in which the first and third words are the same and the second and fourth words are semantically related to each other. They can be seen in the following table.

Table 10: *Repetition in Black Tai four-word phrases (from Fippinger 1975)*

bɔn ²	kin ¹	bɔn ²	ju ²	
place	eat	place	dwell	‘a place to live’ (140)
pa:n ³	miəŋ ¹	pa:n ³	fa:i ¹	
build	dam	build	dam	‘construct irrigation dams’ (141)
jet ⁵	suəŋ ³	jet ⁵	siə ³	
make	trousers	make	shirt	‘to sew some clothes’ (144)
laŋ ¹	fai ⁴	laŋ ¹	fɪn ⁴	
build	fire	build	fire	‘to build a fire’ (144)

The other example from Fippinger’s article illustrates a four-word phrase which is knit by its semantic pattern rather than by rhyme or specific word repetition. This phrase is seen in the example below.

Table 11: *Semantic linkage in a Tai Dam phrase*

<i>Semantic pattern:</i>				
0	0	0	0	
1s	2s	1s	2s	
nam ⁶	loŋ ⁴	fon ¹	tok ²	
water	descend	rain	fall	‘the rains come’ (141)

From the examples presented in this section we can see an increasing number of ways in which four-word phrases have been discussed as being woven together in various Tai languages; the Tai data used was gathered and published between 1914 and 1975, and each author saw some of the aspects of the structure of Tai four-word phrases. We find phonetic similarity, rhyme, repetition and semantic similarity discussed in these materials.

Tai Comparisons

The same or similar four-word phrases can be found in many other Tai languages, often with the same or similar meanings. For example, Seree Weroha (1992:182) provides us the same phrase, with the phonological differences, for Vientiane Lao and Lue; he provides the same English translation for both.

Table 12: *Lao Lue comparison*

saaj ²	pai ¹	saaj ²	maa ⁴	(Lao Vientiane)
expert	go	expert	come	‘knowing when to visit’
caaj ⁵	pai ¹	caaj ⁵	maa ⁴	(Lue)
expert	go	expert	come	‘knowing when to visit’

Many similar phrases may be found between Lao and Thai as well. One occurred in the data presented earlier in this paper.

Table 13: *Lao Thai comparison*

<i>Lao</i> (from Roffe field notes)				
muu ¹	het ³	pet ³	kai ⁴	‘miscellaneous meats’
pig	mushroom	duck	chicken	
<i>Thai</i> (from Haas 1964)				
mũu	hèd	pèd	kàj	‘meats of various kinds’
pig	mushroom	duck	chicken	

6 Four-Word Phrases in Other Southeast Asian Languages

Many other Southeast Asian languages appear to make use of similar patterns of four-word phrases. In his 1965 article, “Parallel Constructions in Vietnamese,” Nguyen-Dinh-Hoa discussed Vietnamese four-word phrases (128). He pointed out a pattern of repetition which is apparently the same as that discussed earlier in this paper for Lao: words one and three are repeated and words two and four have a close semantic link. He also noted that in some four-word phrases in Vietnamese “rhymes are utilized for the second and the third word of the construction” (128); this is the same pattern found for standard internal rhyme in Lao four-word phrases.

Judith Jacob has written about similar patterns in Khmer. In her 1975 article, “Observations on the Uses of Reduplication as a Poetic Device in Khmer,” some of her examples are similar to types found in the Lao data. One type of repetition she presents for Khmer is “repetition of a word with the interpolation of the components of a compound word” (117); this type appears to be the same as that presented for the Lao data in Table II, 3, of this paper as “Repetition with a (Split) Compound.” Another type she lists for Khmer is “repetition of a word with the interpolation of two words which are lexically opposites” (117). This type appears to be the same as the Lao type presented in Table II, 4, “Repetition with Opposites.”

Writings on the Hmong language also give evidence that four-word phrases with constructions similar to that of Lao are to be found in Hmong. In her 1992 book, *Meaningful Tone*, Martha Ratliff includes a Hmong phrase that illustrates this. This Hmong four-word phrase with its English gloss appears to be of the same type of repetition as that presented in this paper in Table II, 1, “Repetition with Close Semantic Link,” in which words one and three are the same and words two and four have a close semantic link. The phrase is presented in the next table.

In a 1999 article, “Metaphorically Speaking in White Hmong,” Elizabeth Riddle also includes some Hmong four-word phrases that show repetition with close semantic links. Two such phrases are presented below.

Table 14: *Repetition in Hmong four-word phrases*

tawv	ncauj	tawv	tsaig	(Ratliff,1992:89)
hard	mouth	hard	jaw	‘indocile’
kub	tes	kub	taw	(Riddle, 1999:383)
hot	hand	hot	foot	‘busy doing many things at once’
cua	daj	cua	dub	(Riddle, 1999:384)
wind	yellow	wind	black	‘storm’

Four-word phrases with patterns of repetition, rhyme and semantic linkage appear not only in Lao and other Tai languages, but apparently in other Southeast Asian languages as well. The few examples discussed for Vietnamese, Khmer and Hmong in this section indicate that this is a potentially rich area for linguistic investigation and comparison within and between the languages of the region

7 Conclusion

There are three main patterns to be found in the four-word phrases looked at thus far for Lao: repetition, rhyme and semantic linkage. These patterns can overlap. As Roffe has noted, there are both fixed expressions and productive patterns to be found in the Lao material. In the connected discourse samples from the *mohlam* performances, it is the productive patterns with repetition dominant that appear most frequently.

In addition to the Lao phrases provided in this paper, we have also given examples of similar patterns in four-word phrases in other Tai languages. Finally, we have presented some material which suggests that similar patterns of four-word phrases can be found in other Southeast Asian languages of the region, specifically Vietnamese, Khmer and Hmong.

Four-word phrases in Lao have been gathered and analyzed for patterns within these phrases. In my continuing work on this project, I will investigate their use, distribution and frequency in a number of types of discourse. Further analysis of the role of four-word phrases and related patterns by linguists, language teachers and translators of literature should be productive in expanding our knowledge of aspects of Lao discourse and poetics and could have value for exploring discourse structures and poetics in other Tai languages as well.

It seems fitting to end this article with a traditional Lao farewell,⁷ which contains a fixed four-word phrase with internal rhyme meaning essentially ‘Be well.’

khoo ¹	hay ⁵	yuu ⁴	dii ²	mii ³	heej ³
request	give	exist	good	have	strength

ຂໍໃຫ້ຢູ່ດີມີແຮງ

Notes

1. There are a number of phrases in Lao made up of four syllables, which actually consist of two words. In this paper, I will not be discussing them. In addition, there are Lao four-word phrases in which one element appears to be a nonsense syllable. Such phrases will not be discussed in this paper.
2. Note that even though this is a fixed expression, there is a variation of it in which the last word in the phrase is *kham*³ rather than *thooj*³ for the word 'gold.' See Kerr (1972: 505) for this variation. In his article on reduplication Roffe did not use a phonemic transcription; the one given in this example is one I am using for this paper. See further explanation in the next note.
3. All of the Lao examples in Roffe's field notes were written in the Lao script; an English gloss was provided, but no phonemic transcription. In his 1946 article on Lao phonemic structure, Roffe uses a numbering system to indicate the six tones he describes there (292). For that 1946 paper, Roffe's informant was a member of the royal family from the Luang Prabang area, and Roffe describes this as a "dialect of North Laos" (289). For more information on Roffe's presentation of tone there, see that article.
In their 1956 language learning textbook, the Roffes use symbols to indicate the six tones of the Luang Prabang dialect which they describe there (11). In that book, the transcription system they use is similar to the one used by Haas for Thai. However, the numbering system I use in this paper for these six Northern Lao tones and the six tones in the Southern Lao dialects in my material is an adaptation of the Gedney tone chart (1972). The value of the tone indicated by each number varies according to the dialect it represents.
4. Double reduplication may be the pattern of the Shan phrases meaning 'with a glistening sheen' (299) or 'off and on for a long time' (301).
5. Haas indicates that there is an alternate form of semi-repeated expression, those having the second and fourth word the same (xvii). I show that pattern in this table; however, she does not provide an example of this variation.
6. Note that in Table II., No. 1, I have considered a Lao phrase with 'ear' and 'eye' to have a close semantic link, rather than the vague semantic relation which Haas attributes to them in her Thai example (xviii).
7. There are a number of possible criteria for deciding that a particular phrase is a "fixed expression." One of these is the high frequency of use of a particular four-word phrase in everyday speech or in particular cultural situations. The phrase *yuu*⁴ *di*² *mi*³ *hɛɛŋ*³ meets this criterion in being heard frequently in daily speech, particularly at times of farewell. Another criterion is finding the phrase listed and defined as a unit in a Lao-Lao or Lao-English dictionary. This phrase can be found in Kerr (1972: 956).

References

- Compton, Carol J. 1975. "Lam Khon Savan: A Traditional Form and A Contemporary Theme." In *A Tai Festschrift for William J. Gedney, April 4, 1975*, edited by Thomas W. Gething, pp. 55-82. Southeast Asian Studies Working Paper No. 8. Honolulu: University of Hawaii.
- Compton, Carol J. 1979. *Courting Poetry in Laos: A Textual and Linguistic Analysis*. Monograph No. 18. DeKalb, Illinois: Center for Southeast Asian Studies, Northern Illinois University.
- Compton, Carol J. 1992. "Lao Poetics: Internal Rhyme in a Performance of Lam Sithandone." In *Papers on Tai Languages, Linguistics and Literatures: In Honor of William J. Gedney on his 77th Birthday*, edited by Carol J. Compton and John F. Hartmann, pp. 231-261. DeKalb, IL: Center for Southeast Asian Studies, Northern Illinois University.

- Fippinger, Jay. 1975. "Black Tai Sentence Types: A Generative Semantic Approach." In *Studies in Tai Linguistics in Honor of William J. Gedney*, edited by Jimmy G. Harris and James R. Chamberlain, pp. 130-169. Bangkok: Central Institute of English Language, Office of State Universities.
- Gedney, William J. 1972. "A Checklist for Determining Tones in Tai Dialects." In *Studies in Linguistics in Honor of George L. Trager*, edited by M. Estellie Smith, pp. 423-37.
- Haas, Mary R., compiler. 1964. *Thai-English Student's Dictionary*. Stanford: Stanford University Press.
- Hudak, Thomas John. 2000. "Introduction." In *Cushing's Shan-English Dictionary: A Phonetic Version*, pp. xi-xx. Tempe, AZ: Arizona State University, Program for Southeast Asian Studies, Monograph Series Press.
- Jacob, Judith M. 1979. "Observations on the Uses of Reduplication as a Poetic Device in Khmer." In *Studies in Tai and Mon-Khmer Phonetics and Phonology: In Honour of Eugénie J.A. Henderson*, edited by Theraphan L. Thongkum, Pranee Kullavanijaya, Vichin Panupong, and M.R. Kalaya Tingsabadh, pp. 111-130. Bangkok: Chulalongkorn University Press.
- Kerr, Allen D. 1972. *Lao-English Dictionary*. Washington, DC: Consortium Press. The Catholic University of America Press. 2 vols.
- Nguyen-Dinh-Hoa. 1965. "Parallel Constructions in Vietnamese." In *Indo-Pacific Linguistic Studies, Part II. Descriptive Linguistics*, edited by G. B. Milner and Eugénie J. A. Henderson, pp. 125-139. Amsterdam: North-Holland Publishing Company.
- Ratliff, Martha. 1992. *Meaningful Tone: A Study of Tonal Morphology in Compounds, Form Classes, and Expressive Phrases in White Hmong*. Special Report No. 27, Monograph Series on Southeast Asia. DeKalb, IL: Center for Southeast Asian Studies, Northern Illinois University.
- Riddle, Elizabeth. 1999. "Metaphorically Speaking in White Hmong." In *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society*, edited by Graham W. Thurgood, pp. 377-392. Program for Southeast Asian Studies Monograph Series. Tempe, AZ: Arizona State University.
- Roffe, G. Edward. 1946. "The Phonemic Structure of Lao." *Journal of the American Oriental Society* 66:289-295.
- Roffe, G. E. 1975. "Rhyme, Reduplication, etc. in Lao." In *Studies in Tai Linguistics in Honor of William J. Gedney*, edited by Jimmy G. Harris and James R. Chamberlain, pp. 285-315. Bangkok: Central Institute of English Language, Office of State Universities.
- Roffe, G. Edward n.d. Field notes. (Late 1960s to early 1970s.)
- Roffe, G. Edward and Thelma W. Roffe. 1956. *Spoken Lao*. Vols. I and II. Washington, DC: American Council of Learned Societies.
- Seree Weroha. 1992. "Semantic Considerations in Interpreting Inscriptions: Illustrations from Lue, Lao and Kammuang." In *Papers on Tai Languages, Linguistics and Literatures: In Honor of William J. Gedney on his 77th Birthday*, edited by Carol J. Compton and John F. Hartmann, pp. 179-185. DeKalb, IL: Center for Southeast Asian Studies, Northern Illinois University.

AUSTRONESIAN ERGATIVITY TRACED THROUGH TWO CYCLES¹

Joseph C. Finney
Monterey, CA
<jcfinney@redshift.com>

0 Introduction

This is a study in the diachronic syntax of the Austronesian (AN) languages. It presents a paradigm of the cycle of ergativity, as seen in that family. We shall be concerned with the syntactic casemarking and the semantic Case relations of the core terms of the verb, in all its voices. Like most in the field, we'll use the accepted definitions from Dixon (1994) that a language is "accusative" (NomAc) if S, the single essential term of intransitive sentences, is marked the same as A, the agent (actor, doer) of the transitive sentences, while O, the object-patient-undergoer bears a different marking, defined as Accusative; and a language is "ergative" (AbErg) if S is marked the same as O, while A bears a different marking, defined as Ergative. Dixon (1994, 183) is to be commended for using the concept "moving a language around the cycle of change", and that concept is basic to our presentation here. Dixon's book defined concepts and basic orientation assumed in the present work.

We'll study how languages move around in cycles. Examine our table (diagram, a short unnumbered page) first and use it as orientation in reading the paper.

As we are specially interested in Polynesian and the complex development within it, let's summarize the successive phases of syntax that come down to it from our theoretically reconstructed PAN.

There are two syntactic states (NomAc and AbErg) in each cycle. Each state lasts for a stage.

There are two phases (A and B) in each syntactic stage (NomAc or AbErg).

That makes four phases in each full cycle (1A, simple NomAc; 1B, NomAc with passive; 2A, simple AbErg; 2B, AbErg with antipassive).

1. The earliest AN we can reconstruct began as NomAc in PAN (easily reconstructable from Basay and other languages considered archaic). Starosta (who died recently) thought Rukai preserves PAN's NomAc (though it might possibly have gone from PAN NomAc through AbErg and back to NomAc).
2. AbErg in many Formosan and Philippine tongues, and through one branch to Padoe's new NomAc. In parallel at greater speed AbErg lasts to about Proto Eastern Oceanic.

¹ The following abbreviations will be used for language groups: AN Austronesian, MP Malayo-Polynesian, CEMP Central-Eastern Malayo-Polynesian, CP Central Pacific, PN Polynesian, EPN East Polynesian. All may be preceded by P for Proto-. In hypothetical words, V and C mean any vowel and any consonant. PCM preposition-or-casemarker. TAM tense-Aspect marker, RED reduplication, DP determiner (noun) phrase, Lin linker. In Tagalog I'll spell the ergative casemarker *nang* as such and not "ng" as is customary.

3. NomAc from roughly Proto Eastern Oceanic through Central Pacific, Fijian, and the earliest reconstructable phase of Proto PN.
4. AbErg by the final phase of Proto PN (through a complex series of changes as proposed by Finney). Staying AbErg in Tongan, Samoan, and largely in the Outliers.
5. Rapanui, Easter Island (parallel to the mainline East PN) staying AbErg but becoming NomAc now, ending cycle 2 and beginning cycle 3 now.
6. NomAc in Proto East PN and persisting as in Hawaiian, Tahitian, Maori. These are 1B of cycle 3.
7. Maori is still within NomAc 1B cycle 3, but its predominant use of passive voice (more often than active) is a sign that it is moving toward AbErg syntax 2A. The passive voice, growing in the present phase, will become the Ergative in the future phase 2A, if the Maori language survives so long.

AN is a good family in which to analyze the diachronic changes of ergativity because for at least one living group, East Polynesian, reconstructions show two complete ancestral cycles (four changes of ergativity state, eight phase-changes), from PAn down: NomAc to AbErg to NomAc to AbErg to NomAc. And because one EPN language is irrevocably headed for AbErg again in a third cycle.

In applying the Dixon test to Stage 2 Ergative AN languages, which construction (voice) shall we use as the transitive? The Agent-Focus, or one of the Non-Agent-Foci (NAF)? As we shall show in examining Mayrinax, the answer is clearly the Patient Focus. That's the one that is the canonical Non-Agent Focus (and is in English, too). In any AbErg language, the Patient Absolutive term is the only one that will survive and have a descendant as a core term in the NomAcc 1A of the new cycle (where it will be the Accusative Direct Object). Analogously, in English the canonical passive is the Patient Passive "An award was given ([to] the teacher) (by the Board)"; and not, for example, the Dative Passive "The teacher was given an award (by the Board)." A complete cycle, shown in the Table, has four stages. They are: 1A, simple NomAc; 1B, NomAc with a passive voice; 2A, simple AbErg; 2B, AbErg with an antipassive voice [whence we keep repeating]. Could we begin just as well with simple AbErg and end four stages down? I think not. The path 1B to 2A is simple, but 2B to 1A is more complex and variable and is better put at the end.

A complete cycle, shown in the Table, has four stages: 1A, simple NomAc; 1B, NomAc with a passive voice; 2A, simple AbErg; 2B, AbErg with an antipassive voice. Could we begin just as well with simple AbErg and end four stages down: I think not. The path 1B to 2A is simple, but 2B to 1A is more complex and variable, and is better put at the end.

A good recent discussion on the general concepts is Christopher Manning's Ph.D. dissertation, published by Stanford (1996). His research was on Inuit, a West Greenland Eskimo language, but he discusses broader issues well. He considers argument structure (something closer to semantics than to pure syntax) the major determiner of syntax in languages throughout the world. He observes that two paths have been proposed for sliding into ergativity: one, from passive voices; the other, from nominalization (which is not discussed in this present paper).

Some events we'll see are (1) evidence about PAN from Basay, Atayalic and other archaic tongues; (2) dramatic return to NomAc through mixed ergativity ending Cycle 1, in van den Berg's series of villages (and confirmed in Central Pacific); (3) complex changes out of Fiji-like NomAc in the PPN period; (4) many types of syntactic change in Polynesian languages; (5) return to NomAc through mixed ergativity in Rapanui (right now!); and (6) one NomAc East PN language (Maori) taking an irrevocable step toward AbErg syntax early in a third cycle.

The term "Focus" refers to the DetP in Absolutive case, in phase 2A or 2B, which, like all Absolutive DetP, descends from ancestral Nominative case subject of a passive in stage 1B. There are three or more because stage 1B normally has three or more passive constructions. It also refers to the "Focus affixes" on the verb which identify those roles of each Abs term as an Agent, a Patient, or a Locative or some other role. Each corresponds to a type of passive voice. In stage 1B, Agent Focus (AF) Abs DetP had been Nominative subjects of active voice sentences, while Non-Agent Focus (NAF) Abs DetP had been Nominative subjects of passive sentences in the recent stage 1B.

The same basic principles should underlie events in all cycles, but very early in the second detectable cycle of AN languages, the Focus affixes in MP had severely eroded, and the term Focus is not used. One could label Tongan's main syntactic (AbErg) type as NAF, or label Fijian's or East Polynesian's NomAc syntax as AF, but nobody does.

Translations of Ergative sentences (stage 2A) into English. If the Ergative agent is translated as the English subject, it shows the new NomAcc syntax that will appear two stages later, 1A, when the old Ergative case is reanalyzed as the new Nominative case. But if the Absolutive patient is translated as the subject, it shows the old passive voice of the NomAc sentence that appeared earlier in stage 1B. English is a language in stage 1B.

1 General characteristics of AN languages

Major words from most ancient forward were CVCCVC > CVCVC > CVCV (in most MP languages and many others now).

The greatly prevailing word order is SVO, and therefore adpositions are prepositions, not postpositions.

All AN languages have Verb Object order, and so they have prepositions, left of the DetP, not postpositions.

In many AN languages, adjectives are not sharply distinguished from verbs. As an example, in Nanumanga, where I did field work, "te tangata fano" translates "the running man", while "te tangata e fano" translates "the man runs or is running", in which "e" is the most neutral tense-aspect marker. In many AN languages "ne" marks past or present perfect tense, and in some "kV" marks future. Many languages in cycle 1 have subject markers and object markers on the verb. They are obviously reduced forms of personal pronouns, reflecting older word order and syntax. If third person we commonly see both the noun and the affix.

The word "focus" means "voice". The focus markers on the verb show its voice. The suffix -un shows that the subject is a patient. The suffix -an showed originally that the subject was a location, but in some languages has come to be used also for a patient subject. Both those markers may be old phrases meaning "at it". Agent subjects in the archaic languages may be shown by prefixes and infixes at the verb.

PAn demonstratives distinguish specific/nonspecific (which Huang calls Rf-Nrf), not definite/indefinite. When English “a” means “a certain” it is specific in AN. In the disagreement between Starosta and Blust, I may occasionally refer to (Austronesian minus Rukai) as Mainline AN.

Some events we’ll cover are (1) evidence about PAn from Basay and other archaic tongues; (2) dramatic return to NomAc through mixed ergativity ending Cycle 1, in van den Berg’s series of villages; (3) complex changes out of Fiji-like NomAc in the PPN period; (4) many types of syntactic change in Polynesian languages; (5) return to NomAc through mixed ergativity in Rapanui (right now!); and (6) one East PN language’s taking an irrevocable step toward AbErg in a third cycle.

We shall see evidence to support a prehistory (1) that the earliest reconstructable PAn had NomAcc syntax; (2) that before moving toward ergativity PAn had to develop passive voice(s) (1B) and developed at least three of them, much like the ones that English has; (3) that through small, identifiable steps, the passive voices became ergative voices in 2A; and (4) by the time the Malayo Polynesian speakers left Formosa, nearly all the AN tongues had changed at least to 1B and some to 2A; and (5) the AbErg, stage 2A was ancestral to all the AN languages but the most highly archaic ones.

2 Basic principles of linguistic change

Lenition (progressive weakening of sounds toward zero) is an inconspicuous process, but it’s the great engine of linguistic change. As phonological clues disappear, redundancy falls, and scarcely enough stimuli remain for interpretation, words are inserted. Unstressed words (especially pronouns, and prepositions or casemarkers) become reduced to clitics, which then are reduced to affixes and then to zero.

Space concepts (and space prepositions) are the source of abstract concepts. The prepositions “from [a physical source]” come to mark agents (sources of action): at first in passives (1B) and then in the ergatives to which the 1B passives are reanalyzed in the formation of stage 2A.

As agent subjects are in some sense the source of the action, and accusative object patients are in some sense the destination or goal or theme of the action, it’s not surprising that ancestral “from, by” prepositions come to mark agents, and ancestral “at” or “to” prepositions come to mark patients, receivers of the action.

Though we tend to think categorically, in major changes of status many syntactic and phonological changes take place in small steps, which are inconspicuous. A change from AbErg to NomAc syntax, or from NomAc to AbErg, takes place as a series of smaller steps.

This standard way to develop an Ergative syntax is to reanalyze a passive voice (stage 1B) to Ergative (thereby entering stage 2A).

When a syntactic innovation occurs, it may begin as a synchronic transformation, secondarily generated. When the innovation is well established, it may become the one that is generated directly, and then the older syntax, so far as it remains, comes to be generated from the newer one. We shall see in Mayrinax, 2A, that the newer NAF patient-focus construction comes to be the standard, and the older (disappearing) AF sentences are generated secondarily and are not the ones to be used in the Dixon test for ergativity.

Nominalization had been proposed by Starosta, Pawley and Reid (1981) as a source of Ergative syntax in AN. That’s not implausible. In some AN languages (notably

Tagalog), “The eat banana got sick” is the way to say, “The one who ate bananas got sick.” And the fact that Mayrinax has a voice that some call passive (in NomAc) and others call ergative (in AbErg) is reason to accept that languages drift imperceptibly from one syntax to the other.

Dixon says (p.186): “What is interesting is the ways in which a language moves from accusative to ergative or from ergative to accusative. There are several kinds of diachronic mechanism that may be involved: the reinterpretation of a passive or antipassive as the unmarked transitive construction type; the development of a new periphrastic system of tense and/or aspect marking based on participial forms; the creation of a new case or the extension in meaning of an existing one; generalization from one tense-aspect to another; shift in constituent order and topicalization; and so on. Path (a), from accusative to ergative, is by no means the mirror-image of (b), from ergative to accusative. It is true that one way of achieving (a) is through reinterpretation of a passive, and of (b) by reinterpretation of an antipassive. But I emphasize that although passive and antipassive appear to be syntactically parallel, with A and O interchanged, in fact they are semantically quite different. As a result, the circumstances in which passive reinterpretation can trigger (a) are quite different from those in which antipassive reinterpretation can trigger (b).” [Our observations in this paper confirm that there are differences.] Dixon’s list of paths is much longer than Manning’s.

He adds (p. 228-229), “These systems are far from being exact complements....We might infer that the S/A links must be stronger and more important from the fact that there are many languages with no trace of ergativity, whereas no language lacks a degree of accusativity.”

It’s also worth while to examine Dixon’s Nominal Hierarchy (p. 85), for which he notes (p. 187), “Accusative marking extends across all types of nominal constituent while ergative is found only on the right-hand side, with inanimates.” First-person pronouns are at his far left. He quotes another scholar’s finding Hittite with an ergative case used only with neuters or inanimates: “the bindings (Erg) clasp the head (Acc).” I am not the only scholar of Polynesian that has found the opposite: that Ergative case is used mostly for human beings who are held morally responsible for what they do (Duranti, 1981 and later). In PPN the Ergative case is created when human agents come to be allowed in what had been a slot for inanimate causes.

Languages of communities that stay put permanently tend to change only slowly. It seems agreed that the indigenous languages of Formosa are as a whole close to the ancestral PAN. In contrast, those of faraway Polynesia have undergone repeated wrenching changes, and are not the ones from which one would reconstruct PAN. In the 6000 or so years since Formosa was settled, Padoe, far to the South, has undergone a full change of cycle and returned back to NomAcc. And the languages that moved still farther East, as far as East Polynesia, have changed even more rapidly. East Polynesian languages have undergone two full cycles, and Rapanui and Maori, in different ways, are in the third cycle already. The ancestors of Polynesia may have completed the first cycle 3000 years ago, somewhere near Birds Head and Halmahera.

Dual inheritance. Notice that when a passive voice is created, the new agent inherits its identity from the Nominative subject of the active form, but inherits its case and marking from the “from, by” prepositional phrase. At the same time, the new patient subject of passive inherits its identity from the Accusative case of the active voice, but

inherits its case and marking from the Nominative subject of the active voice. The agent in the passive has been demoted from a core term to an oblique adjunct that can be omitted. [Puzzle: in Gault's (1999) account of one Ergative language, any term can be omitted, even the Absolutive patient subject or the Ergative agent.]

3 Sharp and fuzzy changes of syntax

Changes from 2B AbErg to 1A NomAc are sharp, though sometimes murky. There is seldom difficulty in labeling the syntax. The change from 1B passive to 2A ergative is fuzzy, as shown by disputes over the labeling of Mayrinax.

4 Reconstruction of PAN from archaic languages, and especially from Mayrinax (and Rukai?)

The earliest form of hypothetical PAN that we can reconstruct is a NomAc language in stage 1A. That conclusion is required by the persistence of its ghost in Mayrinax's demoted Agent Focus construction. We can't say much about what preceded that stage because 1A is preceded by a murky change involving reanalysis of Ergative to Nominative case. The two examples in which we see the evidence both work through a state of mixed ergativity, but that may not be the only route. It's hard to posit an ergative marking that would be preserved as the essentially zero marking on the Nominative noun phrases reconstructable in PAN, where the identification of the Nominative subject agent is done by the focus-marking morphemes. At the end of this paper we'll suggest some approach to reconstructing an ancestor of PAN by comparison with forms found in families related to Austronesian.

The most obviously archaic AN attested languages (except Basai and Rukai) are languages in stage 2A. As such, they have AbErg syntax; at least in their primary transitive voices, the Non-Agent Focuses, NAF, of which the Patient Focus is the canonical one. But they also show an "Agent Focus" [AF] construction, which has suffered demotion and has become an archaic relic. It no longer counts as a transitive construction, despite the preservation of the apparently transitive wording, including affixes on the verb that identify the focus. Its words (seen most clearly in 2A Mayrinax) preserve the actual wording (including focus marking on the verb) of the earlier stage 1A, which had had transitive NomAc syntax. That's the wording of the basic syntactic structure of Proto-Austronesian's NomAc syntax. Why don't we call Mayrinax NomAc? Because that construction in Mayrinax, though worded the same, no longer functions as the transitive form, nor as the basic syntactic form. It's a derived construction, just as the passive voice is a derived construction in English and other NomAc languages. In the syntactic change from late stage 1B to early stage 2A, what had been the basic form was reanalyzed as a derived form. And what had been a derived form (passive) was reanalyzed as the the basic form. Had the old usage continued, and not been eclipsed by the old passive form (reanalyzed as AbErg construction), Mayrinax would still be NomAc, phase 1B with passive. What the predominating evidence shows is that the immediate ancestor of Mayrinax was in stage 1B, and before that, stage 1A, and that that easily reconstructible NomAcc 1A stage is the earliest form of Proto-Austronesian that can be recovered and perceived, as it was preceded by a sharp, murky change.

Thereafter, the archaic AF form must disappear before the new AbErg construction can develop an antipassive in 2B.

So, the highly archaic voice, from PAN, seems to persist at least in the ancient wording and form and focus markers, but reanalyzed so that it cannot be correctly called

NomAc today, even though it still looks like it. It's a demoted voice. That's why it's not the form to be used in Dixon's test. (That's an issue that Dixon failed to make clear.)

When one looks at the AN languages that are obviously archaic, most seem to show, in general, AbErg syntax. For that reason, most AN linguists (though not Starosta) have mistakenly said that PAn must have had AbErg syntax. The common conclusion does not follow. As in those languages the AF is the demoted form of the of the ancestral syntax from state 1A, it's the one that shows the easily reconstructible earlier ancestral NomAc syntax, and that's what deserves to be called PAn. Those who have called Mayrinax AbErg are basically correct, but it's also basically correct that, following the general principle of the repeating cycle, we can easily reconstruct the earlier phase 1A. And the 1A phase is the appropriate starting point for the diachronic account of the Austronesian languages.

But on the other hand the conventional majority has been correct in the view that (despite what Huang has sometimes said) the canonical unmarked voice in Mayrinax today (and for a substantial part of the past) has been the NAF, and specifically the patient focus. In other words, Mayrinax today is basically AbErg, despite Huang's vacillation about it.

Let's restate the matter, because it's fundamental. By Dixon's definition, the archaic construction in PAn must be called NomAc in Pan. It won't survive into the next 1A phase that the language (if it were to survive) would enter some day, ushered in by the creation of a new Nominative case. In Mayrinax today the NAF construction (specifically the patient voice) is the basic voice and is the one to use in Dixon's test, and its S=O finds Mayrinax to be an AbErg language. What is shown by the S=A agreement in Mayrinax's demoted AF is that Mayrinax's ancestor was PAn NomAc.

Among Western MP languages, from AbErg, some move around the cycle and become NomAc again. Berg's Padoe is striking in being the specific village at which the return to NomAc takes place. The evidence is clear. We are very fortunate that Berg gave so much detail. The languages that pushed farther East changed more rapidly. So, in the present day we see both Padoe completing its first full cycle and Rapanui, Easter Island, completing its second.

Perhaps nobody has said so before, but it seems clear that mixed ergative, no matter whether lexical or fluid, is not a mere curiosity. It is the normal and usual way in which a language changes back from AbErg 2B to NomAc 1A. It does so by reanalysis of the old Ergative case into the new Nominative case. Berg's Padoe is a clear example.

As noted, some Polynesian languages completed a second full cycle in the recent 3000 years and have entered a third. Yes, though the markers have changed, it turns out that the way Rapanui is ending the second full cycle and beginning the third, at this very time is, like Padoe, by creating mixed intransitivity (fluid ergativity, in this case).

At the end of the PPN period a new Ergative was created in an unusual way. The passive that it reanalyzed was not a normal passive but a lexical passive: a set of verbs that had reanalyzed an involuntary cause construction ("The building was hit by a storm") into a passive-like construction that used casemarker "e" for the cause or agent. When this clause was borrowed into the Fiji-like NomAc construction in Proto-Central Pacific, whose verbs had -Cia verbal suffix, it wreaked havoc. We've described some of the complex series of changes that thereupon took place, in various divisions and subdivisions of Polynesian, from that point on. Fischer's discoveries and hypotheses based on Mangareva are of interest. And Maori, NomAc, 1B, has taken the first irreversible step toward a new 2A AbErg syntax, in that third cycle.

5 View of the whole course from basic NomAc PAN onward

In this study of syntactic change we shall see an account of how languages move through each full cycle, from stage 1A simple NomAc, 1B Nom Acc with passive voice, 2A simple AbErg, 2B AbErg with antipassive, thence to 1A again. It's shown in the Table page. Each stage 1A ends one cycle and begins the next cycle. Many languages stay a long time in stage 1A or 1B.

When we look at the AN languages that are obviously archaic, many of them seem show AbErg Syntax, whose unmarked voice is NAF. It is for that reason that most AN linguists [though not Starosta] have mistakenly said that PAn must have had AbErg syntax. As the AF is the demoted form of the ancestral syntax from stage 1A, it's the one that shows the ancestral NomAc that is to be called PAN. Some scholars have rightly called Mayrinax AbErg, while others have hesitated to say so, for reasons that are easily understandable. When we look at the flow as part of a cycle that inexorably repeats itself, it is clear that Mayrinax in Stage 2B, still in the first cycle, is not far removed from the easily visible proto-language that was NomAc.

There can be little doubt that the canonical ("unmarked") voice in 2A Mayrinax is the NonAgent Focus, and specifically the Patient Focus, which had been the subject of passive voice in stage 1B.

In Mayrinax, as we saw, there is an obviously archaic voice, the AF, which persists and is nowadays derived secondarily, from the NAF. Remember the ancestral 1B, where the passive voice (marked) had been derived from the active. Examination of Mayrinax AF shows that S (subject of intransitive sentence) is marked the same as A (agent of transitive sentence). This demoted archaic construction had been NomAc.

Among AbErg Western MP languages, too, some have moved around the cycle and become NomAc again. Berg's Padoe is striking in being the specific village at which the return in one branch of MP takes place slowly, ending its cycle 1.

With ancestral speeding up, ProtoEastern MP completed its first cycle in perhaps 3000 years, somewhere near Halmahera and the Bird's Head. In a second full cycle. in another 3000 years or so, Rapanui is doing so at this very time. It's interesting that the two that are completing a cycle just about now are doing so in the same way, by mixed intransitivity (fluid ergativity in this case), even though one is ending the first cycle and the other ending the second cycle.

At the end of the PPN period a new Ergative voice was created in an unusual way. The passive that was reanalyzed was not a normal passive but a lexical passive: a set of verbs that had reanalyzed an involuntary cause construction ("the building was hit by a storm") into a full passive marked by casemarker "e". When this agent phrase was borrowed into the Fiji-like Cia NomAc construction that early PAn had inherited, it wreaked havoc. We look at some of the complex series of changes that took place in various divisions and subdivisions of Polynesian from that point on. Fischer's discoveries based on Mangareva are of interest. Rapanui (like Padoe) is at the end of a cycle (the second full cycle in this case), and the beginning of a third. Slightly ahead of it, Maori, NomAcc, 1B, has taken the first irretrievable step toward a new 2A ergative syntax, in a third cycle.

Let's summarize what we have noticed about the sharp differences in the speeds of change. In general, the languages still on Formosa have changed least from PAn, and are

still in the earliest attested cycle. Those that have moved farthest East, notably Polynesian, have changed their syntax most rapidly. At present, some Polynesian languages have completed two full cycles, while Padoe has just completed only one cycle down from PAn.

6 The table of the ergativity cycle in AN 1A > 1B > 2A > 2B > 1A

The Table shows the general path within the single cycles in Austronesian. Each row represents a stage of syntax. In the archaic AN tongues, stage 2, the ergative voices (NAF, non-agent-focus) resemble passives, of course, because that's what they were in stage 1B before reanalysis; and the simple intransitives are marked the same as the Agent-Focus constructions (thus treating alike what had come down from the S & A of stage 1). So, even if no living archaic language showed precisely a simple Nominative-Accusative (NomAcc) syntax, it still would be easy to reconstruct the ancestor that did, and that's what we'll call PAn (Proto-Austronesian).

This is the canonical cycle of ergativity, somewhat idealized. Each row (or "phase") 1A is simple NomAc syntax; Row 2A, simple AbErg. Rows 1B and 2B show the syntaxes of the passive voice of NomAc and the antipassive voice of AbErg. At the end of 2B, the Ergative marking is extended first to subjects of active intransitive sentences (making mixed or fluid ergativity); and then to all intransitive sentences (creating, by definition, a phase 1A in a new NomAc syntax). Padoe has just done this at the end of its first cycle, and Rapanui is in the process of doing it now at the end of its second cycle down from PAn.

Creation of a passive voice does not make movement to AbErg inevitable. A NomAcc language with a passive voice may stay that way a long time. I suggest that change to AbErg becomes inevitable if and when the passive voice comes to be used oftener than the active voice (as it has now in New Zealand Maori).

In the syntactic change to phase 2A, the 1B oblique agent is promoted back to status as a core term, the Ergative case, and its "from, by" preposition nV is reanalyzed as the Ergative casemarker. At the same time, the Nominative patient subject of the former passive is reanalyzed as the Absolutive patient (classed by many as subject) of the new AbErg sentence. This is the "patient focus", which is the most prominent among the NAF (Non-Agent Focus) constructions, for the same reason that English patient passives ("An award was given.") are commoner than dative or locative passives ("John was given an award." and "This house has been lived in."). Both the Ergative agent and the Absolutive patient are core terms.

