



**Pacific Linguistics**

---

**Electronic Publication E-2**

# **SEALS VIII**

**Papers from the 8<sup>th</sup>  
annual meeting of the  
Southeast Asian  
Linguistics Society  
(1998)**

**edited by  
Mark Alves,  
Paul Sidwell  
and David Gil**

**Pacific Linguistics  
Research School of Pacific and Asian Studies**



**The Australian National University**

## Pacific Linguistics E-2

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 McConville, *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*  
Melenaita Taumoefolau, *University of Auckland*  
Tasaku Tsunoda, *University of Tokyo*  
John Wolff, *Cornell University*  
Elizabeth Zeitoun, *Academica Sinica*

# **SEALS VIII**

## **Papers from the 8<sup>th</sup> annual meeting of the Southeast Asian Linguistics Society 1998**

edited by  
Mark Alves,  
Paul Sidwell  
and David Gil



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

## Pacific Linguistics E-2

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 (8th : 1998

Kuala Lumpur).

SEALS VIII : papers from the 8th annual meeting of the  
Southeast Asian Linguistics Society.

Bibliography.  
ISBN 9780858835757.

1. Asia, Southeastern - Languages - Congresses. I. Alves,  
Mark J. II. Sidwell, Paul J. III. Gil, David. IV.  
Australian National University. Research School of Pacific  
and Asian Studies. Pacific Linguistics. V. Title. VI.  
Title : Southeast Asian Linguistics Society 8th.

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	ix
Note	xi
Notes on Thanh-Chuong Vietnamese in Nghê-an Province <i>Alves, Mark J. and Nguyêñ Duy Huong</i>	1
Nominal auxiliaries in Lai <i>Bedell, George</i>	11
Lexical decomposition and locative predicates in Bonggi <i>Boutin, Michael E.</i>	25
Fusion and alignment: the case of Malay <i>Delikhan, Ann</i>	45
Reduplication In The M'nong Language <i>Dinh Le Thu</i>	57
Prosody and the segmentation of Malay discourse <i>Don, Zuraidah Mohd</i>	65
Toward reconstruction of demonstratives in Proto-Austronesian <i>Finney, Joseph C.</i>	79
Grammaticalization of deverbal markers: toward a cross-linguistic study in the semantic extension of motion verbs <i>Kessakul, Ruetaivan and Toshio Ohori</i>	93
Morphological structure of Malay: using psycholinguistic analyses of rated familiarity <i>Lee, Lay Choo, Susan J. Rickard Liow and May-Ling Olivia Wee</i>	109
The direction of monosyllabicity in Raglai <i>Nguyen Van Hue</i>	123
Topic selection and organization and discourse strategies in Malay news in four countries <i>Poedjosoedarmo, Gloria R.</i>	125
Wh-interrogatives in the Malay classical text of Sulalatus Salatin <i>Razak, Rogayah A.</i>	135

Malay lexicalized items in Penang Peranakan Hokkien <i>Lim Beng Soon and Teoh Boon Seong</i>	149
A unified analysis of some Vietnamese reduplication forms <i>Vu, Sonny X.</i>	165
The emergence of 'give' passives in East and Southeast Asian languages <i>Yap, Foong Ha and Shoichi Iwasaki</i>	193

**SEALS papers not appearing in this volume:**

The syllabification of high vowels in Malay: a constraint-based analysis  
*Ahmad, Zaharani*

The acquisition of long distance *wh*-questions in Singaporean Malay  
*Aman, Norhaida and Gabriella Hermon*

The phonology and syntax of Mizo language  
*Chhangte, S.L.*

Are there 'precategorial' morphemes in Austronesian languages?  
*Clynes, Adrian*

Malayic variants in Southwestern Borneo: Gerai and Sepotong  
*Collins, Jim*

Aslian and the other branches of Mon-Khmer  
*Diffloth, Gerard*

Is Chakma language facing extinction?  
*Ganguly, S.*

Kuala Lumpur Malay as a Mainland Southeast Asian language  
*Gil, David*

Tai and Vietnamese irrigated rice technology and culture: a preliminary comparative linguistics study of borrowing  
*Hartmann, John*

Taboo in Malay: a relevance theoretic approach  
*Jalaluddin, Nor Hashimah*

Concordances in the study of English and Philippine languages  
*McFarland, Curtis D.*

The ‘softeners’ in some modal structures in Vietnamese when teaching overseas students  
*Nguyễn Thín Nam*

Problems in Tagalog morphological categorization: the Tagalog derivational clitic  
*Rubino, Carl*  
 (published as “The Tagalog Derivational Clitic” in *Linguistics* 36.6:1147-1166 1998)

LSP and the need for monolingual dictionaries  
*Schaarschmidt, Gunter*

On valence relations in Malay verbal prefixation  
*Sew, Jyh Wee*

The accessibility hierarchy revisited: the syntax and semantics of resumptive pronouns in Malay  
*Schindler, Patrick A.*

The development of contrastive accent in Makassarese  
*Tadmor, Uri*

Verb serialization in Ambonese Malay  
*Tjia, Johnny*

A comparison of some Old Thai orthographies in West Nghê An, Vietnam  
*Trần Trí Dõi*

Word stress in Malay  
*Yong, Janet Y.*

Evolution of the scale of notation in the Austronesian languages  
*Zarbaliyev, H. M.*



# 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 Southeast 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 Eighth Annual Meeting and publication of the proceeding***

The Eighth Annual Meeting of the Southeast Asian Linguistics Society was held at Brisdale Value Inn in Kuala Lumpur, Malaysia, on 20-22 July 1998. It was organised by David Gil with the assistance of the Department of Audiology and Speech Sciences at the Universiti Kebangsaan Malaysia.

The publication of the proceedings was unfortunately delayed until this year (2007), creating some difficulties in the production of this volume. In the case of several papers we were unable to contact the authors and we took the decision to go ahead with publication on the basis of those authors' submissions as we held them. Our working assumption is that even a delayed and potentially imperfect publication is more useful than none at all, and we beg the patience and understanding of contributors and readers in this matter.

Mark Alves  
Paul Sidwell  
David Gil

## Note

---

Starting with SEALS VIII, XII & XIII, *Pacific Linguistics* (with generous support of the *Centre for Research in Computational Linguistics*) will become 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 distributed free of charge electronically, in addition to the print version being available for purchase. 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



# NOTES ON THANH-CHƯƠNG VIETNAMESE IN NGHỆ-AN PROVINCE<sup>1</sup>

Mark J. Alves

*Montgomery College, Maryland*

*<Mark.Alves@montgomerycollege.edu>*

Nguyễn Duy Hương

*Institute of Linguistics, Hanoi*

## 1 Introduction

Vietnamese dialectal variation in terms of pronunciation is generally divided into three main regions, namely Northern, Central, and Southern Vietnamese (NV, CV, and SV hereafter).<sup>2</sup> Typically, each region is characterized by the varieties<sup>3</sup> of Vietnamese spoken in major cities of those regions, Hà-Nội (considered in Vietnam to be ‘Official Vietnamese,’ OV hereafter), Huế, and Hồ-Chí-Minh respectively, though Vietnamese in Vinh is also sometimes recognized as a Central variety distinct from that of Huế. However, few dialect studies on Vietnamese have concentrated on the numerous tiny and isolated villages scattered throughout Vietnam. Reports of their conservative characteristics (cf. Châu 1989, Nguyễn T. C. 1995) suggest that some of these geographically remote varieties

---

<sup>1</sup> We wish to thank all the consultants who participated in this study. The data was collected in September of 1997. One of the researchers in this study, Mr. Nguyen, is a native of that region and was crucial in eliciting the local vocabulary. Speech samples were recorded using the Summer Institute of Linguistic’s ‘Cecil’ acoustic phonetic software. To elicit the data, three word-lists were used, including one-syllable words (46 words), two-syllable words (57 words), and complete sentences (15 sentences). The word-lists were chosen primarily for tonal contrasts, though substantial segmental variety was also included. The sentence list--which contained simple, daily sentences--was created based on known regional vocabulary. Most of the recorded data came from two native speakers of TC Vietnamese. Recordings made with the help of speakers from other regions other than Nghệ-An were representative of Official Northern and Southern Vietnamese, which allowed a point of reference for comparison.

We recognized the hazard of using wordlists to elicit natural speech, so we employed a few devices to reduce this interference as much as possible. First, we persuaded our consultants to use their most natural ‘family’ pronunciation, which most readily admitted was more comfortable than reading in OV. Next, Mr. Nguyễn, as a native speaker of TCV with linguistic training and familiarity with dialect differences, helped to identify the use of OV. Finally, the sentence list was actually ‘translated’ by the consultants into their native variety, and was not read by the consultants, which created a point of reference to help screen out reading pronunciations.

<sup>2</sup> See Thompson 1984-85 for a general phonetic description of those regions.

<sup>3</sup> The term ‘variety’ is used throughout most of the paper rather than ‘accent’ or ‘dialect’ in order to avoid the controversy in deciding which category is accurate. Based on the significant phonological and lexical differences between TCV and OV, TCV is more like a ‘dialect’ than a minor regional ‘accent,’ though this is still relative to the region in which it is spoken where the differences may be said to be simply accent. Perhaps the TCV region in Nghệ-An province, including different accents spoken in neighboring areas can, altogether, be considered a ‘dialect’ with varying local ‘accents,’ though further research is needed to clarify the situation.

of Vietnamese are largely untapped storehouses of historical linguistic information about the Vietnamese language.

This paper describes both the modern and historical characteristics of one variety, Thanh Chương Vietnamese (TCV hereafter), a variety of Vietnamese from the dialectally diverse region of North-Central Vietnam.<sup>4</sup> Thanh Chương is a district in the western hills of Nghê-An province, North-Central Vietnam, a short distance from the border of Laos. Field data has shown TCV to be a highly conservative variety of Vietnamese. In the following sections, the archaic nature of TCV is illustrated in terms of Vietnamese phonology and etymologies. After a summary of the phonological characteristics of TCV, those features are compared with Mường, Rục<sup>5</sup> (a Minor Vietic<sup>6</sup> language), and Pacoh (a language of the Katuic branch of Mon-Khmer). Then, the etymological layers of TCV (regional and non-regional Vietnamese, Vietic, and Mon-Khmer) are discussed, with highlights of Mon-Khmer vocabulary not present in modern OV.<sup>7</sup>

## 2 TCV Phoneme System

TCV has maintained all of the segmental distinctions as represented by the national Quốc-Ngữ orthography. This is generally true among varieties of North-Central Vietnamese, possibly making this the regions with the highest degree of conservatism, according to currently available data. The sound systems of Vietnamese in Hồ-Chí-Minh, Hà-Nội, and Huế have all undergone various phonological mergers resulting in smaller phonemic inventories. While TCV has 24 consonants, SV and mid-CV have 22, and NV has only 20. NV has no retroflex consonants, which merged with palatals, and /j/ and /r/ merged with /z/. From Huế to Hồ-Chí-Minh city, [v] and [z] have merged with [j], and dentals have merged with velars in syllable-final position. TCV has retained all of those phonemic distinctions, as shown in Table 1.

Most varieties of Vietnamese have preserved all the vowel phoneme categories seen in the Quốc-Ngữ orthography, though phonetic differences exist.<sup>8</sup> TCV vowels are, as are NV vowels, phonetically close to the indicated orthographic representation, having few allophonic variants. This differs from the phonetic characteristics of SV and mid-CV

<sup>4</sup> Both North and Central Vietnam show a higher degree of dialectal diversity than the much younger South, which had only been inhabited by the Vietnamese after the fall of the Champa Empire in the end of the fifteenth century. Whether the North or Central region contains more diversity is not something that has been investigated to our knowledge, though our studies of the Vietnamese of several villages in Nghê-An suggest that this area is indeed highly diverse and may contain at least a few distinct dialects.

<sup>5</sup> The primary source is Nguyễn V. L. 1993.

<sup>6</sup> The term ‘Vietic’ (Hayes 1992) refers to the group of languages including Vietnamese and Mường, which together form a subgroup of Vietic, and the two dozen or so groups of archaic languages (see Ferlus 1974 and 1975 for a list and subgrouping), which have been called ‘Minor Vietic’ (Alves, 2003).

<sup>7</sup> Tai-Kadai cognates are not discussed as they are neither genetically nor statistically relevant. This paper was written with the assumption that sufficient data has been collected to date in order to firmly place Vietnamese in the Mon-Khmer language group. Gage (1985) claimed to have found only 3% of Vietnamese vocabulary as possibly being cognate with Tai-Kadai forms, while Thomas and Headley (1970) found about 2X% of Mon-Khmer forms.

<sup>8</sup> Again, see Thompson 1965 for a discussion of the differences.

(e.g. vowel centralization in closed syllables, such as /i/ to /ɨ/ and /e/ to /ə/ and monophthongization of diphthongs).

**Table 1: Thanh-ChưƠng Consonants<sup>9</sup>**

	<b>lab</b>	<b>den</b>	<b>pal</b>	<b>ret</b>	<b>vel</b>	<b>glot</b>
<b>voiceless stop</b>	-p	-t-	-c-	t-	k-	
<b>voiced stops</b>	b-	d-			g-	
<b>Voiceless continuants/aspirates</b>	f-	t <sup>h</sup> -	s-	ʂ-	x-	h-
<b>voiced continuants</b>	v-	z-				
<b>nasal</b>	-m-	-n-	-ɳ-		-ɳ-	
<b>glides/liquids</b>	-w-	l-	-j-	ɻ-		

Besides consonants and vowels, TCV tones also differ from other mainstream Vietnamese varieties. Table 2 lists the phonetic realizations of tones in Hà-Nội, Nha-Trang, and Thanh-ChưƠng. Among many varieties of Vietnamese, only five tonal phonemes remain of the six represented orthographically, most often due to a collapsing together of the hỏi and ngã tones. Such is the case for TCV as well as SV and CV. The sá̂c tone does have more than one phonetic realization in TCV and Hà-Nội Vietnamese depending on the syllable final, whether a voiceless stop or a sonorant (vowel or nasal). The exact value of the open-sá̂c tone in TCV is left undetermined in Table 2 because that tone showed two different phonetic forms, a low-level tone (11) and a low-rising glottalized tone (13g).

**Table 2: Tone Systems of Vietnamese as Spoken in Hà-Nội, Nha-Trang, and Thanh-ChưƠng<sup>10</sup>**

<b>TONE</b>	<b>HN</b>	<b>NT</b>	<b>TC</b>
ngang	33	33	35
huyền	32	32	33
sá̂c, open	24	45	11/13g
sá̂c, closed	45		45
nặng	22g	23	22
Hỏi	31	24	31
Ngã	35g		

<sup>9</sup> IPA is used for much of the paper, though the Vietnamese Quốc-Ngữ orthography has been used for ease of reading and reference. Hyphens are used in the table to indicate the range of distribution of the phonemes in syllables, whether strictly in word-initial position, such as the continuant series (e.g. /x-/), word-final position (only /-p/), or both positions, such as the nasal series (e.g. /-ɳ-/).

<sup>10</sup> The Y.R. Chao system is being used to represent tone throughout this paper. The number 5 represents the highest pitch level, and 1 is the lowest. The first number represents the starting point and the second, the end point. In this paper, the letter 'g' indicates that the tone is clearly glottalized.

**Table 3: Correspondences of Standard and TCV**

Gloss	Comparison QN <sup>11</sup> IPA		OV	IPA	TCV	IPA
1. gourd	âu : u	əw : u:	bầu	bəw <sup>21</sup>	bù	bu: <sup>11</sup>
2. bear			gấu	gəw <sup>24</sup>	gụ	gu: <sup>11</sup>
3. deep			sâu	səw <sup>11</sup>	su	su: <sup>11</sup>
4. buffalo			trâu	cəw <sup>11</sup>	tru	tu: <sup>11</sup>
5. road	uro : a	iə : a:	dường	d̥iəŋ <sup>21</sup>	dàng	daŋ <sup>33</sup>
6. person			người	ŋjəj <sup>21</sup>	ngài	ŋa:j <sup>33</sup>
7. grill			nướng	n̥iəŋ <sup>24</sup>	náng	naŋ <sup>11</sup>
8. fire	ura : a		lửa	liə <sup>31</sup>	lả	la: <sup>31</sup>
9. itchy			ngứa	ŋjə <sup>24</sup>	ngá	ŋa: <sup>11</sup>
10. female	ay : ây	aj : əj	cái	ka:j <sup>24</sup>	cáy	kəj <sup>11</sup>
11. urinate			dái	da:j <sup>24</sup>	dây	dəj <sup>11</sup>
12. fruit			trái	ca:j <sup>24</sup>	tráy	təj <sup>11</sup>
13. you	ay : i	aj : i:	mày	maj <sup>21</sup>	mi	mi: <sup>35</sup>
14. this			này	naj <sup>21</sup>	ni	ni: <sup>35</sup>
15. same	ô : u	o: : u:	giống	zəuŋ <sup>m24</sup>	giົm	zu:m <sup>31</sup>
16. knee			gối	go:j <sup>24</sup>	cúi	ku:j <sup>11</sup>
17. I			tôi	to:j <sup>33</sup>	tui	tu:j <sup>35</sup>
18. goat	ê : ia	e: : iə	dê	ze: <sup>33</sup>	dia	jiə <sup>35</sup>
19. son-in-law			rê	ze: <sup>31</sup>	ria	riə <sup>33</sup>
20. pastry	anh : eng	əŋ : εŋ	bánh	baj <sup>24</sup>	béng	bε:ŋ <sup>11</sup>
21. soup			canh	kaj <sup>11</sup>	keng	kε:ŋ <sup>35</sup>
22. fishy			tanh	taj <sup>11</sup>	teng	te:ŋ <sup>35</sup>
23. thatch			tranh	caj <sup>11</sup>	treng	tε:ŋ <sup>35</sup>
24. blue /green			xanh	saj <sup>11</sup>	xeng	xε:ŋ <sup>35</sup>
25. pile	-ông/- ôông	əŋŋ <sup>m</sup> : oŋ	chồng	cəwŋ <sup>m2</sup> 1	nhôông	ŋno:ŋ <sup>35</sup>
26. trunk			gốc	gəuk <sup>p45</sup>	côôc	ko:k <sup>11</sup>
27. to plant			tròng	ʈəŋŋ <sup>m21</sup>	lôông	lo:ŋ <sup>35</sup>

A quick examination of the features of earlier stages of Vietnamese and Vietic<sup>12</sup> shows how conservative TCV phonology is. There are a few general patterns of

<sup>11</sup> Some of the Quốc-Ngữ symbols used in this table are part of a non-standard, vernacular convention used to transcribe differences in regional pronunciations.

<sup>12</sup> For a description of earlier stages of Vietnamese and Vietic, see Nguyễn T.C. 1995.

phonological correspondences in certain lexical items between OV and TCV,<sup>13</sup> which are illustrated in table 3.<sup>14</sup> First, TCV has preserved forms that have not undergone diphthongization (examples 1 to 9, 13, and 14). Next, as mentioned, TCV has preserved initial retroflex sounds, a characteristic of SV and CV (examples 4 and 12). Then, there are two cases of final vowel-consonant assimilation (examples 21 to 25). TCV has preserved a few cases of the earlier Vietic sequence /-ɛŋ/ for which OV has only the palatalized /-æŋ/. Interestingly enough, two of the forms are actually old Chinese loanwords (examples 20 and 24). Also, TCV has not completed the process of rounding final consonants after background vowels (examples 25 to 27), a process completed in OV but not in Muòng dialects (cf. Nguyễn M.D. 1972).

### 3 TCV Etymological Layers

TCV vocabulary has at least five identifiable etymological layers. These include (1) non-regional vocabulary (Vietnamese common to all regions of Vietnam), (2) regional vocabulary generally used in Central Vietnam, (3) local vocabulary used in and around Thanh-ChưƠng district, (4) Vietic vocabulary not seen in OV, and (5) some Mon-Khmer vocabulary not seen in OV. The significance of the latter four categories is that they are in contrast with OV, which shows that vocabulary in TCV's region has historically gone down a somewhat different etymological road. In some cases, the development is very localized with no identifiable external source. In other cases, TCV appears to have preserved genetically higher-level forms (i.e. from Vietic and Mon-Khmer) that were not maintained in mainstream Vietnamese. Finally, some TCV words are archaic forms that may predate Chinese loanwords and help identify an early layer of Chinese loanwords not belonging to the literary Tang-Song era Sino-Vietnamese.

#### 3.1 Non-Regional Vietnamese Vocabulary

For the most part, the TCV lexicon consists of basic and culturally specific words used throughout Vietnam. This pan-Vietnam vocabulary is also multi-layered, including that which is Vietic, Mon-Khmer,<sup>15</sup> Chinese, Tai-Kadai, and purely Vietnamese in origin. As a result, though Vietnamese speakers from outside the North-Central region may have difficulty comprehending the TCV accent, there is overall mutual intelligibility. Furthermore, speakers of TCV are aware of the lexical differences between their local vocabulary that of OV, which is often used on Vietnamese TV, radio, and in schools. Speakers of TCV are even aware of the lexical and phonological differences between their speech and that of nearby neighbors. Regardless of the differences between TCV and OV, all TCV speakers recognize their linguistic and ethnic connections, a fact supported by lexical and phonological evidence.

<sup>13</sup> Maspero 1912 included data on various dialects throughout Vietnam. Some of the characteristics that occurred in our data were still present in Maspero's time.

<sup>14</sup> The acronyms used in the table are as follows. QN (Quốc-Ngữ), IPA (International Phonetic Alphabet), OV (Official Vietnamese), and TCV (Thanh-ChưƠng Vietnamese).

<sup>15</sup> For a good list of the core Mon-Khmer vocabulary in Vietnamese, see Huffman 1977.

### 3.2 Regional Vocabulary: Central Vietnamese

TCV contains numerous vocabulary items that are associated with Central Vietnam. Speakers in both Vinh and Huế either use such forms or are aware that these are lexical items restricted to Central Vietnam. Table 4 contains a few examples.

**Table 4: Central Vocabulary in Thanh-Chương Vietnamese**

Gloss	OV	IPA	TCV	IPA
1. (reciprocal)	nhau	ɲaw <sup>33</sup>	chắc	cak <sup>11</sup>
2. no	không	xauŋ <sup>m33</sup>	nó	nɔ: <sup>31</sup>
3. this	này	naj <sup>21</sup>	ni	ni: <sup>35</sup>
4. where	đâu	dəw <sup>33</sup>	mô	mo: <sup>35</sup>

These forms are not seen with complete consistency throughout the region. The negation word was claimed by some speakers in urban areas to be restricted to rural areas.

### 3.3 Local Vocabulary: Thanh-Chương

Some TCV vocabulary is restricted further to the region in and around Thanh-Chương. A few examples are shown in Table 5.

**Table 5: Local Thanh-Chương Vocabulary**

Gloss	OV	IPA	TCV	IPA
1. we (excl.)	chúng ta	cu:ŋ <sup>m24</sup> ta: <sup>33</sup>	choa	cwa: <sup>35</sup>
2. no need	không bảo	xəuŋ <sup>m33</sup> ba:w <sup>31</sup>	hung nhủ	hung <sup>m35</sup> nu: <sup>31</sup>
3. which	nào	na:w <sup>21</sup>	mò	mo: <sup>33</sup>
4. remain	còn	kɔ:n <sup>21</sup>	nưng	ni:ŋ <sup>35</sup>

More definitive statements about the precise geographic range of the usage of these forms can only be made after more dialect studies are conducted in this region. Tables 4 and 5 are just a taste of the wide-ranging lexical diversity that has been witnessed in Vietnamese dialect studies.

### 3.4 Vietic Vocabulary: Muong and Minor Vietic

TCV contains some lexical forms that appear to be cognate with Muồng and a Minor Vietic language, Rục. Some of those cognates illustrate phonological changes that occurred between Proto-Vietic and modern OV. In Table 6, the change from the cluster \*[tl]<sup>16</sup> to a single retroflex [ʈ] is seen in example 2. Examples 3 and 9 illustrate the change from a final \*[l] to [n] in TCV, which is lost in OV. Example 8 illustrates the rounding of finals after back round vowels.

Examples 2 and 7, and possibly 4 and 5, are examples of lexical preservations in which the original Vietnamese forms have been replaced by Sino-Vietnamese terms. The words ‘head’ and ‘tiger’ are lexical preservations in TCV, while the modern OV form is

<sup>16</sup> These clusters are attested in seventeenth century Vietnamese as seen in Alexandre de Rhodes’ dictionary (1651).

indisputably Chinese in origin.<sup>17</sup> Rục, an extremely conservative language even among Minor Vietic languages, lends support to the claim that these TCV lexical forms are genuine Vietnamese words and not Chinese loans.

**Table 6: Thanh-Churong Cognates with Mường and Rục**

Gloss	OV	IPA	TCV	IPA	Others
1. day	hôm	ho:m <sup>33</sup>	bựa	bɪ:ə <sup>22</sup>	bưa (Rục)
2. head	đầu	dəw <sup>21</sup>	trôoc	tɔ:k <sup>13</sup>	tlok (Mường) kulōok <sup>4</sup> (Rục)
3. light	nhẹ	ŋe: <sup>22</sup>	nhen	ŋen <sup>22</sup>	nhel <sup>3</sup> (Rục)
4. place	chỗ	co: <sup>35</sup>	lõ	lo <sup>31</sup>	lô <sup>4</sup> (Rục)
5. still	còn	kɔ:n <sup>21</sup>	lúa	liə <sup>35</sup>	lɔ <sup>1</sup> (Rục)
6. there	kia	ki:ə <sup>33</sup>	nó	nə: <sup>13</sup>	na <sup>3</sup> (Rục)
7. tiger	hổ	ho: <sup>31</sup>	khái	xai:j <sup>13</sup>	khal (Mường)
8. tired	mệt	met <sup>22</sup>	nhọc	ŋauk <sup>p2</sup> <sub>2</sub>	nhôok <sup>4</sup> (Rục)
9. tree	cây	kəj <sup>33</sup>	cơn	kə:n <sup>35</sup>	käl (Mường)

**Table 7: Comparison of TCV and Non-OV Vietic Vocabulary**

Gloss	HN	IPA	TC	IPA	Katuic/Vietic
Far	xa	sa: <sup>33</sup>	ngái	ŋa:j <sup>11</sup>	*səŋa:j (Proto-Katuic) cŋaay (Khmer) chŋongaj <sup>35</sup> (Rục)
to slice	thái	tʰa:j <sup>24</sup>	sát	ʂat <sup>11</sup>	*ciət (Proto-Katuic) klat (Pacoh) sat (Bru)
which	nào	na:w <sup>21</sup>	mô	mo: <sup>35</sup>	*ʔdəmə: (Proto-Katuic) chamo/mô (Rục)
defecate	iả	?iə <sup>35</sup>	e	?ɛ: <sup>35</sup>	*?eh (Proto-Katuic)
to like	thích	tʰic <sup>45</sup>	ưng	?iŋ <sup>35</sup>	ính (Pacoh)
small pot	niêu	niəw <sup>33</sup>	tréc	ʈek <sup>11</sup>	trek (Pacoh)
to prune	thiến	tʰiən <sup>24</sup>	lăt	lat <sup>22</sup>	lăt (Pacoh)

### 3.5 Mon-Khmer Vocabulary not in OV

The final etymologies discussed here include some possible Mon-Khmer cognates that are not present in OV. In table 7, the TCV forms are compared with apparent Katuic

<sup>17</sup> Such Sino-Vietnamese forms are readily checked in Sino-Vietnamese dictionaries, though the possibility still remains that these are coincidental look-alikes. However, contrasting evidence in other Vietic languages suggest otherwise.

cognates,<sup>18</sup> and when notable, other non-Katuic forms. This evidence is suggestive of a close historical relationship between Vietic languages and the Katuic branch, though whether it is a genetic relationship, contact relationship, or both certainly cannot be determined by this meager amount of data.

#### 4 Concluding Comments

Clearly, the amount of lexical and phonological difference between OV and TCV suggests that further research on Vietnamese dialects would turn up more diversity that could broaden our understanding of the relationship between Vietnamese and other neighboring languages and language groups in the region. Unfortunately, recent years have seen the neutralization of dialects along the increasingly modernized roads along the Vietnamese coastline. It is only a matter of time before that neutralization reaches places like Thanh-Chương and thereby eradicates what is a wealth of historical information.

#### References

Alves, Mark J. 2003. Ruc and other Minor Vietic languages: Linguistic strands between Vietnamese and the rest of the Mon-Khmer language family. *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society*, ed. by F. K. Lehman (Chit Hlaing).

Hoàng, Thị 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.

De Rhodes, Alexandre. 1991 (Originally published in 1651). *Từ điển Annam-Lusitan-Latinh (Thường gọi Từ điển Việt-Bồ-La)*. Hồ Chí Minh: Nhà Xuất Bản Khoa Học Xã Hội. (Both original Latin and Vietnamese translation).

Ferlus, Michel. 1974. Le groupe Viet-Muong. *Asie de Sud-Est Monde Insulindien* 5.1:69-77.

\_\_\_\_\_. 1975. Vietnamien et Proto-Viet-Muong. *Asia Sud Est un Monde Insulindien* 6.4:21-55.

Gage, William W. 1985. Vietnamese in a Mon-Khmer perspective. *Southeast Asian Linguistics Presented to Andre-G. Haudricourt*, ed. by S. Ratankul, D. Thomas, and S. Premisirat, 493-524.

Hayes, La Vaughn H. 1992. Vietic and Việt-Mường: a new subgrouping in Mon-Khmer. *Mon-Khmer Studies* 21:211-228.

Hoàng, Thị Châu. 1989. *Tiếng Việt Trên Các Miền Đất Nước (Phương Ngữ Học)* (Vietnamese in regions of the nation (dialectology)). Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.

Huffman, Franklin. 1977. An examination of lexical correspondences between Vietnamese and some other Austroasiatic languages. *Lingua* 43:171-198.

Nguyễn, Minh Đức. 1972. *Một vài nét về các thổ ngữ của tiếng Mường Hòa Bình* (Some notes of the indigenous languages of the Muong of Hoa Binh), *Tìm hiểu ngôn ngữ các dân tộc thiểu số ở Việt Nam* (Understanding the languages of the ethnic minorities of Vietnam): 6-24. Hà Nội: Viện Ngôn Ngữ Học.

<sup>18</sup> All reconstructed forms are taken from Pejros 1996.

Nguyễn, Tài Cẩn. 1995. *Giáo trình lịch sử ngữ âm tiếng Việt* (Textbook of Vietnamese historical phonology). Hà Nội: Nhà Xuất Bản Giáo Dục.

Nguyễn, Văn Lợi. 1993. *Tiếng Rục* (The Ruc language). Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.

Pejros, Ilia. 1996. *Katuic Comparative Dictionary*. The Australian National University, Pacific Linguistics Series C-132.

Thomas, David D and Robert K. Headley Jr. 1970. More on Mon-Khmer subgroupings. *Lingua* 25:398-418.

Thompson, Laurence C. 1984-85. A Vietnamese reference grammar. *Mon-Khmer Studies* 13-14. University of Hawai'i Press.



# NOMINAL AUXILIARIES IN LAI<sup>1</sup>

George Bedell

*International Christian University*  
<gbedell@yahoo.com>

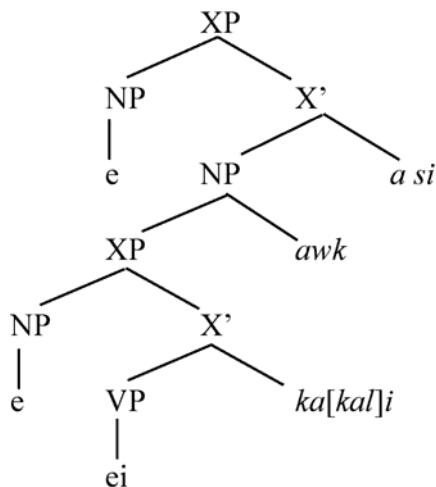
This discussion is concerned primarily with the Lai word *awk* as in (1) to (4) below. In particular, we want to determine the lexical category to which it belongs and the syntactic structure(s) in which it appears. The conclusion will be that it is a noun which functions as an auxiliary predicate.

(1) *Ka kal awk a si.*  
(2) *Kal awk ka si.*  
‘I should go.’

In (1) and (2), *awk* carries the meaning of obligation, corresponding to English ‘should’, ‘must’ or ‘have to’. The preceding *ka kal* ‘I go’ or *kal* ‘go’ appear to be a sentential or verbal complement. The following *a si* ‘it is’ or *ka si* ‘I am’ are apparently the Lai copula. The verb particles *ka* and *a* mark agreement of a finite verb with its subject, respectively first and third person singular.<sup>2</sup> The construction in (1) is impersonal, with the complement marked for subject agreement, while that in (2) is personal, with the copula marked for subject agreement. There is a difference in meaning, whereby the subject is contrasted or emphasized in (2).

We take the structure of (1) to be as in (i).

(i)



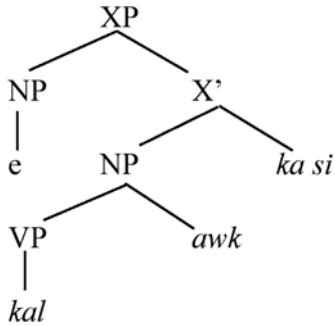
<sup>1</sup> I am grateful to Samuel Ngun Ling for teaching me what I know about Lai, and to Albert Ceuhlun, Kenneth Van Bik, Antony Ngun Uk and F. K. Lehman for discussion of some of the examples in the preliminary version of this paper.

<sup>2</sup> For fuller discussion of the system of agreement in Lai, see Bedell 1998.

In (i), *awk* is the head noun of a predicate noun phrase which also contains a sentential complement. Like other nouns, *awk* cannot be marked for tense/aspect or agreement and such markers appear on the copula *si*.

Similarly, we take the structure of (2) to be as in (ii).

(ii)



The difference between (i) and (ii) is that in the latter, the complement of *awk* is a VP rather than an XP. The semantic subject of *kal* in (ii) must be interpreted to be the subject of *awk*; this presumably characterizes the VP complement structure.<sup>3</sup> In (i), where *kal* has its own independent subject, the subject of *awk* must be interpreted as non-referential.

*Awk* appears in a distinct construction, illustrated by (3) and (4), where it seems to lack the meaning of obligation.<sup>4</sup>

- (3) *Kal awkah ka trih.*
- (4) *Kal awk ka trih.*  
'I am afraid to go.'

In these sentences, the main predicate is the verb *trih* 'be afraid' and the noun phrase headed by *awk* is a complement to it. (3) differs from (4) in the presence of the postposition *ah*, which cannot be used when *awk* is the main predicate.<sup>5</sup> Thus (5) and (6) are ungrammatical as compared with (1) and (2).

<sup>3</sup> Sentences like (2) represent a case of the clitic climbing construction as discussed in Bedell 1996a.

<sup>4</sup> Some may wish to say on semantic grounds that the *awk* in (1) and (2) is a different word from the *awk* in (3) and (4). Our position is that they are the same word, interpreted differently according to the syntactic context. The situation may be compared to English sentences like the following.

- (a) John is to go.
- (b) John is afraid to go.

In (a) there is a meaning associated with the infinitive which is not present in (b). English has no counterpart of Lai *awk* here, but we think it undesirable to identify the meaning in question with the copula *be* or the infinitive marker *to*. Whether *awk* is one word or two has little bearing on the issues raised here.

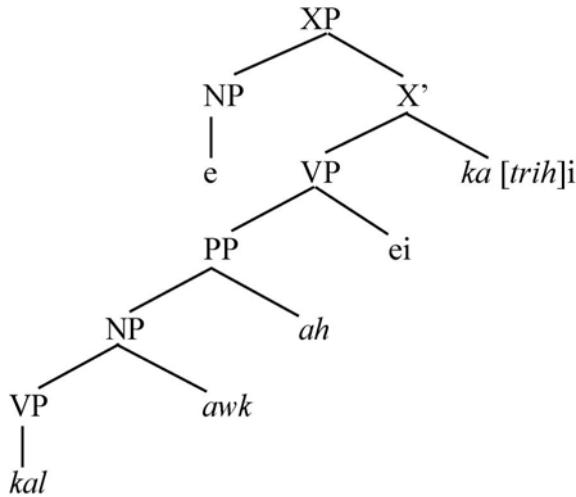
<sup>5</sup> The postposition *ah* is written as a suffix to certain words, including *awk* in our examples. It is however separable (see for example sentence (76) below) and syntactically a postposition.

(5) \**Ka kal awkah a si.*  
 (6) \**Kal awkah ka si.*

The possibility of *ah* provides additional evidence for the nominal nature of *awk*, since the complement of a postposition is typically a noun phrase.

We take the structure of (3) to be as in (iii).

(iii)



The noun phrase *kal awk* in (iii) is syntactically identical to that in (ii); since *awk* is not a predicate in (iii) it has no subject associated with it, and its subject is interpreted to be the subject of the main verb *trih*. The structure of (4) is the same as (iii), except that no postpositional phrase (and therefore no postposition *ah*) is present; here the complement of *trih* is a noun phrase.

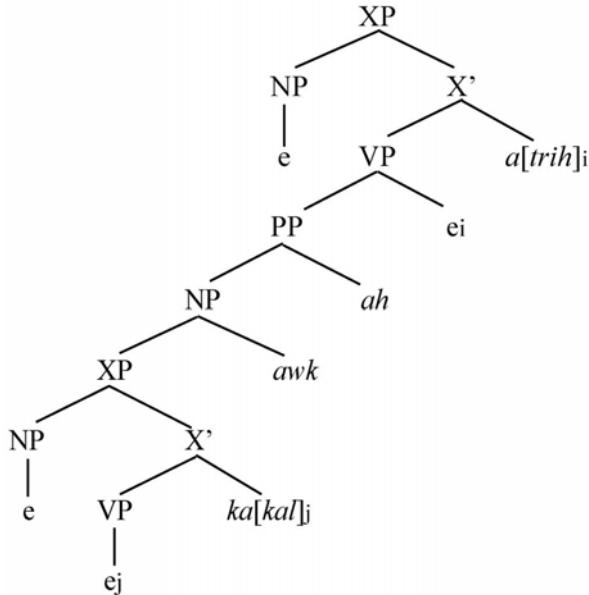
The construction in (3) and (4) differs from (1) and (2) in a second respect: there is no impersonal variant. Sentences (7) and (8) are grammatical, but interpreted as having two distinct referential subjects, one for *kal* and one for *trih*.

(7) *Ka kal awkah a trih.*  
 (8) *Ka kal awk a trih.*  
 ‘He/she is afraid I will go.’

This is clearly necessary, since unlike the copula *si*, *trih* is an independent predicate and must have a referential subject.

We take the structure of (7) to be as in (vii).

(vii)



In (vii), the noun phrase *ka kal awk* is syntactically identical to that in (i); as with (4), the structure of (8) differs from (vii) in the absence of PP over it.

In sentences (1) through (8), the verb *kal* of the complement is intransitive. But transitive verbs may equally appear in this position, as illustrated in (9) to (16) for *awk* as a predicate.

(9) *Kan chawnh awk a si.*

(10) *In chawnh awk ka si.*

(11) *Ka chawnh awk na si.*

(12) *Chawnh awk kan si.*

‘I must speak to you.’

(13) *Na ka bawmh awk a si.*

(14) *\*Ka bawmh awk na si.*

(15) *Na bawmh awk ka si.*

(16) *Bawmh awk na ka si.*

‘You must help me.’

(9) through (16) are parallel to (1) and (2) except that the verbs *chawnh* ‘speak’ and *bawmh* ‘help’ take an object in addition to a subject.

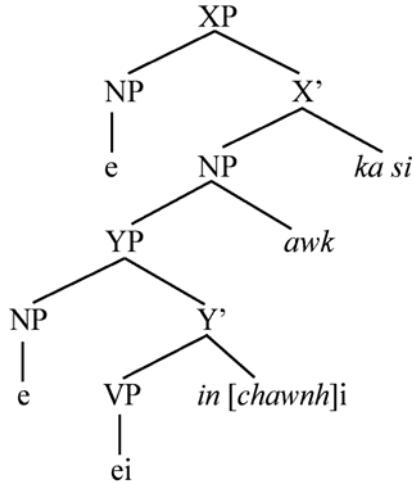
(9) and (13) are impersonal, with the complement verb marked for both subject and object agreement. *Kan* marks a first person singular subject simultaneously with a second person singular object. *Na* marks a second person singular subject and *ka* a first person singular object; in this case the two arguments are marked independently.<sup>6</sup> As shown by (10), (11) and (15) either the subject or object may appear as the subject of the copula in the personal construction. As in the case of (1) versus (2), these choices signal differences in contrast or emphasis. The ungrammaticality of (14) follows from the interpretation of *ka*

<sup>6</sup> See footnote 2.

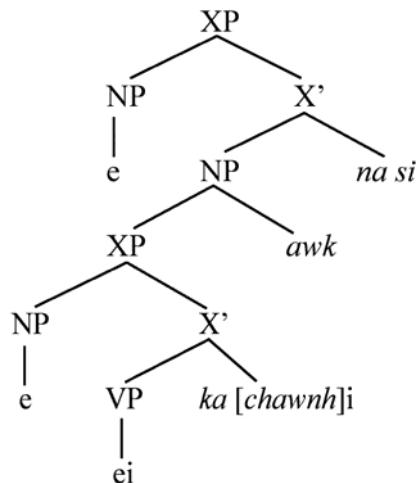
as a subject agreement marker unless accompanied by a preceding subject marker. It is grammatical (parallel to (11)) with the meaning 'I must help you'.<sup>7</sup> It is also possible, as illustrated in (12) and (16), to have a personal construction in which both subject and object agreement appear on the copula. Such complex copula agreement is found only with nominal auxiliaries. To accommodate such sentences, we need to allow *awk* to take complements consisting of an object agreement phrase (YP) as well as to allow the missing object of a transitive verb to be interpreted as the subject of a higher clause.

Thus we take the structures of (10), (11) and (12) to be as in (x), (xi) and (xii).

(x)



(xi)



Corresponding examples of *awk* as the head noun of a complement are (17) to (22).

<sup>7</sup> Albert Ceuhlun reports (personal communication) that in his Lai, sentences like (14) are grammatical with *ka* interpreted as object agreement. For him, both (11) and (14) are ambiguous. He interprets the meanings in which *ka* represents the subject of *chawnh* or *bawmh* as possessives with nominalized *chawnh-awk* 'someone to speak to' or *bawmh-awk* 'someone to help'. Thus (11) may be glossed 'you are my someone to speak to' and (14) as 'you are my someone to help'. (10) and (15) are not similarly ambiguous due to the distinction between *na* and *in*.

- (17) *Kan ton awkah ka ra lai.*
- (18) *In ton awkah ka ra lai.*
- (19) *(Nangmah) ton awkah ka ra lai.*  
‘I will come to meet you.’
- (20) *Kan ton awk ka ra lai.*
- (21) *In ton awk ka ra lai.*
- (22) *(Nangmah) ton awk ka ra lai.*  
‘I will come to meet you.’

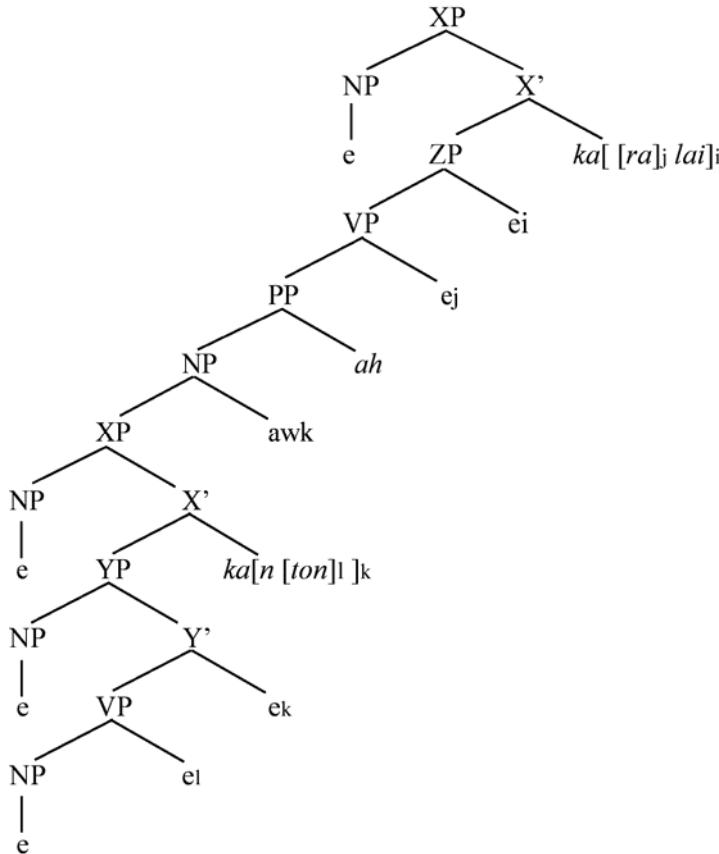
Like *trih* in (3) and (4) or (7) and (8), *ra* ‘come’ has a fixed argument structure, and cannot occur in the impersonal form. Agreement on the head verb of the *awk* complement is possible, but not required just as in the earlier examples. If subject agreement is absent, the subject of that verb, here *ton* ‘meet’, is interpreted as the subject of *ra*. If object agreement is absent, the interpretation will be that the object is third person singular, unless an overt pronoun such as *nangmah* ‘you’ appears, as in (19) or (22).

We take the structure of (17) to be as in (xvii).<sup>8</sup> The structures of (18) to (22) may be derived from (xvii) by omitting the lower XP, the YP (whose head is the object agreement marker *in* or *n*) or the PP (whose head is *ah*). Examples like (18) or (21) show that *awk* may take a YP complement.

---

<sup>8</sup> Albert Ceuhlun reports (personal communication) that in his Lai, sentences like (17) to (19) are ungrammatical. He rejects (23) below, and interprets (24) as containing a possessive *ka* with nominalized *kal-awk* as in footnote 8.

(xvii)



The structure and meaning also depend on the particular main verb. Notice first that it is quite possible with *ra* to repeat the subject agreement on the complement verb, as in (17) and (20). It is possible to do this also with *trih*, as in (23) or (24).

(23) *Ka kal awkah ka trih.*

(24) *Ka kal awk ka trih.*

‘I am afraid to go.’

With *ra* the meaning of the *awk* clause in (17) to (22) is similar if not identical to a purpose clause as in (25) to (27).<sup>9</sup>

(25) *Kan ton awk caah ka ra lai.*

(26) *In ton awk caah ka ra lai.*

(27) *(Nangmah) ton awk caah ka ra lai.*  
‘I will come in order to meet you.’

<sup>9</sup> The purpose meaning in (17) to (22) is similar to the meaning of obligation discussed in footnote 4. There an English parallel in this case as well:

(c) John came to see me.

(d) John came in order to see me.

We would not attribute the purpose meaning either to *awk* or to derivation by deletion from paraphrases like (25) to (27). But this matter also has little bearing on the issues here.

These differ in the appearance of an additional noun *ca* following *awk*. With *trih* the purpose meaning is absent, and there is no paraphrase as in (28).

(28) \**Kal awk caah ka trih.*  
 ‘\*I am afraid in order to go.’

*Awk* clauses may appear with a variety of main verbs; two additional examples are *fial* ‘order’ and *tlak* ‘deserve, be worthy’. *Fial* is transitive as in (29) to (32), and *tlak* (usually) reflexive as in (33) to (36).

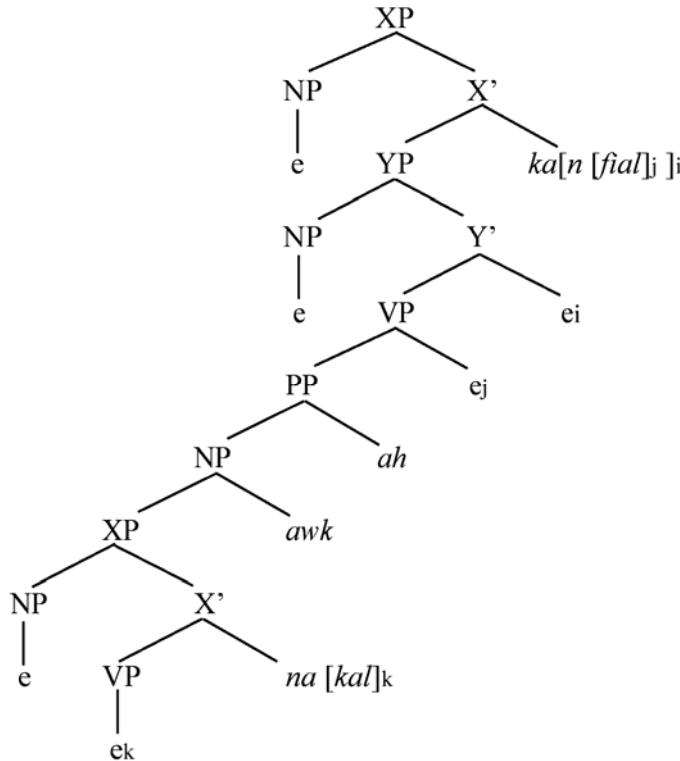
(29) *Na kal awkah kan fial.*  
 (30) *Kal awkah kan fial.*  
 (31) *Na kal awk kan fial.*  
 (32) *Kal awk kan fial.*  
 ‘I order you to go.’

(33) *Ka kal awkah kaa tlak lo.*  
 (34) *Kal awkah kaa tlak lo.*  
 (35) *Ka kal awk kaa tlak lo.*  
 (36) *Kal awk kaa tlak lo.*  
 ‘I don’t deserve to go.’

The reflexive *tlak* behaves like intransitive *trih* or *ra*, but with *fial*, the subject of the complement verb *kal* is understood to be the object rather than the subject.

We take the structure of (29) to be as in (xxix). As with (xvii), structures for (30) to (32) correspond to omitting either the lower XP or the PP.

(xxix)



Lai verbs may have more than two arguments, as illustrated in (37) to (44). The verb *pek* 'give' takes an indirect object in addition to a subject and direct object.

- (37) *Tangka kan pek awk a si.*
- (38) *Tangka in pek awk ka si.*
- (39) *Tangka ka pek awk na si.*
- (40) *Tangka pek awk kan si.*  
‘I must give you some money.’

The verb *trialpiak* 'write for' is a benefactive, and takes a beneficiary in addition to the subject and object of *trial* 'write'.<sup>10</sup>

- (41) *Cakuat na ka trialpiak awk a si.*
- (42) *\*Cakuat ka trialpiak awk na si.*
- (43) *Cakuat na trialpiak awk ka si.*
- (44) *Cakuat trialpiak awk na ka si.*  
‘You must write a letter for me.’

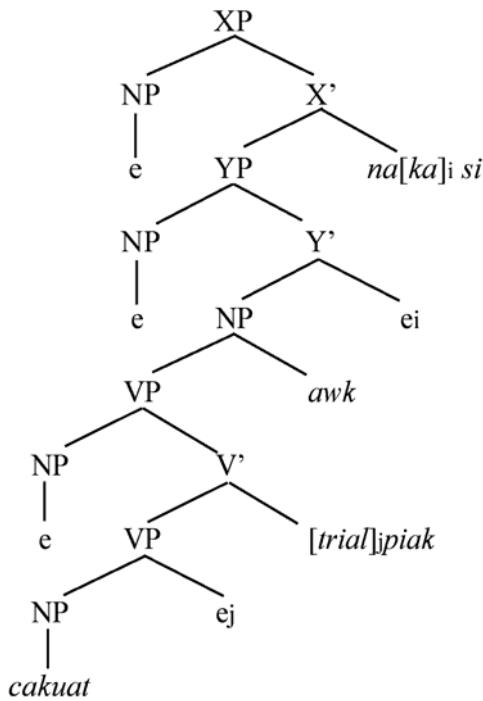
But Lai verbs may agree with no more than two arguments, and verbs like these introduce nothing new.<sup>11</sup>

We take the structure of (44) to be as in (xliv).

<sup>10</sup> For discussion of benefactive sentences in Lai, see Bedell 1997.

<sup>11</sup> Albert Ceuhlun interprets sentences like (38), (39), (41) and (42) in the same way as the examples in footnote 8.

(xliiv)



Sentences (45) through (60) illustrate the occurrence of transitive verbs in *awk* complements to transitive and reflexive verbs. The patterns are parallel to those illustrated in (17) to (22) and (29) through (44). (46) and (50) are ungrammatical in the intended meaning for the same reason as (14) and (42). However, given the semantics of the main verb *fial*, no grammatical interpretation is available in this case. Just as in (19) or (22), if object agreement is absent in the complement clause, the relevant object will be understood as third person unless an overt pronoun appears.

- (45) *Tangka na ka pek awkah kan fial.*
- (46) \**Tangka ka pek awkah kan fial.*
- (47) (*Keimah*) *tangka na pek awkah kan fial.*
- (48) (*Keimah*) *tangka pek awkah kan fial.*
- (49) *Tangka na ka pek awk kan fial.*
- (50) \**Tangka ka pek awk kan fial.*
- (51) (*Keimah*) *tangka na pek awk kan fial.*
- (52) (*Keimah*) *tangka pek awk kan fial.*  
‘I order you to give me some money.’
- (53) *Cakuat kan trialpiak awkah na ka fial maw?*
- (54) *Cakuat in trialpiak awkah na ka fial maw?*
- (55) (*Nangmah*) *cakuat ka trialpiak awkah na ka fial maw?*
- (56) (*Nangmah*) *cakuat trialpiak awkah na ka fial maw?*
- (57) *Cakuat kan trialpiak awk na ka fial maw?*
- (58) *Cakuat in trialpiak awk na ka fial maw?*
- (59) (*Nangmah*) *cakuat ka trialpiak awk na ka fial maw?*

(60) (*Nangmah*) *cakuat trialpiak awk na ka fial maw?*  
 ‘Do you order me to write a letter for you?’

In addition to *awk*, there is at least one other Lai word which may be regarded as a nominal auxiliary, at least in some of its uses. Sentences like (61) to (64) are variants of (1) to (4); the meaning of *ding* here is very close to that of *awk* in the former examples.

(61) *Ka kal ding a si.*  
 (62) *Kal ding ka si.*  
 ‘I should go.’  
 (63) *Kal dingah ka trih.*  
 (64) *Kal ding ka trih.*  
 ‘I am afraid to go.’

However, *ding* also appears as an adverbial particle within the Lai verb complex, as in examples like (65).

(65) *Ka kal ding.*  
 ‘I will certainly go.’

Since the copula *a si* may be used to emphasize any predication, a sentence like (61) will be ambiguous, interpretable also as (65). But (62) to (64) cannot be so understood. In any case, *awk* is not used as a parallel adverbial particle.<sup>12</sup>

(66) *\*Ka kal awk.*

Some examples follow of the constructions discussed, taken from the 1978 translation of the Bible. (67) to (69) are impersonal sentences with the predicate *awk*, parallel to sentence (1). The agreement particles are underlined (the verb *thlacam* ‘pray’ is intransitive, while *thrit* ‘marry’ is transitive). (69) shows a copula sentence as complement to the predicate *awk*. No examples have been found illustrating the remaining possibilities in examples like (9) to (16).

(67) *Hi bantuk hin thla cu nan cam awk a si.* (Mt. 6: 9)  
 ‘You should pray like this.’

(68) *Herodias cu na thrit awk a si lo.* (Mt. 14: 4)  
 ‘You must not marry Herodias.’

(69) *Ahohmanh pakhatnak si a duhmi cu nan sal a si awk a si* (Mt. 20: 27)  
 ‘Whoever wants to be first must be your slave.’<sup>13</sup>

<sup>12</sup> There are varieties of Lai in which *awk* is used as an adverbial particle, and sentences like (66) are grammatical.

<sup>13</sup> In structures like (i), (ii), (x), (xi), (xii) and (xliv), we treated the copula *si* as an element of X which appears only when agreement must be specified for a nominal auxiliary. But as shown by sentences like (69) and (75) below, this is not the case in general. The copula may appear

(70) and (71) by contrast are personal sentences with the agreement markers (again underlined) appearing with the copula *si*, parallel to sentence (2).

(70) *Pathian nih tuah awk a si a timi paoh cu* (Mt. 3: 15)  
 ‘all the things which God says must be done’

(71) *Mi nih zeitindah kan cungah an tuah lai ti kha phan awk kan si fawn.* (Mt. 21: 26)  
 ‘We should fear what people will do to us.’

Sentences (72) and (73) are examples of *awk* heading complement clauses with and without the postposition *ah*, parallel to sentences (3) and (4).<sup>14</sup>

(72) *Cuka i va um awk cu a trih.* (Mt. 2: 22)  
 ‘He was afraid to go and stay there.’

(73) *Mari hi na nupi ca i lak awkah trih hlah.* (Mt. 1: 20)  
 ‘Do not be afraid to take Mary as your wife.’

Sentences (74) to (77) are similar examples with a variety of different verbs which may take *awk* complements. In all these cases the verb directly under *awk* (underlined) is without agreement.

(74) *Herod nih hin ngakchia hi thah awkah a kawl lai.* (Mt. 2: 13)  
 ‘Herod will order the child to be killed.’

(75) *Ka zul u law, mi tlaitu si awk kan cawnpiak hna lai.* (Mt. 4: 19)  
 ‘Follow me, and I will teach you to be fishers of men.’

(76) *An cawnpiaknak kha tlinter awk tu ah ka ra.* (Mt. 5: 17)  
 ‘I come to fulfill their teaching.’

(77) *Hel mei chung i tlak awkah aa tlakmi nan si lai.* (Mt. 5: 22)  
 ‘You will deserve to fall into Hell.’

---

without agreement in infinitive constructions under *duh* ‘want’ or *awk* itself. Presumably these instances of *si* must be verbs rather than dummy Xs.

<sup>14</sup> Sentences like (72) and (73) provide a further argument in favor of the nominal nature of *awk*. The phrases *cuka i* and *ca i* are variants of *cuka ah* and *caah*, respectively. Presumably the variation affects the postposition *ah* (see also footnote 5 above), which changes to *i* when the PP it heads modifies a noun rather than a verb. Compare (c) and (d).

(e) *Cuka ah a um.*  
 He was there.  
 (f) *Mari hi a nupi caah a lak.*  
 He took Mary for his wife.

The variation is apparently triggered by *awk* in (72) and (73), but is not really accounted for by our analysis (in which *um* and *lak* retain their verbal nature). An additional example is (77) below.

Examples with *ding* are (78) to (82). (78) is an impersonal sentence parallel to (61): *ding* here could be either the nominal auxiliary or the adverbial particle (though in context the meaning seems to fit better with the former analysis).

(78) *kannih cu na kan chuah ding a si ahcun* (Mt. 8: 31)  
 ‘if you have to expel us’  
 ‘if you will definitely expel us’

(79) is parallel to (62); that is a personal sentence in which *ding* must be analyzed as a nominal auxiliary because the verb below it (*pek* ‘give’) lacks agreement.