7 Lack of analogy or symmetry: antipassive in stage 2B

The move to NomAc at the end of stage 2B fails to follow the analogy of the move to AbErg at the end of 1B, in two ways. (1) The demotion of the Nominative agent to an adjunct in formation of the passive is not fully followed in the demotion of the Absolutive patient to an adjunct in formation of the antipassive. In the former, the resultant clause is marked as intransitive and the passive agent is marked as oblique adjunct by a "by" preposition. But in the latter, the demotion may or may not be marked by a preposition. Once Blust complained to me that in somebody's published paper, a sentence is marked as intransitive, but it looks transitive with the patient simple Abs. In such cases, the only way to test whether the antipassive demotion of patient has taken place is to ask a native speaker whether the patient can be omitted or not. If it can, move has been made into stage 2B, and the the construction

is intransitive, antipassive. (2) The subsequent change from 2B to 1A is canonically a reanalysis of old Ergative to new Nominative, going through a condition of mixed ergativity. We'll see two instances.

8 Acquisition of casemarkers from PAN zero

We've seen clues that the NomAc syntax visible in PAN had, like English, zero casemarking for both Nominative and Accusative. I have previously shown that the "kV" marker for Nominative was ancestrally a marker of Topic. The candidate for what became Accusative marker in some languages is "cV" in PAN and Mayrinax and "tV" in PMP. See it partly reanalyzed from preposition to Accusative casemarker in Liao's example. See it also in Finney, 2001b, where a discussion gives many examples showing its evolving from preposition to Accusative casemarker. For Amis, Chen (1987, 127) shows the *cu > tu* particle (*u > o*) clearly as having become an Accusative object marker:

Topic	Nominative	Genitive	Accusative	Locative
<i>o</i>	<i>ko</i>	<i>no</i>	<i>to</i>	<i>i</i>
<i>ia</i>	<i>kia</i>	<i>nia</i>	<i>tia</i>	<i>itia</i>

9 Development

At an early stage, PAN had NomAc syntax, phase 1A. Nominative and Accusative casemarkers were zero. PAN entered phase 1B by developing a passive voice, with the demoted agent marked by preposition *nV*. As in all such languages, the canonical element raised to Nominative subject was the patient, though, like English, the language allowed other passive voices in which the dative beneficiary or the location or something else could be the one promoted to Nominative subject. AN ("Mainline" if we exclude Rukai) divided into nine branches, as Blust (1999) defined them, on classical principles of lexicon and phonology. Blust (pers.comm.) does not favor using any syntactic changes in the grouping of related languages. One branch (Malayo-Polynesian) spread throughout more than half the world's longitude to Madagascar and to Easter Island and gave rise to countless languages and dialects. Before the MP dispersion, all these branches probably developed not only the canonical passive from patients but also passive subjects from dative, locative, and instrument. The lowest common ancestor of Mayrinax and Tagalog had Agent-Focus markings on the verb that were the same for S and A, and so the syntax still bore signs of ancestral NomAc, stage 1. Focus markers "ma" (more patient-like) and infix "um", both AF markers on the verb, share the consonant, but it's not clear that they have common ancestry. Topics at the left of the sentence had been marked with "kV" at their right, and the topic (later, Nominative) markers "kV" developed from old topic markers. AN casemarkers lie left of their Det Phrases because they are mostly ancestral prepositions in this VO family. In Mayrinax, most of the casemarkers add the syllable "kV" to show specificity ("the" or "a certain"). A reasonable explanation is that this specificity marker comes down from the older topic marker, as topics are inherently specific. That's what established the association between Case 1 (Nom, Abs) and specificity. In Tagalog the Abs casemarker "a-ng" shows specificity, and indeed when the speaker needs to express the specificity, the way to do so is to put the DetP in the "a-ng" case, (Absolute subject, ancestral Nominative). When a location is put in Tagalog's Abs "a-ng" (subject, specific) case, both the agent and the patient are put in the "n-a-ng" case (Ablative-Genitive-Ergative), making a potential ambiguity. The "-ng" linker on the right end of casemarkers may be an MP innovation, as it is shared by

Tagalog and Old Javanese, but not Mayrinax. Old Javanese uses the vowel “i” (anciently for proper names of persons) left of the linker, while Tagalog uses the “a” vowel (anciently for common nouns). We note Blust’s (1999) harsh remarks in the 8-ICAL volume about Starosta’s view of Rukai’s position in AN.

10 General course of the first full cycle

We’ve anticipatorily mentioned already and will give full evidence now on the Mayrinax Atayal language which is often regarded as a fully AbErg language, stage 2. The conclusions about its Agent Focus voice may surprise some people. As an MP language, Tagalog agrees with Old Javanese in showing an innovation not found in Mayrinax: the insertion of linker “ng” as a suffix on the casemarkers.

We describe Rene van den Berg’s work on the move from AbErg to NomAc in Sulawesi, in which we can identify exactly which village today (Padoe) still preserves the final step ending Cycle 1, still standing on the line between AbErg and the new NomAc.

We compare Berg’s Padoe in Sulawesi with Fisher’s Rapanui, Easter Island, showing that both of them, today, are reaching state 1A NomAc in the same way (through mixed ergativity, reanalysing final 2B to beginning 1A). But Berg’s Padoe is doing so at the end of its first cycle, while Rapanui is doing so at the end of its second cycle. Obviously the ancestors of Oceanic, if they passed near Padoe at all, had done so many centuries earlier. They had ended their first cycle perhaps 3000 years ago, perhaps near the Bird’s Head. This is an example that commonly the languages that have move farthest change more rapidly.

11 Concepts of Nom-Ac vs AbErg in regard to Proto-Austronesian and some of the most archaic AN tongues

Stanley Starosta (who died a few weeks after this SEALS meeting) was said to have a predilection for finding ergativity where others don’t see it. And yet in regard to PAn, Starosta found it NomAc, while most AN scholars have said it must be AbErg (because Mayrinax and Tagalog and a majority of the AN languages that are credibly archaic show AbErg syntax). Starosta pointed out that the dialects of Rukai are basically NomAc. He thought that the oldest version of PAn that we could reconstruct must be much like Rukai. For this paper I have not reviewed my previous studies of Rukai, which were compatible with that thought. I gave the name Mainline AN to all the other branches of AN, and the name Proto-Mainline AN to the common ancestor of all but Rukai. I don’t take a position on the question whether the oldest PAn (ancestor of both Rukai and Mainline AN languages) was more like Proto-Rukai or more like Proto Mainline AN. Either way, PAn was NomAc. But as we’ll not discuss Rukai in this paper, the “PAn” that I reconstruct from Mayrinax and other languages is essentially the ancestor of all the (Mainline) AN languages.

We’ll examine the relic-structures in the archaic AN languages (adding Basay to those already mentioned) to show how they developed from the NomAc syntax of Proto (Mainline) AN, which for convenience we’ll call PAn.

12 What syntax can be seen in the long known AN languages

Exploring this issue, let's compare some AN languages, all known to be conservative: Mayrinax, Old Javanese, Tagalog, and Rukai. I've quoted in an earlier paper Becker and Hunter's (1988) table of "deictic prepositions" (casemarkers) in Old Javanese. Tagalog and Old Javanese are MP languages, and both have suffixed a "linker", "-ng", to the basic forms of the casemarkers. Results in Tagalog are a-ng and na-ng (the latter of which is customarily written "ng").

13 Old Javanese

As noted in my earlier reports, this is Becker and Hunter's table of "deictic prepositions" in Old Javanese:

	[Case 1]	[Case 2]	[Case 3]
	[Nominative]	[Genitive]	[Dative-Loc]
	[Absolutive]	[Ergative]	
	direct	oblique	directional
indefinite, nonspecific	<i>i</i>	<i>ni</i>	<i>ri</i> [<i><*di</i>]
definite	<i>i-ng</i>	<i>ni-ng</i>	<i>ri-ng</i>
definite and specific	<i>i-ka-ng</i>	<i>ni-ka-ng</i>	<i>ri-ka-ng</i>

Old Javanese had lost the kV and cV casemarkers, and Mayrinax had lost the Nom & Loc dV casemarkers which shows above as ri in the normal change from PAn *d to r, which becomes a zero consonant in Tongic, as we shall see in the Absolutive marker for common nouns in Niue.

Two key points to notice are these: (1) Old Javanese resembles Tagalog in that it attaches the "linker" -ng to the casemarking particle under certain very common circumstances. (2) Old Javanese agrees with Mayrinax in adding kV to the casemarkers to show specificity. (True, it has vowel u in Mayrinax and vowel a in Old Javanese, but that's within the normal variation of the vowel in the PCMs.) The two languages have no history of geographic proximity that could allow borrowing. So, the most reasonable explanation of the occurrence of this marker of specificity is that it was present in Proto Austronesian.

As topics are inherently specific, and PAn "kV" began as a topic marker, it's not surprising to find kV as a marker of specificity in Old Javanese as well as Mayrinax and Rukai.

Those authors' concept of 'definite' differs from the usual; theirs equates to the suffixation of the "ng" linker found also in Central Philippine.

14 Old controversy about Atayal

For good reasons, Mayrinax Atayal is the Formosan language most often cited in efforts to probe back toward PAn. It's clearly a conservative language. It's been described in detail by Lillian M. Huang (1994b with elaboration in 2000). She had previously reported (as her Ph.D. dissertation) a study of Wulai Atayal, a much less conservative dialect, and hence one of less value for reconstructing PAn. The essentials of her dissertation were published (1994a) in *Oceanic Linguistics* (University of Hawai'i, where Starosta taught). In that work,

she described Atayal as an Ergative language. But her Mayrinax work was done after reestablishing permanent residence in Taiwan, and the word “Ergative” does not appear in her book. People have said that she deleted the concept to please Paul Li. She published another detailed analysis in *Oceanic Linguistics* (Huang, 2000). We all owe her a tremendous debt for asking good questions and providing great amounts of valuable data. I couldn’t have done the present study had she not gathered the detailed observations and published them.

Wulai Atayal and Mantauran Rukai are not nigh kindred. So when both of them have a particle from PAn *kV (Mantauran ka, Wulai qu) that serves to mark the right edge of a topic, one must believe that that particle was in PAn. In some languages, glottal plus vowel became a topic marker, but glottal stop did not occur in PAn. Glottal often comes from earlier “k”.

My rearrangement of the tables of Mayrinax “casemarking” in Huang (1994, 109) shows what seems to represent an early stage not far from PAn:

glottal+V	kV	nV	cV	kV	(zero)+V	
[Abs]	[Abs]	[Erg]	[prepositional]			
Topic	Topic	Gen	Allat	Comit	Loc	(PAn?)
Nom	Nom	Instr	Dat	Allat		
		Ben	Acc			
(glottal) V	kV	nV	cV	kV	zero V	

All the “casemarkers” are ancestrally either topic markers or prepositions, as I have found in earlier reports. The range of their use in Mayrinax is shown in Finney, 2001b, pages 412-413.

Old Javanese has lost the kV and cV. Mayrinax loses the dV casemarker (which becomes rV in Oceanic languages and PPN and, as we shall see, simple V in Tongic).

Where the vowel is ‘a’ and not ‘i’ the tables actually show zero and not glottal as the initial consonant of the Topic or Nominative. Has the “I” vowel suffered preglottalization?

Prepositions dV, tV, and sV don’t occur in Mayrinax, but were present in PAn. The dV and tV had space meaning; also Nom for dV in some MP tongues. The si particle for proper names of persons occurs in some languages, but is glottal + i in Mayrinax. In some languages, when that person marker occurs it replaces whatever would be the normal preposition or casemarkers in a function.

An interesting fact is that kV serves both as an allative-comitative preposition and as a topic marker that can become a Nominative (and later Absolutive) marker. It’s not clear that the two uses had ancestry in common.

The word Topic in the table stands for usage either as a topic marker or as a marker of specificity. Topics are always specific, and so many instances can be labeled either way. kV is used after prepositions to mark specificity.

15 More on Mayrinax Atayal

Here are some of Huang's (1994) examples of Agent Focus (AF) and NAF (Non-Agent, i.e., patient or locative):

Agent Focus (AF) (intransitive): p. 40

m - ingilis 'i' yaya

AF - cry Nom mother

Mother is crying (or: Mother cried).

Agent Focus (AF) (transitive): p. 41

c <um> abu' cu' qulih ku' nabakis

wrap <AF> wrap Acc.Nrf fish Nom.Rf old:man

The old man wrapped a fish.

These examples confirm the key point that S and A are case-marked the same, and so the apparent syntax is NomAc. But we note elsewhere in this paper that Cu began as a preposition in PAn. So, PAn accusative objects had no marker and only later took prepositions that became casemarkers. The Nominative marker ku was earlier a marker of topic at the left (better called an afterthought when moved to the right, where it stays in Malagasy). We could translate: "(he) wrapped a fish, the old man." But earlier the kV had occurred at the right edge of the left topic.

Patient (theme) passive or ergative (NAF, PF): p. 45

nubuw -un ni' yaba' ku qusia'

drink -PF Gen father Nom.Rf water

The water was drunk by father. or: Father drank the water.

Location passive or ergative: (NAF, LF) p. 53

kabax -an ni' yumin ku' kulu'=mu.

borrow-LF Gen Yumin Nom.Rf car my

Yumin borrowed (at) my car. or: My car was borrowed (at) by Yumin.

For those of us who take the sentences as AbErg (and Mayrinax's basic status as an AbErg language), substitute the label "Ergative" for "Genitive", and substitute the label "Absolutive" for "Nominative". The diachronic point to make is that these sentences were NomAc when the language was in phase 1A & 1B, and they became AbErg in phase 2A. In this voice the change is not obvious. The drastic change in syntax comes in the descendants of the passive voice that arose in phase 2B. It's that agent that becomes the Ergative agent, which thereafter will be reanalyzed as the new Nominative agent in the phase 1A of the following cycle, thousands of years later.

These sentences show why Mayrinax Atayal is so often cited on both sides of issues in AN diachronic syntax. Starosta was right in saying that Huang should have stuck to her earlier position that Atayal has ergative syntax. But Mayrinax shows evidence of an older [PAn] syntax in which neither the Nominative Subject nor the Accusative Object had a casemarker. This language's AF sentences can be taken at two levels. On the surface (ignoring their demoted status) they reflect the ancestral NomAc transitive

sentences and thus clarify the nature of PAN. But because they've been demoted, they are not the transitive sentences in Mayrinax that are to be used in the Dixon test for ergativity. Starosta's (1999) objections to Huang's analysis of Mayrinax are well known.

If (correctly) we use Patient Focus to apply Dixon's test to the Mayrinax sentences, the decision is AbErg. The best known observer, Huang, had vacillated on the point.

Here's an interesting sentence (p. 65): [Terms in square brackets are mine, not Huang's]

kal -un=cu	ni' watan	c=ku sinubilan	cu' kai'
say-PF=1S.BNom	Gen Watan	Acc.Rf legend	Lin language
	[Abs Patient] [Erg]	[Acc Patient]	

Watan told me a legend. Or: (my added translations):

I was told a legend by Watan. A legend was told me by Watan.

An oddity of this (1B dative passive?) sentence is that the Nom/Abs subject (I) of the verb in Patient (Dative?) Focus must be the patient but also the (retained?) Accusative object (legend) must be the patient. English solves it by putting "me" in the dative in active voice; and in dative passive voice having "legend" be a retained accusative object. Huang does not account for the accusative. She gave an active-voice translation into English and I added passive translations. If we regard the construction as dative passive, the solution is the same as in English. It's easy to see why Huang could consider this either a 1B passive or a 2A ergative construction. Generative grammarians like to say that the dative is promoted to accusative before the passive transformation. Hudson's (1992) dissent is persuasive.

The construction still appears to survive today in the AF sentences, though synchronically the generation must be NAF first, then AF. Just look at the AF sentences and treat them as non-derived active voice, which was their ancestry, and you are looking at Proto-AN with its NomAc syntax.

Both Bloomfield (many years ago for Tagalog) and Huang (for Mayrinax) have regarded as passives the constructions that almost everyone today calls ergatives. Both languages show high portions of Agent-Focus sentences that still show clearly the NomAc structure of phase 1B. The high frequency of the (NAF) sentences that could be read as passives invites comparison with Maori's high frequency of passives that we take here as an irrevocable step toward ergativity. For Mayrinax and Tagalog the evidence justifies calling them Ergative languages, despite their clear evidence of NomAc ancestry, which we have called PAN.

We are lucky to have Mayrinax and the detailed work that has been done and published by Huang, as well as by others who have not always agreed with her. Mayrinax is conservative enough to show undeniably (in its archaic AF sentences) the basic NomAc 1A structure of its ancestors. It also shows undeniably the basic 1B passive sentence structure. From reanalysis, Mayrinax itself is AbErg, stage at least 2A. Some might contend that the demoting reanalysis of AF, though not exactly an antipassive, creates stage 2B. If that were so, all four stages or phases of a full cycle could be glimpsed in some sense in Mayrinax.

16 Zero-marked nominative & accusative in Proto-Austronesian

In a paper given in 2002 at the 9th International Conference on Austronesian Linguistics, H.-ch. Liao shows (in a *cu > tu* language):

tmuqiq	tu	punuz	na	paRibunan	tu	bawbi'
pierce	Acc.	backside	of (the)	watcher	(of, at)	(the) garden

Note that in this sentence the first occurrence of “tu” is best treated as an Accusative object marker (a late development), while its second occurrence can only be a preposition (basically, “at”).

This sentence shows a very old PAn locative preposition, glossed “at” in English, whose object here is “garden”, but it also shows at the left in the same sentence a later development, in which it is less like a preposition and more like a casemarker for the Accusative Direct Object [backside]. As we can easily see here, it’s an old preposition “at” that is becoming an Accusative casemarker. We conclude that in its ancestor, PAn, the Accusative case (as in English nouns) bore no marker. And the *kV* particle that becomes a Nominative marker in some AN languages, and a marker of specificity in others, was ancestrally a topic marker. So in PAn both

Nom and Acc (as in English nouns) bore no marking. This is a new conclusion, not previously made by others.

17 Tagalog

Something like Mayrinax is found in Tagalog and related Philippine languages, though complicated by the fact that the languages suffer from impoverishment of case-marking. With only three case forms and no preposition, some ambiguity and some paradoxical marking is found in Tagalog and some other Philippine languages. When the Absolutive focus (ancestral Nominative) is a location, both agent and patient are marked “*nV*” (ancestrally “from” or agent). Even when the English sentence has “to” the wording in Tagalog can be “*nV*”, whose basic ancestral meaning was “from”.

The Philippines came into American possession as an almost accidental effect of the Spanish-American War (and Hawaii was admitted, too). Leonard Bloomfield took a quick look at Tagalog (the dialect of the capital, and, later, the national language). He called the Ergative voice “passive” (not unreasonably, as its phase 2 grammar descended from the passive of the earlier phase 1B).

The Absolutive case 1, marked by “*a-ng*”, descends from the PAn Nominative, and behaves much like the same case in Mayrinax. Case 3, *sa*-marked here, is basically locative and general oblique. Case 2, the ancestral Genitive case, Ergative in Mayrinax, has that role here, too. But (because three is a very small number of cases for a language-- we can call Tagalog case-impoverished) that case, marked by “*n-a-ng*” has come to be used for a broader syntactic and semantic range than it has in most other tongues. See McGinn’s (1988) comments.

Most remarkably, when case 1 is used for locative or other oblique function, Ergative “*ng*” (oddly spelled [“*ng*”]), case 2, “*ni*” with a proper name, is used for both the doer agent subject and the undergoer patient object in the same clause, as follows (Schachter & Otnes, 1972, 315):

Hinahaluan ni Rosa ng asukal ang kape sa kapitera
 stir by Rosa CM sugar Abs coffee loc pot
*Rosa stirs sugar into the coffee in the pot. Or literally:
 The coffee in pot is stirred sugar into by Rosa.*

18 Basay, an extinct Formosan language Nom Ac, stage 1B

In a valuable and outstanding paper, Li (1999, 641) shows Basay’s syntax. The Agent Focus markers are used not only in transitive but also in intransitive sentences, showing S marked the same as A, thereby NomAc, as in this sentence:

isu c-um-aka-cakai.
 thou AF-RED-walk
You (sg) kept walking. (not demoted)

Li’s careful study of Basay was based on notes taken by Japanese in 1936. Japanese does not mark word boundaries. The language had been called extinct since 1900 but two old women who could still speak some of it were found. As we’ll see, Basay is still in Stage 1, NomAcc. A crucial test is to find sentences that are passive/ergative and see if the agent is deletable passive (stage 1B) or not deletable (Ergative in stage 2A).

Of Li’s two examples, # 47 LocFocus “You simply play the flute”, has the agent is not deleted. Indeed, the agent “you((sg)” is expressed twice: most archaically by a “Genitive” pronoun suffixed to the verb, and by an independent “Neu” (Nominative) pronoun:

siamammang isu tuLni- an - su. (Tr Text 4)
 simply thou/Neu play- LF- thou/Gen
*You (sg) simply play-the-flute.
 [thou simply playest by thee]?*

This not truly transitive, as no object is expressed. We see dual contradictory marking of agent, which we’ll also see in some Pohnesian Outlier tongues. Agents expressed as affixes sometimes are relics of an earlier syntax (RErg?).

The other, # 11, shows the crucial deletion of agent:

vla - vlai - ana kimu u saquL
 Red give LF you/Nom Obl fruit (Ret.Acc.obj)
You were given fruit [at]. Thou wast (etc).

As the agent is deleted, the syntax is 1B, not 2A. That shows that Basay is 1B passive and is essentially PAn. Both these sentences are Locative Focus, which is commonly used with weaker transitivity. Li also shows some words with MP phonology, probably loans.

19 Seediq

Seediq and Atayal are grouped together within Atayalic. The passive translations with dangling prepositions seen here are the same as those we see for Mayrinax. But we’ll see that Seediq is more archaic and closer to PAn than Mayrinax is.

Seediq is described by Chang (1999). First he gives us a sentence in NomAc phase 1 or Agent Focus (Active Voice):

ciga	wadu-ku	m-ekan	ido
yesterday	Past-1sgNom	AV-eat	rice

I ate rice yesterday.

This looks like a normal transitive sentence in NomAc in a Phase1 language; except for the fact that the agent subject is not expressed as an independent word, but only by a so-called agreement marker suffixed to the verb-- because it's a personal pronoun. Is "rice" an incorporated noun? My suggestion is that it's the PAn direct object which has not yet had the "tu" preposition inserted as its casemarker as in Liao's sentence. So this seems to represent a very early form of PAn.

Notice that the AV/AF is not a demoted voice in Seediq. So in that respect Seediq is more conservative and Mayrinax shows later changes. Seediq looks like stage 1B, NomAc with a passive voice, an intransitive voice, agent deletable. Or one could call this sentence intransitive because the rice is incorporated into the verb. [This is clearly a simple active voice in a NomAc language, in stage 1. It's a straight retention from the earliest NomAc language, surely must be called PAn, because it is obviously in cycle 1 and not a later cycle. Chang does not give examples of AF with a noun subject.

The patient undergoer "rice" has no preposition nor casemarker. It appears to be a simple direct object, unless it's incorporated into the verb, which it would make intransitive.

More problematic is sentence (5) said to have bound pronouns undergoing clitic climbing and base generated as agreement affixes:

kuxun	-	su	-	mu	q-um-ita	isu
like (PV)		2sgNom		1sg Gen	see (AV)	2sg Acc

I like to see you
[my reanalysis: literally, choice:]

like	thee	I	see	thee
like	thou	by me	see	thee

20 The genitive singular may be a relic of earlier Erg.

The so-called agreement marker on the verb is an old fossilized reduced form of a personal pronoun. For a noun subject the personal pronoun would be in third person. But evidently (from this example) if the subject agent is in first person, it's expressed only as a suffix on the verb, and the equivalent expression by an independent word is suppressed.

Or perhaps there are examples in which (as with third person noun subjects) the first person independent pronoun is expressed, though Chang did not give any examples thereof. How about passive voice and Non-Agent-Focus?

In Seediq as in Mayrinax Atayal, and many Formosan languages, the patient (non-active)-voice verbs have focus-marking -un on the verb.

Chang quotes from Holmer, (1996):

subet-	un-	mu	ka	ricahj
hit -	PatVerb	Gen 1 sg	Nom	plum

The plum will be hit by me. (I will hit it)

subet-	an-	mu		
hit	LocVb	Gen 1sg		

The field will be hit plums in by me. (I will there)

biq-	un-	saku (*su-ku)_	pila	
give	PatV	2sGen + 1s Nom	Nom	money

The money will be given to you by me.
(? I will give you the money.)

Also it's odd to see Genitive (from, by) translated 'to' you. The reason may be that, like Mayrinax, Seediq has impoverishment of number of cases. Finding two non-coreferential Nominative elements (not equational) in the same clause is puzzling. Now see:

biq-	i-	maku (*namu-mu)	pila	
give	PV(Imp)	2P.Nom - 1S.Gen	money	

The money should be given to you (pl) by me!

Note the English translations with dangling prepositions, as we had to give for Mayrinax. Chang doesn't say whether the agent can be omitted. If they can, the syntax is 1B and NomAc and close to PAn.

All Chang's agents of passives are genitive pronouns. The study raises some interesting questions. More data would help. The crucial question is whether the agents are deletable (passive) or core terms (ergative).

21 Conclusions on archaic languages

Basay is clearly 1B NomAc and close to PAn. Seediq's sentences pose problems, as the analysis hangs on the deletability of its agents, which we are not told.

The important point is that in at least some Formosan languages the set of focus markers on the verb for intransitive sentences is the same as the set of focus markers for transitive Agent-Focus sentences, which had ancestrally been the NomAc sentences in stage 1A's active transitive voice. That fact groups the A and S elements powerfully and fundamentally together in contrast to the O element. So, by definition, the ancestral syntax is NomAc. It's the active voice NomAc syntax found in ancestral stage 1A, and the words are still there, though a reanalysis may have taken place.

The whole derivation process we have shown, from stage 1A to 1B to 2A, involves the Nominative cases of 1A and 1B becoming the Absolutive cases of stage 2A and 2B. The Nominative cases from the old active voice become the Absolutive cases in the AF, Agent-focus sentences, and the Nominative cases from the old passive voice become the

Absolutive cases in the NAF, Non-Agent-Focus sentences, the basic voice in Ergative languages.

As noted, the prevailing opinion had been that, because most of the more ancient-appearing languages in most of the primary branches of AN have Ab-Erg syntax (in their NAF voices), Proto-AN must have been AbErg. But, as we see, the syntax preceding those AbErg languages (and still visible there as a dying archaism) was NomAc PAN. The NomAc appearance still survives today in the AF sentences, where the Absolutive element is the Agent. But *Marinax Atayal* is clearly AbErg, state 2A.

As Dixon has noted, NomAc languages predominate, and are likely to stay NomAc for long periods of time. How far is *Mayrinax* from the PAN NomAc syntax that shows so clearly in its AF sentences? Unlike Maori, whose ancestry has gone through two full cycles since PAN's Nom Ac syntax and has moved toward late1B near 2A in the third cycle, *Mayrinax* is still in its first cycle, still so close to its ancestral PAN NomAc that the person who has studied it most can still describe it as NomAc. What has frozen *Mayrinax*'s syntax for six thousand years? Freezing in NomAc is common, but freezing in AbErg is less so. It may be that the freezing between PAN and *Mayrinax* occurred in the time when the syntax of early *Mayrinax* was still NomAc 1B.

There's no reasonable way to avoid calling PAN a Nom-Acc language.

Furthermore, my analysis of the Rukai languages, not given here, is not incompatible with Starosta's view that proto-Rukai, NomAc, was one of two basic branches of PAN, also NomAc. One very ancient feature found in Rukai is a particle ancestrally *ka that was the right boundary of a topic at the left of a sentence, before the particle moved to the left of the phrase. There is a possibility that Rukai's NomAc is a whole cycle earlier than the NomAc ancestor that can be seen in the Basay study. On the other hand, there is cogency in Blust's harsh disagreement with Starosta.

22 Padoe in Van Den Berg's work: an exciting finding

We've mentioned it above, but Berg's work is so important and exciting that we give more details.

This involves a chain of dialects, but not the kind from a group settling on a coast line, along which the newcomers spread. In time, people in adjoining villages still understood one another, but not the ones at the two ends.

In Berg's case, MP AN tribes were migrating South-East from West Central Sulawesi (Celebes) toward where Muna and Wolio are now, leaving a series of villages. The one at the Northwest, at the origin, kept the oldest syntax. The next village changed its syntax slightly, moving from an AbErg syntax toward a NomAc one. Each successive village moved a bit more toward NomAc. (This series is a great place to study the series of small changes that add up to a change in syntactic type of sentence.) Oddly, each village kept its own innovation and made no more, with only the innovators moving farther SouthEast. One village, Padoe, achieved the point of balance between AbErg and NomAc, with mixed ergativity (mixed intransitivity), in which the active intransitive subjects took old-Ergative-new-Nominative marking, while the less dynamic ones kept old Absolutive marking. It's the only place known in the world in which each successive village has fossilized its own step through phase 2B toward NomAc syntax.

Other Western MP languages, including the ancestors of Malaysian and Indonesian, made their own steps into Accusativity in less striking ways.

23 Contributions of the Wouk-Ross 2000 book and the aftermath of the first reanalysis from ergative in 2B to nominative in 1A

None of the three authors reviewed here uses the concept of the diachronic repeated movements of languages around the cycle of Ergativity, the basic concept that Dixon proposes. But their findings support that concept and the approach used in the present study.

24 PAn's ergativity: support from an unexpected source

In January, 2002, four months before this SEALS meeting, I attended and gave a paper at the 9th International Conference on Austronesian Languages (as well as one the following week at the COOL-5 Conference on Oceanic Linguistics), both sponsored by RSPAS at Australian National University. Citing lack of budget, to everyone's surprise, the host group, RSPAS of Australian National University, announced that the papers will not be published. The lack of publication is a disservice to the scientific community. Each of the eight previous ICAL meetings had seen publication of the papers by its sponsoring organization. Those who had submitted papers had done so with an implied promise that at least half the papers given would be published.

In presenting a paper at ICAL9, I mentioned my view that PAn was NomAc. As usual, several members objected because most of the archaic AN languages, including most Formosan ones, are AbErg and therefore it seemed that PAn must have been AbErg. That's not a logical conclusion. It's clear that at some points, between one time phase and the next, a change of syntax takes place. All we can reasonably expect is that the later phase should show traces of the earlier. And it does.

At the meeting, copies were sold of the just-printed 474-page paperback book, *The History and Typology of Western Austronesian Systems*, edited (and partly written) by Fay Wouk and Malcolm Ross. On page 34, inconspicuously and without emphasis, the book supports the conclusion that PAn had NomAc syntax. Ross says:

“The Formosan data indicate that intransitive verbs had the same morphology as actor voice [AF] transitives.” (and in a footnote remarks that [as I have noted, too], the situation is less clear in Tagalog and other Philippine languages). It follows (though Ross doesn't specifically say so) that by Dixon's rule, the AF sentences of Formosan languages are NomAc. Welcome, Malcolm!

In the present paper we have seen that the AF constructions in Mayrinax Atayalic Formosan (and Basay and other tongues) are descendants of NomAc constructions in the preceding Stage 1A, which must be called PAn. The part of the reasoning that Ross didn't use is what Dixon (op. cit.) called “moving a language around the cycle of change”.

Following, in the same 2002 book, is Bob Blust's “Notes on the history of ‘focus’ in Austronesian languages”. Like some others, Blust finds the AN languages odd in having three passives while English and other IE languages are said to have only one. We can only wonder why so many English speakers are unaware the English has multiple passives much like those of AN languages. Our common passive uses the direct Accusative object: “I saw the cat. The cat was seen (by me)”. We also have a dative passive: “The board gave John an award. John was given an award (by the board).” In the latter, “an award” is a “retained accusative object” also found in some AN languages. (Mayrinax uses the preposition / casemarker that has been moving toward Acc casemarker.) In English we also have a prepositional passive, which with various prepositions uses the location, the

instrument, etc., as new Nominative subjects: “This house has been lived in.” “This knife has been cut with.” “This one has been spoken for.” English and the AN languages differ slightly in the actual or potential number of passives. A minor difference is that “This house has been given a party in” (with retained accusative object) is acceptable in Tagalog (Dick McGinn, 1988 and pers. comm., and the section on Tagalog in the present paper), but is not accepted in English by many speakers. It’s true that English doesn’t put focus affixes on the verb. Instead, we have prepositions dangling at the end of some sentences. It has not been determined what ancient phrases have been collapsed into those Focus markers.

25 Note on some forms from Donohue

Also, in the same book, the set of forms shown in Donohue’s (2000) report is intriguing. How does it fit into our paradigm of four phases per cycle? P. 85 shows:

Transitive verb without [O] enclitics: [his number 11a]

No-kiki’I te iko’o na beka
 3R-bite CORE you NOM cat
The cat bit you.
 This appears to be:
 3s-bite Acc you Nom cat

Tukang Besi is off the Southeast corner of Sulawesi, near Wolio and Muna and about the point at which Central MP languages begin. We have noted a tV preposition that easily approximates an Accusative casemarker or becomes one. And we have noted that the reanalysis in moving from 2B to 1A (new cycle) that changes the Ergative marker (nV) to a Nominative marker. It seems that the agent “cat” has been reanalyzed from Ergative to Nominative (as typically happened in the move from 2B to 1A) and, of course, keeps the (reanalyzed) nV marker that had anciently been the Ergative agent marker as well as a “from” preposition. So in the second full cycle, the marker “nV” is not primarily an Ergative marker but a Nominative marker. (As we’ll see, Reid and his student forgot that and misinterpreted a nominative nV marker in Fijian as Ergative. Forgive them, because the nV preposition “from” persists and can confusingly recreate nV ergative phrases in the second cycle.)

So Tukang Besi seems to be a language in stage 1A early in its second cycle. Not all the sentences that Donohue presents seem fully consistent with expectations. It would be good to see more work on the language within the framework of “moving around the cycle of change”.

26 General course of the later full cycles

We don’t discuss Central MP nor Eastern MP outside Oceanic, nor Oceanic outside Central Pacific, nor Fijian nor Rotuman, all of which I have written about elsewhere.

We’ll see an enormously complex sequence of syntactic changes that took place within the PPN period (from the point when the earliest form of PPN began to diverge from the NomAc syntax of Fijian). I have discussed them in more detail in earlier papers than here.

A new Erg construction arose in PPN from an unusual source and wreaked havoc when it was borrowed into the -Cia NomAc construction, which it reanalyzed to a passive.

We'll review Fischer's promising and interesting proposal, including a new theory of the primary divisions within East PN. Mainline EPN ends Cycle 2 and begins Cycle 3, becoming NomAc 1A by reanalyzing a look-at construction as Accusative. Going farther, Rapanui (Easter Island) has within our lifetime moved to the very border of a new NomAc syntax, 1A, and stands on the border, the end of the line, today. It now has the mixed intransitive construction (in this case fluid ergativity) that is often found when an ergative language in stage 2B is in the process of being reanalyzed as NomAc stage 1A, beginning its third cycle down from PAN. Maori is taking the irrevocable step through the new 1A, 1B, toward ergativity (2A) in the third cycle.

27 Mixed marking after reanalysis into 1A

When the old ergative agent (marked from PAN with "nV", "from" > "of" [genitive], "from" > "by" [agent]) becomes the new Nominative in the new 1A phase, below, we'll note dual contradictory A marking and even S marking in the Polynesian outlier Kapingamarangi. Nouns may use one system of casemarking while pronouns use another.

In a Ph.D. dissertation this past year at the University of Hawai'i Ritsuko Kikusawa (under Lawrence Reid) found some ergative-looking nV agents in Fijian and evidently saw Proto-Fijian as an ergative language. That seems unlikely. We know that the ancestor of Eastern Oceanic had completed cycle 1 by reanalysis of Ergative case to Nominative. If there had been a new Ergative in Fijian earlier than today's NomAc, Fijian would need to have completed Cycle 2, and therefore East PN have completed Cycle 3. That would require an extremely rapid full cycle after Proto Oceanic just before Fiji's NomAc Cia construction. It would mean that the recreation of NomAc in East PN completed not the second but the third full Cycle down (from the level of PAN that is glimpsed through the Agent Focus in Mayrinax). That's not easy to believe. So it's a matter that needs to be investigated skeptically.

Clearly the "nV" particle that Kikusawa and Reid found is a Nominative (not Ergative) marker, as that is what one can expect to find after the reanalysis of old Ergative marker "ne" to new Nominative case marker at the end of Cycle 1 and beginning of Cycle 2. It does not show a whole new ergative phase within Fijian; that's a false impression that could arise from forgetting that Cycle 1 "n" ergative marker had already been reanalyzed as a Nominative marker. And because the old "from" preposition nV persists into Polynesian, there is always the possibility of a new formation of passive agent and then ergative agent marker from it. We shall see that in North Tuvalu (my fieldwork) the ergative agent is marked by "e", from PMP nominative marking, while in South Tuvalu it is marked by "nee" from ancestral ergative marking.

28 The remarkable formation of PPN: fascinatingly complex process

Pawley (1966) discovered the major divisions of PN.

Fijian, Rotuman and Polynesian are the three groups constituting Central Pacific. There can be little doubt that it's basically Fijian that is close to Proto CP. For years a battle raged between those who thought PN began as AbErg and some parts of it drifted to NomAc and those who thought the reverse. Oddly, those who thought PN began with NomAc (including Chung) assumed that PN began with the same NomAc syntax and

marking that East PN languages have today: a NomAc very different from Fijian's. Oddly, nobody tried the obvious approach: to begin PPN with a syntax essentially like Proto-Fijian's and show how all the PN syntaxes can have evolved from it. The solution I developed is as follows.