(79) *Tangka cu misifik hna kha pek khawh ding a si trung.* (Mt. 26: 9)  
 ‘the money should be able to be given to the poor’

In (80), the nominalized clause *chim ding* ‘what you should say’ lacks agreement entirely, but the meaning is consistent with a nominal auxiliary analysis.<sup>15</sup>

(80) *A can a phak tikah chim ding cu pek nan si te ko lai.* (Mt. 10: 19)  
 ‘When the time comes, what you should say will be given to you.’

(81) is parallel to (63), with the presence of the postposition *ah* inconsistent with an adverbial particle analysis.

(81) *Nang ka inn i rat dingah hin kaa tlak lo.* (Mt. 8: 8)  
 ‘I do not deserve for you to come to my house.’

Sentences like (82), on the other hand, are inconsistent with a nominal auxiliary analysis.

(82) *Hi nawlbia cu ka zulh ding cang hna.* (Mt. 19: 20)  
 ‘I definitely have followed these laws.’

Hay-Neave (1948) mentions *awk* in two places. It is listed under ‘Auxiliary Verb Particles’ (p. 21) as meaning ‘obligation’, with the example (83).

(83) *A hngal lo sihmansehlaw hngalh awk a si.*  
 ‘He doesn’t know but he ought to.’

It is also listed as a suffix under ‘Nouns Formed from Verbs’ (p. 38) as meaning ‘something to’ or ‘something for’, with the examples (84) to (86).

(84) <i>trih-awk</i>	something to fear, i. e. a dread
(85) <i>ei-awk</i>	something to eat, i. e. an eatable
(86) <i>din-awk</i>	something to drink, i. e. a drink

---

<sup>15</sup> The main verb of (80) is passive, as signalled by the presence of the copula. See Bedell 1996b for discussion of Lai passives.

Clearly these correspond to the two uses of *awk* distinguished in our examples (1) and (2) as a predicate versus (3) and (4) as a complement head. Though Hay-Neave does not raise the question of whether they are the same word or not, he assigns them different grammatical status and different meanings, and so seems not to identify them. No example has been found of a clearly lexicalized noun parallel to (84) to (86).<sup>16</sup> Hay-Neave may have had in mind cases like (87) or (88).

(87) *Nanmah nih ei awk zeilo pe hna u.* (Mt. 14: 16)  
 'You give them something to eat.'

(88) *atu cu rawl ei awk an ngeih ti lo caah* (Mt. 15: 32)  
 'because now they have no food to eat'

In these sentences, *ei awk* is a kind of nominalized clause parallel to *chim ding* in (80) and not a derived noun. That is, it has the same analysis in sentences like (89) or (90), to which Hay-Neave's analysis cannot apply.

(89) *Khuazei in dah ka ei awk ka rawl cu ka hmuh lai?* (Mt. 6: 31)  
 'Where will I find my food that I should eat?'

(90) *Rawl ei awkah khan an thru.* (Mt. 26: 20)  
 'They sat down to eat.'

Hay-Neave does not mention *ding* at all in his grammatical discussion.

## References

Bedell, George. 1996a Clitic Climbing in Lai. In *Pan-Asiatic Linguistics, Proceedings of the Fourth International Symposium on Languages and Linguistics* Volume II, Thailand: Mahidol University at Salaya.

Bedell, George. 1996b Passives and Clefts in Lai. *Paper presented to the Sixth Annual Meeting of the Southeast Asian Linguistics Society*, Eugene, Oregon.

Bedell, George. 1997 *Benefactives and Clause Union in Lai*. Paper presented to the Seventh Annual Meeting of the Southeast Asian Linguistics Society, Urbana, Illinois.

Bedell, George. 1998. Agreement in Lai. In Chelliah and De Reuse, eds. *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society* 1995, Tempe: Arizona State University.

*Lai Baibal Thiang* (The Holy Bible in Lai). 1978. Calcutta: United Bible Societies.

---

<sup>16</sup> In this connection, see the interpretations of Albert Ceuhlun discussed in footnotes 8 and 11. His Lai seems to contain such nominalizations.

# LEXICAL DECOMPOSITION AND LOCATIVE PREDICATES IN BONGGI

Michael E. Boutin

*Graduate Institute of Applied Linguistics*  
<[michael\\_boutin@gial.edu](mailto:michael_boutin@gial.edu)>

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

The purpose of this paper is twofold: 1) to propose an alternative analysis of locative predicates within the theoretical framework of Role and Reference Grammar (RRG); and 2) to provide a semantically based classification of Bonggi locative predicates including both verbs of location and change of location.<sup>2</sup> These two purposes are interconnected in that the proposed alternative analysis is illustrated via the classification of locative predicates.

Within the theory of RRG, locative predicates have traditionally been treated as having two arguments (the located entity and the location), e.g. Foley & Van Valin (1984:53), Jolly (1993:277), Van Valin (1993a:39) and Van Valin & LaPolla (1997:115). Although these predicates have a semantic valency of two, they have a syntactic valency of one in a large number of languages. In order to account for the discrepancy between semantic and syntactic valency, locative predicates have been treated as exceptions in terms of valency correlation within RRG. In the analysis presented here, the location is treated as a predicate, not a referring expression, and therefore these verbs are necessarily intransitive and not exceptions in terms of valency correlation (cf. Van Valin & LaPolla 1997:156).

§2 introduces some key concepts in RRG. §3, the heart of this paper, provides a semantically based classification of Bonggi locative predicates. §4 discusses the advantage of the proposed analysis over previous RRG analyses of locative predicates.

## 2 Overview of RRG<sup>3</sup>

Predicates are classified into different *Aktionsart* types on the basis of a series of tests which have cross-linguistic validity (Van Valin & LaPolla 1997:93ff.). The tests I use to determine *Aktionsart* types in Bonggi are given in Table 1 (cf. Table 3.1 in Van Valin & LaPolla 1997:94).

---

<sup>1</sup> I would like to thank Eugene Casad, John Dillon, Mark Miller, David Moody, Bhuvana Narasimhan, Chuck Walton and, especially, Robert D. Van Valin, Jr. for their comments and discussion on earlier drafts of this article.

<sup>2</sup> Bonggi is a Western Austronesian language spoken by approximately 1,400 people on Banggi and Balambangan islands in the Kudat District of Sabah, Malaysia.

<sup>3</sup> The reader is referred to Van Valin (1993b) and Van Valin & LaPolla (1997) for elaboration of the theory.

**Table 1:** Tests for determining *Aktionsart* type in Bonggi

Criterion	States	Achievements	Accomplishments	Activities	Active accomplishments
1 Occurs with adverb <i>kosog</i> 'vigorously'	No	No	No	Yes	Yes
2 Occurs with adverb <i>peladn-peladn</i> 'slowly'	No	No	Yes	Yes	Yes
3 Occurs with X for an hour	Yes	No	irrelevant	Yes	irrelevant
4 Occurs with X in an hour	No	No	Yes	No	Yes

A nonverbal locative clause whose predicate is a prepositional phrase is illustrated in (1). The Bonggi clause in (1) differs from its English translation in that there is no copula verb in Bonggi. The absence of a copula verb accounts for the nonverbal nature of such clauses.

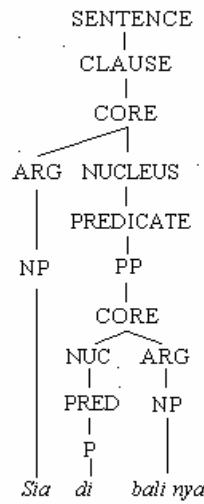
(1) *Sia di bali nya.*  
 3s.NOM at house 3s.GEN  
 'He is at his house.'<sup>4</sup>

RRG takes the position that clause structure is layered. Table 2 illustrates the relationships between semantic elements and syntactic units involved in the layered structure of the clause (Van Valin & LaPolla 1997:27). The layered structure for (1) is represented by the tree in Figure 1. The clause in Figure 1 consists of the core which contains the nucleus and the argument (*sia* '3s.NOM') of the predicate. The predicate in (1) is a prepositional phrase. Predicative prepositional phrases have a layered structure similar to clauses (Van Valin & LaPolla 1997:53).

**Table 2:** The layered structure of the clause

SEMANTIC ELEMENT(S)	SYNTACTIC UNIT
Predicate	Nucleus
Argument in semantic representation of predicate	Core argument
Non-arguments	Periphery
Predicate + arguments	Core
Predicate + arguments + non-arguments	Clause (= Core + Periphery)

<sup>4</sup> Abbreviations used: ACC accusative, ACH achievement, ACL accomplishment, ACT actor, ACY activity, ARG argument, CAU causative, GEN genitive, INGR ingressive, IRR irrealis, LS logical structure, [MR1] one macrorole, NOM nominative, NP noun phrase, NUC nucleus, P preposition, PP prepositional phrase, pred predicate, PRF perfective aspect, PSA privileged syntactic argument, REAL realis, RRG Role and Reference Grammar, s singular, SR semantic representation and ST stative. The PSA in Bonggi occurs in the English free translation in **bold**. Underlying forms are enclosed in brackets following verbs. The symbols ← and ↔ mean 'assigned/linked.'



**Figure 1: Layered structure of the clause in (1)**

In RRG the relationship between a predicate and its arguments is expressed by **Logical Structures** (LSs). LSs provide a formal semantic representation for each verb and they consist of predicates, their arguments and a small set of operators (Van Valin 1990:223). Semantic representations in RRG are based on Dowty's (1979) theory of verbal semantics in which verbs are classified into states, achievements, accomplishments and activities. Table 3 represents the logical structures for the four basic *Aktionsart* classes (cf. Table 3.3 in Van Valin & LaPolla 1997:102).

**Table 3: Logical Structures for basic *Aktionsart* classes**

<i>Aktionsart</i> type	Logical structure <sup>5</sup>
State	<b>predicate'</b> (x) or <b>predicate'</b> (x, y)
Achievement	INGR <b>predicate'</b> (x) or INGR <b>predicate'</b> (x, y)
Accomplishment	BECOME <b>predicate'</b> (x) or BECOME <b>predicate'</b> (x, y)
Activity	<b>do'</b> (x, [ <b>predicate'</b> (x)]) or <b>do'</b> (x, [ <b>predicate'</b> (x, y)])

Although verbal semantics is primarily concerned with the classification of verbs, nonverbal clauses such as (1) are also given a formal representation. The standard LS for locative statives is shown in (2a). The LS in (2a) indicates that locative stative predicates have the two-place abstract predicate **be-LOC'** (x, y) in their logical structure with 'x' and 'y' being the two arguments (cf. Jolly 1993:277; Van Valin & LaPolla 1997:115). The alternative analysis proposed in this paper is shown in (2b). Both LSs in (2) conform to the second LS for states in Table 3; i.e., **predicate'** (x, y). In (2b) the second argument position 'y' is filled by a predicate which means it cannot function as an argument. (2b) involves an embedded predicate; that is, **pred'** is embedded under **be'**.

<sup>5</sup> The conventions for LSs are as follows: predicates are represented in boldface followed by a prime (**pred'** is an abbreviation for **predicate'**); variables are filled by lexical items from the language being analyzed; and elements in small caps are modifiers of the predicate.

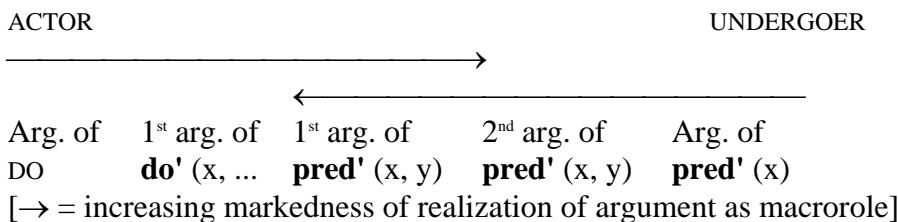
(2) a. LS for locative statives from Van Valin & LaPolla (1997:115): **be-LOC'** (x, y)  
 b. alternative LS analysis for locative statives: **be'** (x, [**pred'**])

Whereas the proposed LS in (2b) would be the LS for all locative statives, the semantic representation (SR) for (1) would be that shown in (3). In (1) the predicate is the prepositional phrase *di bali nya* 'at his house' and *sia* '3s.NOM' is the entity which is located at the site specified by the predicate.

(3) SR for (1): **be'** (3s, [**be-at'** (*bali* 3s)])<sup>6</sup>

**Actor** and **undergoer** are the two primary arguments of a transitive predicate, either one of which may be the single argument of an intransitive verb (Van Valin 1993a:43). "Actor and undergoer are generalizations across classes of specific argument positions in logical structure" (Van Valin & LaPolla 1997:142). The relationship between macroroles and argument positions in LS is captured in the Actor-Undergoer Hierarchy in (4) (Van Valin & LaPolla 1997:146). This double hierarchy states that the argument position that is leftmost on the cline will be the actor and the argument position that is rightmost will be the undergoer. This is the unmarked situation; marked assignments to undergoer are possible.

(4) **Actor-Undergoer Hierarchy**



The number of macroroles a verb takes is either  $\emptyset$ , 1 or 2, and is largely predictable from the LS of the verb (Van Valin 1993a:46-47). Default principles for macrorole assignment are shown in (5).

(5) Default Macrorole Assignment Principles:

- a. Number: the number of macroroles a verb takes is less than or equal to the number of arguments in its LS.
  - 1. If a verb has two or more arguments in its LS, it will take two macroroles.
  - 2. If a verb has one argument in its LS, it will take one macrorole.
- b. Nature: for verbs which take one macrorole,
  - 1. If the verb has an activity predicate in its LS, the macrorole is actor.
  - 2. If the verb has no activity predicate in its LS, the macrorole is undergoer.

<sup>6</sup> (3) provides a rough account of the relationship between the elements of the possessive phrase *bali nya* 'his house'. A more detailed SR would represent this relationship in terms of the predicate **have'** as in **be'** (3s, [**be-at'** (**have'** [3s, *bali*])]) where the underlined item (*bali* 'house') functions as the head of the NP.

Since the second argument in (2b) is a predicate, it cannot function as an argument. Thus, despite having two argument positions ('x' and 'y'), locative statives have only one macrorole.<sup>7</sup> This follows from the principle in (5a.2). The nature of the single macrorole is predictable from (5b); that is, the single macrorole in (1) is an undergoer since there is no activity predicate in its LS in (2b).<sup>8</sup>

In the LS configuration **be'** (x, [pred']), **pred'** corresponds to the 'y' argument position in **predicate'** (x, y). The undergoer corresponds to the first argument of **predicate'** (x, y) since it is the rightmost available argument configuration on the cline of the Actor-Undergoer Hierarchy in (4).<sup>9</sup> Therefore, 'x' is the undergoer in the LS configuration **be'** (x, [pred']). In (1) the located entity (*sia* '3s.NOM') is assigned the macrorole status undergoer.

Macroroles provide the primary link between semantic representation and syntactic representation. Once arguments have been assigned to macroroles, actor and undergoer are assigned to specific morphosyntactic statuses (Van Valin 1993a:76). The most important morphosyntactic status is the **privileged syntactic argument** (PSA) which includes both pivots and controllers.

Part of the process involved in assigning actor and undergoer to specific morphosyntactic statuses is case and preposition assignment. Case marking rules make crucial reference to macroroles and direct core argument status (Van Valin 1993a:72). The case marking rules for Bonggi are given in (6). The rules in (6) apply only to direct core arguments in main clauses.<sup>10</sup>

#### (6) Case marking rules for Bonggi

- a. The PSA takes NOMINATIVE case.
- b. Non-PSA actors take GENITIVE case.
- c. Non-PSA undergoers take ACCUSATIVE case.
- d. Non-macrorole arguments take DATIVE case as their default case.

Only personal pronouns are inflected for case; otherwise, overt case marking is analytic. Analytic case markers include proclitics (which occur with personal nouns) and prepositions. Only personal pronouns and personal nouns receive overt nominative case marking. Common nouns are not overtly marked for nominative case. For example, because the PSA in (1) is a pronoun, it is inflected for nominative case; i.e., *sia* '3s.NOM'.

To summarize, an RRG analysis of clauses (e.g. (1)) includes a syntactic representation as in Figure 1, a semantic representation as in (3), and a small set of principles for linking the two types of representation. These principles include the default macrorole assignment principles in (4) and (5) and language specific principles for selecting a PSA and assigning case as in (6) (cf. Van Valin & LaPolla 1997:177).

<sup>7</sup> Compare Van Valin & LaPolla's discussion of internal experience, attributive and identificational stative predicates (1997:125ff., 156).

<sup>8</sup> Activity predicates are predicates with **do'** in their LS (cf. §3.4).

<sup>9</sup> The configuration **pred'** (x) is unavailable because there are two argument positions. Furthermore, although there are two argument positions, there is no 'y' argument; thus, the first available argument configuration in (4) is the first argument of **pred'** (x, y).

<sup>10</sup> Core arguments are arguments represented in the LS of the verb. Direct core arguments are non-oblique syntactic arguments which correspond to arguments in the LS.

### 3 Classification of Locative Predicates

This section provides a classification of locative predicates in terms of *Aktionsart* classes. §3.1 describes locative states. §3.2 discusses locative accomplishments while §3.3 deals with locative achievements. §3.4 describes activities involving a change of location, §3.5 summarizes active accomplishments and §3.6 introduces other types of predicates whose semantic structure includes locatives.

#### 3.1 Locative States

States last or endure through time and are homogenous throughout the period of their existence. Stative situations are basic in the sense that the semantic structure of accomplishments and achievements are derived from states. A general characteristic of states is that they attribute some property to an entity. When the property attributed to an entity is the location of that entity, the result is a locative clause. The location is realized as either a locative prepositional phrase, a deictic adverb or a locative stative verb. This results in two basic types of locative stative clauses in Bonggi, nonverbal and verbal.

##### 3.1.1 Nonverbal Locative States

Nonverbal clauses are defined as clauses whose predicate is not a verb. In nonverbal locative clauses, the location is the clause predicate which is realized in syntax as either a locative prepositional phrase (e.g. (1), (7a) and (7c)) or a deictic adverb (e.g. (9)). The semantic representations (SRs) for (7a) and (7c) are provided in (7b) and (7d). In (7b) the embedded predicate (i.e., **be-inside'**) has its own argument (i.e., *bali* 3s).<sup>11</sup> This argument of the embedded predicate is a possessive phrase whose LS is **have'** (3s, *bali*) in a more detailed SR (cf. footnote 6; cf. also Van Valin & LaPolla 1997:190).

(7) a. *Sia di soig bali nya.*  
           3s.NOM at inside house 3s.GEN  
           'He is inside his house.'

b. SR for (7a): **be'** (3s, [**be-inside'** (*bali* 3s)])

c. *Sia di soig.*  
           3s.NOM at inside  
           'He is inside.'

d. SR for (7c): **be'** (3s, [**be-inside'** ( $\emptyset$ )])<sup>12</sup>

Based on the overview of RRG presented in §2, (8) is a summary analysis of (7a).

(8) a. LS for locative statives: **be'** (x, [**pred'**])  
     b. SR for (7a): **be'** (3s, [**be-inside'** (**have'** [3s, *bali*])])  
     c. Assign macroroles: undergoer  $\leftarrow$  1<sup>st</sup> argument of **pred'** (x, y)  
     d. Assign syntactic status: PSA  $\leftarrow$  undergoer '3s'  
     e. Assign case: PSA (*sia* '3s')  $\leftarrow$  nominative case

<sup>11</sup> Cf. the prepositional phrase in the tree in Figure 1; cf. also Figure 2.20a in Van Valin & LaPolla (1997:53).

<sup>12</sup> Since the argument of the preposition is unspecified in (7c), it is represented as  $\emptyset$  in the SR in (7d).

(8a) provides the LS for all locative statives. (8b) is the SR for (7a). Locative statives (e.g. (7a)) have two argument positions, but only one argument. Thus, according to principle (5a.2), they have only one macrorole. The nature of the single macrorole is predictable from (5b.2). (8c) assigns the first argument of **pred'** (x, y) (i.e., '3s') to undergoer according to the Actor-Undergoer Hierarchy in (4). (8d) assigns the undergoer '3s' to the syntactic status of PSA. Finally, (8e) assigns nominative case to the PSA following (6a).

Deictic adverbs can also function as nonverbal locative predicates.<sup>13</sup> There are two sets of spatial deictic adverbs in Bonggi. The first set refers to specific locations which are relative to the speaker as shown in Table 4. The second set, which is shown in Table 5, refers to nonspecific spatial deictics which have more approximate locations than their counterparts in Table 4.

**Table 4: Specific spatial deictics**

<i>diti</i>	'here' (near speaker)
<i>dioo</i>	'there' (not near speaker or addressee, but usually visible)
<i>dia</i>	'there' (used to track referents in discourse)
<i>dii</i>	'yonder' (not visible)

**Table 5: Nonspecific spatial deictics**

<i>kati?</i>	'somewhere here' (near speaker)
<i>kenoo</i>	'somewhere there' (not near speaker or addressee, but usually visible)
<i>kana?</i> / <i>kono?</i>	'somewhere there'
<i>kuii?</i> / <i>kii?</i>	'somewhere yonder' (not visible)

Any of the spatial deictics in Tables 4 or 5 can function as the predicate in a nonverbal locative clause. For example, the predicate in (9) is the spatial deictic *kati?* 'somewhere here'. A summary analysis of (9) is provided in (10).

(9) *Sia kati?*  
 3s.NOM here.somewhere  
 'He is somewhere here.'

(10) a. LS for locative statives: **be'** (x, [pred'])  
 b. SR for (9): **be'** (3s, [here'])  
 c. Assign macroroles: undergoer  $\leftarrow$  1<sup>st</sup> argument of **pred'** (x, y)  
 d. Assign syntactic status: PSA  $\leftarrow$  undergoer '3s'  
 e. Assign case: PSA (*sia* '3s')  $\leftarrow$  nominative case

<sup>13</sup> Although locative prepositional phrases and deictic adverbs belong to different word classes, they have the same functions. Deictic adverbs can be seen as abbreviated prepositional phrases.

### 3.1.2 Verbal Locative States

The predicate of verbal locative stative clauses is a stative verb. All locative statives share the same LS: **be'** (x, [pred']), regardless of whether they are verbal or nonverbal. (11) illustrates the locative stative verb *me-loub* 'ST-lying.chest.down'. A summary analysis of (11) is provided in (12).

(11) *Sia me-loub [m-loub].*  
 3s.NOM ST-lying.chest.down  
 'She is lying face down.'

(12) a. LS for locative statives: **be'** (x, [pred'])  
 b. SR for (11): **be'** (3s, [lying-chest-down'])  
 c. Assign macroroles: undergoer  $\leftarrow$  1<sup>st</sup> argument of **pred'** (x,y)  
 d. Assign syntactic status: PSA  $\leftarrow$  undergoer '3s'  
 e. Assign case: PSA (*sia* '3s')  $\leftarrow$  nominative case  
 f. Mark verb class via affixation: *loub* 'lying.chest.down'  $\leftarrow$  *m-* 'ST'<sup>14</sup>

(12f) shows that the verb is affixed with the prefix *m-* to indicate that the predicate is a stative verb. *m-* corresponds to **be'** in the LS. Previous discussion did not include marking the verb class via affixation because nonverbal predicates are not affixed.

Three types of prepositions are distinguished in RRG: argument-marking prepositions, adjunct prepositions and argument-adjunct prepositions (Van Valin & LaPolla 1997:159).<sup>15</sup> Prepositions which mark arguments defined by the LS of the verb have a grammatical or case marking function whereas prepositions which mark non-arguments or adjuncts have an adverbial function, e.g. *di* 'at' in (13). The locative phrase *di katil na* 'on the bed' in (13) is not part of the LS of the verb *me-loub* 'ST-lying.chest.down'; instead, it is a locative adjunct which takes the LS of the verb as one of its arguments as seen in (14a). The LS of the event ('she is lying face down') is treated as an entity being located with respect to a spatial reference point ('on the bed') (cf. Van Valin & LaPolla 1997:159). Since the locative adjunct modifies the core as a whole, it takes the LS of the verb as one of its arguments. (14c) shows that locative adjuncts are assigned to the clause periphery and (14d) indicates that the locative adjunct is marked by the preposition *di* 'at'.

(13) *Sia me-loub [m-loub] di katil na.*  
 3s.NOM ST-lying.chest.down on bed the  
 'She is lying face down on the bed.'

<sup>14</sup> The prefix vowel in (11) is epenthetic.

<sup>15</sup> Argument-adjunct prepositions are defined in §3.3. A fourth type of preposition is the head of PPs which function as clause predicates in nonverbal locative clauses such as those described in §3.1.1.

SR of *meloub*  
 'ST-lying.chest.down' & 1<sup>st</sup> argument of locative adjunct      2<sup>nd</sup> argument of locative adjunct

#      # #

(14) a. SR for (13): **be'** ([**be'** (3s, [**lying-chest-down'**])], [**be-at'** (*katil*)])  
 b. Assign macroroles: undergoer  $\leftarrow$  1<sup>st</sup> argument of **pred'** (x, y)  
 c. Assign syntactic status: PSA  $\leftarrow$  undergoer '3s'  
 periphery  $\leftarrow$  *katil na* 'the bed'  
 d. Assign case and prepositions: PSA (*sia* '3s')  $\leftarrow$  nominative case  
 locative adjunct (*katil na* 'the bed')  $\leftarrow$  *di* 'at'  
 e. Mark verb class: *loub* 'lying.chest.down'  $\leftarrow$  *m-* 'ST'

When the predicate of a locative stative clause is a verb, locative adjuncts can be either locative prepositional phrases as in (13) or deictic adverbs as in (15a). As seen in (15b), peripheral adverbs are treated as one-place predicates which take the LS of the core as their argument (Van Valin & LaPolla 1997:162).

(15) a. *Sia me-loub [m-loub] dioo.*  
 3s.NOM ST-lying.chest.down over.there  
 'She is lying face down over there.'

b. SR for (15a): **over-there'** [**be'** (3s, [**lying-chest-down'**])]

Locative stative verbs are formed by prefixing the verb with *m-* (e.g. *me-loub* 'ST-lying.chest.down' in (15) and *m-ingad* 'ST-near' in (16)). The prefix *m-* indicates that the predicate is a stative verb and corresponds to the logical predicate **be'** in the LS.

(16) *Sia m-ingad.*  
 3s.NOM ST-near  
 'It is near.'

To summarize, the LS for locative statives is **be'** (x, [**pred'**]). Locative states occur in Bonggi syntax as either verbal or nonverbal clauses. Locative stative verbs are marked by *m-* (e.g. (11), (13), (15) and (16)) which corresponds to **be'** in the LS. In nonverbal locative stative clauses, the predicate is either a prepositional phrase (e.g. (1), (7a) and (7c)) or a deictic adverb (e.g. (9)).

### 3.2 Locative Accomplishments

Accomplishments are [-punctual] and contain an underlying stative in their LS. They are derived from states by the addition of the logical operator BECOME which indicates change over some temporal span (Van Valin & LaPolla 1997:104). The LS for accomplishments varies depending upon the type of stative from which a particular accomplishment is derived. Since this paper is concerned with locatives, only accomplishments which have locative statives (i.e., **be'** (x, [**pred'**])) as part of their LS are described. This section shows how the addition of the logical operator BECOME to locative states affects both their semantic and morphological structure.

The addition of the logical operator BECOME to a locative stative indicates a change in location. For example, (17) is an accomplishment which corresponds to the locative stative in (16). The accomplishment predicate *kim-ingad* 'ACL-near' in (17) is derived by adding the prefix *kəm-* to the locative root *ingad* 'near'.<sup>16</sup>

(17) *Sia kim-ingad* [kəm-ingad].  
 3s.NOM ACL-near  
**'It** became near.'

The SRs for (16) and (17) are provided in (18a) and (18b). There is no change between locative states and accomplishments in terms of the assignment of macroroles, syntactic status or case. The difference lies in the addition of the operator BECOME to the LS of the accomplishment clause and a concomitant change in verb morphology. Whereas the difference between states and accomplishments is indicated paraphrastically in the English free translations, the difference is indicated morphologically in Bonggi where *m-* occurs with locative stative verbs and *kəm-* occurs with accomplishments.

(18) a. SR for (16): **be'** (3s, [near']) Verb affix: *m-*  
      b. SR for (17): BECOME **be'** (3s, [near']) Verb affix: *kəm-*

### 3.3 Locative Achievements

Locative achievements are [+punctual]. They are derived from states by the addition of the logical operator INGR 'ingressive' which indicates instantaneous change (Van Valin & LaPolla 1997:104). Most achievement verbs whose LS contains an underlying locative stative are lexicalized in Bonggi; i.e., there is no corresponding locative stative verb. (19) illustrates the achievement verb *ndabu?* 'fell'. A summary analysis of (19) is provided in

(19) *Sia*      *n-dabu?* [*in-dabu?*].  
     3s.NOM    ACH.REAL-fall  
     'He fell.'

(20) a. LS for locative statives: **be'** (x, [pred'])  
 b. LS for achievements with underlying locative: INGR **be'** (x, [pred'])  
 c. SR for (19): INGR **be'** (3s, [fall'])  
 d. Assign macroroles: undergoer  $\leftarrow$  1<sup>st</sup> argument of **pred'** (x, y)  
 e. Assign syntactic status: PSA  $\leftarrow$  undergoer '3s'  
 f. Assign case: PSA (*sia* '3s')  $\leftarrow$  nominative case  
 g. Mark verb class: *dabu?* 'fall'  $\leftarrow$  *n-* 'ACH.REAL'

Because locative statives have a single macrorole, achievements which contain an underlying locative stative in their LS also have a single macrorole. As was pointed out for accomplishments in §3.2, there is no change between locative states and achievements in terms of the assignment of macroroles, syntactic status or case. The difference lies in

<sup>16</sup> The prefix vowel /ə/ is realized as [i] due to vowel harmony. *Kəm-* is a phonologically conditioned allomorph. The underlying form is *-m-* which is realized as *kəm-* before vowel-initial roots and roots whose initial consonant is bilabial.

the addition of the operator INGR to the LS of achievements and a concomitant change in verb morphology. Achievement verbs are prefixed by *n-* if realis and *mə-* if irrealis (cf. (20g)).

Locative adjuncts were introduced in §3.1.2. Locative adjuncts have an adverbial function. They are either locative prepositional phrases (e.g. *di katil na* 'on the bed' in (13) and *di gimbatadn* 'on the dock' in (21)) or deictic adverbs (e.g. *dioo* 'over there' in (15a) and (22)).

(21) *Sia n-dabu?* [in-dabu?] *di gimbatadn.*  
 3s.NOM ACH.REAL-fall on dock  
 'He fell on the dock.'

(22) *Sia n-dabu?* [in-dabu?] *dioo.*  
 3s.NOM ACH.REAL-fall over.there  
 'He fell over there.'

In §3.1.2, I pointed out that RRG distinguishes three types of prepositions; however, only adjunct prepositions were discussed. The remainder of this section summarizes the distinction between adjunct prepositions and argument-adjunct prepositions.

Adjunct prepositions, which are predicates in their own right, take the LS of the verb as one of its arguments. This is illustrated in (25a) where the LS of the verb (i.e., [INGR **be'** (3s, [fall'])]) is the first argument in the LS configuration **be'** (x, [be-on' (gimbatadn)]). Whereas adjunct prepositions are two-place predicates, peripheral adverbs are one-place predicates which take the LS of the core as their argument as illustrated in (25b).

Argument-adjunct prepositions introduce an argument into the clause and share an argument with the LS of the verb rather than taking the LS of the core as an argument (Van Valin & LaPolla 1997:159). Like adjunct prepositions, argument-adjunct prepositions are predicative. Argument-adjunct prepositions are illustrated in (23) and (24) and their SRs in (25c) and (25d). For example, in (25c) the argument-adjunct preposition *kindi* 'to' (represented as INGR **be'** (3s, [be-at' (y)])) where 'y' is 'dock') shares the argument '3s' with the LS of the verb.

(23) *Sia n-dabu?* [in-dabu?] *kindi gimbatadn.*  
 3s.NOM ACH.REAL-fall to dock  
 'He fell down to the dock.'

(24) *Sia n-dabu?* [in-dabu?] *tidi gimbatadn.*  
 3s.NOM ACH.REAL-fall from dock  
 'He fell from the dock.'

(25) a. SR for (21): **be'** ([INGR **be'** (3s, [fall'])], [be-on' (gimbatadn)]) (cf. (14a))  
 b. SR for (22): **over-there'** [INGR **be'** (3s, [fall'])] (cf. (15b))

- c. SR for (23): INGR **be'** (3s, [fall']) & INGR **be'** (3s, [be-at' (gimbatadn)])<sup>17</sup>
- d. SR for (24): INGR **be'** (3s, [fall']) & INGR NOT **be'** (3s, [be-at' (gimbatadn)])

LSs in RRG are designed to identify aspects of semantic structure which affect the assignment of macroroles. They are not designed to capture the various shades of meaning which differentiate different members of the same verb class. Thus, other achievement verbs which share the LS INGR **be'** (x, [pred']) and thus belong to the same verb class as *n-dabu?* 'ACH.REAL-fall' include *n-tumang* 'ACH.REAL-stranded' (cf. (26)) and the forms shown in (27).

(26) *Sia n-tumang [in-tumang] di Kudat.*  
 3s.NOM ACH.REAL-stranded at Kudat  
 'He was stranded in Kudat.'

(27)	Realis	Irrealis <sup>18</sup>	
	<i>i-kisad</i>	<i>mi-kisad</i>	ACH.IRR-slip.off
	<i>i-kusut</i>	<i>mu-kusut</i>	ACH.IRR-step.in.hole
	<i>i-palis</i>	<i>m-palis</i>	ACH.IRR-blown
	<i>i-pupu?</i>	<i>m-pupu?</i>	ACH.IRR-fall.off
	<i>i-reba?</i>	<i>me-reba?</i>	ACH.IRR-collapse
	<i>n-suat</i>	<i>mu-suat</i>	ACH.IRR-incur
	<i>n-sulukng</i>	<i>mu-sulukng</i>	ACH.IRR-caught
	<i>n-tabukng</i>	<i>me-tabukng</i>	ACH.IRR-fall.into
	<i>n-togob</i>	<i>me-togob</i>	ACH.IRR-capsize

To summarize, both locative accomplishments and locative achievements contain an underlying locative predicate in their LS and involve movement of a located entity with respect to a location. The LS for locative accomplishments is BECOME **be'** (x, [pred']), whereas the LS for locative achievements is INGR **be'** (x, [pred']). There is no change between locative states, locative accomplishments and locative achievements in terms of the assignment of macroroles, syntactic status and case. All three types of verbs are intransitive in RRG terms; i.e., they have only one macrorole which is an undergoer.

### 3.4 Activities

Activities are situations which have arbitrary endpoints; i.e., they are inherently unbounded. On the other hand, accomplishments (cf. §3.2) and achievements (cf. §3.3) have natural endpoints; i.e., they are bounded. "For the most part, activity verbs are not derived from stative predicates but are represented as primitive predicates in their own right" (Van Valin 1990:224). The LS for activity verbs is shown in (28a). (28b) illustrates a simple English activity clause and its SR.

<sup>17</sup> '&' means 'and then' and implies temporal sequence (Foley & Van Valin 1984:51).

<sup>18</sup> Irrealis achievement verbs are marked by the prefix *mə-* indicating a hypothetical situation, e.g. *me-dabu?* 'ACH.IRR-fall'. The prefix vowel is deleted before vowel-initial roots and roots whose initial consonant is a bilabial.

(28) a. LS for activity verbs: **do'** (x, [**predicate'** (x)])  
 b. He swims. **do'** (3s, [**swim'** (3s)])

In (28) **do'** refers to a generalized unspecified activity predicate. **Do'** has two argument positions. The first argument position in (28a) is occupied by 'x', the second by another LS, i.e., [**predicate'** (x)]. Most activity verbs have a single argument which is the first argument of **do'**. The variable 'x' in (28a) refers to both the first argument of **do'** and the only argument of **predicate'**. Because the same variable 'x' is used in both places, these arguments are coreferential. Coreferential arguments are counted as a single argument in LSs. Therefore, the LSs in (28a) and (28b) apply to single argument (one-place) activity predicates.

By (5a.2) one-place activity verbs take one macrorole. By (5b.1) the macrorole must be an actor because the LS contains the activity predicate **do'**. According to (4), '3s' in (28b) is linked to actor. Since this paper deals with locatives, only motion activities are described since they involve movement of an entity with respect to a location.

The activity verb *swim* in (28b) is an example of a motion activity verb. The Bonggi clause which corresponds to (28b) is shown in (29) with a summary analysis in (30).

(29) *Sia l-em-ongi [-m-longi].*  
 3s.NOM -ACY-swim  
 'He swims.'

(30) a. LS for motion activity verbs: **do'** (x, [**predicate'** (x)])  
 b. SR for (29): **do'** (3s, [**swim'** (3s)])  
 c. Assign macroroles: actor  $\leftarrow$  1<sup>st</sup> argument of **do'** (x, ...)  
 d. Assign syntactic status: PSA  $\leftarrow$  actor '3s'  
 e. Assign case: PSA (*sia* '3s')  $\leftarrow$  nominative case  
 f. Mark verb class: *longi* 'swim'  $\leftarrow$  -m- 'ACY'<sup>19</sup>

The infix *-m-* indicates that the predicate is an activity verb and corresponds to the generalized activity predicate **do'**. That is, there is a direct relationship between **do'** in the LS and the infix *-m-* in non-imperative activity verb clauses.

### 3.5 Active accomplishments

The addition of a definite goal to motion activity verbs results in an active accomplishment because the definite goal provides a temporal boundary for the event. For example, the addition of a definite goal ('to the other side of the river') to the motion activity verb in (29) results in the active accomplishment in (31). A summary analysis of (31) is provided in (32).

(31) *Sia l-i-m-ongi [-in--m-longi] kindi seborokng sungi na.*  
 3s.NOM -PRF-ACY-swim to other.side river the  
 'He swam to the other side of the river.'

<sup>19</sup> *-m-* is realized as a prefix before vowel-initial roots and roots whose initial consonant is a bilabial obstruent; otherwise, it is infixed after the initial consonant of the stem.

(32) a. LS for motion activities: **do'** (x, [predicate' (x)])  
 b. LS for active accomplishments with locative goal:  
**do'** (x, [predicate' (x)]) & BECOME **be'** (x, [pred'])  
 c. SR for (31): **do'** (3s, [swim' (3s)]) &  
 BECOME **be'** (3s, [be-other-side' (sungi na)])  
 d. Assign macroroles: actor  $\leftarrow$  1<sup>st</sup> argument of **do'** (x, ...)  
 e. Assign syntactic status: PSA  $\leftarrow$  actor '3s'  
 argument-adjunct  $\leftarrow$  *sungi na* 'the river'  
 f. Assign case and prepositions: PSA (*sia* '3s')  $\leftarrow$  nominative case  
*sungi na* 'the river'  $\leftarrow$  *kindi seborokng* 'to other side'  
 g. Mark verb class: *longi* 'swim'  $\leftarrow$  -m- 'ACY'

Active accomplishments can be either goal-oriented or source-oriented. Although active accomplishment verbs have two argument positions in their LS, they have only one argument and one macrorole (cf. Van Valin 1990:227; 1993a:47). By (5b.1) the single macrorole must be an actor because the LS contains the activity predicate **do'**.

The addition of a locative prepositional phrase to motion activity verbs does **not necessarily** result in an active accomplishment. For example, the addition of the locative PP ('in the river') to the motion activity verb in (29) does **not** result in an active accomplishment clause in (33a) because the locative PP is only the site of the activity. In (31) the locative PP is an argument-adjunct (cf. (32e)), whereas in (33a) the locative PP is a locative adjunct which takes the LS of the verb as one of its arguments. The SR for (33a) is provided in (33b).

(33) a. *Sia l-em-ongi [-m-longi] di sungi.*  
 3s.NOM -ACY-swim in river  
 'He swims in the river.'  
 b. SR for (33a): **be'** ([**do'** (3s, [swim' (3s)])], [**be-at'** (*sungi*)])

Besides the site at which a motion activity takes place (e.g. (33a)), motion activity verbs can have a source (e.g. (34)), a goal (e.g. (31)) and a path.<sup>20</sup> Like the addition of a definite goal, the addition of a definite source to a motion activity verb also results in an active accomplishment since the definite source provides a temporal boundary for the event. For example, the addition of a definite source ('from the other side of the river') to the motion activity verb in (29) results in the active accomplishment clause in (34).

(34) *Sia l-i-m-ongi [-in--m-longi] tidi seborokng sungi na.*  
 3s.NOM -PRF-ACY-swim from other.side river the  
 'He swam from the other side of the river.'

The only morphosyntactic difference between (31) and (34) is the difference in preposition. This difference is captured in the LS. The LS for source-oriented active accomplishments is: **do'** (x, [predicate' (x)]) & BECOME NOT **be'** (x, [pred']). The

<sup>20</sup> Discussion of **path** is beyond the scope of this paper.

difference between this LS and that found in (32b) is that the LS for source-oriented active accomplishments includes the logical operator NOT, whereas the LS for goal-oriented active accomplishments does not. The SR for (34) is: **do'** (3s, [**predicate'** (3s)]) & BECOME NOT **be'** (3s, [**be-other-side'** (*sungi na*)]).

### 3.6 Induced states of affairs

The *Aktionsart* classes described in §3.1-§3.5 depict spontaneous states of affairs; however, states of affairs can also be induced. Induced states of affairs are complex in that one state of affairs brings about another. The LS for induced states of affairs is  $\phi \text{ CAUSE } \psi$ , where  $\phi$  is a causal state of affairs which induces another state of affairs  $\psi$ . The logical operator CAUSE expresses a causal relationship between two states of affairs. The remainder of this section deals with induced locative accomplishments.

§3.2 showed that the LS for locative accomplishments is BECOME **be'** (x, [**pred'**]), while §3.5 pointed out that the LS for active accomplishments is either **do'** (x, [**predicate'** (x)]) & BECOME **be'** (x, [**pred'**]) or **do'** (x, [**predicate'** (x)]) & BECOME NOT **be'** (x, [**pred'**]) depending on whether the clause is goal-oriented or source-oriented. Induced accomplishments differ from locative accomplishments and active accomplishments in that induced accomplishments include at least one CAUSE logical operator in their LS. Causal chains are possible resulting in more complex constructions.

The distinctions between motion activities, active accomplishments, induced accomplishments and causal chains are nicely illustrated by the verb root *uhad* 'to move'. (35) illustrates a motion activity, (36) an active accomplishment, (37) an induced accomplishment and (38) a causal chain. The LSs for these four clause types are contrasted in (39) and their SRs are provided in (40).<sup>21</sup>

(35) *Sia m-i-uhad [-in--m-uhad] na.*  
 3s.NOM ACY-PRF-move PERFECT  
 'She has moved.'

(36) *Sia m-i-uhad [-in--m-uhad] tidi Kudat.*  
 3s.NOM ACY-PRF-move from Kudat  
 'She moved from Kudat.'

(37) *Sia i-ng-uhad [-in--ng-uhad] dahi sindoidn nya.*<sup>22</sup>  
 3s.NOM PRF-ACT-move dirt fingernail 3s.GEN  
 'She removed the dirt underneath her fingernails.'

(38) *Sia i-p-uhad [-in-p-uhad] diaadn tidi bali nya.*  
 3s.NOM PRF-CAU-move 1s.ACC from house 3s.GEN  
 'She made me move from her house.'

<sup>21</sup> The accomplishment LSs in (39) are source-oriented.

<sup>22</sup> Since common nouns are not overtly case marked, there is no accusative case marker in (37).

(39) a. LS for motion activities: **do'** (x, [**predicate'** (x)])  
 b. LS for active accomplishments:  
**do'** (x, [**predicate'** (x)]) & BECOME NOT **be'** (x, [**predicate'**])  
 c. LS for induced accomplishments where  $\psi$  is a locative accomplishment:  
**do'** (x, [**predicate'** (x)]) CAUSE [BECOME NOT **be'** (y, [**predicate'**])]  
 d. LS for causative constructions with an embedded locative accomplishment:  
**do'** (w, [**predicate'** (w)]) CAUSE (**do'** (x, [**predicate'** (x)]) CAUSE [BECOME  
 NOT **be'** (x, [**predicate'**])])

(40) a. SR for (35): **do'** (3s, [**move'** (3s)])  
 b. SR for (36): **do'** (3s, [**move'** (3s)]) & BECOME NOT  
**be'** (3s, [**be-at'** (Kudat)])  
 c. SR for (37): **do'** (3s,  $\emptyset$ ) CAUSE [BECOME NOT  
**be'** (*dahi*, [**be-at'** (**have'** [3s, sindoidn])])]<sup>23</sup>  
 d. SR for (38): **do'** (3s,  $\emptyset$ ) CAUSE (**do'** (1s,  $\emptyset$ ) CAUSE  
[BECOME NOT **be'** (1s, [**be-at'** (**have'** [3s, bali])])])

Active accomplishments (e.g. (39b)) have one macrorole (an actor), whereas induced accomplishments (e.g. (39c)) have two macroroles. Causative constructions with an embedded locative accomplishment (e.g. (39d)) have a superordinate CAUSE (cf. Van Valin 1993a:85).

The actor is the PSA in (35), (36), (37) and (38). Three classes of verbs are morphologically derived: *-m-* for motion activities and active accomplishments, *ng-* for induced states of affairs whose actor is the PSA, and *p-* for causative constructions.<sup>24</sup> Differences in morphology and corresponding accomplishment type have to do with the degree of control which the actor exercises. When the actor and the entity being moved are coreferential, *-m-* is used indicating a motion activity as in (35) or active accomplishment as in (36). When the actor has direct control over the entity being moved, *ng-* is used indicating an induced state of affairs as in (37). When the actor is a causer who has indirect control over the entity being moved, *p-* is used indicating a causative construction as in (38). Briefly, actors which can be construed as agents have more direct control than causers.

The notion of control explains why some verb roots cannot be used to form induced states of affairs marked by *ng-*. For example, the root *longi* 'swim' cannot be used to form an induced state of affairs marked by *ng-* since the actor does not control another entity while swimming. However, a causative verb can be derived from the root *longi* 'swim' since a causer can make someone else swim although he cannot directly control their swimming, e.g. (41).

23 In (40c) and (40d) the second argument position in the  $\emptyset$  portion of the SR is  $\emptyset$  (i.e., not specified) since the causing activity is not specified (cf. Van Valin 1990:225).

24 The prefix *ng-* is realized in different ways. The relevant phonological processes are: vowel epenthesis, vowel harmony, vowel weakening and consonant coalescence. The consonant coalescence rule replaces /ŋ-/ and root-initial voiceless consonants with a nasal homorganic to the root-initial consonant. With the exception of a few borrowed words, root-initial voiced bilabials also coalesce with /ŋ-/.

(41) *Sia i-pe-longi [-in-p-longi] anak nya.*  
 3s.NOM PRF-CAU-swim child 3s.GEN  
 'She made her child swim.'

The accomplishment verbs which have been described thus far in this section are derived from motion activity verb roots (e.g. (36), (37), (38) and (41)). However, there are verb roots (e.g. *ipa?* 'put' in (42)) which can only be used to form induced accomplishments and no activity verbs can be related to them by any surface derivational process. The SR for (42) is provided in (43).<sup>25</sup>

(42) *Sia i-ng-ipa? [-in-ng-ipa?] badi? nya di tana?.*  
 3s.NOM PRF-ACT-put machete 3s.GEN on ground  
 'He put his machete on the ground.'

(43) SR for (42): **do'** (3s, Ø) CAUSE [BECOME **be'** (*badi?*, [**be-at'** (*tana?*)])]

This section began by pointing out that the LS for induced states of affairs is  $\phi$  CAUSE  $\psi$ , where  $\phi$  is a causal state of affairs and  $\psi$  is an induced state of affairs. I have provided a survey of some of the variation that exists in the  $\phi$  and  $\psi$  portions of the LS of induced states of affairs and how these differences in meaning correspond to differences in form; for example, the differences between active accomplishments, induced accomplishments and causative constructions (cf. (39)). Although this survey has not been exhaustive, it has provided an overview of some of the most frequent types of induced states of affairs whose  $\psi$  portion includes a locative accomplishment.

Actor-PSA are marked by *-m-*, *ng-* or *p-*; *-m-* is used for activities and active accomplishments; *ng-* for induced states of affairs; and *p-* for causative constructions.

#### 4 Conclusion

Previous RRG analyses of locative statives (e.g. Foley & Van Valin 1984:53; Jolly 1993:277; Van Valin 1993a:39; Van Valin & LaPolla 1997:115) assume that these predicates have two arguments, the located entity and the location. On the one hand, because semantic valence in RRG refers to the number of arguments a verb has in its logical structure, locative stative predicates traditionally have a semantic valence of two. On the other hand, locative statives normally have a syntactic valence of one.

The discrepancy between semantic and syntactic valence is normally accounted for by analyzing locative statives as having one macrorole; thus, reducing the semantic valence to one. Although this violates principle (5a.1) in that verbs with two arguments in their LS should take two macroroles, it does not contradict the more general principle in (5a) which states that the number of macroroles a verb takes is less than or equal to the number of arguments in its LS. However, verbs which are exceptions to (5a.1) are marked by a [MR $\alpha$ ] feature where 'MR' stands for 'macrorole' and  $\alpha$  can have a value of Ø, 1 or 2 depending on the number of macroroles the verb takes (Van Valin & LaPolla 1997:154). In such an analysis, all locative stative verbs and motion verbs end up being marked as [MR1] indicating that they are idiosyncratic and have only one macrorole. Van Valin &

<sup>25</sup> Although (42) is an actor-PSA clause, there is a corresponding undergoer-PSA form.

LaPolla (1997:153ff.) suggest one solution which follows this analysis without having to mark all these verbs as exceptions.

In early 1998, Van Valin suggested to me that it might be possible to analyze locative statives like identificational and attributive constructions. The analysis of locative statives presented in this paper is based on this suggestion and has an advantage over previous analyses in that locative statives are no longer exceptions to (5a.1). Thus, no appeal need be made to the more general principle in (5a) and since locative statives are no longer an exception, they need not be marked in the lexicon as such.

Schwartz (1993:447) posited a lexical rule of predicate creation as another means of accounting for the discrepancy between semantic and syntactic valence of attributive and identificational constructions. In my analysis, identificational, attributive and locative stative constructions have the LS **be'** (x, [pred']). In all three types of constructions the second argument position is filled by a predicate which means it cannot function as an argument. In locative stative constructions, the location is treated as a predicate. Since there is only one argument, there can only be one macrorole and there is neither valence discrepancy (cf. Van Valin & LaPolla 1997:125ff., 156) nor a need for a special lexical rule.

The analysis of locative statives in §3.1 was extended to cover both accomplishments (§3.2) and achievements (§3.3) which are derived from locative statives. Furthermore, the analysis of locative prepositional phrases in §3.1.1 was expanded to handle locative adjunct prepositional phrases in §3.1.2 and locative argument-adjunct prepositional phrases in §3.3 and §3.5. Finally, §3.6 introduced different types of induced states of affairs whose resultant state is a locative accomplishment like the accomplishments described in §3.2.

## References

Dowty, David Roach. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.

Foley, William A., and Robert D. Van Valin, Jr. 1984. *Functional syntax and universal grammar*. Cambridge Studies in Linguistics 38. Cambridge: Cambridge University Press.

Jolly, Julia A. 1993. Preposition assignment in English. *Advances in Role and Reference Grammar*, ed. Robert D. Van Valin, Jr., 275-310. Amsterdam and Philadelphia: John Benjamins.

Schwartz, Linda. 1993. On the syntactic and semantic alignment of attributive and identificational constructions. *Advances in Role and Reference Grammar*, ed. Robert D. Van Valin, Jr., 433-63. Amsterdam and Philadelphia: John Benjamins.

Van Valin, Robert D., Jr. 1990. Semantic parameters of split intransitivity. *Language* 66(2):221-60.

Van Valin, Robert D., Jr. 1993a. A synopsis of Role and Reference Grammar. *Advances in Role and Reference Grammar*, ed. Robert D. Van Valin, Jr., 1-164. Amsterdam and Philadelphia: John Benjamins.

Van Valin, Robert D., Jr., ed. 1993b. *Advances in Role and Reference Grammar*. Amsterdam and Philadelphia: John Benjamins.

Van Valin, Robert D., Jr. 1995. *Role and Reference grammar. Handbook of pragmatics manual*, eds. Jef Verschueren, Jan-Ola Östman, and Jan Blommaert, 461-69. Amsterdam: John Benjamins.

Van Valin, Robert D., Jr. and Randy J. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.



# FUSION AND ALIGNMENT: THE CASE OF MALAY<sup>1</sup>

Ann Delikan

*New York University*

*<ad17@nyu.edu>*

## 1 Issues

### 1.1 *The Original Problem*

The facts of prefixation in Malay, in the case of nasal-final prefixes (e.g. 'meN-' (transitive)), follow in (1). (Roots are underlined.)

1.	a. məN + <u>pukol</u>	[məmukol]	'hit', v.t.
	b. məN + <u>kupas</u>	[məŋupas] ('mengupas')	'peel', v.t.
	c. məN + <u>təntu</u> + kan	[mənəntukan]	'determine', v.t.
	d. məN + <u>buwat</u>	[məmbuwat]	'make', v.t.
	e. məN + <u>garam</u>	[məŋgaram] ('menggaram')	'salt', v.t.
	f. məN + <u>dəsu</u>	[məndəsu]	'resound', v.i.

Rule-ordering analyses of this data would claim that that nasal substitution in Malay involves the prefix-final nasal (N) assimilating to the place of articulation of the first sound of the following root (Rule 1), followed by deletion of that onset obstruent in the case of voiceless-obstruent-initial roots (e.g. /p/, /t/, /k/) (Rule 2). Thus, for instance, [məmukol] ('mempukul', *hit*, v.t) is correct, but \*[məmpukol] is ungrammatical.

### 1.2 *OT and the Problem*

Such an analysis of the facts raises the following concern. Assimilation and deletion are two separate and unrelated processes, and there is no lack of cross-linguistic data to demonstrate this fact. Stipulating that the Malay data represents the product of a two-step process is not explanatorily adequate because it fails to explain why the processes occur together.

Optimality Theory avoids such stipulation, since it does not support a processual analysis of the data. In Optimality Theory, possible pronunciations (output candidates) of a word (input) are evaluated simultaneously against a set of ranked constraints. The pronunciation that incurs the fewest violations of the constraints is selected as the 'optimal' one (See Appendix A for the fundamental tenets of Optimality Theory). A recent Optimality theoretic treatment (Pater 1996) involves claiming that 'fusion' is a more desirable account of the above prefixation facts. A fusion version of  $məN_1 + p_2ukol$  follows in (2), where a two-to-one mapping is seen to hold between the two input segments,  $N_1p_2$ , and the single output segment,  $m_{1,2}$ .

<sup>1</sup> I am grateful to Keith Fernandes, Adamantios Gafos, Fran Gulinella, Harry van der Hulst, Young-Kook Kwon, and Nancy Ritter for their feedback and kind support, although I hasten to add that the material in this paper may not necessarily reflect their views.

2.	input	$m \ \partial N_1 + \underline{p_2} \ u \ k \ o \ l$
	output	$m \ \partial \ m_{1,2} \ u \ k \ o \ l$

Pater's one-step analysis is meant to eliminate the explanatory gap inherent in two-step analyses of the Malay facts. Pater sees fusion as one of many 'repair' strategies adopted by various languages to avoid the sequence Nasal + Voiceless Obstruent (henceforth, 'NT').<sup>2</sup> He cites the faithfulness constraint LINEARITY (McCarthy and Prince 1995), which stipulates that S1 reflect the precedence structure of S2, and vice versa, where S1 is the input and S2 the output. (See Appendix B for full statement of all constraints employed in this paper). Bearing this in mind, consider once again the facts of 'fusion', depicted in (3). (The segment resulting from fusion has two indices associated with it.)

3.  $məN_1 + \underline{p_2}ileh$  (input)  
       \ /  
       mə m<sub>1,2</sub> ileh (output) (Pater 1995:6)

As explained by Pater,  $N_1$  precedes  $p_2$  in the input, but not so in the output. This means that the output violates LINEARITY. The ranking that Pater thus proposes is responsible for the output selected is stated in (4).

#### 4. \*NT > LIN (where LIN = LINEARITY)

The tableau in (5) derives from Pater 1996, and shows the result of this ranking. (‘✓’ denotes the optimal candidate.) The candidate in (5b), in which no fusion has occurred, violates \*NT and is eliminated. Fusion, in (5a), violates LINEARITY but obeys the more high-ranked \*NT, and is therefore the preferred option.

5.	Fusion: *NT >> LIN (Pater 1996:9, #7) <sup>3</sup>	
	Input: mə N <sub>1</sub> + <u>p<sub>2</sub>ileh</u>	*NT      LIN
a.	məm <sub>1,2</sub> ileh ✓	*
b.	məm <sub>1</sub> p <sub>2</sub> ileh	* !

Root-internal data in Malay does not, however, seem to involve fusion. (See Appendix C.c) Pater accounts for this fact by claiming that it is more undesirable to allow fusion root internally than it is to permit the undesirable sequence, but that the constraint against the sequence, \*NT, is a more powerful constraint than a constraint (LINEARITY-ROOT) that militates against non-root-internal fusion. Thus the constraints responsible for

<sup>2</sup> Pater finds that other attested strategies are nasalization of the obstruent, denasalization of the nasal, and deletion of the nasal.

<sup>3</sup> For typological simplicity, I have chosen to use neither Pater's 1996 \*N constraint nor his 1995 version, \*NC, but \*NT. The check denotes the optimal candidate

selection of the optimal candidate in this analysis include both a root-specific and non-root-specific version of a constraint against fusion. The root-specific constraint outranks \*NT, which in turn outranks the general LINEARITY constraint. The relevant ranking is expressed in (6). The tableau in (7) illustrates the selection made. Candidate (7a) fails to get selected because the high-ranked root-specific LINEARITY is violated since fusion has occurred within a root. The non-fusion candidate, (7b), which violates the lower-ranked \*NT, is selected instead.

6. LINEARITY-ROOT >> \*NT >> LINEARITY

7. **Root-internal NT tolerance: LINROOT >> \*NT>> LIN**

Input: <u>əm<sub>1</sub>p<sub>2</sub>at</u>	LINROOT	*NT	LIN
a. əm <sub>1,2</sub> at	* !		*
b. ə m <sub>1</sub> p <sub>2</sub> at✓		*	

(Pater 1996:10, #8 and Pater 1995:8)

### 1.3 Further Complications

Pater writes, ‘Clearly a lot of work needs to be done to determine the empirical coverage of root-specific LINEARITY constraints’ (Pater 1995) — and it does. There is further data in Malay that fails to be accounted for by this analysis. Specifically, the boundaries a) between prefixes, b) between suffixes, c) between reduplicants, and d) between roots and suffixes do not pattern along the dichotomy Pater suggests. (See Appendix C for data in question.) Thus, for instance, fusion does not occur between prefixes, but there is nasal place assimilation, as shown in (8). (Again, italics denote the result of fusion.)

8. məN + pər + səmbah (+kan)  
məmpərsəmbahkan, (\*məmərsəmbahkan)

\\  
lab ‘present (verb)’ + performativ

These facts mirror the root-internal facts of the language. In the environments (b-d) (above), neither fusion nor place assimilation is in evidence. Taken together with the prefix-root boundary, which is the only spot where fusion does seem to occur, these facts indicate the need for some new explanation.

## 2 Analysis

### 2.1 A New Dichotomy

I maintain that parameterizing a fusion constraint according to environment would fail to capture the generalization underlying all the environments, and would be theoretically uneconomical. I suggest instead that there is a dichotomy between sequences of segments that are homogenous with respect to morphological category-type (e.g. root-internal, prefix-prefix, suffix-suffix, reduplicant-reduplicant) and those that are heterogenous with respect to morphological category type (e.g. prefix-root, root-suffix). Fusion is blocked in the former type of sequences, but may be permitted in the latter. I take my cue from

McCarthy and Prince (1995), extending their UNIFORMITY constraint against coalescence (fusion) by specifying the domain of its application as in (9b).

9. a. UNIFORMITY (M&P 19995): S2 may not have multiple correspondents in S1.  
 b. UNIF-HMG: S2 may not have multiple correspondents in S1, where S1 an environment that is homogenous with respect to morphological category-type

The ranking of constraints that I propose follows in (10).

10. UNIF-HMG >> \*NT

I omit irrelevant candidates (e.g. post-nasal voicing, epenthesis, etc.) in (11), which illustrates the result of the ranking in (10).

11. **Prefix-prefix juncture: No fusion**

	UNIF-HMG	*NT
Input: məN <sub>1</sub> + p <sub>2ər</sub> + (buwat)		
a. məm <sub>1</sub> p <sub>2ər</sub> .. ✓		*
b. məm <sub>1,2ər</sub> ...	*!	

The fusion candidate, in (11b), violates UNIF-HMG and is eliminated. The selected non-fusion candidate, in (11a), honors this constraint although it violates the low-ranked \*NT. All other environments are tested in (12), using the same constraint ranking as in (11). The ranking selects grammatical outputs in all three cases. Root-internally (in (12a)) and between suffixes (in (12c)), fusion is blocked because it would involve violating the high-ranked UNIF-HMG. The opposite holds in the heterogenous environment that obtains at the prefix-root juncture, fusion selected over an \*NT violation.

12. a. **Root-internally: No fusion**

	UNIF-HMG	*NT
Input: təm <sub>1</sub> p <sub>2at</sub>		
i. təm <sub>1</sub> p <sub>2at</sub> ✓		*
ii. təm <sub>1,2at</sub>	*!	

b. **Prefix-root juncture: Fusion**

	UNIF-HMG	*NT
Input: məN <sub>1</sub> + p <sub>2ileh</sub>		
i. məm <sub>1,2ileh</sub> ✓		
ii. məm <sub>1</sub> p <sub>2ileh</sub>		*!

c. **Suffix-suffix juncture: No fusion**

	UNIF-HMG	*NT
Input: ..+kan <sub>1</sub> + k <sub>2ah</sub>		
kan <sub>1</sub> k <sub>2ah</sub> ✓		*
kaŋ <sub>1,2ah</sub>	*!	

## 2.2 Fusion And Alignment

One section of the data still needs to be accounted for, as shown in (13).

13. \*EXCEPTION: Root-Suffix

	UNIF-HMG	*NT
Input: <u>rakam</u> <sub>1</sub> +k <sub>2</sub> an		
a. rakam <sub>1</sub> k <sub>2</sub> an		*
b. *rakaŋ <sub>1,2</sub> an✓		

The posited ranking selects the ungrammatical candidate in (13). The fact that fusion fails to occur at the root-suffix boundary is, I suggest, the result of an interaction of the constraint against fusion with an alignment constraint that protects the integrity of left edges of affixes more than it does their right edges. The alignment constraint in question follows in (14).

14. ALIGN Affix L, PrWd R: The left edge of every affix must be aligned with the right edge of some prosodic word.

Fusion at the boundary referred to in (14) would constitute a violation of this constraint, as it would ‘damage’ the edge in question. For fusion to be prohibited at this juncture, \*NT must be ranked lower than the alignment constraint, as shown in (15). (The comma between UNIFHMG and ALIGN AffixL, PrWdR indicates that these two constraints are not ranked with respect to each other.) The tableau in (16) shows the results of this ranking.

15. UNIF-HMG, ALIGN Affix L, PrWd R >> \*NT

16. Root-suffix juncture: No fusion

input: <u>rakam</u> <sub>1</sub> + k <sub>2</sub> an	UNIF-HMG	ALIGN Affix L,PrWd R	*NT
a. [[rakam <sub>1</sub> ] k <sub>2</sub> an] <sup>4</sup> ✓			*
b. [[raka]ŋ <sub>1,2</sub> an]		*!	

The fusion candidate, in (16b), violates the alignment constraint and is dispreferred even though it avoids violation of UNIF-HMG and \*NT. The preference is for candidate (16a), in which the root right edge does not interact with the suffix left edge, despite its violation of the low-ranked \*NT.

The tableau in (17) now illustrates all environments tested against this ranking. (Roots are underlined.) Within roots (cf. (17a)), between prefixes (cf. (17c)), between

<sup>4</sup> I assume exhaustive prosodization (McCarthy and Prince 1993) of all morphemes into recursive prosodic words.

suffixes (cf. (17e)), and between roots (cf. (17f)) fusion is blocked by the high-ranked UNIF-Hmg. In each case, a low-ranked \*NT violation is preferable. Between a root and a suffix (cf. (17d)), fusion is blocked to avoid a costly violation of the alignment constraint. Fusion is selected in (17b) because it avoids violation of all three constraints, in preference of a \*NT violation.

### 17. All other environments

	UNIFHMG,	ALIGN (Aff L, PrWd R)	*NT
a. Input: <u>təm<sub>1</sub>p<sub>2</sub>at</u>			
i. təm <sub>1</sub> p <sub>2</sub> at ✓			*
ii. təm <sub>1,2</sub> at	*!		
b. Input: <u>məN<sub>1</sub> + p<sub>2</sub>ileh</u>			
i. məm <sub>1</sub> p <sub>2</sub> ileh			*
ii. məm <sub>1,2</sub> ileh ✓			
c. Input: <u>məN<sub>1</sub> + p<sub>2</sub>ər (+ bu..)</u>	UNIF-HMG	ALIGN( Aff L,PrWd R)	*NT
i. məm <sub>1</sub> p <sub>2</sub> ər .. ✓			*
ii. məm <sub>1,2</sub> ər ..	*!		
d. Input: <u>rakam<sub>1</sub> + k<sub>2</sub>an</u>			
i. rakam <sub>1</sub> ]k <sub>2</sub> an] ✓			*
ii. raka] $\eta$ <sub>12</sub> an] <sup>5</sup>		*	
e. Input: <u>... + kan<sub>1</sub> + k<sub>2</sub>ah</u>			
i. ... ]kan <sub>1</sub> ]k <sub>2</sub> ah] ✓			*
ii. ... ]ka] $\eta$ <sub>12</sub> ah]	*!	*	
f. Input: <u>kawan<sub>1</sub> + k<sub>2</sub>awan</u> <sup>6</sup>			
i. [[kawan <sub>1</sub> ] k <sub>2</sub> awan] ✓			*
ii. [[kawa] $\eta$ <sub>1,2</sub> awan]	*!	*	

### 2.3 Prosodic Implications

If the left edge of an affix must be aligned perfectly alongside the right edge of a prosodic word, this suggests that the way in which prosodic words are built in Malay is [[prefix+root]+suffix], i.e. with the prefix and root forming a complex. This possibility runs

<sup>5</sup> I assume the insertion of prosodic word boundaries occurs to show the grouping I propose preliminarily in 2.3, namely that suffixes fall outside the prosodic word occupied by the root and prefix(es) I assume also that each suffix projects its own prosodic word.

<sup>6</sup> I assume reduplication is suffixal.

counter to the widely-held view that prosodic words for such a language are built with the root and suffix forming a complex, as shown in (18a). The new prosodic structure I propose is repeated in (18b).

18. a. Traditional analyses: [prefix + [root+ suffix] ]  
 b. Proposed prosodic structure: [ [prefix + root] + suffix]

Nasal assimilation facts in Malay seem to pattern along this tendency in the language to ‘protect’ the left edges of its affixes, lending credence to the implied asymmetry between left and right edges of affixes. The alignment constraints used thus far interact with not only \*NT but also a constraint that penalizes codas that are not place-linked to their adjacent onsets (cf. CODA CONDITION (Ito and Mester 1994)). As stated in (19), the alignment constraint outranks both the Coda Condition and \*NT constraints. The results of this ranking. appear in (20)

19. ALIGN AFF L, Pr Wd R >> CODACOND >> \*NT

20. **Alignment outranks CodaCond**

Input: <u>rakam</u> + kan     lab vel	ALIGN AFF L,PrWd R	CODA COND	*NT
a. [[rakam] kan] ✓     lab vel		*	*
b. [[raka $\eta$ ] kan] \\   vel	*!		

Place assimilation, in candidate (20b), violates the high-ranked alignment constraint. The candidate in which no interaction occurs between the right edge of the lower prosodic word and the left edge of the suffix is preferable, therefore, despite its violation of CODACOND and \*NT.

In (21), all other environments are tested against this ranking. ((In (21b), I omit candidates that involve variations in the juncture between the root and the first suffix.)

## 21. Other environments: Alignment outranks CODA-COND

a. Input: məN + pər + <u>buku</u> + kan   lab	ALIGN AffL,PrWdR	CODA-COND	*NT
i. [[məm.pər. <u>bu.ku</u> ] kan] ✓ \\  lab			*
ii. [[məŋ.pər. <u>bu.ku</u> ] kan]     vel lab		*!	
Input: <u>tuhan</u> + kan + kah     cor vel			
i. [[[tuhan] kan]kah] ✓     cor vel		*	*
ii. [[[tuhan]kan]kah] \\  vel	*!		
c. Input: <u>əmpat</u>     lab lab	ALIGN Aff L,PrWdR	CODA- COND	*NT
i. əm.pat] ✓     lab lab			*
ii. [əŋ.pat]     vel lab		*!	*

In (21a), nasal assimilation between the prefixes avoids violation of CODA-COND and cannot be blocked by the alignment constraint as the relevant boundary for its application is not involved. The candidate in which, for instance, the prefix final nasal assumes a velar place specification, falls afoul of CODA-COND and is therefore dispreferred. Candidate selection in (21c) parallels that in (21a). In (21b), the alignment constraint blocks selection of a place assimilated candidate (cf. (21bii)), the candidate in which no interaction occurs between the suffixes selected instead, despite its CODA-COND violation.

### 3 Conclusion, implications and direction for further research

The claims of this paper are two-fold. First, it shows that the data under investigation are less irregular than they seem. There is a dichotomy amongst the data that is captured by referring to a constraint like UNIF-HMG, which governs behavior of segments according to

their morphological affiliation. This account of the data is achieved without parameterization of the constraint according to every environment involved.

Second, I have shown that apparent exceptions to this are the result of asymmetrical strength of edges. I maintain that the correct edges to consider require looking from the perspective of affixes, not roots.

Significantly, this last claim involves a commitment to the way in which prosodic words are built in Malay such that the prefix and root form a complex, with suffixes falling outside in separate prosodic words.. Nasal assimilation facts support the posited asymmetry between prefixes and suffixes, but an investigation of stress assignment would be vital to determine the strength of the current claim about prosodic word formation, and to motivate the boundaries involved more thoroughly. It would also be crucial to determine whether any other segmental processes in the language interact with the proposed boundaries. Certainly, too, much work still needs to be done to determine, for instance, whether a re-ranking of the constraints under review will account for dialect differences in Malay, before it can be ascertained how far an OT treatment of such data improves upon previous analyses.

## Appendix A

### Fundamentals of OT (From McCarthy and Prince 1993a)

- a) Universality. Constraints are universal; and universally present in all grammars.
- b) Violability Constraints are violable, but violation is minimal.
- c) Ranking Constraints are ranked on language-particular basis: The notion of minimal violation is defined in terms of this ranking. A grammar is a ranking of a universal constraint set.
- d) Inclusiveness The constraint hierarchy evaluates a set of possible candidates that are admitted by very general considerations of structural well-formedness.
- e) Parallelism Best-satisfaction of the constraint hierarchy is computed over the whole hierarchy and the whole candidate set. There is no serial derivation.

An Optimality-Based Grammar, schematically:

$$\begin{aligned} \text{Gen}(\text{in}) &= (\text{cand 1}, \text{cand 2} \dots) \\ \text{Eval} (\{\text{cand 1}, \text{cand 2} \dots\}) &\rightarrow \text{cand n (the output)} \end{aligned}$$

GEN determines the set of candidate analyses consistent with a given input.(GEN may freely delete, insert or link segments as well as assign prosodic structure). GEN is conceived as a function that, for each particular input form, generates the range of all possible candidate linguistic analyses. These candidates are evaluated simultaneously by a function EVAL. EVAL rates the members of the candidate set in terms of their relative harmony, or degree of success with regard to the language's ranking of the constraints, imposing an ordering on them. A maximally harmonic member of the candidate set is optimal.

## Appendix B

### Constraint list

\*NC (Pater 1995): No nasal-voiceless obstruent sequences (Replaced in Pater 1996 by \*N, and referred to in this paper as \*NT)

LINEARITY (McCarthy and Prince 1995):  $S_1$  reflects the precedence structure of  $S_2$ , and vice versa.

MAX (McCarthy and Prince 1995): Every element of  $S_1$  has a correspondent in  $S_2$ . “MAX allows an interpretation of fusion as a two-to-one mapping from Input to Output: Two Input segments stand in correspondence with a single Output segment” (Pater 96: 7)

UNIFORMITY (McCarthy and Prince 1995): No element of  $S_2$  has multiple correspondents in  $S_1$ ,  $S_1$  the input,  $S_2$  the output.

UNIF-HMG: No element of  $S_2$  has multiple correspondents in  $S_1$ , where elements of  $S_1$  are homogenous with respect to morphological category-type,  $S_1$  the input,  $S_2$  the output.

ALIGN Affix L, PrWd R: The left edge of every affix must be aligned with the right edge of some prosodic word.

ALIGN Affix R, PrWd L: The right edge of every affix must be aligned with the left edge of some prosodic word.

CODA COND (Ito and Mester 1994): Codas are disallowed unless linked to a following onset.

## Appendix C

### DATA (a-b, Pater 1995/6)

#### a. Nasal substitution (NT) b. Failure of substitution (ND)

məN+ pileh	məmileh ‘choose’, v.t.	məN + bəli	məmbəli ‘buy’, v.t
məN + tules	mənules ‘write’, v.t.	məN + dapat	məndapat ‘get’, v.t
məN+ kesah	məŋesah ‘relate’, v.t.	məN + ganti	məŋganti ‘(ex)change’, v.t

#### c. Failure of substitution root-internally (N + any T)

təmpat ‘place’, n.	tambah ‘add on’, v.t.
hantar ‘send’, v.t.	təndaj ‘kick’, v.t
munken ‘poss./maybe’, adj.	tiŋgal ‘live/remain/die/leave’, v.i.

#### d. Failure of substitution elsewhere (new DATA)

i) məN + pər + hamba (+kan)	məmpərhambakan, *məmərhamba.. ‘enslave’, v.t.
ii) məN + tər + balek + kan	məntərbalekkan, *mənərbalekkan ‘overturn’, v.t.
iii) rakam +kan	rakamkan, *rakajan ‘record’, v.t.
iv) di + tuhan + kan + kah	dituhankankah, *dituhaŋajah ‘deified’, pass/attrib. + interrogative

## References

Cohn, Abigail & John McCarthy. 1994. *Alignment and parallelism in Indonesian*. MS, Cornell University and University of Massachusetts, Amherst.

Ito, Junko and Armin Mester. 1994. Reflections on Coda Conditon and Alignment. In Jason Merchant, Jaye Padgett and Rachel Walker (eds.) *Phonology at Santa Cruz (PASC)* III:27-46.

Ito, Junko and Armin Mester, and Jaye Padgett. 1995. Licensing and Underspecification in Optimality Theory. *Linguistic Inquiry* 26:571-614.

McCarthy, John and Alan Prince. 1995. Faithfulness and Reduplicative Identity. In Jill Beckman, Laura Walsh Dickey and Susan Urbanczyk (eds.) *University of Massachusetts Occasional Papers: Papers in Optimality Theory* 18:249-384.

----- 1993a. *Prosodic Morphology I: Constraint Interaction and Satisfaction*. University of Massachusetts, Amherst, and Rutgers University, New Brunswick, NJ, MS. [Forthcoming MIT Press].

----- 1993b. Generalized Alignment in Geert Booij -Jaap van Marle (eds.), *Yearbook of Morphology 1993*. Dordrecht, Kluwer: 79-153

Onn, Farid M. 1980. *Aspects of Malay Phonology and Morphology*. Bangi: Universiti Kebangsaan Malaysia. [Published version of 1976 University of Illinois dissertation].

Pater, Joe. 1996. Austronesian Nasal Substitution and other N effects. To appear in Kager, René, Harry van der Hulst, and Wim Zonneveld (eds.), *The Prosody Morphology Interface*. Cambridge University Press

Pater, Joe. 1995. *Austronesian Nasal Substitution and other NC effects*. Draft of paper presented at proceedings of the Utrecht Prosodic Morphology Workshop.

Prince, Alan S. and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms, Rutgers University and University of Colorado at Boulder. [Forthcoming MIT Press].



# REDUPLICATION IN THE M'NONG LANGUAGE

Dinh Le Thu

*Viet Nam National University  
Ho Chi Minh City*

## 0 Introduction

The M'nong language belongs to the South Bahnaric language subgroup of Mon-Khmer, part of the greater Austroasiatic language family. M'nong is native to the southern and southwest parts of Dak Lak province and the northern part of neighboring Song Be and Lam Dong provinces, all in the Central Highlands of Vietnam. The M'nong language is spoken by 67,062 people in Viet Nam (1989 census). This paper on morphological reduplication in M'nong is based on the dialect of M'nong Preh spoken in the villages of Cu Jut and Dakgan in Dak Lak province.

## 1 The Forms of Reduplication in the M'nong Language

The M'nong language has several forms of reduplication. Each form expresses a distinct meaning and follows a phonetic rule. Reduplication in M'nong is used in some of the following ways: to describe an emotion or an attitude of the speaker in a particular situation, to indicate the degree or the characteristic nature of the base word, or to express onomatopoeia, among others. The reduplicants often consist of either a series of two-syllable or four-syllable units, though some three-syllable units do occur.

### 1.1 Two-Syllable Constructions

The first (or the second) syllable of a two-syllable series generally is a free root and does not change as reduplication occurs, and its meaning also remains the same. This reduplication occurs mainly with adjectives, verbs, nouns and descriptive roots, but it also occurs to a lesser degree with other classes of roots. In onomatopoetic words, the reduplicated syllable has no lexical meaning. There are two types of reduplication which form two-syllable constructions: total and partial reduplication.

#### 1.1.1 Total Reduplication

Total reduplication involves the complete repetition of the original free root. The free root has a syllable structure which falls into one of three structural types: an open syllable (ending in a vowel), a sonorant syllable (ending in a nasal or semivowel), or a closed syllable (ending in a unreleased stop). The following are examples.<sup>1</sup>

---

<sup>1</sup> The script used in this paper is based on the Vietnamese Quoc Ngu alphabet, commonly used to represent other indigenous languages spoken in Vietnamese. Many consonants and are pronounced as they are in English, while the vowels used in this paper require some explanation: 'a' is IPA /a/, 'â' is /ʌ/, 'e' is /ɛ/, 'ê' is /e/, 'i' is /i/, 'o' is /ɔ/, 'ô' is /o/, 'o' /ʌ:/, 'u' is /u/, 'u' /i/. Among the notable exceptions for consonants, 'ng' is /ŋ/ and 'nh' is /ɲ/. Long vowels can be indicated by a line above.

- a) Open syllables: *bro' bro'* 'continuously', *bru bru'* 'homeless, neglected', *mhaor' mhaor'* 'every afternoon'
- b) Sonorant syllables: *blao blao* 'stagger, reel', *iɔm iɔm* 'quietly, softly', *blun blun* 'very fast', *hol hol* 'smoulder', *vir vir* 'mooching around/moping about', *đưng đưng* 'flame, blaze'
- c) Closed syllables: *bik bik* 'thumps', *groc groc* 'flock', *hop hop* 'open the mouth wide, gape'

The initial consonant of a reduplicated syllable can be a single consonant or a consonant cluster. The combinations of bilabial voiced stops with liquids such as: [br], [bl], [-br], [-bl] and the single consonant [r] are particularly common in the reduplicated root. Other consonant clusters found in reduplicated forms include: [gr], [gl], [kr], [kl], [dr], [pl], [tr], [nd], [n'h], [r'h], among others.

### 1.1.2 Partial Reduplication

This reduplication can be divided into several types: (a) changes of the initial consonant, (b) changes of the vowel, (c) changes of the final consonant, and (d) changes of both the vowel and final consonant. Each type is discussed and illustrated below.

#### (A) *Changing of the initial consonant*:

In this construction, the vowel and the final consonants of the syllable remain unchanged, while the initial consonant or consonant cluster is changed according to a particular pattern. For the majority of the data fitting this form, the initial consonant of the first syllable is a liquid, either an [r] or an [l], and the second syllable can begin with any consonant. The following are examples of these alternations.

- r - b** : *rik bik* 'swarm, teem - insects', *rôc bôc* 'whisper', *rok bok* 'in a hushed tone'
- r - c** : *rîp cîp* 'sound of small bird- tweet twect', *rûc cûc* 'bend over some work'
- r - d** : *rup dup* 'of house : lowly, shabby', *run dun* 'stooping'
- r - h** : *raâng haâng* 'scorching, burning-hot', *ro'k ho'k* 'wait'
- r - kh** : *rik khik* 'giggle', *rok khok* 'roar of laughter'
- r - ng** : *ro'k ngok* 'swarming, teeming (with people)', *rôc ngôc* 'clean, tidy'
- r - s** : *rap siap* 'rustling of leaves, of gravel', *rap suap* 'rustle, rustling'
- r - bl** : *rip blip* 'shine, glitter of eyes', *rêp blêp* 'twinkle'
- r - kl** : *ruk kluk* 'silence of nature', *răk klăk* 'in silence, quiet of people'
- l - h** : *lût hut* 'flying low', *lûq huq* 'sound of agreement'
- l - v** : *lah vah* 'rock', *lûng vuñg* 'supple of movements'

#### (B) *Changing of the vowel*

(B.1) Change in the pitch only: This change can be predicted for the majority of the data on the basis of the vowel pitch. The tongue height never changes between the first and second syllable, but the pitch always changes from high to low between the two syllables. The change can also be considered as a change from a front vowel to a back vowel with the same tongue height. This applies equally whether the vowel is long or short. Some examples follow.

a. Alternation between long vowels [i-u] and short vowels [ĩ – ũ]:

**i - u** : *plip plup* ‘chubby, plump’, *hik huk* ‘untidy’, *ntik ntuk* ‘be hanging in the air’,  
*plih pluh* ‘wag’, *bit but* ‘shake, vibrate’

**ĩ - ũ** : *dĩn dũn* ‘stooping, bending’, *cĩt cút* ‘sound of pounding rice’, *bĩt bút* ‘smile, smiling’

b. Alternation between long vowels [e–o] and short vowels [ě-ǒ]:

**e - o** : *en on* ‘feeble, weak’, *lo le* ‘be abandoned, helpless’, *ceq coq* ‘timid; abashed’,  
*eng ong* ‘be furious; irascible’

**ě - ǒ** : *ěm ǒm* ‘threaten’, *krěp krōp* ‘noise of ox cart running on stone-road’

(B.2) Change in the tongue height only: This is a second type of vowel change between the high and close vowels [i, u] and the one low and open vowel [a]. Some examples for the alternation [i-a] and [u-a] follow.

**i - a** : *brim bram* ‘closely-planted’, *mhip mhap* ‘very dirty, soiled’, *gik gak* ‘strut; go slowly like an important person’, *suit suat* ‘dark, gloomy’

**u - a** : *hur har* ‘thirst for, hunger for’, *bru bra* ‘be dispersed, flee in all directions’, *rju rja* ‘stay too long, linger’, *ruq raq* ‘drip drop’

#### (C) *Changing of the final consonant*

In this case, the final consonant of the first syllable must be a voiceless stop: [p], [t], [c], [k], [q] or a voiceless glottal fricative [h]. The final consonant occurring in the second syllable is either a liquid [l] or [r], a nasal [m], [n], [nh], [ng], or it is zero.

**p - l** : *nōp nōl* ‘durable, be lasting’

**c - l** : *khoc khol* ‘tattered, in rags’, *khuc khul* ‘beat about the bush’

**k - l** : *hok hol* ‘in a mess’, *hik hil* ‘playful, mischievous’

**p - r** : *blip blir* ‘boast, talk big’

**t - r** : *ndēt ndēr* ‘helpless’

**c - r** : *ngăc ngăr* ‘nimble, briskly’, *hōc hōr* ‘panic-stricken’

**k - r** : *gok gor* ‘for a long time’

**p - ng** : *brip bring* ‘a very large quantity’

**t - ng** : *hit hing* ‘burning, hot’

**c - m** : *grăc grăm* ‘honest, good’

**q - n** : *luq lun* ‘grow a paunch’, *dhiq dhin* ‘fat, corpulent’

**h - nh** : *bhioh bhionh* ‘have the gift of the gab, loquacious’, *jojh jojh* ‘uneven’

**h - l** : *bhiah bhial* ‘dripping wet’

**t - zero** : *vēt vē* ‘confused, embarrassed’, *kut ku* ‘cuckoo’

Of the different combinations listed above, the most common pairs found are [k-l], [c-l], [c-r], [t-r]. These account for 51.8% of all final consonant changes found in our field work data.

*(D) Changing of both the vowel and the final consonant*

There seems to be no regularity or predictability to either the vowel quality or the final consonant change. However, there is a tendency for the second syllable to be more sonorant than the first syllable.

*blăk blo'* 'wander', *răk rong* 'take care', *roh rai* 'disperse'  
*blênh blu* 'appear and disappear', *laq lêng* 'roll of a drum'

One syllable words can also reduplicate just the initial consonant to form a presyllable. The first syllable is more sonorant than the second syllable.

<i>bok</i> 'white'	>	<i>bo'bok</i> 'off white, whitish'
<i>duh</i> 'hot'	>	<i>do'duh</i> 'luke warm'
<i>nhot</i> 'tough'	>	<i>nhɔ'nhot</i> 'slightly tough'
<i>kot</i> 'strong, thick'	>	<i>kakot</i> 'a little strong/ thick'

**1.2 Four-syllable constructions**

Four-syllable constructions occur with two-syllable free roots, which may be completely reduplicated or partially reduplicated. The reduplication formula is expressed with the following symbols.

- A- one syllable in the root
- B- a second syllable in the root different from A
- A'- partial reduplication of A
- B'- partial reduplication of B
- x- a syllable, completely different from A or B
- x'- partial reduplication of X

(A) **AB > ABAB** (A two-syllable root is reduplicated completely.)

<i>mbrêh nhaâp</i> 'scintillate'	>	<i>mbrêh nhaâp mbrêh nhaâp</i>
<i>blê blêp</i> 'shine'	>	<i>blê blêp blê blêp</i>
<i>du hŭ</i> 'in turn'	>	<i>du hŭ du hŭ</i>

(B) **AB > AABB** (A and B are both free roots, each one is reduplicated completely. AA always occurs before BB.)

<i>kho ao</i> 'clothes'	>	<i>kho kho ao ao</i>
<i>ur sai</i> 'wife and husband'	>	<i>ur ur sai sai</i>
<i>en on</i> 'mincing'	>	<i>en en on on</i>
<i>go'm ngo'i</i> 'laugh and speak'	>	<i>go'm go'm ngo'i ngo'i</i>
<i>guq dôk</i> 'sit and stand'	>	<i>guq guq dôk dôk</i>

(C) **AB > ABAB'** (AB is a two-syllable compound word.)

<i>êng ang</i> 'strange'	>	<i>êng ang êng rang</i>
<i>cê lêt</i> 'small, unimportant'	>	<i>cê lêt cê lanh</i>

(D) AA' > AA'BB' (AA' is a two-syllable reduplicated root, and BB' is a partial reduplication of AA').)

<i>luk lăk</i> 'disorderly'	>	<i>luk lăk tăk tai</i>
<i>vêt vê</i> 'confused'	>	<i>vêt vê cū cuăng</i>
<i>êt êl</i> 'compel'	>	<i>êt êl dêl dôl</i>

(E) AA' > xAxA' (AA' is a two-syllable reduplicated root, x is an open syllable: *mô*, *ma*, *dê* etc.)

<i>laq lăng</i> 'not serious'	>	<i>mô laq mô lăng</i>
<i>ngo'q ngor</i> 'lazy'	>	<i>mô ngo'q mô ngor</i>
<i>blem blom</i> 'corpulent, fat'	>	<i>ma blem ma blom</i>
<i>-brôc -bro'c</i> 'timid'	>	<i>dê -brôc dê -bro'c</i>
<i>eng ong</i> 'be furious'	>	<i>dê eng dê ong</i>

## 2 Meanings of reduplicated forms

Two factors which influence the meaning of a construction are the word class of the free root and the form of reduplication. In each construction, the form can express a distinct meaning which depends on the nature of the free root, for example, whether it is a noun, verb, adjective, or descriptive root.

There is often stark contrast between the initial consonants, the vowels, and the final consonants of the two syllables in partial reduplication formations. This contrast in sounds makes the reduplicated word very rich in imagery and description. If the root is a verb, the partial reduplication formation expresses the action while at the same time also expressing the means by which that action occurs. Thus, *rêp blêp* not only means 'twinkle', but it also expresses the means by which this process occurs (i.e. it is the periodic repetition of 'twinkle' with different degrees of light). If the root is an adjective, the partial reduplication formation expresses the meaning more precisely. Thus, *vil rdit* 'round' becomes *vil rdit* 'perfectly round'.

In the M'nong language, the pitch or method of articulation of a vowel or a consonant in total or partial reduplication often has a relationship with the meaning of the construction. Constructions often resemble natural sounds. The shape of the articulators can be associated with the shape, form, or appearance of things. In addition, pitch can indicate colors, varying degrees of light or darkness, or similar distinctions. This is similar to the reflection of the natural world in the sound of music.

It is hard to define exactly how many types of meaning reduplication expresses in the M'nong language. However, based on data collected in the field and from the M'nong-Vietnamese dictionary, the most common meanings are listed and described below.

**2.1** There are a large number of onomatopoetic words formed as a result of total reduplication: *crak crak* 'sound of the rain', *dûng dûng* 'sound of the drum', *hurê hûr* 'sound of the wind', *krao krao* 'sound of baby crying', and other similar types. The original free root of the onomatopoeic words is often found with the initial consonant /p/: *păng păng* 'sound of firing', *pin pin* 'sound of car horn', *plop plop* 'sound of applause',

*pak pak* ‘sound of cutting down trees’, and so on. There are also a large number of two-syllable onomatopoetic words formed as a result of partial reduplication.

*rek phek* ‘noise of firewood burning’  
*rek bhek* ‘noise of mouse eating rice’  
*ing iang* ‘noise of frog croaking’  
*kêng kõng* ‘noise of the bells ringing’

**2.2** Noun constructions are commonly used for the names of plants or animals: *ka ka* ‘beech-tree’, *ket ket* ‘tree with rugged leafs’, *bul bul* ‘small ophi cephalidae fish’, and others. Several names of animals and plants are also a result of partial reduplication.

*kô kim* ‘butterfly’, *mrok mro* ‘cricket’,  
*geng gong* ‘praying mantis’, *hêp tu ru* ‘grass rush’.

**2.3** Two-syllable total reduplication formations with a noun as the original free root express the meaning of plurality ‘every x’.

<i>ngih</i> ‘house’	>	<i>ngih ngih</i> ‘every house’
<i>buynh</i> ‘person’	>	<i>buynh buynh</i> ‘every person’
<i>ntôk</i> ‘place’	>	<i>ntôk ntôk</i> ‘every place’
<i>ôi</i> ‘morning’	>	<i>ôi ôi</i> ‘every morning, morning after morning’
<i>nar</i> ‘day’	>	<i>nar nar</i> ‘every day, day after day’
<i>khay</i> ‘month’	>	<i>khay khay</i> ‘every month, month after month’

**2.4** Verb constructions carry a sense of repetition and continuation.

<i>đút</i> ‘to nod’	>	<i>đút đút</i> ‘nod repeatedly’
<i>par</i> ‘fly’	>	<i>par par</i> ‘fly repeatedly’
<i>kdăt</i> ‘jump’	>	<i>kdăt kdăt</i> ‘jump repeatedly’

**2.5.** Two-syllable total reduplicative formations with adjectives or verbs as the original free roots express lightened meaning (e.g., ‘a little x’, ‘somewhat x’).

<i>no'iñ</i> ‘easy’	>	<i>no'iñ no'iñ</i> ‘a little easy’
<i>m'bro'i</i> ‘slow’	>	<i>m'bro'i m'bro'i</i> ‘a little slow’
<i>rgai</i> ‘thin’	>	<i>rgai rgai</i> ‘a little thin’
<i>bêñ</i> ‘familiar’	>	<i>bêñ bêñ</i> ‘somewhat familiar’
<i>bo</i> ‘to love’	>	<i>bo bo</i> ‘to love less’

This lightened meaning is also expressed by two-syllable partial reduplication constructions that can be categorized semantically.

(A) Lightening of smell and taste

<i>ciăt</i> ‘salty’	>	<i>cia ciăt</i> ‘a little bit salty’
<i>nhôt</i> ‘sweet’	>	<i>nhô nhôt</i> ‘a little sweet’
<i>sõt</i> ‘tasteless’	>	<i>so' sõt</i> ‘plain’

<i>kah</i> 'delicious'	>	<i>ka kah</i> 'tasty'
------------------------	---	-----------------------

## (B) Lightening level of character, colour, temperature

<i>dăng</i> 'hard'	>	<i>da dăng</i> 'a little hard'
<i>kloh</i> 'clean'	>	<i>klo klot</i> 'a little clean'
<i>ndrêh</i> 'green'	>	<i>ndrê ndrêh</i> 'light green'
<i>guh</i> 'red'	>	<i>go' guh</i> 'light red, reddish'
<i>duh</i> 'hot'	>	<i>do' duh</i> 'warm, a little hot'
<i>kăt</i> 'cold'	>	<i>ka kăt</i> 'cool, chilly, a little cold'

## (C) Lightening of size, quantity, weight

<i>jéê</i> 'small'	>	<i>jê jéê</i> 'a little small'
<i>téh</i> 'large'	>	<i>tê têh</i> 'somewhat large'
<i>prêh</i> 'tall'	>	<i>prê prêh</i> 'somewhat tall'
<i>gleh</i> 'short'	>	<i>gle gleh</i> 'somewhat short'
<i>jõk</i> 'heavy'	>	<i>jo' jõk</i> 'somewhat heavy'
<i>thông</i> 'light'	>	<i>thô thông</i> 'somewhat light'
<i>õk</i> 'much, many'	>	<i>o' õk</i> 'not too much/ many'

**2.6** The two-syllable partial reduplication form with the rhyme [-en ] in the reduplicant occurs in nouns, verbs, and adjectives, giving them general meanings with negative connotations.

<i>tok</i> ' hair'	>	<i>tok ten</i>
<i>môp</i> 'hats'	>	<i>môp men</i>
<i>sa</i> 'to eat'	>	<i>sa sen</i>
<i>nti</i> ' to study'	>	<i>nti nten</i>
<i>ueh</i> ' beautiful'	>	<i>ueh uen</i>
<i>bok</i> 'white'	>	<i>bok ben</i>

**2.7** Some partial reduplication forms occur with [-a-] between the two syllables. The first syllable is an adjective root, while the final syllable is a reduplicant with the initial consonant [r-]. This reduplication formation intensifies the meaning.

<i>jõk</i> 'heavy'	>	<i>jõk a rõk</i> 'very heavy'
<i>prêh</i> 'tall'	>	<i>prêh a rêh</i> 'very tall'
<i>gleh</i> 'short, low'	>	<i>gleh a reh</i> 'very short, very low'
<i>toyh</i> 'big'	>	<i>toyh a roh</i> 'very big'
<i>õk</i> 'many, much'	>	<i>õk a rõk</i> 'too many, too much'

**2.8** Four-syllable reduplication forms emphasize the meaning more fully. The following are examples.

## (A) A sense of plurality:

<i>kho ao</i> 'clothes'	>	<i>kho kho ao ao</i> 'many clothes'
<i>ngih vo'l</i> 'houses'	>	<i>ngih ngih vo'l vo'l</i> 'many houses'

(B) A sense of repetition:

<i>sa nhêt</i> 'eat and drink'	>	<i>sa sa nhêt nhêt</i> 'eat and drink again'
<i>go'm ngo'i</i> 'laugh and speak'	>	<i>go'm go'm ngo'i ngo'i</i> 'laugh and speak again'

(C) A sense of degree of character or characteristics:

<i>en on</i> 'walk or behave flirtatiously or affectedly'	>	<i>ren en ron on</i> 'walk or behave more flirtatiously or affectedly'
<i>iep iop</i> 'lean, thin'	>	<i>rep iep rop iop</i> 'very thin with long arms & legs'
<i>săng</i> 'obscene'	>	<i>săng blah săng blo'i</i> 'completely obscene'

### 3 Concluding Thoughts

It is worth noting that the M'nong language sometimes deploys affixation, the main function of which is to change the class or subclass of the word to which it is affixed: e.g., it may turn a verb into a noun, an intransitive verb into transitive verb, etc. The reduplicative formations and their meaning probably have a relationship with affixation in M'nong. It is the hope of the author to touch upon this subject in the future.

### References

Elizabeth, M. Bunker. 1964. Bahnar Reduplication. *Mon-Khmer Studies I*: pp. 99-119 . Saigon: SIL.

Moravcsik, Edith A. 1978. Reduplicative constructions. In *Universals of human language, Vol. 3. Word structure*, ed. by Joseph H. Greenberg, pp.297-334. Stanford, Calif.: Stanford University.

Bùi, Khánh Thé and Đinh Lê Thu 1994. *Từ điển M'nong-Việt* (A Mnong-Vietnamese Dictionary). Daklak.

Hoàng Văn Hành. 1985. *Từ láy trong tiếng Việt* (A dictionary of reduplicative forms in Vietnamese). Hanoi: Social Sciences Publishing House.

Hoàng Văn Hành. 1987. *Góp phần luận giải về cách cấu tạo từ láy trong các ngôn ngữ Mon-Khmer* (An explanation of reduplication among Mon-Khmer languages). *Ngon ngữ* (Language), 1987, no. 1 and 2.

Hoàng Văn Hành. 1993. *Từ láy trong các ngôn ngữ Katuic ở Việt Nam* (Reduplication among the Katuic Languages of Vietnam). *Ngon ngữ* (Language), no. 4.

Nguyễn Tài Cẩn. 1996. *Ngữ pháp tiếng Việt* (A Vietnamese Grammar). Hanoi: National University in Hanoi.

Nguyễn Thịện Giáp. 1995. *Từ vựng học tiếng Việt* (Vietnamese Lexical Studies). Hanoi: Đại học và THCN, Hà Nội.

Nguyễn Thị Ánh Miêu. 1985. *Hiện tượng láy trong tiếng m'nong* (The Phenomenon of Reduplication in Mnong). Dissertation. Ho Chi Minh: Central University of Ho Chi Minh City.

# PROSODY AND THE SEGMENTATION OF MALAY DISCOURSE

Zuraidah Mohd Don

*University of Malaya*

*<zuraida@um.edu.my>*

## 1 Introduction

The aim of this paper is to examine the issues of segmentation in continuous discourse and present a theoretical framework that somewhat eases its segmentation into fragmented parts. The difficulty encountered in segmenting a stretch of speech into separate tone groups provides the impetus for carrying out research of this nature. The paper presents arguments for segmenting discourse into units of speech whose boundaries are defined by audible prosodic cues. The conclusion is that this less restrictive framework enables the identification of prosodic cues in segmentative work and the roles they play in discourse development.

The standard approach to the description of intonation, especially in the British tradition of intonation, is to establish a unit of phonological organization within which the nucleus or focus can be defined. The assignment of tonal features in turn depends on the necessity of having appropriate information points pre-established. The recognition of this central unit in the study of intonation is described succinctly by Scuffil (1982:34) as follows:

All analyses of intonation postulate a unit which is central in the sense that it provides the framework within which intonational features can be described.

These units share a theoretical orientation and characterize some units of intentional description, the neutral and unmarked case coinciding with a clause. The intimate relationship between prosody and segmentation of speech is expressed in the words of Gardiner (1977:4) who postulates that intonation segments utterances into 'phrases signaling to what extent the phrases are related to one another and element within the phrase is the center of attention.'

The fact that it is impossible to utter an extended stretch of speech without some kind of break, and that it is impossible for the hearer to interpret what is said unless what is perceived is chunked into manageable units makes segmentation into divisible unit obligatory. Nevertheless, the decision as to how the verbal content of his discourse should be segmented is optional in the sense that it lies with the speaker. The belief that segmentation of discourse is prosodically identifiable and that often (but not always) segmentation is based on speaker decision in pursuit of a purpose provides the impetus for the investigation into the relationship between segmentation and the role that the segmented chunks play in discourse development.

## 2 A proposed theoretical framework

Bolinger (1989) regards segmentation as 'universal' such that 'all languages use intonation, rate, and pause to mark divisions' (p.82). For example in Cayuvava, 'a fall and a trail off after the final accent signals the wish not to continue and a maintenance of the level of the final accent indicates incompleteness' (1982:82). According to Bolinger a speaker indicates a break in the discourse by a shift in pitch, in particular a drop in pitch. Likewise Brazil (1975, 1978), Lehiste (1975) and Bolinger (1989) note that the speaker indicates the beginning of a new utterance or topic by a change in prosody, i.e. raising the pitch height and/or increasing loudness.

The difficulties encountered when trying to segment Malay discourse using a definition of the nucleus and its domain which was finely tuned for English have led to the abandonment of dividing stretches of speech into tone groups. In practice, however, tone-group boundaries are sometimes not set off by audible prosodic cues and that segments which are demarcated by audible breaks may not contain a nucleus (Knowles: 1991). Crystal (1969) himself admits that he sometimes takes recourse to syntactical or semantic criteria to place boundaries. What seems to make segmentation difficult is the presence of hesitation phenomena brought about by planning and production difficulties. The very high proportion of performance errors in Brown et al's (1980) data of spontaneous informal conversations, e.g. false starts, hesitation, slips of the tongue, overlapping and incomplete utterances contribute to making segmentation problematic.

When a speaker encounters problem in producing what he wants to produce he may be forced to pause or lengthen a syllable at an inappropriate place, thereby disrupting the prosodic flow of the utterance. This break in the prosodic flow causes a stream of speech to be realized as consisting of fragmented chunks whose boundaries may not coincide with syntactic boundaries. A speaker may, for example, pause before reaching the nucleus resulting in segments which have no nucleus. Although intonationally insignificant, these planning units are important in the sense that they fulfill the speaker's interactional purpose, i.e. gaining him time to prepare his subsequent contribution.

The researcher has therefore adopted a less restrictive approach to segmentation; division into segmented parts is based on the presence of the prosodic cues which causes a break in the prosodic flow of utterances. The methodology for identifying the segmented parts thus rests on identifying the audible prosodic features whose presence contributes to making the units they bound hearable as segmented chunks. The question of whether these segmented parts contain a nucleus or not is not at issue here because the aim is not to describe the intonation patterns of Malay. After the identification of these segments, the contribution they make to discourse development will be examined. These segmented parts which I refer to as 'speech units' can be realized as follows:

1. a unit of information containing an item of information which is made prominent by a combination of prosodic cues, i.e. a speech unit containing a nucleus. The study interprets the term 'information' broadly so as to include not only content-information but also information of a social, pragmatic and interactional in nature.
2. a unit of performance such as a slip of the tongue, a false start, an incomplete lexical item, etc.

Although most linguists recognize the tone group as the basic intonational unit only Crystal (1969:204-7) gives the most complete discussion of the phonetic cues which signal the boundaries of the tone group. He claims that in most cases its boundaries can be determined by the following phonetic/phonological cues:

1. audible change in pitch at the level of boundary depending on the direction of the nucleus,
2. the presence of audible pause, final syllable lengthening or aspiration at the end of tone-group. (1969:204-7)

Crystal further adds that in the absence of these phonetics cues one can then resort to relying on 'grammatical or semantic criteria to place the boundary' (p. 207), but such cases are few. Couper-kuhlen (1986) and Cruttenden (1986) list out the following external criteria for boundary identification: pause location (either real or potential), final syllable lengthening, rhythmic discontinuity and the presence of anacrusis. Another boundary signal is the relative tempo of unstressed syllables.

To overcome the problem of segmentation, Ladd (1986) proposes the recognition of two types of intonational phrasing: Major Phrase (MP) and Tone Group (TG). The former is delimited by the presence of overt phonetic cues such as pause, anacrusis, syllable lengthening, pitch change, change of tempo, etc. whose presence is agreed upon by most linguists. Tone group is 'merely a structural unit of phonology – the domain within which the nucleus is defined', whose existence is identified solely on the basis of tonal structure (i.e. the presence of nucleus). These two types of domain form hierarchical structure and that they are recursive in the sense that a constituent is allowed to dominate a constituent that is higher in rank, parallel to the situation in syntax. In Ladd's opinion this theoretical framework will help overcome problem in cases where no phonetic cues bound segments whose internal structure satisfies the minimum requirement of an independent tone group.

Pike (1962) presents a detailed description of how the boundaries of phonological units can be identified using rhythmic criteria. According to Pike while prenuclear unstressed syllables are characterized by crescendo loudness and faster tempo due to the relative shortness of the syllable, post nuclear syllables are marked by decreasing level of loudness (decrescendo), lenis articulation (including devoicing) and relative lengthening of the syllable. O'Connor (1973) adds that one can distinguish prenuclear syllables from post nuclear unstressed syllables by relative tempo such that the former is uttered relatively faster than the latter.

Table 1 below draws together the various observations that linguists have made as regards the phonetic criteria that contribute to making continuous stretches of speech realized as separate units of prosody.

**Table 1:** *External Cues to Tone group boundaries*

Linguists	Pitch parameters	Tempo parameters	Other parameters
Crystal/ Cruttenden/ Couper- Kuhlen	change of pitch level or pitch direction	pause; final syllable lengthening; anacrusis; rhythmic discontinuity	semantic/ syntactic criteria
Pierrehumbert	boundary tones	pause; final syllable lengthening	
Pike		relative allegro tempo of pre-nuclear unstressed syllables; relative lengthening of post nuclear unstressed syllables	crescendo loudness of pre-nuclear unstressed syllables; decrescendo loudness of post nuclear unstressed syllables; lenis devoicing
Ladd	pitch change	anacrusis; final syllable lengthening; pause	major group-marked by audible prosodic cues; minor group- marked by the nucleus
O'Connor		prenuclear unstressed syllables uttered faster than post nuclear unstressed syllables	

### 3 Truncated segments

Based on the external cues described above, a preliminary analysis of the data reveals the occurrences of the following truncated segments whose audible prosodic cues at the boundaries give them the auditory effect of being cut off or incomplete. This incompleteness is brought about by the fact that these speech units are syntactically, semantically and/or intonationally incomplete. They can be categorized as follows:

1. incomplete speech units whose boundaries are set off by prosodic cues. Syntactically and semantically they are part of the subsequent chunk of speech from which they are separated
2. abandoned speech units whose realization is the consequence of unfluent speech. Unlike the former, the speaker leaves the segment incomplete and starts fresh. The boundaries are demarcated by audible prosodic cues. They do not cohere syntactically or semantically with the segment following them
3. parenthetical speech units which interrupt the prosodic flow of the primary utterance. The boundaries of these ‘wedged in’ segments are set off by prosodic cues

### 3.1 Incomplete Segments

Extracts 1 through 3 present a dilemma in intonation analysis. It does not require a very close examination to notice that the highlighted truncated segments of speech (arrowed) in Extracts 1, 2 and 3 below share similar characteristics such that they are marked off from the segments following them by a pause, and they are intonationally incomplete in the sense they lack a nucleus. The speaker breaks off before reaching the nucleus, a point that is illustrated by the absence of the nucleus (noted by the absence of capitalization).

#### Extract 1

1B kita more `INDIVIDUALISTIK (0.23)

< >(0.76)

< >(0.63)

→ dia: (0.42) daripada (0.43)

<>(0.55) < >(0.35)

<>(0.41)

< >L

< >p

↑`AWAL lagi `TADI: didik secara berkumpul

< >(1.15)

<>(0.46)

< >(1.2)

B WE are more individualistic. From young, they have been brought up in groups.

#### Extract 2

2A how is it different

< >(0.93)

→ ataupun ada (0.29) `KESAMAAN

< >(0.58)

<>(0.27)

A How is it different or is there any similarity?

#### Extract 3

→ 3B kita (0.36) pun tukar MODEL

<>(0.25) < >(1.2)

<>L < >L

B So we change the model

Although the lack of nucleus contributes to making the highlighted segments intonationally insignificant, they are significant in the sense that the pause which bounds them affects the prosodic flow of the discourse and causes discontinuity. Pauses of this type are generally regarded as hesitation phenomena whose occurrences are quite common in the corpus analyzed. The relative lengthening of the final syllable further supports the assumption that the speaker is planning what to say next.

Despite the break, the speaker maintains a link with what follows by producing the final syllable of *dia* 'they' with a level pitch which neither falls nor rises, thereby communicating the intention to continue. Although the overall pitch of *dia* is low, the speaker signals the wish to continue by keeping the pitch level and lengthening the final syllable. One can easily detect that the speaker is facing planning difficulties by his relatively slower speech rate, i.e. 550 msec. in producing a two syllable utterance in comparison with 350 msec. that he takes to utter a four syllable utterance located subsequent to it. The rush to produce *daripada* 'from' causes the speaker to pause to replan what he wishes to say. Again he indicates incompleteness by producing the final syllable with a non-low pitch level. Besides the pause which causes *daripada* to be broken off from the subsequent talk, the speaker also distinguishes the latter from the former by a step-up in pitch when producing *awal* 'young', the initial word in the subsequent segment.

Likewise in Extracts 2 and 3 *ada* ‘there is’ and *kita* ‘we’ respectively are separated from the subsequent segments by a brief pause, yet are prosodically bound to them by their terminal pitch which stays non-low. This cue is used by the speaker as a floor keeping strategy while planning his utterance. Gumperz (1992:235) calls this type of pitch that suggests completeness ‘a holding intonation’. In a recent study on the management of talk, Local (1992:275) observes that these pitch characteristics along with loudness are used by speakers to hold turn or project more talk by the current speaker (cf. Zuraidah Mohd Don: 1996, 1998, 2006).

It is inevitable that in producing spontaneous speech speakers may produce segments whose prosodic boundaries do not coincide with syntactic-boundaries. This is so because unlike the latter, a phonetic entity is basically some unit of performance. Speakers, for example, pause in the middle of noun phrases, verb phrases and other close knit syntactic constituents, changing pitch height and direction on syllables for the purpose of keeping the floor while planning what to say next. In Extracts 4 and 5 below, a close knit noun phrase *dua ketul besi* '2 pieces of iron' is separated from each other by a pause, defying syntactic cohesion, and in Extract 6 a modal *boleh* 'can' is separated from its verb *membawa* 'bring'.

#### Extract 4

4B kalau kita `AMBIK (0.32)  
 → a: `DUA (0.24) ketul `BESI  
     <    >CRES  
 B If we take two pieces of iron

### Extract 5

5B dan yang `TUA baru `DARJAH (0.28)  
 < >(1.46)  
 < >f

`EMPAT  
 < >(0.34)  
 < >H  
 < >CRES  
 B and the eldest is only in standard four

## Extract 6

6B maknanya kalau saya `BOLEH (0.34)

&lt; &gt;f

a: `MEMBAWA (0.32) a:

&lt; &gt;(0.56)

seorang `JEPUN (0.28)

&lt; &gt;f

ke `MALAYSIALAH

&lt; &gt;f

B Meaning if I can bring back a Japanese to Malaysia

In Extract 4, *dua* 'two' is separated from a quantifying noun by a brief pause. The final syllable is relatively high signaling completion and the word is uttered with crescendo loudness. Likewise in Extract 5 the last syllable of *darjah* 'standard' is uttered with a pitch higher than the preceding one and accompanied by forte loudness. In Extract 6 the speaker's short utterance is segmented into four prosodically marked segments. The separation of the prior segment from the subsequent one indicated by a brief pause is done at a syntactically inappropriate point. The word 'boleh' (can) is uttered loud and with a non-low pitch height signaling incompletion. The occurrence of 'a:' whose level pitch seems to serve no function other than to maintain the speaker's turn while he plans his utterance further supports the assumption.

These 'planning units' as Cruttenden (1986) calls them are the realizations of planning problems which often occur in natural spoken data and are usually separated from the corrected utterance by a brief pause. According to Brown et al (1980:68) these 'search pauses' which are brief in duration (i.e. usually between 0.28 and 0.38 seconds) are the results of planning problems and such an assumption is correct since the units whose boundaries they demarcate are incomplete syntax.

### 3.2 Abandoned Segments

Abandoned segments are actually made up of incomplete syntax. What is observed in the data is that the speaker usually goes back to the beginning of abandoned segments and reproduces them again, this time more fluently, e.g. *mama jum* in 7A, *maksud sa* in 8A and *bukanan* in 9A:

## Extract 7

7A `MAAFLAH (0.15)  
 < >(0.43)  
 gurau sikit `DOKTOR (0.3)  
 < >(1.03)

a→ mana jum  
 < >(0.39)  
 < >L  
 < >p

b→ ↑mana jumpa `JEPUN tu  
 < >(0.41)  
 < >(0.59)  
 < >f

A Sorry. Just kidding doctor. Where did you meet the Japanese?

## Extract 8

a→8A maksud sa (0.40)  
 < >(0.43)  
 < >DEC  
 < >L

↑`MAKSUD saya:  
 < >(0.51)  
 < >f

sekarang ni kereta `PROTON  
 < >(0.40)  
 < >(0.39)

A What I mean is that nowadays PROTON cars

## Extract 9

9A bukanan (0.21)  
 < >(0.43)

↑bukan nak congkil `PERIBADI  
 < >(0.36)  
 < >(0.97)

A Not that I want to inquire into your private life

In Extracts 7, 8 and 9, each of the speakers produces an incomplete utterance, pauses briefly and proceeds to repair the preceding contribution by repeating it, and thereon continues with the utterance. In Extract 7, the speaker distinguishes the speech 'error', *mama jum* (arrowed a) 'where' from the 'corrected' segment *mana jumpa Jepun tu* 'Where did you meet the Japanese' by uttering the repair with a pitch-step up and producing the latter relatively louder than the former. This prosodic marking is also accompanied by a shift in tempo whereby the corrected utterance is uttered at a faster pace than the prior talk, i.e. the speaker takes 410 msec. to utter a seven- syllable utterance as compared to 390 msec. to produce a three- syllable utterance.

In 8, again the change in prosody contributes to the hearing of a stretch of speech as two segmented chunks. The speaker signals the start of a new contribution by a marked increase in loudness and a pitch step-up from a low-pitched *sa* to a relatively high-pitched *maksud* 'mean'. The brief pause which contributes to the abandoned segment hearable as a segmented portion gives the speaker time to reformulate his utterance. In 9 the contribution which initiates repair *bukan nak* is separated from a prior segment by a brief pause and a noticeable shift in prosody. The repair begins with a correction of the mispronounced *bukanan* in combination with a marked increase in pitch height and loudness.

The self correction gives the impression that the speaker is consciously aware that he has not got right what he was trying to produce and repaired it. The speaker contextualizes the repair with certain prosodic cues in order to indicate its status as a new contribution. Sometimes instead of correcting the false start by repeating it, the speaker abandons this incomplete segment and proceeds to make a fresh start with a new utterance.

### 3.3 Parenthetical Segments

Parenthetical segments are brought about by the insertion of a secondary utterance within a primary one causing the latter to be divided into segmented chunks. Parenthesis requires clear prosodic chunking so that the embedded portion can be distinguished from the segmented primary portions. A preliminary analysis of the data reveals a number of these disruptive segments whose boundaries are clearly demarcated by audible prosodic cues.

A common example of this type of 'wedged in' segment found in our data is one that is introduced for the purpose of ensuring that the turn is kept yet allowing the speaker time to plan what to say next. This 'turn-keeping' and 'buying-time' strategy is clearly reflected in the following extracts whereby the speaker uses 'wedged in' phrases like *apa ni* 'what's this' in Extract 10, *apa tu* 'what's that' in Extract 11 to signal that he is in search of words to complete his prior talk and he still intends to continue.

## Extract 10

10A jadi: masa saya:

< >(0.55)  
 < >(1.22)  
 ↓apa ni: (0.32)  
 <>(0.44)  
 < >(0.76)  
 <H><L  
 ↑attachment `TRAINING  
 < >H  
 < >f  
 salah `SEBUAH a:  
 < >(1.41)  
 < >L  
 syarikat DISANA:  
 <l >l>(0.93)  
 < >L

A So when I .. what's this ... attachment training in one of the firms ... a firm there

## Extract 11

11A bila dah masuk dua kali salam dia sebutlah (0.21)

↓apa tu (0.41)  
 <l >l>(0.51)  
 < >p  
 dia punya nama: khulafah arrasyidin  
 <al >al>(0.29)

A After giving two greetings, he mentioned, what's that, the name of Khalifah Arrasyidin

## Extract 12

12A saya diberitahu:

< >(0.95)  
 bahawa apabila kita membaca  
 < > (1.27)  
 walaupun satu huruf (0.23)  
 < >(1.16)  
 kita akan dibagi: a:  
 < >(0.85)  
 <>(0.26)  
 ↓apa nama ni (0.23)  
 < >(0.30)  
 < >L  
 < >p  
 sepuluh kali ganda pahalalah  
 < >(0.17)

*apa ni* (Extract 10), *apa tu* (Extract 11) and *apa nama ni* (Extract 12) which are 'wedged into' the main utterance do not only disturb the syntactic and semantic cohesion of speech but also its prosodic flow. These parenthetical segments are incomplete by themselves. In these instances, they are set off from the primary utterance by a pause and their overall low pitch. In Extract 10, the speaker distinguishes the first segment *jadi masa* *saya* 'so when I' which is syntactically and semantically incomplete, from the embedded segment *apa ni* 'what's this' by a shift in speech rate and pitch. The lento tempo of the first segment gives the impression that the speaker is facing some sort of planning difficulty. The lengthening of *ya* allows him planning time; however, the speaker still needs planning time as a result of which he produces an utterance which not only breaks the prosodic flow of the prior talk but also its syntactic and semantic flow. This inserted segment which is uttered relatively fast and begins with a step up in pitch gives him planning time as well as allows him to keep his turn. The subsequent segment, i.e. 'attachment training' is set off from the inserted segment by a pause and a shift in pitch, i.e. with a pitch step up from low and level pitched 'ni'. It is also marked by a high overall pitch and forte loudness.