From a construction with agent marker *ni* plus personal article *a*, developed an agent phrase *nia* > *nea* > *nee*. That last is the agent marker in the Southern part of Tuvalu in 1970, some in Funaafuti (Southern) though mostly in Nanumanga (Northern).

Also, early PPN had an involuntary cause construction (ancestrally intransitive) studied by Hooper (1984) showing:

Patient + verb + preposition *i* + inanimate cause.
She showed remnants of it in current languages.

With passive-appearing marking, she considered it a lexical passive, both semantically and syntactically. Chung (1978) considered it semantically passive, but syntactically active intransitive, a distinction she needed in her work.

When voluntary agents came to be allowed in the inanimate cause slot, followed by the personal article *a*, the construction became one with a human ergative agent *ia* > *ea* > *ee* > *e*: the same vowel sequence as for the nV ergative agentmarker in South Tuvalu. Thus PPN acquired an ergative construction, which progressively came to be usable with transitive verbs in general. Then the new ergative construction was borrowed into the major class of verbs, which had kept the Fiji-type active transitive construction, having suffix *-Cia* [ancestrally "at it"], and it wreaked havoc there, reanalyzing the construction into a passive one with *-Cia* suffix. (It is not related to the "passive" construction that some students of Fijian recognize and others don't.) Thus late PPN got two constructions with "e" agents: passive with the *-Cia* suffix and ergative without it. Experts debated whether the suffix was added to the ergative form to create the passive form, or subtracted from the passive form to create the ergative form, but neither of those proposals was true, and neither was even minimally plausible. Ergative became the dominant construction in Tongan and in terminal PPN as a whole, which had also a look-at construction that was almost transitive.

Samoic-Outlier PN languages got all those constructions. Samoan kept all varieties. The outliers had a broad choice of syntaxes to choose among, with various results. In Proto East PN Rapanui may have been the first to break away from the main line, and possibly the only EPN tongue that has not discontinued using the AbErg syntax till now. Mainline EPN let the AbErg construction fall out of use and reanalyzed the look-at construction as a normal Nom-Ac (with the locative preposition reanalyzed as Accusative casemarker). Mainline EPN divided into a Marquesan-Hawaiian and a Tahitian-Maori, etc., branch.

In summary, some early scholars believed in a drift from (Tongan) AbErg to (EPN) NomAc. Others believed in a drift from (EPN) NomAc to (Tongan) AbErg. Neither side got the point because neither side began with a "PPN" NomAc form much like the Proto-Central Pacific syntax which became the syntax of early Proto-Fijian and early Proto-Polynesian. My proposal of the late PPN ergative arising from the involuntary cause construction is the only one that accounts for all the outcomes and is free of implausibilities.

29 What's remarkable about Niue

pronouns & proper names	common nouns	
a	e < *re	Absolutive
e	he < *se	Ergative

The “e” Absolutive form descends from PPn “rV” from PAn “dV”, which had both Locative and occasional Nominative uses.

Seiter (1982), in his dissertation under Chung, reveals interesting things about the Niue language, which with Tongan constitutes the Tongic primary division of PN. Some of the most fascinating things Seiter never became aware of, because neither he nor anybody else then understood the diachronic phonology. Niuean has two syntactic classes (1) pronouns and proper names, and (2) common nouns. What amazes people is that the form e, which is an Ergative marker in Niue and most PN languages, is the Absolutive marker for common nouns in Niue. The explanation, as we are about to see, is that the two “e” particles came from different ancestries. Pronouns and proper names take “a” to mark absolutes and “e” for ergatives: as is true in other PN languages. The odd forms are those for common nouns: he for ergatives and e for absolutes. What can these be? Knowing that se > he is the singular nonspecific article in PPN, I guessed that “e” (with its ancestor *re) was likely to be the singular specific article. Checking with forms in Tongan and other PN languages, I confirmed it. Samoan has le < *re for singular specific article where most PN languages have forms from *se. PPn had a form *re < PAn, PMP *di. What was PMP *dV became *rV in PPn and in the Tonga-Niue (Tongic) branch of PN the *r zeroed out. In Tagalog’s case system (where the basic locative casemarkers combine two ancestral cases and appears as both di and sa) shows its three cases as i-an, ni-yan, di-yan: deictic pronouns in Absolutive, Ergative, and Locative cases. So Niue’s case forms are not so odd, after all. They all come down regularly from Proto-Malayo-Polynesian.

30 What's remarkable about Tuvalu

When I took a sabbatical leave from the University of Kentucky in 1970-1971 and spent the greater part of it on the island of Nanumanga in what is now Tuvalu, I taught the group of eighteen 11-year-olds who had been selected as potential passers of the British eleven-plus examination. One day near the beginning, I asked the class how to say, “I see you.” in their language. Even though I knew that they were taught to answer as a group, it was amazing to hear a chorus:

“Aku e kau matea ngina koe.”
 Twelve syllables for what was three in English.

Under British administration, with British-trained native teachers, children had been instructed in English but never had any formal instruction about their own language. And the children had not been taught to recite anything in the Polynesian language in a chorus.

It turned out that they gave me what seemed to them the canonical translation, even though in practice the whole thing was almost never said. Nobody had ever analyzed the

matter for them, and they must have acquired intuitively the syntactic analysis and basic theoretical representation (Nom-Ac in this language that is predominantly ergative today). And that's what they gave me (spontaneously, with no time to discuss it among themselves) as the fundamental translation of an English transitive sentence.

It is illuminating that the immediate response of the Polynesian children was to give the basic underlying ancestral NomAc structure in this language in which most sentences today are said with ergative syntax. So in my paper published in the *Journal of the Polynesian Society*, I gave that remarkable conclusion.

With help from an adult I got the parsing: [I added the numbering later]

1	2		3	4	5	6
aku	e		kau	matea	ngina	koe
I	[tense-aspect]		I	see	Cia	thee

The more genuine form of “ngina” was “ngi-a”, which is used in the neighboring island of Niutao. Nanumanga's odd form contaminates it with another suffix, *ina*, used in Samoan, which we need not discuss. The “ngi-a” seems to be something like “at that”, “at it” or “at him”, in which (as often happens in languages) a grammaticalized form takes the old 3d person singular form (a < PAn *ia.).

Indeed, the Nanumanga agent could also occur in one other slot, elements 7 and 8, at the right, as “e aku” (North Tuvalu) or “nee au” (South) [by me]. So the agent “I” or “by me” could appear in any one or two of the three slots but not in all three.

Whenever the maximum sentence was shortened to

2	4	5	6	7	8
e	matea	(ngina)	koe	(n)e(e)	aku

it was an AbErg sentence.

Whenever it was shortened to
 (aku) e (kau) matea koe
 [with at least one of the (aku, kau) pair]

it was NomAc.

Whenever it was shortened to
 e matea koe

it was passive.

The flexibility of choice of syntax was amazing.

Element 3 is *e* in north Tuvalu and *nee* in south Tuvalu. The governor's grammar (Kennedy, 1945) wrote *ne*, as Kennedy was unable to distinguish long vowels (which have shortened to one-and-a-half length).

Elements 3 and 5 are well-preserved ancestral pronouns for nominative and accusative cases. Examination of Fijian (not done here) shows much the same with slightly earlier appearance in some ways, and less so in others.

Nobody had ever taught the children how to answer a question such as the one I asked. But evidently the reply they gave me was so deeply imbedded unconsciously in their linguistic competence that it came out simultaneously in all eighteen children.

The 3-4-5 sequence (agent, verb, patient; or if you like, SVO) is what I have called the inner clause. It's a fossilized version of the subject-verb-object expression which appears (not fossilized) in some of Berg's Sulawesi languages (as terms in main clauses), and in Yap and other Oceanic languages. I was the first to propose (orally, at the ICAL at Leiden) that Yap is an Oceanic language. Somebody who heard me rushed to make that proposal in a publication without citing me as the source. Eventually, as in Samoan today, the subject and object pronouns (elements 3 and 5) become clitics. Synchronic linguists ignore clitics as nonexistent, but diachronically the clitics, as former words, deserve close attention from linguists, as they are the remnants of core words of an earlier phase, and thus show the ancestral syntax. In Samoan, deletable element 3 is the last remnant of the ancient Nominative element, but a Samoan expert misleadingly calls it "Ergative clitic" because it's co-referent with the 78 ergative phrase.

The British sub-governor for the Ellice Islands (the traditional name Tuvalu was restored on independence) had chosen Nanumanga Island for me because it had not been much studied by anthropologists or linguists, and because he admired its local government. By good luck, it turned out to be the one in which the "I see you" question revealed, better than Samoan, the whole range of NomAc and AbErg constructions that were in use at the breakup of PPN.

I was also lucky in the verb that I happened to use in my question. Only a small class of verbs in Nanumanga keeps the full use of the PPN syntax. Most transitive verbs there (and all verbs in most of Tuvalu) use an almost pure AbErg syntax without keeping the older NomAc options that are still used, in one set of verbs, in Nanumanga.

Niko Besnier, in oral communication a few years later, minimized the value of my finding and interpretation. He pointed out that the verb "matea" belongs to a small class (most verbs and nouns are CVCV). Although small or large parts of what I found in Nanumanga are also found in two islands of Tuvalu near Nanumanga, and in some islands of the Outlier group, he was not moved from his feeling that the only important fact of Nanumanga is that most of its verbs use ergative syntax with only a few signs of the ancestral NomAc syntax that may have developed soon after the ProtoPolynesian period.

I didn't deny Besnier's facts and he didn't deny mine. Is it only a matter of taste about which set of facts is more worth trumpeting?

My suggestion is that the final stage of ProtoPN, the beginning stage of Proto-Mainline PN, and the beginning stage of the Samoic and Outlier group (to which Naumanga belongs), kept a state that had both the AbErg syntax (prevailing in the final stage of Proto-Polynesian and in the Tongic group) and the almost transitive look-at construction that in the earliest ancestral phase of East Polynesian was developing toward a new NomAc syntax.

31 What's remarkable about Samoan

From terminal PPN Samoan inherited a look-at construction that was potentially transitive. Samoan is also a language in which the syntax with ergative casemarker “e” for agents is used in speeches in the fono (local council) to attribute moral praise or blame (Duranti, 1981 and later). Nearly always taking a human subject, it can also be used for storms and other causes of disaster, in a personification.

When the missionaries translated the Bible, they mistakenly tended to mark all agents with e and all patients with the locative and look-at marker, the ninth letter of the alphabet. The combination was an innovation and must have puzzled the native Samoans at first. A normal Samoan clause, (like Rapanui's) might have either the quasi-accusative look-at patient, marked with i, or the ergative agent marked with e, but not both. But because of the high prestige of the London Missionary Society's translation of the Bible in Samoan, such sentences have been gaining use in Samoan. The change is consistent with the general diachronic linguistic tendency to lose cues through lenition and then lessen ambiguity by insertion of words.

Studies of look-at in English today have also shown acquisition of some Accusative Object qualities, thus beginning a syntactic change.

32 Dual contradictory agent marking in Polynesian

Elsewhere I have discussed the dual contradictory casemarking seen in Marian Klamer's work on a Central MP language. Here are examples of it, from my teacher Sam Elbert, in Outlier Polynesian languages among those noted in Finney, 1998b (in Kapingamarangi):

au	e	moinaa	e	au	koe
I	TA	cherish	by	me	thee
au	ku	kite	e	au	a mee (Elbert)
I	TA	see	by	me	ART him
au	te	iloo	e	au	
I	not	know	by	me	

Dual marking (like mixed intransitivity) is a process occurring when Ergativity is changing. The Outlier language cited here is one coping with the excessive number of alternate syntaxes left available at the end of the PPN period.

33 What's remarkable about Rapanui

Often regarded as the first branch off Proto-East PN, Rapanui is the only East PN language proven to preserve the PPN glottal stop. Found in Tongan, that stop seems to have zeroed out in the Samoic-Outlier group. It must have been preserved in Proto-East PN to reach Rapanui. In recent years Rapanui requires either marking the agent with e or marking the patient with (former “at” preposition) i but seldom both. Finney & Alexander (1999) called Rapanui “a language at the end of the line” in two senses. In a good sense, as one that fascinates linguists, an opportunity to observe Rapanui moving through fluid ergativity to become NomAc. In a bad sense, as disappearing because all under 25 now have been reared to speak Chilean Spanish.

As we have noticed, the phenomenon of “mixed intransitive” or “mixed ergative” is a situation in which the more active intransitive verbs have AbErg syntax and the less active ones have NomAc syntax. In some languages the mixture is lexically determined; in other cases it’s “fluid” ergativity, i.e., casemarking and syntax are determined by the degree of activity implied in the particular sentence. The mixed situation (no matter whether fluid or lexical) had been described only as a curiosity. But our present paper asserts that it is a transient state found only when the syntax is changing from 2B AbErg to 1A NomAc. And on the Table page, the reader can easily see how it develops and why.

From Alexander (1981:142) here is an example of split ergativity in Rapanui:

He	tahuti	e	te	poki
present	run	TAM	the	child

The child runs.

It is mixed and fluid because when “tahuti” is used for less active running, the subject is not marked with the Ergative “e” marker [new Nominative], nor any other casemarker.

34 An original proposal

I propose that mixed intransitivity (mixed ergativity), no matter whether lexical or fluid, occurs ONLY when a language is changing from AbErg (stage 2B) to NomAc (stage 1A). Of course, it does not occur in languages when they are changing from NomAc to AbErg. Dixon is right in seeing the cycle as moving around in only one direction. It doesn’t reverse course.

We catch the syntactic change and the phase and cycle change (2B to 1A) both in Padoe and in Rapanui (Easter Island). Nobody else has discovered the relation between mixed ergativity and the process of changing from NomAc to AbErg.

35 Steven Fischer’s discoveries in Rapanui and Southeast Polynesian

Mainline EPN is the main group of EPN which divides into two: the Hawaiian-Marquesan and the Tahitian-Maori group. Because of shared phonology, Mangareva and the smaller islands around it had been thought to have branched off from Marquesan (even though Mangareva is on the obvious path to Rapanui from anywhere else in Polynesia.) The traditional view, held by Pawley and others, was that Rapanui was the first branch off protoEPN.

Steven Fischer is the most creative person working on Polynesian today. He knows linguistics from Sanskrit and Greek onward. His books on the history of language are classics. He (2001a,b) made an important discovery in East Polynesian. Comparison of Mangarevan wordlists from different dates show that Marquesan changes appeared only later. So the Marquesans must have invaded Mangareva and partly settled it. He calls it “a dominant intrusive language subjugating a vulnerable indigenous language”. The settlers of Rapanui must have passed through Mangareva before it was invaded by Marquesans. Rapanui’s preserved glottal stop came down from PPN, but was lost independently in most other PN lines.

So Fischer proposes that Rapanui and the Mangarevan group together separated from Mainline EPN as the first separation from Mainline EPN. A possible objection is that

in the EPN group, Rapanui alone keeps the ancestral glottal stops. But glottal stops are easily lost independently; nor is it sure that Mangareva does not have the glottals, unwritten.

Fisher advances the concept of (proto)SouthEast Polynesian, to consist of Mangareva and smaller islands nearby. In his view it's a slightly earlier Protolanguage consisting of the Mangareva group plus Rapanui that first separated from Mainline EPN. He includes Rapanui in SEPN.

One interesting point is that in Fischer's reconstruction of early development of SouthEast PN phonology is a lenition from "h" to glottal stop, one that I have not encountered elsewhere. Rapanui differs from many PN tongues both in keeping the PPN glottal stops and in refraining from lenition of other sounds (k) to glottal stops. Not all would call "h" a fricative. If a fricative it must be a laryngeal (glottal) fricative. Both sounds are about the end of the line in lenition: both of them weaken directly to zero. Normally a stop weakens to a fricative, not the reverse. I'd like to hear from anybody who knows of a change from glottal stop to "h".

It has been noticed that Rapanui preserves two features from terminal PPN that are lost in Mainline EPN: the ergative case and the glottal stop. Fischer's theory, that Rapanui and Mangarevan formed a group that left Mainline EPN together, poses a question. In separating from Rapanui, did Mangarevan (known from dictionaries not written by linguists) lose both the ancestral glottal stop and the ergative case? Will Fischer's theory be stronger if it turns out that Mangarevan keeps the ergative case? Fischer notes that Mangarevan had a particle "e" (with or without glottal?) used before nouns, but in the absence of sentences it is not known whether the particle marked a passive agent (as in Mainline EPN), which would not have a glottal, or an ergative agent (as in Rapanui) which would. Somebody could do a good masters or doctoral thesis by sailing to Mangareva, resolving these issues, and clarifying its phonology and syntax.

36 What's remarkable in Mainline EPN: entering the third full cycle

Mainline EPN, which includes all of East PN except Rapanui and the Mangareva group, divides into two branches, one that includes Marquesas and Hawai'i and another that includes Tahitian, Maori and much of Tuamotu.

One early and excellent study was done by Patrick Hohepa (1967a,b), who was a native speaker of Maori, and one of the first scholars of Polynesian to use generative grammar. Caught in the argument whether early Polynesian languages began as AbErg and drifted into NomAc, or began as NomAc and drifted into AbErg, Hohepa chose drift from East PN NomAc to Tongan AbErg. Neither scenario was precisely true, but Tongan must have come close to its present form (AbErg) before the East PN languages came close to theirs (reinventing NomAc by replacing the simple direct object with a new structure, inserting a locative preposition). This may have been facilitated by the fact that, like Samoan, Proto East PN may have inherited the incipiently transitive look-at preposition which used the same marker, *i*.

Hohepa correctly concluded that Maori was a NomAc tongue that showed a characteristic that seemed to point to a change toward AbErg. That was, the use of passive voice more often than the active. He was right in principle. Maori has begun that path. But he was mistaken in thinking that Tongan was ahead of Maori in the same movement. In fact, Tongan reached ergativity in phase 2A of the second full cycle. Maori's ancestor,

Proto-Mainline PN, at the end of the PPN period, must have had an AbErg voice in common with Tongan and all PN. But it dropped out of use long ago in Proto-Mainline EPN and the ergativity that Maori is headed toward now will be the phase 2A of the third full cycle down from PAn.

As we have seen, the phenomenon of “mixed intransitive” or “mixed ergative” is a situation in which the more dynamic intransitive verbs have AbErg syntax and the less active ones have NomAc syntax. In our survey, the only two examples thereof occurred when a change was taking place from AbErg to NomAc syntax to begin a new full cycle. Like Padoe, a whole cycle earlier, Rapanui is at that point today. The S had been marked the same as the O, with AbErg syntax, but when the former AbErg marking comes to be used with all the intransitive sentences, by definition the old Ergative case in Rapanui has just become a Nominative (just as Padoe has done).

It is impressive that Padoe and Rapanui are the only examples we found of mixed intransitive (mixed ergative) syntax; and that in both cases, just now though a full cycle apart, it occurred only in the reanalysis of old Ergative to new Nominative case that creates now the new state 1A, ending one full cycle (of thousands of years) and beginning another.

37 Construction rejected for English passive but fully acceptable as Tagalog ergative (or passive)

English accepts “This house has been lived in” but not with a retained Accusative object, as in: “This house has been given a party in.”

But in Tagalog, as Dick McGinn (1988 and pers.comm.) points out, that construction with retained object is fully acceptable.

The following examples of locative focus (LF) are given by Schachter & Otnes (1972:317). My translation keeps the Abs focus as subject. Schachter’s (in parens here) makes the Tagalog ergative phrase the English subject. [These are sentences in which Tagalog puts both agent and patient in the ergative case.] English “in” or “on” is needed to translate the locative focus marker on the verb.

Pinagbayuhan nila ng palay ang lusong.
 pounded (LF) Erg-them Erg rice the mortar’
The mortar was pounded rice in by them.
 (They pounded rice in the mortar.)

Pinaglalaruan namin ng tses ang mesang maliit.
 played (LF) Erg-us Erg chess the small table
The small table is played chess on by us.
 (We play chess on the small table.)

Pinaglutuan ni Helen ng karne ang kawali.
 cooked (LF) Erg Helen Erg meat the frying PAn
The frying pan was cooked meat in by Helen.
 (Helen cooked some meat in the frying pan.)

All these sentences show retained direct objects marked in Ergative case in this language with too few cases.

Note also in Chang's (1999) Seediq, p. 361, in an AbErg sentence with a retained Accusative object for which no case is named (in this Ab-Erg language):

subet - an - mu ricah ka neepahah
 hit loc vb mrkr 1SGGen plum Nom field
The field will be hit plums in by me.

We could say (as we say of Maori later in this paper) that this is stage 1B with passive becoming more common than active voice, and thus moving toward reanalysis into stage 2A. Or we could say that (like Mayrinax's AF) this is an archaic remnant of 1B syntax in a 2A language. If the latter, it is comparable to the occurrence of the obviously archaic AF construction in Mayrinax.

Typical conservative Formosan and Philippine AN languages have ergative / passive voices much like those three basic types. The patient passive has the verb suffixed with -un or -in, presumably from the PAn obscure vowel. The locative passive has the verb suffixed -an. The element 'an' also occurs as an independent locative word; and as a noun-forming suffix that ancestrally formed locational nouns (where the activity takes place) but in some languages has come to form general verbal nouns. In some languages the dative focus takes the locative focus form. In others, the locative focus noun represents a patient (theme), with little or no implication of location, but with (perhaps) a lesser degree of transitivity than is conveyed by the patient focus suffix. My tentative hypothesis is that the -un, -in, and -an elements were ancestrally phrases Vna meaning something like 'at that' or 'at it' [coindexed with the (Focus) (Nominative) (Absolutive) NP].

38 Preposition nV from PAn to Polynesian

The ablative directional preposition nV "from" (a concrete location) came to be used abstractly with objects that were sources in any sense, causes in general.

When an ablative 'from' preposition acquires both genitive (possessive and 'of') uses and agent uses (in passive or ergative) it often loses ablative 'from' uses. But 'from' usage persists in some tongues all the way down to East Polynesian. In Maori, for instance, nV translates 'from', though not used with verbs of motion (Williams and Williams, 1956).

Partial survival of 'from' meaning amid changes is also seen in Southeast Ambryon in Vanuatu, an Oceanic language (Parker, 1970). Preposition ni is glossed 'with, by means of, for (in acquiring), than', while raNi is 'from'. So we see that PAn nV (taking mostly i vowel, as other prepositions do in CEMP) has lasted in the AN family many thousands of years, even when branches of it have diverged into Ergative, whose descendants can be reanalyzed into Nominative cases that keep the "nV" marking that meant "from" in earliest PAn and still does in many AN languages to this day.

Languages that lost the ablative 'from' meaning of the nV particle were left with two usages thereof: Genitive and Ergative (i.e., agent not in the unmarked case).

The result of all this is that even after the casemarker "nV" has been reanalyzed from an Ergative marker to a Nominative marker, the "from" meaning often remains and in the next cycle can be used to create a new marker of passive agents and then a new Ergative case marker.

39 Incidental sidelight

There is an interesting sidelight to the canonical method of creation of ergative case by reanalysis of passive agents. One AN language was reported (in a masters' thesis) to be at least weakly ergative with no evidence of derivation from a passive. The same demonstrative marks S and O, while for A, the marker is zero. For specifics, contact my source, David Ross Clark, a born Canadian who, after getting his Ph.D. in California, moved to Auckland University in New Zealand.

40 Starosta's statement of position

For our purposes, the best account by Starosta of his view is Starosta (1995).

As we have seen, Dixon, like most linguists, takes NomAc as the unmarked form for languages in general, but Starosta's position seems to differ. Dixon says that NomAc seems intrinsically favored. Is Starosta contradicting that when he classifies all intransitive subjects as semantic patients? Note, however, that he labels those subjects (syntactically) as "actor", not as "undergoer".

In response to a question, Starosta wrote to me (pers. comm., July 24, 2000, two years before his death which occurred about the time of this 2002 SEALS meeting). I have not checked his references, and so they are not given at the end of this paper.

"Here are some answers to your question. Starosta, 1988, 126-128. Patient (formerly also Object or Theme): The perceived central participant in a state or event. Patient centrality. Lexicase assumes there is a Patient in the case frame of every verb, where "Patient" corresponds to Halliday's MEDIUM.

"Every process has associated with it one participant that is the key figure in that process; that is the one through which the process is actualized, and without which there would be no process at all. Let us call this element the MEDIUM, since it is the entity through which the process comes into existence...in a material transitive clause...the medium is obligatory in all processes; and it is the only element that is. [1985, 146]. The idea of requiring every verb to have a Patient...is partly a lexicase internal development (Taylor 1972, 37) but was strongly influenced by Gruber's (1965) work which I first encountered in Jackendoff's *Semantic Interpretation in Generative Grammar* (Jackendoff 1972; esp. pp. 29-31), by John Anderson's analogous use of nom as an obligatory term in every case frame (1961:37), and by Halliday's (1967-1968) transitivity and theme analysis..."

Though my 1998 paper on Chamorro refuted Starosta's inclusion of that language in his controversial "treetop" set, a great deal of Starosta's work is fundamentally sound. Blust once looked with me at a phase 2B antipassive sentence marked intransitive, and objected that the sentence was obviously transitive with Ergative subject and Absolutive object. The problem was that it was one of the languages in which the preposition for the demoted antipassive can be zero. I told Blust that the only way to make the distinction is to ask a native speaker whether the presumably demoted Abs patient could be deleted. If so, the construction is antipassive and hence intransitive (as passives are), and the phase has moved from 2A to 2B.

41 Conclusions

1. Languages that have a passive or ergative voice vary in the number of such voices. They always have the canonical one that puts the Accusative patient undergoer into the Nominative slot while demoting the active Nominative subject to an adjunct. English passive voices differ somewhat, but not fundamentally, from the ancestral passive voices (stage 1B) that underlie the ergative voices found in stages 2A and 2B in Austronesian languages.
2. By Dixon's definition, languages such as Mayrinax Atayal and Tagalog are ergative, stage at least 2A and often 2B. To reach that conclusion we must first decide which voice is the basic one for applying the definition: the AF or an NAF. In stage 1B the Nominative subject of the Active is the basic grammar, the one generated first, and Dixon's rule confirms the syntax as NomAc. The passive is secondarily derived. But in stage 2A and 2B, the descendant of that old construction, the AF (Agent Focus), which still looks like NomAc, with identical marking for intransitive subject and for agent, is a demoted and a derived construction. So what the marking shows is not that Mayrinax is NomAc, but that Mayrinax's ancestor (PAn) had NomAcc syntax.
3. Evidence from Li's brilliant study of the dead Basay language clearly shows it to be 1B NomAc with deletable passive agents, very close to NomAc PAN as argued by Starosta.
4. Despite its obvious strong ergativity, Mayrinax Atayal also clearly shows ancestral PAn NomAc in its archaic dead-end AF (Agent-Focus) construction, in which intransitive S and the transitive Agent are marked alike, fulfilling the definition we all follow from Dixon. We call Mayrinax AbErg because of what holds true in its basic voice, the NAF, Non-Agent Focus (and specifically the patient-focus). The NomAcc appearance shown in Mayrinax's Agent Focus is a demoted relic of the ancestral PAn NomAc syntax. Some of this can be seen in other languages such as Basay and Tagalog. The demoted AF construction must fall into disuse before the next 1A.
5. From a generative point of view, today's AbErg patient-focus construction (in AbErg languages such as Mayrinax) must be generated first and the NomAc appearing Agent Focus construction be derived from it synchronically. This is true even though historically the Ergative arose as a reanalysis of a passive voice which, of course, was secondary, synchronically derived at the time from the ancestral active voice NomAc construction that survives as the AF.
6. What we show as the normal route to ergativity is not the only route. One AN language created a (weakly) ergative construction with no sign that it developed from a passive. Dixon was right. There is more than path.
7. Return to NomAc syntax (typically after creation of an antipassive) differs sharply from return to AbErg syntax after creation of a passive voice.
8. Stage 1A shows no obvious trace of its immediately preceding 2B syntax. That was wiped out in passing through mixed ergativity, reanalyzing Erg case to Nom. The Reanalysis of Ergative agent as Nominative is the sharp change that conceals it. That is one reason that it makes sense to define that each new full cycle begins at the point of restoration of NomAc in stage 1A, not at the fuzzy restoration of AbErg in stage 2A. Mayrinax and other 2A tongues keep sentences looking like the

- passive clauses in their recent ancestor 1B. Mayrinax also keeps archaic demoted Agent Focus clauses that keep the wording that had been NomAc in stage 1.
9. Change back to NomAc, ending the first full cycle, is most obvious in van den Berg's work. His showing of five villages is the only case I know of anywhere in the world where a dialect chain remains geographically in order from the most archaic syntax to the most changed. It is made even more remarkable because it clearly identifies one village, Padoe, in the middle, having mixed ergativity, at the point where AbErg changes to NomAcc. That author is one of the most sophisticated linguistic experts working on Austronesian languages today.
 10. Proto-Polynesian acquired a new AbErg 2A system in a complex way, but still kept other syntaxes as the common ancestor lost the old Fijian-like NomAc. It became neither an exclusively AbErg nor a fully NomAc language before it split up.
 11. It's not easy to summarize briefly the complex sequence of changes that took place within the Proto-Polynesian period. Its ergative does arise from a passive, but it's a lexical passive, not one with an active voice. It seems fair to say that at the end of the PPN period Ergative was the predominant syntax, though PPN kept the optional use of Fijian NomAc syntax that we see in Nanumanga, and had a look-at construction that could become accusative.
 12. One sub-branch of PEPn, Rapanui, at the end of cycle 2, has reached the point of return to NomAc by acquiring mixed ergativity (specifically, fluid ergativity), and then reanalyzing the old Ergative as the new Nominative to begin cycle 3. This is exactly what Padoe has done in ending its slow-moving first full cycle to begin its cycle 2.
 13. Perhaps the major discovery reported in this paper is that only two instances of mixed ergativity were found (Padoe and Rapanui) and both occur at the point of a major change in syntax, and specifically at the end of a full cycle and beginning of the next, as 2B changes to the new 1A.
 14. Mainline EastPn returned to NomAc by letting the ergative syntax drop out of use, and by letting a locative construction be reanalyzed as a NomAc transitive.
 15. It's by a different mechanism that Maori is interesting in that it is just at the point in 1B when passive voice comes to be used more often than active voice. That may be the normal event that is to be followed by the reanalysis of the passive term (a mere adjunct) into Ergative (a core term), the reanalysis that changed the phase from 1B to 2A, the path of inevitably moving toward a new Nominative. The fact that we don't see other languages in Maori's late 1B state may mean that languages pass rapidly through it by entering the reanalysis of passive to ergative that changes the state from 1B to 2A.
 16. Steven Fischer is on firm ground in his discovery that Mangareva's language was severely impacted in the phonology of its vocabulary. He worked with dictionaries and not with elicited sentences. His view is plausible that Rapanui belongs in the group with Mangareva (and not in the mainline EPN group which includes the Marquesas-Hawaiian and the Tahitian-Maori groups). He denies that Rapanui and Mainline East PN formed a group in opposition to Mangarevan.
 17. Both in Nanumanga, Tuvalu, where I did fieldwork, and in Rapanui, for which I wrote a review of DuFeu's (1996) grammar, we find confirmations that after a syntactic change, the remnants of use of the old syntax have come to be generated

from the new syntax (a reversal of the direction of the generating process). This may be a rule of diachronic generative grammar.

42 Degrees of confidence about the conclusions

Conclusions vary in the degree to which they are backed up by evidence.

When the antipassive transformation uses an “at” preposition, that’s certainty. In languages that make that development (demoting the patient to an adjunct) without inserting a preposition, we can only confirm that it has taken place by asking a native speaker whether the patient DP can be deleted in the sentence.

There is more direct evidence in deciding whether a construction has developed from 1B passive to 2A ergative. The agent in a passive is oblique and deletable, while the Ergative agent in an ergative construction is not. In Huang (1944b) I didn’t find any complete sentence that was passive without an agent expressed. (Page 79 has a patient focus expression glossed “the person was beaten by someone” but it was one in a set of five incomplete sentences all of which [the others active] were filled in with “the person” or “someone” or “something” to replace missing parts of the sentence.) That carries some weight toward calling “AbErg” the NAF sentences, which she translates active voice in English.

Were we justified in calling Mayrinax’s AF construction a demoted and dying one, and thereby using patient NAF as the essential transitive construction for applying Dixon’s test and thereby classifying Mayrinax as AbErg? We do so because in each move we observe from 2B AbErg to 1A NomAc, the new NomAc is not a survival of the earlier NomAc but is a new formation (the reanalysis of something else). Languages drift around the circle in one direction only. Both in Padoe and in Rapanui an earlier 2B Ergative case is being reanalyzed into a new Nominative case, moving through mixed intransitivity. In Mainline East PN (after PPN developed more constructions than any language needs, and the Fiji-like NomAc was bombed out) the new active NomAc was a reanalysis of a look-at construction into an Accusative object, while the ergative construction simply dropped out of use.

There is something intrinsically elusive about the definition of antipassive as the demotion of the Absolutive patient from a core term to an oblique adjunct that is deletable. One problem is that the Absolutive is in some sense the Subject of the sentence and as such can’t be zero. Gault’s (1999, 398) examples of antipassive show preposition *ma* (OBL) inserted to demote the Abs subject DP that had come down from 1A’s Accusative object. Gault says, “It [the antipassive construction] significantly reduces the transitivity of the underlying transitive clause. In many instances the resulting construction is clearly an intransitive with the patient in an oblique phrase; in other instances the patient does not receive oblique case marking, but its topicality has been significantly reduced. It may be possible to argue that because of the reduced topicality of the patient, the antipassive is a true intransitive even when the patient does not receive oblique case marking. ...[I]t is clear that the antipassive significantly reduces the topicality of the patient and so for the purposes of this paper it will be considered a type of intransitive.”

Mayrinax is clearly beyond phase 1B and Ergative because the passive agent adjunct of 1B has been promoted to a term. That’s enough to show that it is not NomAc, though its ancestral NomAc [PAn] has been so and the ancestral PAn NomAc still shows through. Huang (sometimes) labels the Abs as Nominative and the Ergative as Genitive (as

indeed they had been in phase 1B). From Mayrinax's present state demoting change of the Abs patient to an adjunct would make it 2B. Thereafter, loss of the AF (with reanalysis of Ergative case to Nominative) could move it into 1A in a new cycle. But we don't see evidence moving Mayrinax beyond 2A. Huang translates her presumably AbErg NAF sentences into both active and passive English sentences. So, AbErg stage 2A is the label that fits Huang's evidence best. Having been burnt in controversy for changing her position on Atayal's NomAc or AbErg status, Huang has cautiously protected herself in recent years with important new data but chosen not to use the words that would assert or deny the ergativity of the constructions.

43 For comparison: relics of ancestral ergativity in English & Latin

Notice that in English as well as Latin, neuter nouns and pronouns have the same forms for Nominatives as for Accusatives. As we know, the former Nominative of a passive becomes Absolutive in phase 2A. Latin neuter nouns and PIE had -m (Greek -n) and -d for Latin pronouns (id, quid, quod, istud, illud, aliquid) whose cognate in English is -t for neuter pronouns (it, that, what). I suggest that that's an old IE Absolutive case ending (d). In some Latin neuter nouns (declensions 2, 3, 4) it's zero after the stem vowel. English nouns are invariant but for genitive, and so not helpful here. Perhaps those mark more than one period in prehistory when Proto IndoEuropean marked Absolutive case (S=O) with zero or -d or -m.

Begin with the thought that (from the reanalysis of old Ergative as new Nominative in Austronesian) the use of the same casemarking for Nominative as for Genitive (Ergative) is a sign that in some ancestral period the Ergative case was reanalyzed as the new Nominative. We find that in Latin the ending is -s for both nominative and genitive in non-neuters in the third declension singular nouns. That may mark a period when Ergative cases, phase 2B, in an ancestor of Latin, were being reanalyzed to Nominative making phase 1A, to begin a whole new cycle, as happened in Padoe and Rapanui. In Classical Latin times, writers Cicero, Caesar, and Vergil, with intentional archaism, used SOV order.

The suffix -s must have been the ergative casemarker in IE as the prefix nV was in AN, and both became Nominative by reanalysis as 2B gave way to 1A. English and Latin's decaying casemarkers at the ends of words are old postpositions from a phase of IE that had OV word order, and the modern prepositions arose in a later period along with VO word order. Essentially all AN is VO.

Note that when English and Latin show relics of ancestral ergativity, they are in certain case conditions and in ergative-looking clause structure. That's because the ancestral ergativity is wiped out in the mixed ergativity and reanalysis at the end of phase 2B.

44 Connection with related language families?

At the age of 76 I have retired from the field and must leave the broader question to others, notably those who have identified which families are related to AN. One thought is that if another family uses the same marking, such as "nV" for "from", that represents what was true of a common ancestral language. But if the other family shows "nV" for a non-subject agent, that may more likely show a similar development that has taken place in parallel with AN's. Reid (1999) is worth reading. In the future, of course, the AN languages that survive will keep diverging from one another in phonology and syntax and more full cycles will follow.

Postscript

In lenition, stops are commonly weakened to fricatives, and then to h, which if it is a fricative, is a glottal or uvular one. Both /h/ and glottal stop are virtual dead ends of lenition; the next step is to zero. English has only three words with phonemic glottal stops, and they all have marginal status as words: (1) uh-uh (“no”); (2) uh-oh (“something bad has just happened”); and (3) “ah-ah-ah” [pitch rising from syllable to syllable and topping and falling in the final syllable] (“Don’t do what I think you are about to do”). A 7 PM weekly radio program, about the year 1940, began: “Ah-ah-ah! Don’t touch that dial. Stay tuned for [name of program].” It’s interesting to find Steven Fischer proposing a phonological change from “h” (a fricative at most, if not a breathing) to glottal stop in Southeast Polynesian. In lenitions, a stop can easily change to a fricative, but not the reverse.

SEALS (number) means Papers from the (Fifth, etc.). Annual Meeting of the Southeast Asian Linguistic Society. OL is Oceanic Linguistics, University of Hawai’i Pr.. PL is Pacific Linguistics, RSPAS, Australian National University. JPS is Journal of the Polynesian Society.