In Extract 11, 'apa tu' is separated from the prior talk by a brief pause of 210 msec. And the subsequent talk by a pause of 420 msec. There is also a step down in pitch from 'lah' the last syllable of the prior talk to 'a' the first syllable of the embedded segment 'apa tu'. The overall pitch is lower than the prior and subsequent segments. Likewise in Extract 12 the speaker introduces a new utterance 'apa nama ni' (what that is) before completing his earlier utterances and indicates a break in syntactic and semantic cohesion by demarcating the embedded segment with pauses. The overall loudness is relatively softer than the prior or subsequent talks. It is also uttered with an overall low pitch and a relatively faster tempo than the segments which bound it. The embedded segment in Extract 11 'apa tu' and in Extract 10 'apa ni' serve the same function respectively.

Another type of parenthetical segment found in our data is what is defined as ‘... a dependent satellite part of the utterance, wedged into a non-compact primary (frame) reference from which it differs’ (O’Connor (1973)). Parenthesis expresses a secondary communication. The parenthetical segments in Extracts 13 and 14 are introduced for the purpose of communicating secondary information. For example in 13 the speaker inserts ‘yalah kata orang tu’ (yes that’s what people say), and ‘tak silap saya’ (if I’m not mistaken) to communicate something other than what is communicated in the primary utterance. The ‘secondary communication’ realized in the data below is prosodically distinguished from the segmented chunks of the primary utterance.

### Extract 13

## Extract 14

14A semalam dilapurkan a: (0.43)  
 ↓tak silap saya (0.39)  
 ↑timbalan mentri kesihatan agaknya mengumumkan  
 akan mengadakan perang terhadap AIDS

A Yesterday it was reported that, if I'm not mistaken,  
 perhaps the Deputy Minister of Health will have  
 campaigns against AIDS

In Extract 13, the segment *yang sebenarnya* 'actually' is broken off from the subsequent segment *yalah kata orang tu* 'yes like what people say' of the utterance by a lengthy pause of 1120 msec. The latter is often used by a speaker when he wants to qualify that whatever is said in the subsequent utterance is described by some people as such. Although *ya* the final syllable of *saya* has a level pitch which neither rises nor falls, its loudness seems to trail off until it reaches inaudibility. There is a slight pitch step-up at the beginning of *yalah kata orang tu* to indicate the beginning of a new utterance. Although the final syllable *tu* uttered low, the pitch is level indicating that the speaker has not yet finished. Thus although the embedded segment is separated from the subsequent talk by a pause of 740 msec., the level pitch *tu* provides the link-up with the subsequent utterances which the prior segmented talk is part of. According to Bolinger (1986:82) the 'maintenance of the level of the final accent indicates the intention to continue'.

#### 4 Conclusion

The findings show the advantages of corpus-based study. Many of the observations described here would never have been noticed without the use of a sample of natural spoken Malay. The difficulties encountered when attempting to analyze Malay using a definition of the nucleus and its domain led to the abandonment of dividing stretches of speech into tone groups.

It is inevitable that in segmenting utterances into phonetically defined units, one would come across instances where boundaries of phonetic entity do not correspond with boundaries of syntactic constituent. This is because unlike the latter, phonetic entity is basically some unit of performance. Pauses, for example, may break up close unit grammatical constituents or bound false starts, slips of the tongue, or incoherent segments. In so far as pauses or some other prosodic cues arise from a real need for verbal planning, their occurrence can be regarded as a direct result of the fact that speech is produced in real time and real settings. To ignore the presence of the audible cues between segments means to ignore the natural phenomena whose presence are welcome as overt indications of processing activity. Prosodic segmentation is a potential correlate of almost any syntactic unit and can be employed by the speaker in accordance with his encoding strategy. The study shows that the subjects segment utterances into segmented chunks of various grammatical units, each playing its role in the development of the discourse, however small or insignificant.

## References

Bolinger, D. L. 1989. *Intonation and Its Uses*. London: Edward Arnold.

Brazil, D. 1975. *Discourse Intonation. Discourse Analysis Monographs 1*. Birmingham: ELR

Brazil, D. 1978. *Discourse Intonation II. Discourse Analysis Monographs 2*. Birmingham:ELR

Brown, G. Currie, K.L. and Kenworthy, J. 1980 *Questions of Intonation*. London:Croom Helm.

Couper-kuhlen, E. 1986. *An Introduction to English Prosody*. London:Edward Arnold.

Crystal, C. 1969. *Prosodic Systems and Intonation in English*. Cambridge:CUP.

Cruttenden, A. 1986. *Intonation*. Cambridge: CUP.

Gardiner, D. 1977. *Two assumptions in the Study of Intonation*. Paper read at the LSA Annual Meeting, Chicago.

Gumperz, J.J. 1992. 'Contextualisation Revisited', in Duranti, A. and Goodwin, C. eds., *The Contextualisation of Language*. Amsterdam: John Benjamins.

Knowles, G. 1991. 'Prosodic Labeling: The Problem Of Tone Group Boundaries', in Johnson, S. and Stenström, A. eds. English Computer Corpora. *Selected Papers and Research Guide*: 149-164.

Ladd, D.R. 1986. Intonational Phrasing: the Case for Recursive Prosodic Structure. *Phonology Yearbook*, 3:311:40.

Lehiste, I. 1975. The Phonetic Structure of Paragraph. In Cohen, A. and Nooteboom, S.G. eds. *Structure and Process in Speech Perception*. Berlin: Springer Verlag.

Local, J. 1992. Patterns and Problems in a Study of Tyneside Intonation. In Johns-Lewis, C. ed. 1986. *Introduction in Discourse*. London: Groom Helm.

O'Connor, J.D. and Arnold, G. F. 1973. Intonation of Colloquial English. 2<sup>nd</sup> edition. Arlow: Longman. 1973

Pike, K. L. 1962. Practical Phonetics of Rhythm Waves. *Phonetics* 8:9-30.

Scuffil, M. 1980. *The Interpretation of English Intonation Patterns by Native Speakers and German Speaking Learners*. Unpublished paper: University of Massachusetts.

Zuraidah Mohd Don. 1996. *Prosody in Malay Discourse: An Analysis of Broadcast Interviews*. Ph.D. Thesis, University Malaya.

Zuraidah Mohd Don. 1998. *Prosody and the management of Interaction in Malay Broadcast Interviews*. Paper presented at the 6<sup>th</sup> International Pragmatics Conference

Zuraidah Mohd Don and Knowles, G. 2006. Prosody and turn-taking in Malay broadcast interviews. *Journal of Pragmatics* 38: 490-512.



# TOWARD RECONSTRUCTION OF DEMONSTRATIVES IN PROTO-AUSTRONESIAN

Joseph C. Finney

Monterey, CA

*<jcfinney@redshift.com>*

## 1 Overview

This is a study in diachronic syntax of the Austronesian (henceforth AN) languages. This is a preliminary study, based on study of published grammars and dictionaries of languages representing major branches of the AN language.<sup>1</sup> The role of visibility in choice of demonstrative needs special study. In some languages (Mantauran Rukai) the differentiation of *da* (most distant) and *na* (intermediate) is one of visibility: if the object is not in sight it is *da*, and if it is in sight (and not here) it is *na*. Some other AN languages differentiate here-visible, here-not-visible, there-visible, and there-not-visible.<sup>2</sup>

Demonstrative adverbs ('here' and 'there') are semantically related to demonstratives ('this', 'that'), and in some languages their lexical forms are systematically related, though in English, 'this' is not lexically related to 'here'. The present study will refer only occasionally to demonstrative adverbs. It deals with demonstratives both as substantives ('I saw that', where 'that' is a whole DetP) and as what used to be called demonstrative adjectives or pronominal adjectives ('I saw that chair', where 'that' is the Determiner within the DetP). Nearly all AN languages have VO order, and so prepositions (and other casemarkers) come left of their Det phrases.

The conservative AN languages have casemarked demonstratives of CVCV form, in which the left syllable is a casemarker (in most cases, a former preposition) and the right syllable is a demonstrative showing distality (distance) comparable to 'this' and 'that' in English. Either consonant, or both, may be zero. While English has only two distances and Latin and Japanese have three, many AN languages have more than three, and we can list six or more monosyllabic CV demonstratives that occur widely in AN and show differing distalities. In some three-distance AN languages, including Hawaiian Polynesian and

<sup>1</sup> The aboriginal languages of Formosa/Taiwan are called Formosan, as the term 'Taiwanese' refers to the South Chinese language that became the native language of most Chinese in Taiwan. Abbreviations used are: C and V for generalized consonant and vowel; and for language groups, IE for Indo-European, AN for Austronesian, MP for Malayo-Polynesian, and PN for Polynesian, all of which can be preceded by P for Proto. For purposes of easy comparison among languages, I respell the Hawaiian [t/k] phoneme as 't'. I spell the Malagasy [u] as 'u' and not 'o'. Despite the tradition in Malagasy of spelling final /i/ confusingly as 'y', I spell it 'i'. I spell Dyen's PAN \*C as 'c', so as not to confuse it with the generalized consonant. It's an affricative like 'ch' in English or Spanish. In Malayo-Polynesian, and in some Formosan languages, this sound merges into PAN/PMP \*t, and so, of course, the attested form of the demonstrative *cu* is *tu* in Tagalog and many other languages. At points where morpheme boundaries are relevant I have inserted hyphens that are not in the quoted sources.

<sup>2</sup> Unlike Indo-European, AN has only one language substantially attested for a thousand years (Old Javanese); all the others became significantly attested only about 200 years ago. In many AN languages the first serious attempt to make a grammar and a dictionary was done by missionaries intending to translate the Bible and to preach.

Tuvalu Polynesian, the intermediate distance (*na* in both those languages) means ‘near you’. Interestingly, the same is true of Latin (*iste*) and Japanese (*so-*). In some of those languages, including Latin and Hawaiian, this ‘second-person’ demonstrative can be used scornfully.

In work not to be reported here, I have found that all the casemarkers attributable to PAN were ancestrally either prepositions or topic-markers in the earliest reconstructable phase of PAN. Many of the examples that we’ll see have casemarker *i* (topic, nominative), *i* (locative) or *di* (locative). For convenience I’ll use the abbreviation PCM for ‘preposition(s) and/or (non-prepositional) casemarkers’. The PCMs used as independent words left of Det phrases are in origin the same as those occurring as the left syllables of case-marked demonstratives (though, of course, they can come to develop differently over time).

In AN as in IE, third-person pronouns typically develop from demonstratives. In AN as in IE, demonstratives are used either as substantives (constituting a whole Det phrase) or as Determiners heading a Det phrase (a use formerly called demonstrative adjective or pronominal adjective in English). The term ‘Determiner’ includes nonprepositional casemarkers as well as articles and non-substantive demonstratives. In AN as in IE, definite or specific articles can arise from further grammaticalization of demonstratives, often with lenition.

As a first approximation, there are at least six basic demonstratives that occur widely throughout AN and are candidates for being reconstructed as PAN. Roughly in order from proximal (this, here) to distal (that, there), they are:

<b>PAN</b>	ni	di	a	cu	na	da
<b>PMP</b>	ni	ri	a	tu	na	ra

The left two and the right two may have been systematically related semantically and historically. The two in the center may have been inherently semantically neutral in regard to distance. It appears that in languages that have lost one or more of the distal demonstratives, the central demonstratives move distally; and in languages that have lost one or more of the proximal demonstratives, the central demonstratives move proximally in their usage. This is especially true of *tu*. Note that [c] weakens to [t] not only in MP but also in some Formosan branches of AN.

The *di/ri* demonstrative is not found broadly in the Formosan languages. Its clearest occurrence in Formosa is in Nataoran Amis (Chen 1985), where it is fossilized as the right framing demonstrative regardless of distality. The variety of PAN \**d* occurring in the demonstratives is *dI*, which in PMP is \**d-r-r* (i.e. [d] initial and [r] medial and final). Thus the *di* demonstrative appears in Tagalog as *re*, at least in the right syllable of casemarked demonstratives. In conservative MP languages, [e] is generally split from PAN \**i*, and [o] is generally split from PAN \**u*, though in major morphemes of two syllables AN final \*-aw and \*-ay become [-o] and [-e] respectively. Schachter & Otanes (1972: 93) say ‘the initial [d] of the *sa* forms of a deictic is frequently replaced by [r] when the deictic occurs in the middle of a phrase, particularly after a vowel.’ Later in this paper we’ll consider other possible PAN demonstratives, *ca*, *ta*, and *ti*, as well as an invader.

Perhaps the most typical language will have only three of the basic six. The order of distality varies somewhat. I’ve placed a left of *cu* (*tu*) because it’s used that way in Amis

Formosan as well as in some MP languages (Malagasy, Bikol). But in Tagalog, *a* is more distal than *cu* [tu], and in fact *tu* serves as the most proximal demonstrative of all for those Tagalog speakers who don't use *ri* [re] at all. In Sebuano (Cebuano) the order is *ri* (near me), *ni* (near you and me), *na* (near you), *tu* (away from us both or invisible to us both), though *a* substitutes for *ri* in a part of the paradigm. Some orders, however, allow no exceptions in any language. In every language that has both *ni* and *na*, *na* is more distal than *ni*. And in every language that has both *di* and *da*, *da* is more distal than *di*.

So it seems very likely that at a very early level, perhaps older than PAN itself, semantic differences were established both between \*d- and \*n- and between \*-a and \*-i. That implies that at some early time each of the forms \*ni, \*na, \*di, and \*da, may have been composed of two morphemes. Did one of the contrasts denote distance and the other contrast denote visibility? That's the sort of hypothesis that must be explored. In some languages, such as Mantauran Rukai in Formosa, the most distal form, *da*, refers to objects that are not visible to me (and, in general, not visible to you, either).

In one group of MP languages, including Malagasy, there is a complex set of demonstratives, distinguished by both visibility and distance, and some demonstratives have been created by compounding. Though each casemarked demonstrative consists historically of two morphemes, we'll find that in some languages the word comes to be treated as a single morpheme. One accidental fact that facilitates that outcome in MP is that the nominative or absolute casemarker *i* is a homonym of the casemarker *i* which is locative, and which comes to add other oblique functions (genitive in some languages, accusative in others). So the combination of *i* plus demonstrative comes to be considered a demonstrative. So, other determiners (article or casemarker or both) come to be inserted at its left. An example is that Hawaiian has a set of demonstratives *tee-ia*, *tee-naa*, *tee-la*, whose right halves were ancestrally \*i-a, \*i-na, \*i-ra. Such occurrences in other AN languages misled Dempwolff (1938) into constructing a PMP one-morpheme demonstrative \*i(y)a. Most of us now construe the demonstrative ancestrally as simple \*a.

In AN languages as in languages in general, demonstratives and third-person pronouns can often be shown to have common ancestry. Note the Romance languages which have differentiated third person pronouns, demonstratives, and definite articles from forms of Latin *ille*, the most distal demonstrative. Within MP, the form \*i-(y)a (nominative and locative casemarker plus intermediate distance \*a) is the usual source of third person singular pronoun, though the 3ps comes from PAN \*da in Amis and from PAN \*na in Kambera. Dempwolff, whose scope of study was MP, reconstructed both a personal pronoun *i(y)a*, 'he, she', and a 'demonstrative pronoun' *i(y)an* 'that'. (I have modernized the notation and translated from the German.) Because -an occurs as a postposed locative marker in some Formosan languages, it seems likely that the -n forms may have earlier meant 'there' before meaning 'that'. It is also worthy of note that while in two of his citations Dempwolff translates the -an form as 'that', in the third one he says 'this'. That's consistent with the intermediate position of *a* in distality.

## 2 Criteria

A hard question is: in how many languages, and in which language groups must we find a demonstrative (or any other feature) in order to justify our attributing it to PAN? The answer must depend in part on the identification of the primary branches of PAN. If we know for sure that both of two groups are primary branches of PAN, the reasonable

attribution of the element to the proto-language of each of the two groups (not by loan) is all that is needed to attribute it to PAN. But obviously, the more primary groups it occurs in, the more comfortable we feel in the attribution to PAN.

Although the MP group includes nearly all the AN languages (98 or 99 per cent of them), the conventional conclusion from Robert Blust's thorough and systematic work has been that at best, MP is one of several primary branches, and perhaps not even that. The aboriginal languages of Formosa have long been considered by most scholars to comprise all three of the primary divisions of AN, and MP may be a subdivision of one of them. Recently, however, in a paper given at the 8th International Conference on Austronesian Languages, Blust has proposed that the Formosan languages comprise nine primary divisions of AN, with MP the tenth.

Blust's nine Formosan branches are: 1. Atayalic; 2. East Formosan (Amis, Siraya and others); 3. Puyuma; 4. Paiwan; 5. Rukai; 6. Tsouic; 7. Bunun; 8. Western Plains; 9. Northwest Formosan (Saisiyat and Kulong-Pazeh). If, indeed, PAN has ten primary branches, it would be a good idea, in diachronic studies such as this, to examine a language in each branch, although, in theory, finding an element in two branches should be enough to reconstruct it in PAN. One principle that has been proposed is to require the element to be found both in an MP language and in a Formosan language, because Formosa is small enough that borrowing of an element from any branch to any other branch there can be suspected.

### 3 Demonstratives in Eastern MP

Eastern MP consists of the (many) Oceanic languages and the (few) SHWNG. With some irregularities, Hawaiian shows the first, third, fifth and sixth of the basic demonstrative set as:

(tee)nei      (\*tee)ia      (tee)naa      (tee)laa

These are from PPN casemarked demonstratives:

\*e-ni      \*i-a      \*e-na      \*e-ra

These PPN forms are also reconstructed for Proto-Eastern Oceanic (though all with the ancestral casemarker *i*-). Forms from other Eastern MP languages are from within this set, though often with more lenition or with fewer of the four forms. Proto-Polynesian was a highly conservative language, despite its phonological mergers and its wobbling between Accusative and Ergative syntax.

The Hawaiian Dictionary (Puku'i and Elbert 1971) indicates that 'nei' and 'ia' are virtually synonymous (both defined 'this'), though *nei* never takes the *tee* prefix, which is a specific or definite article. Though I'll not discuss it here, the *tee* element also seems to be ancestrally two morphemes, \**ti-a*.

#### 4 Demonstratives in Central MP

The most thorough study of a Central MP language is that of Kambera by Klamer (1994), in which the demonstratives are given on page 59. The forms are:

ni        nai        na        nu

These include the first (*ni*) and fifth (*na*) of our basic six. Semantically, Kambera's *ni* is at me, near me; *nai* is near me but farther than *ni*; *na* is at you, near you; and *nu* is far from both me and you. (I prefer to word them this way and not use the common awkward wording, 'near the speaker', 'near the addressee'.)

Kambera (and perhaps other Central MP languages) has a system for number that I have not seen elsewhere. Each of these demonstratives is followed, in the same word, with a third person pronoun: *na*, singular, or *da*, plural. A sentence given by Klamer is:

'Nu-na        atau        'ni-na ?  
That one        or        this one ?

We need not be amazed that the 3ps pronoun element added (*na*) is of the same ancestry as its homolog which is one of the basic demonstratives, and so it occurs twice as '*na-na* for one of the distances. The Kambera set (ignoring '*nai*' which as bisyllabic might be a compound) suggests the possibility of an ancestral paradigm that could be as follows:

<b>Visible:</b>	ni	na	nu
<b>Invisible:</b>	di	da	du

It is of interest that of the two in Kambera that are not among our basic six, *nai* and *nu*, each, though uncommon in the AN family, has enough presence there to suggest the possibility of other PAN demonstratives, perhaps \**Cay* / *Cai* or *Cu*.

#### 5 Question: is there a PAN demonstrative 'ay'?

In Li's study of Tanan Rukai in Formosa (his Ph.D. dissertation under Starosta) (Li, 1973, 87) the casemarked demonstrative for the most proximal distance is *kay* or *kayvay* in the Nominative and *ki-kay* in the non-Nominative. In Tagalog, the so-called *sa* case (Case 3) marker for personal nouns is *kay* (singular) and *ki-na* (non-singular). Malagasy has a demonstrative *iai* which seems to be formed irregularly: its ancestry and its formation within the language are problematical. Why doesn't an epenthetic /z/ develop there as it did in *i(y)a* > *izy*? The possibility of other PAN \**Cu* demonstratives in a paradigm with Kambera's *nu* is a hypothesis to be considered, too.

#### 6 Questions about 'ta' and 'ca'

As Mayrinax shows *-ca* and Tsou shows *-ta*, it would appear that both forms may have been demonstratives in PAN. The two forms, of course, must merge in Malayo-Polynesian, and also in Bunun and in East Formosan, which includes Amis, Siraya, and the Kavalan and Bazsay-Trobiawan group. What evidence there is in languages that keep /c/ and /t/ separate, suggests that *ta* is nearer, more proximate, and more visible than *ca*. I am not aware of any language that separates /t/ and /c/ and has both *ta* and *ca* demonstratives.

## 7 Demonstratives in some Western MP tongues

Tagalog shows the second, third, and fourth of the basic set of demonstratives as:

(i)re (iy)a(n) (i)tu

though the locative (*sa*) case has not *re* but *ne*, the first, leftmost of our basic six, and adds, as its most distal, one that is not in our list of six:

i(y)o(n)

but in locative (*sa*) case:

du'on

So we have two models for the far-distant non-basic demonstrative in Tagalog; both seem to have labialization, but the *i*-case suggests that it occurs in the final syllable, while the *d*-case form suggests that labialization began in the left syllable and spread to the right. A labial element /u/ or /w/ seems also to be shown in the Wolio demonstrative, Anceaux (1987) as follows: *iwe* ‘there’. We’ll return to Tagalog and Wolio’s labials later.

Cebuano Visayan (Bunye and Yap 1971) shows four distances (five, with problematic ‘*ha*’), whose order is *di-ni-na-tu* (with left syllable *di*- and *nga*-). Changing their spelling from ‘*nh*’ to ‘*n*’ and from ‘*dt*’ to ‘*t*’, the distances are:

-ri < -di	near me (not found with <i>ri</i> - prefix)
-a	near me (not found with <i>nga</i> - prefix)
-ni	near both you and me
-na	near you
-ha	far from me or from me and you (not with <i>nga</i> -)
-tu < cu	very far from both me and you

But we’ll revisit its *dtu* later.

## 8 Demonstratives in some Formosan languages

For Mayrinax Atayal, a fairly conservative Formosan language, Huang (1994: 133) shows the demonstrative pronouns as follows:

<u>Distance</u>	<u>Visibility</u>	<u>Form</u>
nigh to me	visible	ha-ni
nigh to you	visible	ya-ni
away from both	(in)visible invisible	ha-ca ya-ni

The proximal *-ni*, of course, is one of the most commonly occurring demonstratives in Austronesian languages. The *-ca*, however, is not. This paradigm shows that the two dimensions, distance and visibility, are independent of each other. But there is no simple

relation between the morphology and the dimensions. Notice that *yani* means both ‘visible and near you’ and ‘invisible and not near you nor me’. Perhaps a way to describe its territory is that it covers the intermediate ground, i.e., neither invisible far away nor visible close.

For Siriya (Adelaar, 1997: 373, which merges /t/ and /c/), the demonstratives are simple (only two, and with no morphological distinction between singular and plural):

<b>Proximal</b>	<b>Distal</b>
(this, these)	(that, those)
a-ta	a-na

For Tsou, Tung (1964) also shows that distance and visibility are independent, though Tsou’s paradigm bears no resemblance to Mayrinax’s:

<b>Visible</b>	<b>Not visible</b>
Near: e	Nearby: co < *cu
Mid-distance: si	Seen by me before: o < *u
Far away: ta	Never seen by me before: na

Linguists have called these Tsou forms casemarkers. In fact, they were case-marked demonstratives, but for most speakers the casemarking syllable *i*- has disappeared and only the demonstrative remains. These are the nominative [‘second conjunctive’]. For the other case or cases, the ancestry is less clear.

For Nataoran Amis (Chen, 1987), note that ancestral *na* and *ni* have merged. That’s my interpretation of Chen’s showing that in at least some constructions it is clear that *na* is an alternate way of saying *ni a* (followed by a noun), where *a* is called a ligature. If the manifest *na* is at least partly of *ni* ancestry, it is understandable that *na* can serve as a proximal demonstrative ‘this’. The distal is *ra* < \**da*, and it serves also as 3sg pronoun. The *i-a* form serves as article ‘the’. The fourth demonstrative form that occurs is *iri* < \**i-di*; it has been specialized as the right framing demonstrative and is no longer regarded as a demonstrative. ‘Framing’ is a term first used by Ed Keenan and Ralalaoherivony (1998) for Malagasy, where an NP often is framed between two demonstratives, left and right, but Malagasy has a constraint that the left and the right must be the same.

Note also that Nataoran Amis has an interrogative pronoun *icoa* < \**i-sua*. It is called impersonal, and occurs only in ‘neutral’ case (predicate nominative), in these forms: *icoa* ‘where?’, *hacoa* ‘how much? how many?’, and *o icoaan* ‘which one?’. Each occurrence of /o/ is ancestral \**u*.

For Rukai, related dialects described by Li (1973), by Starosta, and by Zeitoun, *da* is the most distal demonstrative; it requires invisibility. In at least one Rukai language, Tanan, the demonstrative for an intermediate distance with visibility has interchangeable forms *na* and *nia*, showing that here as in Amis, the *ni* and *na* forms have merged, and perhaps for the same reason. The most proximal demonstrative determiner in Tanan Rukai is *kay*, as we have noted. A table modified from Li (1973:87) is:

<b>Proximal</b> 'this'	<b>Intermediate</b>	<b>Distal</b> 'that'
-NM      ki-kay	i-ni-a, ina, na	i-Da-a
+NM      kay(vay)	ku-a-ni	ku-a-Da

## 9 An intriguing discovery: A PAN demonstrative in '-wa' or '-ua': The logical steps in establishing it

Let's begin, as I did, with noticing some peculiarities in one or two MP languages; and then examine more languages, one at a time, till we find a set of hypotheses and a diachronic scenario that comes close to accounting for them.

These observations are consistent with the fact that in some languages such as Tagalog the *-an* form is not the most distal, but is accompanied by a more distal form *-un*. Thus in Tagalog, corresponding to *iyan* and *diyan* we find the corresponding more distal forms *iyon* and *doon* [du'on]. Schachter and Otanes (1972: 91) show this table of 'deictic pronouns':

<b>Ang Form</b>	<b>Ng Form</b>	<b>Sa Form</b>
ire /'ireh/	nire /nireh/	dine /di:neh/
ito /'itoh/	nito /nitoh/	dito /di:toh/
iyan /('i)yan/	niyan /nyan/	diyan /dyan/
iyon /('i)yon/	niyon /nyon/	doon /du'on/
	noon /nu'on/	

Note that the top line, for the most proximal 'this', mingles forms from the first two (*-ri* and *-ni*) of the basic six demonstratives. Even so, that combined distance form is fading out of the language. Some speakers, especially in Manila, no longer use it at all. So the generally neutral distance form, *-to* < \**cu*, has had to move into the proximal 'this' distance. The a distance form (*iyan*, *niyan*, *diyan*) holds the mid distance; it's a 'that' but not very far. And both forms that in general in AN show greater distance than a (*na* and *da*) are also missing. They lost out to a demonstrative from outside the basic set. Its varying forms show a labial vowel < \**u* in the left or right syllable or in both. Whatever its origin, it shows regularities, and must have undergone some changes not easily understood.

For the second language, see the demonstratives in Wolio, a language of Southeast Sulawesi [Celebes], as given in Anceau's (1987) Wolio Dictionary:

iwei	here
iwe	there
itu	that, yonder, over there

And some compound uses

iwe itu	there, yonder
iwei itu	over there, yonder

So far, not very promising. In the two languages the demonstratives have shown some odd forms. At this point, we have a hunch that PMP may have had a demonstrative for great distance with *-u-* or *-w-* in it. But still only a hunch.

Now, a third language, Malagasy, a language with a great number of demonstratives (seven, or as many as eleven, depending on who counts them). Malagasy (Keenan, Pearson, Randriamasimanana) has a demonstrative *iru*, or, as I prefer to write, *i-ru*. Still not very promising. Another source for Malagasy (Randriamasimanana, a linguist who is a native speaker of that language) shows what seems to be another version of the one we have just seen. It is *irua*, or as I'll write it, *i-rua*.

But three syllables are too many by one for a casemarked demonstrative. Has the first or the last syllable been added later? If the casemarked demonstrative is *i-ru*, the casemarking is the normal nominative and locative, and the demonstrative syllable must be *ru*. That doesn't fit the Wolio. If the casemarked demonstrative is *ru-a*, the casemarking is the locative /d/ as in the Tagalog, and the Wolio form suggests that the PMP form ancestral to the Malagasy and Tagalog forms may have been *ri-we*, while the Wolio came from an alternate PMP form *i-we*, using a different locative preposition. Both hypotheses of ancestral prepositions are consistent with evidence from other attested languages.

The next step is a lucky accident. At the March 1998 meeting of the Austronesian Formal Linguistics Association (AFLA) I happened to encounter a paper on the language of Madura by William D. Davies of the University of Iowa. In it are the demonstrative adjectives (determiners):

jiya	this
juwa	that

In a swift oral consultation, Bob Blust said that he had never seen AN demonstratives of that form. He identified the initial consonant /j/ as PAN \*Z.

The paired words appear to belong to a paradigm. If so, the left syllable should be a casemarker and the right syllable a demonstrative showing distance. But there's a problem. With initial PAN \*d-, a casemarking syllable \*di occurs widely, and a variant \*du is plausible. But no PAN PCM \*Zi nor \*Zu is reconstructable from other AN languages.

The final piece of evidence that solved the problem came when I looked again at Ferrell's (1981) Paiwan Dictionary. There lay an entry that I had long forgotten: a locational noun *zua* with some uses as a verb and as a demonstrative pronoun. Among its 37 entries are:

i zua	there; over there
i-zua	there is; there exists
sa-zua-u	go there!
a-zua	that, those (also contracted to: a-za)
pi-zua	to put there
pa-ka-zua-n	(1) place; (2) route, itinerary
ma-i-ta-zua	thus, in that way; (Western Paiwan) gratis
sa-zua-in	destination
zua-zua-n	(Western Paiwan) furthest

Here I use bold face for what I think is the most ancient use, 'far away place', and for the demonstrative use which it has in MP languages.

Now it becomes possible to solve the problems of the *-u-* and *-w-* demonstratives not only in Madurese but in other languages.. The word *\*zu(w)a* must have occurred as such both in Proto-Paiwan and in Proto-MP. In both, it must have had, as one of its meanings, the far distant demonstrative ‘that’. A hypothesis to consider is that *\*zu(w)a* ‘that, far distant’ was inherited from PAN or a lower-level ancestor common to Proto-Paiwan and Proto MP. But that’s not possible, at least not for PAN, because the ancestor of Madurese *j-* is PAN *\*Z*, and the ancestor of Paiwan’s *z-* is not. Notice that the word has not been reported in other branches of AN.

Another hypothesis might be that Proto-Paiwan and Proto-MP shared a common ancestor below PAN, which innovated this word after all the relevant phonological changes had taken place. That’s not the most plausible choice, either.

Notice the range of semantics in Paiwan for *zua*. In accordance with general linguistic principles the semantic ‘a distant location’ is almost certainly the earliest of the attested meanings, and the demonstrative ‘that (far away)’ is a later meaning evolved by grammaticalization. So the reasonable conclusion is that the word *zu(w)a* was borrowed from Proto-Paiwan into Proto-MP as a demonstrative pronoun (and possibly with some other uses). So the word was not a member of the class of PAN words we are reconstructing, the class of true demonstratives, each of which was a single syllable, and, indeed, a single mora. In *zu(w)a*, just as in *i(y)a*, the semivowel was nonphonemic (unless the language had a constraint that every syllable must begin with a non-zero consonant).

Notice that the two words in Madura, though made to look alike with initial */j/*, are not morphologically or syntactically analogous. The casemarked demonstrative *j-i(y)-a* is made of at least two morphemes: (1) a PAN casemarker consisting of */i/* or possibly */Ci/* (in earliest times either a preposition or a topic marker), and a demonstrative *a*. But *zu(w)a*, as befits a noun, is a single morpheme of two syllables.

When the bisyllabic *zu(w)a* became fully a demonstrative in MP, it evidently posed a challenge in that it disrupted the system of AN demonstratives to which it did not conform. Let’s explore the hypothesis that the coping with such challenge can account for the irregular demonstratives that we have noted. First, what happened in the ancestor of Madurese? It must have lost all the other demonstratives, and had none left but *i(y)a* and *ju(w)a*. The latter was the distant ‘that’; and so the former, which has been a neutral demonstrative, took the territory of ‘this’. Note that Madura’s ‘that’ word is identical phonologically and morphologically with the word in Paiwan, but for the regular shift from */z/* to */j/*. Then the language forced the two Madura words into a morphological paradigm by changing the initial consonant of ‘this’ from zero to */j/*, i.e., the language’s definition of the demonstrative category came to include that the word begin with */j/*, which became the element marking a demonstrative.

Among the languages we have seen, Madura was alone in changing the true ancestral demonstrative to fit the pattern of the invader *zua*. Apparently most languages rejected *zua* and let it go obsolete. All the other four MP languages in which we have found a *zua* descendent have changed it to fit into the paradigm of the true ancestral demonstratives. But the four tongues we are about to examine did it in four different ways.

In Wolio, the form *ewe* was reached by zeroing the initial *\*Z* and weakening the final */a/*. The normal AN *\*i* vowel was inserted into the left syllable, making it a casemarking syllable, but the *u(w)* was reduced to */w/*, so that the word kept only two syllables.

Tagalog is another in which the invader *zua* was reformed to fit (at least in part) the paradigm of the true ancestral AN demonstratives. As we see in the Tagalog table's bottom lines, in some syntactic cases the left syllable's vowel changed to the standard /i/ (and, if so, kept the nonphonemic /y/ glide), while in other syntactic cases the left vowel kept the invader's /u/ (or made it /o/). The initial consonant of the initial syllable of all syntactic cases shifted from *z* to the standard marker of the appropriate case. By backward assimilation, the vowel of the second syllable shifted from /a/ to a labial vowel. All this accounts for the forms seen in Tagalog's set of the most distal demonstratives, which without this reconstructed history are highly confusing and inconsistent.

The proposed scenario for Malagasy is also fascinating. First, the *zua* took the casemarking prefix and became *izua*. Then the consonant changed and so the word became *irua*, which in Malagasy's standard spelling is 'iroa'. One authority today lists it as *iroa*, while another lists it as *iro*; so apparently the language today is in the process of losing the final vowel and beginning to conform to the bisyllabic paradigm of the true ancestral demonstratives. In Malagasy, the scenario that we have established shows that (between the alternate forms *iru* and *irua*) it is *irua* and not *iru* that stays closer to its ancestral pattern (\**zua*). We might say that through the sequence *zua* > *izua* > *irua* ('iroa') > *iru* ('iro') the word has succeeded in faking its ancestry. Now it comes to mimic the pattern of the Malagasy demonstrative *iri*, which descends normally from PAN \**i-di*. That gives it an appearance of legitimacy. But the final step in that sequence is still in process. We are fortunate to have caught it before *irua* is totally replaced by *iru*.

Incidentally, Malagasy's *i-ri* may also descend from PAN \**i-da*, with the same final lenition that produced *iti* from ancestral \**i-ta*.)

One other MP language turned up with what must be a descendant of \**zua*, and it is in the right position, as the most distal of the three demonstratives in the language. That language is Manobo (Elkins, 1968). The proximal is given as *kayi*, *he'i*, and *he'ini* (PAN \*-ni); the intermediate 'that, not far' is *ne'eya*, *he'eyan*, *heya'* (PAN \*-a); and the distal 'that, far' is *haza*, *he'aza*'. The reasonable interpretation is that the element -za is a lenition of \**zua*. Again, the phonology was adjusted so as to fit the paradigm of the genuine PAN demonstratives, each of which is a monosyllable with a single short vowel (with a monosyllabic casemarking determiner on its left). So, the theory given here purports to show that four MP languages, in series of steps, altered the invader *zua* to conform with the paradigm of the ancestral PAN demonstratives. And each of the four did it in its own distinctive way.

So this is a nice detective story with a happy ending. Everything puzzling has been explained. The diachronic scenario works.

In summary, the scenario is as follows: In Proto-Paiwan was an old noun *zua* meaning distant location, distant place; and it was beginning to be used somewhat like a demonstrative. In PMP its use became clearly that of a demonstrative. Speakers had to reconcile its use with that of the normal PAN demonstratives. Each member of that set consisted of a right syllable that was one of 6 or 8 distance demonstratives (*ni*, *di*, *a*, *tu* < *cu*, *na*, *da*; plus probably *ta* < PAN \**ta*, \**ca*, and possibly *ti*), and a left syllable that was a casemarker (commonly nominative *i* or locative *i* or *di*). The Madura language responded by altering the morphology of the single remaining PAN true demonstrative to conform to that of the *zua* derivative. But Tagalog, Wolio, Manobo, and Malagasy, each in a different

way, altered the morphology of the *zua* derivative to conform to that of the PAN true demonstratives.

## 10 Marking of plurality in the demonstratives

In some AN languages the plural demonstrative is formed by inserting the syllable *-re-* between the casemarking syllable and the demonstrative syllable. Thus in Malagasy a singular casemarked demonstrative is *i-tu* (routinely misspelled *i-to*) and its plural is *i-re-tu* (*i-re-to*).

Independent 3pl pronouns in Rukai (Li, op. cit. 83) are as follows. Note that the intermediate demonstrative is used for definite and the distal demonstrative for indefinite in these pronouns:

<b>def</b>	ku-l-i ni	l-i-ni-a	zero	-l-i-ni
<b>indef</b>	ku-l-i-Da	l-i-Da-a	zero	-l-i-Da

1. Some languages don't mark plurality in the demonstratives.
2. Some languages mark plurality by inserting left or right of the demonstrative, a plural marker. This may or may not be related to the 3pl personal pronoun.
3. Some mark plurality for demonstrative pronouns as well as for common nouns by inserting a collective noun (or a former collective noun).
4. One Central MP language overtly marks number for both singular and plural by adding on the right the third person singular or plural pronoun.

## 11 Vowels 'i' for high and 'a' for far and the possibility of ancient 'u'

As we have noted, *di* is proximal and *da* distal in every language that has both. Likewise, *ni* is proximal and *na* distal in every language that has both. Does that principle hold for all consonants?

In Bunun, a Formosan language (Jeng, 1977), the principle holds for *-ti* and *-ta*. Thus (p 244) we see *?iti?* glossed 'this place' in a sentence 'I touch this place'. As determiners, the demonstratives come after the nouns in Bunun. On page 154 we see the phrase *Iudun ta?* glossed 'in the mountain there'. The book has many examples of determiner *di* 'this' after nouns (busul *di* 'this rifle' page 154); I didn't find examples of *\*da*. Interestingly, Bunun (Jeng, op.cit.) shows the proximal *i*, distal *a* also with consonant *p*, which seldom occurs in monomoraic function words in the AN family. Examples, page 131, are for the 3<sup>rd</sup> person singular pronoun, animate, six choices, three of which are:

- c) ?aipa? (far)
- d) ?aipi? (near)
- e) ?aipu? (far)

That example is intriguing for two reasons. First, because it shows the proximal-distal distinction by vowel with morphemes other than the *d*, *n*, and *t*. That strengthens the hypothesis that in a very ancient level, possibly earlier than PAN, each of those demonstrative syllables was built of two morphemes, with the vowel marking the distance. Thus, possibly, *\*ni* was earlier *nV + i*.

The other intriguing point is that a third vowel, u, occurs in the demonstrative paradigm and denotes far distance. That's also what it denotes in Klamer's Kambera in the Central MP area. It occurs as an interrogative pronoun in some other languages. Such examples suggest the possibility that similar use of the third vowel may have occurred in PAN, though it survives now only in a very few uses in a very few languages.

## 12 Conclusions

1. Proto Austronesian had at least six, and perhaps as many as nine, demonstrative pronouns (also serving as demonstrative determiners). All show distance in some sense.
2. All these basic root demonstratives were single light syllables CV in which the C could be zero.
3. In their common and normal use, the root demonstratives have attached at their left a casemarking syllable, which is also CV, in which the C can be zero.
4. In the relatively conservative AN languages today, the (casemarked) demonstratives still have that CVCV structure. The potential number of such is the language's number of casemarkers times its number of root demonstratives.
5. Each conservative language uses a subset of the PAN root demonstratives, sometimes as few as two (English 'this' and 'that') but typically three or four, and a few languages use more than four of them.
6. In the AN demonstratives, visibility is one of the key semantics. In some languages two demonstratives may differ only in visibility. In some languages, visibility crosscuts distance. Languages also vary in the extent to which (if at all) the demonstratives also denote visibility by the second person and/or proximity to the second person.
7. Variations are discussed in the ways in which (if at all) the demonstratives are marked for plurality.
8. I have reconstructed the prehistory of a bisyllabic invader of two syllables which entered the semantic territory of demonstratives late in the PAN period. I have traced in four languages the ways in which the invader was altered step by step to conform to the pattern of the standard casemarked demonstratives (very different ways in each of the four). And I have found one language in which it was the standard demonstrative that was reshaped and reanalyzed to conform to the pattern of the invader.

## References

Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36(2): 362-397.

Anceaux, J. C. 1987. *Wolio Dictionary*. Dordrecht, Netherlands: Foris.

Bunye, Maria V. R. and Elsa P. Yap. 1971. *Cebuano Visayan Dictionary*. Honolulu: U. Hawaii Pr.

Chen, Teresa. 1985. *Verbal Constructions and Verbal Classification in Nataoran Amis*. Pacific Linguistics C-85. Australia National University Press,

Davies, William D. 1998. Paper read at the annual meeting of the Austronesian Formal Linguistics Society, Honolulu.

Dempwolff, Otto. 1938. *Vergleichende Lautlehre des Austronesischen Wortschatzes*.

Finney, Joseph C. 1997. Casemarking and the \*ia Demonstratives. *Proceedings of the Seventh International Conference on Austronesian Linguistics*, Cecilia and Wim STOKHOF (Eds.). Amsterdam.

Huang, Lillian M. 1994. *A Study of Mayrixan Syntax*. Taipei: Crane.

Jeng, Heng-hsiung. 1977. *Topic and Focus in Bunun*. Taipei: Special Publication 72 of the Institute of History and Philology, Academic Sinica.

Keenan, Ed and Balohisoa Ralalaoherivony. 1998. *Raising from NP in Malagasy*. Presented at the Fourth Annual Meeting of the Austronesian Formal Linguistics Association, UCLA 1997

Klamer, Marian. 1994. *Kambera: A language of Eastern Indonesia*. [HIL dissertation 11]. The Hague: Holland Academic Graphics.

Li, Paul Jen-kuei. 1973. *Rukai Structure*. Institute of History & Philology, Special Publication, No. 64. Taipei: Academia Sinica. (University of Hawaii at Manoa Ph.D. dissertation)

Puku'i, Mary Kawena and Samuel Elbert. *Hawaiian Dictionary*. 1971, Honolulu: University of Hawai'i Press.

Rich Schachter, Paul and Otanes. 1972. *Tagalog Reference Grammar*. Berkeley and Los Angeles: University of California Press.

Starosta, Stanley. 1996. *Formosan clause structure: transitivity, ergativity, and case marking*. Symposium series of the Academia Sinica, Number 2: Chinese languages and linguistics IV, Academia Sinica. Taipei, Taiwan. 18-20 July 1994.

Tung, T'ung-Ho. 1964. *A Descriptive Study of the Tsou Language, Formosa*. Taipei: Special Publication 48 of the Institute of History and Philology, Academica Sinica.

Zeitoun, Elizabeth. 1995. *Issues on Formosan Linguistics*. Paris, France: Universite Denis Diderot Ph. D. dissertation (Paris 7).

# GRAMMATICALIZATION OF DEVERBAL MARKERS: TOWARD A CROSS-LINGUISTIC STUDY IN THE SEMANTIC EXTENSION OF MOTION VERBS<sup>1</sup>

Ruetaivan Kessakul

*Chulalongkorn University*

*<ruetaivan@yahoo.co.jp>*

Toshio Ohori

*The University of Tokyo*

*<tohori@boz.c.u-tokyo.ac.jp>*

## 0 Introduction

The development of serial verbs into other grammatical categories has been widely studied, e.g. Matisoff (1991), Heine et al. (1991), Lord (1993), Bisang (1996). In this paper we will take a close look at some motion verbs in Thai, namely, *paj* ‘to go’, and *maa* ‘to come’, *càak* ‘to leave’, *thúng* ‘to arrive’, *løøj* ‘to pass, exceed’, and examine directions of their semantic extensions. The presentation is mainly data-oriented, but the cognitive basis of semantic extension will also be discussed.

## 1 Background and Proposed Patterns

Concerning the functions of serial verbs, Thepkanjana (1986) has stated the following characterizations: (a) complementing the initial verb, which results in the semantic implications of causative, passive, or resultative; (b) indicating direction and aspect; (c) acting as grammatical markers and/or case markers, i.e. coverbs; and (d) indicating purposive and simultaneous actions (for the functional extensions of Thai verbs, also cf. Kolver 1984, Areemit 1986, Bilmes 1995). However, she did not fully discuss the pathways of semantic extension of the verbs we are going to discuss. We start from Bisang (1996: 575), who proposes two pathways of grammaticalization in the domain of verb (a list of abbreviations is given at the end of this paper; CV=causative verb).

(1)			-->	TAM	-->	CONJV	
	(XXI) V	-->	RES	-->	Vd	-->	TAM
	(XXII) V	-->	CV	-->	CONJV		
		-->	COV	-->	CONJV		

These pathways are found across Mainland Southeast Asian languages, but it is also known that they are not strictly unidirectional processes. Rather, the Thai verbs in question still maintain their original meanings and verbal functions while developing into

<sup>1</sup> We are indebted to Shigeru Sakahara (University of Tokyo), Yo Matsumoto (Meiji Gakuin University) and Makoto Minegishi (Tokyo University of Foreign Studies), for their helpful comments on earlier drafts of this paper. We would also like to thank Tasanee Methapisit (Tokyo University of Foreign Studies) and Kiyoko Takahashi (Chulalongkorn University) for their cooperative discussions. Also a word of thanks to the individuals who asked questions and gave us valuable suggestions at the conference. Any errors are, of course, our own.

grammatical units. Keeping this in mind, we propose two directions of change in (2), which show mapping from prototypical meanings to extended grammatical meanings. Note that the extended meanings are in the hierarchical relations to each other.

(2) Type I:  
 V --> Vd --> TAM --> COV --> CONJV --> CONV --> LEX  
 Type II:  
 V --> COV --> CLM (LEX)

We divide the five motion verbs we take up into two groups based on our proposed patterns in (2). In the next section, we will illustrate them with examples taken from narrative data. They are natural spoken discourse by Thai native speakers based on the animated film ‘Pingu: The Most Cheerful Penguin in the World’. Also, some examples are taken from two Thai narrative stories as supplementary data.<sup>2</sup>

## 2 Two Directions of Grammaticalization

The two directions of grammaticalization in the domain of motion verbs we wish to propose are based on the idea that synchronically observed polyfunctionality patterns should reflect processes of semantic extension (which we call “motivation approach”, cf. Ohori 1998). Below are relevant examples:

### 2.1 Type I: V --> Vd --> TAM --> COV --> CONJV --> CONV --> LEX

The first array of functional extensions is summarized in Table 1. Type I features various uses of *paj* and *maa* where the function of a lexical word extends into a grammatical word with such functions as directional verb, tense-aspect-modality marker, coverb, conjunctional verb, and converb. Then through the process of lexicalization, it shifts into a particular meaning which functions as an eventual marker.

**Table1:** *The Chains of Polyfunctionality with paj and maa*

Categories	<i>paj</i> ‘to go’	<i>maa</i> ‘to come’
V	to move away from the speaker’s viewpoint	to move towards the speaker’s viewpoint
Vd	away from the speaker’s viewpoint	towards the speaker’s viewpoint
TAM	pass away progression imperative sense 1PL ‘let’s’ 2SG/PL direction	up to the present continuity imperative sense 1PL ‘let’s’ 2SG/PL direction
COV(+NP)	‘to’ indicating what is	the same as <i>paj</i>

<sup>2</sup> The two Thai narrative stories are:

(a) ‘taam sadèt klaj bâan’ or Retour the route to Europe as visited by King Rama V, Amporn Jirattikorn, Bangkok, Matichon Publisher, 2540 (1997)  
 (b) ‘phûtuan’ or Friends, prabhatson sewikun, Bangkok, dôok-ya Publisher, 4th edition, 2540 (1997)

	reached or approached	
CONJV	purposive	the same as <i>paj</i>
	sequential	
CONV <sup>3</sup>	V <i>paj</i>	V <i>paj</i> not applicable
	indicating two on-going actions	
LEX	V <i>paj</i> V <i>maa</i> --> indicating repetition of the action --> shifting to a particular meaning which denotes the eventuality and functions as an eventual marker (showing adverb-like function)	

(3) *paj*

## V

a. khòpkhun māak phōm paj kòðn  
thank you very much I (MASC) GO:V before  
ná? kráp  
SFM POLITE  
'Thank you very much. I have to go now.'

## Vd

b. pingu khó? pràtuu hōøñáam léew pèød døøn  
Pingu knock door toilet CONJ open walk  
khâw paj  
enter GO:Vd  
'Pingu knocked on the door, opened it and walked into the toilet.'

## TAM

c. pingu kô? nøøn kòøt túkkata thîi mëe  
Pingu CONJ lie hold teddy bear REL M mother  
hâj maa con làp paj  
give come CONJ sleep GONE:pass away  
'Pingu held the teddy bear that his mother had prepared for him until he fell asleep.'

d. léew lôok pengün kô? døøn paj døøn  
then child penguin CONJ walk go walk  
paj rûiway rûiway  
GO:progesssion continuously  
'Then little penguin (Pingu) kept on walking continuously.'

e. paj raw klàp kanthè nawa  
let's we return together Nawa  
'Nawa, let's go back.'

f. paj khâw hōøñáam dâj léew  
direction enter toilet possibility now  
'Go to the toilet now.'

<sup>3</sup> The Term CONV (converb) is defined here as a nonfinite verb form whose main function is to mark adverbial subordination (cf. Haspelmath 1995:3).

COV

g. wanníi pen wan thiíi pingu cà dəənthaan̄  
 today COPULA day REL M Pingu will travel  
paj bāan pâa khondiaaw  
TO:COV house aunt alone  
 'Today is the day that Pingu will travel to his aunt's house alone.'

CONJV

h. pingu dəən paj cəə bāan lăŋ nūŋ  
 Pingu walk GO:CONJV find house CLF one  
 'Pingu walked and then found a house.' (sequential)  
 i. phr̄? wan níi pingu cà dəənthaan̄ paj hăa khun pâa  
 because day this Pingu will travel GO:CONJV see HO aunt  
 'Because Pingu will go to see his aunt today.' (purposive)

CONV

j. raw dəən paj khui paj kanthà  
 we walk GO talk GO together  
 'Let's walk along and talk together.'

LEX

k. pingu tham taa klîŋ klîŋ paj maa  
 Pingu make eyes roll roll GO COME  
 'Pingu rolled his eyes round repetitively.'  
 l. t̄e kòon mēe kô? mâjdâj sōncai  
 from before mother CONJ NEG be interested in  
 duon̄daaw r̄òok t̄e... (omitted) kruu hâj<sup>4</sup>  
 stars PRT but teacher give  
 tham raaȳaa... (omitted) paj paj maa maa  
 do report GO GO COME COME  
 kô?ləej hän maa ch̄ôp duu daaw paj dûay  
 CONJ turn come like look star go too  
 'I hadn't been interested in the stars before, but after my teacher ordered me to  
 write a report (about the astronomer), at last I began to like the star too.'

(4) *maa*V

a. kháw maa rooŋrian săay thûk wan  
 s/he COME:V school late every day  
 'S/he comes to school late everyday.'

Vd

b. mii rót khõn khõŋ khâp phâan maa  
 there is car carry thing drive pass COME:Vd  
 'There is a freight car coming by.'

TAM

<sup>4</sup> r̄òok is a particle used after a negative statement to make the sense milder.

c. maa wâaj câawthîi câawthaaj kan kòon  
COME:TAM to pay respect the guardian spirit of one's land together before  
'Let's make the sacrifice to the guardian spirit first.'

d. maa kin aahâan khöoŋ pâa kòon  
COME:TAM eat food POSSESSIVE aunt before  
'Come on and try my dish first.'

## COV

e. kháw cà bin maa jiipùn aathít nâa  
s/he will fly TO:COV Japan week next  
'S/he will fly to Japan next week.'

## CONJV

f. pingu pít pràtuu sianj daŋ nóŋ kôŋlæŋ tûun  
Pingu close door noise loud brother CONJ awake  
khûn maa róŋhâj  
ascend COME:CONJV cry (sequential)  
‘As Pingu closed the door loudly, his brother awoke and cried.’

g. súu khóŋ maa faak dûay ná  
buy thing COME:CONJV make gift too PRT  
khráp phii  
POLITE elder brother/sister  
‘Please buy a souvenir for me too.’

The examples in (3) and (4) above illustrate various functions of the two motion verbs. Although the surface forms of these verbs are the same, their functions differ according to the environment in which they occur. Let us take *paj* for instance. We can see that the positions of *paj* in the serial verb constructions (henceforth SVCs) are different. That is, the verb can develop into a Vd when modifying the preceding verb as path or direction markers (cf. (3 and 4b)). In (3 and 4g), it can take a prepositional argument and function as a COV. Further, it functions as a CONJV, i.e. it links the following action to the preceding one, expressing such relations as purpose and temporal sequence, while conversbs express the simultaneity of two actions. Moreover, we can see that both *paj* and *maa* can denote the sense of repetition that leads to a particular meaning, functioning as a near-lexicalized eventual marker.

## 2.2. Type II: $V \rightarrow COV \rightarrow CLM (LEX)$

This type of grammaticalization pattern has three typical examples, namely *càak* ‘to leave’, *thúŋŋ* ‘to arrive’, and *løej* ‘to pass, exceed’. These tend to be natural choices, since there are corresponding usages of motion verbs in other languages, for example, Vietnamese (cf. *đến* ‘to arrive’).<sup>5</sup>

<sup>5</sup> The similar phenomenon can be seen in Vietnamese too. The examples of *đến* ‘to arrive’ are cited below (based on Nguyen 1979):

COV --> ALLATIVE: Xe chạy đến ga mới đỗ khách.

'The bus will not discharge its passengers until it arrives at the station'

**TEMPORAL:** *The bus will not discharge its load báo đến hai giờ sáng*

dọc bao dell hai giờ sáng  
‘to read newspaper until 2 am.’

**Table 2:** the chain of polyfunctionality with *càak*

<u>Categories</u>	<i>càak</i> 'to leave'
V	to depart from a reference point
COV	ablative 'since' agent instrument (tool, force, material, means)
CLM (LEX)	cause and reason <i>nûtaaq càak</i> 'because' <i>nôok càak</i> 'except for' <i>lăj càak</i> 'after' <i>càak nán</i> 'after that'

(5) *càak*V

a. kháw càak bâan kđøt paj tâñtèe jaŋ dèk  
s/he LEAVE:V home to be born go since still child  
'S/he left his (her) hometown when s/he was still a child.'

COV

b. tøønnán kô? mii khôøŋ tòk càak rót prajsanii  
at that time CONJ there is thing fall FROM:COV car post  
'At that time there was something falling from the post-car.'

c. càak wanní paj túkuàaŋ cà klàp kuuw  
FROM:COV today go everything will be back return  
maa müwan dæem  
come like before  
'From now on, everything will be alright as before.'

d. lëew pingu kô? dâj khôøŋ kin càak  
then Pingu CONJ receive thing eat AGENT:COV  
khun pâa khâaŋ bâan  
HO aunt side house  
'Then Pingu received some food from the neighbouring aunt.'

e. lûuk sâap càak kaan àan năŋsüuuphim  
I know BY MEANS:COV NOM read newspaper  
'I know it by reading the newspaper.'

FOCUS PARTICLE:	tiêu đến 20 đồng 'to spend as much as twenty dong'
DEGREE MARKER:	đọc đến rúc đầu 'to read so much as to have a headache'
EVEN:	đến (cả) ông, nó cũng không sợ. 'He is not afraid even of you.'
TERMINAL:	biết đến 'to know about/of'; nghĩ đến 'to think about/of'; nói đến 'to remember'

f. lûuk súus khɔ̄ŋ phûuak níi càak ŋøən s̄insðøt  
 I buy thing NUMBER this INSTRUMENT:COV money betrothal  
 'I bought these things with the betrothal money.'

g. phrɔ? nùuway càak kaan døənthaŋ  
 because be tired CAUSE:COV NOM travel  
 'Because (Pingu) is tired due to his travelling.'

CLM

c. nûaŋ càak sòŋ sian̄ daŋ kô?løeŋ tham hâj  
BECAUSE (OF) send noise loud CONJ make/do give  
 nôoŋ tûuŋ  
 brother awake  
 'Because of a loud noise, his brother was awoken.'

Next, we will list up the interconnection between various functions of *thŭŋ* 'to arrive' as shown below.

**Table 3: The Chain of Polyfunctionality with *thŭŋ***

<u>Categories</u>	<i>thŭŋ</i> 'to arrive'
V	to get to a place
COMPLEMENT V	indicating the achievement
COV	allative 'until' recipient terminal
FOCUS MARKER	to emphasize the amount
DEGREE MARKER	to indicate degree of an action or state by preceding the modificational phrase <i>thŭŋ káp; thŭŋ khanàat</i>
PREVERB	a colloquialism of <i>cuiŋ</i> , introducing the resultant statement. It can be used as a clause-linkage marker, too.
CLM (LEX)	'even' <sup>6</sup> concessive <i>thŭŋ mée wâa; thŭŋ jâaŋ raj</i> <i>thŭŋ krâñánkɔ? taam</i>

(6) *thŭŋ*

<sup>6</sup> The difference between EVEN and CONCESSIVE of *thŭŋ* can be briefed as follows:

	<b>Subordinate clause</b>	<b>Main clause</b>
EVEN:	<i>thŭŋ</i> (pen) [NORM]	won't do what one is expected to do.
CONCESSIVE:	<i>thŭŋ</i> [EXPECTATION]	will do what one is not expected to do.

V

a. pingu trɔŋ paj r̄ūuayr̄ūway naj th̄isùt k̄?  
 Pingu walk-straight go keep on in the end CONJ  
th̄ūŋ bāan khun pāa  
ARRIVE:V house HO aunt  
 'Pingu kept on going straight ahead, and at last (he) arrived at his aunt's house.'

COMPLEMENT V

b. l̄ɔŋ ùthaan j̄aaŋ khít māj th̄ūŋ  
 she exclaim like think NEG ARRIVE:COMP  
 'She exclaimed surprisingly.'

COV

c. khun luŋ khàp maa th̄ūŋ thaŋ yēek  
 HO uncle drive come ALLATIVE:COV way separate  
 'The uncle drove along until the crossroad.'

d. thúk thii hēn yùu th̄ūŋ hāathūm s̄ɔŋyaam  
 every time see be UNTIL:COV 11 pm 12 pm  
 'I usually see him here until 11 or 12 pm.'

e. faak khwaam khítth̄ūŋ th̄ūŋ khun pāa d̄uaj  
 deposit NOM regards RECIPIENT:COV HO aunt too  
 '(Please) give my regards to your aunt.'

f. mūuwa thorasàp paj th̄ūŋ bāan l̄éew  
 when telephone go TERMINAL:COV house finish  
 'When (Pingu's aunt) telephoned (Pingu's) house, ...'

FOCUS MARKER

g. kaan d̄əəŋthaŋ kin weelaa th̄ūŋ c̄et d̄uwan  
 NOM travel eat time EMPHASIS 7 months  
 'The travel took times as much as 7 months.'

DEGREE MARKER

h. kháw kr̄ot māak th̄ūŋkàp ph̄ut māj ð̄ok  
 he be angry much DEGREE speak NEG come out  
 'He was so angry that he could not speak.'

PREVERB

i. thammaj th̄ūŋ d̄âj khít wâa ph̄uak ph̄ii  
 why PREVERB possibility think say NUMBER this  
 pen nákmuay lâ? khráp  
 be boxer particle POLITE  
 'Why do you think we are the boxers?'

j. ph̄om paj th̄ii borisàt k̄ɔŋ l̄éew th̄ūŋ  
 I (MASC) go to company first then PREVERB  
 maa th̄iinii  
 come here  
 'I went to the company first, then came here.'

CLM

k. th̄ūŋ pen khon k̄e kháw k̄? māj lúk

EVEN IF:CLM be person old s/he CONJ NEG get up  
 hâj nâŋ ('even')  
 give sit  
 'He never offer the seat to anyone, even to old people.'

1. thǔŋ jaŋ dèk khâw kô? mii khwaam  
THOUGH:CLM still child s/he CONJ have NOM  
 rápphìtchâop (concessive)  
 responsibility  
 'Though he is still a child, he has a sense of responsibility.'

The last member of this type is *ləəj* 'to pass, exceed'. This item also shows its polyfunctionality in line with the other two items illustrated above.

**Table 4: The Chain of Polyfunctionality with *ləəj***

<u>Categories</u>	<i>ləəj</i> 'to pass, exceed'
V	to pass or exceed the reference point
Vd	past, through
COV	excessive
POSTVERB	past the reference time
EMPHASIS MARKER	manner adverbial 'immediately'
CLM	to emphasize the degree of an action or state determined by the speaker's attitude.
	showing the resultant state or the following action

(7) *ləəj*

V

a. thâa ləəj baŋpuu paj léew cà hăa  
 if PASS:V Bangpuu go PERFECT will find  
 ráan aahăan jâak dûay  
 shop food difficult too  
 'If (we) go beyond Bangpuu, it 's difficult to find the food shop.'

Vd

b. phõm mœŋ ləəj paj thîi koosõn  
 I (MASC) look PAST:Vd go at Koson  
 'I looked at Koson.'

COV

c. rótmee léen ləəj pâaj

bus run PAST:COV bus stop  
 ‘The bus ran past the bus stop.’

### POSTVERB

d. pâa bòøk pingu wâa “kin løøj câ”  
 aunt tell Pingu say eat IMMEDIATELY POLITE  
 ‘His aunt told Pingu that “Please have it (at once).”’

### EMPHASIS MARKER

e. Hûh! jaŋ cháaw yùu løøj  
 interjection still morning be EMPHASIS MARKER  
 ‘Huh! It’s too early to get up.’

### CLM

f. pingu tháŋ nùuway lé? híw løøj kin mó  
 Pingu both be tired and be hungry PASS:CLM eat all  
 ‘Pingu was tired and hungry, so he ate everything.’

g. pingu kô? làp paj lëew pâa løøj pít  
 Pingu CONJ sleep go PERFECT aunt THEN:CLM switch off  
 faj hâj nõõn há?  
 light give lie POLITE  
 ‘After Pingu went asleep, his aunt then switched off the light for him.’

Considering examples (5) to (7) together with (3) to (4) exemplified earlier, we can safely state that these verbs have extended their semantics from lexical words to grammatical words and developed polyfunctionality following the course of extension as in other serial verb languages.

In the next section, we try to explicate differences in the functions of these verbs as lexical verbs and grammatical markers by looking at their syntactic properties based on the analysis of *càak* and *thuŋŋ*.

### 3 Some Evidence from the Syntactic Side

The grammaticalized verbs exemplified in the preceding section do not appear with agent-oriented adverbials and take neither an obligation marker nor a negation marker, whereas fully lexical verbs do. These points are borne out by the following tests. First let us start with an obligation marker.

(8) a. kháw tõõŋ càak bâan këøt  
 s/he OBLIG LEAVE:V home to be born  
 ‘S/he had to leave her/his hometown.’

b. \*pingu dâjráp khanõmpaŋ tõõŋ càak khon khâaŋ thaŋŋ  
 Pingu receive bread OBLIG FROM:COV person wayside  
 ‘Pingu received some bread \*must from the wayside people.’

c. \*tõõŋ càak kaan thîi kháw khaj n  
 OBLIG BECAUSE NOM REL M s/he diligent

kháw cuŋ sòop dâj  
 s/he CONJ test POSSIBILITY  
 ‘\*Must because of his/her diligence, s/he can pass the exam.’

Second, the contrast can be attested by using the subject-oriented adverbials such as *rew* (early; quickly; fast) and the postverbal NP ellipsis rule.

(9) a. kháw thǔŋ bâan rew  
 s/he ARRIVE:V house early  
 ‘S/he arrived her/his house early.’

b. \*faak khwaam khítthǔŋ thǔŋ khun paa rew  
 deposit NOM regards TO:COV HO aunt quickly  
 ‘Please give my regards to your aunt \*quickly.’

c. \*thǔŋ rew jaŋ dèk kháw kô? mii khwaam  
 THOUGH quickly still child s/he CONJ have NOM  
 rápphitčhôp  
 responsibility  
 ‘Though \*early/quickly s/he is still a child, s/he has the sense of responsibility.’

These examples show a clear-cut distinction in the words *càak* and *thǔŋ* between lexical uses as in (8a) and (9a) and deverbal uses as in (8b, c) and (9b, c). Third, the same point can be made with regard to the negative marker *mâj*.

(10) a. kháw jaŋ mâj thǔŋ (bâan)  
 s/he yet NEG ARRIVE:V (house)  
 ‘S/he has not arrived yet.’

b. \*kháw thoorasâp mâj thǔŋ phûuan  
 s/he phone NEG TO:COV friend  
 ‘S/he phoned \*not to her/his friend.’

c. \*mâj thǔŋ jaŋ dèk kháw kô? mii khwaam  
 NEG THOUGH still child s/he CONJ have NOM  
 rápphitčhôp  
 responsibility  
 ‘\*Not though s/he is a child, s/he has the sense of responsibility.’

From these pieces of evidence, we may assume that the two patterns of grammaticalization of motion verbs which we proposed earlier are well-established.

#### 4 Crosslinguistic Validity of Image Schemas

The motivation for these motion verbs to develop polyfunctionality can be attributed to the similarity of schematic meanings. That is, this phenomenon can be understood in terms of the ‘gestalt preserving’ nature of semantic extension (cf. Ohori 1995: 703, who discusses Japanese *kara* ‘from/since’ and other case marker-clause linkage marker parallels; see Talmy 1988; Sweetser 1988).<sup>7</sup> For a recent argument that image schematic meanings are

<sup>7</sup> The word ‘polyfunctionality’ adopted here is used to mean that a lexical word in the isolating languages, such as Thai, Vietnamese, Cambodian etc., commonly has a versatile usage without

not strictly preserved, see Matsumoto (1998) on Japanese *o motte* and *ni yotte*.

Now, let us draw out the image schemas of *càak* ‘to leave’, *thúŋg* ‘to arrive’, and *løøj* ‘to pass, exceed’. Based on our analysis above, these three verbs seem to preserve their basic image schemas in accordance with the general tendency. But at the same time, we find some aspects of the image schemas are modified to suit the communicative needs. The basic patterns of extension are:

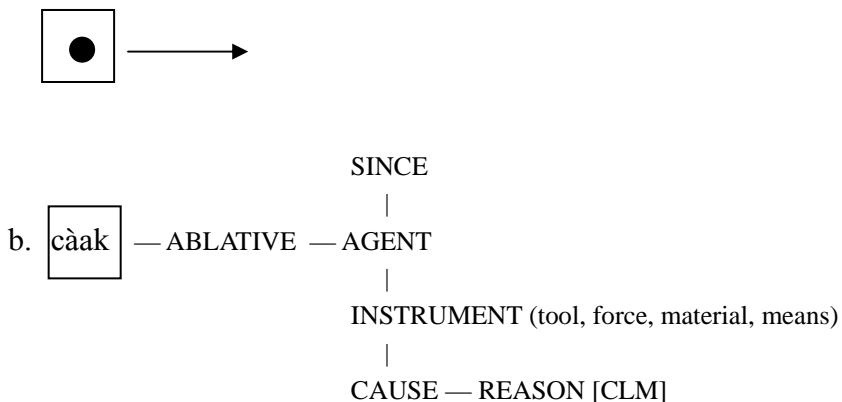
(11)	<b>SOURCE</b>	<b>GOAL</b>	<b>PATH</b>
SPATIAL	ablative	allative	excessive
TEMPORAL	since	until	past
ANIMATE	agent	recipient	-
LOGICAL	cause/reason	concessive	resultative

*Løøj* can take ANIMATE ARGUMENT for PATH as well, but there is no difference in function whether *løøj* takes ANIMATE or SPATIAL as its argument.

(12) rótmeé léen løøj pâaj/Mary  
bus run exceed bus-stop/Mary  
'The bus passed the bus-stop/Mary'

The image schemas of *càak* ‘to leave’, *thúŋg* ‘to arrive’, and *løøj* ‘to pass; exceed’, with illustrations of conceptual networks, are given in (13) through (15) respectively.

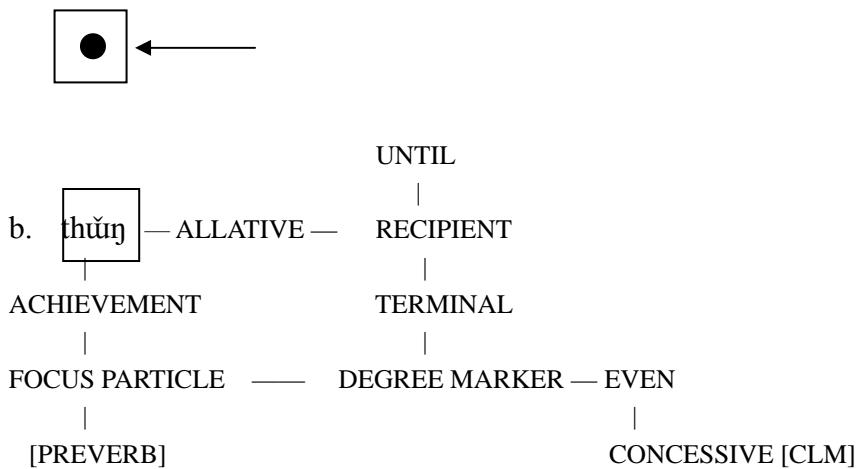
(13) a. *càak* ‘to leave’



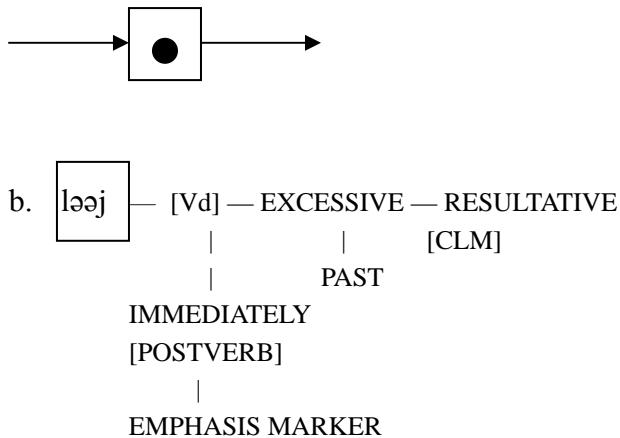

---

any overt morphological marking. Therefore, the lexicon has no clear-cut boundaries in grammatical categories, and yields a chain of interlinking loop as its meaning slightly changes due to the practical use. Note that the notion of polyfunctionality is different from polysemy in the sense that the semantic net or meaning chains of the so-called polyfunctional word are associated with various grammatical categories, whereas the various meanings of a polysemous word are connected with one grammatical category alone.

(14) a. *thǔŋj* ‘to arrive’



(15) a. *ləəj* ‘to pass, exceed’



## 5 General Remarks

Based on the above schemas, a number of interesting observations arise, some of which are given below.

First, does the same happen in other languages as well? Given the generality of image schemas, we would say that may be true in a lot of cases, but there are certain language-specific differences. For example, the use of the source marker to encode agents is limited in Thai. In a well-developed case marking language such as Japanese, it is possible to mark the agent of speech event verbs with *kara* ‘from’, but this is not the case in Thai.

(16) watasi-kara kodomo-ni chuui si-masu  
 I-FROM kid-DATIVE complain do-POLITE  
 ‘I will advise the kids.’

(17) \*càak chán cà dùu dèk  
 FROM I will scold kid

It would be interesting to see how broad the coverage of the adpositional uses of motion verbs can be in non-case marking languages.

Secondly, the concessive use of *thűŋ* ‘to arrive’ seems at first surprising. One may expect, for example, that it would mark result. We are not yet certain why this is the case, but one possibility is that a concessive interpretation implicitly assumes some scale. The form *thűŋ* in its deverbal use may mark an intermediate goal, and there may be a further, implicit goal which is the true endpoint. See the schema in (18).

(18) X does something or is at state Y with a further reference point of Y'

X ----- Y -----> Y'

In this connection, it seems interesting to see origins of concessive markers in typologically similar languages.

Finally, the present study is expected to promote our understanding of serial verbs (cf. Lefebvre 1991) and contribute to the typology of event integration (cf. Talmy 1991). According to Talmy's study, there are two basic types of language, namely verb-framed languages, where the notion of path is incorporated in the verb (as Spanish), and satellite-framed languages, where path is realized as a satellite, the verb incorporated manner (as in English). Talmy also notes that Japanese is a verb-framed language while Chinese a satellite-framed language. Then, what about other East Asian languages? Kessakul (1998) shows that Thai has both properties, though it is perhaps closer to a verb-framed language. In fact, this distinction seems best understood as a typological continuum. Li (1997) has argued that Modern-Day Chinese has changed from a verb-framed language into a satellite-framed language over the past millennium (and maybe more), showing that the verb-complement constructions encode path, state change, and realization. Although verbs we have taken up retain lexicality in Thai, this may not always be the case in other languages. For example, Chinese *shàng*, ‘to go up’ is used as a main lexical verb less frequently than its counterpart in Thai. Further inquiry into the details of event integration from the viewpoint of areal typology, including the study of serial verbs as a site of typological change, is expected in the near future.

## Abbreviations

CLF	classifier
CLM	clause-linkage marker
COMP	complement
CONJ	conjunction
CONJV	conjunctional verb
CONV	converb
COV	coverb
CV	causative verb
HO	honorific
LEX	lexicalization
MASC	masculine
NEG	negative
NOM	nominalizer

OBLIG	obligation
PRT	particle
REL M	relative marker
RES	resultative verb
SFM	sentence-final marker
TAM	tense-aspect-modality
Vd	directional verb
1SG/PL	1st person singular/plural
2SG/PL	2nd person singular/plural

## References

Areemit, Salee. 1986. *Functions and meanings of the word /laɔj/ in the Thai language*. Master's Thesis. Chulalongkorn University.

Bisang, Walter. 1996. Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and Mainland South East Asian languages. *Studies in Language* 20: 519-597.

Bilmes, Leela. 1995. The grammaticalization of Thai 'come' and 'go'. *BLS 21: Special Session on Discourse in Southeast Asian Languages*: 33-46.

Haspelmath, Martin. 1995. The converbs as a cross-linguistically valid category. In Haspelmath, Martin and Ekkehard König (eds.), *Converbs in Cross-Linguistic Perspective*: 1-55. Berlin: Mouton de Gruyter.

Heine, Bernd, Ulrike Claudi, Friederike Hünnemeyer. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: The University of Chicago Press.

Kessakul, Ruetaivan. 1998. Rethinking the Manner Predicates of Motion Verbs in Thai and Japanese. *Journal of Language and Information Sciences* 3. The University of Tokyo.

Kolver, Ulrike. 1984. Local Prepositions and Serial Verb Constructions in Thai, *AKUP* Nr. 56.

Lefebvre, Claire (ed.). 1991. *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*. Amsterdam: John Benjamins.

Li, Fengxiang. 1997. Cross-linguistic lexicalization patterns: diachronic evidence from verb-complement compounds in Chinese. *Sprachtypologie und Universalien Forschung* 50: 229- 252.

Lord, Carol. 1993. *Historical Change in Serial Verb Constructions*. Amsterdam: John Benjamins.

Matisoff, James A. 1991. Areal and universal dimensions of grammaticalization in Lahu. In: Traugott, Elizabeth and Bernd Heine (eds.) *Approaches to Grammaticalization*, vol. 2:383-453. Amsterdam: John Benjamins.

Matsumoto, Yo. 1998. Semantic change in the grammaticalization of verbs into postpositions in Japanese. In Ohori, Toshio (ed.), *Studies in Japanese Grammaticalization*: 25-60. Tokyo: Kuroshio.

Nguyen Dinh-Hoa. 1979. *201 Vietnamese Verbs*. New York: Barron.

Ohori, Toshio. 1995. Case markers and clause linkage: toward a semantic typology. In Casad, Eugene H. (ed.), *Cognitive Linguistics in the Redwoods*: 693-712. Berlin: Mouton de Gruyter.

\_\_\_\_\_. 1998. Polysemy and paragrammatic change in the Japanese conditional marker ba. In Ohori, Toshio (ed.), *Studies in Japanese Grammaticalization*: 135-162. Tokyo: Kuroshio.

Sweetser, Eve E. 1988. Grammaticalization and semantic bleaching. *BLS* 14:389-405.

Talmy, Leonard. 1988. The relation of grammar to cognition. In: Rudzka-Ostyn, Brygida (ed.), *Topics in Cognitive Linguistics*: 165-205. Amsterdam: John Benjamins.

\_\_\_\_\_. 1991. Path to realization: A typology of event conflation. *BLS* 17: 480-519.

Thepkanjana, Kingkarn. 1986. *Serial verb constructions in Thai*. Ph.D Dissertation. University of Michigan.

# **MORPHOLOGICAL STRUCTURE OF MALAY: USING PSYCHOLINGUISTIC ANALYSES OF RATED FAMILIARITY**

Lat Choo Lee, Susan J. Rickard Liow & May-Ling Olivia Wee

*National University of Singapore*

*<swksusan@nus.edu.sg>*

## **0 Introduction**

Psycholinguistic databases, listing properties of English words such as frequency and concreteness, have proved useful for the design of language assessment tools and experimental investigations of cognitive processes. These databases are now widely available for several other languages, including French, German, Dutch and Spanish, thereby facilitating work on bilingualism and cross-linguistic research. In this paper, we describe *how* and *why* we are developing a psycholinguistic database for Malay, which takes into account the relationship between morphological structure and rated word familiarity.

We started the database by asking 35 Malaysian first language speakers of Malay to rate a corpus of 4,328 words, comprising stem and affixed forms, on a 7-point familiarity scale. From this main database, we then extracted the mean ratings for 36 stem verbs and compared them to the mean ratings of their affixed forms: *teR-*, *meN-*, *di-*, *beR-kan*, *meN-...-kan* and *di-...-kan*. Our results provide support for several predictable patterns: stems are rated as more familiar than affixed forms, active forms are rated as more familiar than passive forms, and informal forms are rated as more familiar than formal forms. The utility of these kinds of analyses and the main psycholinguistic database is then illustrated with particular reference to the design of spelling and reading tests in Malay, and empirical studies of written language processing.

## **1 Psycholinguistic Analyses of Writing Scripts**

Words have many properties that can be quantified and stored in a psycholinguistic database, e.g., frequency, number of syllables, number of letters, number of phonemes, and type of morpheme. These databases are useful for the development of clinical and educational assessments as well as for the design of psycholinguistic experiments. By manipulating one property of a word with precision, while holding the others relatively constant, the clinician or researcher can make a more accurate assessment of a person's language processing skills.

For English, and many other alphabetic scripts, psycholinguistic databases have already provided the foundation for reliable diagnostic tools and coherent models of cognitive processing. The importance of calibrating word properties has led to computerized compilations, e.g., Coltheart's (1981) *M.R.C. Psycholinguistic Database* for English, and the Max Planck Institute's *CELEX Lexical Database* for Dutch, English, French and German (Center for Lexical Information 1995). Before these computerized

---

Mark Alves, Paul Sidwell & David Gil, eds. *SEALSVIII: papers from the 8th meeting of the Southeast Asian Linguistics Society (1998)*. Canberra, Pacific Linguistics, 2007, pp.109-119.

© Lat Choo Lee, Susan J. Rickard Liow, and May-Ling Olivia Wee.

resources became available, databases comprised a large number of printed pages which required painstaking word-matching by hand. Carroll, Davis and Richman's (1971) detailed work is a good example. These authors assembled a sizeable corpus of 5,088,721 English words, with different affixed forms listed separately. The resulting 805 pages of text, entitled *The American Heritage Word Frequency Book*, lists objective word frequencies for 86,741 different forms from 1,045 published materials. Other more modest databases, such as Toglia and Battig's (1978) *Handbook of Semantic Word Norms*, have been compiled using subjective 7-point ratings, rather than objective word counts, for a range of word properties including concreteness, number of associations, and familiarity.

Although databases utilizing subjective ratings are often smaller in size and they are not founded on word counts from printed matter, they can incorporate a broader range of properties that are salient for cognitive processing, such as imageability, and they retain the scaling used by a specific sample of readers. In a recent study of Chinese character processing, Rickard Liow, Tng and Lee (1999) used Toglia and Battig's method to calibrate semantic and phonetic regularity for Mandarin across subject pools from China, Singapore and Taiwan. Their results suggested that empirical research on the processing of Chinese characters demands more precision than has been observed to date, and that ratings are best obtained from *local* subject pools.

Given some of the advantages of subjective ratings, we reasoned that Toglia and Battig's method of scaling word properties would be a good starting point for a psycholinguistic database of the Malay language. In what follows, we describe *how* and *why* we collected and analyzed *familiarity* ratings.

## 2 Towards a Psycholinguistic Database for Malay

Malay languages belong to the Austronesian family and are used by more than 100 million people in South East Asia. Their respective standard forms are amongst the national languages for Indonesia, Malaysia, Singapore and Brunei, and there are also a large number of colloquial versions of Malay spoken throughout the region. The standard, or *formal* version<sup>1</sup> of Malay, is more commonly used for writing. Compared to English, formal Malay is very heavily inflected and the relationship between the orthography and phonology is much more predictable, both in the Arabic script and the more common Romanized script.

Despite the widespread use of informal and formal Malay and the potential for interesting cross-linguistic studies, no psycholinguistic database is yet available. Inevitably this has limited the development of standardized language assessment tools and the scope of empirical research on cognitive processing in Malay. For undergraduate projects, some pioneering students at the National University of Singapore gauged familiarity for 530 words (see Pereira, Rickard Liow & Mohd Saniff, 1992) and a search of the international literature yielded just *one* psycholinguistic study of Malay language processing published in English. This was a Stroop experiment conducted by Baluch and Abdullah-Darlan (1991) which compared reading in the Arabic and Romanised forms of the script. The scarcity of psycholinguistic research papers published in English on the Malay language suggests a brief review of the script would be useful here.

---

<sup>1</sup> We have sometimes used the terms standard and formal interchangeably but recognize that there are formal and informal forms of particular words within standard Malay.

### 3 The Malay Writing Script

Both forms of the Malay writing script are alphabetic, but they are used in different domains. The Arabic form (Jawi) is now used almost exclusively for the teachings of Islam, whilst the newer Romanized form (Rumi) is used for all other kinds of education and almost all types of printed materials. For this reason, and because comparisons with English are more feasible, we restricted our study to the Romanized form.

*Orthography and Phonology.* We noted earlier that the relationship between Malay orthography and phonology is more predictable (or regular) than it is for English. In Singapore and Malaysia, the grapheme-phoneme (letter- to- sound) correspondence is perfect for non-loan words with the exception of the letter ‘e’, which carries two phonological forms in standard Malay (but only one Indonesia). Although the grapheme-phoneme mappings are sometimes different from those in English (e.g., ‘c’ sounds like ‘ch’ in church), the same set of vowels and consonants are in use (except the letter ‘x’ which is found in foreign words).

Unlike English, there are only three diphthongs in Malay (*ai*, *au* and *oi*) and most syllables are very short. Possible syllable structures include *v*, *vc*, *vcc*, *ccv*, *cv*, *cvv*, *cvc*, *cvcc*, and *ccvc*, but when the vowels ‘a’ and ‘i’ or ‘a’ and ‘u’ are adjacent in a closed syllable, they are pronounced with a syllable boundary between them, e.g., *lain* (other) and *laut* (sea) are both two syllable words.

*Morphology.* We also noted earlier that affixation is more common in Malay than in English. At least nine prefixes (*meN*<sup>2</sup>-, *beR*-, *teR*-, *di*-, *peN*-, *peR*-, *se*-, *ke*-, and *mempeR*-), three suffixes (-*kan*, -*i*, -*an*), four infixes (-*el*-, -*er*-, -*em*-, and -*in*-) and thirteen circumfixes (*meN*...-*kan*, *di*...-*kan*, *meN*...-*i*, *di*...-*i*, *beR*...-*an*, *beR*...-*kan*, *peN*...-*an*, *peR*...-*an*, *ke*...-*an*, *mempeR*...-*kan*, *dipeR*...-*kan*, *mempeR*...-*i*, and *dipeR*...-*i*) are widely used. Like affixation, reduplication<sup>3</sup> is also common in Malay, and so the nature of the morphology makes the number of letters and syllables per word much higher than in English: average word length is about seven letters, with about three syllables, but some words are much longer. For example, the 18-letter, 8-syllable word *membahagi-bahagikan* (to divide into smaller parts!) comprises a prefix, a reduplicated stem, and a suffix. See Karim, Onn, Musa and Mahmood (1996) for further details. Each of these affixes serves at least one semantic function depending on the word class (noun, adjective or verb) of the stem, and the word class often changes after affixation. For example, when the prefix *beR*- is attached to a noun *payung* ‘umbrella’, the resulting affixed word *berpayung* is a verb ‘using the umbrella’.

This kind of morphological structure suggests that the meaning of a relatively unfamiliar affixed word such as *menderas* ‘move quickly’ can usually be predicted by putting the meanings of its more familiar stem, *deras* ‘quick’ and its affix together, and conversely, the meaning of a relatively unfamiliar stem such as *sarap* ‘have breakfast’ can be predicted by splitting the more familiar affixed word *sarapan* ‘breakfast’ into its stem and affix. However, exceptions occur when the meaning of a stem or affixed form cannot

<sup>2</sup> Depending on the initial letter of the stem to which the affix is attached, there may be morphological assimilation, e.g., *meN* + *buka* = *membuka*. In some instances, the initial letter of the stem word is dropped after affixation, e.g., *peN* + *tari* = *penari*. For more details on morphological assimilation, refer to Koh (1978).

<sup>3</sup> Reduplication is used to express indefinite plurality, intensity and repetition. Refer to Heah (1989), for more details.

be deduced by adding or subtracting the meanings of the components, e.g., *salin* 'to change' and *bersalin* 'to give birth'.

These patterns, and the exceptions, strongly suggest that the relationship between familiarity and morphology in Malay is likely to be particularly important for the development of language assessment tools and empirical research (see Lee, 2000). For this reason, we decided to start developing a psycholinguistic database by focusing on familiarity and morphology. In this paper, we describe how we collected mean familiarity ratings from a large, representative sample ( $N=4,328$ ) of stem and affixed words, and then for illustrative purposes, we analysed the subset of 36 stems which carried the greatest number of different inflectional forms.

#### 4 Method

Using Toglia and Battig's methodology (Ibid.), a 7-point familiarity rating scale was constructed in questionnaire form, in which 1 = very unfamiliar and 7 = very familiar. All the Malay words were typed in lower case and order of presentation in the list was randomised.

*Participants.* Thirty-five Malaysian first language speakers of Malay between the ages of 18 and 39 participated in this study and were paid RM 50.00 (USD \$14) each for their help.

*Materials.* To ensure that the participants would have been exposed to most of the words at some time, a corpus of 4,328 different lemma were compiled from the Primary 1 to Primary 3 textbooks used in Malaysian schools plus and from a Malay-English dictionary (Kamus Harian Federal, Daud 1945) that is used by Primary 4 pupils to lower Secondary school level (i.e., vocabulary of 6 – 14 year olds). In the final corpus, 2,510 items were stem words, 1,625 were stem words with affixes, 123 were reduplicated words, and 70 were reduplicated words with affixes. Word length ranged from two to eighteen letters (mean = 6.924; s.d. = 2.237), and from one to eight syllables (mean = 2.77; s.d.= 0.869).

*Procedure.* Participants were asked to rate each of the 4,328 words in the corpus on the 7-point familiarity scale using the following instructions (written in Malay) which are similar to those of Toglia and Battig:

For every individual, there are words that are frequently encountered and experienced in daily life. The aim of this experiment is to collect ratings on the following list of words according to their familiarity. For words that you frequently encounter or experience, a high rating should be given. Conversely, for words that you seldom encounter or experience, a low rating should be given. Circle one of the numbers given on the 7-point scale to indicate your own familiarity with the words given in the list. If you do not recognize a given word, you are required to cancel it out. Please try to use all 7 numbers on the scale whilst giving your familiarity ratings. Do not limit yourself to using only 3 and 4, or 1 and 7. There are no correct answers in this familiarity rating task.

#### 5 Results

For each of the 4,328 words, the ratings given by all 35 subjects were pooled together, and means and standard deviations were calculated. A subset of 116 mean ratings, comprising the 36 stem verbs with their affixed forms, was then extracted to enable a more detailed

analysis of the relationship between inflectional morphology and rated familiarity. The rank order of these stems and affixed forms is shown in Table 1.

**Table 1: Rated Familiarity of Stem and Affixed Forms in Descending Order**

Type	Characteristics	Mean	Std. Deviation
Stem	Stem	5.900	.7270
-kan	Suffix; Informal; Active	5.129	.3795
teR-	Prefix	5.100	.3932
meN-	Prefix; Transitive; Active	4.973	.3777
di-	Prefix; Passive	4.847	.2662
beR-	Prefix; Intransitive	4.837	.8830
di-...-kan	Circumfix; Formal; Passive	4.697	.2214
meN-...-kan	Circumfix; Formal; Active	4.664	.1816

Overall, the stems, e.g., *buat* ‘make’, ranked higher in familiarity than their affixed forms *membuat* ‘make’, and within the affixed forms, active forms, e.g., *menghantar* ‘deliver’, ranked higher than passive forms *dihantar* ‘to be delivered’, transitive forms, e.g., *mengejar* ‘chase’ ranked higher than intransitive *berkejar* ‘rush’, and informal forms, e.g., *ingatkan* ‘to remember’, ranked higher than formal *mengingatkan* ‘to remember’.

Paired comparisons between stems at the individual word level and their respective affixed forms yielded the results shown in Table 2. In general, stems were rated as significantly more familiar than their corresponding affixed forms.

**Table 2: Comparison of 36 stem verbs and their respective affixed forms**

Affixed form is		Prefix				Inconsistency <sup>4</sup>	Informal active	Circumfix	
		Trans Active	Intrans	Passive	Informal active			Formal Active	Formal Passive
Word	Stem (S)	meN-(A)	beR- (B)	di- (C)	teR- (D)	-kan (E)	meN-...-kan (F)	di-...-kan (G)	
<i>ada</i> (have)	6.63		5.32*** <sup>5</sup>				4.85***	4.70***	
<i>bakar</i> (burn)	5.90	5.12**		4.76** *	5.32**				
<i>balas</i> (answer)	5.79	5.18*	5.09**	4.64** *					
<i>balik</i> (return)	6.88		4.47***		5.26***	4.97***			
<i>balut</i> (wrap)	5.43	4.76*	4.90*	4.65					
<i>bawa</i> (bring)	6.48	5.09***		4.68** *		4.97***			
<i>beli</i> (buy)	6.76	5.53***		5.12**		5.63***	4.76***		

<sup>4</sup> *teR-* has an inconsistent effect such that when it is added to an intransitive or transitive stem, the verb retains its base form but when it is added to a stem, the form can become either active or passive.

<sup>5</sup> \* indicates the two variables are significantly different at .05 level, \*\* significance at .01 level, and \*\*\* significance at 0.001 level.

				*				
Word	Stem (S)	meN- (A)	beR- (B)	di- (C)	teR- (D)	-kan (E)	meN-...-kan (F)	di-...-kan (G)
<i>belit</i> (coil around)	4.59	4.50	5.24*		4.32			
<i>benam</i> (soak)	3.37	4.06**	3.29		4.29***			
<i>beri</i> (give)	5.50	5.12		4.85		5.37	4.71*	4.67*
<i>buang</i> (throw)	6.40	5.35***		5.00** *	5.09***			
<i>buat</i> (make)	6.32	5.35**		5.52** *		5.40**	4.53***	
<i>buka</i> (open)	6.43	5.00***		5.15** *	5.65***	5.44**		
<i>celup</i> (dip)	5.32	4.29**	2.97***	4.27** *				
<i>dapat</i> (get)	6.56	4.82***			4.97***		4.71***	
Word	Stem (S)	meN- (A)	beR- (B)	di- (C)	teR- (D)	-kan (E)	meN-...-kan (F)	di-...-kan (G)
<i>fikir</i> (think)	6.12		5.97		5.26**		4.48***	
<i>hantar</i> (deliver)	6.38	5.09***		4.73** *				4.41***
<i>hias</i> (decorate)	5.38	5.06		4.74*			4.36***	
<i>hirau</i> (care)	4.33					4.21	4.62	
<i>ingat</i> (remember)	6.29				5.56**	5.62**	4.82***	
<i>jadi</i> (happen)	6.29				5.18***	4.85***	4.38***	4.85***
<i>jahit</i> (sew)	5.68	5.21*		4.74** *		4.91**		
<i>kejar</i> (chase)	5.88	5.12*	4.76***	4.74** *				
Word	Stem (S)	meN- (A)	beR- (B)	di- (C)	teR- (D)	-kan (E)	meN-...-kan (F)	di-...-kan (G)
<i>letak</i> (put)	6.15				4.97***	5.24***	4.59***	4.41***
<i>lipat</i> (fold)	6.21		5.41**	5.00** *				4.50***
<i>main</i> (play)	6.32		5.77				4.62***	4.61***
<i>masuk</i> (enter)	6.71				4.97***	5.61***	5.03***	4.82***
<i>perlu</i> (need)	5.53					4.91*	4.79**	4.94*
<i>pulang</i> (return)	5.50					4.94	4.59***	
<i>sedia</i> (prepare)	5.85		5.62			4.79***	4.97***	5.06**
<i>serah</i> (submit)	5.50		4.94		5.21		4.68**	
<i>siap</i> (complete)	6.45					5.47***	4.65***	
Word	Stem (S)	meN- (A)	beR- (B)	di- (C)	teR- (D)	-kan (E)	meN-...-kan (F)	di-...-kan (G)
<i>siram</i> (water)	5.29	4.62**	3.30***	4.63*				
<i>susun</i> (arrange)	5.59	5.21	4.97	4.91*		4.76**		

<i>tukar</i> (change)	6.15		5.53** *	4.88** *		5.24***	4.47***	
<i>tutup</i> (close)	6.44		4.68***	5.09** *	5.35***			
Mean	5.900	5.130	5.100	4.973	4.847	4.837	4.697	4.664
Std. Dev.	0.727	0.38	0.393	0.378	0.266	0.883	0.221	0.182
N	36	18	14	19	19	17	10	19
Min	3.37	4.21	4.29	4.06	4.27	2.97	4.41	4.36
Max	6.88	5.63	5.65	5.53	5.52	5.97	5.06	5.03

In addition to the single verb analyses shown in Table 2, four pairs of comparisons amongst the affixed forms were of interest: *meN-* (active prefix) and *di-* (passive prefix), *meN-* (transitive prefix) and *beR-* (intransitive prefix), *-kan* (informal suffix) and *meN-...-kan* (formal circumfix) and *di-...-kan* (passive circumfix) and *meN-...-kan* (active circumfix). Each affixed form in a pair differs from the other form in terms of one characteristic, for example, *meN-...-kan* is active and formal whilst *-kan* is active but informal.

As Table 3 shows, only two of these overall pairwise comparisons showed statistically significant differences in their familiarity ratings: words with the prefix *meN-* (active form) were rated significantly more familiar than those with the prefix *di-* (passive form), and words with the suffix *-kan* (informal form) were rated more familiar than those with the circumfix *meN-...-kan* (formal form). Without a larger sample of affixed words, we cannot be sure whether the other two pairwise comparisons would yield significant differences, but our preliminary data suggest that while *meN-* (transitive form) might be rated as more familiar than *beR-* (intransitive form), the addition of the *-kan* suffix to make the circumfixes *meN-...-kan* and *di-...-kan* seems to negate the difference between the active and passive forms.

**Table 3: Overall comparison of affixes (paired t-tests)**

Mean familiarity ratings for affixed pair		Characteristics	N	t	p
meN- 4.718	beR- 4.315	Transitive / intransitive	8	1.6	.154
meN- 5.069	di- 4.821	Active / passive	16	4.533	.001
-kan 5.175	meN-...-kan 4.685	Informal / formal	13	4.274	.001
meN-..-kan 4.743	di-...-kan 4.758	Active / passive	8	-.190	.855

## 6 Discussion

The purpose of this study was to describe how and why we are developing a psycholinguistic database for Malay. Recall that we collected mean familiarity ratings on 4,328 words from 35 first language speakers. From this main database, we then extracted the mean familiarity ratings for 36 stem verbs and compared these to a series of affixed

forms: *teR-, meN-, di-, beR-, -kan, meN-...-kan and di-...-kan*. Our results suggest that there is a relationship between familiarity and morphology: stems are rated as more familiar than affixed forms, active forms are rated as more familiar than passive forms, and informal forms are rated as more familiar than formal forms. Now we will illustrate the utility of the detailed analyses, and the main database, with particular reference to our current work on standardized language assessments and cognitive models of reading.

### 6.1 Standardized Language Assessments

With the aid of our familiarity ratings, Lee developed spelling and reading tests for children attending primary schools in West Malaysia (Lee, 2000; Lee & Rickard Liow, 1999).

*Malay Spelling Test.* For literacy tests in English, lists of items are usually ordered roughly in terms of frequency and they include regular and irregular words. For Lee's test, the spelling list was initially ordered using our familiarity ratings, and it included words with the letter 'e' (to test the child's knowledge of the single irregular grapheme-phoneme correspondence) and some words which re-syllabify after affixation (to test the child's knowledge of morphological processes).

*Malay Reading Tests.* Lee has also used the familiarity ratings to devise single word and text-based reading tests. The single-word reading test was developed along the same lines as the Malay Spelling Test (described above) and her text-based test is similar to the (English Ibid.) Neale Analysis of Reading Ability (NARA). The NARA, which is used to assess reading rate, reading accuracy and reading comprehension, comprises a series of passages that are graded in terms of word frequency and sentence complexity. Likewise, Lee varied the level of difficulty in the passages by controlling the type of inflectional morphemes and their familiarity ratings such that the difficult passages have more complex morphological structures (more affixation and reduplication), and the words are generally less familiar. Given that the Romanized script is almost totally regular, we think that reading accuracy may be relatively easy to attain in Malay compared to English, whereas comprehension of complex morphological structures will be a major obstacle to independent reading for children, especially those with limited aural exposure to formal Malay.

Lee also developed standardized versions of these spelling and reading tests for 7- to 9-year-olds, with separate norms for Malaysian children from Chinese- and Malay-first-language backgrounds. Without a psycholinguistic database, these kinds of language assessment tools, and many others that await development, e.g., tests of receptive and expressive vocabulary, could not have been designed with any confidence<sup>6</sup>.

### 6.2 Cognitive Models of Reading

Dual route models of reading (e.g., Morton, 1979; Morton and Patterson, 1980; Coltheart, Curtis, Atkins and Haller, 1993) provide a coherent account of the cognitive processing of the English alphabetic script by skilled unilingual readers. However, there are two kinds of evidence to suggest that these models may not be applicable to other alphabetic scripts,

---

<sup>6</sup> Since the 8th SEALS conference, Tye & Rickard Liow (submitted) have used the familiarity ratings to investigate language representation with case studies of Malay-speaking dysphasic adults.

especially those with more regular orthography-phonology or more regular orthography-morphology relationships.

First, Wimmer and Goswami (1994) found that children reading German (can and do) make greater use of the non-lexical route than children reading English. Second, no reading models (including the dual route type) can yet account for the visual search strategies we have observed in biscriptal bilingual (Malay-English) readers, e.g., Green, Rickard Liow, Tng and Zielinski (1996), and Rickard Liow, Green & Tam (1999).

Rickard Liow and Wee (1999) predicted that these different visual search strategies might be attributable to skilled readers' sensitivity to higher level psycholinguistic variables such as inflectional morphology which is well contrasted between the English and Malay scripts despite their graphemic similarity. They designed a cross-linguistic experiment to investigate affixation and lexicality using two paradigms: *visual search* to look at prefix versus stem position effects directly, and *lexical decision* to strengthen the links with previous research on unilingual English readers. The results showed support for our hypothesis that top-down processing reflects knowledge of morphological structure, especially in skilled readers of Malay.

For this kind of cross-linguistic research, it is important to control (or manipulate) word properties such as familiarity and number of letters, across the two languages. The corpus we described above was drawn from school textbooks only and it proved insufficient for these experiments. However, using the same principles, Rickard Liow and Wee were able to collect familiarity ratings from Malay-English bilinguals on sets of eight-letter stem and affixed words in the two languages. The raw familiarity ratings cannot be used directly for selecting words (because a rating of '1' on the English scale may not be the same as a value of '1' on the Malay rating scale) so we first centile-ranked the words in each language, and then matched their ranks. Clearly a psycholinguistic database for Malay can increase the validity and reliability of empirical studies of reading by enabling greater precision in the selection of experimental stimuli.

To summarize, there are many descriptive books on the Malay language, for example, Hassan (1974), Koh (1978), and Karim et. al. (1996), but none provide a suitable resource that can form the foundation for empirical work. In this paper, we have argued that a psycholinguistic database is crucial for the design of clinical and educational assessments and for extending cognitive models of language processing hitherto based on unilingual English-speaking populations. The corpus we have documented here, together with its implications, is limited in that we calibrated only familiarity and inflectional morphology for a small sample of words. However, we have shown *how* and *why* a database for Malay can and should be developed.

### Acknowledgements

The first author collected these familiarity ratings as a foundation for her Masters thesis on children's reading and spelling skills in Malay. We are grateful to the National University of Singapore for scholarship support, to the staff and students of Universiti Kebangsaan Malaysia (Kuala Lumpur), Sekolah Menengah Kebangsaan King Edward (Taiping, Perak), and Sekolah Menengah Kebangsaan St. George (Taiping, Perak), for their help with data collection, and to Nurai'sha Bte Hassan for her involvement at the SEALS conference.

Correspondence concerning this paper and request for a soft-copy of the familiarity database, should be addressed to A/P Susan J. Rickard Liow, Department of Social Work

and Psychology, National University of Singapore, Kent Ridge Crescent, Singapore 119260, fax (65) 67781213 or email: <swksusan@nus.edu.sg>

## References

Bahasa Malaysia Tahun 1, 2, 3. *Kurikulum Bersepadu Sekolah Rendah*. 1995. Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka & Pustaka Cipta.

Baluch, B. & Abdullah-Darlan, K. 1991. On processing the Roman and Arabic orthographies of Malay: Some implications from the Stroop colour-word test. *Psychological Reports*, 69, 2, (Oct) 562-569.

Carroll, J.B., Davies, P. & Richman, B. 1971. *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin.

Centre for Lexical Information. 1995. *The Celex Lexical Database Release 2*. Nijmegen, The Netherlands: Max Planck Institute for Psycholinguistics.

Coltheart, M. 1981. The Medical Research Council Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 33A, 497-505.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. 1993. Models of reading aloud: Dual route and parallel distributed approaches. *Psychological Review*, 100, 589-608.

Daud, M.S. 1995. *Kamus Harian Federal*. Kuala Lumpur, Malaysia: Federal Publications.

Green, D.W., Rickard Liow S.J., Tng, S.K. & Zielinski, S. 1996. Are visual search procedures adapted to the nature of the script? *British Journal of Psychology*, 87, 311-326.

Hassan, A. 1974. *The Morphology of Malay*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Heah, C.L-H. 1989. *The Influence of English on the Lexical Expansion of Bahasa Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Karim, N.S., Onn, F.M., Musa, H.H., & Mahmood, A.H. 1996. *Tatabahasa Dewan. (Edisi Baharu)*. Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.

Koh, B.B. 1978. *The Teaching of Malay Affixes*. Siri Pengarajan Fajar Bakti. Kuala Lumpur, Malaysia: Penerbit Fajar Bakti.

Kucera, H. & Francis, W.N. 1967. *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

Lee, L.C. 2000. *Development of Literacy Skills in the Malay Language: Comparison of L1 (Malay) and L2 (Chinese) children in Malaysia*. Masters Thesis, National University of Singapore.

Lee, L.C. & Rickard Liow, S.J. 1999. *Malay Standardised Reading and Spelling Tests*. Paper presented at the SHA(S) & NUS Joint Workshop on Malay Language Acquisition. National University of Singapore. (Jan).

Morton, J. 1979. Word Recognition. In J. Morton and J. Marshall (eds), *Psycholinguistics Series*, Vol. 2. London: Elek.

Morton, J. & Patterson, K. 1980. A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson and J.C. Marshall (eds), *Deep Dyslexia*. London: Routledge and Kegan Paul.

Neale, M.D. 1989. *Neale Analysis of Reading Ability*. (Revised British Edition). Great Britain: NFER-NELSON.

Pereira, V., Rickard Liow, S.J. & Mohd Saniff, S. 1992. Familiarity Ratings for 530 Malay Words. *Working Paper No. 42. Department of Social Work and Psychology*, National University of Singapore, Singapore.

Rickard Liow, S.J., Green, D.W. & Tam, L-J, M-M. 1999. The development of visual search strategies in Malay-English and Chinese-English bисcripts. *International Journal of Bilingualism*, 3, 333-349.

Rickard Liow, S.J., Tng, S.K. & Lee, C.L. 1999. Chinese characters: Semantic and phonetic regularity norms for China, Singapore and Taiwan. *Behaviour Research Methods, Instruments & Computers*, 31, 155-177.

Rickard Liow, S.J. & Wee, M-L., O. 1999. *Reading Strategies of Skilled Bilingual Bисcripts*. Paper presented at the SHA(S) & NUS Joint Workshop on Malay Language Acquisition. National University of Singapore. (Jan.)

Toglia, M.P. & Battig, W.F. 1978. *Handbook of Semantic Word Norms*. Hillsdale, New Jersey: Lawrence Erlbaum.

Tye, W.P. & Rickard Liow, S.J. (submitted). *Malay language processing: Evidence from case studies of dysphasia*. Paper submitted for the International Symposium of Linguistics & Speech-Hearing Sciences, to be held in Kuala Lumpur (Oct, 2002).

Wimmer, H. & Goswami, U. 1994. The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition* 51: 91-103.



# THE DIRECTION OF MONOSYLLABICITY IN RAGLAI

Nguyen Van Hue

*National University  
Ho Chi Minh City*

## 1 Introduction

Raglai is a Chamic language spoken by more than 72,000 people in the highlands of South Central Vietnam. Raglai can be divided into two main dialects: the Northern Raglai (located mainly in Khanh Hoa and some parts of Ninh Thuan Province) and the Southern Raglai (located in Ninh Thuan and Binh Thuan Province). Of the Chamic languages, such as Cham, Ede and Churu, Raglai is considered most representative of Proto-Chamic due to its relatively conservative characteristics, especially its presyllables.

Although there is still an abundance of polysyllabic words in the Raglai lexicon (about 40 percent in Southern Raglai and up to 43 percent in Northern Raglai), Raglai is now changing into a monosyllabic language like its neighboring languages in the highlands of Vietnam. The cause of this important change is ‘the areal pressure in the direction of monosyllabicity’ (Lee 1966: 76). By studying what is changing in the direction of monosyllabicity in this language, we can know more clearly the way that other Chamic languages have gone.

## 2 The Phonological Word in Raglai

Like other neighboring languages, such as the Mon-Khmer languages Mnong and Koho and the Chamic languages Cham, Ede, and Churu, a phonological word in Raglai can be monosyllabic, disyllabic, or trisyllabic. The polysyllabic word consists of one main-syllable with primary stress and from one to two presyllables preceding the main-syllable. In the case of trisyllabic words, the initial presyllable has a secondary stress.

The presyllable in Raglai has a simple phonological structure. It consists of only one initial consonant and one vowel.

Initial consonants of presyllables: The list of initial consonants of presyllable in Northern Raglai is shown in Table 1. In this list, voiceless stops (except [c]) occur most often. Voiced stop initials are found in fewer words.

**Table 1:** *Raglai presyllable initials*

<b>Voiceless stops</b>	p-	t-	c-	k-	?
<b>Voiced stops</b>	b-	d-	ɟ-	g-	
<b>Nasal stops</b>	m-				
<b>Fricatives</b>		s-	ʃ-		h-
<b>Liquids</b>		l-, r-			

Vowels of presyllables: One characteristic that distinguishes Raglai from other Mon-Khmer languages which only have a schwa [ə] in the presyllable is the vowel system of presyllables. Phonetically, there are 3 basic vowels [a], [u], [i], and the mid-vowel [o]. In some phonetic contexts, the schwa [ə] can be heard instead of [i], [u], or [a]. The mid-vowel [o] is found in some disyllabic words which have a mid-vowel in the main-syllable (Ex: *horəj* ‘day’, *lopeh* ‘thin’).

There is no contrastiveness in the presyllabic vowel, and in some cases, an alternation is found.

‘thinking’	sining ~ saning
‘tale’	julukal ~ jalakal

Because of the loss of phonemic contrastiveness of vowels in presyllable, the presyllable in Raglai (and also in many languages in the area) has a rather special status in the phonologic word. Phonologically, the structure of the disyllabic word is /CCVC/.

### 3 Changes in the Direction of Monosyllabicity in Raglai

By comparing the phonological words of Raglai with those of Proto-Chamic and Proto-Austronesian and between the two dialects of Raglai, phonetic changes in the direction of monosyllabicity in Raglai can be grouped into two stages.

Stage One: The changes are in Proto-Chamic language because these changes were shared innovations in all daughter languages. The most important change in this stage is the shift of presyllabic initial voiced stops to voiceless stops when they were followed by a voiceless stop of the main-syllable, as in Table 2. In this stage, some other shared innovations were found such as the shift of initial voiced stops to preglottalized stops (e.g., \**buhuk* > <sup>?</sup>*buk* ‘hair’; \**bahuh* > <sup>?</sup>*bəu* ‘smell’).

**Table 2: Devoicing in Proto-Chamic**

Proto-Austronesian	Proto-Chamic	Raglai	Gloss
* <i>batuk</i>	* <i>pituk</i>	<i>pituk</i>	‘cough’
* <i>bitis</i>	* <i>patih</i>	<i>patih</i>	‘thigh’
* <i>gatel</i>	* <i>katal</i>	<i>katal</i>	‘itchy’

Stage Two: In this stage, Raglai underwent some phonetic changes like other neighboring languages. Generally speaking, Southern Raglai changes faster than Northern Raglai because the former has a close contact with Cham that has gone very far towards monosyllabicity. In particular, some peripheral areas of both dialects have many changes and their phonological words are very similar to those of Cham.

There are several main changes in the presyllable in this stage. There was a loss of phonemic contrastiveness of presyllabic vowels, although there are still three vowels phonetically. Because of the loss of phonemic contrastiveness, vowel alternation is common. Moreover, there is the assimilation with the main-syllabic vowel. Also common was the deletion of vowels especially when the initial consonant of the main-syllable is [h], [l], or [r]. By dropping the presyllabic vowel, aspirated consonants or consonant clusters

were formed, as shown in Table 3. Finally, as noted above, the initial voiced stops changed into voiceless stops when the following stop was a voiceless one, as shown in Table 4. In this stage, the change was non-conditional in Southern Raglai.

**Table 3: Telescoping**

	<b>N. Raglai</b>	<b>S. Raglai</b>
‘old’	taha	tha
‘feather’	bilu	plu
‘husked rice’	barah	rah

**Table 4: Initial devoicing**

	<b>N. Raglai</b>	<b>S. Raglai</b>
‘mushroom’	bumau	pimau
‘eight’	dalapan	talapan

#### 4 Conclusion

Phonetic changes in Raglai, in general, are similar to those of many languages in the area. In Raglai, due to its relative isolation, the changes are just beginning. Thus, we can easily follow its changes and know how one Malayo-Polynesian language changes into a monosyllabic language and how the gradual loss of phonemic contrastiveness of presyllabic vowels has proceeded.

#### References

Blust, R. A. 1988. *Austronesian root theory: An essay on the limit of morphology*. Amsterdam/Philadelphia: John Benjamin.

Dempwolff, O. 1934-38. *Comparative phonology of Austronesian word list* (Translation by Ateneo de Manila University, Loyola Heights, Quezon City, 1971).

Lee, E. W. 1966. *Proto-Chamic Phonologic Word and Vocabulary*. Indiana University, dissertation.

Moussay, G. 1971. *Dictionnaire Cam-Vietnamien-Française*. Phan Rang, Vietnam: Cham Cultural Center.



# TOPIC SELECTION AND ORGANIZATION AND DISCOURSE STRATEGIES IN MALAY NEWS IN FOUR COUNTRIES

Gloria R. Poedjosoedarmo

*National Institute of Education Singapore*

*<poedjo@singnet.com.sg>*

## Introduction

Malay is an official language in four Southeast Asian countries: Malaysia, Indonesia, Brunei and Singapore. MABBIM (Majalis Bahasa Brunei, Indonesia, Malaysia), a regional body, in conjunction with local government bodies, has attempted to formulate a 'standard' form of the language. Work to date has consisted mainly in standardizing the spelling (two and a half decades ago), and proposing vocabulary adoption, especially in connection with recent scientific and technological advances.

Despite these efforts, Malay speakers from different countries still complain of difficulty in understanding each other. How much of this difficulty is due to differences in lexicon, to differences in grammatical structures or to differences in discourse organization? This paper compares some of these features in similar texts originating from the countries in question in order to determine the nature of differences in discourse structure and organization as well as discourse-related grammatical features which might partially account for problems in cross-national intelligibility. Specifically the paper examines TV news broadcasts and news articles found on the internet, as the topics discussed in the various countries within these genre are similar and the genre features in each case (news broadcasts and Internet articles) are also fairly constant.

Features examined include (1) selection and ordering of topics, and length of each item, (2) similarities between English and Malay broadcasts in Singapore (to determine to what extent translation might be responsible for any features of the Malay broadcasts in Singapore which differed from those of the other Malay-speaking countries), and (3) grammatical choices (i.e. active vs. passive) in Malaysian vs. Indonesian Internet news articles.

## Some Notes on Problems in Data Collection and Methodology

The original plan for this project had been to collect TV news broadcasts in the four countries during the same time period, on the assumption that the topics would thus be similar. TV news broadcasts were selected as the genre because broadcasting constraints result in discourses of similar length and structure. It was felt that any differences between them would thus truly reflect differences in discourse strategies used by speakers of the language in the four countries (as well as differences in national preoccupations in terms of topic selection and topic prominence).

The recordings were made on 3, 4 and 6 February 1998. During this week topics on the international news included (1) the US balanced budget, (2) Clinton's sex scandal, (3) an airline crash in the Philippines, (4) UN weapons inspection in Iraq, (5) the Asian

economic crisis. Regional topics included (1) the double Chinese New Year and Moslem Hari Raya holiday and (2) the hot dry weather.

As I was not able to record the news myself in all four countries simultaneously, I relied on friends and relatives to assist me.<sup>1</sup> Unfortunately the video tape on which recordings were made in Indonesia turned out to be damaged and unusable. The recordings made in Brunei were on audio tape only and the sound quality was somewhat inferior. It was not possible to transcribe the contents of this tape word for word. However, topics could be identified and timed.

As damage to the Indonesian tape was only discovered months after the recordings were made, the initial comparison of TV news broadcasts was limited to three countries: Malaysia, Singapore and Brunei. An opportunity arose several months later to record some TV news broadcasts in Indonesia but it was not possible to obtain simultaneous recordings of news broadcasts in the other countries. General features of topic selection, arrangement and timing found in the Indonesian news broadcasts were, however, noted.

Due to the problems encountered in collecting news broadcasts simultaneously from the four countries, Internet news resources were explored as an alternative source of texts. The advantage of using these is that, since they are archived, it is possible to examine selection and length of topics from previous dates. This source, however, like newspaper articles, has the disadvantage over TV news broadcasts, that there are less stringent restrictions concerning overall length of an issue.

In this paper TV news broadcasts will be examined first. Then Internet sources will be discussed. Finally, there will be a discussion of conclusions that can be extracted from this analysis regarding overall tendencies in use of the formal style of language in the four countries and implications of these tendencies for cross-national intelligibility.

### **The TV News Broadcasts in four Malay-speaking Countries**

In this section, three sets of data will be analysed. These include (1) TV news broadcasts from Malaysia, Singapore and Brunei on 3 February 1998, (2) English and Malay news broadcasts from Singapore on 4 February 1998, and (3) an Indonesian TV news broadcast of 11 July 1998.

### **Malaysian, Singaporean and Bruneian News of 3 February 1998**

I will list here the topics covered by each country in the order of coverage and indicate the time devoted to each topic. A comparison of similarities and differences will then be made.

#### Malaysian news (TV 3) 3 February 98

- Traffic accidents over the double holiday (nearly 15 min)
- Clinton's sex scandal (1 minute)
- Cebu Pacific plane crash in the Philippines (1 min 5 sec)
- UN weapons inspection in Iraq (2 min 20 sec)
- Civil strife in Somalia (1 min)
- Medical breakthrough in Europe – using neutrons to attack damaged brain cells (53 sec)

---

<sup>1</sup> I wish to thank Peter Martin for recording the news programmes in Brunei and my son Krisnadi Poedjosoedarmo for recording the news programmes in Indonesia.