References

- Alexander, James D. 1981. Case-marking and Passivity in Easter Island Polynesian. *OL* 20: 131-149.
- Becker, Alton L. & Thomas Hunter. 1988. Cohesion in Old Javanese prose. Pp. 475-488 in Richard McGinn (Ed.), *Studies in Austronesian Linguistics*. Ohio University Press, with distributor Chicago: University of Chicago Press.
- Berg, Rene van den. 2002, discussion with this writer at the International Conference on Austronesian Linguistics, elaborating his earlier published papers. .
- Blust, Robert. 1999. Subgrouping, Circularity and Extinction: Some Issues in Austronesian Comparative Linguistics. Pp. 31-94 in Elizabeth Zeitoun and Paul Jen-kuei Li (Eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica.
- Blust, Robert. 2002. Notes on the history of ‘focus’ in Austronesian languages. Pp. 63-78 in Wouk & Ross (Eds.).
- Carroll, Vern & Topias Soulik. 1973. *Nukuoro Lexicon*. Honolulu, Univ. Press of Hawaii.
- Chang, Yung-Li. 1999. Seediq bound pronouns: pronominal clitics or agreement affixes? Pp. 355-370 in Elizabeth Zeitoun and Paul Jen-kuei Li (Eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica.
- Chen, Teresa. 1987. *Verbal constructins and verbal classification in Nataroan Amis*. Canberra, Pacific Linguistics C-85.
- Chung, Sandra. 1978. *Case Relations and Grammatical Relations in Polynesian*. Austin and London: University of Texas.
- Dixon, Robert M. W. 1994. *Ergativity*. Cambridge, New York & Melbourne: Cambridge University Press.
- Donohue, Mark. 2002. Voice in Tukang Besi and the Austronesian focus system. in Wouk & Ross (Eds.)
- DuFeu, Veronica. 1996. *Rapanui*. London & New York: Routledge.

- Duranti, Alessandro. 1981. *The Samoan Fono: A Sociolinguistic Study*. Canberra, Pacific Linguistics.
- Finney, Joseph C. 1983. Nanumanga verbal syntax: a study in variation, with implications for Proto-Polynesian. *JPS* 92(1): 107-115.
- Finney, Joseph C. 1994. Casemarking and the *ia demonstratives. Pp. 551-574 in *Proceedings of the International Conference on Austronesian Linguistics*. Cecilia Ode & Wim Stokhof, Eds., Leiden, Netherlands.
- Finney, Joseph C. 1997. The Yap language: whence and how? *SEALS* 5, pp.97-114.
- Finney, Joseph C. 1998a. Fiji's 'o' and 'ko' in context-- some points in the diachronic syntax of Fijian. *SEALS* 6, pp.149-170.
- Finney, Joseph C. 1998b. Dual contradictory agent-marking: solutions to Proto-Polynesian mysteries. Pp. 103-114 in Matthew Pearson, Ed., *Recent Papers in Austronesian Linguistics: Proceedings of the Third and Fourth Meetings of the Austronesian Formal Linguistics Association (AFLA)*. Los Angeles: UCLA Department of Linguistics.
- Finney, Joseph C. 1998c. Chamorro: Whence and How? *ibidem*.
- Finney, Joseph C. 2000a. Two "e" studies in Polynesian. Pp. 405-416 in: *Proceedings of the 4th Conference on Oceanic Linguistics*. Wolfgang Sterling & Steven R. Fischer, Eds. Auckland: Leo Pacifica.
- Finney, Joseph C. 2000b. The True Prepositions/ Casemarkers in Proto-Oceanic. (Given at the Second International Conference on Oceanic Linguistics.) *Pacific Linguistics*. (Bill Palmer and Paul Geraghty, Eds.)
- Finney, Joseph C. 2000c. A language at the end of the Line. in Stevenson, Lee. & Morin, 20001.
- Finney, Joseph C. 2001a. General Diachronic Course of Proto-Austronesian Casemarkers-- Prepositions and Topic Markers: Sources of Austronesian Casemarkers. *SEALS* 9.
- Finney, Joseph C. 2001b. A Language at the End of the Line. *Pacific 2000*: Pp. 405-416 In: Stevenson, Lee, Morin.
- Finney, Joseph C. (forthcoming). An applicative process persisting through Austronesian. *SEALS* 7.
- Finney, Joseph C. (forthcoming). Toward reconstruction of demonstratives in Proto-Austronesian. *SEALS* 8.
- Finney, Joseph C. (forthcoming) Will the real Proto-Austronesian please stand up? *SEALS* 10.
- Finney, Joseph C. and James D. Alexander. 1998. *The Rapanui language of Easter Island: Where does it fit in the Polynesian family tree?* Easter Island Foundation, Los Osos, CA.
- Fischer, Steven Roger. 2001a. Mangarevan doublets: preliminary evidence for Proto-Southeastern Polynesian. Pp. 417-424 in Stevenson, Lee, Morin.
- Fischer, Steven Roger. 2001b. Mangarevan doublets: preliminary evidence for Proto-Southeastern Polynesian. *OL* 40(1), 112-124.
- Gault, JoAnn. 1999. Ergativity, Subject and Focus in Sama Bangingi'. Pp. 393-422 in Elizabeth Zeitoun and Paul Jen-kuei Li (Eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Academia Sinica.
- Hohepa, Patrick. 1967a. A profile generative grammar of Maori. Indiana U. Pub. in *Anthropology & Linguistics*, M20.

- Hohepa, Patrick. 1967b. The accusative-to-ergative drift in Polynesian languages. *JPS* 78:295-329.
- Hooper, Robin. 1984. Neuter verbs, stative aspect and the expression of agency in Polynesian. *JPS* 93:39-70.
- Huang, Lillian M. 1994a. Ergativity in [Wulai]Atayal. *OL* 33, 1:129-143.
- Huang, Lillian M. 1994b, *A Study of Mayrinax Syntax*. Taipei: Crane.
- Huang, Lillian M. 2000. Verb classification in Mayrinax Atayal. *OL* 39: 364-390.
- Kennedy, Donald Gilbert. 1945. *Te Ngangana a te Tuvalu* [The language of the group of eight]. Suva: Government Printer.
- Kikusawa, Ritsuko. 2002, unpublished thesis, U, Hawai'i.
- Li, Paul J.-k. 1999. Some problems in the Basay language. Pp. 637-665 in *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. (Elizabeth Zeitoun & Paul J.-k. Li, Eds.) Taipei: Academia Sinica.
- Manning, Christopher D. 1996. *Ergativity*. Stanford, CA: CSLI Publications (Dissertations in Linguistics series).
- McGinn, Richard. 1988. Government and Case in Tagalog. Chapter 9 in Athens, OH: *Studies in Austronesian Linguistics*. Ohio University Center for International Studies.
- Pawley, Andrew. 1966. Polynesian languages: a subgrouping based on shared innovations in morphology. *JPS* 75:39-64.
- Pawley, Andrew and Lawrence A. Reid. 1979. The evolution of transitive constructions in Austronesian. Pp. 103-130 in *Papers from the Second Eastern Conference on Austronesian Languages*. Michigan Papers on South and Southeast Asia. University of Michigan.
- Rau, Der-Hwa Victoria. 1992. *A Grammar of Atayal*. Taipei: Crane.
- Reid, Lawrence A. 1999. New linguistic evidence for the Austric hypothesis. Pp. 5-30 in Zeitoun and Li (Eds), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*. Academia Sinica, Taipei.
- Ross, Malcolm D. 1995. Reconstructing PAN Verbal Morphology: Evidence from Taiwan. Pp. 727-791 in Paul J.-k. Li et al. (Eds), *Austronesian Studies Relating to Taiwan*. Taipei: Academia Sinica.
- Ross, Malcolm D. 2002. The history and transitivity of western Austronesian voice and voice-marking. Pp. 17-62 in Wouk & Ross (Eds.).
- Schachter, Paul, and Fe T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley, University of California Pr.
- Seiter, William J. 1980. Studies in Niuean Syntax. in *Outstanding Dissertations in Linguistics*. Jorge Hankamer, Ed. New York: Garland. 1-348.
- Starosta, Stanley, Andrew K. Pawley & Lawrence A. Reid. 1981. *The evolution of focus in Austronesian*. in Pacific Linguistics C-75 (Eds. Amran Halim, Lois Carrington & S. A. Wurm).
- Starosta, Stanley. 1995. A grammatical subgrouping of Formosan Languages. Pp. 683-726 in *Austronesian Studies Relating to Taiwan* (Eds: Paul J-k Li, C-h Tsang, Y-k. Huang, D-a. Ho, and C-y. Tseng). #3 in Symposium Series of the Institute of History & Philology, Academic Sinica, Taipei.
- Starosta, Stanley. 1999. Transitivity, ergativity, and the best analysis of Atayal case marking. Pp. 371-392 in: *Selected Papers from the 8ICAL*. Elizabeth Zeitoun and Paul J.-K. Li, Eds.

- Stevenson, Christopher M., Georgia Lee, and F. J. Morin. 2001. *Pacific 2000: Proceedings of the Fifth International Conference on Easter Island and the Pacific*. Easter Island Foundation: Los Osos.
- Williams, W. L. & H. W. Williams. 1956. *First Lessons in Maori*. (First published 1862, later editions revised by W. W. Bird). Christchurch: Whitcombe and Tombs.
- Wouk, Fay and Malcolm Ross. 2000 (Eds.). *The History and Typology of Western Austronesian Voice Systems*. Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

PRAGMATIC INFLUENCE ON PRONOUNS IN LAI (HAKHA) CHIN, WITH ESPECIAL REFERENCE TO FOCUS AND CONTRAST

Ceu Hlun

University of Illinois at Urbana-Champaign

0 Introduction

Grice's (1975) maxim of quantity says that people generally follow the rule of "say no more than you must" This maxim seems to provide a possible explanation on the ellipsis of some particles or phrases in many of *pro*-drop or Free Empty Category languages. This interface between pragmatics and syntax is obviously shown in the relation between pragmatics and argument structure (on the relationship between the verb and its argument, Goldberg, *forthcoming*) in Lai, a language spoken in Myanmar. In this paper I will attempt to describe the pragmatic influence on the choice of pronouns in Lai.

First I will briefly describe the morphology of argument structure, describing the agreement system of the verb phrase that plays a vital role in the argument structure of Lai sentences. Then I will describe the distribution of pronouns in general. I will discuss the constraint on argument ellipsis in the third section, where I will make my claim that all pronouns in Lai are either focal or contrastive, or, arguments in Lai are overtly expressed solely for what is conventionally called 'focus'. Nevertheless, I will treat focus and contrast rather as different pragmatic notions having opposite polarity, as opposed to Schwarzschild (1999). In the later sections of the paper, I will focus on the contrast between focus and contrast.

Not much research has been done on this language. George Bedell (1996) described the agreement systems of Lai, which was a major breakthrough, as the pronominal agreements, with which Lai is rich, were thought to be pronouns before that time. Bedell, however, was not focussing on the other part of argument structure, and that is, the reason pronoun ellipsis was not explained. A few researchers, such as Melnik and K. Van Bik (of UCB), and F. Lehman have been working on the morpho-syntax of this language, but none of these scholars has worked on this topic so far. This work is intended to provide an insight into the argument structure of Lai as well as Burmese, which has similar phenomena of argument ellipsis (but not verbal agreements). But I will not discuss the Burmese case here.

1 Language Information

Lai (often known as Hakha Chin) is a Tibeto-Burman, spoken in Chin State, Myanmar (formerly Burma). It has been categorized as an ergative-absolutive language where the direct object takes an empty case marking element. Basically the language has a word order of SOV; but the word order is very flexible. It can be OSV, or the VP alone. Some scholars, F.K. Lehman for example characterize it as an Free Empty Category language while others might prefer to say it is a *pro*- drop language. Pitch and stress do not effect the logical semantics interpretation of an utterance.

2 Lai Agreement

Lai has a paradigm of pronominal agreement on the verb stem, which are clitics functioning to recover the ‘phi-features’ of the arguments of the verb (Bedell, 1996).

Table 1: *Pronominal agreement*

		1 st	2 nd	3 rd
Subject	sg	ka-	na-	a-
Subject	pl	kan-	nan-	an-
Object	sg	ka-	in-	-∅
Object	pl	kan-	in..V..hna	-hna

These pronominal agreements are also possessive when prefixed to nominal expressions.

The basic word order SOV is maintained in the morphology of the VP, but the syntax of the arguments is controlled by discourse structure. The VP can represent the whole sentence as long as the arguments are recoverable, and the pronominal agreements are correctly affixed to the verb stem.

Lai VPs have the basic structure: subj-agr > (obj-arg>), V (> obj-arg). For example:

(1) a ka zoh
 3sS 1sO look
 She/He/It looks at me.

(2) ka zoh hna
 1sS look 3pO
 I look at them.

(3) kan in zoh
 1pS 2O look
 We look at you.

There is no pronoun in the examples above, but the pronominal agreements are what stand for the arguments in their absence. This is the strategy to minimize the amount of utterance to conform to Grice’s maxim of quantity. Pronouns are required only when need arises to focus on an argument.

3 Lai Pronouns

Lai pronouns are mainly applicable to human or personified subjects and objects. The chance of having pronouns diminishes with the inanimacy of the entity. In reciprocals, pronouns are used for everything—human or non-human. Lai pronouns can be divided into two main categories, based on their pragmatic roles—focus pronouns and contrast pronouns.

Table 2: Focus and contrast pronouns

First Person		
	Sg	Pl
Focus	keimah	kanmah
Contrast	kei	kannah

Second Person		
	Sg	Pl
Focus	nangmah	nanmah
Contrast	nang	nannah

Third Person		
	Sg	Pl
Focus	amah	anmah
Contrast	anih	annah

4. Distribution of Pronouns in Lai

Pronouns occur overtly only as focused NPs and otherwise, in general, their features are recoverable¹. Generally, the pronouns can occupy either the subject or object position with some constraints on the type of the pronouns. The choice of the pronoun is, however, constrained by the discourse structure. Thus the pronoun can occur:

1. As agent in transitive sentences, having ergative case marked by postposition *nih*. The focus is on the agent.

(4) anmah *nih* an hal
 they erg 3pS ask
 THEY asked him/her/it.

2. As patient in transitive sentences, having absolutive case when the focus is on the patient.

(5) keimah an ka hal
 I 3pS 1sO ask
 They asked ME.

It is less common that two pronouns overtly occur in the same clause. It is sometimes claimed that languages tend to avoid more than one “new” argument per clause (Goldberg, *forthcoming*, citing Dubois 1987). However the occurrence of two pronouns is not uncommon, as the language allow two focused elements in the same clause.

(6) Nanmah nih amah hal u.
 2pS erg 3sO ask IMP(erative)
 YOU ask HIM.

3. As subject in the intransitive clause, having absolutive case when the focus is on the subject NP.

- (7) Keimah ka kal lai.
 I 1sS go will
I will go.

4. Focus pronouns occur as reciprocal NPs (contrast pronouns cannot).

- (8) Keimah le keimah ka i hal.
 I and I 1sS RECIPR ask
I ask myself.

- (9) Anmah le anmah an i al.
 They and they 3pS RECIPR debate
 They are arguing with one another/themselves.

5. Pronouns occur as possessive markers before the nouns when they are focused.

- (10) Nangmah kedan maw keimah kedan?
 YOUR shoes or MY shoes?

As against unfocused

- (10') na kedan maw ka kedan

5 Argument Omission

It has long been known that subjects are more likely to be omitted (Goldberg *forthcoming*, citing Bloom 1970; Chomsky 1982, Hyams 1986; Jaeggli and Hyams 1988; Uziel-Karl and Berman 2000), since subjects are supposed to be topical in most cases. In Lai, however, the objects are, on such grounds possibly just as topical; any of the two or arguments, if present, can occupy the topic position that is in most cases, by default, that of the subject of the sentence. Present or absent, they are essentially represented by their agreement clitics, and it is these that take fixed order, subject always being first. The claim that subjects are more likely to be omitted is therefore weakened by evidence shown by Lai. This is possibly because either/any of the arguments can be topicalized depending on the discourse structure. This is possibly a compensation for a passive construction that is very rare in the language. I will not go further on this point. Thus a simple transitive sentence (11), where the agent is topicalized, is also uttered as (12), where patient is topicalized, depending on the speaker's choice of topic.

- (11) ui nih me an dawɪ hna
 dog erg goat 3pS chase 3pO
 The dogs chase the goats.

- (12) me ui nih an dawi hna
 goat dogs erg 3pS chase 3pO
 The goats the dogs chase,
 i.e. The dogs chase the goats.

One or both of the arguments can be omitted as in (13) where the verb with the agreements represents the full clause.

- (13) [ui nih] (me) an dawi hna
 They chase them

Although it appears that the simplest instances of sentences with overt subjects are mostly intransitive, as with *wh*-questions, overt subject and object commonly occur in other simple sentences such as exclamations, depending on the information structure. Either the subject or the object, or both, in that case, is likely to be indefinite, like *mi* ‘people’ in (15). There is no argument omission in such utterances.

- (14) Zei dah a cang?
 What Q 3sS happen
 What happens?
- (15) Ngal nih mi a tai!
 boar erg person 3sS slash
 A boar slashed somebody!

6 Focus Pronouns vs. Contrast Pronouns

Schwarzschild (1999) discusses the notions of “novelty focus” and “contrastive focus,” and claims that both of them are the outcome of the combination of the Givenness Constraint and Avoid F. In his paper he treats focus and contrast as (more or less) the same category, which is contrary to the evidence given by the choice of pronouns in *Lai*.

Lai has distinct pronouns for focus and contrast, which indicates that focus and contrast are different pragmatic notions. I will show their different occurrences in different discourse situations, to demonstrate the pragmatic constraints on their occurrences.

6.1 Morphosyntactic Evidence

As opposed to focus pronoun,

1. Contrast pronoun cannot occur in object position in imperative sentences.

- (16) *Amah* va hal ko
 **Anih* va hal ko
 Just go and ask HER.

2. Contrast pronoun cannot occur as reciprocal NPs, as the object is not in contrast with other persons but is fixed to the subject. The pronoun is essentially a focus pronoun.

(17) *Keimah* le *keimah* kaa hal
 **Kei* le *kei* kaa hal
 I ask myself.

(18) *Anmah* le *anmah* an ido.
 **Annih* le *annih* an ido.
 They are fighting one another/themselves.

3. Contrast pronoun cannot occur with emphasized expressions marked by prefix *a-*.

(19) Nangmah rilmal i *a-keimah*/**a-kei* tehna ke tawlcu ka ruat kho lo.
 you great p emph-me leg wash D 1sS think can NEG
 I can't imagine that You, the Great One, would wash the feet of a layman like ME.

Where p marks a sort of truncated relative construction, so that 'rilmal i' amounts to "one who is great"; D = determiner; HUM is a humiliative marker.²

4. Contrast pronouns cannot occur before intensifiers, particles like *ko* 'particularly' and *hrimhrim* 'the very' (these are translations only). *Cu* (a generalized determiner) in (20) marks the topic, not as it marks contrast in other cases.

(20) *Amah* ko cu pei a zual cu!
 **Anih* ko cu pei a zual cu!
 HE is the extreme!

(21) Nangmah hrimhrim nih va pe hna.
 *Nang hrimhrim nih va pe hna.
 (YOU) Go yourself and give [it to] them.

6.2 Pragmatic Difference

Context: An army commander is urging the soldiers to go to the battlefield, telling them how important it is to fight this battle. In fact, the coward does not intend to go.

(22) *Amah* (caang) cu a kal lai lo.
 [caang 'but he'] (Foc)
 HE will not go, or
 Him, HE won't go.

Another commander has to stay at camp while others are going to the battlefield because he is expecting an important guest who is coming to visit the camp, otherwise the commander would go enthusiastically to the battlefield.

(23) *Anih* cu a kal lai lo. (Contr)
 **Anih* caang cu a kal lai lo.
 HE will not go (he will not go with them)

The two structures are similar morphologically, but the discourse determiner *cu* (a determiner that fundamentally ‘points’ not in real space but in the space of the existing discourse) in the two sentences has different roles. It marks the topic in (24) while it marks contrast in (25), as it does in the following examples.

(24) *Keimah* (taktak) *cu* kaa lio lai lo. (Focus)
 actually
 I myself (for one) am not going to actually swim.
 [I’ve just come to watch the kids.]

(25) *Kei* *cu* kaa lio lai lo. (Contr)
 **Kei* taktak *cu* kaa lio lai lo.
 I am not going to swim. [*You* go, if you want.]
 (Different subjects for one event)

The difference between the two expressions is obvious when negation (*lo*) is deleted.

(26) **Keimah* *cu* kaa lio lai.
 (27) *Kei* *cu* kaa lio lai.
 I’m taking a swim [You will not!]

Focus and contrast have opposite polarity—contrast being the negative—according to the choice of Lai pronoun in different contexts. That is, focus is related to the “peculiarity” of an entity to the rest of its type, whereas contrast would be related to the “exceptionality.” I will illustrate this point in an example.

Context: In an airport lounge where passengers are waiting, the flight attendant announces that they need a volunteer to delay his/her flight as they have oversold tickets. She asks, “Who would like to volunteer?” Sam raises his hand and says:

(28) *Keimah* kaa pe lai.
 I will do it. (I will commit myself and I do not know, maybe not even care, about what anyone else may do.)

At the time of boarding, another passenger, who is expecting Sam to move toward the boarding gate along with them, tells Sam to move. Sam refuses, saying:

(29) *Kei* *cu* ka kal rih lai lo.
 I am not flying yet. [But you are!]

In (28) Sam, by his peculiar behavior, is committing himself to do something. That is a positive-minded action. In (29) he is making himself an exception to other passengers. That is a negative-minded action. Here the focus pronoun *keimah* is used for a commitment, whereas the contrast pronoun *kei* is used for avoidance or excuse. In (28) Sam is not opposing himself to anyone else. In (29) he is contrasting himself to other people.³

Context: Husband and wife

- (30) Rawl chuang ning. (Non-F)
 food cook 1st-IMP
 I will let myself cook, *or* I will cook (lit. Let me cook..)
- (31) *Kei* rawl chuang ning. (Contr)
 I will cook [and YOU take care of the baby.]
- (32) *Keimah* rawl chuang ning. (Focus)
 I food cook 1st-IMP
 I will cook (because *I* cook better than all of you.)

Note that we can see with this example that, in some sense, Focus often amounts to a limited case of Contrast, namely, as against the world in general.

Example (30) is not focusing on anything. In (31) the supposed husband is contrasting himself with his wife in terms of their division of labor. He is not, however, contrasting himself with other specifiable people in (32).

As has been argued, pronoun ellipsis marks the argument and, if the omitted argument is the subject, especially if the predicate is intransitive, the clause as non-focal. Non-F clauses use pronominal agreement only. Similar examples (30)-(32) support the different readings determined by the choice of pronoun, e.g., to mark possession.

Context: At a picnic; Talking about a camera

- (33) *Ka* ta a si lo. (Non-F)
 1sS possession 3sS be NEG
 This is not mine. It does not belong to me.
- (34) *Keimah* ta a si lo. (Focus)
 my (foc) possession 3sS be NEG
 This is not MINE. It's my FRIEND's.
- (35) *Kei* ta a si lo. (Contr.)
 my (contr) possession 3sS be NEG
 This is not MINE. I don't know whose this is. [MINE is not black.]

In (33), the speaker is making himself an exception from the set of all people that might possibly own that camera. This expression is not possible if it is non-negative declarative sentence.

- (36) **Kei* ta a si
 It's MINE!

This indicates that the contrast pronoun is a negative polarity item as well. That is compatible with the fact that that pronoun *kei* is used to make a contrastive *exception* on

one person to the rest of the population in the discourse. On the other hand, the focus pronoun is used in non-negative sentences in both cases—where the pronoun is focused or contrasted.

- (37) Keimah ta a si
It's MINE.

The focus pronoun is used again in positive statements, whereas the contrastive pronoun is chosen for negative ones.

Context: Talking about going to a party when the daughter is demanding to go.

- (38) *Ka duh ah ka kal lai; **ka nawl a si.**
my wish at 1sg go fut; my choice it-is (Non-focal)
- (39) Ka duh ah ka kal lai; **keimah nawl a si.**
I will go if I want; **it's up to ME.** (Focal)
- (40) Na pa hal. **Kei nawl a si lo.**
your dad ask (Contrastive)
Ask your dad. **It's not up to ME.**

7 Summary

In this paper, I have claimed that, in *Lai*, any pronominal arguments can, indeed, must be omitted, unless they are either focal or contrastive, as long as the argument is recoverable, or indeed, in the case of a three place predicate, whether or not it is recoverable. In that case, the basic word order is maintained in the morphology of the verb complex. Thus the verb with its agreements can represent the whole sentence if no argument is in focus. Only the focused pronoun is overtly expressed as the subject or the object. I have also tried to show that contrast is different from focus, as shown by the choice of *Lai* pronoun morphologically, syntactically, and pragmatically, contrary to the implicit claim of Schwarzschild. I made a claim that *focus* and *contrast* are of opposite polarity, where contrast has the negative polarity. In this paper, I think I have been able to point out the interface between syntax and pragmatics, as it describes how the omission of the argument of the verb is sensitive to pragmatic factors such as the maxim of quantity or recoverability of the topic. On the other hand, the infelicitous choice of pronoun can change the interpretation of the utterance, and even the whole discourse eventually.

Notes

Work for this paper was supported by a grant from the University of Illinois Urbana-Champaign Campus Research Board to the *Lai* Chin Dictionary Project.

1. Actually not every such empty argument pronoun is recoverable. With a three place predicate, the indirect object is mirrored by an agreement clitic, but *not* the direct object. It is this that makes *Lai* and related languages a Free Empty Category language rather than a pro-drop

language, and for languages like Burmese, without any agreement system, none of the empty argument positions are recoverable.

2. amounts to ‘a mere I’.
3. Comparison with Burmese is helpful here. The distinction parallels that between the postnominal particle *ka* and the postnominal particle compound *ka.taw.* . In Upper Burmese at least, keeping in mind that, like Lai, Burmese is a Free Empty Category language that ‘avoids pronoun’ unless it is needed for focus or contrast, subject pronouns in particular tend quite generally to be followed by one or other of these postnominal elements. If I say

i. cunnaw-*ka.* thwa::me
 I KA go fut.

It means that I, for one, at any rate, am going to go, regardless of what others may do. If I say

ii. cunnaw *ka.taw.* thwa::me

It is understood that I mean to go *instead of* anyone else, regardless. For evidence that *ka.* is, like Lai postnominal *cu*, a determiner, see Lehman 2000; for similar evidence for Lai Chin *cu*, see Lehman 2002.

References

- Bedell, George. 1996 Agreement in Lai. Paper for the 1996 Pan Asiatic Linguistics Conference. Bangkok: Mahidol University.
- Goldberg, E. Adele. *forthcoming*. Discourse and Argument Structure in *Handbook of Pragmatics*. Larry Horn and Gregory Ward eds. Oxford: Blackwell.
- Grice, H.P. 1975 Logic and Conversation. *Syntax and Semantics* Volume 3: Speech Acts. P. Cole & J. Morgan, Eds. New York: Academic Press. pp. 41-58.
- Lehman, F. K. 2000 Burmese (A Grammatical Sketch). In *Facts about the World's Languages*, Eds. Jane Garry and Carl Rubino. New York: H. W. Wilson. pp. 105-110.
- _____. 2002 Number Marking in Lai Chin and its Theoretical Consequences. In *Papers from the Tenth Annual Meeting of the Southeast Asia Linguistic Society*. Tempe: Arizona State University Program for Southeast Asian Studies, Monograph Series. pp. 167-188.
- Schwarzschild, Rogers. 1999 Constraints on the Placement of Accent. *Natural Language Semantics* 7:141-177.

SINIFICATION OF THE ZHUANG PEOPLE, CULTURE, AND THEIR LANGUAGE

Pingwen Huang

Northern Illinois University

<paulhuang2000@hotmail.com>

0 Introduction

Most Zhuang people live in the Guangxi Zhuang Autonomous Region, the Wenshan Zhuang-Miao Autonomous Prefecture, Yunnan Province. There are also some Zhuang isolates in Guangdong, Guagnzhou, Hainan, and Hunan provinces. In Guangxi, there are 52 counties populated by the Zhuang people. The Zhuang people make up 90% of the total population in nine counties, 50% of that in 39 counties. According to “The Fourth Survey” of the national census conducted by the CCP in 1990, the total population of the Zhuang people was 15,555,802; now it is estimated to be 17 million, ranking as the largest ethnic minority in the People’s Republic of China.¹

The autononyms of the Zhuang vary from place to place. They are variously called Bouxcuengh, Bouxyaex, Bouxnoengz, Bouxdoj, Bouxmbanj, Bouxraeuz, Bouxrungh, and so on. Zhuang is the official government name for this native ethnic group that has its origins in South China. The Zhuang language is a member of Tai language family and of Sino-Tibetan stock, according most Chinese linguists. It is divided into the Northern and Southern dialects with two rivers, the Youjiang and the Yongjiang, as boundaries. Northern Zhuang speakers make up 68% of the total Zhuang population; Southern Zhuang speakers make up 32%.

Zhuang is the main means of communication in daily life, but Chinese is the only language that is used on official occasions. Among the nine counties where the Zhuang make up 90% of the total local population, the Zhuang language is spoken on most occasions, except those that are official in nature. In eighteen counties where the Zhuang make up 70-80% of the total population, Zhuang language is the farmers’ daily language. According to a random survey, Zhuang monolinguals make up 42.29% of total Zhuang population, while Zhuang-Chinese bilinguals make up 54.72%.²

The number of Zhuang-Chinese bilinguals has grown rapidly in the past 50 years of CCP (Chinese Communist Party) rule, especially since the 1980s, when China opened its doors to foreign countries. The Communist Party administrative government felt it was so hard to work in the Zhuang region in the 1950s because of the language barrier. However, nowadays it is easier because half of the total Zhuang population can speak Chinese. My parents’ village (Manlongrang) is a perfect example. It is a pure Zhuang village where there are only three family names, Huang, Ban, and Lu. Before the communist liberation in 1949 there was only one educated man, Huang Dingmen who went to high school in Nanning (the capital of Guangxi Zhuang Autonomous Region) and could read and speak Mandarin Chinese. Another man who worked as a porter in Baise city could speak Cantonese. They were the only two Zhuang-Chinese bilinguals living in the village at that time. The number of bilinguals has rapidly grown since the 1950s, as shown in Table 1.³

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society* (2002). Canberra, Pacific Linguistics, 2007, pp.89-100.

© Pingwen Huang

As the data in the table show, the total number of bilinguals by the end of the 1990s was almost three times as many as those by the end of 1980s. By the end of the 1980s it was almost double that of the end of the 1970s, and eight times as many as by the end of the 1950s. One new phenomenon that has been of growing concern since 1980s is that Mandarin Chinese gradually became the first language of some of the bilinguals. By the end of the 1990s some teenagers had already lost their mother tongue and are now monolingual Chinese speakers.

There is no doubt that the Zhuang language is an endangered language, although it will not disappear in the next one or two generations. Still, we cannot ensure it will not face and end as a human language in the next few hundreds years. In an attempt to understand the historical process of Zhuang language change and endangerment, this paper will discuss the forces that have influenced the use of the Zhuang language over time.

Table 1: *Change in Languages Spoken in Manlongran Village: 1940s-1990s*

Time (end of)	Population of Man- longrang	Languages used				Bilingual	Over three languages
		Zhuang	Guiliuhua*	Cantonese	Mandarin		
1940s	120	120	0	1	1	2	
1950s	167	167	3	1	2	5	1
1960s	234	234	5	4	6	12	3
1970s	387	387	7	13	14	21	12
1980s	500	490	9	31	29	41	34
1990s	600	570	11	46	140	130	72

(*Guiliuhua, is commonly spoken in Liuzhou prefecture and Guilin Prefecture in the Guangxi Zhuang Autonomous Region. It is a subdialect of the Chinese dialect called Guanhua.)

1 Sinification of the Zhuang Language

According to historical records, the Zhuang territories were to be loyal to their Han rulers as “one member of United China.” In 214 B.C., the Qinshi Emperor waged a strong military campaign against the Baiyue in Lingnan and occupied the area called “South of the Mountains.”⁴ From then on, the Emperor began to administer Lingnan and ruled the minorities there directly or indirectly.

The motivations that led the Chinese Imperial government to rule the minority territories was obviously not only to occupy the land but also to make profit from it. Despite these motivations, the central imperial government began to promote a series of policies for governing the ethnic minorities. The one with the greatest impact was Sinification. The goal of Sinification was to control the minority groups forever in all fields - politics, economics and culture - by importing Chinese culture and then weakening and eliminating the ethnic nationalistic consciousness that could lead to the desire to be independent states.

Sinification issues historically have involved: 1) forcing vast numbers of Han Chinese to migrate and mix with minority groups so as to allow the Han culture natural penetration; 2) encouraging Han Chinese to marry with minority peoples so as to let Chinese culture penetrate local family structures; 3) forcing minority groups to learn Han language and culture directly, especially from the formative teenage years on.

After the cessation of war on minority groups in Lingnan, the Qinshi Emperor began his first step to Sinicize the original inhabitants of Lingnan by moving some Han people from the central China and “forcing them to live with the Baiyue”. The Qinshi Emperor sent 15,000 women from central China to Guangxi. A year later (222 B.C), he sent 500,000 military agricultural colonists to the Zhuang region in eastern Guangxi and western Guagndong. These Han people were the people in the main who imported Chinese culture to the Zhuang people when they penetrated the area and lived among the Zhuang. After the Qin dynasty ended, there were many times that vast numbers of Han people moved into the Zhuang territories. According to historical records, there were four large waves of immigrants who fled to Guangxi to avoid warfare and famine in the north. The first time this happened was during the Yongjia years of the Western Tsin Dynasty (307-312 A.D.) at the beginning of the fourth century A.D. At that time in central China sixteen states were engaged in wars against each other. There were numerous Han refugees who emmigrated into the Lingnan region as a result. The Guangxi population increased quickly to ten times the size of what it was in the early Sui dynasty Daye Five year (609 A.D.) from Liu Song Daming Eight year (464 A.D.). The second time ocured in the later Tang dynasty (618-907 A.D.). Again the states in central China were engaged in wars, and a large number of Han escaped from their “homeland” and fled to Guangxi. The third time occurred at the end of the Southern Song period in the13 century. The Song and the Yuan (Mongol Dynasty) made war against each other to gain power. Again, as a result, there were a lot of Han immigrants who moved to Guangxi as refugees and made the population increase by about 440,000 homes in the Sun-Yuan year (1330 A.D.)⁵ when there were only about 280,000 homes in the North Song Yuan-Feng Three Year (1080 A.D.). The forth time happened during the end of Yuan dynasty and the beginning of the Qing dynasty. There were large numbers of Han from Hunan and Hubei provinces who moved into the Zhuang region to avoid famine.

The main reason that the Guangxi population increased so quickly was that, in addition to the natural increase in the minority population, the Han immigrants from north China infiltrated the south in such large numbers. The Han population penetration of the Zhuang area was in eastern Guangxi and western Guagndong first and then later in western Guangxi. The penetration into the Zhuang heartland in western Guangxi did not start until the Tang dynasty (618-907 A.D.). Han Chinese immigrated into the region not only from northern China but also from Guangdong. They moved along the Xijiang River toward the Yuejiang River, the Zuojiang River and the Youjiang River. Even In Xilin county which is located at the border of western Guangxi, there were Han immigrants by the time of the Ming dynasty (1368-1644.) From then on almost all of the Zhuang counties had Han Chinese immigrants. In the following dynasties, the Han population in Guagndong increased quickly. The Han population constituted less than 20 percent of the Guangxi population in the mid-sixteenth century. By the end of the nineteenth century this percentage had increased to over 50 % and, by the mid-twentieth century, was nearly two-thirds.

Besides promoting migration, the Qinshi Emperor forced all China to use “only one written script,” eliminated local written scripts, and forced minority people to learn and use Chinese. Liang Tingwang, a professor from the Central University for Nationalities, said (p.c.) that the ancient Zhuang had their own proto writing system but had to give it up because of the Qinshi Emperor’s tough policy and to adopt the Han Chinese writing

system, which ultimately developed into the old Zhuang demotic script (Figs. 1 & 2) alongside classical Chinese writing during the Tang dynasty (618-907 A.D).⁶

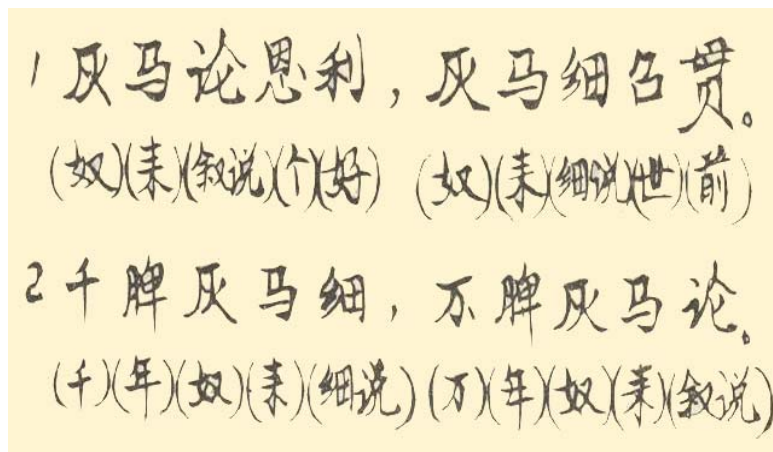


Figure 1: Old Zhuang Script

- 1 **Hoiq ma lwnh aen ndi,
Hoiq ma si ciuh gonq.**
- 2 **Cien bi hoiq ma si,
Fanh bi hoiq ma lwnh.**

Figure 2: New Zhuang Script Transcription of the Fig. 1 Text⁷

Towards the end of Qin dynasty (221-207 B.C.), a period of many peasant rebellions, Zhao Tuo claimed independence from the imperial government and created the “Nanyue Country” (Southern Yue country) in 207 B.C. in the Lingnan area and called himself “Emperor of the Southern Yue.” He led the peasants to rise up against the Qinshi Emperor. During the years when Zhao ruled Lingnan, he promoted a kind of minority policy which was known as “He Ji Baiyue” (“Unify the Baiyue”). He practiced power politics by using force on the one hand to threaten the Min-Yue, Xi’ou, and Luoyue and bribes on the other hand. He called himself “Head of the Yi-Man”⁸ He respected the Yue people’s customs, rallied their local rulers, and forced the local chiefs to be controlled by the central government administrators, but let them continue their old policies and local political traditions. In addition, he encouraged Han immigrants to marry Yue⁹ and promoted the central China Han culture among the Yue.