- Sports (7 min)
- Weather (less than 1 min)

Singapore Malay news – Prime 12 – 3 February 98

- PM Goh's visit to Indonesia (1 min 30 sec)
- Min. of Ed. Teo in Switzerland says Eastern Asia will not be defeated by the current economic crisis (1 min 10 sec)
- Asian economic crisis in Japan and Malaysia (1 min 25 sec)
- Asian stock market and Dow Jones (2 min 11 sec)
- America's balanced budget (1 min 25 sec)
- Seagate (Thailand) lays off workers (37 sec)
- German factory begins production of silicon wafers in Singapore (1 min 1 sec)
- Local crime story – man attacked in coffee shop later dies (1 min)
- A bag containing money and a passport was stolen from a man waiting at the bus terminal to go to Malaysia (50 sec)
- Two brothers (in Singapore) are accused of receiving bribes worth millions (53 sec)
- “Gong Xi Raya” celebrations in Singapore schools (53 sec)
- Cebu Pacific airline crash (1 min 12 sec)
- Iraqi crisis – UN weapons inspection (4 min 22 sec)
- Death penalty to convicted woman in Texas (56 sec)

Sports:

- tennis (37 sec)
- soccer (1 min 10 sec)
- riots in Italy (1 min)
- Local stocks and foreign exchange (1 min 20 sec)
- Weather (41 sec)

Brunei – RTB – 3 February 98

- Hot dry weather
- causes forest fires
- handling of oil in refineries to prevent fires
- asthmatics should avoid dust (6 min 50 sec)
- Number of new converts to Islam (1 min 10 sec)
- Some companies continue to celebrate the dual holiday (20 sec)
- Local weather (27 sec)
- Cebu Pacific airline crash (2 min 19 sec)
- Iraqi crisis – UN weapons inspection (3 min 16 sec)
- Videos of Princess Diana's car (1 min 15 sec)
- Cigarette smoking causes fires (1 min 5 sec)
- Regional weather (18 sec)
- South Korean economy (1 min 32 sec)
- US balanced budget and more spending on education (1 min 28 sec)
- Sports (1 min 36 sec)

Common topics between the news broadcasts of the three countries included (1) the Cebu Pacific airline crash, (2) US – Iraqi relations and the UN weapons inspection, (3) sports and (4) the weather. A comparison of time spent on each of these topics in each country is given in the following chart:

(Topic)	Malaysia	Singapore	Brunei
<b>Cebu Pacific airline crash</b>	1 min 5 sec	1 min 12 sec	2 min 19 sec
<b>US – Iraqi relations &amp; UN weapons inspection</b>	2 min 20 sec	4 min 22 sec	3 min 16 sec
<b>Sports</b>	7 min	2 min 47 sec	1 min 36 sec
<b>Weather</b>	(less than one min)	41 sec	7 min 35 sec

In addition to identical topics found in the three news broadcasts there were some topics which were similar in certain respects but which were given different focus in different countries. These included the double holiday – Hari Raya and Chinese New Year – and the US presidency.

In connection with the double holiday, the Malaysian news spent 15 minutes discussing traffic accidents, the first item on the news that evening. Singapore had a 1 minute 16 second item on what they called ‘Gong Xi Raya’ celebrations in schools, celebrations combining the two holidays with cultural shows from each ethnic group. Brunei had two related items, one 20 second one on the fact that some companies continued to celebrate after the holiday had officially ended and a 1 minute 10 second item on new converts to Islam.

In regard to the US presidency, Malaysia had a 1 minute item on Clinton’s sex scandal which was followed on 4 February with a 54 second sequel. Singapore and Brunei both had items on the balanced budget, Singapore’s being 1 minute 25 seconds in length with a follow-up on 4 February of 54 seconds, and Brunei’s being 1 minute 28 seconds in length. Though the tape of the Indonesian news for that evening was not available, the Internet source of Kompas online had one article on 3 February and two on 4 February relating to the sex scandal. There was nothing in the archives on the balanced budget.

## Summary

If we compare the news broadcasts from Malaysian, Singaporean and Bruneian TV, we can make the following observations:

- Singapore’s news items tend to be quite short, generally under 2 minutes
- Brunei and especially Malaysia tend to have fewer items but some much longer (4 Feb TV3 had a story about rabies which lasted 5 min 36 sec)
- Brunei and Malaysia tend to have local news first while Singapore has “headline news” first which may be either local or international

### English and Malay News in Singapore

To what extent are the patterns found in Singapore's Malay news influenced by Singapore's English news? Baker (1992) has noted that frequent translation can affect the structure of the language and recent work on transfer suggests that transfer of discourse level features are more likely to persist than are surface level syntactic features (Gass & Selinker 1983). With this in mind, a comparison was done of the Malay and English news broadcasts in Singapore on 4 February 1998. Topics covered and time devoted to each are listed here first:

#### Singapore Malay news 4 February 98

- Prime Minister Goh back from Indonesian visit (1 min 43 sec)
- Regional exchange rates (54 sec)
- Economic crisis in Indonesia (45 sec)
- Who will be Indonesia's next VP? (35 sec)
- South Korea's financial crisis (54 sec)
- Clinton's balanced budget (45 sec)
- Textbooks to be selected from commercial companies (1 min 26 sec)
- Delegates from Crescent Girls' School to attend model UN meeting in Russia (53 sec)
- ASEAN nations to adopt common system for registration of new drugs (48 sec)
- Dealing with the aging population (50 sec)
- Heroin found – crackdown on drugs (59 sec)
- Local traffic accidents (1 min 3 sec)
- Events sponsored by Mendaki (37 sec)
- Stocks to be sold by Second Chance Enterprise (1 min 55 sec)
- Crisis in Iraq (1 min 21 sec)
- Sri Lanka's independence celebrations, riots and Prince Charles' visit (37 sec)
- Chinese defense minister and Japanese counterpart meet (50 sec)
- Cable accident in Italy – Clinton apologizes (57 sec)
- Cebu Pacific air crash – all presumed dead (44 sec)
- Hong Kong fire (27 sec)
- Texas woman receives death penalty (1 min 15 sec)
- Sports (26 sec)
- Stocks & exchange rate (1 min 14 sec)
- Weather (42 sec)

#### Singapore English news 4 February 98

- Textbooks to be published by commercial companies (1 min 22 sec)
- Local traffic accidents (1 min 27 sec)
- District court sentences 16 year olds who clashed in rival gangs (43 sec)
- Crackdown on drugs (42 sec)
- ASEAN health officials agree on common system for drug registration (1 min 26 sec)

- President Goh back from Indonesian trip (1 min 42 sec)
- Who will be Indonesia's next VP? (44 sec)
- South Korea's financial crisis (1 min 6 sec)
- Clinton's balanced budget (1 min 19 sec)
- Local traffic (less than 1 min)
- Chinese defense minister and Japanese counterpart meet (33 sec)
- Gulf crisis (1 min 32 sec)
- Sri Lanka's Independence celebration, riots and Prince Charles' visit (54 sec)
- Cable accident in Italy (57 sec)
- Cebu Pacific air crash (about 1 min)
- Local housing agents ask home owners to chip in to cover cost of advertising (1 min 40 sec)
- Crescent Girls' School delegation to model UN in Russia (1 min 16 sec)
- Sports (2 min 23 sec)
- Weather (1 min 19 sec)
- Stocks and exchange rates (36 sec)
- Repeat headlines (16 sec)

If we compare the time devoted to similar topics in the Malay and English news broadcasts in Singapore, we find that there is not a great deal of difference, as seen in the following chart:

(Topic)	Malay	English
Textbooks	1 min 26 sec	1 min 22 sec
Traffic accidents	1 min 3 sec	1 min 27 sec
Crackdown on drugs	59 sec	42 sec
Drug registration procedure for ASEAN	48 sec	1 min 26 sec
PM Goh back from Indonesia	1 min 43 sec	1 min 42 sec
Next VP in Indonesia	35 sec	44 sec
South Korea	54 sec	1 min 6 sec
Clinton's balanced budget	45 sec	1 min 19 sec
Chinese defense minister meets Japanese counterpart	50 sec	33 sec
Gulf crisis	1 min 21 sec	1 min 32 sec
Sri Lanka's Independence celebrations	37 sec	54 sec
Cable accident in Italy	57 sec	57 sec
Cebu Pacific air crash	44 sec	About 1 min
Crescent Girls' School model UN delegation to Russia	53 sec	1 min 16 sec

**Comparison of Order of Ideas**

Although the topics covered and the amount of time devoted to each in the English and Malay broadcasts in Singapore are quite similar, as seen above, they are not necessarily presented in the same order. A comparison of transcribed texts for parts of two news items, the decision of the Singapore Ministry of Education to commercialize textbook publication and the cable car tragedy in the Italian mountain resort, shows that the sequencing of ideas within the discourse is also rather different for the two languages. Samples of these transcribed texts are given here, the Malay one being translated (more or less literally) into English:

**(1) Commercialization of Textbook Publication (English)**

The Ministry of Education has announced a move to commercialise textbook publication. It wants the private sector to design and produce textbooks for all exam subjects so that schools will have a wide range of teaching materials to choose from.

**(2) Commercialization of Textbook Publication (Malay, translated)**

Schools will soon have a wider range of choices of teaching materials if the Ministry's plans for privatizing the publication of textbooks succeeds. The Ministry wants the private sector to publish textbooks for all examinable subjects.

**(3) Cable car tragedy (English)**

Now the tragedy in an Italian mountain resort: a US warplane clipped the overhead wires of a cablecar, killing twenty people. President Clinton has expressed deep regret over the mishap.

**(4) Cable car tragedy (Malay, translated)**

President Clinton apologized for the tragedy of an American military plane which ran into the cables of an Italian cable car while on a low level training mission.

In both of these stories (as well as in others that occurred in both the Malay and English news) most of the same information occurred in both stories but not necessarily in the same order. If film clips were shown, the same film clips were used in both languages. However, if those filmed spoke, the speakers were heard in the English version but in the Malay version the reporter simply continued to talk about the story (and did not translate what the speaker in the film clip said).

Conclusions concerning English influence on Malay news stories in Singapore are that the influence on the language is minimal. Writers of the Malay news stories select some of the same news items and use any film clips prepared for or acquired by the English program but they reorder items, delete some, add their own and reorder points within a story.

### Indonesian news broadcasts – *Dunia Dalam Berita* – 11 July 1998

Here I will list the topics covered and the amount of time devoted to each. This data cannot, of course, be compared directly with that from the other countries as the recording was made at a different time. However, some generalizations about the general structure of the program can be made. The items and time devoted to them were as follows:

- Golkar convention (3 – 5 min)
- World Bank loan to South Korea (1 min 30 sec)
- US loan to the Philippines (1 min 25 sec)
- Asian economic crisis – World Bank urges developed countries to help those most in need (1 min 20 sec)
- ITB is looking in to the development of cooking oil, margarine and possibly even motor fuel from a new vegetable source (45 sec)
- Elections in Japan (25 sec)
- Delegation to Iraq (45 sec)
- A new zoo is being built where animals including domesticated cattle can roam freely (55 sec)
- Sports (World Cup) (5 min)
- Weather (1 min)

### Main observations on Indonesian broadcast

The Indonesian news broadcasts differs from those of the other countries investigated in that the total length of the program is only about 20 (rather than 30) minutes. In terms of the organization of ideas, the Indonesian program was more like Singapore than like Malaysia and Brunei in that ‘top stories’ could be either local or international.

### Indonesian and Malay Internet News Stories – a comparison of the use of active vs. passive verb forms

One important feature of Indonesian texts is the frequent use of passive verb forms. (See Poedjosoedarmo 1986 for an explanation of why this pattern is so common.) If Malaysian Malay, partially under the influence of English and/or Chinese, uses passives less frequently than Indonesian does, this difference in discourse strategy might be one factor in occasional communication breakdown between speakers of the two varieties.

In the present study, two news items were reported in both *Kompas* (Indonesian) and *Berita Harian* (Malaysian) on the 5 and 7 July. These items were Croatia winning in soccer (World Cup) over Germany and President Habibie asking Indonesians to fast on Mondays and Thursdays in order to save rice. For the first of these stories, the Malaysia item was a translation from Reuters while the Indonesian one was the story of an Indonesian reporter who had been present during the game. The statistics are as follows:

*Kompas*:

17 active transitive verbs

9 passive verbs

*Berita Harian*

- 31 active transitive verbs
- 5 passive verbs

These figures may suggest that English is indeed influencing verb form choices in Malaysian journalistic writing more than it is in Indonesian. However, in the second news story, where no translation from English was involved in either country, the figures were as follows:

*Kompas*:

- 9 active transitive verbs
- 4 passive verb

*Berita Harian*:

- 18 active transitive verbs
- 7 passive verb

Though the Malaysian article was longer, the proportions of active to passive verb forms in the two articles was similar. The conclusion then is that the disproportionate number of active forms in the Malaysian soccer story is probably due to the fact that it was a direct translation. In stories composed by local reporters there seems not to be much difference between the two varieties in this respect. The major linguistic difference between news stories from the two countries does, in fact, appear to be lexical. This then is probably the source of problems in cross-national comprehension of formal texts.

### Overall Conclusions

The structure of news broadcasts, a borrowed genre, seems to have been adapted similarly in all the Malay-speaking countries. The influence of English in either live news broadcasts or news stories found on the internet is only apparent where direct translation has taken place. No obvious or consistent structural differences were found in comparing news articles from Indonesia and Malaysia, suggesting that reported difficulties in comprehension are probably due to lexical differences between the two varieties.

Language variation occurs where change has taken place and change may be motivated by either external influence or by an imbalance in internal features. In the case of the structure of Malay news broadcasts, English is the obvious external influence, in fact, the source of the genre. However, regional influences can also be a factor in change resulting in variation. An earlier study of pronunciation features in TV news conducted in Brunei (Poedjosoedarmo 1996), found that while Bruneian newscasters were attempting (not with total success) to follow the pronunciation conventions they had observed on Malaysian TV (such as using schwa for a final spelt a), Malaysian newscasters at the same time, influenced by the internal decision to follow 'bahasa baku' (standard language) 'rules', attempted (also not with total success) to pronounce final a as /a/.

According to accommodation theory (Trudgill 1986), individuals tend to modify speech features to be like those they feel positively towards and different from those they feel negatively towards. In the same way, a particular style, a particular genre in a given location may either verge towards or away from a similar style and genre in a neighbouring

location, depending on the current cultural and political climate. Bruneians look to Malaysia in matters of language but to Singapore in matters of economic management. Relations between Singapore and Malaysia during the current year have been tense and yet many Singaporean Malays feel their closest affinity to be with Malaysia. Indonesia is culturally and politically very different from the other Malay-speaking countries and very diverse internally. All of these factors no doubt contribute to tendencies to be like or different from neighbouring countries in terms of language use patterns. A more detailed exploration of the relationship between these features and variation in the structure of discourse, however, will have to await further study.

## References

Baker, M. 1992. *In Other Words: A Coursebook on Translation*. London & New York: Routledge.

Gass, S. and L. Selinker. 1983. Language Transfer: A Conference Report. In F. Eppert (ed.), *Transfer and Translation in Language Learning and Teaching*. Singapore: RELC.

Poedjosoedarmo, G. 1986. Subject Selection and Subject Shifting in Indonesian. In Soenjono Dardjowidjojo (ed.), NUSA: *Miscellaneous Studies of Indonesian and Other Languages in Indonesia*. Part 8: 1-18. Jakarta: Badan Penyelenggara Seri NUSA.

\_\_\_\_\_. 1996. A Preliminary Look at Phonetic Variation in the Vowels of Brunei Malay. In J. Collins (ed.), *Language and Oral Traditions in Borneo*. Williamsburg, VA: Borneo Research Council.

Trudgill, P. 1986. *Dialects in Contact*. Oxford: Basil Blackwell.

# WH-INTERROGATIVES IN THE MALAY CLASSICAL TEXT OF SULALATUS SALATIN

Rogayah A. Razak

*Faculty of Allied Health Sciences*

*Universiti Kebangsaan Malaysia*

*<rogayah@medic.ukm.my>*

## 1 Introduction

This study is part of an on-going research project which looks at the interrogatives used in Malay classical texts. The Malay classical texts are written documents produced before the 1900s. The classic Malay genre reflects a state of the Malay language through time. The data of the classic Malay genre consists of *hikayats*, official announcements of Malay rulers, official letters of sultans to colonialist administrators or official letters to foreign rulers. Through the examination of these documents and the style of language used, we can describe the patterns of constructions used and trace the development of the forms used in these constructions.

*Sulalatus Salatin* or *Sejarah Melayu* used in this study is the version edited by A. Samad Ahmad (1996) which is nearer to the original manuscript compared to the versions edited by Shellabear (1909) or Winstedt (1938). *Sulalatus Salatin* (SS hereafter) is a text which was written or compiled by Tun Seri Lanang during the times of Sultan Abdullah Ma'ayah Shah, a ruler in the southern state of Johor from approximately 1613 to 1623. The purpose of writing this collection of stories of the Malay sultans and Malay Sultanate was to inform the future generation of their heritage, culture, and roots of their lineage.

'Bahawa hamba minta diperbuatkan hikayat pada Bendahara, peraturan segala raja-raja Melayu dengan istiadatnya supaya didengar oleh anak cucu kita yang kemudian dari kita dan diketahuinya segala perkataan, syahadan beroleh faedah-lah mereka itu daripadanya'

SS contains interrogatives in the form of *wh*-questions, yes-no questions, alternative questions, and a small sampling of other forms of interrogatives.

### wh-question

(1) Mengapa maka anakku hendak pergi bermain jauh? (p36)  
Why that child-my wan go play far  
'Why would my child want to go play far away?'

### **yes/no question**

(2) Adalah kita memanggil paman ini, kita hendak beristeri,  
 is-EMP we all uncle Dem, we want marry  
 'We summon uncle here, (because) we want to marry,  
 carikan kita, adakah paman? (p36).  
 find us is-QPRT uncle  
 find a wife for us, are there any, uncle?'

### **Alternative question**

(3) Akukah Pak Si Bendul engkaukah Pak Si Bendul? (p21)  
 I-QPRT the clown you-QPRT the clown  
 'Am I the clown, (or) are you the clown?'

Our focus in this discussion is on one type of interrogatives, i.e. the *wh*-questions. We will look at the various issues regarding *wh*-movement in *wh*-questions with the purpose of trying to account for the characteristics of *wh*-movement in classical Malay of *Sejarah Melayu*.

We will be employing assumptions of Government and Binding as in Chomsky (1981, 1986) in describing our data. We could not look at degrees of ungrammaticality of sentences in the data as our only source of judgment is the corpus in the book. Therefore, we will be restricting our discussion to account for the properties of *wh*-movement in the variety of Malay used in the classic text.

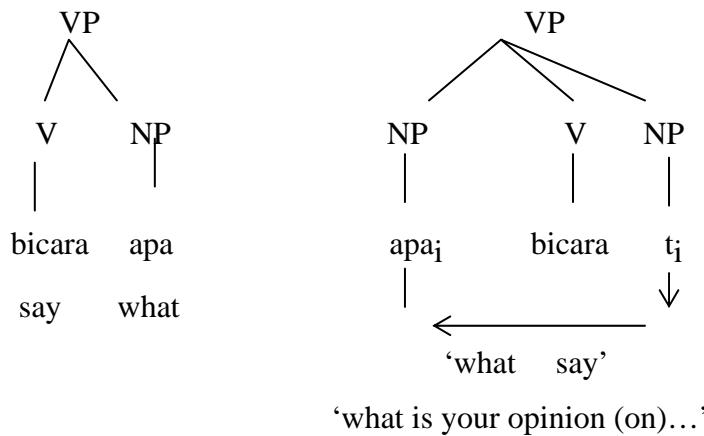
## **2 Properties of *Wh*-Interrogatives in SS**

Basically, the *wh*-movements in SS do not differ much from *wh*- movements in modern Malay. Nevertheless, there are characteristics which give *wh*-movement in classical Malay as in SS a distinctive and idiosyncratic flavor.

### **2.1 Absence/Presence of Internal Fronting in [+wh] constituents**

#### **(a) Verb phrases (VP [+wh])**

Verb phrases which have the feature [+wh] tend to undergo an internal fronting of the word containing the [+wh] feature in the VP constituent prior to movement to [Spec,CP] position.

**Figure 1**

now what say all people because we want to there  
 ‘Now, what does everyone think, as we want to go there,

terlalu sekali kita berkenan memandang dia.] (p38)  
 very much we like look her  
 (we are) very much infatuated looking (at) her’.

Other examples of this type of internal-VP movement are as in (5a) to (5f).

(5) a. apa mahu?  
 what want  
 ‘What (do you) want?’

b. apa rasa?  
 what think  
 ‘What(do you) think?’

c. apa kerja?  
 what work  
 ‘What (are you) doing here?’

d. apa khabar?  
 what news  
 ‘How are (you) doing?’

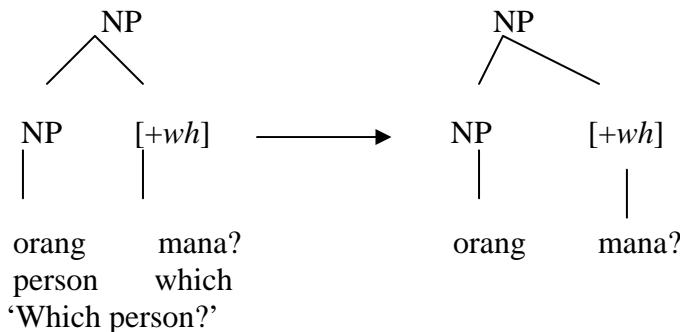
e. apa sebab?  
 what reason  
 ‘What is the reason?’

f. apa jadinya?  
 what become  
 ‘What happened?’

Following the internal movement in the VP constituent, the whole VP<sub>[+wh]</sub> undergoes *wh*-movement to [Spec,CP] position.

(b) Noun Phrases (NP [+wh])

There is no internal movement of the *wh*-word within the NP; instead the *wh*-word remains at the original position in the NP constituent even after *wh*-movement in the VP.



**Figure 2**

(6) [Orang mana [kamu ini], dan [hendak ke mana [kamu]]?]  
 people where you DEM and want to where you  
 'What is your origin and where are you heading to?'

Other examples of NP<sub>[+wh]</sub> are in (7a) to (7f).

(7)	a. bahasa mana? language which 'Which language?'	b. destar mana? head-gear which 'Which head-gear?'	c. daging apa? meat what 'What meat?'
	d. gendang apa? drum what 'Which drum?'	e. menteri apatah? minister what-EMP 'What kind of minister?'	f. orang siapa? person who 'Whose people?'

(c) Clause (IP [+wh])

IP<sub>[+wh]</sub> in SS has examples of internal fronting of the [+wh] word in the clause, as in (8a) to (8c), (9), and Figure 3.

(8)	a. [IP apa janji paman?] what promise uncle 'What is uncle's promise?'	b. [IP apa pekerjaan lain?] what work other 'What other job?'
	c. [IP apa namanya?] what name-it 'What is its name?'	

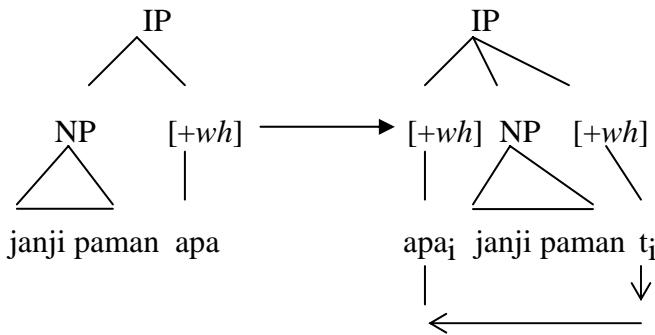


Figure 3

(9) [Apa janji paman [hendak diperjanjikan dengan hamba itu?  
 what promise uncle want promise-PASS with servant ]])(P30)  
 'What is uncle's promise to be promised to me?'

We discovered that some constituents with [+wh] features display internal fronting in the constituents prior to *wh*-movement. We find that NPs [+wh] at [VP V NP [+wh]] and NPs at predicative positions particularly of the [IP NP NP [+wh]] structure allows internal movement. On the other hand, NPs at modifier positions in nominal constructions [NP NP NP [+wh]] do not allow movement.

This apparently corresponds to the fact the NPs [+wh] in the former group are NPs in governed positions, whilst the NPs [+wh] in the latter group are NPs in an ungoverned positions. We propose in (10) a condition on extraction for the internal fronting.

(10) Condition on Extraction for Internal fronting  
 (i) NPs [+wh] in governed positions permit internal fronting.  
 (ii) NPs [+wh] in ungoverned positions does not permit internal fronting.

This condition explains the nature of *wh*-constituents in interrogatives in SS. NPs [+wh] in governed positions are free to move after being assigned theta-role by its governor whereas NPs [+wh] in ungoverned positions are modifiers which are crucially needed by its noun head to complete its structure and meaning.

## 2.2 Categories of Wh-Movement

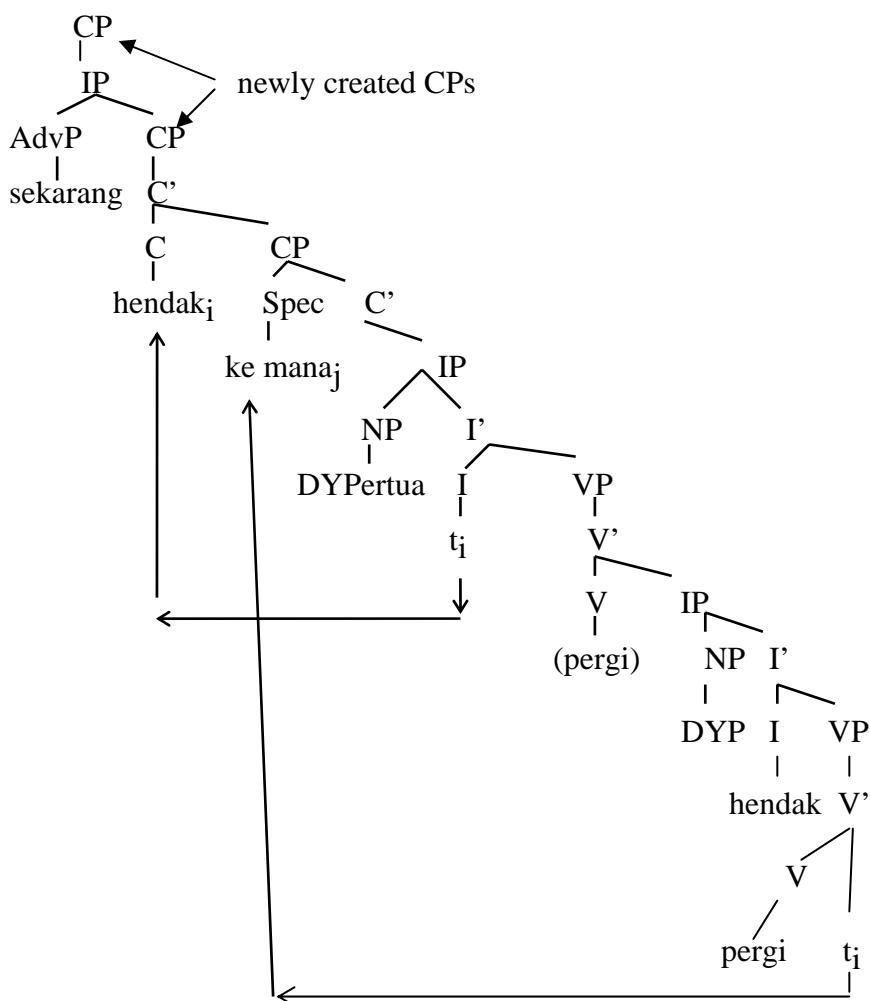
Wh-movement in SS can be categorized as either head movement or maximal projection movement. The majority of the data that we have on *wh*-interrogatives consist of examples of maximal projection movements. Head movements in *wh*-interrogatives are only a small sample of the data.

Maximal projection movement is movement of the NP, VP, PrepP, AdvP, or InflP(clause). The maximal projection movement will move to a position which allows an XP to be adjoined. This position is the [Spec, CP] position. Head movement is movement of X<sup>0</sup> such as the noun apa 'what', siapa 'who', berapa 'how much', verbs such as kabul 'agree', or auxiliaries such as hendak 'want' to a position which allows an X<sup>0</sup> to fill the adjoined position. This position is the [Comp, CP].

In order to test these generalizations, we have to look at interrogative sentences which have both maximal projections and head movements.

(11) Sekarang      hendak ke      mana      Duli Yang Dipertuan      hendak pergi? (p284)  
 now            want to        where    Your Highness            want    go  
 'Now, where does Your Highness want to go?'

(12) DS: [ADVP Sekarang [IP Duli Yang Dipertuan hendak pergi  
 now            YHighness            want    go  
 [IP DYP hendak ke mana ]]  
 YH    want    to where



**Figure 4**

From the surface structure in (11), we can conclude that the interrogative sentence involves two movements: the movement of hendak 'want' and the *wh*-movement of ke mana. Accordingly, hendak 'want', being an  $X^0$ , has to move to a position which allows an  $X^0$ , whilst PrepP ke mana 'where to' has to move to a position which allows an  $XP$ . We

also have to cater for the AdvP *sekarang* 'now', located sentence-initially at surface structure. These requirements mean that we have to create two new landing sites: one for the  $X^0$  kehendak and the other for AdvP sekarang. The PrepP ke mana 'where to' moves to the [Spec,CP] of the matrix clause. The new landing sites are adjunctions to the CP of the matrix clause. Figure 4 gives us the surface structure of the interrogative sentence.

The auxiliary hendak 'want' which is from I, the INFL head moves to Comp of CP. Meanwhile the PrepP ke mana 'to where' which is a maximal projection goes to [Spec,CP].

Another sentence which demonstrates these two kinds of movements is (13).

(13) Lagi pula pada Duli Yang Dipertuan belum kita menjunjung  
more over to Royal Highness not we salute  
'Moreover, before we salute His Royal Highness;  
  
duli; bagaimana akan kita berjabat tangan dahulu (p254)  
respect how will we shake hand before  
how are we to shake hands first?'

(14) SS: [AdvP Lagi pula pada Duli Yang Dipertuan belum kita menjunjung duli;  
CP bagaimana<sub>i</sub> [C akan<sub>j</sub> [IP kita<sub>j</sub> berjabat tangan dulu t<sub>i</sub> ]]]]

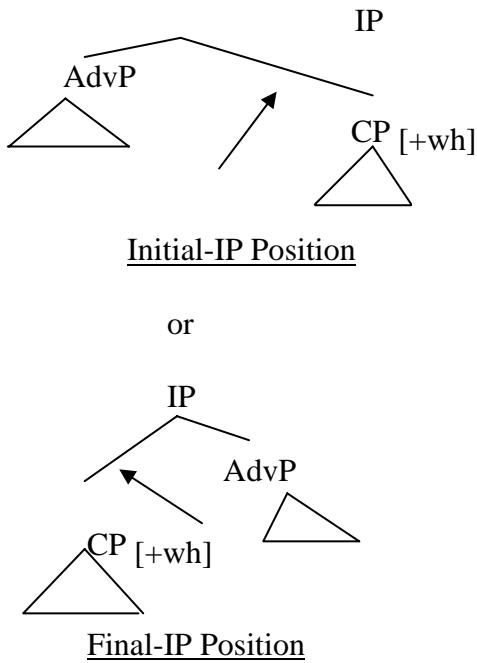
In (14), bagaimana 'how', an AdvP, moves to [Spec,CP], whilst the auxiliary akan 'will', which is the head of INFLP, moves to [Comp,CP].

### 2.3 The Landing Sites for Wh-Movements

Interrogative sentences in SS demonstrate interesting possibilities for landing sites of *wh*-movements. There is the position [Spec,CP] of the matrix clause, as in (15) and (15').

(15) Gunung mana yang kelihatan dua itu?  
Mountain which that appears two that  
'Which mountain appears to be two (mountains)?'  
  
(15') SS: [CP gunung mana<sub>i</sub> [IP yang kelihatan dua itu t<sub>i</sub> ]]

Another possible landing site is a position between the ADVP and the CP matrix. This site is a common site for *wh*-movement in the data. The AdvP is usually some form of salutation, indicating time, condition, an introductory comment, etc which precedes the matrix CP. The structure is as produced in Diagram 5.

**Figure 5**(16) Examples of initial IP position

(a) [AdvP Jikalau tuan hamba daripada anak cucu Iskandar  
 if sir slave from child grand Iskandar  
 'If you are from the generation of Alexander'...  
 (conditional) [CP ...]] (p21)

(b) [AdvP Sekarang [CP ...]] (p38)  
 'Now, ...' (time)

(c) [AdvP Hei, Maharajah Dewa Sura [CP ...]] (p100)  
 'Hei, King Dewa Sura ... (salutation)

(d) [AdvP Pada perasaan patik sekelian, [CP ...]] (p275)  
 at feeling subject all  
 'We all feel ,... (introduction)

(17) Examples of final IP position

(e) [CP ... [AdvP jikalau apa sekalipun tiada kita tahani]] (p194)  
 if what once-EMP no we stop  
 '...., whatever it is we will not stop you'.

(f) [CP ... [AdvP jikalau demikian sia-sialah yang dibeli  
 if like that waste-EMP that buy-PASS  
 ... 'if that's the case, that which was bought has gone to waste

oleh bapaku sekati emas di benua keling itu].  
 By father-my one kati gold at continent Indian the  
 by my father the one kati of gold at the Indian continent'.

These structures are characteristic of clausal structures in the classical Malay text. These types of structures produce sentences which are unusually complex and long-winded due to the positioning of these AdvPs.

(18) [Jikalau kita dikehendaki Allah Taala antara siang dan malam,  
 if we want-PASS God between day and night  
 'If we are called to Allah (anytime) between night and day,  
 [siapa [Bendahara dan orang-orang sekalian sembah [akan ganti kita  
 who Bendahara and people all prostrate will replace us  
 who (will) Bendahara and the other subjects prostrate to replace us,  
 [supaya Pahang ini jangan binasa?]]]]]  
 so Pahang this not destroy  
 so that Pahang will not be destroyed?'

In some cases, the landing sites of the *wh*-movement could be the [Spec,CP] of the respective embedded clauses. This is especially so in the cases of complex NPs in subject or object positions and interrogative sentences which have embedded sentences conjoined by connectives.

(19) [CP Apa kehendaknya[IP menyuruh Orang Kaya Kedua ini  
 what wish-his ask People rich second DEM  
 'What (are) his wishes [asking both of Orang Kaya  
 (dan)[CP apa hendak [IP dicari?]]]](p98)  
 and what want find-PASS  
 (and) what desires (does he)sought?'

In (20), a conjoined sentence with two embedded sentences, the *wh*-word goes to the respective immediate CPs.

(20) [CP 'Anak siapa [IP engkau ini ], [CP siapa [IP namamu ], dan  
 child who you DEM who name-you and  
 'Whose child are you, what is your name, and  
 [CP apa sebabnya [IP engkau hanyut dengan sekeping papan perahu ini]]]? (p112)  
 what reason-it you afloat with a-piece wood boat DEM  
 why were you swept away with only this piece of perahu wood?'

In (21), the sentence is a conjoined sentence with three embedded IPs. The *wh*-words move to the respective embedded CPs.

(21) ['Apakah [dosa hamba ke bawah duli YDP] (adalah) kerana  
 what-QPRT sin servant to under Royal Highness (is) because  
 'What are my sins towards Your Royal Highness,

salah hamba yang sedikit inikah maka hamba hendak dibuangkan?'] (p201)  
 wrong servant that little DEM-QPRT that servant want banish-PASS  
 is it for my small misdeed that I am going to be banished?"

We conclude that possible landing sites of *wh*-movement could be [Spec,CP] of either matrix or embedded clauses or an A' position between the AdvP and the CP matrix clause.

#### 2.4 Types of Movements – Long or Short Distance Movements

The types of *wh*-movement in the data could be categorized as either long-distance or short-distance *wh*-movement. Short distance movements are when the XPs [+wh] are moved to the [Spec,CP] of the matrix clause. On the other hand, long distance movements are when the XPs [+wh] are moved and crossed over more than one embedded IP.

(22) '[IP Hendak [CP ke mana [IP Laksamana baharu-baharuan  
 want to where Laksamana just  
 datang [IP mengadap kita]]]' (p128)  
 come (have) audience us  
 'Where are you going? Didn't you just arrive to have an audience with us?'

(23') DS: [IP Laksamana baharuan-baharuan datang [IP Laksamana mengadap kita  
 [IP Laksamana hendak ke mana?]]]

(23') SS: [IP Hendak<sub>i</sub> [CP ke mana<sub>j</sub> [IP<sub>1</sub> Laksamana baharu-baharuan datang  
 [IP<sub>2</sub> pro mengadap kita [IP<sub>3</sub> pro t<sub>i</sub> t<sub>j</sub> ]]]]]

The PrepP ke mana 'where' originated in the IP<sub>3</sub> and is moved to the [Spec,CP] of the matrix clause. On its way to its destination, the PrepP [+wh] crosses over 3 IPs. Although this distance would normally result in an ungrammatical sentence, sentence (23) is grammatical due to the fact that PrepP ke mana is an adjunct and its trace in the extraction site need not be governed.

We did not come across any long distance movement of argument XPs [+wh]. All instances of argument XPs are short distance movements like the examples in (24) to (26).

(24) [Siapa<sub>i</sub> [ t<sub>i</sub> bersetia dengan engkau orang derhaka?]]  
 who befriends with you people traitors  
 'Who befriends traitors like you?'

(25) [Apa<sub>i</sub> [Kita buat t<sub>i</sub> [duduk saja ini?]]]  
 what we do sit only this  
 'What are we doing sitting around like this?'

(26) [Orang mana<sub>i</sub> [kamu ini t<sub>i</sub>] dan [hendak<sub>j</sub> [pro t<sub>j</sub> (pergi) ke mana, kamu?]]]  
 people where you DEM and want (go) to where you  
 'What people are you, and where do you want to go?'

We find that adjuncts XPs[+wh] move relatively further than argument XPs[+wh]. An adjunct need not be governed and can be as far from its extraction site compared to an argument which needs its governor nearer for government.

## 2.5 The Cyclic Application of Transformations

Complex sentences in SS follow the cyclic application of transformational rules. The cycle begins from the most deeply embedded IPs progressing upwards and leftwards to the matrix clause.

(27) Apa mau Raja hendak bertemu dengan kita? (p277)  
 what want King want meet with us  
 'What does the king wants to want to meet with us?'

(27') DS: [IP1 Raja hendak bertemu dengan kita [IP2 raja mahu apa?]]

The first cycle is the most embedded IP – which is fronting of apa in the VP clause.

Cycle 1: (i) [IP2 raja [VP apa? mahu t<sub>j</sub>?]]

Then the VP clause is moved to the [Spec,CP] of the matrix clause.

Cycle 2: (ii) [CP Apa mau<sub>i</sub> [IP raja hendak bertemu dengan kita [IP2 pro t<sub>i</sub>]]]

Another example of a cyclic application of transformation is in (28).

(28) Jikalau ia hamil sekalipun, apatah orang kaya takut menaruh  
 If she pregnant anyhow, what-EMP people rich afraid keep  
 dia kerana orang kaya, orang tua kepadaanya?  
 her because people rich people old to-her  
 'Even if she is pregnant, why are you frightened to have her (in your house) because you are older than her?'

(28') SS: [AdvP Jikalau ia hamil sekalipun, [CP apatah<sub>i</sub> [IP orang kaya takut t<sub>i</sub> [IP pro menaruh dia kerana orang kaya orang tua kepadaanya?]]]]

Cycle 1: No transformation.

Cycle 2: *wh*-movement of apatah to [Spec,CP] of matrix clause.

We have shown the cyclic nature of transformation rules in SS. This accounts for the systematic ordering of constituents in the sentences regardless of their complexity.

### 2.6 Examples of Wh-Interrogatives In-Situ

Even though most *wh*-word or phrases move overtly to [Spec,CP], there exists a small number of *wh*-words or phrases which remain *in-situ* at base position in interrogative sentences. Consider (29) to (33).

- (29) Siapa mengarang surat ini?  
who write letter DEM  
'Who composed this letter?'
- (30) Raja yang mana?  
king that which  
'Which king?'
- (31) Si Putih dengan Si Khatijah mana dia?  
Si Putih and Si Khatijah where her  
'Where is Si Putih and Si Khatijah?'
- (32) Manda ini orang siapa?  
Uncle DEM people who  
'Who are you (Uncle) associated with?'
- (33) Kamu (pergi) ke mana?  
You go to where  
'Where did you go to?'

### 3 Wh-Interrogatives in SS: A Summarization

One peculiarity of *wh*-interrogatives in SS is the internal fronting of the *wh*-word which has the [+wh] feature in the VP constituent prior to movement to [Spec,CP]. We find that clauses with [+wh] feature also demonstrate this kind of internal fronting prior to *wh*-movement. Only NPs [+wh] do not have such internal fronting. We postulated the condition on extraction for internal fronting to account for this phenomenon. This initial fronting preceding *wh*-movement could be due to Indian stylistic influence of expressions considering that Hindu rituals and the Indian language has a great influence on the Malay language and culture of the Malay natives.

The other interesting feature was the possible landing sites for *wh*-movements in SS. We found that the landing sites could be [Spec,CP] of either matrix or embedded clauses or an A' position between the AdvP and the CP matrix clause. The possibility of landing sites in SS certainly departs radically from modern Malay interrogatives. It is certainly these intricate and complex structures which made classical Malay a more genteel variety of the Malay language. It allows room for creativity and thus accounts for the complexity of sentences produced. This appropriately describes this variety of Malay language used in SS, which is termed bahasa istana 'a language used within the inner circle of the palace'. This variety of Malay is the top-most ranked variety of language used in the feudalistic society depicted in SS. It represents Malay at its best, with beautiful phrasing of utterances yet very subtle in its meanings.

In general, we found that *wh*-movement in SS has characteristics which modern Malay *wh*-interrogatives have. They demonstrate short distance and long distance types of movements and the application of transformations are cyclical in nature. Adjuncts also have relatively longer movements than arguments. The *wh*-movements possess head and maximal projection movements, and it also has *wh*-interrogatives which are *in-situ*, albeit not many.

We hope further research will bring out further characteristics of interrogatives of classical Malay. An analysis of the whole gamut of interrogatives used in classical Malay would be useful in trying to provide an explicit account of *wh*-interrogatives in Malay.

### References

Ahmad, A. Samad. 1996. *Sulalatus Salatin*. Sejarah Melayu. Kuala Lumpur: DBP.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Barriers*. Cambridge, Mass.: MIT Press.

Razak, Rogayah. 1995. *The Syntax and Semantics of Quantification in Malay – A Government and Binding Approach*. Unpublished Ph.D. Dissertation. Universiti Sains Malaysia.



# MALAY LEXICALIZED ITEMS IN PENANG PERANAKAN HOKKIEN<sup>1</sup>

Lim Beng Soon

*Regional Language Centre (RELC) Singapore*  
<Beng\_Soon\_Lim@relic.org.sg>

Teoh Boon Seong

*Singapore Management University*  
<bsteoh@smu.edu.sg>

## 1 Introduction

This paper aims to document instances of lexical items in Penang Peranakan Hokkien (henceforth PPH) which are Malay loanwords. There has been some research into Peranakan (Baba) Malay, namely that of Shellabear (1913), Vaughan (1879) and more recently Clammer (1980) and Khoo (1996). However, not surprisingly, most academic work so far on Peranakan Chinese (viz. Baba Malay) has been restricted to that of the Malacca or Singapore varieties because of their similarities to Malay superficially.

PPH on the other hand is, for all intents and purposes, a Chinese dialect. Nevertheless, this dialect in question contains numerous Malay lexical items even though its syntax is underlyingly Hokkien. A brief description offered by Khoo (1996) in differentiating PPH from its other cousins is illuminating.

The Baba Hokkien in Penang is closer to this Tang Min (a type of Min dialect) than is the type of Hokkien spoken by the Sinkheks (recently arrived immigrants), who use the Standard Amoy variety. PPH is even less closely related to Standard Amoy because of the Malay words it has assimilated. It is different from other variations of Hokkien spoken on the Mainland or in Malacca and Singapore. (Khoo 1996:112)

The Malay words assimilated in PPH appear in various linguistic levels, including morphemes, words, phrases and clauses. Evidence of the assimilation of Malay in those linguistic levels is copious, for example, the prolific use of the Malay particles 'lah' and 'pun'. The assimilation in the level of morphemes is, however, less evident but nevertheless present, typically in the prolific use of 'nya' in PPH, which is a contracted form of, and semantically similar to, the original Malay term 'hanya', meaning 'only'.

## 2 Historical Background

This paper does not purport to provide a definitive or authoritative historical overview of the Peranakan Community in Penang; however, we believe that a short introduction to the origins of the Peranakan community in Penang might be helpful. The genesis of the Peranakan community in Penang can be traced back to the establishment of Penang as a

<sup>1</sup> This article is a revised version of a paper presented at the 8th Annual Meeting of the South East Asian Linguistics Society, 20-22 July 1998, Kuala Lumpur, Malaysia. We would like to thank the State Chinese Association of Penang and the State Museum Board of Penang for their financial support and encouragement that has made the writing of this paper possible.

Mark Alves, Paul Sidwell & David Gil, eds. *SEALSVIII: papers from the 8th meeting of the Southeast Asian Linguistics Society (1998)*. Canberra, Pacific Linguistics, 2007, pp.149-165.

© Lim Beng Soong & Teoh Boon Seong

British colony in 1786. In the early years after it was made a British outpost, trade flourished and there was much commercial activity and regional trade with the North Malaysian states of Kedah, Perak, the Sumatran Kingdoms and Southern Thailand.

Penang as a focal point of commerce in this part of the Malay Archipelago attracted an influx of Chinese businessmen to the colony, particularly from Southern Thailand and North Sumatra. The Peranakan Chinese in Penang were thought to have been "descendants of Chinese males marrying or cohabiting with Malays or Siamese or Burmese; the last mentioned, however, appear to be rather rare" (Khoo 1998:6). It must be stressed that the Peranakans of Penang still have extensive familial ties with both these regions.

PPH is spoken to greater or lesser degree by many of Penang's Hokkien community, but the prevalence of Malay lexicalized items in PPH is found among older members of the erstwhile Straits Chinese community. Typically, this group of speakers is above the age of 50 and many are English-educated. Although their speech is liberally peppered with lexical items of Malay origin, this does not predispose them to understanding standard Malay, which has gained currency among the younger generation. In fact, many PPH speakers above the age of 50 have never had formal instruction in standard Malay; they are hence unable to understand formal Malay even though bazaar Malay is easily understood by all. Most of the respondents in this study were initially unaware that they had been using words of Malay origin in their everyday use of PPH.

### 3 Procedure

This study was conducted through observation of naturally occurring speech in a PPH family for a period of two months, from May to June 1998. Some samples have been transcribed for ease of presentation. We also collected some taped conversations of spontaneous speech in the said Peranakan family.

The focus of this paper is to analyze the features of Malay lexical items in PPH to see if they are loanwords; exhibit phonological diversity; are different from Malay lexically and semantically; and finally, whether there are loan translations above the level of individual lexicon. We are therefore of the opinion that it is unnecessary to describe longer stretches of speech.

### 4 Framework of Analysis and Discussion of Findings

As an initial study into PPH, we will discuss the individual occurrences of Malay words in PPH. This paper also hopes to shed light on the patterns of code-switching that occur in PPH which might in turn stimulate further investigation into why certain linguistic categories are code-switched more often than the others.

The current paucity of studies undertaken on PPH is regrettable; nevertheless, studies done on other Peranakan Chinese communities in the region provide an outline for relevant areas of investigation. Among the earliest studies into Baba Malay is that of Shellabear (1913): he provides a description of Baba Malay in terms of its distinctive aspects as compared to "...the spoken language of the pure Malays" (Shellabear 1913:50). These aspects include a description of the evolution of Baba Malay; Baba Malay words of Chinese origin; Malay words which are unknown to the Babas; Malay words which are mispronounced by the Babas; and finally, the claim that the Baba idiom is Chinese rather than Malay.

In terms of language use and identity, we find that this issue is particularly relevant to the Peranakan Chinese. This is due to their ability to express their solidarity with both

the indigenous community and the Chinese by their emphasis of either the Malay or Chinese elements of their language. This was not only prevalent among the Peranakan Communities in Peninsular Malaysia but also amongst the Peranakan Chinese in Java and Sumatra. Tan (1979, 1980) did extensive studies into the Babas of Malacca and remarked that the relation between the Baba Malay language and identity was significant. The Babas could manipulate their use of Baba Malay to associate themselves with the Malays, or conversely emphasize the non-Malay elements of Baba Malay to stress their solidarity with the Chinese (Tan, 1979:125-127). Writing on the Peranakan Chinese in Kelantan, Teo (1996) extends the scope of borrowings from the Malays not only in the domain of language but also other cultural borrowings (e.g. attire) from the Malays by the Peranakan in Kelantan. Teo (1996) suggests that for the urban Peranakan, cultural borrowings of attire and cuisine remain back stage, where their manifestation is restricted to the confines of the household. We concur with this, as the Penang Peranakan who are wholly urbanized have similar tendencies. In the case of the rural Kelantan Peranakan Chinese, cultural borrowings are not only back stage but front stage as well. This is entirely plausible, though an equivalent example in Penang is unobtainable.

In terms of a framework of analysis, Tan (1979) has formulated a framework for the comparison of Baba Malay and Standard Malay. He maintains that there are five main patterns of diversity, namely the use of loanwords (including loan-translations), phonological differences, lexical differences, semantic differences and syntactic differences (Tan, 1979:267).

Among the studies elaborated above, we will discuss the observations of Shellabear (1913) and Tan (1979). It must be emphasized that PPH is different from Baba Malay in that it is essentially a Hokkien dialect with a significant number of Malay lexicalized items in use. Some remarks of Shellabear (1913) merit attention. He mentions that the then Baba Malay of the Straits Settlements was different from colloquial Malay in terms of the following:

1. Words introduced which are of Chinese origin; which for PPH is the inverse: where words introduced in the Hokkien dialect are Malay.
2. Babas were unacquainted with a large number of Malay words which are in common use among the Malays. We find that this holds true for speakers of PPH.
3. Babas mispronounce or reconstruct phonologically many Malay words and in some cases have altered the pronunciation so much that the word is almost unrecognizable. This is particularly true of many of the Malay lexicalized items in PPH.
4. To a great extent the Peranakans use the Chinese idiom rather than the Malay in putting their sentences together. This results in the Babas producing Baba Malay at the sentential level which is quite different from the colloquial language of the Malays.

Although Shellabear's observations above are noteworthy, we believe that the framework of analysis provided by Tan (1979) is more productive for the focus of this paper. Hence we discuss how the four main patterns of diversity (i.e. the use of loanwords (including loan-translations), phonological differences, lexical differences and semantic differences (Tan, 1979:267)) between the Malay lexicalized items found in PPH contrast with their original functions and roles in standard Malay. This study, however, excludes the analysis of syntactic differences between Malay in PPH and standard Malay because PPH is basically a Chinese dialect with superimposed Malay lexical items. Nevertheless, this does not preclude the potential analysis of code-mixing and code-switching of Malay in PPH. In this paper, the operational framework of Tan (1979) is used with adaptations to analyze the functions and roles in PPH.

## 5 Loanwords

The Malay lexicalized items defined as loanwords in PPH are words that have been retained in their entirety in terms of semantic content, phonology (excepting the superimposition of tones to the words pronounced) and formal classes. Examples of these words generally include nouns like the names of certain flora and fauna (e.g. Bunga Melor, Katak Puru, Pacat, etc.). Some loanwords are restricted in their functions as grammatical categories. For example, 'tarik' in PPH is used to describe actions related to one afflicted with convulsions, which has the imagery of one's muscle pulling and jerking. Although this is also used by the Malays in a similar situation, it does not have the significance of the commonly used meaning by the Malays, which is 'pull', except in a limited sense of one inhaling cigarette smoke. The above peculiarity of loanwords not functioning in the whole spectrum of classes as they would in the source language is common among Creole languages. This tendency is not restricted to verbs and nouns alone but is also common among the other formal classes, such as adverbs and adjectives.

Sometimes loanwords borrowed from Malay function in the whole spectrum of classes as they would in Malay. For example, the root word 'buah', which as a noun means 'fruit', with proper affixation it also functions as an adjective. This is the same in PPH, although the use of the proper affixes is dispensed with. Another word which undergoes the same class shifts without any use of affixation is 'kerabu', which as a noun means 'salad' and as a verb means 'to make something into salad'.

Tan (1979) included another category of loanwords which he labeled loan-translations. These are direct translations into Baba Malay from Chinese: for example, 'hari satu' which corresponds to the Chinese 'Day One' instead of the Malay equivalent of 'Isnin' (Tan, 1979:275). For this study, loan-translations would be Malay concepts directly translated into PPH. The speech samples collected for this study do support the presence of loan-translation from Malay into Hokkien. However, we are unable to produce a comprehensive elaboration of these manifestations, primarily because this study is not aimed solely at such an objective. Secondly and more importantly, it is only through a comprehensive comparative study of PPH with the other Min dialects that one could ascertain if loan translation from Malay has taken place.

Possible candidates for loan translations would be culturally loaded terms describing actions that are common among the Malays that have been borrowed by the Peranakans in Penang. For example, the expression 'I'll stuff your mouth with chilies for uttering obscenities', which when used in PPH has the verb 'cabai' which means 'to put or stuff chilies into...', a common form of punishment for Malay children due to the same

form of indiscipline. We have found that there is use of Malay above the level of the lexicon which, we believe, represents some form of loan translation. We will provide a cursory discussion of this manifestation below, which we describe as the use of Malay above the level of individual lexicon.

## 6 Phonological Diversity

In this paper, we define phonological diversity as the different phonological manifestation of Malay words in PPH. Generally this phonological diversity involves the use of tones in PPH, which is a common characteristic of the Chinese Min dialect. Thus for example, the Malay borrowings 'kawan' meaning friend and 'kahwin' 'to marry' are rendered respectively as [kawan] and [kawen]. Here we witness the superimposition of a hi-lo or falling tone onto the disyllabic loanwords of Malay. Due to the influence of the Min dialect, which usually uses monosyllabic words, we also find that Malay, being an agglutinative language, often has words that have more than three syllables. These words when borrowed in PPH are often truncated or clipped for ease of pronunciation. A case in point is the pronunciation of the word 'barangkali', a four syllable word, as [bangkali] a three syllable word; compound words like 'buah guli' simplify to [bakguli]. The exception is the word 'hormat', a disyllabic word which is rendered as [horomat] in PPH. An explanation for this seeming counterexample would be that the contiguity of the alveolar trill and the bilabial nasal presents difficulty in pronouncing the word for PPH speakers and s/he reconstructs it as [horomat], with an epenthetic vowel insertion to break up the consonant cluster as it were.

It is understandable that PPH, being a type of creole, often displays syllable reduction: 'satu' > [sau] 'hari' > [ari], 'apa macam' > [amcam], 'bawa pergi' > [buat pi], and 'buat apa' > [buat pa] (Gwee 1993). The examples from PPH make our analysis even more exciting; not only do the borrowings conform to certain phonological constraints of Hokkien in terms of tone and syllabic simplification, but also because PPH is spoken in the northern region of Peninsula Malaysia, it conforms to the phonological characteristic of the north-west Malay dialect of Penang and Kedah. In general, there is the rule of high vowel lowering in final closed syllables in Malay, and indeed we have them as [sireh], [abo?], etc. In the Penang and Kedah Malay dialects, there are the phonological rules of  $l > \emptyset$  ( $l$  deletion) followed by diphthongization. Indeed, borrowings from Malay are then remolded as such in conformity to the phonological structuring; 'bisul' rendered as [bisoi], 'sanggul' as [sanggoi]. First, we have high vowels in final closed syllables lowered, and then the diphthongization is followed by the  $l$  deletion. We also note that not all high vowels in final closed syllables lower: for example 'tumis' > [tumis]. Indeed, this is what Pakir (1986) discovered in Peranakan Malay in Singapore. She went a step further in offering an explanation: namely, that the affrication of the final /s/ phoneme somewhat preempted the vowel lowering, which applies across the board in other instances. Words with diphthongs which monothongize are also evident in 'hairan', which is pronounced as [heran] in PPH. We notice also that words like 'binatang' are rendered as [manatang]. This is not surprising since the bilabial voiced stop is substituted with a nasal, but with the place of articulation is retained in the PPH form; Chinese speakers often have this negative transfer where Malay words like 'banyak' are rendered as [manya?], with the palatal nasal presenting a problem because it is non-existent in Chinese.

## 7 Lexical and Semantic Diversity

Lexical and semantic diversity between Malay lexical items in PPH and Standard Malay takes the form of the different meanings attached to the same lexical item in PPH and Malay. For example, the word 'hangat' in PPH is used to describe anger or express that one is angry, rather than the common Malay meaning that one is hot. The rather restricted meaning of words is common among Creole speakers, and in fact, most speakers of PPH do not know the other multifarious meanings of a Malay word. In other words, this relates to the broadening of semantic values in a lexical item. The lexical item 'senduk' refers to a ladle in Malay, and spoons are 'sudu'; however, for PPH speakers, 'senduk' refers to both spoon and ladle and it is other factors, such as the situation and shared knowledge of speakers, that disambiguate the referent of 'senduk' to the PPH speaker. In fact, this issue was brought up by Tan (1979) who states that the Baba Malay speaker used the adjectives 'big' or 'small' to indicate if s/he was referring to a ladle or a spoon (Tan, 1979: 274-282). Interestingly, here the PPH speaker employs similar devices by using the Hokkien adjectives which are the equivalent of 'big' and 'small' to refer to a ladle in the former (tua senduk) and a spoon (se senduk) in the latter. Another example of semantic broadening in the Malay lexicon adopted in PPH is the use of 'longkang' for drains, just as it is done in Malay. However, bigger drains are also referred to as longkang, with the Hokkien adjective 'big' modifying it. The appropriate Malay term would be 'parit'. However, this term does not exist in PPH.

We found that lexical and semantic differences between Malay and PPH can also be found in the semantic narrowing of certain borrowed Malay lexicon in PPH. We are aware that some of the lexical items cited in this analysis might have had wider semantic significance in the past and might not be regarded as instances of semantic narrowing if a diachronic study of PPH is undertaken. However, as this is a synchronic study of PPH, we will consider the issue of semantic narrowing and other features of lexical and semantic narrowing in light of the current PPH usage. The occurrence of semantic narrowing is found in all classes of words and also in all the six categories in which we have grouped the words. For example, the use of 'balai' in PPH means 'police station' as it does in Malay. However, there are no other uses of 'balai' in PPH, unlike in Malay, which has terms such as 'balai bomba' for 'fire station' or 'balai rakyat' for 'community hall'. Likewise, the root word 'amok' in Malay can function as a noun or a verb; but in PPH, it only functions as a verb where 'to (a)mok a person' is to beat a person without restraint characteristic of one running amok. In the same manner, the word 'racun' is used to describe someone who is a bane or one who disrupts other people's plans, and is not used to describe poisonous substances. Semantic narrowing also takes place in what we classify as actions-verbal and physical. For example, 'bicala' is only used for a trial and not used in its other sense, which is 'to speak'. The adjective 'haram' also undergoes semantic narrowing as it describes 'bearing a grudge against someone' or 'disliking someone'. It does not have the religious connotations of being 'forbidden' as it does in Malay. On the other hand, the word 'asap' in PPH has a strong religious connotation of burning incense for religious practices but does not function as an adjective as it does in Malay.

The speakers of PPH have also innovated on the meaning of certain Malay lexical items that they have borrowed. For example, the use of the lexical item 'buta' which means 'all for nothing' is not known among the Malays. Similarly, the term 'lumak' is used to describe a person as being flirtatious but is not used among the Malays.

## 8 Use of Malay above the level of individual lexicon

Under this category, we have found that the use of Malay at the sentential level is also present in PPH. We found that the use of Malay at this level seems to be fossilized and the terms used represent some form of formulaic expression. For example, the use of the saying 'pasang kuat' means dressing to impress. While 'pasang kaki' refers to the act of tripping someone by intent. Further examples of these are similar in use to colloquial Malay: for example, the term 'gatal' for 'itchy' / 'itch' is also used in Malay and PPH as an adjective to describe an amorous person. Other formulaic expressions function as adjectives with a preponderance for describing the negative nature of humans: examples are 'kepala angin' for 'temperamental', 'naik angin' for 'to lose one's temper', 'tiga suku' for 'crazy', 'ketam batu' for 'a stingy person', 'kurang ajak' for 'ill mannered' and 'muka binci' for 'hateful face'. Anybody who talks too much is referred to as 'cakap banyak' for 'talkative', while those adjectives that describe the positive in humans are few and far between, for example, 'muka sayang' for 'lovable face'.

We believe that at this level of language use, the speakers' creativity is shown, and there is more than just borrowings of singular lexical items necessitated by the lack of such concepts in the natal language of the speaker. We suggest that some form of loan translation takes place when the PPH speaker uses the saying 'The Batak is masquerading as a Nonya' or 'You sit like the worm when it has struck ash'. The former has the Malay equivalent of 'Batak menyamar Nonya' and the latter 'Cacing kena abu'. The above sayings are Malay sayings, and the fact that these sayings have found their way unaltered into PPH testify to the fact that there are loan translations.

Another use of Malay above the level of individual lexicon involves sayings which are essentially Malay although their meanings might be altered. For example, the saying that one is a 'tali barut' in PPH means 'to be insensitive to other people's desire for privacy'. However, in Malay the meaning is completely different; it means that one is a 'hated collaborator of the enemy'. While the saying 'naik jaki (cf. Malay 'dengki')' means to get angry with someone.

## 9 Conclusion

There is no evidence that points to the fact that the Peranakans in Penang never developed a language that was as heavily influenced by Malay as did the Malacca and Singapore Peranakans. In fact, early Peranakans in Penang were described as speakers of Baba Malay, although they were also said to be more fluent in Hokkien as compared to their southern cousins. Khoo states that the Penang Peranakans 'did not forget their native tongue (in general Hokkien) although all of them, like their Melaka counterparts, spoke Baba Malay' (Khoo, 1998:6). In fact, there are books printed in Baba Malay by Baba writers residing in Penang in the early twentieth century. However, there are very few Baba Malay speakers among the Penang Peranakans now; instead, there is a predominance of PPH speakers among the Penang Peranakans, and as proven by this study, PPH is essentially a Hokkien dialect with some Malay elements.

It is our belief that there are two possible scenarios leading to the situation where PPH could have gained currency among the Penang Peranakans to the extent that it supplants Baba Malay, although certain key Malay lexical items were maintained and introduced into PPH. This situation of language attrition (which in this case results in Baba Malay being subjected to the influences of Hokkien) was commented on by Clammer (1980) who cited Freedman (1962) that '...in 1870 it is estimated that only 20 per cent of

the Malacca and Singapore Babas could speak Chinese. But by 1975 the number was approximately 42 per cent in Malacca *and higher in Penang*, and rising in both cases (Clammer 1980:134, italics added).<sup>1</sup> This presupposes that there are Penang Peranakans who speak Baba Malay and that their numbers are decreasing. We will not speculate if there are still Penang Peranakans who still speak Baba Malay, but we are certain that their numbers are negligible if any.

This situation could have resulted from the different situation in which the Penang Peranakans found themselves, in contrast to their southern cousins. Clammer (1980) suggests that this difference (i.e. the use of Hokkien by the Peranakans in Penang as compared to the use of Baba Malay by the Peranakans in Singapore and Malacca) between the Penang Peranakans and their southern cousins could be that Penang had no significantly large native population before the British acquired it. Therefore, the Peranakan community was much less influenced by the Malays but much more influenced by the influx of Chinese immigrants who came to or passed through Penang since the early nineteenth century (Clammer 1980:8). Clammer (1980) suggests that Baba Malay evolved through the situation and circumstances of the early Chinese presence in Malacca since the late fourteenth century (Clammer 1980:8), and as Penang did not share many of the conditions prevailing in Malacca, the Penang Peranakans developed differently from those in Malacca.

Conversely, it could be that Baba Malay, like the other components of Baba culture in Penang might have absorbed increasingly significant Hokkien elements over time due to assimilation and integration into mainstream Chinese Culture after the Peranakan heyday. As Clammer (1980) observed and questioned '...are they (the Babas) in fact a perpetually marginal group because they are always being assimilated to someone else -- the Malays, the mainstream Chinese culture, or anglicized Chinese culture? (Clammer 1980:130)' This situation is entirely credible as the Peranakan community in Penang shifts from their initial association with the Malays in favor of the mainstream Chinese culture, as the initial predominance of the Malays in the former Straits Settlements was replaced by that of the Chinese. This shift towards the mainstream Chinese culture or anglicised Chinese culture could have been further facilitated by the fact that the Babas tend to look to Chinese culture as the main tradition with which they can identify (Clammer 1980:132). This is further compounded by the increasing pressures of westernization aggravated by the Babas' anglophilic persuasions. This would then account for the Baba Malay of Penang taking on more Chinese (Hokkien) and English lexicon, hence resulting in the birth and currency of present day PPH.

## References

Abdullah Hassan and Ainan Mohd. 1993. *Kamus Peribahasa Lengkap Utusan Utusan* Publications and Distributors. Kuala Lumpur, Malaysia.

Clammer, J. R. 1980. *Straits Chinese Society*. Singapore University Press.

Gwee, Thiam Hock William. 1993. *Mas Sepuloh: Baba Conversational Genres*. Singapore: Armour Publications.

Khoo, Joo Ee. 1996. *The Straits Chinese*. Kuala Lumpur: The Pepin press.

Khoo, Kay Kim. 1998. Malaysia: Immigration and the growth of a plural society. *Journal of the Malaysian Branch of the Royal Asiatic Society* 71.1: 1-25.

Pakir, Anne. 1986. *A Linguistic Investigation of Baba Malay*. Hawaii: University of Hawaii. Ph. D. Dissertation.

Pakir, Anne. 1989. Linguistic alternants and code selection in Baba Malay. *World Englishes* 8.3:378-388.

Shellabear, W. G. 1913. Baba Malay: An introduction to the language of the Straits born Chinese. *Journal of the Straits Branch of the Royal Asiatic Society* 65: 49-63.

Tan, Chee Beng. 1979. *Baba and Nonya: A study of the ethnic identity of the Chinese Peranakan in Malacca*. Ithaca: Cornell University. Ph.D. Dissertation.

Tan, Chee Beng. 1980. Baba Malay Dialect. *Journal of the Malaysian Branch of the Royal Asiatic Society* 53.1:150-165.

Tan, Chee Beng. 1993. *Chinese Peranakan Heritage in Malaysia and Singapore*. Kuala Lumpur: Fajar Bakti.

Teo, Kok Seong. April 1996. *Masyarakat Cina Peranakan Kelantan satu pengenalan sociolinguistik*. Jurnal Dewan Bahasa: 344-354.

Vaughan, J. D. 1879. *The Manners and Customs of the Chinese in the Straits Settlements*. Singapore, reprinted in Kuala Lumpur, OUP, 1971.

## Appendix: (Malay Lexicalised Items in Penang Peranakan Hokkien)

### 1 Trinkets

anting -- 'ear studs'  
 berlian -- 'diamonds'  
 bintang -- 'a star broach'  
 cincin -- 'ring'  
 gelang kaki -- 'anklet'  
 gelang -- 'bracelet'  
 gigi -- 'holder of gems/things and teeth of keys'  
 intan -- 'low grade diamond'  
 kelip -- 'sparkle'  
 kerosang -- 'broach' cf. kerongsang in Malay  
 liante -- 'chain', cf. rantai in Malay.  
 satu -- 'solitaire'  
 suasa -- 'low grade (9-carat) gold'  
 subang -- 'earrings'  
 tapak sireh -- 'sireh box'

### 2 Animals

babi -- 'pigs'  
 belalang -- 'grasshopper'  
 buaya -- 'crocodile'  
 burong kuku -- 'pigeons'  
 burong -- 'birds'  
 cacing -- 'worm'  
 cangkerik -- 'cricket'  
 cicak -- 'lizards'  
 katak puru -- 'frogs', cf. katak in Malay.

kongkiak -- 'big black ants'  
 kutu -- 'lice'.  
 manatang -- binatang 'animal'.  
 mosang 'bobcat' cf.-musang in Malay.  
 pacat -- 'leach'  
 sipot -- 'cockles'  
 sotong -- 'octopus'

### 3 Things

(h)abok -- 'dust'  
 abu -- 'ash'  
 adek~alek -- 'younger siblings'  
 bakuli (buah guli) -- 'marbles'  
 bangsa -- 'race' / 'ethnic origin'  
 bangsat -- 'a derogatory term for a disliked person'  
 balai -- 'police station'  
 baleh -- 'bench'  
 bangku -- 'stool'  
 barang-barang -- 'things'  
 batang -- 'a stem or stemlike thing'  
 batu -- 'stone'  
 bayang-bayang -- 'shadows'  
 bedak -- 'powder'  
 bedak sujuk -- 'rice powder'  
 bidan -- 'midwife'  
 biji -- 'seed'  
 botoi - 'botol' in Malay (a borrowing from the English 'bottle')  
 buah -- 'fruit'  
 buai-buai -- cf. buian in Malay -- 'swing'  
 bunga melor -- 'jasmine flower'  
 butut -- 'cork'  
 cawat -- 'loin cloth'  
 curut -- 'cigar'  
 gaji -- 'pay/wages'  
 gambit -- 'gambir'  
 geling rumpah -- 'stone used to grind spices'  
 getah -- 'sap of fruits' / 'rubber band'  
 hamba -- 'slave'  
 jamban -- 'toilet'  
 kapok -- 'cotton', cf. kapas in Malay.  
 kapok -- 'lime'  
 kasut manek -- 'beaded shoes'  
 kasut kodok -- 'a type of beaded shoes'  
 kawan (noun/verb) kawan (where there is a superimposition of hi-lo or falling tones)  
 ketam batu -- 'a stingy person'  
 kole -- 'a big steel mug'  
 lalang -- 'wild grass'

lampeng -- 'diapers' cf. lampin in Malay  
 lili -- cf. lidi in Malay  
 longkang -- 'drain'  
 mata -- 'police' cf. mata-mata in Malay, where 'mata in its base form means 'eyes'  
 mesegate -- 'mosque' cf. mesjid in Malay  
 orang gaji -- 'servant, maid'  
 orang jaga -- 'a maid who specializes in looking after mothers in confinement'  
 pagar -- 'fence'  
 piring -- 'plates, saucers'  
 puluk 'bolster'  
 racun -- 'a baneful person'  
 rokok (chau [Hokkien] -- grass) -- 'tobacco'  
 sabun -- 'soap'  
 sampah -- 'rubbish'  
 sanggoi -- 'bun of hair of a lady', cf. sanggul in Malay  
 sarong  
 sepatu -- 'shoes'  
 senduk -- 'spoon/ladle'  
 sireh -- 'betel nut leaf'  
 sulam -- 'embroidery'  
 tali -- 'string/rope'  
 tangkai -- 'amulet'  
 tanjong -- 'cape'  
 tembok -- 'wall'  
 timba -- 'pail'  
 tuala -- 'towel'  
 tumbuk -- 'pestle'

#### 4 Methods of cooking/food

acat -- cf. acar in Malay -- 'chutney'  
 air-mawar -- 'essence'  
 apom-bakuah -- cf. apom berkuah in Malay -- 'pancake with sauce'  
 apom -- balek -- 'a pancake'  
 asam pedai 'asam pedas' in Malay -- 'a dish'  
 asam -- 'tamarind'  
 bangkuang -- 'turnips', cf. mengkuang in Malay  
 bawang -- 'onion'  
 bayam -- 'spinach'  
 belacan -- 'shrimp cake'  
 buah betai 'buah petai' in Malay -- 'a type of vegetable'  
 buah kerai -- 'buah keras' in Malay  
 bunga-kantan -- 'blue ginger'  
 bunga telang -- 'a type of flower'  
 cabai -- 'chili'  
 cabai burung -- 'small chili'  
 cincalo -- 'preserved shrimps'  
 daun kaduk -- 'a type of herb'

goreng pisang -- 'a dessert'  
 goreng -- 'to fry'  
 gulai -- 'curry'  
 jambu-air -- 'guava (soft)'  
 jintan-manis -- 'coriander'  
 kacang-botoi -- 'a type of vegetable'  
 kacang nyenyak -- 'ladies finger'  
 kerabu -- 'salad' (this could also function as a verb, i.e., 'to make something into salad')  
 kueh-bangkek -- 'a type of cake'  
 kueh -- bengkah (bengkang) -- 'a type of cake'  
 kueh-kapek -- 'a type of cake, 'love letters'  
 kueh koci -- 'a type of cake'  
 kueh kodok -- 'a type of cake'  
 kueh-lapis -- 'layered cake'  
 kunyet -- 'tumeric'  
 lengkuas -- 'a type of plant' cf. lengkuas in Malay  
 lobak -- 'white carrots'  
 lumak -- 'creamy'  
 lumpa -- 'rempah' in Malay  
 lun-pandan ~ daun pandan in Malay -- 'pandan leaves'  
 nasik-kunyet -- 'yellow ginger rice'  
 nasi ulam -- 'a dish'  
 nona -- 'custard apple'  
 otak-otak -- 'steamed fish wrapped in banana leaf'  
 panggan -- 'to grill' cf. panggang in Malay  
 pulut ikan -- 'fish maw curry' cf. perut ikan in Malay.  
 pulut -- 'glutinous rice'  
 pulut inti -- 'glutinous rice cake'  
 roti -- 'bread'  
 roti jala -- 'a dish'  
 sambai-belacan/udang -- 'a type of chili paste' or sambal udang in Malay  
 santan -- 'coconut milk'  
 timun -- 'cucumber', cf. mentimun in Malay  
 tumis~tumi -- 'to saute', cf. tumis in Malay  
 ubi-kayu -- 'sweet potato'  
 unde-unde -- 'a type of cake'  
 ulam -- 'salad'

## 5 Body terms

bak ketiak -- 'arm pits'  
 batang -- 'penis'  
 bisoi -- 'wart' cf. bisul in Malay  
 buah peliak -- 'scrotum'  
 daki -- 'dirt'  
 gigi -- 'teeth'  
 jantan -- 'male'  
 jari -- 'finger'

liba -- 'lap'  
 lutut -- 'knee'  
 mata sepek -- 'squint eyed'  
 nanah -- 'puss'  
 paha -- 'lap'  
 pinggan -- 'waist'  
 puki -- 'female genital'  
 tetek -- 'breasts'

## 6 Actions-verbal + physical

agak -- 'almost' / 'to estimate'  
 antak -- 'to send', cf. hantar in Malay  
 asap -- 'to light incense'  
 batuk -- 'to cough'  
 bangkai -- 'carcass or bad-smelling'  
 bangkali (barangkali) -- 'possible'  
 bantai -- 'to thrash'  
 bantang -- 'superstition', cf. pantang in Malay  
 baru -- 'new, just'  
 belang-belang -- 'stripes'  
 bengkok -- 'bent'  
 betui-betui -- 'really', cf. betul-betul in Malay  
 bicala -- (noun) 'court proceedings', cf. bicara in Malay  
 bilai (bilas) -- 'to rinse'  
 bising -- 'noisy'  
 botak -- 'bald'  
 bukumbak -- 'twins' cf. berkembar in Malay  
 bukumang (berkembang) -- 'wearing the sarung at the armpits'.  
 bunting -- 'pregnant'  
 burok (adj) -- 'bad'  
 busuk -- 'foul smelling'  
 buta -- 'all for nothing'  
 cabai -- 'to punish one with the application of chili'  
 cakap banyak -- 'talkative'  
 campur -- 'to mingle'  
 celaka -- 'a curse'  
 cicit -- 'great grandchildren'  
 comek -- 'cute', cf. comely in Malay  
 comot -- 'all crumpled'  
 cubit -- 'to pinch'  
 cucuk -- 'to instigate'  
 dan-dan -- 'immediately'  
 gaji -- 'to employ'  
 galek -- 'scratch, lacerations'  
 galoh -- 'to quarrel', cf. gaduh in Malay  
 ganggu -- 'to disturb'  
 gasak -- 'to deride'

gatai -- 'amorous' cf. gatal in Malay  
 gatai -- 'itch'  
 gayat -- 'frightened'  
 geram -- feelings of 'hate'(negative) / feeling of fondness (positive)  
 gili -- 'tickle', cf. geli -- 'disgusting'  
 gomoi -- 'wrestle', cf. gumol in Malay  
 gugur -- 'abortion'  
 gulong -- 'to roll up'  
 had -- 'limit'  
 hangat -- 'angry'  
 haram -- 'bear a grudge'  
 harap -- 'depend'  
 hentam -- 'to hit'  
 heran -- 'curiously', cf. hairan in Malay  
 hitam manis -- 'dark and sweet'  
 horomat -- 'respect', cf. hormat in Malay  
 ikat -- 'a bunch of fruit tied together'  
 jadi bo -- 'successful, did it turn out right?'; bo=a question particle in Hokkien  
 jalan-jalan -- 'to stroll'  
 janji -- 'to promise'  
 jarang-jarang -- 'sparse, but not hardly'  
 jelak -- 'nauseated from overeating a type of food'  
 jijik -- 'nauseating'  
 jilat -- 'to lick'  
 juling -- 'to stare'  
 kacau -- 'to disturb'  
 kaku -- 'awkward'  
 kangkang -- 'astride'  
 karut -- 'nonsense'  
 kawen -- 'to marry', cf. 'kahwin' in Malay  
 kejap -- 'sarong worn tightly and neatly'  
 keliling (kiah=walk) 'around'  
 kepala -- 'supervisor'  
 kepala angin -- 'temperamental'  
 kepala pusing -- 'headache'  
 kerakot -- 'crooked'  
 kesian -- 'pity', cf. kasihan in Malay  
 kedukut -- 'stingy'  
 ketak-ketak -- 'frightened out of one's wits'  
 kilat -- 'shinny'  
 kuat -- 'strong'  
 kuasa -- 'control / authority'  
 kunduri -- 'a party held by the Malays' cf. kenduri in Malay  
 kurang ajak -- 'ill-mannered'  
 laju -- 'withered'  
 laku -- 'saleable'  
 lalai -- 'forgetful'

lalu -- 'appetite'  
 lama-lama -- 'eventually'  
 langkah -- 'stride-over'  
 langsung -- 'absolutely'  
 lapat -- 'to obtain' cf. dapat in Malay  
 lapek -- 'to cushion'  
 latah -- 'to curse uncontrollably'  
 lelong -- 'to auction off'  
 lembek -- 'flaccid'  
 lembut -- 'soft spoken/petite'  
 lengan-lengan -- 'to stroll'  
 letak -- 'to put'  
 lumak -- 'flirtatious'  
 lumpang -- 'to hitch a ride' cf. tumpang in Malay  
 luput -- 'rot' cf. reput in Malay  
 mabok -- 'drunk, intoxicated'  
 macam-macam -- 'all sorts'  
 makan gaji -- 'wage earner'  
 mana -- 'where'  
 manja -- 'to pamper'  
 masak-masak -- 'a child's game'  
 merepek -- 'talk nonsense'  
 muka binci ~ muka penyala -- 'hateable face'  
 muka sayang -- 'loveable face'  
 (a)mok -- 'to hit another without restrain' (characteristic of running amok)  
 naik angin -- 'to lose one's temper'  
 naik jaki -- cf. naik dengki in Malay -- 'to get angry with'  
 naiya -- 'to be victimized', cf. aniayai in Malay  
 (ha)nya -- 'only'  
 nyanyok -- 'senile'  
 padan -- 'serves one right'  
 pakat -- 'to collaborate'  
 pandai -- 'smart in a pejorative sense'  
 pasai -- 'because', cf. pasal in Malay  
 penyek -- 'squashed' / 'flat'  
 pereksa -- 'to examine'  
 puas -- 'satisfied'  
 pun -- 'also'  
 pula -- 'also'  
 rajin -- 'hardworking'  
 rasa -- 'taste/feeling'  
 rindu -- 'miss'  
 salah-(adj) -- 'wrong'  
 sambai -- 'until' cf. sampai in Malay  
 sangkot -- 'entangled'  
 satu macam -- 'eccentric'  
 sayang -- 'to love'

sayang -- 'what a pity'  
 sedap -- 'tasty'  
 sekali -- 'one shot, at one go'  
 selongkah -- 'to search', cf. selongkar in Malay  
 semak -- 'a mess'  
 sembang -- 'to chit-chat'  
 sembunyi -- 'to hide'  
 sendak -- 'lean' cf. sandar in Malay  
 sengek -- 'crooked lopsided'  
 senja -- 'dusk'  
 senyom -- 'to smile'  
 senyom manis -- 'a sweet smile'  
 sepak -- 'to slap'  
 silap -- 'a mistake'  
 som-som -- 'lukewarm' cf. suam-suam in Malay  
 suka -- 'to like'  
 suku -- 'a quarter'  
 sunyi -- 'quiet'  
 tabik -- 'salute'  
 tahan -- 'to withstand'  
 tak tentu -- 'uncertain'  
 tali barut -- 'to be insensitive to other people's need for privacy'  
 tambah -- 'to add'  
 tapi -- 'but', cf. tetapi or tapi in Malay  
 tapih -- 'filter' cf. tapis in Malay  
 tarik -- 'convulsion'  
 tekejut -- 'shocked'  
 te(r)keliat -- 'sprained/twisted'  
 tenggelam -- 'to sink'  
 tepelanting -- 'to collapse from being slapped'  
 terbalik -- 'to overturn/overturnd' or 'to make a u turn'  
 tehrap -- 'to lie on the stomach' cf. tiarap in Malay  
 tiga suku -- 'three quarters, or crazy'  
 tolong -- 'help'  
 ulu -- 'being provincial'  
 ulut -- 'massage' cf. urut in Malay

## 7 Interjections

adoi  
 alah  
 alamak  
 amboi  
 ceh  
 lah

# A UNIFIED ANALYSIS OF SOME VIETNAMESE REDUPLICATION FORMS

Sonny X. Vu

Massachusetts Institute of Technology

<sonny@sonny.vu>

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

In Vietnamese, one of the main forms of derivational morphology is reduplication. Although reduplicative forms in Vietnamese have been extensively catalogued (Hoàng Văn Hành<sup>2</sup> (1997)) and at least partially described (Emeneau (1951), Thompson (1965), Ngô Thanh Nhàn (1984), Hoàng Văn Hành (1985), Nguyễn Tài Cản (1996), Nguyễn Kim Thán (1997), et seq.), they have been subjected to limited analysis (Ngô Thanh Nhàn (1984), Agbayani (1997)) and not examined for general principles. A crucial distinction that often has been overlooked in previous studies is that only a few of the patterns identified as reduplications are productive. In this study, I describe, characterize, and present an OT analysis of four major *productive* patterns. I show that rerankings of a core set of well-attested constraints provide a natural account for the data that I present.<sup>3</sup> Most importantly, I show that a number of seemingly disparate properties in these four processes follow from differences in the prosodic structures (Selkirk (1978) et seq.) of the output forms.

The four processes I describe<sup>4</sup> and analyze are: Full Reduplication (FR), reduplication involving Emergence of the Unmarked effects (TETU) (McCarthy and Prince (1994b)), and two processes involving Melodic Overwriting (Alderete, et al. (1997))

---

<sup>1</sup> This paper was supported in part by the Vietnamese Advanced Summer Institute 1997 Program. Special thanks to Nguyễn Thị Minh-Phuong for clarifying judgments on all the data in this paper, Cheryl Zoll for guiding this study, and my classmates Calixto Agüero-Bautista and Philippe Schlenker for extensive discussions. Thanks to the participants in the 1997 Workshop in Phonology and Morphology at MIT, the audiences at the 14<sup>th</sup> Northwest Linguistics Conference and at the 8<sup>th</sup> Southeast Asian Linguistics Conference, Brian Agbayani, Mark Alves, Cao Xuân Hạo, ñinh Lê Thu, Ken Hale, Morris Halle, Michael Kenstowicz, Alec Marantz, Ngô Nhu Bình, Rob Pensalfini, Phạm Hoa, Shinko Tamura, and Joseph Thai for useful comments and discussions. Usual disclaimers apply.

<sup>2</sup> For the sake of clarity, all Vietnamese names mentioned in this paper will be written in the traditional format and orthography: last, middle, first name. All Vietnamese data will also be presented in the traditional Vietnamese orthography for ease of reference.

<sup>3</sup> Hence, Vietnamese presents a case of conflicting rankings of constraints within a language.

<sup>4</sup> In the data presented, subscripts denote tones (see below for numbering scheme) and superscripts denote stress levels on the following syllable. For example:

<sup>0</sup>nhà<sup>1</sup>toán<sup>0</sup>hoc = 'mathematician'

bears the stress pattern 010. The base (BASE) is underlined when it can be determined. Other abbreviations: emph = emphatic; pej = pejorative; att = attenuative; int = intensive; RED = reduplicant. Some of the data in this paper were taken from a number of sources such as Hoàng Văn Hành (1997)'s *Từ Điển Từ Láy* (a dictionary of reduplicative forms produced by the National Linguistics Institute of Vietnam) or from Ngô Thanh Nhàn (1984). All data in this paper are either provided or confirmed by native speaker consultants including, Cao Xuân Hạo, Nguyễn Thị Minh-Phuong, Phạm Hoa, and myself.

which I will refer to as a-reduplication (a-Red) and iec-reduplication (iec-Red). A summary of some of the differences is provided in Table 1.

<b>FR:</b>	sạch → <sup>0</sup> sạch <sup>1</sup> sạch = ‘clean’ → ‘clean’ (+attenuative, informal) <sup>5</sup>
<b>TETU:</b>	đẹp → <sup>0</sup> đèm <sup>1</sup> đẹp = ‘pretty’ → ‘pretty’ (+attenuative, formal)
<b>a-Red:</b>	<sup>1</sup> lò <sup>1</sup> mò → <sup>0</sup> lò <sup>2</sup> mà <sup>0</sup> lò <sup>1</sup> mò = ‘to grope’ → ‘to grope’ (+emphatic)
<b>iec-Red:</b>	bán → <sup>1</sup> bán <sup>1</sup> biệc = ‘to sell’ → ‘to sell’ (+pejorative)

**Table 1: Differences between four productive reduplication processes**

Type	Number of syllables in input (BASE)	Stress pattern of output	Syntactic separability of constituents	Avoidance of identical syllables
FR	1	01	no	no
TETU	1	01	no	yes
a-Red	2	0201	no	yes
iec-Red	1 or 2	11, 0101	yes, yes	yes

In Table 1, ‘syntactic separability’ refers to the observation that some of the reduplicative forms in Vietnamese can have syntactic material inserted between BASE and RED if they can be determined. The last column in Table 1 refers to the fact that some processes do not allow the copying of identical material while others do. A shared feature of the four processes is the fact that the number of syllables in the base is equal to the number of syllables in the RED. Thus, for example, the output of a-Red always has four syllables whereas the output for FR is always disyllabic since a-Red always takes a disyllabic base and FR a monosyllabic base. Another shared feature is that all four processes preserve the grammatical category of BASE.

In addition to exploring the differences between these four processes, I also examine the common features shared among them. These include: complete preservation of the base in the input, lack of double reduplication (the application of a reduplicative process on a reduplicated form), and respect of regular tone-coda co-occurrence restrictions that appear elsewhere in the language.

One of the main empirical observations made in this study is the distinction between productive and non-productive reduplication patterns. Although both types obey some of the constraints I posit, only the productive patterns are assumed to be produced by the grammar, whereas the non-productive cases (the majority of reduplicative forms in Vietnamese) are not. Although this distinction has been alluded to in the Vietnamese linguistics literature (e.g., Ngô Thanh Nhàn (1984)), many of the descriptions of Vietnamese reduplication (e.g., Nguyễn Tài Cẩn (1996), Nguyễn Kim Thán (1997)) assume that all reduplicative forms are formed in the grammar by various rules; however, a cursory review of the data shows that these rules over-generate the possible forms in the language.

<sup>5</sup> All final obstruents in Vietnamese are unreleased. I do not indicate this in the data for typographical simplicity.

## 2 Vietnamese Phonology and Theoretical Assumptions

### 2.1 Relevant Vietnamese Phonology

**Tones:** In most Northern dialects, the dialects of focus in this study, Vietnamese has six phonologically contrastive tones. Every syllable is associated with one of these tones. Although I will use the same notation as Burton (1992)'s, I depart from his categorization by viewing the tone pairs {1, 4}, {2, 5}, and {3, 6} to be related not by contour shape (since they are clearly not similar in shape, e.g., tone 2 is rising whereas tone 5 is dipping) but rather, in terms of their 'bundles' of phonetic features.<sup>6</sup> See Wannamacher (1997) for a related discussion on voice quality as a potential tonal feature. Tones 1 and 4 are both lax in terms of glottal stricture and medium in terms of length (for more discussion of length as a tonal parameter, see Alves (1997b)). Tones 2 and 5 have tense glottal stricture, are glottalized at the end, and are short. Tones 3 and 6 are also both tense but are long, in fact, longer than all the other tones. Perhaps one way to characterize tonal markedness in Vietnamese is to say that tones that are lax are unmarked and the ones that are not lax are more marked. For a detailed phonetic description of Vietnamese tones, see Han (1969). See Edmondson and Lợi (1997) for a detailed phonetic study of the tones in Northern Vietnamese. The main motivations for the coupling of tones as shown in Table 2 are: their synchronic phonological correlations (e.g., their related distributions in reduplication as described in this paper, see Võ Xuân Hạo (1997) for a detailed discussion) and their tonogenetic history.<sup>7</sup>

**Table 2: Vietnamese tone registers and contours**

		Contour <sup>8</sup>		
		1	h	lh
Register	+High	1	2	3
	-High	4	5	6

There is no phonological tone sandhi except in a few domains of the grammar, e.g., reduplication. This characterization is motivated not only in terms of their phonological 'well-behavedness' but also by a number of tonogenetic accounts such as Haudricourt's hypothesis (Haudricourt 1954) of tonal splitting due to the presence of voiceless obstruent codas lowering the pitch of tones in early Vietnamese.

<sup>6</sup> Rolf Noyer has mentioned in personal communication that perhaps the characterization based on the contour and register parameters can be maintained. The actual contour shape of tone 5 might then just be a phonetic reflex of the tone being in the lower register and having the glottalization at the end.

<sup>7</sup> Historically, Vietnamese developed phonetically conditioned allophonic features in different syllable types, A (open syllables), B (syllables with final fricatives), and C (syllables with final stops p/t/c/k/?). As the coda consonants in Vietnamese changed (fricatives and glottal stops, which are well attested in other Mon-Khmer languages, were lost in Vietnamese), those phonetic categories became phonemic (i.e., unpredictable based on environment). The next step was the conditioning of height based on the voicing of the initial. Tonal categories A, B, and C each developed two allophones, with voiceless onsets resulting in higher tones and voiced, lower tones. Eventually, Vietnamese initials changed, thereby masking the original voicing, but leaving the tonal height, which became phonologically contrastive. Thus, 3 tonal phonemes became 6. (Haudricourt (1954), Alves (1997b))

<sup>8</sup> Or better put, grouping with respect to voice quality.

In terms of tonal features, I adopt Agbayani's correlation of [α vocal fold] features with tone contours on vowels and voice distinctions on consonants:

[stiff]  $\Rightarrow$  C[-voice] or V[h] and [slack]  $\Rightarrow$  C[+voice] or V[l]

**Vowels:** Vietnamese has nine vowels. Although vowels obligatorily lengthen in open syllables, two of the vowels /ɛ/ and /a/ are contrastive in terms of length in other environments as well. The place features of the vowels are the most relevant to the present discussion. For a detailed phonetic study of Vietnamese vowels, see Trần (1967) and Han (1966).

Vowel:	[i]	[e]	[ɛ]	[u]	[o]	[ɔ]	[ɯ]	[ɛ̄]	[a]
Place:	+cor	+cor	+cor	+lab +dor	+lab +dor	+lab +dor	+dor	+dor	+rad

**Tone-coda co-occurrence restriction:** Syllables with final obstruents, i.e., [p, t, k], can only be assigned tones with contour value **h**, i.e., tones 2 and 5. Otherwise, any of the six tones can be assigned to any other syllable type. For typographic reasons, diphthongs and triphthongs are represented here the two and three vowels, respectively.

**Other:** Vietnamese lacks consonant clusters and has only a limited number of possible codas. The possible syllable types are: C(G)V(:)C and C(G)V: where G is a glide. Note that C(G)V is not allowed, i.e., syllables with only an onset (without or without a glide) and a short vowel. For typographic reasons, diphthongs and triphthongs are represented here the two and three vowels, respectively.

Although any consonant from the phonetic stock can be an onset, the only possible codas are the voiceless unreleased obstruents [p, t, k], nasals [n, m, ɲ] and glides [w, j]. There are also no co-occurrence restrictions on vowels (or vowel clusters) and onsets. The most noticeable phonotactic effects are the co-occurrence restrictions of certain vowels and coda consonants and the co-occurrence restriction on certain tones and codas. The matching of alveolar and palatal place features between main vowels and coda consonants is the most salient pattern of the former: the vowels [i.e., ɛ] cannot occur with the velar codas [k, ɲ].<sup>9</sup> For a detailed discussion of the general nature of the Vietnamese phonological system, see Thompson (1965) and Đoàn Thiện Thuật (1977).

## 2.2 Theoretical Assumptions

**Optimality Theory:** My analysis is presented in the Optimality Theory (OT) framework in which universal but violable constraints are posited. Within this theory, languages differ according to the way these constraints are ranked with respect to one another. The main mechanisms in OT are:

*Gen:* function which generates all the possible candidates for a given input

*Eval:* function which evaluates the optimality of each of the candidates with respect to a language-specific constraint ranking for the given input

<sup>9</sup> Pham (1997) gives an account of the distribution of final consonants in 4 dialects (Hanoi, Saigon, Hue, Quangnam) and for the co-occurrence restriction of velars after front vowels.

*Con*: set of universal constraints

The optimal candidate selected by Eval is the one which violates the highest ranking constraints (taken from *Con*) the least number of times, i.e., the one which satisfies the most number of constraints starting from the most highly ranked one. In the case where candidates tie, in the sense that the highest ranking constraint they violate is the same one, the candidate which violates this constraint the least number of times wins; if they tie on this measure, then the one which satisfies the next highest constraint the least number of times wins, and so on.

One notion in OT which is crucial to my analysis is Correspondence Theory (CT) (McCarthy and Prince (1994a), McCarthy and Prince (1995)), an approach which provides a mechanism for the general purpose of string comparison. A correspondence relationship is said to measure the *faithfulness* of the output to the input, i.e., how much deviation exists between the output and the original input form. In the theory of reduplication, this notion of Correspondence is also useful in the comparison of how different the reduplicant (RED) is from the base (BASE) as well as how faithful the output is to the input. The correspondence relation between two strings is defined in 0.

**Correspondence:** Given two strings  $S_1$  and  $S_2$ , correspondence is a relation  $\mathfrak{R}$  from the elements of  $S_1$  to those of  $S_2$ . Segments  $\alpha \in S_1$  and  $\beta \in S_2$  are referred to as correspondents of one another when  $\alpha \mathfrak{R} \beta$ . (McCarthy and Prince (1995))

**Prosodic structure:** In this study, I rely heavily on making reference to prosodic structure. I posit these structures as a part of my analysis because a number of generalizations, such as stress patterns and syntactic separability, of these processes can be accounted for purely in prosodic structural terms. In following this approach, I have assumed the Prosodic Hierarchy Hypothesis (Selkirk (1978), Nespor and Vogel (1982), Hayes (1989), et seq). According to this theory, prosodic domains are divided into a number of salient levels on which various phonological and morphological processes can take place. The hierarchy is given in 0:

Prosodic Hierarchy	Domain
$\phi$	Phonological Phrase
$\omega$	Phonological Word (or Prosodic Word)
$F_t$	Foot
$\sigma$	Syllable
$\mu$	Mora

The levels that are particularly relevant to this discussion are the foot and the phonological word. In Vietnamese, as in many languages, each foot is assigned one stress. It is assumed that in Vietnamese, this level ordering is strict and exhaustive in the sense that there are no other levels intervening between any two given levels and that for a given utterance all of the levels exist in the prosodic structural representation. Therefore, for example, given a phonological word, one cannot insert another phonological word inside

of it. However, within a phonological phrase, a phonological word may be inserted in between two other words. This last fact will be the basis of one of my arguments for different processes having different prosodic structures.

**Reduplication:** In this paper, I assume ideas from Generalized Template Theory, an approach couched in the OT framework that views the shape of reduplicants as the result of morphological specifications. This approach views reduplication as a process of concatenating a RED affix that is phonetically empty in the underlying representation which is later filled in by various phonological and/or morphological processes. (McCarthy and Prince (1994a), Urbanczyk (1997)) Proceeding along these line, Alderete et. al. (1996) makes a distinction between two different kinds of processes which provide the reduplicant with phonetic content. Both processes involve the insertion of fixed material; they are characterized as: TETU processes and Melodic Overwriting (henceforth, MO) (See Yip (1992)). TETU arises from phonological constraints on markedness (segmental and tonal) whereas MO derives from morphological sources. There are examples of both kinds of processes in Vietnamese.

### 3 Data: Descriptions and Generalizations

#### 3.1 Productive processes

In this section, I describe in detail the four types of reduplication processes that I will be investigating in this paper. The reason for selecting these processes is that they are the most productive, if not the only ones, in the language. As shown below, many of the forms which do not fall in any of these four processes should be treated as lexicalized forms. Independent evidence for this proposal includes the fact that they have irregular changes in meaning as well as non-productive, non-predictable phonological and morphological form, and that they can themselves undergo reduplication.

**Full reduplication:** One of the most common types of reduplication in Vietnamese is full or total reduplication. In these cases, RED is an exact copy of BASE both segmentally and tonally. The semantic change in these cases is usually attenuative. There are a number of cases given in Thompson (1965) which report that plural and distributive meanings are possible when nouns undergo FR as seen in examples 0; however, these cases only exist in certain dialects and are very unnatural in many dialects, including most of the Northern dialects. In general, FR can only operate on stems that are adjectives (a-d, i, j), adverbs (g, h), or verbs (e, f). The stress patterns that obtain in these forms are regularly 01 (personal judgment, Cao Xuân Hạo (1998a)). As shown below, in terms of meaning changes and stress assignment, these forms most closely resemble the TETU forms mentioned above.

#### Full reduplication examples

a.	nhanh → <sup>0</sup> nhanh <sup>1</sup> nhanh	(+H,J)(+H,I)	fast (+att)
b.	đẹp → <sup>0</sup> đẹp <sup>1</sup> đẹp	(-H,h)(-H,h)	pretty (+att)
c.	cười → <sup>0</sup> cười <sup>1</sup> cười	(-H,I)(-H,I)	smile (+att)
d.	nhỏ → <sup>0</sup> nhỏ <sup>1</sup> nhỏ	(+H,lh)(+H,lh)	small (+att)
e.	sách → <sup>0</sup> sách <sup>1</sup> sách	(+H,h)(+H,h)	book (+dist)

These forms do not have the separability effects. Also, they cannot form the base for any reduplication process (or any other derivational morphological process in general, e.g., nominalization) and only monosyllabic words can serve as the base for FR. (Exceptions are usually more literary forms.)

*Reduplication via replacement with unmarked material:* As mentioned above, a significant number of reduplication cases in Vietnamese can be accounted for by the claim that RED is filled with material that is unmarked where possible and is minimally different from the base. In this way, these cases are productively formed by the grammar. The examples in 0 - 0 (data taken and readjusted<sup>10</sup> from Agbayani (1997)) illustrate the proposal that [+stiff vocal cord] as well as [+labial] and [+dorsal] vowel place features (as opposed to [+coronal] and [+radical]) are marked structures in Vietnamese. The consequence of these marked structures is detailed in and 0 and 0 and illustrated in examples 0 - 0. Independent evidence for the marked status of these forms include the relative sparseness of these forms indicated in various statistical studies of text corpora (Ngô Thanh Nhàn (1984), Võ Xuân Hạo (1997)) and the forms' relative non-occurrence in loan phonology as shown in studies such as Nguyễn Đinh-Hòa (1980).

#### Consequences of tonal markedness

The marked tonal feature [+stiff vocal cord] is associated with tones 2, 3, 5, and 6, thus making these tones more marked than the ones that do not have this feature, i.e., tones 1 and 4.

#### Consequences of vowel place markedness

The markedness of vowel place features [+labial] and [+dorsal] results in the vowels [u, o, ɔ] being the most marked for having both of these, followed by [i, e, ε] for having only one of these features, and finally by the least unmarked vowels [ɯ, ɤ, ɑ] for having neither of these features. This is deduced from the vocalic stock whose features are summarized in section 2.

To illustrate how unmarked material emerges in reduplication, I present a set of data exemplifying the regular tonal changes that occur in TETU reduplication. The examples below differ minimally on their basic syllabic structure: syllables with no coda in the base 0, nasal codas 0, and voiceless obstruent codas 0. See Agbayani (1997) for a more detailed discussion of marked (and unmarked) structures in Vietnamese.

In examples 0 and 0, the register of the tone associated with the base is preserved in RED and the contour of the tone in RED changes to 1. Facts concerning tone harmony were also observed in Vũ Thé Thạch (1994). Note that in the case where BASE already has an unmarked tone, the tone does not change in RED.

#### Syllable type: no coda in BASE

(-H, c) in BASE → (-H, l) in RED; no segmental changes

i.	nhiều	(-H,l)	be a large amount
ii.	<sup>0</sup> nhiều <sup>1</sup> nhiều	(-H,l)(-H,l)	be a rather large amount

<sup>10</sup> Glosses readjusted based on native speaker judgments; stress patterns added.

iii.	nhẹ	(-H, <b>h</b> )	be light (in weight)
iv.	<sup>0</sup> nhe <sup>1</sup> nhẹ	(-H,I)(-H, <b>h</b> )	be rather light

(+H, c) in BASE → (+H,I) in RED; no segmental changes

i.	hiu	(+H,I)	(breeze) is gentle
ii.	<sup>0</sup> hiu <sup>1</sup> hiu	(+H,I)(+H,I)	blow very lightly
iii.	khá	(+H,h)	rather good
iv.	<sup>0</sup> kha <sup>1</sup> khá	(+H,I)(+H,h)	rather mediocre
v.	nhỏ	(+H,lh)	small
vi.	<sup>0</sup> nho <sup>1</sup> nho	(+H,I)(+H,lh)	rather small

(8) Syllable type: sonorant (nasal) coda in BASE

(-H,c) in BASE → (-H,I) in RED; no segmental changes

i.	buồn	(-H,I)	be sad
ii.	<sup>0</sup> buồn <sup>1</sup> buồn	(-H,I)(-H,I)	be a little sad
iii.	chậm	(-H,h)	be slow
iv.	<sup>0</sup> châm <sup>1</sup> chậm	(-H,I)(-H,h)	be somewhat slow
v.	loãng	(-H,lh)	diluted
vi.	<sup>0</sup> loãng <sup>1</sup> loãng	(-H,I)(-H,lh)	rather diluted

(+H,c) in base → (+H,I) in reduplicant; no segmental changes

i.	xanh	(+H,I)	be blue, green
ii.	<sup>0</sup> xanh <sup>1</sup> xanh	(+H,I)(+H,I)	be bluish, greenish
iii.	trắng	(+H,h)	be white
iv.	<sup>0</sup> trắng <sup>1</sup> trắng	(+H,I)(+H,h)	be whitish
v.	ẩm	(+H,lh)	be humid
vi.	<sup>0</sup> ẩm <sup>1</sup> ẩm	(+H,I)(+H,lh)	be slightly humid

In 0, note that the same tonal change occurs as seen in the above examples. However, in order to obey the tone-coda co-occurrence restriction, the change also drives a segmental change when the base has an obstruent in the coda: the coda in the reduplicant is the homorganic nasal corresponding to the obstruent in the coda of the base.

Syllable type: obstruent coda in base

( $\alpha$ H,h) in BASE → ( $\alpha$ H,I) in RED; obs] $_{\sigma}$  in base → nasal] $_{\sigma}$  in RED

i.	đẹp	(-H,h)	be beautiful
ii.	<sup>0</sup> đèm <sup>1</sup> đẹp	(-H,I)(-H,h)	be rather pretty
iii.	sạch	(-H,h)	be clean
iv.	<sup>0</sup> sành <sup>1</sup> sạch	(-H,I)(-H,h)	be rather clean
v.	tốt	(+H,h)	be good, fine
vi.	<sup>0</sup> tôn <sup>1</sup> tốt	(+H,I)(+H,h)	be rather good
vii.	chắc	(+H,h)	certain
viii.	<sup>0</sup> chẳng <sup>1</sup> chắc	(+H,I)(+H,h)	more or less certain

Note that there are cases where a base can be reduplicated via FR as well as TETU. The difference in these cases would be the reduplicative forms' formal register. For example, although both đèm đẹp and đẹp đẹp are possible reduplications of đẹp, the former is more formal than the latter. This difference can also be understood in terms of a colloquial versus literary distinction where the FR form is the more colloquial of the two.

*Reduplication via Melodic Overwriting I: iec-reduplication:* One of two productive forms of reduplication in which fixed segments appear in the reduplicant is iec-Red (henceforth,

$\text{RED}_{\text{iec}}$ ). This suffixing process adds a pejorative meaning to base. 0 and 0 provides a summary of the generalizations on the form and distribution of this process with illustrative examples in 0.

(10) Examples of iec-Red forms

a.	<sup>1</sup> bạn <sup>1</sup> biệc	(-H, <b>h</b> )(-H, <b>h</b> )	friends (+pej)
b.	<sup>1</sup> ngủ <sup>1</sup> ngiếc	(+H, <b>lh</b> )(+H, <b>h</b> )	to sleep (+pej)
c.	<sup>0</sup> uông <sup>1</sup> thuốc <sup>0</sup> uông <sup>1</sup> thiếc	(+H, <b>h</b> )(+H, <b>h</b> )(+H, <b>h</b> )(+H, <b>h</b> )	to take medicine (+pej)
d.	* <sup>1</sup> liếc <sup>1</sup> liếc	(+H, <b>h</b> )(+H, <b>h</b> )	to scam (+pej) <sup>11</sup>
e.	* <sup>1</sup> tiếc <sup>1</sup> tiếc	(+H, <b>h</b> )(+H, <b>h</b> )	to regret (+pej) <sup>12</sup>

(11) Generalizations on the form of iec-Red

- a.  $\text{RED}_{\text{iec}}$  can only be attached at the right edge of BASE.
- b.  $\text{RED}_{\text{iec}}$  is identical to BASE except the last rhyme of  $\text{RED}_{\text{iec}}$  must be iec with tone (+H, **h**). In a number of Southern and Central dialects, the register of the tone that is associated with the rhyme replacement matches the register of the corresponding rhyme in BASE.
- c. The stress pattern for disyllabics formed by this process is 11; the pattern for quadrasyllabics is 0101.
- d. This process cannot operate on BASE with rhyme = iec with tone ( $\alpha$ H, **h**). In these cases, the form simply does not exist, nor is there any regular alternative form. The meaning would be conveyed either paraphrastically or by the use of a replacement lexical form (see the starred examples in 0). This is in contrast to analogous cases in Abkhaz (m-reduplication; see Bruening (1996)) and Telegu (p-reduplication; see Alderete et. al. (1997)) where repetition is avoided by the use of a replacement allomorphic form. For example, in Abkhaz, for bases which have *m*- in the onset, an m-reduplication process is replaced by a p-reduplication process to avoid identical copying.

(12) Generalizations on the distribution of iec-Red

- a. The process can occur on a base of any lexical grammatical category (noun, verb, adjective, etc.) of one or two syllables. In the case of disyllabic verbs, it can only apply to non-simple verbs (verb-verb or verb-object compounds) or borrowed disyllabic verbs.
- b. There may be additional material intervening between BASE and RED of a reduplicative form that was produced by iec-Red. A discussion of the constraints on what kinds of syntactic structures can be inserted is beyond the scope of this paper; however, a number of illustrative examples are provided in 0.
- c. In general, iec-Red can only apply to bases with a maximum of two syllables.

In general, in a significant number of reduplicative forms in Vietnamese BASE and RED may be separated from each other and syntactic material can be inserted. The only process out of the four considered which creates forms that have this effect is iec-Red.

<sup>11</sup> There is no replacement form for this; the meaning would have to be expressed paraphrastically.

<sup>12</sup> <sup>1</sup>tiêng <sup>1</sup>tiec would be a replacement form.

Other lexicalized reduplicative forms have this property as well.<sup>13</sup> Below, I account for the fact that iec-Red forms exhibit these effects while a-Red ones do not.

(13) Examples of separability of iec-Red forms

- a. Con chẳng chịu học hành gì cả cứ ban với biếc suốt ngày.  
you NEG try study what all just friends and RED<sub>iec</sub> all day  
“You never study or do anything, you’re just with your friends all day.”
- b. Tại sao con cứ khóc hoài vậy không ngủ không ngủ bao giờ.  
why you just cry all the time Q NEG sleep NEG RED<sub>iec</sub> ever  
“Why are you crying all the time? You’re never sleeping or anything!”
- c. Anh ấy chẳng may áo mà cũng chẳng may iếc gì cả.  
he that NEG sew shirt and also NEG RED<sub>iec</sub> what all.  
“He doesn’t sew shirts or anything at all.”

*Reduplication via Melodic Overwriting II: a-reduplication:* Another regular process of Melodic Overwriting in Vietnamese is the replacement of a coda rhyme by *-a:* in RED in the reduplication of some disyllabic verbs. I refer to the reduplicant in this process as RED<sub>a</sub>. 0 presents a summary of the generalizations about the forms created by this process. Examples in 0 are representative of forms produced by this process.

(14) Example of a-Red forms

- a. rộn ràng → <sup>0</sup>rộn <sup>2</sup>rà <sup>0</sup>rộn <sup>1</sup>ràng = to be tumultuous (+emph)  
(-H,h)(-H,l)(-H,h)(-H,l)
- b. may mặc → \*<sup>0</sup>may <sup>2</sup>mà <sup>0</sup>may <sup>1</sup>mặc = to sew and to wear  
(+H,l)(-H,l)(+H,l)(-H,h)
- c. ăn cơm → \* ăn cơm ca ăn cơm = to have a meal<sup>14</sup>  
(+H,l)(+H,l)(+H,l)(+H,l)

(15) Generalizations on the form of a-Red

- a. RED<sub>a</sub> can only attach to the left of BASE.
- b. All of BASE is copied exactly (segments as well as tones), except the rhyme of the second syllable in RED has the following form:
  - i. vowel: always *a*<sup>15</sup> and no coda
  - ii. tone: register is the same as the register of the corresponding tone in the base; the contour of the tone must be **I**.
- c. The stress pattern in the quadrasyllabic output is: 0201

The process can only occur on bases that are simple verbs (i.e., verbs that are not a compound of two verbs or an incorporation of an object and a verb) that are of two syllables. Thus, a-Red targets much more specific bases than iec-Red. Examples 0 show

<sup>13</sup> Compounds are also analogous to these i%oc-Red and lexicalized forms in that, in addition to the stress properties mentioned above, they also have this separability effect.

<sup>14</sup> Literally, ‘eat’ + ‘rice’. Can also have the meaning ‘to eat rice.’

<sup>15</sup> Here, we see an Emergence of the Unmarked effect within a Melodic Overwriting scheme. Also, there are a number of cases of ρ: (back, middle, unrounded) being an acceptable rhyme, but there only a very limited number of these cases (~3-4). In any case, ρ: is fairly unmarked so an extension on the Emergence of the Unmarked argument to vowels could account for this.

that a-Red is not possible on serial or compound verbs and verb-object incorporations. The process cannot be applied to forms in which the second syllable has rhyme *-a*. As in the similar exceptional case for iec-Red, the meaning of the avoided form is expressed either paraphrastically or is replaced by another lexical form with the same meaning. This is seen in examples 0. Forms produced by a-Red do not exhibit any separability effects as seen in compounds, iec-Red forms, and other polysyllabic forms.

(16) Examples of a-Red in which certain outputs are prohibited

- a. \*lê la lê la (lê la = to wander drunkenly)
- b. lê lê la la = to wander drunkenly (+emph, +crit)

### 3.2 Non-productive (Lexicalized) Forms

One of the main empirical observations that I make in this paper is that many of the forms that traditionally have been thought to be the result of reduplication processes are actually lexicalized forms. By “lexicalized,” I am referring to the fact that a form has an arbitrary form, an arbitrary meaning, or both, and that it is not actively produced by grammatical processes (phonological, morphological, etc.). Thus, the motivation to refer to these cases as patterns may have been because many of the forms do indeed pattern together in that there are distinct sets of cases where only the onset is reduplicated, only the onset and the tone is reduplicated, etc. However, the strongest evidence for the idea that these forms are lexicalized is that they are all non-productive in their phonological form and/or their meaning. The choice of whether the onset, coda, tone, or some combination thereof is preserved is not predictable from the form of (what would be) the ‘base.’

For cases in which a form exhibits reduplicative phonology in some way, i.e., some salient subpart of the constituent morphemes resemble each other, it may simply be that the form was produced by some sort of reduplicative process in the past but has changed over time both in form and in meaning, and now behaves just like any other polysyllabic word. In cases where neither of the morphemes in the form are free morphemes, it seems even clearer that the form was not generated by the grammar but rather stored in memory as a single lexical entry.

These lexicalized forms exhibit many of the same properties, listed in 0, as compounds do, and hence are similar to iec-Red forms.

(17) Shared properties of lexicalized reduplicatives and compounds

- a. Stress pattern is 11.
- b. Exhibit separability effects.

Note that the examples listed below exhibit nearly every possible type of partial reduplication given the assumption that the syllable is the minimal unit of analysis. (Ngô Thanh Nhàn (1984))

#### Examples of different types of lexicalized reduplicative forms

(18) Onset replacement:

- a. rõi (to entangle) → bối rõi (to embarrass)
- b. lanh (clever, quick) → lanh chanh (hasty, hurried)

- (19) Rhyme and tone replacement:
  - a. rộn (busy) → rộn ràng (happily busy)
  - b. mò (to grope) → lò mò (to grope for a long time)
- (20) Onset and tone replacement:
  - a. chơi (to play) → chơi bời (to play mischievously)
  - b. tấp (to rub up against) → tấp nập (to be crowded)
- (21) Rhyme replacement:
  - a. cắm (to plant into) → cắm cự (to concentrate)
  - b. tỉnh (refreshed) → tỉnh táo (awake, not sleepy)
- (22) Nuclear vowel replacement:
  - a. mập (fat) → mập mạp (fat (only concerning children))
  - b. rao (to advertise) → rêu rao (to be obnoxiously loud)
- (23) Coda replacement:
  - a. xanh (blue) → xanh xao (pale)
  - b. chắc (solid) → chắc chắn (sure, certain)
- (24) Cases where BASE and RED are not discernible:
  - a. lững thững = to walk slowly and aimlessly
  - b. lay láy = to sparkle

### 3.3 Summary

The similarities between the four processes are listed below. Differences between the individual processes are summarized in Table 1.

- BASE is always preserved (segmentally and tonally) in the output.
- The number of syllables in RED is equal to that in BASE.
- The forms all respect the tone-coda co-occurrence restriction.
- There can be no double reduplications in a productive sense.

## 4 Analysis

I first show that a number of highly-ranked constraints are active in all four processes (including TETU as partially analyzed in Agbayani (1997)). This accounts for the similarities. Next, I show that two constraints are active in all four processes, but are ranked differently depending on the process. For each process, I present arguments for the relative ranking of these constraints. I do this by first positing a prosodic structure for the outputted reduplicative form, and then illustrating that the properties mentioned above follow from these structures by computing a number of example tables based on the constraints proposed.

I am assuming that RED is phonetically empty in the input and is filled in via correspondence relations which attempt to map as many similar segments as possible from

the base without the insertion of additional material. I posit prosodic structure for candidates provided by GEN but not for the underlying representations. EVAL determines the optimal candidate based on segmental, tonal and prosodic content with respect to the constraints from CON.

The analysis naturally follows from the observation that the surface properties of the different processes vary according to their prosodic structures. For FR and TETU processes, since RED is an affix and is only one syllable long, under the assumption that the Prosodic Hierarchy is strictly layered, the prosodic structures of the output forms follow from alignment and foot-binarity constraints. In these two processes, the structure of the output is a prosodic word with one foot consisting of two syllables as shown in 0. Similarly, for a-Red, RED is an affix with two syllables, and so the output structure, a prosodic word with two disyllabic feet, is predicted from the same alignment and foot-binarity constraints. In these three processes, properties such as differential stress patterns (i.e., 01 in TFR and TETU, 0201 in a-Red) within a single word and the non-separability of the constituent morphemes follow from the proposed prosodic structure.

In the case of iec-Red, non-differential stress patterns (i.e., 11 and 0101 for mono- and disyllabic bases, respectively) and the syntactic separability of the base and the reduplicant support the hypothesis that RED is itself a prosodic word (structure given in 0).

I use an Optimality Theoretic approach via Correspondence Theory because it can provide a natural account of the properties of all four processes, thereby giving a unified analysis of a large portion of the (productive) Vietnamese reduplication system.

#### 4.1 Highly-ranked (Undominated) Constraints

I posit the constraints listed below to be active and highly-ranked in all four processes. In fact, these constraints are undominated except for the fact that there is one process-specific constraint (alignment of  $\text{RED}_{\text{iec}}$ ) which is ranked above these constraints in iec-Red. These constraints account for the prosodic structures in the output forms, whereas the relatively ranked constraints posited in the following sections account for the segmental and tonal content of the forms.