In the succeeding dynasties, the central Imperial rulers imitated and advanced the Qin dynasty and Han dynasty’s political policy and intensified the process of Sinification, tightening their control over the Zhuang. Though the imperial government’s administrative power in the Zhuang region began with the Qin dynasty, control over the Guagnxi area was limited. The rulers exercised direct control over the eastern portion of Guangxi, running from the eastern border of Nanning Prefecture to the western border of Guangdong. Here Han and Hakka Chinese eventually dominated and continued efforts to consolidate their administrative control over the region.

The central government maintained its normal control of western Guangxi through the *jimi* (“loose reins”) system during the Song dynasty. The *tusi* system began during the Tang dynasty and lasted until the early part of the Qing (Manchu) dynasty in Guangxi. It required local hereditary chiefs to pledge loyalty to the imperial regime but did not dictate how these rulers should govern within their independent realms.

For the more than 2,000 years that the central imperial leaders, starting with the Qinshi Emperor, who united China, ruled over Guangxi on to the end of Qing dynasty, their policy of Sinification gradually weakened the Zhuang ethnic consciousness. More and more Zhuang gave up their language and culture, chose Han Chinese culture and “became Han.” In the eastern portion of Guangxi, where the Han mixed with Zhuang, some Zhuang were forced out of their fertile land into the mountainous areas in the west. Many of the original inhabitants who still lived there but were Sinicized to the degree that they became indistinguishable from the Han. Today, there are fewer and fewer Zhuang language speakers in eastern Guangxi.

2 The Modern Period: Defects in CCP Policy and the Limitations of the New Standard Zhuang Writing System

The Communist Party began governing the Zhuang territories on November 8, 1949. On March 8, 1958 the CCP created the Guangxi Zhuang Autonomous Region, influenced by Marxist-Leninist theory of Nationalities and the Soviet model. When the CCP assumed power in Guangxi, it knew the Zhuang educational levels were lower than the national average in the province. Illiteracy rates in Zhuang areas were over 80 percent.¹⁰ The CCP realized that it was hard for them to develop Zhuang economic, cultural and political life in order for them to “catch up” to the Han and achieve its goal of “all nationalities are equal.” If the Zhuang educational level was far below the national level, it would also be hard to strengthen the consolidation of the nationalities and central control over the province. In addition to training Zhuang cadres who could effectively relate to the Zhuang and inspire the local population to comply with party directives, the party also sought to improve the general level of education throughout the Zhuang areas. Educational reform was mandatory. During the first few years of CCP rule, eradicating illiteracy and developing educational opportunities for the public were two of the primary goals of educational work. Devising and using a modern Zhuang writing system was one of the primary tools for fighting illiteracy. With the help of Soviet linguists, the central government helped the Zhuang create the new Zhuang alphabet, which was based on the Latin alphabet. In December of 1955, the National Council ratified the plans for the alphabet and promoted it in 1957. In 1981 the new alphabet was revised so that only Latin letters were used.

In August 1958 the scope of the project was expanded, and Zhuang peasants were trained for two years in night schools, while cadres received a year of instruction. The Zhuang language was promoted as “going on to reach its first ‘high tide’.” The content of the courses changed radically during the Cultural Revolution, as might be expected. Although most of the schools remained open during the early years of the Cultural Revolution, they concentrated on spreading the party’s political message and on pronunciation rather than on reading skills. Unfortunately, all the schools were shut down at the end of the “special period.”¹¹ The peasants who were trained in the schools finally lost their reading skills and became “the new illiterate.”

Zhuang language promotion rose to a second “high tide” level in the 1980s. Efforts to promote the Zhuang written script were revived in 1984. Language committees were

reestablished, along with new experimental Zhuang language schools in twenty-two counties in Guangxi. Over the next four years the number of counties that were experimenting with Zhuang doubled to fifty-two. The Department of minority languages of ten Central College for Nationalities in Beijing was set up to recruit the first class of students to learn the Zhuang language in 1980, just after the Cultural Revolution. Over the next eight years it opened one class every two years and then closed. The Guangxi Zhuang Language School (College) opened and recruited four classes of 200 students or so each year. Most of its graduates taught in primary school. However, the plan had to be modified to run as a high school program at the end of 1990s because no more students were interested in studying the Zhuang language after the local governments announced that no special rights were available for Zhuang students to work in the minority areas. Although the college still remains open, the number of students and the content of courses radically changed to concentrate on studying in Chinese. The Guangxi Nationalities Institute recruited its first class of students for its newly established Zhuang language department in 1984. However, students expressed little interest in majoring in Zhuang. It was harder for students to find a job in comparison with other majors, since the CCP broke down the old work system of job guarantees and now let students find a job by themselves. The Zhuang language department was closed down in 2002 and merged with the Chinese department after seeing that it could not recruit enough students to open classes. As for the primary schools or middle schools, by 1988 country governments were required to subsidize the Zhuang language instruction schools, if they wanted to keep them running, and as a consequence, more than half the schools closed. By 2000 there were no real Zhuang language schools in the Zhuang areas. Zhuang Language promotion went to the level of “low tide”.

The central government provided funding in Guangxi to promote the Zhuang language in the 1950s and in early 1980s. However, as the statistics show, Zhuang language could not really develop Zhuang education to a state where it was “naturally flourishing.” Educational levels are still lower in the province than for the national average. What are the reasons?

1. The government’s minority policy is defective. As we now realize, in the past most people only recognized the functional aspects of educating students in Zhuang but did not realize its cultural role and value. The common opinion was that minority languages were simply a “bridge to a Chinese education.” As some officials and cadres said, the Zhuang language is a “walking stick” or crutch for the Zhuang people to use in learning Chinese. They can throw it away when they master Chinese. That is why so many officials and cadres believed that it was enough to open Zhuang language classes during grades 1-3 only in primary schools. From grade four on, Zhuang students would begin to learn Chinese. When we look back on this history we realize that in the Guangxi Autonomous Region, language policies could not be implemented. Some cadres within the province and some of the masses did not give the cooperation and support they should have, thus affecting the implementation of autonomous rights. According to the Law of Regional Autonomy, minority nationalities are free to use, develop, and promote their language. However, in Guangxi, autonomy rights are only words not backed up by action. For certain, the Zhuang language can not be an official language. The Guangxi Zhuang Autonomous Region government did not pass a series of autonomy rights to ensure that the Zhuang people would be free to use their language anywhere in the province. According to

CCP policies on Zhuang rights, we discover that the policies were defective. That is, they did not set up a clear law to command regional and local governments to support, promote, and use the Zhuang language. Moreover, the government did not make enough workers and complete units (departments) available to promote Zhuang. Zhuang writing appears prominently on government gate signs alongside the Chinese equivalent, but it cannot be admitted alongside Chinese as an official language. In the Zhuang areas, some Zhuang speakers who can use the new Zhuang written script complain that they cannot write in Zhuang on pieces of mail because there are no postal officers who can deal with Zhuang writing. The central government and the autonomous government did not set up a complete educational system from primary to higher education and autonomous rights for the Zhuang to learn to use their own language and culture. As the peasants say, “Since the Zhuang language cannot be used equally alongside Chinese in standard examinations, why should we let our children waste their time learning the Zhuang language?” Education remains the primary means of social and economic upward mobility in Zhuang areas and of escaping from the poverty of the countryside. Zhuang students and their families, therefore, have an intense interest in excelling in the Han-language examinations. Through the standard state examinations, any student can theoretically rise to the top ranks of the educational system. The educational system, therefore, has proved an important integrating force now that Zhuang peasants have access to schools and hope to benefit from education. The standard examination system has made the average Zhuang less willing to learn the Zhuang language and more interested in joining in the privileges that state education has to offer.¹²

2. The Zhuang were not interested in promoting the new Zhuang written script. Nor were they interested in creating the unified Zhuang ethnic nationality. The pressure to create Zhuang Autonomous Region clearly did not come from the grassroots level. i.e. by the Zhuang themselves. Rather, the proposal was actively promoted by the central government and the Zhuang middle-level cadres. Consequently, the Zhuang students and their families have little interest in learning the new Zhuang written script because they “believe” they can not get any benefit from it.

3. The numerous dialects within the Zhuang speech region also complicates the promotion of a Zhuang written language. There are two dialects in the Zhuang areas: Northern Zhuang and Southern Zhuang.¹³ Furthermore, there are seven subdialects in Northern Zhuang: Guibei, Liujiang, Hongshuihe, Yongbei, Youjiang, Guibian, Qiubei, Lianshan. There are five subdialects in Southern Zhuang: Yongnan, Zuojiang, Dejing, Yanguang, Wenma. The population of Northern Zhuang speakers makes up 68% of the total Zhuang population, while that of Southern Zhuang speakers make up 32%. Phonologically, one of the most distinctive difference between the two dialects is that there is a set of aspirated initials in the Southern dialect that are absent in the Northern dialect. Speakers from the two dialect areas, sometimes even between two different subdialects, cannot interact effectively. They have to communicate in Mandarin, local Southwestern Chinese or Cantonese. The so-called “Standard Zhuang” is based on the Wuming county dialect which is only 40 kilometers away from Nanning, the capital of Guangxi Zhuang Autonomous Region. Wuming speech is neither Northern nor Southern Zhuang and is spoken nowhere outside Wuming county. The difference between Wuming speech and the many different dialects cannot be systematically correlated. For these reasons, students and their families feel it is too hard to learn. My former students in Guangxi Zhuang Language

College in Wuming complained that Standard Zhuang is a “new” language far different from their dialects and hard and unnatural to learn. Some students said they tried to speak standard Zhuang to their families but nobody understood what they said. These Zhuang speakers were trained in school to teach the Standard Zhuang language in primary schools in their areas after they graduated and went to work as teachers. However, some of them could not master it even after four years of study. Some of them even complained that the “Standard Zhuang” is harder than Mandarin Chinese. How can the government and university educators make people interested in learning a language which is far removed from the language of their daily life?

4. Han Cultural and Educational Pressures on the Zhuang. As mentioned before, the imperial government actively promoted a nationality policy of Sinification in Lingnan ever since the beginning of the Qin dynasty. This policy established the Han culture and educational system in minority areas and forced minority people to learn Han culture once the Han had penetrated the territories “South of The Clouds.”

In reality, Zhuang education is Han education. Formal schooling of the Zhuang began in the Qin dynasty. The Qinshi Emperor transplanted the central Han feudal policy in the Zhuang territories after his military forces occupied Lingnan. From then on the central imperial government set up its bureaucratic administration to rule the Zhuang and began to set up some schools there to popularize Confucianism. As recorded in the historical annals of Ling Hu Xi Zhuan in the chapter called *Sui Shu* (The Sui Dynasty Book): “Build cities and towns, open schools! The minorities will thank the Han and, in their response, they will called it “a great corrective training.” Since the beginning of the Tang and Song dynasties, Han education had a strong impact on the Zhuang through the Han feudal county system which fostered and enhanced it. More and more historical records about education in the Zhuang-speaking region appeared. Schools were opened by the *Fu* (prefecture), the *Zhou* (an administrative division which smaller than a prefecture but bigger than acounty), and the *Xian* (county), and were called *fixue* (Fu School), “Zhouxue” (Zhou School), *xianxue* (Xian School), respectively. The earliest *fixue* in Guangxi was the Liuzhou Fuxue, which was built during the early Tang dynasty, followed by the Guilin Fuxue, which was built during the middle of the reign of the Dali Emperor (766-779 A.D.). During the Song dynasty, many *fixue* were developed. In Guangxi alone, six new *fixue* were opened: Qingyuan Fuxue, Pingle Fuxue, Wuzhou Fuxue, Nanning Fuxue, Xunzhou Fuxue and Lianzhou Fuxue, were opened. There were more new *fixue* established during Ming dynasty and Qing dynasty. Furthermore, there were some other small schools built in the *Fu*, *Zhou* and *Xian*, such as, Confucian Schools, Shuyuan (general education schools), and Minjian Shexue (local county schools). Although these schools were patterned after the old style of education, they were effectively the same as the central Han educational system. Their goal was training officials for administrative work. The contents of the text books all were concerned with Confucian thought and values. Chinese was the only language used in class.

The new style Zhuang education began in the twentieth century with the establishment of institutes for higher education for local political chiefs, agriculture, advanced normal (teacher) training, army survey personnel, cadre training, and included other universities and colleges. From the 1930s to the end of 1940s, high schools, middle schools and primary schools were set up. By 1949, in Guangxi, there were over 17,000 primary schools, which had 1.4 million students enrolled and 17 high schools, which had a

total of 25,200 students enrolled. According to a 1990 survey, there were 460,000 high school students, who made up 32% of total students in Guangxi. The Zhuang had 2.45 million primary students; 660,000 middle or high school students; 190,000 vocational school students and 18,000 university or colleges students.

Since Zhuang education is Han education, all these Zhuang students were trained in Han language and culture. They are the main source for transplanting Han language and culture in Zhuang areas. They are trained to accept Han culture and thereby weaken their ethnic roots, slowly losing Zhuang ethnic consciousness and rapidly changing conditions in the direction of accepting Han culture. They are bilingual. However, most of the new generation will not be Zhuang speakers. In addition, the opinions and values of these new intellectuals have a powerful influence on the Zhuang around them.

3 The Cultural Prestige of Han Influences Over the Zhuang

Every “nationality”¹⁴ has its own culture. Although differing cultures cannot be labeled as good or bad, the cultures of China’s nationalities are viewed by the government as “highly developed” and “less developed.” In a system of “United Ethnic Nationalities” the culture of a “nationality” which has a “highly developed social production” enjoys a higher prestige when compared to the other “nationalities.” It seems to be human nature to choose to adopt a culture that enjoys higher prestige. Han “social productions”¹⁵ were usually considered to be higher than those of other nationalities in China. The Han culture was promoted as “highly developed,” while other nationalities’ cultures were “less developed.” Consequently, Han culture became the main standard in China which guided and influenced other nationalities’ cultures. The Han culture penetrated and influenced minority nationality cultures because of its larger population and their dominant culture.

Han culture began its penetration of the Zhuang region when the Han began to administer Lingnan in the Qin dynasty. The Zhuang culture was in an “inferior” position for the many centuries it was in contact with Han culture. The Han influenced Zhuang language and culture deeply. More and more Zhuang envied the greater prestige and power of the Han. This fundamental psychological attitude led to the acceptance of Han culture by the Zhuang. The Zhuang were intentionally Sinicized by the Imperial government for hundreds of years. Because the Zhuang culture was put in an inferior position when it came in contact with the Han culture, the Zhuang people’s cultural psychology changed radically. They had a general sense of inferiority about their culture, looking down on their traditions, believing they were too backward or too much of a “village culture” and not urban and modern enough. This sense of inferiority forced them to hide their real ethnic identity to avoid the Han racial discrimination. Meanwhile, they tried to meet the need for psychological “equality” by believing they were “Han” or “Han who spoke Zhuang.” Moreover, when the CCP wanted to help them to create a “United Zhuang” in the 1950s, they not like the idea of being singled out on the one hand and forced into a category they did not feel was part of their cultural heritage and background on the other. Even today many peasants do not know what it is to be a Zhuang, and they do not know they are Zhuang. Some do know they are ethnic Zhuang, but they are ashamed to tell an ethnic Han that they are Zhuang because it puts them in an asymmetrical power position by doing so.

Cultural prestige controls language prestige. Generally speaking, Chinese has high prestige and it has become the common language in China. It is now a “world language.” Zhuang envy the Han and want to learn the language of the Han, i.e., Mandarin Chinese. Zhuang peasants believe their language is a “backward language” in comparison to the

Han language. They hope that they and their children will become real “Han” by speaking Chinese to others. Zhuang students have a growing incentive to learn the Han language and not “waste their time” in studying their “backward language”. Chinese is the second language in Zhuang daily life, and it now is even the first language for the new generation in some areas. According to a random survey, Zhuang monolinguals make up 42.29% of the total Zhuang population, while Zhuang-Chinese bilinguals make up 54.72%.

4 Economic Pressures Compel the Zhuang to Learn Chinese

Another major reason that compels the Zhuang people to learn Chinese is the pressure of earning a living. China opened its door to foreign countries in the 1980s. The new “free market system” has been pushing Chinese economic development. However, because of a lack of technology and shortage of transportation infrastructure, the Zhuang areas are economically backward areas. There is a big gap between the Zhuang Autonomous Region and the eastern Chinese provinces. The Zhuang standard of living is far lower than it is for the Chinese in the eastern parts of China, even lower than the Chinese in the Guangxi province. According to a 1990 survey, there were almost two million peasants living under the poverty line in Guangxi. Most of them are Zhuang. It is far different for the Han who fled into the Zhuang areas to avoid famine hundreds of years ago. Hundreds of thousands of Zhuang peasants have been forced to leave their home villages to avoid poverty and flood into Chinese cities to look for jobs. Because of the new economic situation they find themselves in, the Zhuang people have to learn and use Chinese.

Another issue within the recent political and economic reforms is changing the old way that the government allotted students jobs after graduation and offered them guidance, if not guarantees. Now students must look for jobs by themselves. Zhuang students don’t want “waste their time” in learning to read and write Zhuang because university students are not able to take their examinations in the Zhuang language. Furthermore, students worry that they might be relegated to a life of employment in Zhuang areas and not be selected for work in the more economically booming Han areas both in and outside Guangxi. At the present, socio-economic reality means that the Zhuang people must master Chinese, if they want to look for jobs in the cities. If Zhuang children want to go on to higher education, they must master Chinese too. In fact, the measurement of their intellectual level depends largely on the level of their Chinese.

The Han language is the only language for the Zhuang to interact with the Han and even with Zhuang in different dialect areas. As it is said, you don’t need to worry wherever you might be in China if you have mastered the Han, i.e., Mandarin language, just as you don’t need to worry wherever you are in the world, if you can speak and read the a-b-c’s of English. Han/Mandarin language is the main language target for the Zhuang. They want and need to learn to learn it for survival.

5 Conclusion

As was discussed before, the Zhuang culture has been influenced by Han culture for over 2,000 years of contact with the Han who entered the Zhuang region of Lingnan. The Zhuang people have had to give up their language and learn the language of the Han imperial rulers not only because of external causes but also because of internal reasons. The imperial government promoted “Sinification.” This is the first and foremost reason. Almost everything

relates to this reason that the Zhuang have adopted Han culture and language and have assimilated to the Han and can be linked to the policy of Sinification.

Minority languages and writing systems in China or in East and Southeast Asia face the same challenges as lesser languages everywhere. They are clearly endangered species. In the age of globalization, there are powerful economic incentives to cast aside local languages. It is clearly possible for an endangered language to vanish in 1-2 generations, swallowed up in the language and culture of a dominant ethnolinguistic majority, such as Chinese or English. In any society in any part of the world, minority languages and scripts can only be kept alive if there is strong leadership from minority leaders and support and encouragement from central governments to keep them from dying - and with them the treasure of indigenous knowledge that they preserve.

Notes

1. Data from the Fourth survey of Chinese Population. 1993.
2. Unpublished survey. Chinese Nationalities Affairs Committee.
3. Huang, Pingwen, 2001. "Different Cultural Contacts and Language Change: A Study of Zhuang Language Change." Unpublished Ph.D. dissertation. Central University for Nationalities, Beijing, PCR.
4. Today, this region is roughly equivalent to Guangdong and Guangxi provinces.
5. The Chinese custom is to refer to dates of historical importance in terms of who the ruler was at the time or some other naming custom. That is to say, unlike in the West, events were recorded not with dates but with names of persons, periods, events or dynasties, e.g., "Warring States" (481-221 B.C.); "Spring and Autumn" (722-481 B.C.). In this case, the year 1330 A.D. was the year when Emperor Sun of the Yuan dynasty was in power. Also, when a census was conducted, only houses were counted, not individuals who lived in them.
6. The history of the system of demotic Zhuang character writing is similar to the development of the Vietnamese *chu nom*.
7. 1. I come to talk about good things. I come to talk about ancient dynasties. 2. Of one thousand years ago, I come to talk in detail. Of ten thousand years ago, I come to talk.
8. "Yi" referred to all of the minority groups in eastern China; "Man" – often translated as "Barbarians" – referred to the minority groups in southern China.
9. These are the indigenous people of southern China, i.e., non-Han. They were the ancestors of the Zhuang (Tai) and most likely included peoples of Austroasiatic stock as well as Tai.
10. Fan Qixu and Qin Naichang. *Zhuang Encyclopedia*. Nanning: Guangxi Ethnic Minority Press. p.233 [in Chinese]
11. This is a Chinese euphemism for the Cultural Revolution and is also called "The Dark Period."
12. Kaup, Katherine P. 2000. *Creating the Zhuang - Ethnic Politics in China*, Boulder, CO: Lynne Rienner Publishers. p.138.
13. Furthermore, in the Tai language family tree, which traditionally has three main branches, Northern Zhuang is listed as a Northern Tai language, while Southern Zhuang is listed as a Central Tai language.
14. Up until quite recently, Chinese publications in English on minority groups in China refer to them as "nationalities." Some newer publications are beginning to use the term "ethnic group" in place of "nationality."
15. By "social productions" is meant achievements of all kinds: in the arts, sciences, and technology.

References

- Fan Qixu and Qin Naichang. 1993. *Zhuang Encyclopedia*. Nanning: Guangxi People's Press. [in Chinese].
- Fourth Survey of Chinese Population*. 1993. Beijing: Chinese Accounting Press.
- Kaup, Katherine P. 2000. *Creating the Zhuang: Ethnic Politics in China*, Boulder, CO: Lynne Rienner Publishers.
- Huang, Pingwen. 2001. "Different Cultural Contacts and Language Change: A Study of Zhuang Language Change." Unpublished Ph.D. dissertation. Central University for Nationalities, Beijing, PCR.

ORIENTATION ORIGINS: THE SOURCE OF JRU' CARDINALS

Pascale Jacq

Centre for Research on Language Change (ANU)

pascalejacq@yahoo.com

Paul Sidwell

Centre for Research in Computational Linguistics (Bangkok)

<paulsidwell@yahoo.com>

0 Introduction

In the course of fieldwork on the Jru' (Laven)¹ Mon-Khmer language of southern Lao P.D.R., we encountered a puzzle when we first elicited terms for the cardinal directions. While we found that Jru' people consistently use the same set of 4 terms, some older informants, and the oldest written sources, gave conflicting sets of directions for these terms.

Investigating this puzzle we developed the hypothesis that the cardinals derived from a set of directionals that originally reflected traditional Jru' village design, which is no longer practiced. It turns out that this idea does offer a good basis to etymologise the Jru' cardinals, although we are still unsure of the explanation for the confusion in some of the data.

1 The data/issues

Early in our Jru' fieldwork we recorded the Cardinal Directions from 2 young informants: in 1998 from a man named Eh, and in 1999 from a woman named Toi. They are cousins, both aged in their mid twenties, and raised in Paksong district (Eh in Ban Thongset, and Toi in Ban Phu Khao Thong). Both learnt Jru' as a second language from their Aunts/Uncles (their parents only spoke Lao to them). They presented the set of cardinals as diagrammed in figure 1. Interestingly Eh freely alternated between initial *p*- and *k*- for 'west' and 'east' (and once pronounced 'west' as *klɨŋ*, which was perhaps an error). We compared these data with the written sources we held, laid out here in Table 1.

Two things are evident, 1) that the oldest source, Prachakij-karacak, which actually records data from the 1890s, records the variants with initial *p*- that we found in the contemporary speech of Eh, and 2) there is apparently some confusion concerning which directions are indicated by each term.

The first point appears to be rather marginal, as such alternations in prefixal consonants are not uncommon in Bahnaric languages, and so far we have assumed that the forms with initial *k*- are basic. However, as this alternation indicates that we are dealing

¹ Jru' /ʃruʔ/ (also known in the literature as Jruq, Loven, Lawen, Boloven) is a West Bahnaric language spoken in southern Lao PDR, predominantly on the Boloven Plateau. (Lao national census gives a figure of 40,000 Jru' 'Lawen' people, although many Jru' in larger towns under the age of twenty do not speak the language competently or only as a second language).

with prefixed forms, it suggests that we can segment and potentially etymologise the morphemes.

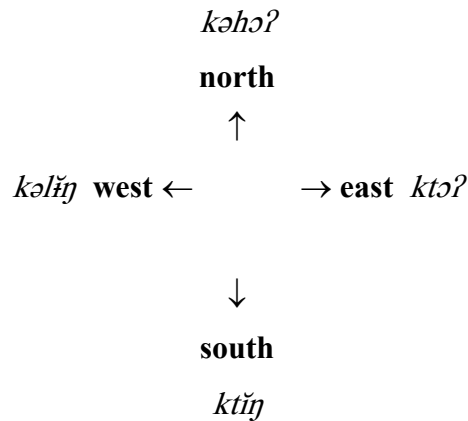


Figure 1: Set of Jru' cardinals

Table 1: Jru' directionals from various sources

	North	South	East	West
Eh & Toi	<i>kəhɔʔ</i>	<i>ktɿŋ</i>	<i>ktɔʔ</i>	<i>kəlɿŋ</i>
Prachakij-karacak (1995 [1919]):	<i>pɿŋ</i>	<i>pɿtɔʔ</i>		
Bondet de la Bernadie (1945): (with our phonemisation):	<i>k'ho</i> /kəhɔʔ/	<i>k'tin</i> /ktin/	<i>k'to</i> /ktɔʔ/	<i>k'lu'ng</i> /klɿŋ/
Huffman (1971):	<i>khɔʔ</i>			<i>klɿŋ</i>
Ferlus (1969-70):	<i>ktɔʔ</i>	<i>klɿŋ</i>		

[We had also recorded *ktɿŋ* 'north' from Eh's father in 1998.]

In the process of writing her grammar of Jru', Jacq set out to investigate the etymology and explain the confusion in the use of these four words. In the first place simple comparison with cardinals recorded for other West Bahnaric languages was not useful, as they generally show a pattern of using compounds meaning 'unstream' and 'downstream' for North and South, reflecting the flow of the dominant rivers on the plains, and compounds on the model of 'sunrise' and 'sunset' for East and West. Some have a term for South that appears to derive from prepositions meaning 'here' or 'middle', but none of these corresponded to Jru' terms under investigation here.

On the basis of form Jacq identified that *ktɔʔ* and *kəhɔʔ* may be analysed as derivatives from prepositions: *tɔʔ* 'to', and *hɔʔ* 'in' with prefixes attached (note that Prachakij-karacak recorded *prhɔʔ* 'in', rather than *hɔʔ* which is consistently recorded in all later sources).

The analysis of *klɿŋ* and *ktɿŋ* took a little more digging. In 1998 Jacq recorded Eh's father using *ɿŋ* to mean 'behind, back', while younger speakers such as Eh and Toi use *pɿt* for this meaning. We may assume that *ɿŋ* is a preposition that has fallen out of use, but survives in the *klɿŋ* formation. At first we were not able to identify a source for *ktɿŋ* within the Jru' lexicon, so we searched further afield within West Bahnaric. Likely cognates were found among wordlists we collected in our 1998 and 1999 fieldtrips: Sapuar *tiŋ* 'following, next (year)'; Oi *tiŋ* 'over, in sky'; and we found two forms in Michel Ferlus' Nhaheun fieldnotes: *pak tiŋ* 'north, upstream', *ŋwɔɔ tiŋ* 'next year' (Ferlus 1998) allowing us to reconstruct Proto West Bahnaric **tiŋ* 'following, next' (Jacq & Sidwell 2000)². In retrospect it also appears that the Jru' form recorded by Prachakij-karachak *prtɿn* 'out' also may reflect this root (assuming that the final *-n* is an attempt to represent a velar nasal).

So on purely formal grounds we appear to have sources for these cardinal terms, but semantic problems remain, as it is not immediately clear how the known meanings of these prepositions come to reflect the compass points. Our suggestion is that originally the four terms were indeed directions but instead of referring to the poles/sun position, they described points of reference within the village layout. Investigation into old anthropological sources and ethnographic surveys made in the late 1800s to the early part of the 20th century revealed that the Jru' did seem to have a rather strict traditional village design, where houses were built in straight lines, e.g.:

After a marvelous walk in the forest, I arrive at a second Boloven village, called Ban Dong Kheuong consisting of some fifteen houses, abandoned like the others, placed on a single line and built on an identical model. (Harmand 1997:92)

Harmand describes the Jru' (Boloven) houses, noting that "they have a verandah on each end. A platform lengthens the floor on the side of the principle entrance" (p.89) unfortunately Harmand does not note alignment with geography or sun positions. The clearest evidence of this is found in Fraisse's (1951) detailed ethnographic study of the indigenous and Lao populations living throughout the Boloven Plateau. In reference to the three Jru' villages north of Paksong in the Nong Lê complex (at that time already established for 40 years), Fraisse (p.60) notes that "L'entrée des maisons est toujours dirigée vers l'Est: c'est un «ordre des Génies»." In comparison, in a Jru' village called Ban Nhik (now called 'Kilometer 30 Village') to the west of Paksong, Fraisse (p.64) remarks that "L'orientation des portes des maisons se fait vers le Nord ou le Sud."

Prachakij-karachak (1995:85) reported that at the time of his investigation in the 1890s the Boriwen (Jru') had a similar house design to the Yaheun (Nhaheun). Wall (1975) described the Se Nam Noi village layout and cardinal direction of the village. The Se Nam Noi river is crucial to the Nhaheun village layout:

Géographiquement le village de Se-Nam-Noi, ou Daak Mih comme l'appellent les Nya Hön, est situé à 60 km à vol d'oiseau de la frontière vietnamienne, au bord de la rivière principale du plateau, rivière dont il porte le nom. Celle-ci a une grande importance dans la "géographie nya hön". (Wall 1975:44)

² Compare also Proto South Bahnaric **tiŋ* 'follow, chase' (Sidwell 1999), and Old Khmer *tiŋ* 'follow, chase' (Pou 1992)

(located by a large pond that feeds the Sepian River) and, from several villages north of Paksong: Ban Thônset, Ban Katuot and Ban Sekhôt (all aligned along the road running north) and from Paksong itself. All but one informant gave forms and directions perfectly consistent with those we recorded from Eh and Toi 4 years before. One 57 year old man from Ban Thôn Wa gave directions which were consistent with those recorded by Prachakij-karacak, although his fellow villages disagreed with him, and gave the 'standard' directions. In subsequent discussions Sidwell was assured by villages that this fact had no special significance, it was simply that older people sometimes get confused!

Conclusion

Our conclusions at this stage are mixed: earlier investigators may have incorrectly recorded the meanings of the cardinals they elicited, or they may well have accurately recorded usages that are/were different from what we have recorded in 7 contemporary villages, for reasons that are not clear. The latter remains an intriguing problem, keeping in mind that such different usage of the cardinals would have caused miscommunication between speakers. On the other hand, we are reasonably satisfied that we have etymologised the forms. It is a reasonable, although unproven, hypothesis that they originate within the scheme of traditional Jru' village layout.

References:

- Bondet de la Bernadie, J. 1949. Le Dialecte des Kha Boloven. *Bulletin de la Société des Études Indochinoises*. 24/3:57-78.
- Ferlus, Michel. 1998. *Nhaheun-French-English Lexicon*. (1969-70 fieldnotes edited and annotated by Pascale Jacq & Paul Sidwell). München, Lincom Europa.
- Ferlus, Michel. 1969-70. ms. Lexique Laven. Collectes réalisées en 1969-1970.
- Fraisse, André. 1951. Les villages du plateau des Boloven. *Bulletin de la Société des Études Indochinoises* (nouvelle séries), 26/1:53-72.
- Harmand, F. J. 1997. (translated by Walter E. J. Tips). *Laos and the Hilltribes of Indochina: Journeys to the Boloven Plateau, from Bassac to Hué through Laos, and to the Origins of the Thai*. White Lotus Historical Works in Translation, Bangkok: White Lotus Press (original publication: Harmand 1878-79).
- Huffman, Franklin E. 1971. *Vocabulary Lists (Mon-Khmer) 20 languages*. Copy of manuscript held at Summer Institute of Linguistics Library Bangkok.
- Jacq, Pascale & Paul Sidwell. 2000a. *Loven (Jruq) Consolidated Lexicon*. Languages of the World Dictionaries. München: Lincom Europa.
- 2000b. *A Comparative West Bahnaric Dictionary*. Languages of the World Dictionaries 21. München: Lincom Europa.
- 1999. *Sapuan (səpuar)*. München, Lincom Europa.
- Jacq, Pascale. 1997-2001. ms. Various fieldnotes on Jru', Nhaheun, Oi, Sapuan, Alak, Talieng, Laveh languages of southern Lao P.D.R.
- Pou, Saveros. 1992. *Dictionnaire vieux khmer-français-anglais: An Old Khmer-French-English Dictionary*. Paris, Cedoreck.
- Prachakij-karacak, Praya. 1995. *Some Languages of Siam*. (Translated and annotated version of original 1919 publication by David Thomas & Sophana Srichampa (eds.)). Bangkok, Mahidol University.

- Sidwell, Paul. 1999. *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra, Pacific Linguistics 501.
- Wall, Barbara. 1975. *Les Nya Hön: étude ethnographique d'une population du plateau des Bolovens (sud-Laos)*. Paris, Mémoire.

ON THE SO-CALLED NOMINALIZER NAAK IN LAI (HAKHA) CHIN, WITH REMARKS UPON ITS OTHER FUNCTIONS IN CHIN LANGUAGES AND ETYMOLOGY

F. K. Lehman (F. K. L. Chit Hlaing) & Ceu Hlun
University of Illinois at Urbana-Champaign
<f-lehman@uiuc.edu>

1 Introduction, the Standard Treatment of *naak* in Lai

In the grammatical literature on Lai Chin up till now, the suffix *naak* has been considered as a nominalizing suffix on verbs. For example, in much Christian literature, such as Bible translation, where the function of *abstract* nominalization seems to be a relatively recent extension of its more basic and traditional use. Thus,

1. $\text{ʈhut } naak \text{ ka-duh lo}$
 sit 1sg want neg
 I don't want to sit/sitting.

In fact, however, when *naak* is suffixed, as in $\text{ʈhut-}naak$, the sense is more ordinarily a locative or instrumental or other 'argumentative' nominalization. The abstract nominalization function appears then to be an extension of this other nominalization function, and we need to consider the question how it came about. Note that the Stem of a verb to which *naak* is suffixed is already Stem II. For example, $dawnaak$, 'love' (cf. *daw* 'to love'). Stem II, (see Lehman 1996 and *to appear*) we have shown to be the perfective stem, i.e., already, in itself, 'nominal', more precisely gerundive~infinitival.

2 Daai and Cho (Southern) Chin Cognates of Lai *naak*

How did this extension of function arise? To see the answer, we need to start by examining the results of an interesting paper by Helga Hartmann (2000) on its cognate in the Southern Chin language, Daai (and in the closely related Cho, on which one of us, Lehman, has also worked). In Daai the cognate is itself a verb, with the usual Chin system of stem alternation (*na~naak*). It can be suffixed to nouns or verbs with what turns out to be an applicative sense, for example, with the noun 'wife' *kkhyu*.

2. kah-ning kkhyu-*naak* vai ni
 1sg-2sg wife-take must !
 I must take you to wife.

Or, consider the adjectival verb 'pleasant', *phyah/phyah*. It can be 'transitivized' by prefixing the root with *k-*, which is then obligatorily suffixed by *naak*, giving the sense of 'love'.

3. kah-kphya-*naak* ni
 1sg pleasant- !
 I love her indeed!

We do not propose to go in to all the material on Daai from Hartmann's paper, but notice that this same function also gives nouns from verbs, as when the verb *ngthei*, 'to learn' is suffixed, giving *ngtheinaak*, 'lesson'. Moreover, it is also the case here that the nominalization often, as in Lai, has a more particular sense rather than purely abstract nominalization, as when the noun *anghmiipnaak*, means 'a place of darkness' (from *anghmiip*, 'dark'). Note that if, for instance, such an expression is in a participial subordinate clause, *naak* takes the stem alternation form *na*, as in

4. nu noh pyen *na* lü ah-ngsuh püi
 mother Erg scold part 3sg quarrel part.
 Mother having scolded [her], quarreled with her.
 Mother scolded her and quarreled with her.

And this is the evidence for the claim that in this language *naak* is indeed a full verb.

Now, for various reasons, it seems that the Daai/Cho *naak* has to be treated as an 'applicative' verb, i.e., a verb which takes as its argument a verb or noun and produces a derived sense that 'applies' the argument to the base. So (2) amounts to 'to take someone to wife'; (3) amounts to 'to find her pleasing', and *anghmiipnaak* amounts to 'applying darkness to some place or other. Moreover, the essentially applicative nature of *naak* is clearly seen in its use with nouns such as 'heart', *mlung*, where *mlungnaak* means precisely 'take to heart'.

3 Lai and Central Chin *nak/na* as a Verb Suffix and as a Complementizer

Now, from this vantage point of comparison we can look again at the Lai use of *naak*.

Central~Northern Chin: takes either stem of Proto-Chin *naa(k)* into a stem-invariant transitive applicative; Mizo takes I, Lai takes II. As we know, for any transitive verb in these languages the 'primary' stem is ambiguous. On the one hand, the etymological stem is II, and arguments from clauses relativizing an object show that semantically the underlying stem is I, which serves as a stative adjectival verb. But it is II that is on the surface the primary stem, the stem used for ordinary, default declarative finite clauses.

3.1 In Lai

(a) *naak* is easily seen to be an applicative element serving to mark a dependency between a clause and a pro (implicit?) nominal argument or adjunct thereof, and thus essentially serves as a Complementizer, C of CP (cf. *de* of Chinese and *i*. in Burmese relative constructions). C in languages of all sorts often use other parts of speech—prepositions like *for*, Chinese *de* ('of', 's) or the *i*. of Burmese; *that* in English (demonstrative), and so on.¹ Of course, it is no great leap from applicative verb, which is an application of a verb to an adjunct or argument DP of its clause, i.e., making the clause a dependent of the head DP, to a more general, grammaticalized affix of dependency.

(b) In Lai *naak* serves also as a valence-changing applicative verbal suffix proper, and unlike the Daai/Cho (co)verb, it cannot attach to nouns, only to verbs. This no doubt

was its first grammaticalization in the development in Lai and other Central Chin languages from the situation still represented in Southern Chin. Thus,

5. mah mawṭaw kha Chicago-ah ka kalnaak cang
 that car that to 1sg.go perf
 I have taken that car to go to Chicago.
6. mah fung khan ar ka-thah
 that stick that-obl chicken 1sg.kill
 I killed the chicken with that stick.
- 6'. mah fung kha ar ka-thahnaak
 I 'took and killed' the chicken with the stick.
 (I took the stick to kill the chicken.)

Note here that in (6) *fung* (stick) is an adjunct, as seen by the oblique inflection of the demonstrative; in (6') it is an argument, which is what we mean by saying that the co-verb *naak* changes the valence of even the transitive verb *thah* (to kill). Note also that *naak* is invariant as to the verb stem it attaches to, regardless of valence-changing, even though, in general, changing the argument structure of a verb (especially a transitive verb) results in a change from Stem I to Stem II.

7. mah fung kha ar ka thahnaak tik-ah
 when I take and kill a chicken ...

Note also that the place where *naak* in Lai has previously been treated as just a 'nominalizer', it is indeed a variety of dependency marker/complimentizer, and that in such cases in fact the affixation even to an intransitive verb does not actually change its valence or argument structure. Thus, consider

8. a kal-naak
 1sg go –

Depending upon discourse context this can mean, for instance,

- 8'. the 'time' of his going

(essentially application of the act of going to a particular time/event—a pseudo event-nominalization, perhaps; or, more usually

- 8". his goal-object of going

We generalize this as follows, keeping in mind that this language is a Free Empty Category language in the sense of Yan Huang (2000), namely, that the default, non-focus/non-contrastive form of a pronoun (see now Ceu Hlun, this volume) is zero-form or *pro*, and that 'one/it' is almost never focused, even in English.

9. [a- [[kal]_{IP} naak]_{CP} pro]_{DP}

where *pro* is to be understood or imagined with the verb,, in this case with *kal*, ‘to go’. Hence it means, in generalized form, ‘his “one” of going’.

So much for *naak* being a ‘nominalizing particle’, and we easily see that the first element, clitic *a-* (third person singular agreement form) is not the subject agreement here at all, in as much as it need not be used here at all. For, *kal-naak* all alone can mean ‘going’ in the sense of anyone’s ‘time /goal/event of going’ (see now Bedell, this volume for the proper syntax of such clitic prefixes, which serve sometimes to mark argument agreement on finite verbs, sometimes as possessives (as in this instance).

4 Conclusions and Further Considerations

A number of remarks, observations, and questions remain. First of all it is clear that the road to the Lai use of *naak* has involved successive grammaticalization from a stage represented by the Daai/Cho *na/naak*. The earlier stage has to have been the conversion of this originally independent verb into a ‘particle’, an invariant form verbal applicative suffix that is essentially an argument valence changing particle. The second stage has been more recent and seems to have been its further conversion into a genitive-dependency particle (a ‘relativizer’) in Complementizer position, Lai showing both results today. In the latter function, note that there is no overt adjunct to refer to, and in particular this is so for the more recent use as a purely abstract nominalizer, applying to only the *event* referred to by the verb, whilst in the former function it applies the verb to a specifiable argument of that verb.

Demonstrating this direction of development is easy. It is the fact that, of the Daai/Chin verb’s two Stem forms, Central Chin languages have chosen, perhaps arbitrarily, one or the other for its invariant particle function: *naak* in Lai, *na* in Mizo and Laizo.

Note now that in its two functions in Lai, *naak* takes two distinct syntactic positions, and this is shown by the fact that the verb, in the first function, where *naak* seems directly suffixed to it, takes Stem I invariantly, appropriately to its being used the verb of a declarative clause, whereas, when *naak* is a complementizer, the verb is a gerund and hence, appropriately, is in Stem II.

Then, it has to be pointed out that the category of ‘relativizer’ is of course not a true category of syntax. It is a species of complementizer. That some complementizers relate a clause/verb/predicate to a *clausal* head whilst others relate it to a nominal head is uncontroversial. And that complementizers with different such relational functions may employ elements from different ‘parts of speech’ is well attested. For instance, in English the *that* complementizer is clearly derived from the demonstrative ‘that’, and it relates a subordinate clause directly to a higher clause, though also relating a finite clause to a nominal head (I said that.../The fact that...). With non-finite complement clauses one uses *for* clearly a preposition by origin (‘waiting *for* him to go...’); the preposition *of* is used if the head to which the clause is a complement is itself a nominal (the fact of his going).

So, that in Lai, that in the first instance a verbal derivational suffix goes on later to appear in Comp need not be surprising. Nor need one be surprised to find that in Lai different complementizers are used for, respectively, propositional complement clauses (*tih*—a *kal tih an ti* ‘They say that he is going’) and for non-gerundive relative clauses (*mi* ‘one’, otherwise a pronominal form—a-*kal mi tlangvaal* ‘the young man that goes’).

Finally, some nominal applications for *naak* seem to be fixed, lexicalized even. Thus take *kal-naak* (*kal* = ‘go’). Speakers disagree upon whether or not it can mean the time of going, but all agree it can, usually has to, mean the place-to-which one goes. And *ihnaak* (*ih* = ‘sleep’) is fully lexicalized as simply the word for ‘bed’. So, in as much as object entities are, as it were, nearer to being true arguments (as against adjuncts) of such predicates—one’s sleeping more directly invokes the place than the time of the sleeping, since every event necessarily has a time in any case—this may well represent yet a further stage of grammaticalization towards full abstract nominalization, directed more to the clause/predicate as an event/state *type* than to just the verb proper as a specific predication; for, *ihnaak* is not a place where any given person sleeps at any given moment, but is rather simply a place where one sleeps in the abstract. This, however, requires much more thought and work to be made firm or even clear.

Notes

This work has been supported in part by a grant from the University of Illinois Urbana-Champaign Campus Research Board to the Lai Chin Dictionary Project. I must acknowledge the major contribution to this paper by my pupil and assistant, Ceu Hlun

1. I am aware that recently there have been arguments made that, for instance, the structure of Chinese relative constructions does not involve clausal complementation (see Aoun and Li MS, * Chapters 4 and 5), but I shall try to show elsewhere that, after all, the structure of relative clause constructions seems to involve complementation universally.

References

- Aoun, Joseph, and Yen-hui Audrey Li. MS. *Essays on the Representational and Derivational Nature of Grammar: the Diversity of wh-Constructions*. Los Angeles: University of Southern California, Department of Linguistics.
- Bedell, George. Forthcoming. *Genitive Constructions in Lai*. Paper for the Twelfth Annual Meeting of the Southeast Asian Linguistic Society (SEALS XII, 2002). Tempe: Arizona State University Program for Southeast Asian Studies.
- Ceu, Hlun. Forthcoming. *Pragmatic Influence on Pronouns in Lai (Hakha) Chin, with Especial Reference to Focus and Contrast*. Paper for the Twelfth Annual Meeting of the Southeast Asian Linguistic Society (SEALS XII, 2002). Tempe: Arizona State University Program for Southeast Asian Studies.
- Hartmann, Helga. 2000 *Functions of naak/na in Daai Chin with Examples from Other Chin Languages*. 33rd International Conference on Sino-Tibetan Languages and Linguistics (Bangkok)
- Huang, Yan. 2000 *Anaphora: A Cross-linguistic Study*. Oxford University Press [Oxford Studies in Typology and Linguistic Theory]
- Lehman, F. K. 1996 *Relative Clauses in Lai Chin, with Special Reference to Verb Stem Alternation and the Extension of Control Theory*. *Linguistics in the Tibeto-Burman Area* 19, 1: 43-58.
- Lehman, F. K. Forthcoming. *Larson Shell VP and Object Agreement, with Remarks on Ergative and ‘Unergative’ Transitive Clauses in Lai Chin*. *Papers from the Ninth Annual Meeting of the Southeast Asia Linguistic Society*. Tempe: Arizona State University Program for Southeast Asian Studies, Monograph Series.

THE MON-KHMER SUBSTRATE IN CHAMIC: CHAMIC, BAHNARIC & KATUIC CONTACT

Paul Sidwell

Centre for Research in Computational Linguistics (Bangkok)
& Australian National University
<paulsidwell@yahoo.com>

1 Introduction

The publication of Thurgood's (1999, henceforth "Thurgood") comparative-historical study of the Chamic languages puts the field of historical Chamic studies on a very solid foundation, yet also raises many new and important questions in relation to the history of language contact in Mainland Southeast Asia. Recent advances in the phonological and lexical reconstruction of Bahnaric and Katuic (Sidwell 2002, Sidwell & Jacq 2003, Sidwell 2005)¹ have clarified the internal genetic classifications of these groups, allowing inferences to be made about their historical locations and contacts with Proto-Chamic (henceforth "PC"). Comparison with the PC lexicon strongly suggests an ancient Sprachbund involving especially strong PC influence upon North and Central Bahnaric, but rather less direct contact with Katuic. The source of a substantial proportion of the apparently MK stratum in PC remains a mystery, and may well have been an extinct language (or languages) spoken in areas now home to Katuic or Bahnaric speakers.

2 Champa & Proto-Chamic

The Chamic languages are a Malayo-Polynesian sub-grouping, with speakers located today in Vietnam, Cambodia, Hainan Island (China) and Sumatra (Aceh Province of Indonesia). It appears that more than 2000 years ago a people speaking a language closely related to Proto-Malayic began to colonise the central Vietnam coast, and by the middle of the first Millennium they had established a series of allied coastal polities that we refer to conventionally as Champa. For around a thousand years the Cham settlements were important religious, military and trading centre of the region, rivaling Cambodia and Vietnam in importance. However, Champa entered an eight century long period of intermittent decline in the face of pressure from Vietnam, beginning with sacking of the northernmost centre of Indrapura (located near present day Da Nang) in 982, ultimately terminating in a rump Cham state persisting at Phan Rang into the 1830s. As the Vietnamese advanced southward Cham speakers variously resettled into the highlands, along the Mekong, and abroad by sea. Today we find the ironic situation that Chamic languages are virtually unknown in the zone of their ancient glory but are more or less alive and well on it's geographical periphery.

¹ These references were unpublished but circulating in manuscript form at the time of the SEALSXII meeting. Taking advantage of the unfortunate delay in the publication of these proceedings I took the liberty of updating the text of this paper to reflect the reality of these publications, as well as useful comments I received from colleagues, particularly Graham Thurgood and Anthony Grant.

It is reasonably assumed that during the first Millennium the ethnic Cham peoples spoke a more or less mutually intelligible language reconstructable as PC. A version of PC was offered by Lee (1966) and Burnham (1976) based upon Highland Chamic languages. Thurgood revised and expanded the reconstruction, taking into account a more representative range of Chamic languages, including Acehnese, which Thurgood controversially includes within the family. This proto-language has extraordinary characteristics: although self-evidently related to Malayic, it contains hundreds of Mainland SEAsian loans (many more than does Malay), and underwent partial phonological restructuring towards Mainland SEAsian type. This creates a very special opportunity for comparative linguistics, including the possibility that the inferred ad-strata may be subjected to internal and comparative investigation, revealing their identities and periodisation of contact influence.

Which were these contact language(s)? —the principle candidate according to Thurgood is Bahnaric, and secondarily Katuic. This is indicated in part by the fact that of the 267 MK-PC lexical parallels identified by Thurgood, the overwhelming majority have excellent Bahnaric matches. For the purposes of this paper I will assume that prolonged and intense Bahnaric-Chamic linguistic contact is accepted as a given, the outstanding issue being to identify the direction of Bahnaric-Chamic borrowings. Understandably Thurgood suggests that:

The early arriving pre-Chamic peoples most likely landed south of Danang and thus probably encountered Bahnarics. Given the restructuring of the arriving Austronesian language that took place, these pre-Chamic people must have become socially dominant, with this dominance leading many most probably Bahnaric-speaking people to shift to Cham, [p.251]

This conclusion is dependent on showing that the substantial common Bahnaric-Chamic lexicon is borrowed from Bahnaric. None-the-less it is clear that these two groups have (and continue) to experience significant linguistic contact. However, it is not so clear just how extensive Chamic-Katuic contact may have been. Thurgood's appendices list some 62 Katuic parallels, which appear to be the principle motivation for Thurgood's confidently assertion that:

...early contact between Chamic and Katuic speakers needs to be recognized anyway to account for the presence of Katuic borrowings in Chamic and Chamic borrowings in Katuic. (p.241)

and

Of particular interest are the Pacoh forms as they indicate intense contact between ancient Chams and the Katuic. (p.240)

Reference is also made to Reid's (1994) claim of an apparent morphological strata in Katuic, which Thurgood presumes is due to Chamic influence. I don't wish to offer a detailed critique of Reid's position here, but I read that paper as arguing for an Austric (and hence relation at much greater time depth) explanation for the MK-An morphological

parallels (which were earlier raised by Schmidt 1905, 1906). Notwithstanding the higher proportion of Chamic loans into Pacoh (clearly a local phenomenon), Katuic morphology is derived via its MK heritage, and not borrowed via later An contacts.

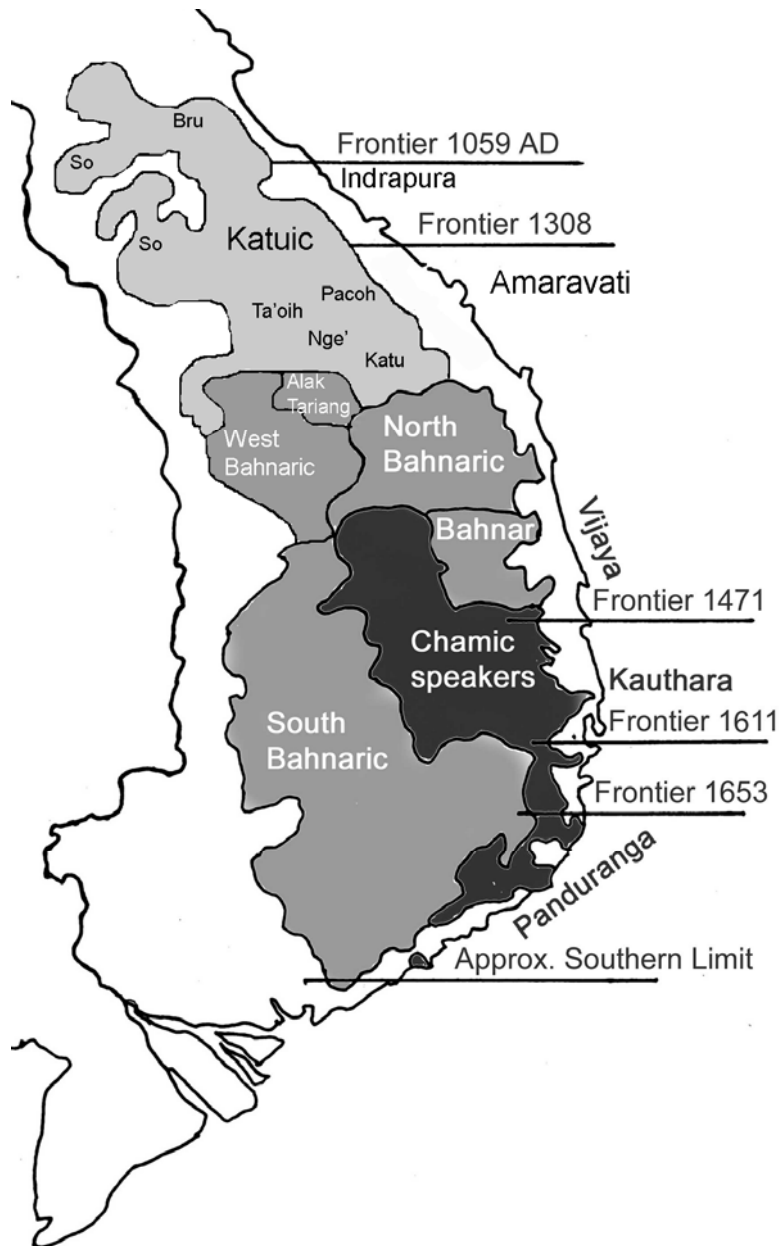


Figure 1: Map of approximate Bahnaric, Katuic and Chamic language distributions in Vietnam, and the historical Cham polities

I strongly suspect a geographical motivation strongly underlies ideas that Katuic speakers had an influence on ancient Cham, and I am sympathetic to the notion that we may gain some important insights by considering a broader synthesis of geographical and historical evidence in combination with linguistic data. It is tempting to place significance on the fact that the north-south distribution of Katuic and Bahnaric languages approximates

rather neatly the known extent of ancient Champa. The presently Katuic speaking areas lay adjacent to the old northern polities of Indrapura and Amaravati, while Bahnaric and remnant Chamic speaking areas are adjacent to Vijaya, Kauthara and Panduranga. Although the Katuic languages are certainly less Chamicised than the Bahnaric, this may simply be consistent with the lack of extensive Chamic resettlement into the Katuic zone, and the fact that the northern Cham polities were the first conquered by Vietnam.

Interestingly the main Highland Chamic intrusion (Ede and Jarai speaking) neatly corresponds in latitude to Kauthara, while the remnant coastal strip of Chamic speakers reflects the survival of an intact Panduranga into modern times.

We cannot assume that the modern distribution of Katuic and Bahnaric languages is unchanged from prehistoric times, but is rather the result of historical movements since the break-up of the respective proto-language speaking communities. Applying principles of dialect geography to the genetic classifications of these languages, we have a chance to suggest the homeland locations and migration paths of the daughter language speakers. These (necessarily tentative) results may then be correlated with what we know and presume concerning the geography of ancient Champa with what may be inferred from the distributions of loanwords. With is in mind I proceed to discuss the internal classifications of the Katuic and Bahnaric families.

3 The Katuic Family

There are approximately 15 Katuic languages spoken today, mostly in the area immediately north of the Bahnaric range and, by a huge number (perhaps a million?) Kui dialect speakers in a discontinuous spread from Laos, across into Isaan Thailand and northern Cambodia. The geographic centre of Katuic linguistic diversity lies in Laos in the hills east and north-east of Salavan, extending across into Vietnam.

It has been recognised consistently in comparative Katuic linguistics that Kui (also called Souei) sub-groups with Bru and So in a West-Katuic sub-branch (Thomas 1967 etc.). Phonologically the group is marked by a register split and vowel restructuring that is strongly parallel to that of Middle Khmer. Thus is it suggested that the Kui dialect chain, and the significant northern range of Bru and So must reflect historically late expansions of Katuic speakers under Khmer influence/dominance.

Although comparativists who have considered the issues using different methods (Thomas 1967, Efimov 1983, Diffloth 1982, Gainey 1985, Peiros, 1996, L-Tongkum 2000, Sidwell 2005) have offered various classifications, they have all agreed on the reality of the western sub-grouping and furthermore the identification of Katu (and dialects) as a distinct sub-group that branches from the highest node in the tree. On the basis my recent historical phonological analysis (Sidwell 2005) I advocate the following classification into four more or less equidistant sub-groups:

Katuic

West Katuic:	<i>Kui, Souei, Bru, So</i>
Ta'oih:	<i>Ta'oih, Katang, Talan/Ong/Ir/nh Nge'/Kriang Chatong</i>
Pacoh:	<i>Pacoh</i>
Katu:	<i>Katu, Kantu, Phuong, Triw, Dakkang</i>

Figure 2: *Classification of Katuic languages by Sidwell (2005)*

It seems likely that the PK home-range was located within the zone still occupied by Ta'oih, Pacoh and Katu speakers today, corresponding in latitude to the old Cham polity of Amaravati. It is therefore possible that Katuic speakers were not living adjacent to Indrapura during most of the first Millennium. The lexicostatistical analysis of 50 Katuic wordlists by Miller & Miller (1996) found Kui and Bru dialects to be rather close, sharing between 60% and 75% of basic vocabulary (depending on the various lists compared), and consistently less than 60% with other Katuic languages. This would suggest a West Katuic dispersal from late in the first Millennium, first northward into the Indrapura hinterland, and then south-west over the Mekong.

4 The Bahnaric family

Bahnaric is an MK family of more than 30 languages spoken in Vietnam, Cambodia and Laos. Most comparative-historical studies of the Bahnaric languages have been focussed on the smaller, self-evident subgroups, and even purported Proto-Bahnaric studies (such as my own PhD thesis of 1999) have not yet considered the totality of the family. Foundational works such as Thomas (1966), the Proto-North-Bahnaric of Smith (1972) and the Encyclopaedia Britannica entry of Diffloth (1974) distinguished a basic north-south split, with the most important discussion concerning whether Bahnar groups with the northern or southern divisions (especially Gregerson, Smith & Thomas 1976).

I recently presented a classification of the Bahnaric languages based upon historical phonology (Sidwell 2002). This scheme divides the family into 3 main branches, as follows:

Bahnaric

North:	<i>Sedang, Hrê, Rengao, Jeh, Halang etc.</i>
West:	<i>Loven, Lawi, Cheng, Oi, Laveh/Brao, Nhaheun etc.</i>
Central:	
West Central:	<i>Taliang/ Kasseng</i>
North Central:	<i>Alak</i>
East Central:	<i>Cua</i>
South Central:	<i>Tampuon, Bahnar, South Bahnaric (Stieng, Chrau etc.)</i>

Figure 3: *Classification of Bahnaric languages (Sidwell 2003)*

Both West Bahnaric (WB) and North Bahnaric show a degree of geographical and linguistic unity indicative of relative stability since the initial break-up of Proto-Bahnaric. By contrast there is geographically discontinuous distribution and greater internal diversity among Central Bahnaric (CB), implying several significant dispersal events.

Highland Chamic speakers of Jarai and Ede separate the South Bahnaric (SB) from the Bahnar, interrupting what once must have been a continuous community. On lexicostatistical grounds one may estimate that the linguistic separation of Bahnar and SB began up to two thousand years before present. According to my analysis the SB languages share about two thirds (or a little less) of basic lexicon with each other, but consistently less than half of basic lexicon with other Bahnaric languages. This suggests that SB began to disperse in the latter half of the first Millennium, having already become distinct from Bahnar etc. significantly earlier.

An even earlier separation from CB is reflected in the positions of Alak and Taliang, spoken in Laos adjacent to WB and Katuic communities. They share less than half their basic lexicons with each other and the other Bahnaric languages. Furthermore Alak and Taliang lack any significant stratum of Chamic loans, instead showing various Katuic and Khmer borrowings, in fact many of the same borrowings as found in WB languages.

This suggests the modeling of a relative chronology as follows:

1. In the earliest phase WB, NB and CB became established in approximately their present locations north of, and perhaps somewhat overlapping, the nowadays Ede speaking area, with a wedge of CB (pre-Alak, pre-Taliang) speakers between the western and northern groups.
2. In the second phase Chamic influence extends inland over the NB and the main CB speaking areas. WB and the northern wedge of CB speakers avoid Chamic influence, forming linkages with Katuic.
3. Chamic influence in the Central Highlands builds, and at some point a major dispersal of CB speakers southward and eastward was triggered.

I speculate that the phases 1 and 2 correlate with the emergence during the first Millennium of rivalries between Cham polities and Cambodia (Funan?). It is also possible that for some time into the first Millennium Katuic and Bahnaric speakers were mainly located in the hinterlands of Chamic Amaravati, Vijaya and Kauthara, and the extent of linguistic contacts and influences were conditioned by the specific circumstances of each of these polities (internal government, economy, external relations etc.). One outcome of these early conditions/events was the emergence of an areal-linguistic divide that ran through the Highlands, between the zones of Chamic and Cambodian influence (corresponding to CB, NB and Highland Chamic versus Katuic, WB and outlying CB speaking zones).

5 Analysis of lexical comparisons and conclusions

Thurgood's search for the MK sources of Chamic lexicon focused specifically upon Bahnaric and Katuic, mindful of their known geographical proximity to Chamic. It is also a practical reality that broad lexical resources for comparative MK research, such as

etymological dictionaries, were not readily available.² Consequently this preliminary exploration of the question missed an important consideration – without knowing the wider MK distributions and phonological forms of the relevant MK etyma it is not possible to reliably assess the origin and direction of borrowing of specific forms.

Focusing specifically on the Katuic issue, I extracted the 62 Chamic etymologies where Thurgood adduced a Katuic comparison, and I subjected these to a broader etymological analysis, so far as my limited dataset would allow. The results are presented in the Appendix, where the items have been organised into categories reflecting which MK form shows the closest matches to PC.

My assumption is that, although a complete treatment of Thurgood's PC etymologies is desirable, an indication of the likely extent of ancient Chamic-Katuic contact should be revealed by such an investigation. I am specifically concerned to distinguish comparisons of the following types:

- Unique PC-Katuic isoglosses, being the best evidence of specific contact.
- Widely distributed MK forms which may have been borrowed from more or less any MK source.
- Forms with specific phonological features marking them as likely borrowed from an identifiable language.

The results are quite striking, the most important summarised as follows:

- For only one item out of 62 was it possible to identify a unique Chamic-Katuic isogloss (item 1.a in the Appendix).
- 21 items are widely distributed MK etyma with no specific phonological features identifying a likely specific source of borrowing into Chamic.
- For 9 items that show a good Katuic match, there is an equally good Bahnaric match, and the restricted distributions indicate borrowing in from Chamic.
- 8 items resemble various MK words, but are not shown to be of MK origin.
- 31 items show equal or better matches with other MK branches, including Bahnaric, Vietic, Khmer, Mon, Northern MK. Allowing for the likelihood that a significant proportion of the Bahnaric and Katuic comparisons reflect inward loans from Chamic, there is no single standout MK language identifiable as the major donor to Proto-Chamic.

It would appear that there is little empirical evidence that would unambiguously “indicate intense contact between ancient Chams and the Katuic” (Thurgood p.240). And on the basis of this brief survey one may well speculate that the contact with Katuic such as it was (some dozens of words only?), was decidedly one-way, with borrowing overwhelmingly from Chamic. Even the hundreds of Bahnaric-PC comparisons appear to be mostly borrowed from Chamic, occurring after important splits among Bahnaric speakers.

This raises a greater riddle – what were the sources of the hundreds of clearly MK loans into PC (not to mention hundreds of words of unknown origin) that helped to

² Fortunately this situation is beginning to change, for example I can report that in 2006 Harry Shorto's *A Mon-Khmer Comparative Dictionary* was published by Pacific Linguistics, Canberra.

condition its dramatic lexical and phonological restructuring? I simply do not yet know. However, the reconstructed linguistic geography discussed above raises some tantalising possibilities, since during early stages of Chamic settlement it is quite likely that much of the Indo-Chinese hinterland now or recently occupied by Bahnaric and Katuic speakers was inhabited by speakers of other MK languages. I can only suggest that much more work needs to be done if we are to properly reconstruct and understand the relexification and restructuring of the ancient Chamic language.

References

- Burnham, Eugene C. 1976. *The Place of Haroi in the Chamic Languages*. M.A. thesis, University of Texas at Arlington.
- Diffloth, Gérard. 1974. Austro-Asiatic Languages. Encyclopaedia Britannica. Chicago/London/Toronto/Geneva, *Encyclopaedia Britannica* Inc. pp 480-484.
- Diffloth, Gérard. 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouïque. *Mon-Khmer Studies* 11:47-82.
- Efimov, Aleksandr. 1983. *A phonological reconstruction of Proto Katuic* (in Russian). PhD thesis, Institute of Far Eastern Studies, Moscow.
- Gainey, Jerry. 1985. *A comparative study of Kui, Bruu and So phonology from a genetic point of view*. Bangkok, Thailand: Chulalongkorn University thesis.
- Gregerson, Kenneth J., Kenneth D. Smith, and David D. Thomas. 1976. The place of Bahnar within Bahnaric. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta (eds.), *Austroasiatic studies, part I*. Oceanic Linguistics Special Publication, 13. Honolulu: University of Hawaii Press. pp.371-406.
- Jacq, Pascale. & Paul Sidwell. 2000. *A Comparative West Bahnaric Dictionary*. München, Lincom Europa.
- Lee, E.W. 1966. *Proto-Chamic Phonologic Word and Vocabulary*. Ph.D. dissertation, Indiana University.
- L-Thongkum, Therapan. 2000. *Languages of the Tribes in Xekong Province Southern Laos: a foundation for research and development*. Bangkok, Chornalonghorn University.
- Miller, John and Carolyn Miller. 1996. Lexical comparison of Katuic Mon-Khmer languages with special focus on So-Bru groups in Northeast Thailand. *Mon-Khmer Studies* 26:255-290.
- Peiros, Iliia. 1996. *Katuic comparative dictionary*. Canberra, Pacific Linguistics C-132.
- Reid, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33(2):323-344.
- Schmidt, Pater Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie*, Braunschweig, 5:59-109.
- Schmidt, Wilhelm. 1905. Grundzüge einer Lautlehre der Mon-Khmer-Sprachen. *Denkschrift der Akademie der Wissenschaften*, Wien, Philologisch-Historische Klasse 51:1-233
- Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra, Pacific Linguistics 551.
- Sidwell, Paul. 2000. *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra, Pacific Linguistics 501.
- Sidwell, Paul. 2002. Genetic classification of the Bahnaric languages: a comprehensive review. *Mon-Khmer Studies* 32:1-24.

- Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich, Lincom Europa.
- Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas, Language Data Series, Summer Institute of Linguistics.
- Thomas David. 1966. Mon-Khmer subgroupings in Vietnam. In Norman Herbert Zide, ed. *Studies in Comparative Austroasiatic Linguistics*. Indo-Iranian Monographs, V. The Hague, pp 194-202.
- Thomas, Dorothy. 1967. *A phonological reconstruction of Proto East Katuic*. MA thesis, University of North Dakota.
- Thurgood, Graham, 1999. *From Ancient Cham to Modern Dialects: two thousand years of language contact and change*. Honolulu, University of Hawai'i Press.

APPENDIX: Etymological analysis of 62 PC-MK comparisons with Katuic citations drawn from Thurgood (1999)

Note: Vetic reconstructions are from Ferlus (pers. com.) other MK reconstructions are my own unless otherwise marked. Abbreviations are as follows:

An	Austronesian	PMK	Proto-Mon-Khmer
MK	Mon-Khmer	PMP	Proto-Malayo-Polynesian
PB	Prot-Bahnaric	PSB	Proto-South Bahnaric
PC	Proto-Chamic	PV	ProtoViet-Muong
PK	Proto-Katuic	SB	South Bahnaric

1) Unique PC-Katuic matches

- a) PC ***baal** ‘mend; to patch’
Thurgood gives a PK form *_p_l ‘to patch’ (from Thomas 1967), no broader etymology apparent.

2) PC-Bahnaric matches

- a) PC ***caɗiŋ** ‘finger’
Cf. Bahnar **deŋ**, Rengao **hədiŋ** lack wider MK etymology, suggesting borrowing from Chamic. PK ***ɗŋ** ‘limb of animal/plant’ is not obviously related.
- b) PC ***ɗaŋ** ‘lie supine; be on back’
Bahnar **pəɗaŋ** ‘lie face up’ is isolated in MK and therefore likely borrowed from Chamic. It is not obvious that Katu **ləlaŋ** ‘lie on back’ can be related.
- c) PC ***kɔʔ** ‘white’
Bahnar **kɔk**, Köho (SB) **kɔ** ‘white fur’ etc. appear to be borrowed from Chamic, other MK suggest PMK ***klɔk**.
- d) PC ***lɔk** ‘peel’
Cf. PK ***lok** ~ ***lɔk**, Bahnar **plɔʔ**, PSB ***plook**. The distribution and irregularities among the MK forms suggests borrowing from Chamic and/or other source. Katuic forms such as Pacoh **lɔt** ‘peel back edges’ are not related.

3) PC-Katuic-Bahnaric matches

- a) PC ***?awaak** ‘spoon; ladle’
Bahnar **?waak** and Katu **waaʔ** appear to match, but other MK forms show front vowel reflexes corresponding to Khmer form **wæek** ‘ladle’, suggesting that Bahnar, Katu may have borrowed from Chamic, cf. Cebuano Bisayan (Austronesian) **luwag** ‘ladle’.
- b) PC ***?ale** ‘medium bamboo’
PK ***?aleɛ** ‘type of bamboo’, Bahnar dialects ***?aleɛ**, ***leɛ**.
South Bahnaric ***glee** ‘small bamboo’, Khmer **khləj** and Viet. **tre** suggest PMK

gli?. Note also North Bahnaric **pələe** ‘type of bamboo’.

On balance the source of ***ʔaləe** forms is not clear, although the **ʔa** presyllable is a typically Katuic feature.

- c) PC ***biluay** ‘gourd’
Cf. PB ***pluuɟ**, PK ***ʔaluuj** - the latter lacks initial labial, but could be secondary in Katuic. Not found elsewhere in MK.
- d) PC ***cuah** ‘sand’
Cf. Bahnar **cwəh**, PNB ***cuas**, PK ***cuah** are clearly borrowed from Chamic on phonological criteria; other MK suggest PMK ***ksaac** ~ ***skaac**.
- e) PC ***haŋɔ** ‘pine’
Cf. NB and Bahnar **həŋɔɔ** ‘pine’, PK. ***sŋɔɔ** ‘pine’.
Katuic initial /s/ suggests the proto-form.
- f) PC ***jaaŋ** ‘spirit; god’
Bahnaric and Katuic suggest ***jaaŋ** ‘spirit’, but An forms (Malay, Javanese etc. **hyang** ‘deity’) suggests origin in Western Austronesian.
- g) PC ***raw** ‘wash’
Cf. PB ***raaw**, PK ***ʔaraaw** versus Khmu’ **raa**, Viet. **rũa**.
- h) PC ***padiəŋ** ‘carry; transport’
Peiros PK ***ptiaŋ** ‘haul, transport’ on the basis of Bru **pətiəŋ** and Pacoh **pətiəŋ**; also SB forms such as Mnong **pədiəŋ**.
These data suggest borrowing from Chamic, with subsequent regular devoicing in Katuic.
- i) PC ***picah** ‘broken; break’
Cf. Malay **pecah** ‘scattered’. Restricted distribution of MK forms: (WestKatuic) Bru **pəcàh**, Kui **kəcàh** (hence Peiros’ PK ***pəc[ə/a]h**~***kəc[ə/a]h** ‘crack’), Bahnar **pəcah**, **cah**, SB e.g. Stieng **cah** ‘break’.
The lack of regular **c > s** shift indicates borrowing into Bahnar and SB indicates borrowing from Chamic.

4) PC-Katuic-Vietic matches

- a) PC ***ʔeh** ‘excrement; defecate’
Cf. PK ***ʔeh**, PV ***ʔeh**; other MK reflect a different root: ***ʔic** ~ ***ʔac**.
- b) PC ***ʔjuɔp** ‘smoke tobacco’
PV ***ʔjuɔp** ‘to smoke’ (attested in Arem and Sáčh), Katuic; Bru **juɔp**, Pacoh **ʔjuɔp** ‘to suck’.

5) PC-Katuic-Khmer matches

- a) PC ***rəh** ‘cut’
Cf. PK ***rəh** ‘cut/slash’, Khmer **rəh** ‘to cut with short, careful strokes’.

6) PC-Katuic-Northern MK matches

- a) PC ***klaj** ‘penis’
Cf. PK. ***klæ** ‘penis’, Kammu-Yuan **tléʔ**, Riang-Lang **ʔkleʔ** etc.

7) PC-Katic-Vietic-Northern MK matches

- a) PC ***krɔŋ** ‘river’
Cf. PB **kroŋ**, PV ***krɔŋ**, Old Mon **kruŋ**, Riang-Lang [ʔom] **_kroŋ**, Kammu-Yuan **krɔŋ** etc., PK ***kruŋ** ‘forest’.
Vietic and Khmuic showing best phonetic match to Chamic.

8) PC-Bahnaric-Vietic matches

- a) PC ***dəŋ** ‘stand; stop’
PV ***təŋʔ** > Viet. **dựng** [dɿŋ] ‘get up’ could be the direct source of Chamic and Bahnaric forms (e.g. Bahnar **dəŋ** ‘stand, stop’) if we can reconcile the timing of the initial voicing change in Vietnamese.
- b) PC ***kaaŋ** ‘chin; jaw’
Cf. PB ***kaaŋ**, PV ***kaaŋʔ**. Note also similar widely distributed Kadai forms such as Thai **k^haaŋ**.
PK ***tbaaŋ** (mentioned by Thurgood) is not related.
- c) PC ***cim** ‘bird’
Only SB and Vietic suggest ***cim** with short vowel, while other MK attest a long vowel (e.g. PK ***ceem**, Bahnar **sɛem** and others).

9) PC-Khmer matches

- a) PC ***buhaj** ‘otter’
PMK ***bheʔ** > OKhmer **bhee** suggests the closest match to Chamic; Vieto-Katuic forms have medial **-s-** rather than **-h-**, reflecting a secondary development.
- b) PC ***padaw** ‘warm, hot’
Mon, Khmer, Katuic, Bahnaric forms suggest PMK ***ktuuʔ**. Voicing/glottalisation shift in OKhmer **ktau** > **kɗau** proves Khmer as source for Chamic.
- c) PC ***kɔŋ** ‘bracelet’
Cf. PB ***koŋ**, PK ***kɔŋ**, Khmer **kɔŋ/kɔŋ**.
- d) PC ***laʔi** ‘basket, winnowing’
Khmer **lʔy** appears to be the source of Chamic and various Bahnaric forms, e.g. Sre **ləʔii**, Rengao **rəʔii** and others.
Thurgood noted Peiros’ PK ***ʔorie**~***korie** ‘k.o. basket’, but may rather be related to Khasi **khri** ‘small basket’.

10) PC-Northern MK matches

- a) PC ***kacuh** ‘spit’
Cf. PB ***kcoh** ‘spit’, PK. ***kcoh** ‘spit’, Mưong **chủ**, Kammu-Yuan **kəcúh**, Riang-Lang **təkchu**.
Northern MK forms agree more consistently with PChamic.
- b) PC ***ruah** ‘choose’
Cf. Bahnar **rəjh**, PSB ***ruəh**, PK ***rəəs**, OKhmer **res /rəəh/**, PWaic **rəs**.
MK forms suggest PMK ***rəəs**. Chamic vowel ***ua** may be explained by diphthongisation of [ɔ] or similar vowel attested in Waic and other NMK. SB clearly back-borrowed from Chamic, since the diphthong cannot be derived from Bahnaric vocalism.
- c) PC ***plum** ‘land leech’
Cf. PB ***pləəm**, PK ***pləəm**, PV ***pleem**, PASlian ***pləm**, Mon **kləm**, Khasi **thliem**, Palaung **plum**, PWaic ***pl̩/om**.
The short back-vowel reconstructed by Thurgood is consistent with various Northern MK forms.

11) PC-Mon matches

- a) PC ***[kulit]** **səʔ** ‘lungs; placenta’
Mon **kəsəʔ** provides a unique phonological match, NMK forms such as Lawa **hməuʔ** may also be related by infixation.
PB ***səh**, PK ***səh** etc. cannot be connected to Chamic.

12) PChamic matches broadly distributed MK forms with no specific indication of source in PC:

- a) PC ***ʔaha**, ***ha** ‘open (mouth to say sthg.)’
Mon, Khmer, West Katuic, Bahnaric, Vietic all suggest PMK ***haʔ**. Note also Viet. **hả** suggests ***hah**.
- b) PC ***ʔəʔ** ‘vomit’
Corresponding MK forms in Mon, Bahnaric, Palaungic, Nicobaric suggest such a PMK root.
PK ***həək** ‘choke on food’ is a different etymon.
- c) PC ***həʔaap** ‘yawn’
Wide distribution in MK suggests PMK ***sʔaap** ‘yawn’, specific Chamic source not obvious (cf. PMP ***uap** ‘yawn’).
- d) PC ***kalaəŋ** ‘hawk; bird of prey’
All MK branches show forms indicating PMK ***klaəŋ** or similar.
- e) PC ***klaas** ‘escape’
Widespread MK comparisons, e.g. PB ***klaas**, Bru (West-Katuic) **rəklah**, Mon **pləh**, Khasi **khlad** ‘to separate’, Kammu-Yuan **kláh** ‘to exceed’, etc.
Etymon is clearly MK in origin

- f) PC ***səna** ‘crossbow’
Khmer, Katuic, SouthBahnaric and Vietic suggest PMK ***snaʔ**, with no specific indication of which may be the source for Chamic, although it is likely that SB borrowed from Khmer or Chamic.
- g) PC ***ruaj** ‘fly; bug; insect’
Katuic, Bahnaric, Khmuic, Palaungic, Viet-Muong and Aslian all suggest PMK ***ruəj**, so there no specific indication of the source for Chamic.
- h) PC ***sagər** ‘drum’
Mon, Khmer, Bahnaric and Palaungic suggest PMK ***sgər**, while Katuic forms support **sgir** or **sgir**.
- i) PC ***tram** ‘soak’
Khmer, Katuic and Bahnaric suggest PMK ***tram**.
- j) PC ***səh** ‘only; empty; free; leisure’
Khmer, Bahnaric, Katuic suggest PMK ***səh**.
- k) PC ***sula** ‘leaf’
All MK suggest PMK ***slaʔ**.
- l) PC ***kuah** ‘shave, scrape’
Widespread MK reflexes suggest PMK ***kuəs** ‘scrape, scratch’, however there are also similar Austronesian forms, e.g. Iban **kukus** ‘to scrape out’.
- m) PC ***luaj** ‘swim’
The root is widespread in MK, although vowel reflexes considerably making reconstruction problematic – no obvious source of /ua/ apparent.
- n) PC ***prək** ‘squirrel’
All MK branches suggest PMK ***prək**, leaving no specific source for Chamic apparent.
- o) PC ***pəh** ‘open’
Cf. Palaung **puh**, Bahnar **pəh**, Nancowry **fuáh**, PV ***pəh/bəh**, PK ***pəh/poh/pooh**, suggesting various possible PMK reconstructions, and no specific indication of source for Chamic.
- p) PC ***hək** ‘pour out; spill’
PK. ***hək** ‘pour’, Biat (SB) **huk** ‘pour out’, Palaung **thəʔ** ‘id.’, Khasi **theh** ‘pour, spill’ suggest Mk etymology, but reconstruction is problematic and specific Chamic source not obvious.
- q) PC ***pər** ‘fly’
Bahnaric & Katuic suggest ***par**, other MK suggest ***pər** (e.g. PV ***pər**, Palaung ***pər** and others).
- r) PC ***haŋ** ‘hot; spicy’
Bahnaric, Vietic and Nicobar suggest PMK ***haŋ**.
- s) PC ***pah** ‘slap’
Cf. Khmer **pah** ‘hit, strike’, PV ***t-pah**, PK ***pah** ‘slap’, PSB ***pah** ‘hit’. Other

MK show back vocalism, e.g. Palaung **poh** ‘slap’. Malay **pepah** ‘strike, beat’ with /a/ < *ə suggests a competing An source for Chamic.

- t) PC ***bala** ‘tusk; ivory’
 Reflexes with /a/ main-vowel attested in North Aslian, Vietic (Arem & Sáčh only) and Bahnaric.
 However, SB, WB and other MK have forms with a back vowel suggesting PMK ***bluk** ~ ***bluək**.
- u) PC ***sapal** ‘arm’
 Viet. **vai** ‘shoulder’, PK **?apaal** ‘shoulder’, Temiar (Aslian) **pal** ‘arm’ are suggestive, but we don’t have indication of the direction of semantic shift ‘shoulder’ ↔ ‘arm’.

13) PC forms resemble, but do not match MK forms:

- a) PC ***buc** ‘uproot, pull up; scratch’
 Cf. PK ***pooc**, PV ***pɔɔc**, WKhmer **baoc** ‘pull out’.
 Voice quality and vowel differences mean that explaining a relation between Chamic and MK forms is problematic.
- b) PC ***gaj** ‘stick; wood’
 Thurgood compares PK ***gir** ‘rod’, but finals do not agree.
- c) PC ***gulac** ‘return; go home; again’
 Thurgood compares (Peiros)PK ***kə[l/h]aɔjh** ‘compensate for, return’, but the comparison is without foundation. Likely MK parallels are not evident.
- d) PC ***guj** ‘carry on back’
 Cf. PK ***kuuj** ~ ***kuj** ‘carry on back/shoulder’, PB **kɔɔj** ‘id.’ (with some regular short vowel reflexes, e.g. Hrê **kuj** ‘carry on head’).
 Voicing of the Chamic initial is not explained by Katuic/Bahnaric forms.
- e) PC ***klah** ‘lose’
 Forms such as Bahnar **klah** ‘to part, split’, PK. ***klah** ‘split’ do not show a reasonable semantic match.
- f) PC ***tubuac** ‘beak’
 PK ***crɔɔh** and PV ***k-ɔɔh** are suggestive, but do not explain the Chamic laminal final.
- g) PC ***siyaam** ‘good; nice; pretty’
 Compared by Thurgood to Katuic and Bahnaric forms suggesting ***liam**, but these cannot be the source of Chamic.
- h) PC ***pə-pət** ‘to fan’
 Thurgood compares various MK forms, but none provide a reasonable phonological match (no MK form with –t final evident).

14) Areal forms possibly via Malay:

- a) PC ***kapaas** ‘cotton’
Similar forms are found throughout Eastern and Northern MK languages, apparently borrowed from Sanskrit **karpāsa**/Pali **kappāsa**.
- b) PC ***ribuu?** ‘storm’
Probably a variant of the An root reflected in Malay **ribut**, and loaned into Bahnar **həbut** ‘wind to blow hard’. PK ***rpuu?** likely borrowed in from Chamic.
- c) PC ***kaduŋ** ‘pocket; pouch’
Cf. Bahnar **kəduŋ** ‘pocket, pouch’, PK. ***kduŋ** ‘bag, pocket’, OKhmer ***kəndaŋ** ‘basket made of leaves’, Malay ***kandung** ‘pocket, pouch’.

FROM MALAYIC TO SINITIC: THE RESTRUCTURING OF TSAT UNDER INTENSE CONTACT

Graham Thurgood

California State University, Chico

<gthurgood@csuchico.edu>

Fengxiang “Frank” Li

California State University, Chico

<FLi@csuchico.edu>

1 Background

Hainan Cham [=Tsat], an Austronesian language of Hainan, is one of the clearest examples in the literature of a language restructuring under intense contact. Some two thousand years ago, traders speaking a Malayo-Chamic language set up trading posts on the coast of modern day Vietnam. Interaction and intermarriage with speakers of Bahnar led to the total restructuring of the language; under intense contact with Bahnar, it became Chamic, a language that differed strikingly from its Malayo-Chamic ancestor in phonology, morphology, lexicon, and syntax (constructions). For roughly a thousand years, this newly restructured Chamic language—the language of the Champa Federation—existed as an only moderately differentiated dialect continuum along the coastline of southern Vietnam, with a small trading post on Hainan Island.

The next major restructuring occurred after the northern capital fell to the Vietnamese in 982. This event led the Northern Cham to split into two groups: the bulk of the merchant class (including many Muslims—Huihui, in Chinese) fled to Hainan becoming the *Utsat* (etymologically, *u* ‘people classifier’ + *Tsat*, **cam* ‘Cham’; note that, a thousand years ago, all Chamic speakers were more than likely called Chams), while the bulk of the non-merchant class fled to south and, in many cases, inland, becoming the Northern Roglai (etymologically, *ra* ‘people’ + *glai* ‘forest’). It is worth noting that the Northern Roglai are the Chamic group reputed to have the royal treasures from the northern capital.

The arrival of the Northern Cham traders on Hainan—an event noted in the Chinese Dynastic records—led to another complete restructuring of the language, this time under the influence of the monosyllabic and tonal Hlai languages (Tai-Kadai), the monosyllabic and tonal Min dialects of Chinese, and, more recently, under the quite intense influence of Mandarin.

2 Restructuring the Phonology

The phonological restructuring was significant, although the initial steps were simple enough. Malayo-Chamic had penultimate stress, but like some dialects of modern Malay, when the penultimate vowel was shwa, it more than likely had final stress. Under influence from Bahnar speakers (and possibly other Mon-Khmer groups), stress switched to final position. This, combined with continued interaction with Bahnar speakers whose languages

were sesquisyllabic (weaker presyllable + stressed final syllable; in terms of stress, iambic) led to the change from disyllabic to sesquisyllabic forms.

While the four unstressed first syllable vowels remained unchanged, the stressed main syllable vowels proliferated. The inherited vowels went from seven to nine (four monophthongs and three diphthongs), with the splitting of the two high vowels into a diphthongized stressed variant and an unstressed, undiphthongized short variant. A number of vowels were borrowed into Chamic from Bahnar sources, although usually one or more inherited forms appear to have first developed the vowel phonetically, with this outlier providing a model for the borrowing. Finally, a length distinction developed, apparently triggered initially by the lowering of the inherited shwa to a short /a/, thus providing a length contrast with the inherited /a/. Continued contact with Bahnar reduced the four-way contrast in the first syllable in many dialects of Chamic, making the structure more sesquisyllabic. (for more on vowels, see Thurgood 1998, 1999)

Another minor change was the proliferation of glottalized stops, again apparently first through changes in inherited forms with native material (see Greenberg 1970; Thurgood 1999:87-94). This opened the way for borrowings, although given the intensity of the contact, the glottalized stops most likely would have been borrowed in any case.

Last, but not least, was the development of a register system, that is, contrasting voice qualities, typically a two-way between breathy voiced vowels (< earlier voiced obstruents) in contrast with modal or clear register. This simple register distinction may be reconstructable to PC. There is some question whether this register distinction developed under Mon-Khmer/Bahnaric influence or not; our initial assessment was that it did, but others such as Sidwell (p.c.) argue that Bahnar was not registral at that point. Be that as it may, register was widespread and certainly existed in the history of Hainan Cham. Other Chamic languages have gone on to elaborate their own systems in various ways (Thurgood 1996).

The final restructuring of Hainan Cham phonology takes place after their arrival on Hainan and had come into more intense contact both with Hlai speakers and with Min speakers, that is, sometime after 982. The two salient features are the much accelerated movement from sesquisyllabic to monosyllabic and the development of tones, but there was also simplification in the vowel system.

The increased monosyllabification sometimes came about through collapsing the two syllables into one. If the medial syllable began with *-h-, the monosyllabification was completed before the arrival on Hainan, perhaps as far back as the PC stage: as nothing has been distorted by doing so, we have used Malay to represent the Malayo-Chamic stage: Malay *tahun* ‘year’ cf. PC *thũn > Hainan Cham *t^hun*³³ (note that in some dialects of Chamic the reflexes of *th- actually are pronounced as clusters, not aspirated stops). Collapse of forms with medial *-l- or *-r- into monosyllables postdates the Hainan Cham arrival on Hainan, as several Northern Rglai sources still have two syllable forms, but otherwise the developmental pattern parallels that for the medial *-h-: Malay *bulan* ‘moon’ cf. PC *bila:n > Hainan Cham *p^hian*¹¹. Finally, where it was not possible to collapse the two syllables into one, a process with an intermediate stage still seen in Rade but completed in Hainan Cham: Malay *basah* ‘wet’ cf. PC *basah > Rade *msah*, Hainan Cham *sa*⁵⁵.

The most discussed development, however, is the development of a tone system, a tone system that parallels the tone systems of the other languages of Hainan, including the

Hlai dialects and Min dialects that Hainan Cham speakers had contact with. The external motivation for the tonogenesis was contact; the internal paths are transparent: forms ending in *-h have a 55 tone (high level); forms ending in a glottal stop have a 42 falling tone if the form began with a PC voiced obstruent (which led to breathiness, which determined the tone class), but with a 24 rising tone, if it did not; and, the remaining forms have a 11 tone (low level tone), if the form began with a PC voiced obstruent (which led to breathiness, which determined the low tone class), and a 33 (mid level) tone, if it did not. The actual picture is slightly more complicated; in PC disyllabic forms with two syllables, if the initial of the first syllable was a PC voiced obstruent, the breathiness from that obstruent spread through the medial and this breathiness determined the tone class of the second syllable. (Haudricourt 1984, Maddieson and Pang 1993, Thurgood 1996)

3 Lexicon

The vocabulary was restructured first in Champa and then again on Hainan. Half of the vocabulary, including much of the core vocabulary, is Bahnaric. In fact, there are often doublets, with one form inherited, the other borrowed. Also a significant number of body part words are borrowed e.g. *cadiəŋ ‘finger’, *sua ‘dead skin’ (*kulit ‘skin’ is inherited). Other borrowings included kinship terms, adjectives, nouns, verbs, and so on. In fact, the borrowing is so massive that the language was sometimes thought to be Mon-Khmer, rather than Austronesian.

There is also the occasional borrowing of a grammatical morpheme, such as the negative imperative marker. More significant for the morphology is the iambic stress pattern; prefixes were reduced and then lost.

The arrival on Hainan led first to the borrowing of a small amount of Hlai vocabulary (and Hlai borrowed some from Chamic, most notably the word *nam ‘six’). Contact with Chinese, in contrast, led to massive lexical borrowing, including and grammatical morphemes (discussed in the next section). Zheng (1997:54) writes that, of some 2428 lexical items, roughly 20% are of Chinese origin: 21% of the nouns, 14% of the verbs, 31% of the adjectives, 31% of the classifiers, and several pronouns.

Much of the original Austronesian vocabulary is now gone, some lost to Bahnaric (Mon-Khmer) borrowings and now some to Chinese.

4 Constructions

The Hainan Cham came to Hainan speaking a language with limited grammatical morphology and with constructions marked with a grammatical morpheme plus word order. All of this is being rapidly restructured under contact with Chinese; we suspect that two aspects of this restructuring are the effects of the mass media and the results of near-universal schooling. Much of this is discussed in Thurgood and Li (2003), but it can be illustrated briefly here.

4.1 Genitives with Full Noun Phrases

In Northern Roglai all full NPs are postposed. In Hainan Cham, all genitive full NPs are preposed, as in Chinese, with the construction marked by *sa*³³ even in the most colloquial, least-Sinicized texts.

- (1) Northern Roglai: Nh GENNP
 gaʔ sa:k (Lee 1966:65)
 roof house
 ‘the roof of the house’
 Hainan Cham (all) GENNP sa³³ Nh
- (2) ...piaɪ³³ sa³³ za:ŋ³². (Zheng 1997:95)
 ...village GEN person
 ...cūn de rén
 ...cūn de rén
 ‘people of the village...’

In the most Sinicized texts, the native genitive marker *sa*³³ has been replaced by *tɪ*³, borrowed from Chinese.

- (3) Hainan Cham (Mandarinized) GENNP *tɪ*³³ HeadNP
 tan³³ k^hua⁵⁵ tɪ³³ si¹¹ha:u²¹, (Zheng 1997:4.1.3)
 arrive daybreak GEN after
 dào tiānliàng de shíhòu ...
 dào tiānliàng shíhòu, ...
 ‘At daybreak, ...’

4.2 Genitives with Pronouns

Genitives with pronouns show a mixed pattern in Hainan Cham. In more colloquial texts, the pronominal genitive is simply preposed without a genitive marker; in the more Sinicized texts it tends to be preposed with a genitive marker, as in Chinese. In Chinese, some variation in the use of a genitive marker occurs, apparently correlating with the transparency of situation being coded.

- Northern Roglai: Nh GENPr
 (4) sa:k hǎ (Lee 1966:65)
 house you
 ‘your house’
 Hainan Cham (colloquial): Nh GENPr
- (5) ko²⁴ ʔbu²⁴ nau³³ sa²⁴. (Zheng 1997:92)
 head.hair she messy
 tóufa tā luàn
 tā de tóufa luàn.
 ‘Her hair is messy.’
 Hainan Cham (Chinese influenced, with *sa*³³)
- (6) nau³³ sa³³ ko²⁴ ʔbu²⁴ sa²⁴. (Zheng 1997:97)
 she GEN head.hair messy
 tā de tóufa luàn
 tā de tóufa luàn.

‘Her hair is messy.’

4.3 Demonstratives and Head Nouns

As with genitive pronouns, both demonstratives and adjectives are postposed (without a genitive marker) in Northern Roglai and colloquial Hainan Cham, but preposed (with and without a genitive marker) in Chinese-influenced Hainan Cham and Mandarin. The tendency is for the genitive marker to show up in the Chinese-influenced Hainan Cham, a construction that reflects Mandarin influence.

- Northern Roglai:
 (7) sa:k gheŋ ʔunĩ (Lee 1966:65)
 house big this
 ‘this big house’
- Hainan Cham:
 (8) ʔai³³ ni³³ sat²⁴ ʔan³³ (Zheng 1997:84)
 water this truly cold
 shuĩ zhè zhēn lěng
 zhè shuĩ zhēn lěng.
 ‘This water is very cold.’
 Hainan Cham (Chinese influenced) (this + GEN) + clf
- (9) ni³³ sa³³ ta¹¹ p^han³² pi¹¹ kiau³³ lu³³. (Zheng 1997:75)
 this GEN one clf CM much
 zhè de yī fèn bǐjiào duō
 zhè yī fèn bǐjiào duō.
 ‘This portion is bigger.’

The alternation is found in the speech of the same speaker and correlates with the text type: the borrowed patterns are found in texts that describe more recent phenomena, whereas the native patterns are used in texts of traditional stories.

4.4 Adjectives and Head Nouns

The pattern for adjectives is the same. The N. Roglai has postposed adjectives as does the colloquial Hainan Cham, while the Mandarinized Hainan Cham has preposed adjectives, both with and without a genitive marker, calquing the Chinese.

- (10) N. Roglai: postposed adjectives
 sa:k gheŋ ʔunĩ (Lee 1966:65)
 house big this
 ‘this big house’
- (11) Hainan Cham: postposed adjectives
 na¹¹tsun³³ pion^ʔ24 poi^ʔ24: (Zheng 1997:1.1.9)
 bird big say
 niǎo dà shuō

dà niǎo shuō:

‘The big bird said:...’

Hainan Cham (Mandarin-influenced) preposed adjectives

- (12) ... kiú³³ san³³ (Zheng 1997:2.1.1)
 ... old village
 ... jiù cūn
 ‘... the old village’
 preposed with *sa*³³, a calque on Mandarin *de*
- (13) na:iŋ³² sa³³ saŋ³³ huat²⁴, (Zheng 1997:2.1.10)
 good GEN life
 hǎo de shēnghuó
 ‘(the) good life’

4.5 Comparative Constructions

There are two distinct Hainan Cham comparative patterns: the native pattern is inherited from Chamic; the other is borrowed from Chinese. As Zheng writes (1997:75), the native construction is: X - Adj - CM/ST, with the adjective preceding the comparative marker (*la:uŋ*³² ‘CM; pass’) and the standard, while the Chinese pattern is: X - CM/ST - Adj, with the adjectives following the comparative marker (*pi*¹¹ ‘CM’ < Chinese) and the standard.

Hainan Cham (colloquial):

- (14) nau³³ ma⁴³ la:uŋ³² ha³³. (Zheng 1997:75)
 he fat CM you
 tā pàng bǐ nǐ
 tā bǐ nǐ pàng.
 ‘He is fatter than you.’
 Hainan Cham influenced by Chinese
- (15) kau³³ pi¹¹ ha³³ tsat²⁴ tso³³ kiə³³ sun³³. (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.’

4.6 Adverbs and Conjunctions from Chinese

As you might recall, some 45% of adverbs, prepositions (like *cong* ‘from’), and conjunctions are borrowed from Mandarin (Zheng 1997:54). In the case of these, what has been borrowed is not a grammatical word, but instead a construction along with the grammatical word that marks it. The rules of usage seem strikingly like those in Chinese.

- Adverbs: (the examples given here are intensifiers)
- (16) t̪^ha:i³³ ʔdiaŋ²⁴. p̪^hai³³ sianŋ²¹ na:iŋ³². (Zheng 1997:76)
 very hot extremely good
 tài rè fēicháng hǎo
 tài rè fēicháng hǎo

‘very hot’ ‘extremely good’

- Conjunctions:
 (17) zi³³ pa³³ ziu³³ ha:i³³. (Zheng 1997:84)
 both hungry and tired
 yòu è yòu lèi
 yòu è yòu lèi.
 ‘Both hungry and tired.’
- (18) zi¹¹ ko¹¹ khiə⁴³ t^ha:i³³ ʔdai²⁴, kau³³ sau⁴³ pu³³ na:u³² lə³³.
 if tomorrow very hot, I then NEG go PERF
 rúguǒ míngtiān tài rè, wǒ jiù bú qù le (Zheng 1997:85)
 rúguǒ míngtiān tài rè, wǒ jiù bú qù le
 ‘If tomorrow is very hot, I won’t go.’

In discussing, not Hainan Cham, but the geographically distant and genetically distinct Tai-Kadai language Mulam, Zheng Guoqiao (1988:173) wrote that in Mulam the degree and quantity adverbs are all borrowed from Han Chinese and that these were subject to the same syntactic rules as in Han Chinese. Whether or not, it is literally true, it is instructive that a good scholar would make such a statement. The Hainan Cham parallels seem striking.

4.7 Other Chinese-influenced Constructions

All sorts of other constructions have been borrowed. In (19) below are three separate constructions showing Chinese syntactic influence. The first, indicated by the initial italicized characters with underline, is the extension of the prehead modification of the *sa*³³ construction to produce a prehead relative clause.

- (19) ʔdi⁵⁵ nan³³ sa³³ mo³³ si¹¹ mai³³ sa³³. (Zheng 1997:73)
 lie.down that GEN cow be female GEN
 tǎng nà de huángniú shì mǔ de
 tǎngzhe de nà tóu huángniú shì mǔ de.
 ‘The yellow cow lying down is female.’

The second is the use of the Chinese borrowing *si*¹¹ to mark the equative construction; the Chamic languages seem to use simple juxtaposition. And the third is the use of a postposed *sa*³³ as a nominalization in the *mai*³³ *sa*³³ ‘female’. All three reflect Chinese influence.

5 Observations

The successive stages in the restructuring of Hainan Cham, first in Champa and now on Hainan, reflect instances of intense contact. One suspects, in fact, that were it not for their identity as Hui, Muslims, Hainan Cham might very well have disappeared by now. As it is, the language has retained little more than words from its origins—the phonology is gone, the constructions are gone, and, although we did not discuss it here, even the rhetorical structures are becoming Chinese.

While the internal paths that brought about these changes are relatively transparent, the directionality and the impetus are provided by social, not linguistic pressures. 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. Much of the language is gone; the rest will follow in another generation. What we fail to record now will be irretrievably lost.

References

- Greenberg, Joseph H. 1970. Some generalizations concerning glottalic consonants, especially implosives. *IJAL* 36:25-45.
- Haudricourt, André-G. 1984. Tones of some languages in Hainan. *Minzu Yuwen* 4:17-25.
- Lee, Ernest Wilson. 1966. Proto-Chamic phonologic word and vocabulary. Ph.D. dissertation. Indiana University. 67-3690.
- Maddieson, Ian and Keng-Fong Pang. 1993. Tone in Utsat [Hainan Cham]. *Tonality in Austronesian Languages*. Edited by Jerry Edmondson and Ken Gregerson. Oceanic Linguistics Special Publication No. 24. Honolulu, Hawaii: University of Hawaii Press. Pp. 75-89.
- Thurgood, Graham. 1999. *From Ancient Cham to modern dialects: Two thousand years of language contact and change*. Oceanic Special Publication No. 28. Honolulu: University of Hawai'i Press.
- _____. 1996. Language contact and the directionality of internal 'drift': the development of tones and registers in Chamic. *Language* 71.1:1-31.
- _____. 1998. The Austronesian and Mon-Khmer components in the proto-Chamic vowel system. *Papers in Southeast Asian Linguistics No. 15: Further Chamic Studies*. Edited by David Thomas. Pacific Linguistics. Series A-89. Canberra: The Australian National University. Pp. 61-90.
- Thurgood, Graham and Fengxiang Li. 2003. Contact induced variation and syntactic change in the Tsat of Hainan. *Language variation: Papers on variation in the Sinosphere and the Indosphere in honour of James A. Matisoff*. Pacific Linguistics. Pp. 185-200.
- 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*. Summer Institute of Linguistics. Pp. 167-177.
- Zheng, Yiqing. 1986. 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 Chubanshe [Shanghai Far East Publishers].

PERCEPTUAL DISCRIMINATION OF THAI TONES BY NAÏVE AND EXPERIENCED LEARNERS OF THAI

Ratree Wayland
University of Florida
<ratree@ufl.edu>

Susan Guion
University of Oregon
<guion@uoregon.edu>

0 Introduction

The present study was conducted to investigate the ability to discriminate the mid and low tone contrast in Thai by two groups of native English (NE) speakers and a control group of native Thai (NT) speakers. The first group was comprised of NE speakers who had no prior experience with Thai, while subjects in the second group were experienced learners of Thai (EE). The variables under investigation were experience with Thai, discrimination of open vs. closed syllables, and the inter-stimulus-interval (ISI) of the presentation (500 vs. 1500 ms).

1 Methodology

Subjects: Sixteen native speakers of American English participated as experimental subjects and eight native speakers of Thai participated as control subjects in the study. All native Thai speakers were from Bangkok and the native English speakers were originally from different regions in the U.S. The native English speaker subjects were divided into two sub-groups with eight subjects in each group: the naïve and the experienced groups. The Native Thai (NT) subjects were recruited from the student population at the University of Florida at Gainesville and the native English speakers were mostly students from the University of Oregon. The NT subjects were between the age of 23 and 28 years (*mean* = 24.5 years). The naïve English (NE) group were between the age of 21 and 47 years old (*mean* = 34 years old), and the experienced English (EE) group were between the age of 20 and 43 years of age (*mean* = 30 years old). Subjects in the NE group had no prior experience with Thai while those in the EE group have been studying Thai (*mean* = 2.5 years, range 1 to 5 years) and have lived in Thailand (*mean* = 4 years, range = 1-12 years). All subjects reported no prior history of speech or hearing impairment.

Stimuli: Stimuli were eight minimal pairs or contrasts (see Table 1 below) of low and mid tone of standard Thai produced by a 36 year-old female native speaker of Thai. Five out of eight contrasts (1-5) are open syllables and the remaining three contrasts (6-8) are closed syllables.

Table 1: *Minimal pairs used in the study*

Mid Tone		Low Tone	
1. [pi:]	‘year’	[pì:]	‘oboe’
2. [pa:]	‘to throw’	[pà:]	‘forest’
3. [k ^h a:]	‘to be stuck’	[k ^h à:]	‘galanga’
4. [t ^h a:y]	‘to guess’	[t ^h à:y]	‘to change’
5. [k ^h a:y]	‘to spit out’	[k ^h à:y]	‘a net’
6. [pa:n]	‘birthmark’	[pà:n]	‘sack’
7. [pan]	‘to share’	[pàn]	‘to pedal’
8. [ʔa:n]	‘saddle’	[ʔà:n]	‘to read’

These contrasts were produced in a Thai carrier phrase “[rau phu:t k^ham wâ: ...]”, “we say the word....” Each contrast was produced three times in random order. The recording took place in a quiet office setting using a high quality DAT cassette recorder (Sony TC-DD8) and a head-mounted microphone (Shure, model SM 10A). The microphone was placed at a 45-degree angle approximately 13 mm from the mouth. The stimuli were later digitized using Cool Edit (Syntrillium Inc.) at 22.05 kHz, with a 16-bit quantization. Each target syllable was then excised out of the carrier phrase and saved as an individual file. All target syllables were normalized for peak intensity (50% of the scale).

Fundamental frequency at the beginning (F_0 onset) and at the end (F_0 offset) of the vowel of all target words were obtained using Pitchworks. These data were analyzed in a series of two-tailed paired-tests examining the difference between closed and open syllables. The results revealed that the mid and the low tones in open syllables were differentiated based on the F_0 onset [$t(15) = 2.14, p < .02$], while both F_0 onset [$t(8) = 2.31, p < .004$], and F_0 offset [$t(8) = 2.31, p < .01$] differentiated the mid and the low tones in closed syllable.

Procedure: The three productions of each word were used in constructing the test. The stimuli were presented in triads designed to test a single contrast. In any given triad, no two stimuli were exactly alike. Two instances of the same word were presented as two different productions.

Each of the eight contrasts was tested by six ‘different’ trials, which consisted of a single token of one word and two tokens of the other word with a different tone. For example, a trial testing the contrast [pi:]/[pì:] might consist of [pi:]-1, [pì:]-3, [pi:]-2 (where the number indicates different productions). In the example given, the tone in the second stimulus is the odd item out because it contains a tone that differs from the first and the third stimuli. The serial position of the odd item out was distributed equally over the three possible positions in the different trials.

Each tone contrast was also tested by four ‘catch’ trials, which consisted of three physically different instances (i.e. different productions) of a single tone. Two catch trials testing each contrast consisted of three instances of one member of the pair (e.g., [pi:]-1, [pi:]-2, [pi:]-3), and the remaining two catch trials consisted of three instances of the other member of the pair (e.g., [pì:]-3, [pì:]-2, [pì:]-1).

To test the effect of ISI, two versions of the test were created. In one version, the interval between the three stimuli in each trial was set at 500 ms, and in the other it was set

at 1500 ms. However, the interval between each response and the presentation of the next trial (ITI) was always set at 1500 ms.

The subjects were tested individually in a quiet room in one session that lasted about 30-45 minutes using a PC. The 160 (8 pairs x 6 different trials + 8 pairs x 4 catch trials x 2 ISIs) trials were randomly presented over headphones at a comfortable listening level. The subjects were told that each trial would be made up of three Thai words spoken by a female native Thai speaker and that they were to focus their attention on the tone or pitch level of the word. They were told to push a button marked “1”, “2”, or “3” if the tone in one word differed from the tone in the other two words, but to click the fourth button, marked ‘none’, if they heard three words produced with the same tone. For example, the button marked “1” will be selected if they think the first word they heard was the one produced with a different tone from the second and the third words. All subjects were tested on both ISIs (500 and 1500 ms) and the order of presentation of the two tests was counter-balanced across subjects. To familiarize subjects with the stimuli and rate of presentations, a short practice session without feedback was provided. Moreover, in each block, the 80 experimental trials were preceded by five practice trials that were not analyzed.

2 Dependent variable

The proportion of ‘hits’ was determined for each contrast by determining how many times, out of a maximum of six, that the odd item out was correctly selected in the different trials. The proportion of ‘false alarms’ was the number of times out of a maximum of four that an odd item out was incorrectly selected in catch trial. An A’ value was then calculated¹ for each of the eight contrast pairs for each subject to provide an estimate of phonetic sensitivity (see Snodgrass, Levy-Berger & Haydon, 1985), taking into account the proportions of both ‘hits’ and ‘false alarms’. An A’ score of .5 would be obtained if the proportion of ‘hits’ equaled that of ‘false alarms’. If the proportion of ‘hits’ was greater than that of ‘false alarms’, then an A’ score greater than .5 would be obtained and vice versa when the proportion of ‘hits’ was smaller than that of ‘false alarms’. Therefore, an A’ score of 1 indicated perfect discrimination while an A’ score of 0.5 or less indicated a lack of phonetic sensitivity.

An inspection of the data suggested that the subjects understood and were able to perform the task. Except for two subjects in the NE groups, all subjects obtained a perfect score of 1.0 on at least one contrast. The highest scores for the two NE subjects were .94 and .96.

3 Results

Effect of ISI: A’ scores were calculated for each subject in each ISI condition. The average A’ scores for each group in each ISI condition are shown in Table 2.

As expected, the native Thai subjects obtained higher A’ scores, on the average, than both groups of native English speakers for both ISI conditions. The EE group also obtained higher scores than the NE group.

Table 2: Mean A' scores for all three groups for each ISI condition.

ISI	Group			
	NT	EE	NE	Mean
1500	.91 (.18)	.85 (.24)	.77 (.26)	.84 (.23)
500	.92 (.15)	.87 (.20)	.79 (.23)	.86 (.19)

These data were analyzed in a Group (3) x ISI (2) analysis of variance (ANOVA). This analysis yielded a significant main effect of Group, but only a marginal significant effect for ISI (Table 3).

Table 3: Results of Group x ISI ANOVA

Group	Tukey's (Group)	ISI
$F(2,89) = 6.96,$ $p < .001$	NT > NE, $p < .001$	$F(1,189) = 3.62,$ $p < .059$

There was no significant interaction between the two factors. A pair-wise comparison using the Tukey's method revealed that the NT speakers obtained significantly higher A' score than the NE only.

Effect of Syllable Type: Mean A' scores for closed and open syllables obtained for each group for both ISI conditions are shown below in Table 4. As predicted, all three groups obtained higher A' scores for closed syllables than for open syllables. Moreover, native speakers of Thai obtained higher scores than the NE group on closed syllables, and higher scores than both the EE and NE groups on open syllables.

These data were analyzed in a Group (3) x ISI (2) x Syllable Type (2) ANOVA. A significant main effect for both Group and Syllable Type was obtained (Table 5).

There was no significant interaction. A Tukey's pairwise comparison revealed that both the NT and EE groups obtained significantly higher A' scores than the NE group.

Table 4: Mean A' scores for closed and open syllables for all three groups of subjects.

Type	Group			
	NT	EE	NE	Mean
closed	.97 (.06)	.97 (.05)	.87 (.15)	.94 (.26)
open	.88 (.20)	.79 (.26)	.72 (.27)	.80 (.24)

Table 5: Results of the Group x ISI x syllable type ANOVA

Group	Tukey's (Group)
$F(2,186) = 6.71, p < .002$	NT, EE > NE, $p < .001, .004$
Syllable Type	
$F(1,186) = 24.97, p < .001$	

Table 6: Mean A' scores for each group by each contrast. SDs are given in parentheses.

Contrast	ISI	NT	EE	NE	MEAN
1.	1500	.91 (.07)	.75 (.25)	.65 (.25)	.77 (.19)
	500	.87 (.12)	.86 (.10)	.62 (.27)	.78 (.16)
2.	1500	.52 (.23)	.33 (.16)	.27 (.09)	.37 (.16)
	500	.60 (.21)	.45 (.15)	.40 (.23)	.48 (.20)
3.	1500	.97 (.08)	.90 (.10)	.86 (.08)	.91 (.09)
	500	.98 (.03)	.93 (.07)	.84 (.11)	.92 (.11)
4.	1500	.96 (.06)	.89 (.13)	.83 (.16)	.89 (.12)
	500	.95 (.08)	0.83 (.23)	.80 (.17)	.86 (.16)
5.	1500	.99 (.02)	.99 (.02)	.96 (.03)	.98 (.02)
	500	1.00 (.0)	0.99 (.03)	.98 (.03)	.99 (.02)
6.	1500	.96 (.07)	.96 (.07)	.87 (.19)	.93 (.11)
	500	.98 (.03)	.98 (.03)	.88 (.06)	.95 (.04)
7.	1500	.99 (.02)	0.99 (.02)	.80 (.23)	.93 (.09)
	500	.98 (.03)	.97 (.05)	.86 (.11)	.94 (.06)
8.	1500	.94 (.08)	.96 (.08)	.92 (.08)	.94 (.08)
	500	.97 (.08)	.99 (.02)	.92 (.08)	.96 (.06)

Effect of Contrasts: Mean A' scores obtained by all three groups of subjects on each contrast in each ISI condition are shown in Table 6. These data showed that subjects in the NT group obtained relatively higher scores on all contrasts than the two native English speaker groups. The EE group also had higher scores than the NE group. Moreover, contrast 1 and especially contrast 2 obtained relatively lower scores than all other contrasts.

These data were analyzed in a Group (3) x Contrast (8) x ISI (2) ANOVA. A significant main effect of both the Group and Contrast factors were obtained (Table 7). Tukey's pairwise comparisons for the Group factor revealed that the NT group obtained a significantly higher score than both the EE and NE groups. More interestingly, the Tukey's analysis also showed that the EE group obtained a significantly higher score than the NE group.

Table 7: Main effect of Group x contrast x ISI

Group	Tukey's
$F(2, 169) = 23.54, p < .001$	NT > EE, NE $p < .02, .001$ EE > NE $p < .001$
Contrast	Tukey's
$F(7, 168) = 64.269, p < .001$	1 > 2 1, 2 < all other contrasts

Tukey's pairwise comparisons for Contrast revealed that while contrast 1 (flute/year) obtained a higher score than contrast 2 (forest/throw), both contrasts obtained significantly lower scores than all other contrasts (p values range from .03 to .001).

Effects of ISI for Difficult Contrasts: In this section effect of ISI in the hardest (two) contrasts was tested. For this reason, a Group (3) x Contrast (2) x ISI (2) ANOVA was conducted. As expected, this analysis yielded a significant main effect of Group as well as of ISI (Table 8). The simple effect of Group was found for both ISIs. Tukey's pair-wise analysis revealed that for both ISIs, the NT group obtained significantly higher score than the NE group only. More interestingly, the simple effect of ISI was found only for the EE group.

Table 8: Main and simple effects of Group x ISI on two difficult contrasts.

Main effects		
Group	ISI	
$F(2, 45) = 4.03, p < .02$	$F(1, 45) = 4.42, p < .04$	
Simple effects		
Group		
ISI 1500	ISI 500	Tukey's (both ISI)
$F(2,45) = 3.64, p < .03$	$F(2,45) = 3.56, p < .04$	NT > NE, $p < .03$
ISI		
NT	EE	NE
<i>ns</i>	$F(1,15) = 6.19, p < .03$	<i>ns</i>

Summary of Results: The results of the analyses presented here revealed that (a) when all contrasts were considered, there was no effect of ISI; (b) when data from only the most two difficult contrasts were considered, a significant effect of ISI was obtained. This effect was significant, however, only for the EE group. For subjects in this group, higher scores were obtained for the shorter ISI. (c) All three groups of subjects were able to discriminate the tonal contrast in question better in the closed syllable condition than in the open syllable condition. Finally, (d) the NT group had higher discrimination scores than the NE and EE groups and (e) the EE group had higher scores than the NE group.

4 Discussion and Conclusion

This study was designed to explore the ability to discriminate between the low and the mid tones of Thai by two groups of native American English speakers, one naïve and one experienced. Subjects in the naïve group were those who had never been exposed to Thai while those in the experienced groups had studied Thai and had lived in Thailand for varying amount of time. The study sought to explore the question of whether or not experience with the target language resulted in improved ability to discriminate the tone contrast in question. The study also asked the question of whether or not the subjects' performance differed as a function of ISI. The interaction between experience and ISI was also explored.

The results indicated that, indeed experience played a facilitative role in tonal discrimination. Subjects who were experienced with the target tones outperformed those who were not. This finding was expected and consistent with results previously found in other studies on segmental features such as vowels and consonants (e.g., Yamada & Tokhura, 1991; Best & Strange, 1992; Flege, Takagi & Mann, 1995). Moreover, the

finding that there was no difference between the NT and the EE, at least in closed syllables, suggested that a native-like discrimination can be achieved among adult L2 learners of a tone language.

The ISI effect emerged only when the two most difficult contrasts were included in the analysis. This effect was found to be significant only for the experienced native English speakers. That is, higher scores were obtained for the shorter ISI (500 ms) for this group of subjects. We hypothesize that the heavy demand on short-term memory posed by the oddity discrimination task used in the study may have been responsible for this finding. In other words, information in short-term memory may dissipate with a longer ISI for the EE group. The fact that this limitation did not have an impact on the native Thai speakers' performance suggested that they were better able to hold in short-term memory the information necessary to differentiate the tones than the experienced native English speakers. This explanation implies that an interaction between the type of cues detected and the duration of their retention in short-term memory may exist. That is, given a difference in the kinds (and/or degree of salience) of acoustic cues, native and non-native speakers may differ in their ability to code the stimuli in terms of tonal categories held in long-term memory. This may be due to more robust long-term memory representations in the NT listeners. This hypothesis, however, deserves further investigation.

As for the naïve native English speakers, the absence of an ISI effect and their relative poor performance suggested they suffered both from the inability to detect appropriate cues and the demand on short-term memory imposed by the discrimination task. The acoustic cues they extracted may not have been sufficiently salient to allow for a categorical processing at shorter ISI, and a combination of a lack of tone representation in long-term memory and the decay of acoustic cues detected contributed to their failure at longer ISI.

Results of this study also revealed that all three groups of subjects obtained higher discrimination scores for closed syllables than for open syllables. While this finding was expected based on acoustic measurements, it represented a new finding in cross-language tone perception. It should be emphasized, however, that this may not have been an effect of syllable-type *per se*, but rather of acoustic salience. As already mentioned (see Stimuli section) the mid and the low tones in open syllables differed only in their F_0 onset, both F_0 onset and offset were significantly different in closed syllables.

In conclusion, results of this study suggest that (a) experience plays a role in tonal discrimination, and (b) ISI affects the discrimination of the most difficult contrast only. These results should be further investigated using other types of discrimination tasks (i.e. AXB, AX), with other tone languages, and extended to other tone contrasts. Future research should also focus on the role of short-term memory, the role of L1 background, the degree of acoustic salience among tones, and the types of acoustic cues used in cross-language tone perception.

Notes

The authors would like to thank J.J. Clark from the University of Oregon for assistance in running subjects and Professor John Hartmann from Northern Illinois University at DeKalb, for his assistance in finding some references cited in the study.

- ¹. If the proportion of hits (H) equaled the proportion of false alarms (FA), then A' was set to 0.5. If H exceeded FA, then $A' = 0.5 + ((H-FA)*(1+H-FA))/((4*H)*(1-FA))$. However, if FA exceeded H, then $A' = 0.5 - ((FA-H)*(1+FA-H))/((4*FA)*(1-H))$.

References

- Best, C and Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 305-350.
- Burnham, D. & Francis, E. (1997). The role of Linguistic Experience in the Perception of Thai Tones. In A. S. Abramson (Ed.), *Southeast Asian linguistic studies in honour of Vichin Panupong*. (pp.29-46). Bangkok: Chulalongkorn University Press.
- Flege, J., Takagi, N. & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /ɪ/ and /I/, *Journal of the Acoustical Society of America*, 99, 1161-1173.
- Snodgrass, J., Levy-Berger, G., and Haydon, M. (1985). *Human Experimental Psychology*. New York: Oxford University Press.
- Yamada, R. & Tokhura, Y. (1991). Age affect on acquisition of non-native phonemes: perception of English /r/ and /l/ by adult Japanese learners of English. In *Proceedings of the XIIIth international congress of phonetics sciences, August 19-24, 1991, Aix-en-provence, France*.

TONOGENESIS IN KHMER

Ratree Wayland
University of Florida
<ratree@ufl.edu>

Susan Guion
University of Oregon
<guion@uoregon.edu>

0 Introduction

Physiological constraints of the articulatory and/or auditory mechanisms have been proposed as the source of sound changes for some time (e.g., Hombert 1977 and references therein, Ohala 1971, 1974, 1981b, 1989, 1993). As pointed out by Hombert (1977), these explanations imply that the speaker's pronunciation may not be perceived as intended. The distortion may occur due to articulatory and auditory constraints, which affect the way the sounds are produced and perceived by listeners. Ohala (1993) reviews many sound changes and proposes a typology of sound change in which variation in speech due to coarticulation (among other things) can fail to be corrected by the listener (hypo-correction) or corrections can be erroneously applied (hyper-correction). As coarticulations or reductions are greater in faster speech (e.g., Guion 1998, Moon, Lindblom & Lame 1995, Lindblom 1990), one might expect more hypo-correction in fast speech forms. Preliminary evidence for such a proposal was given in Guion (1998) in which faster speech forms were found to be more similar acoustically to post sound change forms than citation speech forms.

The goal of this paper is to report on tone development in Khmer (Cambodian). Based on previous research, it will be proposed that this development is largely phonetically motivated. Moreover, we propose that the sound change has its origins in faster, colloquial speech.

1 Tone Development in Khmer

Recently, Thach (1999) has reported a sound change in dialects of Khmer spoken in Vietnam in which consonant + [r] clusters in onset position of main syllables lose the [r] and gain a falling tone on the following vowel (e.g., [krɔː] > [kɔ̀ː] 'poor'). The sound change is quite advanced in these dialects, especially among younger speakers. Through this sound change, tone has been introduced and there are now minimal pairs such as [kɔ̀ː] 'poor' (from [krɔː]) and [krɔː] 'neck'. Additionally, Thach (1999) reports that the trill [r] has become a glottal fricative [h] in syllable initial position in main syllables. Informal observation of the Khmer spoken in Cambodia suggests that a similar sound change is happening there as well.

The sound change in Khmer spoken in Cambodian to be investigated here involves monosyllabic words¹ with a consonant cluster onset. The first consonant in the cluster is an obstruent and the second member is an alveolar trill [r]. This sound change occurs only in colloquial pronunciation, but not careful reading pronunciation. In colloquial pronunciation, the alveolar trill [r] in a voiceless obstruent + [r] sequence is dropped or becomes a fricativized voiceless aspirated [r]. In most cases, the initial voiceless unaspirated stop also changes into a voiceless aspirated stop. If the vowel involved is a

Ratree Wayland, John Hartmann & Paul Sidwell, eds. *SEALSXII: papers from the 12th meeting of the Southeast Asian Linguistics Society (2002)*. Canberra, Pacific Linguistics, 2007, pp.145-152.

© Ratree Wayland & Susan Guion

short low front [a] vowel, it becomes the diphthong [ea]. More importantly, a distinct rising pitch contour is also evident in words having undergone the changes.

2 Acoustic Analysis

The goal of this analysis is to determine acoustically whether or not the above mentioned sound changes have indeed occurred. To this end, an acoustic analysis of a subset of words undergoing these changes is undertaken.

2.1 Method

Speaker: A male native speaker of Khmer from Phnom Penh served as the speaker. He is in his 40's and has been living in the United States for approximately 5 years at the time of recording. No known speech or hearing impairment was reported.

Stimuli: Stimuli were 13 Khmer words (see Appendix) elicited from the speaker. Nine out of the thirteen words begin with a voiceless stop consonant, three with a voiceless affricate and one with a voiceless fricative. The speaker was instructed to read the words from the wordlist at a normal speaking rate. For comparison purposes, the speaker was instructed to first say the word as he normally would if he were to read it from a book (spelling or reading pronunciation), and second as he would in a conversation to another Khmer speaker (colloquial pronunciation). Each word was repeated three times in random order.

2.2 Procedure

Recording of the wordlist was digitized on a Kay Elemetrics CSL station (Model 4300B) at a sampling rate of 25 kHz. Each word was edited and stored as a separate file for further analysis using Cool-Edit (www.syntrillium.com). Subsequent acoustic analyses were performed using Kay Elemetrics MultiSpeech. Acoustic measurements included voice-onset-time (VOT) for words with stop initials, frication duration for words with affricate initials, vowel fundamental frequency (F_0) and frequencies of the first and second formants (F_1 , F_2).

2.3 Results

VOT: Mean VOT duration of initial stop consonants for both modes of pronunciation is shown in Table 1. With the exception of [krɔː] 'poor', VOT duration of initial stops in colloquial speech was generally longer than that of the reading pronunciation (mean = 44 vs. 27 ms). Result of a paired t-test supported this observation ($t = 2.06$, $p < .0001$, $df = 26$). This finding confirmed that, in most cases, a voiceless stop [p, t, k] initials in reading pronunciation became aspirated in colloquial pronunciation.

Table 1: Mean VOT duration (in ms) for stop initials in reading and colloquial modes of pronunciation.

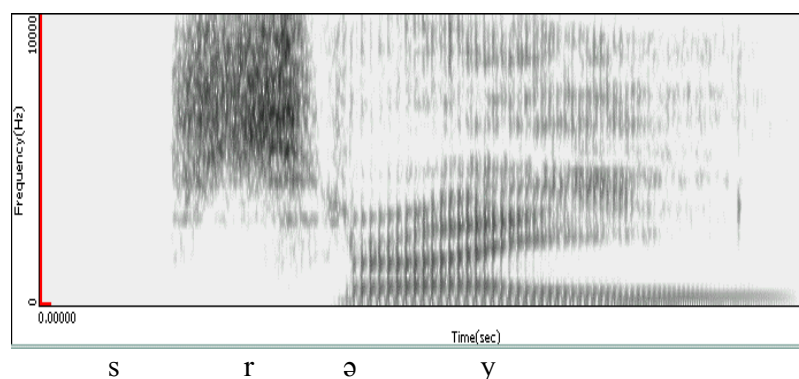
Word	Colloquial	Reading
[pram] ‘five’	39	13
[praə] ‘use’	31	15
[prap] ‘tell’	34	13
[pruəy] ‘sad’	42	33
[priəy] ‘spirit’	34	21
[triw] ‘correct’	52	45
[kru] ‘teacher’	51	36
[krɔː] ‘poor’	50	51
[krɔ̃əŋ] ‘enough’	66	21
Mean	44	27

Fricative Noise Duration: Mean fricative noise duration after the burst of a voiceless affricate initial is shown in Table 2. As expected, for all three words, fricative noise duration was longer for colloquial than for reading pronunciation (mean = 82 vs. 66 ms.). This difference was significant in a paired t-test ($t = 2.31, p < .01, df = 8$). This finding suggested that a voiceless affricate [tɕ] in reading pronunciation became a voiceless aspirated affricate [tɕ^h] in colloquial speech.

Table 2: Mean fricative duration (in ms) of voiceless affricate [tɕ] in reading and colloquial pronunciation.

Word	Colloquial	Reading
[tɕrout] ‘harvest’	82	69
[tɕriw] ‘deep’	80	62
[tɕrɔ̃əŋ] ‘muddy’	84	67
Mean	82	66

As for [s] in [srəy] ‘female’, an auditory as well as a spectrographic examination (Figure 1 and 2) suggested that, in most cases, it became a voiceless aspirated affricate [tɕ^h].

**Figure 1:** Spectrogram of [srəy] ‘female’ in reading pronunciation.

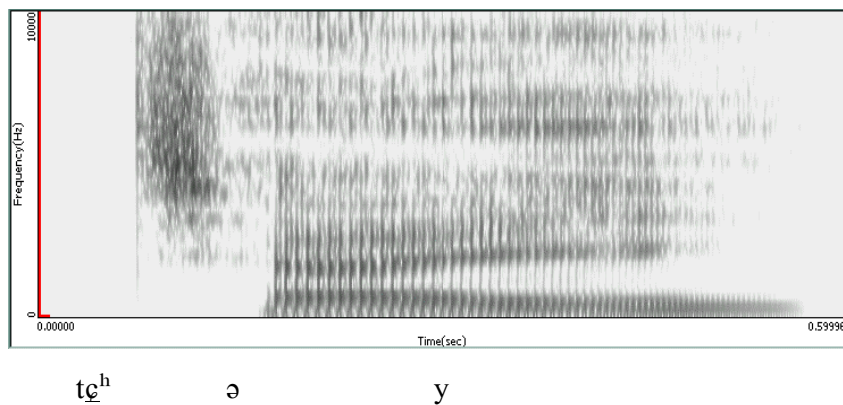


Figure 2: Spectrogram of [srəy] 'female' in colloquial pronunciation.

It should be mentioned, however, that the replacement of the alveolar trill [r] with aspiration was not always complete. In some cases, it became a fricativized voiceless aspirated [r].

F₀ Contour: F₀ contours (in Mels) for colloquial and reading pronunciation averaged across all words is illustrated in Figure 3. In general, F₀ values at vowel onset (182 vs. 181 mels) and at vowel offset (244 vs. 249 mels) are comparable in both types of pronunciation. However, while F₀ continues to rise in reading pronunciation, there is a sharp decrease in F₀ value at 50% in the vowel for colloquial speech pronunciation ($t = 2.18$, $p < .006$, $df = 12$). Thus, while rising pitch contour is the characteristics of reading pronunciation, colloquial pronunciation is best characterized with a falling-rising contour.

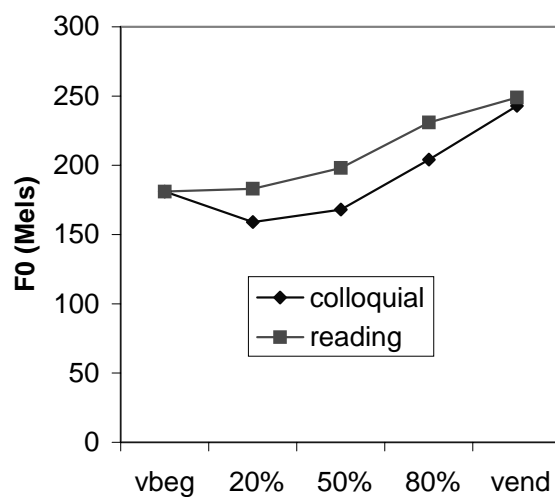


Figure 3: Normalized F₀ contours for reading and colloquial pronunciations.

A Vowel Quality Change: The change from [a] to [ea] is evident in the spectrograms derived from the two different pronunciations of [prap] 'to tell' included in Figure 4 and 5. The locations and movements of the first two formants of the vowels in the two spectrograms clearly showed a change from a monophthong [a] to a diphthong [ea].

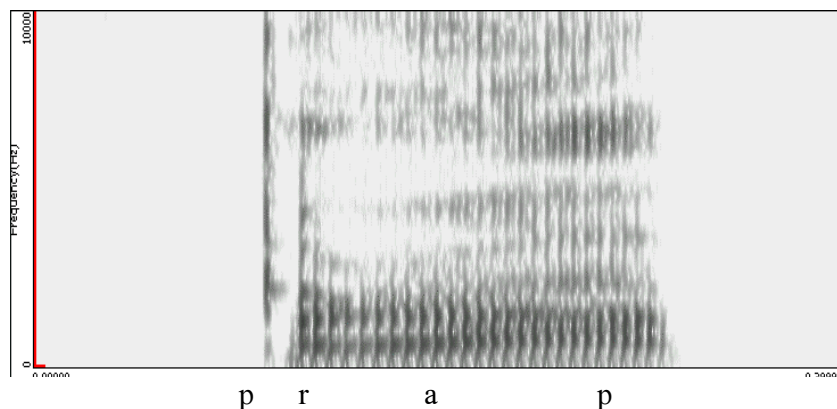


Figure 4: Spectrogram of [prap] ‘tell’ in reading pronunciation.

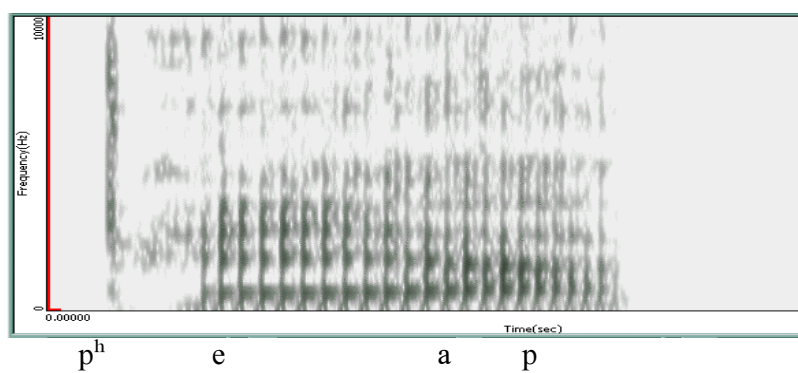


Figure 5: Spectrogram of [prap] ‘tell’ in colloquial pronunciation.

This observation is supported by acoustic measurements of the formant frequencies of the two vowels in all three tokens of [pram] ‘five’ and [prap] ‘tell’. The first formant (F1) and second formant (F2) values of the vowel in [pram] ‘five’ in the reading pronunciation is 832 Hz and 1364 Hz, and 766 Hz and 1248 Hz for [prap] ‘tell’ respectively.

In colloquial pronunciation, mean F1 and F2 values of the first element of the diphthong in [pram] ‘five’ are 572 Hz. and 1816 Hz, while those of the second element are 861 Hz. and 1382 Hz. Those of [prap] ‘tell’ are 504 Hz, and 1819 Hz vs. 706 Hz and 1586 Hz. respectively. Thus, it is evident that the [pram] ‘five’ and [prap] ‘tell’ in reading pronunciation were pronounced as [p^heam] and [p^heap] in colloquial speech.

In summary, the following sound changes occurred when words beginning with an obstruent + r sequence are produced in colloquial style. (1) The alveolar trill [r] was either dropped or became a fricativized voiceless aspirated [r]. (2) In most cases, voiceless stop [p, t, k] becomes voiceless aspirated stops [p^h, t^h, k^h]. (3) Voiceless affricate [tʃ] and voiceless fricative [s] become voiceless aspirated affricate [tʃ^h]. (4) Low vowel [a] becomes the diphthong [ea]. (5) The whole syllable is now produced with a falling-rising pitch contour. Note that this finding is somewhat different from the Khmer dialect spoken in Vietnam reported in Thach (1999). Perhaps the falling tone reported for this dialect is a description of the falling part of the contour, or perhaps the tone development is different

in the two dialects of Khmer. An acoustic investigation of the Khmer dialect spoken in Vietnam would need to be undertaken to decide the matter.

In the next section, we turn to a discussion on plausible phonetic explanations for these sound changes in Khmer.

3 A Phonetic Account

We hypothesize that all the changes are induced by a relatively faster speaking rate in colloquial speech in comparison to a more careful reading pronunciation. First, we propose that the alveolar trill [r] becomes a voiceless [r̥] and assimilates its voicing feature with the preceding voiceless obstruents. This may be due to the temporal overlap or carry-over of the glottal gestures of the initial voiceless obstruent into the following [r]. That is, the vocal folds remain open throughout the initial cluster and do not adduct to begin voicing until the onset of the following vowel. It is also possible that, for aerodynamic reasons, voicing during a trill [r] is difficult to achieve. A relatively strong and high volume of airflow is needed to set the tongue tip into the trill motion. However, this can only be achieved, especially in a short period of time, if the glottis is relatively wide open. In turn, the wide-open glottis makes voicing difficult. Either of these potential explanations could have a plausible origin in faster, colloquial speech if a view of speech production is taken in which energy expenditure on the speaker's part tends toward economy to the extent allowed by the speaking conditions (see e.g., the Hypo and Hyper-Speech theory of Lindblom 1990). On this view, the costly processes of laryngeal timing or aerodynamic adjustments are not reliably used in casual speech. Presumably, in casual speech, there is more signal redundant information allowing the lower cost form of behavior while still permitting sufficient information transfer for effective communication.

In the next step in the sound change, the voicelessness of the [r] is heard as aspiration and becomes perceptually associated (phonologized) with the initial obstruent consonants. This process accounts for the change from voiceless unaspirated stops and affricates to voiceless aspirated stops and affricates. When producing aspiration², it is advantageous for the size of the oral constriction aperture of the initial consonants to be decreased. The smaller or complete closure of the oral cavity will result in an increase in oral pressure and thus, a delay in obtaining a transglottal pressure drop suitable for voicing. This mechanism results in a change from an alveolar fricative [s] to a voiceless aspirated affricate [tʃ^h].

Another aerodynamic measure adopted by Khmer to insure a desired degree of aspiration is the tongue raising gesture at the onset of the vowel, leaving only a small oral constriction aperture, thereby delaying the VOT. This explains why the low short vowel [a] becomes a diphthong [ea], as in [p^heam] 'five' and [p^heap] 'tell'. This mechanism appears to be needed only when a low vowel is involved. A strong degree of aspiration achieved through these aerodynamic mechanisms may also be responsible for a slight degree of perceived breathiness in these syllables

A change in F₀ pattern in colloquial pronunciation may also be attributed to the aerodynamic factor. As shown in Figure 3, there is a significant drop in F₀ value from the beginning of the vowel to vowel mid-point resulting in a falling-rising pitch contour for syllables produced with this style of pronunciation. The decrease in F₀ may be due to the cessation or reduced rate of airflow when voicing begins after aspiration. That is, the reduction of high volume airflow results in a weak Bernoulli effect, thus a slower rate of

vocal fold vibration. This effect was not found in reading pronunciation where initial obstruents are not aspirated.

In conclusion, all sound changes in Khmer including the aspiration of initial obstruent consonants, a change from a voiceless fricative [s] to a voiceless aspirated affricate [tʃ^h], and a change in vowel quality from [a] to [ea] may be attributed to an aerodynamic effect. All of these changes occur as a result of aerodynamic mechanisms adopted to secure the desired degree of aspiration in colloquial pronunciation once the aspiration, which has its origin from the voiceless trill [r̥] is phonologized and becomes part of the initial consonant. A reduction in high volume airflow after aspiration cause the F0 to drop once voicing begins leading to a distinct falling-rising pitch contour.

Notes

The authors would like to thank Sopheak Son for the stimuli used in the study. Many thanks also go to Mohamed Al-Khairy for obtaining some of the references cited.

1. Even though the changes also occurred in second syllables of bi-syllabic words such as [ba:rəy] ‘cigarette’, [bɔŋriən] ‘teach’ etc., they will not be included here. Bi-syllabic words often become sesqui-syllabic words e.g., [bərəy] or [pərəy] ‘cigarette’ or [pəriən] ‘teach’ in colloquial pronunciation. The main (second) syllable which begins with an alveolar trill [r] will undergo the same changes to be described here. The phonetic explanation proposed here should be applicable to those cases as well.
2. There is evidence that the aspiration is further strengthened and becomes [h] in bi-syllabic words. For example, [bɔŋriən] ‘to teach’ is also heard as [pəhiən] (Pisitpanporn, 1999), or in monosyllabic words as in [riən] > [hiən] ‘to learn’ (Thach, 1999).

References

- Guion, S. G. (1998). The role of perception in the sound change of velar palatalization. *Phonetica* 55, 18-52.
- Hombert, J. M. (1977). Development of tones from vowel height? *Journal of Phonetics* 5, 9-16.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory in Mardcastly, W. & Marchal, A., (eds.), *Speech production and speech modeling*, 403-439. Dordrecht: Kluwer.
- Moon, J-S., Lindblom, B., & Lame, J. (1995). A perceptual study of reduced vowels in clear and casual speech. In Elenius, K., & Branderud, P., (eds.), 670-677. Proceedings ICPhS 95, vol 2., Stockholm.
- Ohala, J. J. & Ewan, W. G. (1973). Speed of pitch change. *Journal of the Acoustical Society of America* 53, 345.
- Ohala, J. J. (1974). Phonetic explanation in phonology. In Bruck, Fox, LaGaly (eds.), *Papers from the parasession on natural phonology*, 251-74, Chicago: Chicago Linguistic Society.
- _____. (1971). The role of physiological and acoustic models in explaining the direction of sound change. *Project on linguistic analysis: Reports, Second Series*, no 15.

- Phonology Laboratory—Department of Linguistics, University of California, Berkeley
- _____. (1975). Conditions for vowel devoicing and frication. *Journal of the Acoustical Society of America* 58, S39.
- _____. (1976). A model of speech aerodynamics. *Report of the Phonology Laboratory* (Berkeley) 1, 93-107.
- _____. (1978). Production of tone. In Fromkin, V. (ed.), *Tone*, 5-32. United Kingdom: Academic Press.
- _____. (1981a). Articulatory constraints on the cognitive representation of speech. In Myer, T., Laver, J. and Anderson, J. (eds.), 111-122. *The cognitive representation of speech*. North-Holland Publishing Company: Amsterdam, New York, Oxford.
- _____. (1981b). The listener as the source of sound change. In Masek, Hendrick, Miller, (eds.), 178-203. *Papers from the parasession on language and behavior*, Chicago Linguistic Society: Chicago.
- _____. (1989). Sound Change is drawn from a pool of synchronic variation. In Breivik, Jahr, (ed.), 171-198. *Language change: contributions to the study of its causes*, NY: Mouton de Gruyter.
- _____. (1993). The phonetics of sound change; in Jones, D. (ed.), *Historical linguistics: problems and perspectives*, 237-278, London/NY: Longman.
- Thach, N. M. (1999). Monosyllabization in Kiengiang Khmer. *Mon-Khmer Studies* 29, 81-95.

Appendix: Wordlist

[pram]	“five”
[praə]	“to use”
[prap]	“to tell”
[pruəy]	“sad”
[priəy]	“spirit”
[triw]	“correct”
[kru]	“teacher”
[krɔ:]	“poor”
[krɔ̃ən]	“enough”
[srəy]	“female”
[tɕrout]	“to harvest (rice)”
[tɕriw]	“deep”
[tɕrɔ̃əm]	“muddy”

A COMPARATIVE STUDY OF RICE CULTURE WORDS IN THE GE-YANG AND KAM-TAI LANGUAGES

Zhou Guoyan

*Central University for Nationalities,
Beijing, China*

1 Introduction

Ecological linguistics is a new branch of linguistic science that has just sprung up in recent years. It studies the connection between human languages, the natural ecological environment, and the socio-cultural environment. Ethnic groups (or speakers of certain languages) living in different natural geographical environments may have different numbers of words expressing the same natural phenomena. For instance, people, such as the Eskimos who live near the North Pole have different words for “snow” with different shapes, different colors, and those that fall in different periods of time within the day. The Mongolians, the Kazaks are traditional nomadic ethnic groups. In their languages, cattle of different sex, different age and even female cattle having given birth to young cattle or not are named with different word forms. Ethnic groups who cultivate rice often have plenty of words or expressions for every detail of the process of rice planting and related cultural activities. The Kam-Tai speakers are traditional rice cultivators having thousands of years of rice planting history. Today, most of them still take up rice planting as their means of livelihood. Ge-Yang comprises a group of languages that are related to Kam-Tai languages. The Ge-Yang people have a rather close historical relationship with speakers of Kam-Tai languages. They all are distributed throughout the same geographical area and belong to the same economy-culture type. This paper attempts to make a comparison between the rice cultivation words between these two groups and investigates the similarities and differences between their respective rice cultures.

2 The Relationship Between the Ge-Yang and the Kam-Tai Groups

Ge-Yang is a new term that has appeared in the works of some Chinese scholars ever since the 1990s. Prof. Liang Min created the term for the first time in one of his papers discussing the matter of which group do such languages as Gelao, Mulao, Lachi, Pubiao, Yuren, Buyang, etc. belong. Some Western scholars called them “Kadai” languages or “Tai-Kadai” languages as part of the Kam-Tai group. As Ge-Yang languages are closely connected to the Kam-Tai group, they are therefore also mentioned as “Outlier Kam-Tai” languages in the works or papers of some Western scholars.

Kam-Tai is a well-known group of languages that includes the Thai, Lao, Shan in Southeast Asia. They have large numbers of speakers and include many other small languages. In China, Kam-Tai includes three branches, that is, the Tai branch (called Zhuang-Dai by Chinese scholars) consisting of Zhuang, Buyi, and Dai, the Kam-Sui branch (called Dong-Shui by Chinese scholars), consisting of Kam (also called Dong), Sui, Muolao, Maonan, Lajia, etc., and the Hlai branch which includes only the Hlai language. The data of this paper are mainly collected from the Kam-Tai languages in China.

The Ge-Yang speakers and the Kam-Tai peoples of mainland China are mostly distributed mainly in southern provinces such as Guangxi, Guizhou, Hunan, Yunnan, Hainan and Guangdong. Most of the Ge-Yang language speakers live in the central and the western part of Guizhou, and only a few of them live in scattered places in Yunnan and Guangxi. Each language has only a small number of speakers. The Gelao, which is the biggest in the group, has about five to six thousand speakers, while other languages have only several hundred or even fewer than a hundred. All of them are seriously endangered languages. The Yuren language of this group became extinct in the late 1980s. Some of the data in this paper are cited from the works of such scholars as Zhang Jimin, Liang Min, and Jerry Edmondson, who have been engaged in the study of the Ge-Yang group for a long time. Most of the data were collected in fieldwork done by myself and my colleague, Li Jinfang, along with some of his students.

The ancestors of the Kam-Tai peoples can be traced back to the Yue peoples of more than 2000 years ago. At the historical stage marked by from the Spring and Autumn Period (722-481 B.C.) up to the Warring States Period (481-221 B.C), the Yue people were distributed all over the areas of southern China. The Yue people lived in the area stretching from the southeastern coast to the northern part of Vietnam. The Yue people had so many branches that they were usually called Baiyue (Hundred Yue) in Han Chinese historical documents. The Kam-Tai peoples today mainly originate from the Yue who were distributed throughout Guangdong, Guangxi, and the southeastern part of the Yunnan and the Guizhou plateau. They were referred to as Luoyue, Xi'ou and Nanyue respectively in historical books. Among Chinese scholars there is a viewpoint claiming that the Kam-Tai peoples (mainly the Tai group) distributed in Southeast Asia and the western part of Yunnan were formed by emigrants from Guangxi and the eastern part of Yunnan about 1500 years ago. The ancestors of the Ge-Yang speakers can be traced back to the Pu people of more than 3000 years ago. Pu was also a group of ancient peoples that were widely distributed throughout southern China and had many branches that were called Baipu (The Hundred Pu) in Han Chinese historical documents. During the Spring and Autumn period and the Warring Period, the Yunnan and Guizhou plateau was an area inhabited by both the Yue and the Pu peoples. The latter were mainly distributed in Guizhou. At the end of the 3rd century B.C, the Qin empire took over Nanyue and Xiyue (both were branches of the Yue peoples) by armed force. Many Yue people were forced to move northward into Guizhou to live together with the Pu people there. In the course of long-term contact and intercommunication, Pu and Yue were gradually mixed, and their cultural features became similar. For instance, the Ganlan style of dwelling house of the ancient Yue people was not only inherited by their descendents (i.e., the Zhuang, the Sui, the Kam and many Tai groups of people in Southeast Asia), but also by some of the Gelao people who originated from the ancient Pu. The custom of knocking out the canine teeth can be found among the Zhuang, the Kam, and the Gelao, It is hard to say from which ancient group of people the custom came. As to the ancient language, there can be found many records of the language of the ancient Yue in Chinese historical books. For instance, in a book named *Shuoyan* (literally translated as "The World of Speech") written in the Han Dynasty (206 B.C.-A.D. 200), a folk song of the Yue people was taken down in Han Chinese characters. It is found that the language structure of the folk song is quite similar to that of the modern Kam-Tai languages. Because of the lack of historical records, people know very little about the language of the ancient Pu. From the differentiation seen

between the language structure of modern Ge-Yang and Kam-Tai before the contact of these two group, these languages should be greatly different, but for hundreds of years of intercommunication and assimilation since the end of the Warring Period, these two groups had not only become similar in culture, but they also had many common linguistic features, and these common features are mainly shown in their vocabulary. Professor Liang Min from the Chinese Social Academy of Science made a comparison between the Ge-Yang and the Kam-Tai languages with Swadesh's 200 word list, and found that the highest ratio of cognates between these two groups are the cognates between the Zhuang language and the Buyang language (32.34%), a little bit higher than the cognate ratio among the languages in the Ge-yang group itself. The cognate ratio between the Buyang language and Gelao in Anshun of Guizhou Province is 31.41%. (Liang 1990) What is worth noticing is that the cognate ratio between the Anshun Gelao language and the Hlai language, and between the Sanchong Gelao language and the Hlai, the speakers of which do not live in the same areas, are both higher than that between the Zhuang and the Kam languages. A reasonable explanation of this phenomena is that among the Yue people who immigrated northward, there were some people who had close relationships with the Hlai people who are now living in Hainan island and had merged into the Gelao group.

In a word, the vocabulary similarities between the Ge-Yang and the Kam-Tai groups are the result of language contact, cultural intercommunication and assimilation.

The Kam-Tai peoples are traditional rice cultivators. Nowadays, many scholars believe that the Kam-Tai ancestors, i.e. the Yue peoples, were the creators and disseminators of rice culture in Asia. Today, most of the Kam-Tai peoples still live in flat areas with rich water resources that make it easy for rice to grow, whereas most of the Ge-Yang people live in high mountain areas. Because of the geographical and climatic condition, they mainly plant such crops as corn, wheat, potatoes, etc., which are cold-resistant and drought-enduring. At the same time, they also plant some rice. A small number of them who inhabit the lower flat areas also plant rice as their main crop. The current living status of the Ge-Yang people is the result of historical movements. At first, most of them also lived in flat areas and were engaged in rice planting, As the Han Chinese moved in and formed settlements with large populations, the Ge-Yang peoples were forced to move into the mountains. Today, the Gelao people (the main part of the Ge-Yang speakers) in most places still consider themselves the oldest ethnic group in Guizhou, because wet rice fields and dry lands there were opened up by their ancestors. Every year in July or August, when the crops are ripe, the Gelao people will go the rice fields to cut the spikes, bring them home, and taste the new crop. They do not only cut their own rice, but can also cut the rice which belongs to other families or even other ethnic groups, and people won't say anything about this.

A change in the living environment must bring about a change of means of living. Because of the moving from low flat land to high mountains, the economic structure of the Ge-Yang ancestors changed from rice cultivating to highland planting and hunting. They usually replenish the shortage of food supply by seasonal hunting because of the impoverished soil. In most of the Ge-Yang languages, we can see that rice culture vocabulary is quite different from those in the Kam-Tai languages, but from the small number of cognates, we can still see the influence of the Yue rice culture on its neighboring ethnic groups.

Notes

Many thanks are given here to the Henry R. Luce Foundation for its generous support of our fieldwork, research, publication, and my personal travel and short-term residence at Northern Illinois University in the U.S.

References

Liang Min. 1990. The Problem of Genealogical Affiliation of the Geyang Language Group.
Minzu Yuwen 6. pp. 1-8.