##### (25) Alignment constraints

**ALIGN:** Abbreviation for the following three alignment constraints which determine the position of RED with respect to BASE and which level RED combines with BASE: (McCarthy and Prince (1993))

- a. **ALIGN-L (RED,  $\omega$ ):** For every RED affix, there is a  $\omega$  such that the left edge of RED coincides with the left edge of  $\omega$ .
- b. **ALIGN-R ( $\omega$ , Stem):** For every  $\omega$  there is a stem (root + affix) (McCarthy and Prince 1993, section 4) such that the right edge of  $\omega$  coincides with the right edge of stem. This forces the constituents in FR, TETU, and a-Red to be within the same stem and therefore not separable by anything that is larger than a root, such as a stem or a morphological word, in the Morphological Hierarchy.<sup>16</sup>

<sup>16</sup> Here, I assume the Morphological Hierarchy to be the traditional one as assumed in (McCarthy and Prince (1994a)):

Morphological Word → Stem; Stem → Stem Affix; Stem → Root

c. **ALIGN-L ( $\omega$ , H(Ft)):** For every  $\omega$  there is a head foot H(Ft) such that the left edge of  $\omega$  coincides with the left edge of H(Ft).

ALIGN-L (RED,  $\omega$ ) prevents cases like *\*léch kékch léch ka* where RED appears on the wrong side of the  $\omega$ . ALIGN-L ( $\omega$ , H(Ft)) determines the placement of the head foot which accounts for the fact that primary stress falls on the first foot and not the second one in a Red output forms.

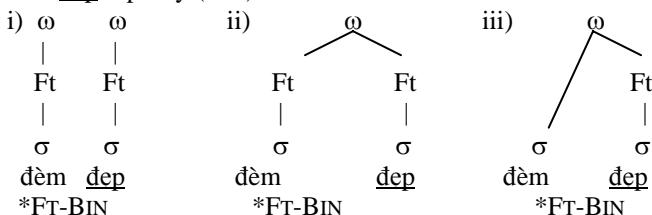
(26) Foot Binarity

**FT-BIN ( $\sigma$ ):** All feet are binary under syllabic analysis.

ALIGN-R ( $\omega$ , Stem) and FT-BIN prevents cases like 0 where the output has a different prosodic structure, which would result in inaccurate predictions for the various properties of the output. Ft-Bin rules out all of the structures in 0 except b),iv) which is ruled out by ALIGN-R ( $\omega$ , Stem). The descriptive facts about differences in the prosodic structure of the different forms are accounted for by alignment and binarity constraints.

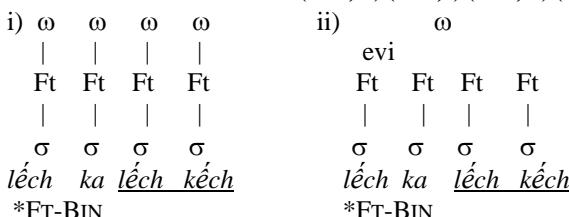
(27) Ill-formed prosodic structures; constraint(s) violated

a.  $\text{dèm } \underline{\text{dep}} = \text{pretty } (+\text{att})$

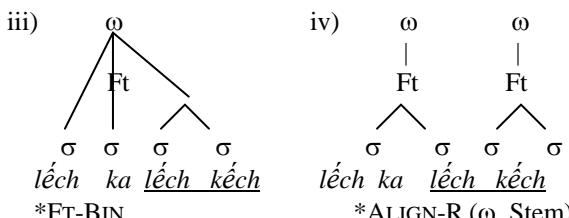


\*ALIGN-R ( $\omega$ , Stem)

b.  $\underline{\text{léch}} \text{ ka } \underline{\text{léch kékch}}$  (+H,h)(+H,l)(+H,h)(+H,h) to clatter (+emph)



\*ALIGN-R ( $\omega$ , Stem)



(28) Input-Output Faithfulness

**IO-FAITH:** The base in the input must not be altered in the output. Technically, this is actually three constraints in one: IO-DEP, IO-MAX, and IO-IDENT, which respectively prohibit insertion, deletion, and non-identity of corresponding segments in the output. (McCarthy and Prince (1993))

This constraint prevents BASE from changing. Therefore, if there is ever a mismatch between BASE and RED, it is not because BASE changed, but rather it is because of partial copying or the insertion of fixed and/or unmarked material. This is why examples such as b) below are good rather than c), which would otherwise have fewer marked features overall.

(29) Example: BASE is always preserved

- a. *đẹp* (-H,**h**) pretty
- b. *đèm đẹp* (-H,I)(-H,**h**) pretty (+att)
- c. *\*đèm đèm* (-H,I)(-H,I)

(30) Phonotactic constraint: Tone-coda co-occurrence restriction

**PHONO<sub>A</sub>**: Highly ranked constraints which induce the tone-coda co-occurrence restriction. Here, I adopt Agbayani (1997)'s analysis which posits a licensing condition on [stiff]. He uses the following two undominated constraints based on Itô, Mester, and Padgett (1995):

- a. **LICENSE [stiff]**: The feature [stiff vocal cords] ([stiff]) must be licensed.
- b. **Obs]<sub>σ</sub> ⊃ [stiff]**: The specification [Obstruent] in coda  $\Rightarrow$  [stiff].

Note that since these constraints are undominated, inputs that violate this constraint simply do not have a phonetic realization. I suggest that the unacceptability of forms such as the ones in 0 is due to blocking effects. Detailed discussion and analysis of these facts are beyond the scope of this paper and are pending future research.

(31) Examples of avoidance of identical structure

- a. *lê la lê la*
- b. *\*lê la lê la*

For the case of iec-Red, the following constraint is ranked above the other alignment constraints and accounts for two facts: 1) Even though the output is a disyllabic form, it is two separate words and thus exhibits the stress patterns and separability properties characteristic of two separate words, and 2) RED appears on the right side of BASE.

**ALIGN-R (RED<sub>iec</sub>,  $\phi$ )**: For every RED<sub>iec</sub>, there is a phonological phrase  $\phi$  such that the right edge of RED<sub>iec</sub> coincides with the right edge of  $\phi$ . This constraint ranks above the other alignment constraints. This accounts for RED<sub>iec</sub> appearing on the right side of BASE and the fact that RED<sub>iec</sub> and BASE are separate prosodic words and therefore have separate stress.

RED<sub>iec</sub> exhibits a number of properties which supports its status as a word (it can be separated from the base like most compounds can in Vietnamese; it receives its own stress, etc.).<sup>17</sup> For more detailed discussion of the syntactic separability of compounds in Vietnamese and the status of RED<sub>iec</sub>, see Vu (1998).

<sup>17</sup> One possibility is that RED<sub>iec</sub> is simply a phonetically underspecified word which becomes specified in the context of another word. In cases where it is separated from the base, one possibility is to say that it raises covertly at LF to check features with the base and in doing so is filled in with phonetic material via correspondence relations.

The constraints ALIGN, IO-FAITH, PHONO<sub>A</sub>, and FT-BIN ( $\sigma$ ) are undominated except for iec-Red in which case ALIGN-R (RED<sub>iec</sub>,  $\phi$ ) out-ranks these constraints.<sup>18</sup> Note that FT-BIN is always violated in iéc-Red for cases where the base is monosyllabic since the output is dysllabic but its prosodic structure is that of two separate words, each with one syllable.

Summary of highly ranked constraints:

Align-R (red<sub>iec</sub>,  $\phi$ ) >> {Align, Ft-Bin, Phono<sub>a</sub>, IO-Faith} >> {relatively-ranked constraints}

#### 4.2 Relatively-ranked (conflicting) constraints

Following the work of Itô and Mester (1995), I employ the concept of a cophonology to model what appears to be conflicting rankings of constraints within a single language. The reduplication processes examined belong to one of two cophonologies. An example is the difference in the status of different kinds of borrowed words in a language with respect to a core-periphery distinction (native words and older borrowed words being considered ‘core’ while recently borrowed words are more at the periphery of a language’s vocabulary) as discussed in Itô and Mester (1995).

(32) Relatively ranked constraint: Base-Reduplicant Identity

**BR-IDENT:** Corresponding elements in the base and reduplicant must match in segmental and tonal features (register and contour values).

(33) Relatively ranked constraints: Dispreference of marked structure

a. **\*MARKED:** No marked features can be associated with segments. I use this as an abbreviation for the following set of constraints against marked features appearing on segments (Agbayani (1997)):

b. Tonal markedness constraint:

**\*[+stiff]<sub>μ</sub>:** No [+stiff] features can be associated with moraic segments. This predicts that the unmarked tonal contour is l.

c. Vocalic markedness constraints:

**\*V-PL/LAB, \*V-PL/DOR:** No [labial, dorsal, coronal, radical] specification for V(owel)-Place, respectively.

I propose that each of the four reduplication processes in Vietnamese belongs to one of these two cophonologies<sup>19</sup>:

(34) Cophonologies in reduplication

a. BR-Ident >> \*Marked

b. \*Marked >> BR-Ident

<sup>18</sup> Another formulation of these facts is to say that the ranking: ALIGN-R (RED<sub>iec</sub>,  $\phi$ ) >> ALIGN, IO-FAITH, PHONO<sub>A</sub>, FT-BIN ( $\sigma$ ) >> {relatively ranked constraints} holds for all four processes.

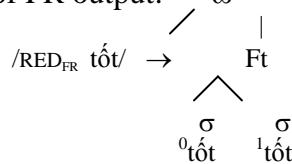
<sup>19</sup> Note that in OT terms, different cophonologies can simply be different rankings of a common set of constraints.

### 4.3 Example tables

Below, I show which of the two cophonologies each process belongs to by computing example tables on data presented in previous sections. For the sake of expositional clarity, the bulk of the argument will be only for which cophonology the process belongs to and hence will not include every constraint mentioned in the previous section. In these tables, only the candidates relevant to arguing the ordering of BR-Ident and \*Marked are considered. These constraints determine the segmental and tonal content of RED but does not bear on the prosodic structure of RED in the output.

**Full reduplication:** In all of these cases, the output appears to be a word with a single disyllabic foot. I refer to the reduplicant in FR as  $\text{RED}_{\text{FR}}$ . In all of these cases, since the output is a single prosodic word there can be no syntactic material inserted in between BASE and RED since that would violate the Prosodic Hierarchy by having a word dominated by a foot. The stress pattern 01 for these forms follows from the prosodic structure since there is only one stress assigned to each foot. The prosodic structure of the output is given in 0. I propose that FR is in the cophonology in which base-reduplicant identity is preferred over not having marked structure.

(35) Prosodic structure of FR output:



(36) Example: FR output form <sup>0</sup>ít <sup>1</sup>ít= ‘small in amount’ (+att)

*MARKED Constraints (unranked)					
	/RED <sub>FR</sub> + ít/ (+H,h)	BR- IDENT	*[+stiff] <sub>μ</sub>	*V-PL/LAB	*V-PL/DOR
a.	in <u>ít</u> (+H,I)(+H,h)	*!*( (n, I)	*( (h)	✓	✓
b.	ít <u>ít</u> (-H,h)(+H,h)	*!(-H)	**( (h, h)	✓	✓
c.	ít <u>ít</u> → (+H,h)(+H,h)	✓	**( (h, h)	✓	✓

Candidates a) and b) in (36) are ruled out because they all have at least one violation of BR-Ident either because of differences in segmental or tonal material. The tension between identity and the dispreference of marked material is manifested in the fact that in order to fully satisfy BR-IDENT, there has to be at least two violations of \*MARKED constraints, namely the **h** in the tones. This example shows that BR-IDENT is ranked above \*MARKED. Candidates with a vowel change in the reduplicant are not considered since any such change will result in a violation of BR-IDENT whether they result in a violation of \*MARKED or not.

(37) Example: FR output form  ${}^0\text{tót} {}^1\text{tót}$  = ‘good, useful’ (+att)

Undominated constraints (unranked)

	/RED <sub>FR</sub> +tót/ (+H,h)	IO- FAITH	PHONO <sub>A</sub>	FT- BIN	ALIGN	BR- IDENT	*MARKED
a.	$\begin{array}{c} \omega \\   \\ \text{Ft} \\ 2 \\ \sigma \quad \sigma \\ {}^0\text{tót} {}^1\text{tót} \end{array}$	${}^*!{}^*$ (n,tone 1)		✓	✓	✓	**** (o,o)
b.	$\begin{array}{c} \omega \\   \\ \text{Ft} \\ 2 \\ \sigma \quad \sigma \\ {}^0\text{tót} {}^1\text{tót} \end{array}$		✓	${}^*!$ (tone 1)	✓	✓	* (tone 1) ***** (o,o, tone 2)
c.	$\begin{array}{c} \omega \\ 2 \\ \text{Ft} \quad \text{Ft} \\   \quad   \\ \sigma \quad \sigma \\ {}^1\text{tót} {}^1\text{tót} \end{array}$	✓	✓	${}^*!$	✓	✓	***** (o,o, tone 2, tone 2)
d.	$\begin{array}{c} \omega \\   \\ \text{Ft} \\ 2 \\ \sigma \quad \sigma \\ {}^1\text{tót} {}^0\text{tót} \end{array}$	✓	✓	✓	${}^*!$ (Align -L (RED, $\omega$ ))	✓	***** (o,o, tone 2, tone 2)
e.	$\begin{array}{c} \omega \\   \\ \text{Ft} \\ 2 \\ \sigma \quad \sigma \\ {}^0\text{tát} {}^1\text{tót} \end{array}$	✓	✓	✓	✓	*! (V)	**** (o, tone 2, tone 2)
f.	$\begin{array}{c} \omega \\   \\ \text{Ft} \\ 2 \\ \sigma \quad \sigma \\ {}^0\text{tót} {}^1\text{tót} \end{array}$	✓	✓	✓	✓	✓	***** (o,o, tone 2, tone 2)

Candidates a) – d) are ruled out by the undominated constraints. e) (here, V = i, e, ε, a, i.e., the vowels with none of the marked V-Place features) is ruled out by a fatal violation of BR-IDENT.

**Reduplication due to Emergence of Unmarked effects:** Agbayani (1997) attributes base-reduplicant mismatch phenomena in Vietnamese reduplication to the interaction of TETU effects (McCarthy and Prince (1994b)) with the Normal Application of Phonological Constraints scheme (McCarthy and Prince (1995)). Specifically, TETU effects refer to the important empirical observation that unmarked effects which would otherwise be invisible in most parts of a grammar arise in unexpected ways in certain domains of a language

(e.g., reduplication). In correspondence-theoretic terms, these effects are accounted for as unmarked constraints out-ranking base-reduplicant identity constraints as shown in (38).

(38) The Emergence of the Unmarked ranking schema:  
 I-O FAITHFULNESS >> PHONO-CONSTRAINT<sub>A</sub> >> B-R IDENTITY

In the above schema, Phono-Constraint<sub>A</sub> are the hidden phonological constraints on unmarkedness. In the case of Vietnamese, for tones, the feature [+stiff] is more marked than [-stiff]; for vowels and codas, [+labial] and [+dorsal] are more marked than [+coronal] and [+radical]. Therefore, the restriction against the marked features may force violations of base-reduplicant identity. This notion can also be used to predict the various vowel and coda replacement in a number of reduplication processes. The result of this mismatch then is partial reduplication of BASE.

This scheme is itself dominated by even more highly ranked constraints on syllable structure, the segment-tone co-occurrence restriction mentioned above, and a faithfulness constraint of BASE in the output to BASE in the input. In my analysis, I show that these constraints are in fact undominated for all reduplication process in Vietnamese. This scheme (given below) is well-attested and referred to in the literature as the Normal application of Phonological Constraints.<sup>20</sup> (McCarthy and Prince (1995)) It is characterized as the following ranking schema:

(39) Normal Application of Phonological Constraints ranking schema:  
 PHONO-CONSTRAINT<sub>B</sub> >> I-O FAITHFULNESS >> B-R IDENTITY

The effect of this schema is that it induces similar effects as in (39), in that it contributes to the mismatch of the base and stem, but only to the extent that the mismatch does not violate a number of highly-ranked phonological constraints which are active throughout the grammar. The schemata (38) and (39) combine to produce the ranking (40). See Agbayani (1997) for the details of this analysis.

(40) Interaction of the two schemata:  
 PHONO-CONSTRAINT<sub>B</sub> >> I-O FAITHFULNESS >> PHONO-CONSTRAINT<sub>A</sub> >> B-R IDENTITY

Although Agbayani (1997)'s analysis correctly predicts the form of some reduplications in Vietnamese, its main limitation is that the data set itself is only a small subset of the overall system of reduplication in Vietnamese. In fact, there are examples in his data set which actually belong to other reduplication processes which are overlooked in his analysis. For example, the form *dẹp dẹp* is ruled out in his analysis but is a possible form in Vietnamese. The point is that although this form is not produced by TETU, it is produced by FR. The main difference is in register in which the forms are used.

Furthermore, his data set is restricted not only to single morpheme reduplications, but only one *type* of such reduplication; the other types such as FR and iec-Red are not accounted for. Other parameters such as stress and syntactic separability are also not

---

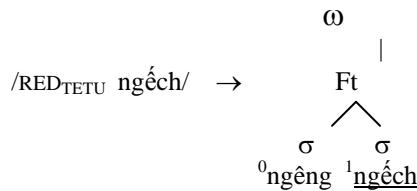
<sup>20</sup> As opposed to the Underapplication or Overapplication of Phonological constraints in which case faithfulness constraints are ranked below identity constraints or above normally highly-ranked constraints, respectively. (McCarthy and Prince (1995))

analyzed. Below, I will show that by a simple extension of some of his arguments and by reconsidering some of the constraints he posits in a more general context, I can account for additional processes as well as parameters along which these processes vary. In doing this, problematic cases such as the large number of cases in which marked structure *does* appear in RED can be explained.

I propose that TETU is prosodically very similar to FR in that it exhibits the same stress patterns (01) and does not have separability effects. However, it crucially differs from FR in that the \*MARKED constraints are ranked above BR-IDENT. Adopting Agbayani's analysis, this forces base-reduplicant mismatch. The mismatch must be minimal in order for the candidate to be successful, i.e., it must maximally satisfy BR-IDENT. I call the reduplicant RED<sub>TETU</sub> to distinguish the two different types of reduplicants.

Like FR, TETU reduplication is occurring at the syllabic level rather than the foot or prosodic word level. This gives TETU the same properties as FR with respect to the stress patterns seen in the output and its syntactic inseparability.

(41) Prosodic structures involved in TETU:



I propose that FR is in the cophonology in (42b) in which BR-IDENT is outranked by \*MARKED.

(42) Example: TETU output form <sup>0</sup>thăng <sup>1</sup>thăng= 'straight' (+att)

Input: /RED<sub>TETU</sub>+ thăng / (tone is (+H,Ih))

\*Marked Constraints (unranked)

	/RED <sub>TETU</sub> + thăng/ (+H,Ih)	*[+stiff] <sub>μ</sub>	*V-PL/LAB	*V-PL/DOR	BR-IDENT
a.	<sup>0</sup> thăng <sup>1</sup> <u>thăng</u> (-H,Ih)(-H,Ih)	**!	✓	✓	✓
b.	<sup>0</sup> thăng <sup>1</sup> <u>thăng</u> → (-H,I) (-H,Ih)	*	✓	✓	*!

Candidate a) in (42) has two violations of \*MARKED constraints which proves to be fatal. b) on the other hand, only has one violation of \*MARKED and is the winning candidate even though it has one more BR-IDENT violation. Again, candidates with vowel changes in RED are not considered because any such change would result in additional violations of BR-IDENT and possibly \*MARKED as well. This shows that the \*MARKED constraints are indeed ranked above BR-IDENT in this process.

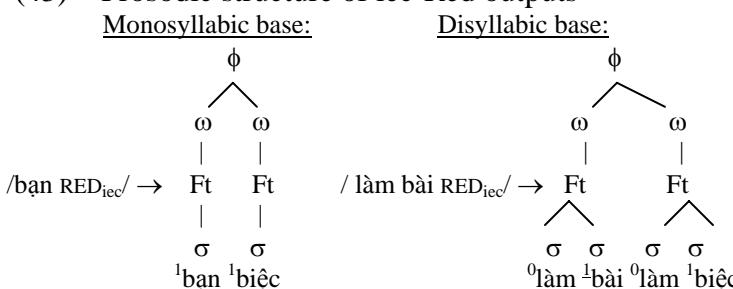
A table similar to (37) can be computed for <sup>0</sup>ngéng <sup>1</sup>ngéch, except in this case, \*MARKED would be reranked above BR-IDENT

**Reduplication via Melodic Overwriting I: iec-reduplication:** iec-Red is probably the most productive of all the reduplication patterns in Vietnamese as it can regularly apply to any lexical word. I show below that the process belongs to the cophonology in which

\*MARKED constraints are outranked by BR-IDENT. This ranking can be argued for using cases where the base has two syllables.

As mentioned above, this process involves the placement of an additional prosodic word (RED<sub>iec</sub>) rather than the appending of phonological material within a prosodic word. Thus, continuing to assume that each foot has one stress assigned, the stress pattern 0101 for disyllabic bases follows from the prosodic structure of the reduplicative form. The syntactic separability of the form is also explained by the fact that it is possible to insert a word in between word (within a phonological phrase  $\phi$ ) without violating the integrity of the Prosodic Hierarchy. Note that this structure also accurately predicts the inseparability of the morphemes in *làm bài* and in *làm biệc* since each of these pairs of syllables is within the same prosodic word. Similar accurate predictions follow from prosodic structure for cases with monosyllabic case as well.

(43) Prosodic structure of iec-Red outputs



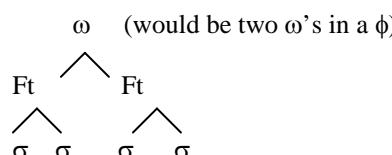
(44) Example: i%oc-Red output form  $^0\text{bún} \quad ^1\text{chả} \quad ^0\text{bún} \quad ^1\text{chiếc}$  = ‘noodle and grilled pork dish’ (+pej)

\*MARKED Constraints (unranked)

	/ bún chả + RED <sub>iec</sub> /	BR-IDENT	*[+stiff] <sub>μ</sub>	*V- PL/LAB	*V- PL/DOR
a.	$^0\text{bún} \quad ^1\text{chả} \quad ^0\text{bún} \quad ^1\text{chiếc}$	****!(iγ, k, tone1,2)	** (tone2 x2)	** (u,u)	*** (u,u,iγ)
b. →	$^0\text{bún} \quad ^1\text{chả} \quad ^0\text{bún} \quad ^1\text{chiếc}$	*** (iγ,c,t one2)	*** (tone 2 x3)	** (u,u)	*** (u,u,iγ)

Candidate a) has four violations of BR-IDENT which is fatal. Candidate b) has one more violation of \*MARKED, but it has one fewer violation of BR-Ident and thus is the winning candidate.

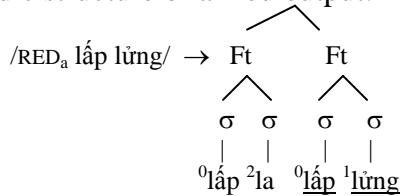
For space considerations a full table for iec-Red will not be presented. The ranking would be: ALIGN-R (RED<sub>iec</sub>,  $\phi$ ) >> {IO-FAITH, PHONO<sub>A</sub>, FT-BIN, ALIGN} >> BR-IDENT >> \*MARKED. The main difference between such a table and a table like 0 would be that the first alignment constraint would prevent structures such as:



*Reduplication via Melodic Overwriting II: a-reduplication:* Although a-Red is productive in the sense that given a form, it is predictable whether this process can operate on the form or not, it is limited to disyllabic bases that are simple verbs. a-Red is in the same cophonology as iec-Red (and hence FR as well).

As was the case above, the stress pattern 0201 and lack of syntactic separability of a-Red forms follow directly from the prosodic structure I propose for the reduplicative forms (given in (45)). Since each foot is assigned one stress, we have the alternation oxoy where  $x,y \neq 0$ . That there is a secondary/primary stress distinction, and that there cannot be a word inserted in between BASE and RED, give evidence to the hypothesis that copying is indeed happening at the foot level and that the entire utterance is dominated by one prosodic word. That BASE and RED cannot be separated by another word is because a word cannot be inserted inside another word.

(45) Prosodic structure of a-Red output:  $\omega$



(46) Example table for a-Red output form  ${}^0$ láp  ${}^2$ la  ${}^0$ láp  ${}^1$ lúng = ‘to bob up and down/equivocate’(+emph)<sup>21</sup> \*MARKED Constraints (unranked)

	/RED <sub>A</sub> + láp lúng/	BR-IDENT	*[+stiff] <sub>u</sub>	*V-PL/LAB	*V-PL/DOR
a.	${}^0$ láp ${}^2$ la ${}^0$ láp ${}^1$ lúng	***! (a,a:,tone1)	*** (tone 2x3)	** (o,o)	** (o,o)
b.	${}^0$ láp ${}^2$ la ${}^0$ láp ${}^1$ lúng →	** (a:, tone 2)	*** (tone 2 x3)	*** (o,o,o)	*** (o,o,o)

In (47), candidate a) has three violations of BR-IDENT, which proves to be fatal since the winning candidate only has two violations of this constraint even though a) has two fewer \*MARKED violations. This shows that BR-IDENT is ranked above the \*MARKED constraints. The table below illustrates the prosodic structural constraints for the same input.

<sup>21</sup> Thanks to Cao Xuân Hạo for pointing out this example.

(47) Example: a-Red output form  ${}^0láp {}^2la {}^0láp {}^1lúng$  = ‘to bob up and down, to equivocate’ (+emph) (full table)

	/RED <sub>A</sub> + láp lúng/	IO-FAITH	PHONO <sub>A</sub>	FT-BIN	ALIGN	BR-IDENT	*MARKED
a.	$\begin{array}{c} \omega \\ ei \\ Ft \quad Ft \\ 2 \quad 2 \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^0láp \ {}^2la \ {}^0láp \ {}^1la \end{array}$		**! (a:, tone 1)	✓	✓	✓	**** (ꝝ,ꝝ, tone 2, tone 2)
b.	$\begin{array}{c} \omega \\ ei \\ Ft \quad Ft \\ 2 \quad 2 \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^0láp \ {}^2la \ {}^0láp \ {}^1lúng \end{array}$		✓	*! (tone 1)	✓	✓	** (tone 1, tone 1, a:) *** (ꝝ,ꝝ, tone 2, tone 3)
c.	$\begin{array}{c} \omega \quad \omega \quad \omega \\   \quad   \quad   \\ Ft \quad Ft \quad Ft \\   \quad   \quad   \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^1láp \ {}^1la \ {}^0láp \ {}^1lúng \end{array}$		✓	✓	*! ALIGN- R (ω, Stem)	**! (tone 1, a:)	***** (ꝝ,ꝝ, tone 2, tone 2, tone 3)
d.	$\begin{array}{c} \omega \quad \omega \\   \\ Ft \quad Ft \\ 2 \quad 2 \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^0láp \ {}^1la \ {}^0láp \ {}^1lúng \end{array}$		✓	✓	✓	**! ALIGN- R (ω, Stem)	***** (ꝝ,ꝝ, tone 2, tone 2, tone 3)
e.	$\begin{array}{c} \omega \\ ei \\ Ft \quad Ft \\ 2 \quad 2 \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^0láp \ {}^2la \ {}^0láp \ {}^1lúng \end{array}$		✓	✓	✓	***! (a,a:, tone 1)	**** (ꝝ, tone 2, tone 2, tone 3)
f.	$\begin{array}{c} \omega \\ ei \\ Ft \quad Ft \\ 2 \quad 2 \\ \sigma \quad \sigma \quad \sigma \quad \sigma \\ {}^0láp \ {}^2lá \ {}^0láp \ {}^1lúng \end{array}$	→	✓	✓	✓	** (a:, tone 1)	***** (ꝝ,ꝝ, tone 2, tone 2, tone 3)

In (47) a) – d) show that violation of any of the first four constraints is fatal; therefore, whether they are ranked relative to one another or not (I give no argument), they are indeed undominated. e) and f) show that BR-IDENT dominates \*MARKED and hence, as much of the segmental as well as tonal material must be copied while preserving the prespecified shape of RED, in this case the fixed segment *a:* in the second syllable in RED.

## 5 Discussion and summary

In the above analysis, I have shown that by examining the prosodic structure of the reduplicative forms in Vietnamese, a number of seemingly disparate properties of these forms can be accounted for. What is particularly important is that copying at different levels in the prosodic structure can result in the output form obtaining different characteristics. For example, copying at the prosodic word level results in the syntactic separability of BASE and RED, whereas copying at any lower level will not. That reference to prosodic structure can give such a unified account supports the Prosodic Hierarchy Hypothesis. Moreover, this analysis brings to light the general importance of such structure in the theory of reduplication cross-linguistically.

I have also shown that FR, a-Red, and iec-Red belong to the cophonology in which constraints against markedness are dominated by constraints forcing an identity relation between BASE and RED. TETU is shown to be in the cophonology in which (as the name of the process implies) constraints against markedness emerge in the output forms. One of the implications of this find is the idea that it may be possible to have conflicting rankings of the same constraints within the same language. This has been shown to be possible in a somewhat different context in Itô and Mester (1995), in which they show that Faithfulness constraints may vary in its ranking depending on how well assimilated a lexical item is in a language's vocabulary. In this case, it is a matter of constraints against markedness arising in certain reduplication patterns and not in others.

## 6 Conclusions and questions

That there may be rerankings of constraints within a single language raises a number of questions for the architecture of OT as an explanatory system. For example, does allowing conflicting constraint rankings make the system too powerful? Although such an alteration to the theory may not be conceptually desirable, what alternatives are there given the assumptions made by OT? With these questions in mind, I suggest that a more productive line of research would be to explore which parts of the phonological system (and the grammar in general) are optimality theoretic and which are not.

One of the issues the data raises, but which was not discussed is that the Melodic Overwriting (MO) processes (a-Red and iec-Red) are blocked from operating on certain special forms which would normally be predicted to be valid inputs. I am referring specifically to the starred examples listed in 0 and 0, repeated here in (48).

### Blocked forms

a.	liéc	(+H,h)	to scam
b.	* <sup>1</sup> <u>liéc</u> <sup>1</sup> liéc	(+H,h)(+H,h)	
c.	tiéc	(+H,h)	to regret
d.	* <sup>1</sup> <u>tiéc</u> <sup>1</sup> tiéc	(+H,h)(+H,h)	
e.	<sup>1</sup> tiêng <sup>1</sup> tiéc	(+H,l)(+H,h)	to regret (+pej)
f.	le: la:	(+H,l)(+H,l)	to wander drunkenly
g.	*le: la: <u>le: la:</u>	(+H,l)(+H,l)(+H,l)(+H,l)	

That there is no alternative form and that the meanings (e.g., 'to scam' (+pej)) would have to be expressed paraphrastically or by the use of an alternative but

unpredictable lexical form suggests a promising topic for future study. It is not clear how such blocking phenomena such as these can be explained within the OT framework since there is no way for the system to ‘crash,’ i.e., to have the winning candidate to be  $\emptyset$ , the empty set. Although Prince and McCarthy (1993) have alluded to a solution employing the notion of a Null Parse, this approach seems unsatisfactory since the result is still an empty category (e). However, e is still something that is being generated (in this case by the phonological system), it just happens to not have any phonetic realization. The output produced by such a process is not the same as the empty set since it is very possible for an empty element to have syntactic and semantic features which would have effects in other areas of the grammar.

In these instances of blocking, it does not seem that there is an empty element being generated. First, there is no syntactic reflex that is caused by such an empty form. Second, there is often an unpredictable replacement form such as in example (69e). Although I do not provide an explanation for these observations, the fact that this sort of blocking in MO processes occurs in other languages such as Turkish (Meltem Kelepir, p.c.) warrants further investigation. Note that this sort of blocking is distinct from what happens in languages like Abkhaz (Bruening (1996)) in which the MO processes that are blocked have a regular alternative process which handles the blocked cases. Therefore, as with the problem of conflicting constraints mentioned above, this need to ‘crash’ rather than to just generate an empty element seems to be a problem that is irreconcilable within the OT system.

Finally, another point of interest is the observation that iec-Red is allowed on any lexical item except simple disyllabic verbs, in which case, a more specific rule, namely a-Red, is applied. Although this sort of pattern has been well attested and is explained by notions like the Elsewhere Condition, what is of interest is that a-Red and iec-Red do not produce the same changes in meaning on a given input. The rule a-Red gives an emphatic meaning to BASE whereas iec-Red produces a pejorative meaning.

## References

Agbayani, Brian. 1997. Phonological Constraints and Base-Reduplicant Non-Identity: The Case of Vietnamese. Ms. University of California, Irvine.

Alderete, John, et al. 1997. Reduplication and fixed segmentism. Ms., University of Massachusetts, Amherst.

Alves, Mark. 1997b. Tonal features and the development of Vietnamese tones. Ms. University of Hawaii.

Bruening, Benjamin. 1996. Abkhaz Mabkhaz: m-reduplication in Abkhaz, Weightless Syllables, and Base-Reduplicant Correspondence. In Benjamin Bruening, Yoonjung Kang, and Martha McGinnis, eds., *MIT Working Papers in Linguistic*:30. 291-329.

Burton, Strang. 1992. Reduplication and the representation of tone in Vietnamese. Ms., Brandeis University.

Cao Xuân Hạo. 1997. Trọng âm và các quan hệ ngữ pháp trong Tiếng Việt. (Stress and its relation to grammar in Vietnamese) Ms., Saigon National University.

Đoàn Thiện Thuật. 1980. *Ngữ âm Tiếng Việt*. [Vietnamese Phonetics] Hanoi: Nhà Xuất Bản Đại Học và Trung Học Chuyên Nghiệp.

Edmondson, Jerold and Nguyễn Văn Lợi. 1997. Tones and voice quality in modern northern Vietnamese: instrumental case studies." *Mon-Khmer Studies* 28.

Halle, Morris and K. N. Stevens. 1971. A note on laryngeal features. *Quarterly progress report No. 101*. 198-213. Research Laboratory of Electronics, MIT.

Han, Meiko S. 1966. Vietnamese vowels. In *Studies in the phonology of Asian languages* technical report 4. Done at Acoustics Phonetics Research Laboratory, University of Southern California.

Han, Meiko S. 1969. Vietnamese tones. In *Studies in the phonology of Asian languages* technical report 8.

Haudricourt, André. 1954. De l'origine des tons en Viêtnamien. *Journal Asiatique* 242: 69-82.

Hayes, Bruce. 1989. The prosodic hierarchy meter. In *Phonetics and Phonology: Rhythm and Meter*, Paul Kiparsky and Gilbert Youmans, eds., 1:201-260. New York: Academic Press.

Hoàng Văn Hành. 1985. *Tùi lây trong Tiếng Việt*. [Reduplicative forms in Vietnamese] Hanoi: Nhà Xuất Bản Khoa Học Xã Hội.

Hoàng Văn Hành, et. al. 1995. *Từ điển từ lây Tiếng Việt*. [Dictionary of Reduplicative Forms]. Hanoi, Vietnam.

Itô, J., A. Mester, and J. Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26:571-613.

Itô, J. and A. Mester. 1995. The core-periphery structure of the lexicon and constraints on reranking. In Beckman, Dickey, and Urbanczyck, eds., *University of Massachusetts Optimality Papers 18*. Amherst, MA.

McCarthy, John and Alan Prince. 1993b. Generalized alignment. *Yearbook of Morphology*, pp. 79-153.

McCarthy, John and Alan Prince. 1994a. An overview of prosodic morphology. Parts 1 and 2: Template form in reduplication; Template satisfaction. Talks at Utrecht Prosodic Morphology Workshop.

McCarthy, John and Alan Prince. 1994b. The emergence of the unmarked: Optimality in prosodic morphology. In Merce Gonzalez ed., *Proceedings of the North-East Linguistics Society 24*. Amherst, MA.

McCarthy, John and Alan Prince. 1995. Faithfulness and reduplicative identity. In UMOP 18.

Nespor, Marina and Irene Vogel. 1982. Prosodic domains of external sandhi rules. In *The Structure of Phonological Representations 1*, Harry van der Hulst and Norval Smith, eds. Dordrecht: Foris.

Ngô Thanh Nhàn. 1984. The Syllabeme and patterns of word formation in Vietnamese. Ph.D. dissertation, New York University.

Nguyễn Đinh Hoà. 1980. Language in Vietnamese Society. *Vietnam Culture Series I*, Asia Books, 63-108.

Nguyễn Kim Thản. 1997. *Nghiên cứu ngữ pháp Tiếng Việt*. [Studies in Vietnamese Grammar] Hanoi: Nhà Xuất Bản Giáo Dục.

Nguyễn Tài Cẩn. 1996. *Ngữ pháp Tiếng Việt*. [Vietnamese Grammar] Hanoi: Nhà Xuất Bản Đại Học Quốc Gia Hà Nội.

Noyer, Rolf. 1995. Vietnamese 'morphology' and the definition of Word. Ms., Brown University.

Phạm Hoa. 1997. The coronal-velar relationship in Vietnamese dialects. M.A. Thesis, University of Toronto.

Prince, Alan and Paul Smolensky. 1993. Optimality Theory: constraint interaction in generative grammar. Ms. Rutgers University and the University of Colorado, Boulder.

Selkirk, Elisabeth. 1978. On prosodic structure and its relation to syntactic structure. In *Nord Prosody II*, T. Fretheim, ed., 111-140. Trondheim, Norway: TAPIR.

Thompson, Laurence C. 1965. *A Vietnamese grammar*. Seattle: University of Washington Press.

Trần Mai Hương Aurelie. 1967. Stress, tones, and intonation in South Vietnamese. Ph.D. Dissertation, Australian National University.

Urbanczyk, Suzanne. 1997. Morphological templates in reduplication. In K. Kusumoto, eds., *North-Eastern Linguistics Society 27*: 369-382.

Võ Xuân Hào. 1997. Nghiên cứu về chức năng của thanh điệu Tiếng Việt. [Studies in the function of tones in Vietnamese]. Ph.D. Dissertation, Hanoi National University.

Vũ Thé Thạch. 1994. Consonant copying and tone harmony in Vietnamese reduplicatives. In Martha Ratliff and Wric Schiller, eds., *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*. Tempe: Arizona State University: 381-395.

Vu, Sonny Xuan. 1998. Doing splits at PF: Compound separability in Vietnamese and Pacoh. Ms. MIT.

Wannamacher, Mark. 1997. Aspects of tone and voice quality: an autosegmental account. Arlington: Texas. University of Texas, Arlington MA thesis.

Yip, Moira. 1992. Reduplication with fixed melodic material. *Proceedings of the North Eastern Linguistics Society 22*: 459-476.



# THE EMERGENCE OF ‘GIVE’ PASSIVES IN EAST AND SOUTHEAST ASIAN LANGUAGES

Foong Ha Yap

*Chinese University of Hong Kong*  
<foong\_ha\_yap@cuhk.edu.hk>

Shoichi Iwasaki

*University of California, Los Angeles*  
<iwasaki@humnet.ucla.edu>

## 0 Introduction

It has been noted that in some languages the same morpheme is used to express both causative and passive meaning. Several linguists have suggested that passive constructions can arise from causative constructions via the permissive and reflexive context (e.g. Keenan 1985; Haspelmath 1990; Knott 1995). Among the languages known to reflect this causative-passive link via the reflexive are Korean (Keenan 1985:262) and Manchu-Tungusic (Knott 1995:57-58). In this paper, we will examine this phenomenon in a number of East and Southeast Asian languages, including a number of Chinese dialects, some Tai and Austroasiatic languages, and colloquial Malay. In particular, we will focus on the causative-passive link involving the morpheme ‘give’.

## 1 Background

We begin with a brief review of the arguments posed for the causative > passive development via the permissive and reflexive contexts in two groups of languages: one that does not involve the ‘give’ morpheme, and one that does.

Korean shows this causative-passive development with the suffix *-I-*, which is realized as *-ki-* in (1) below. The etymology of this suffix is not known.

(1) Korean (Keenan 1985:262<sup>1</sup>)

- a. *Nuna-ka emeni-eke ai-lil an-ki-ess-ta*  
sister-NOM mother-DAT child-ACC embrace-CAUS-PAST-IND  
‘Sister had Mother embrace the child.’
- b. *Ai-ka emeni-eke caki mom-lil an-ki-ess-ta*  
child-NOM mother-DAT self body-ACC embrace-CAUS-PAST-IND  
‘The child had Mother embrace him.’
- c. *Ai-ka emeni-eke an-ki-ess-ta*  
child-NOM mother-DAT embrace-PASS-PAST-IND  
‘The child was embraced by Mother.’

<sup>1</sup> We have slightly modified the glosses for the sake of consistency throughout the paper. Specifically, we have chosen the case markers NOM, DAT and ACC in place of Keenan’s SUBJ, IO and DO, and we use the term IND(icative) in place of DECLAR(ative).

Note that both (1a) and (1b) are causative. The latter, of course, is also reflexive (hence, the term “reflexive-causative”). (1c) has passive meaning. Keenan (1985) suggests that the causative suffix *-ki-* in (1a) can come to be associated with the passive meaning in (1c) through the deletion of the reflexive direct object in (1b), and through “the reinterpretation of the subject *child* as an Experiencer rather than an Agent” (p. 262). Keenan further points out that this causative-passive development is possible because of the availability of a permissive or non-interventive ‘let’-type interpretation for the *-ki-* suffix. We summarize this causative-passive development as follows:

causative > reflexive permissive causative > passive

The Manchu-Tungusic languages provide evidence of a similar development involving the suffix *-bu-* (for Manchu) or *-v(u)-* (for the Tungusic languages), which is claimed to be etymologically derived from a verb meaning ‘give’ (e.g. Gabelentz 1861:518, cited in Haspelmath 1990:48 and I. Nedjalkov 1993:194; I. Nedjalkov 1978:73 and Sunik 1962:130, both cited in Knott 1995:58).

The causative and passive uses of the *-bu-* suffix in Manchu are illustrated in (2a) and (2b) respectively. Note that a causative interpretation emerges when the subject is construed as agentive, and a passive interpretation emerges when the subject is construed as an affected patient.

(2) Manchu (I. Nedjalkov 1993:194)

a. *i bata-be va-bu-ha*  
he-NOM enemy-ACC kill-CAUS-PAST  
'He made (somebody) kill the enemy.'

b. *i (bata-de) va-bu-ha*  
he-NOM (enemy-DAT) kill-PASS-PAST  
'He is/was killed (by the enemy).'

Zaxarov (1879, cited in Knott 1995:57) reports that in classical Manchu the suffix *-bu-* was also used to convey permissive meaning. As discussed in V. Nedjalkov (1964:301-310; 1971:165-171, cited in I. Nedjalkov 1993:193-194), this allows us to posit a causative > reflexive causative > passive development, again via a permissive or non-interventive ‘let’-type causative, as follows:

(i) *He caused/let somebody kill somebody else* (causative)  
(ii) *He let somebody kill him (by his own negligence)* (reflexive-causative)  
(iii) *He was killed* (passive)

In Evenki (Northern Tungusic), the cognate for *-bu-* is expressed as *-v-/p-/b-/mu-/vuv-/muv-/mup-* (these variants being phonologically conditioned) (I. Nedjalkov 1997:218). These suffixes can also be used to express causative and passive meanings, as illustrated in (3a) and (3b) respectively.

(3) Evenki (I. Nedjalkov 1993:195)

a. *mit homoti-va eme-v-re-p*  
 we-NOM bear-ACC come-**CAUS**-NONFUT-1PL  
 'We brought the bear with us.'

b. *mit homoti-du eme-v-re-p*  
 we-NOM bear-DAT come-**PASS**-NONFUT-1PL  
 lit. 'We are come by the bear.' =  
 'We were affected by the bear's coming.'

I. Nedjalkov (1993) reports that whereas in Manchu the causative use of the suffix *-bu-* is more productive than the passive use, the reverse situation holds in Evenki, such that the passive use of the suffix *-v(u)* and its variants is more productive than the causative use. Indeed, the causative use of the suffix *-v-* and its phonological variants is found only in archaic and non-productive causatives (i.e. "fossilized" forms), with productive causatives being formed with the suffix *-vkAn/-mukAn/-pkAn*, as illustrated in (4) below.

(4) Evenki (I. Nedjalkov 1997:231)

*Nungan min-du sulaki-l-ve iche-vken-e-n*  
 he-NOM I-DAT fox-PL-ACC see-**CAUS**-NONFUT-3SG  
 'He showed me foxes.'

Knott (1995:57, citing Sunik 1962:130) points out that this productive causative suffix *-vkAn* consists of the suffix *-v-* and an element *-kAn*, which is etymologically derived from a verb meaning "say". It appears therefore that Evenki has recruited the newer causative suffix *-vkAn* because the older causative suffix *-v-* has lost much of its semantic specificity.<sup>2</sup>

Evidence that the causative suffix *-v-* has become semantically generalized (and thus highly grammaticalized) can be seen in its great versatility, or multifunctionality. I. Nedjalkov (1997:233-235) reports that in Evenki the suffix *-v-* functions as a valence-extender in a variety of constructions, namely certain causatives, instrumentals and resultatives. In these contexts, use of the suffix *-v-* allows for the introduction of an additional argument in the form of an agent, instrument, or location. At the same time, I. Nedjalkov (1997:226-228) also reports the productive use of *-v-* as a valence-reducer, particularly as a passive marker, and with a restricted set of transitive verb stems as an anticausative marker as well. These different uses of the suffix *-v-* may represent instances of homophony, as pointed out in I. Nedjalkov (1997). At the same time, it is in fact possible and highly likely that they point to a diachronic development in which a lexical verb meaning 'give' has evolved multiple functions in the following direction:

<sup>2</sup> Li and Whaley (forthcoming) also report a similar conclusion, based on reconstructive analyses of *-w(u)* constructions in a cluster of Oroqen dialects, which also form part of the Manchu-Tungusic language family.

<u>Lexical source</u> >>	<u>Valence-extenders</u> >>	<u>Valence-reducers</u>
Verb meaning 'give' (i.e. <i>bu</i> )	<b>causative suffix -v-</b> resultative + suffix -v- etc.	anticausative suffix -v- <b>passive suffix -v-</b> etc.

Indeed, I. Nedjalkov (1978:73, cited in Knott 1995:58), has posited that the passive suffix *-v-* in Evenki developed from a verb meaning 'give' via the permissive-causative. This development is similar to the one proposed by Keenan (1985) discussed above. We reiterate this causative-passive development as follows:

causative > reflexive permissive-causative > passive

In support of this proposed diachronic development, Knott (1995) offers corroborating examples from two Tungusic languages that are closely related to Evenki. Citing Novikova (1968), Knott points out that in Even, "the cognate suffix can have either passive or permissive meaning; moreover, in its permissive function it specifically denotes 'unwilling permission' as a result of negligence, or inability to prevent the action" (1995:58; see also Malchukov 1993:378-383<sup>3</sup>). It is interesting to note that the passive uses of the suffix *-v-* in Even likewise tends to be associated with adversative contexts, usually denoting an action that is unfavorable to the subject (e.g. Malchukov 1993:378; 1995:14).

Furthermore, citing an example from I. Nedjalkov (1978:73), Knott also points out that in Solon, "the cognate suffix *-u:-* has the meaning 'let something happen to oneself', without being accompanied by any overt reflexive marker" (1995:58). The Solon example is illustrated in (5) below:

(5) Solon (I. Nedjalkov 1978:73)  
*zaw-u:-sa*  
 catch-**PERMISSIVE-RECENT PAST**  
 'He let himself be caught.'

It is easy to see how this type of construction comes to be chiefly associated with inadvertent contexts such as 'He inadvertently let himself be caught' and from thence to passive interpretations such as 'He was caught'. Crucial to this type of permissive-causative > passive development is an inherent reflexive meaning, regardless of whether overt co-referential marking is obligatory or optional. In the case of Solon, as shown in (5), reflexivity can be morphosyntactically realized via "zero anaphora" (i.e. omission of a

<sup>3</sup> Malchukov (1993) briefly presents a causative > reflexive > passive analysis, among others, then goes on to advocate a functionally (i.e. semantically and pragmatically) motivated account for the observed correlations between factitive causatives, permissive causatives, and passives. (Included among the factitive causatives are the manipulative, coercive and directive type causatives.) Given that permissive-causatives in Even involve non-volitional subjects, Malchukov views both permissive causatives and passives as subsets of adversative constructions, with permissive causatives occupying one end of the continuum and prototypical passives the other. In our view, the diachronic account (i.e. the causative > reflexive > passive analysis) that is being highlighted in this paper is not incompatible with Malchukov's functional account.

reflexive pronominal; see Malchukov 1993:384) in the sense of 'He let (someone) catch (him)'. The prevalence of zero anaphora in the form of non-overtly expressed reflexive pronominals could then easily pave the way for the emergence of passive reanalyses such as 'He was caught'.

It is significant that in Even, even though overt co-referential marking involving a reflexive pronoun is obligatory in the case of factitive (i.e. manipulative, coercive, directive) causatives, a reverse situation tends to hold in the case of permissive-causatives, in that the use of "zero anaphora" (i.e. co-referential deletion) is very common among permissive-causatives. It appears that, since semantic reflexivity need not be syntactically realized, the structure of these Tungusic languages can easily facilitate the emergence of passive interpretations. That is, a passive interpretation can easily emerge when the subject is construed, not as the agent, but as the affected patient.

Comparative evidence from the Manchu-Tungusic languages thus provides us with some important insights into how the causative > passive development could have evolved via the permissive and reflexive context, in this particular case involving a suffix that is claimed to have been etymologically derived from the lexical verb 'give'. In the following section, we will examine if, and to what extent, a similar development is attested in languages beyond the Manchu-Tungusic borders. We will first focus on the causative-passive relationship of 'give' constructions in some Chinese dialects and in colloquial Malay, and will then account for the absence of a similar phenomenon in Thai, Khmer and Vietnamese.

## 2 Causative-passive link in the 'give' constructions of other East and Southeast Asian languages

It has also been noted that among the Chinese dialects, verbs meaning 'give' are frequently used as an agent marker in passive constructions. According to Norman (1982:245), this is attested in most Min, Hakka and Yue dialects, and also in Mandarin dialects, including the dialect spoken in the Beijing area. Indeed, Norman considers the use of a 'give' verb to express passive meaning to be "pan-Chinese".

Xu (1994) identifies twenty-three Chinese dialects which make use of verbs meaning 'give' to express agent marking in passive constructions. Hashimoto (1988) also points out that many southern Chinese dialects use 'give' verbs as passive markers, and that these 'give' verbs closely resemble the most common forms for the verb 'to give' or 'to give back' *in the southern minority languages* (p.347, emphasis added). Matthews, Xu and Yip (2005:270) provide additional examples from a number of southern Min dialects, namely, Shantou *pug*, Chao'an *k'x?*, Chaoyang *k'i?* and Jieyang *k'e?*. Table 2 lists some examples for ease of reference.

**Table 1:** *Passive markers derived from 'give' verbs in various Chinese dialects (data from Hashimoto 1988:347-348; Matthews, Xu & Yip 2005:270; Xu 1994:366).*

<u>Chinese dialects</u>	<u>verbs meaning 'to give'</u>		<u>passive markers derived from 'give' verbs</u>	
Min (Chaoshan: Shantou)	分	<i>pug</i>	分	<i>pug</i>
Min (Hainan0	分		分	
Hakka (Meixian)	分	<i>pun<sup>1a</sup></i>	分	<i>pun<sup>1a</sup></i>
Yue (Cantonese)	畀	<i>pei<sup>2a</sup></i>	畀	<i>pei<sup>2a</sup></i>

Wu (Jiangsu: Suzhou)	撥	<i>pe<sup>2a</sup></i>	撥	<i>pe<sup>2a</sup></i>
<b>Hubei (Hankou)</b>	<b>把</b>		<b>把</b>	
Mandarin (Hankou)		<i>pa<sup>3a</sup></i>		<i>pa<sup>3a</sup></i>
Mandarin (numerous dialects)		<i>gei</i> (& variants)		<i>gei</i> (& variants)
Min (Chaoshan: Chao'an)		<i>k'yr?</i>		<i>k'yr?</i>
Min (Chaoshan: Chaoyang)		<i>k'i?</i>		<i>k'i?</i>
Min (Chaoshan: Jieyang)		<i>k'e?</i>		<i>k'e?</i>
Wu (Pingyang: Zhejiang)		<i>k'a<sup>4a</sup></i>		<i>k'a<sup>4a</sup></i>
Wu (Jiangsu: Yangzhou)	給		給	
Anhui (Hefei)	給		給	
Gan (Yumen)	給		給	
Gan (Anxi)	給		給	
Wu (Wenzhou)	丐		丐	
Min (Amoy or Xiamen)	互	<i>ho<sup>3b</sup></i>	互	<i>ho<sup>3b</sup></i>
Min (Yongchun)	互	<i>ho<sup>3b</sup></i>	互	<i>ho<sup>3b</sup></i>
Min (Fuzhou)	乞		乞	
Gan (Gao'an)	乞	<i>hou<sup>3a</sup></i>	乞	<i>hou<sup>3a</sup></i>
Wu (Wenzhou: Zhejiang)		<i>ha<sup>3a</sup></i>		<i>ha<sup>3a</sup></i>
Wu (Jiangsu Nantong)		<i>ha<sup>3a</sup></i>		<i>ha<sup>3a</sup></i>
Anhui (Shexian)		<i>xe</i>		<i>xe</i>
Anhui (Jixi)		<i>xā</i>		<i>xā</i>
Anhui (Taihu)		<i>t'ai</i>		<i>t'ai</i>
Qinghai (Xining)		<i>tɛ<sup>4</sup></i>		<i>tɛ<sup>4</sup></i>
Wu (Jiangsu: Rugao)		<i>te<sup>3a</sup></i>		<i>te<sup>3a</sup></i>
Xiang (Hengyang)	得		得	

Minority dialects verbs meaning 'to give' or 'to give back'

Yao		<i>pa<sup>2</sup></i>
Maonan		<i>poi<sup>2</sup></i>
Zhuang		<i>poi<sup>2</sup></i>
Dong	<i>pər<sup>3</sup></i>	<i>sa:i<sup>1</sup></i>
Miao	<i>pɛ<sup>33</sup></i>	<i>kang<sup>42</sup></i>
	<i>pə<sup>55</sup></i>	
Zhuang	<i>hau<sup>3</sup></i>	
	<i>hut<sup>3</sup></i>	
Dai (Dehong)	<i>hau<sup>3</sup></i>	<i>sa:i<sup>3</sup></i> ('give back')
Dai (Xishuangbanna)	<i>hut<sup>3</sup></i>	<i>sa:i<sup>3</sup></i> ('give back')
Buyi	<i>hau<sup>3</sup></i>	
	<i>hai<sup>3</sup></i>	
Sui		<i>ha:i<sup>1</sup></i>

It is thus not surprising that Chinese linguists often note that the verb 'give' is a "favorite candidate" for agent marking in passive constructions (e.g. Xu 1994; Yuan 1983; Zhan 1981). The question that interests us here is how the verb 'give' comes to acquire an agent marking function. Xu (1994:366) claims that the crucial link between the lexical use of 'give' and its passive use is the causative. Xu presents four arguments for her position, two of which are discussed here. First, it is noteworthy that some Chinese dialects also make use of the verb 'give' to convey causative meaning. Second, diachronic data suggest that the early uses of *gei* 'give' in eighteenth century Chinese texts have more of a causative nuance, while more clear-cut passive uses appeared later in texts from the

nineteenth century onward. Based on both synchronic and diachronic evidence, Xu posits the following grammaticalization pattern for the 'give' passive in Chinese:

lexical 'give' > causative 'give' > passive 'give'

Chappell and Peyraube (2001) make similar claims for southern Min dialects. This development is also noted in Matthews, Xu and Yip's (2005) discussion of passive and unaccusative constructions in the Jieyang dialect of Chaozhou. Zhang (2000) also posits a similar development, not only for Chinese but for other languages as well.

Below we illustrate the causative and passive uses of a 'give' morpheme in a northern as well as a southern Chinese dialect, namely an example from Mandarin involving *gěi* and an example from Cantonese involving *béi* respectively.<sup>4</sup>

(6) Mandarin (Xu 1994:368, 374)

a. *wǒ gěi nǐ cāi ge míyú* (permissive-causative)  
 I GEI you guess CL riddle  
 'I (will) let you guess a riddle.'

b. *wǒ gěi zhèi wèi ... yě nǚzi jiǒngzhu le* (passive)  
 I GEI this ... wild woman embarrass ASP  
 'I was embarrassed by this wild woman ...'

(7) Cantonese

a. *léih dimgaai mh bēi ngóh zou zau ne?* (permissive-causative)  
 you why NEG BEI me early leave Q  
 'Why won't you let me leave early?'

b. *kéuih chamjat yauh bēi yàhn laau* (passive)  
 s/he yesterday again BEI people scold  
 'S/he got scolded again yesterday.'

Haspelmath (1990:48) has suggested the possibility of a reflexive link between the causative and the passive uses of the 'give' morpheme in Chinese, noting however that no overt signs of reflexive morphology appear to be involved. The example he used is reproduced in (8) below.

(8) Mandarin (Haspelmath 1990:48)

*Lǐsì gěi Zhāngsān kànjiàn-le.*  
 Lisi give Zhangsan see-ASP  
 lit. 'Lisi gave Zhangsan see (him).' =  
 'Lisi was seen by Zhangsan.'

<sup>4</sup> Whereas Hashimoto (1988) transcribes the Cantonese 'give' morpheme as *pei<sup>2a</sup>*, here we follow the transcription convention used in Matthews and Yip (1994).

Semantic reflexivity is syntactically more transparent in the Cantonese dialect, however, where co-referential pronominals are sometimes used in reflexive-causative constructions, as in (9).

(9) Cantonese

(léih) chinkei mhou béi yàhn ganjung (léih) aa  
 you make.sure don't BEI people follow you SFP  
 'Make sure you don't let anyone follow you.' (reflexive-permissive)  
 'Make sure you are not followed.' (reflexive-passive)

This suggests the following grammaticalization path for the emergence of passive 'give' in the Chinese languages:

lexical 'give' > causative 'give' > reflexive-causative 'give' > passive 'give'

A similar development can be seen in colloquial Malay spoken in peninsular Malaysia.<sup>5</sup> Below we illustrate some constructions in Malay involving the morpheme *bagi* 'give'. These include the following range of functions: lexical verb, causative, reflexive-causative, reflexive-passive, and passive uses.<sup>6</sup>

(10) Malay

a. *Cikgu selalu bagi kita banyak kerjarumah.* (lexical verb)  
 teacher always GIVE us a.lot.of homeworK  
 'The teacher always gives us a lot of homework.'

b. *Cikgu akan bagi kita pulang awal esok.* (permissive-causative)  
 teacher will LET us go.home early tomorrow  
 'The teacher will let us leave early tomorrow.'

c. *Cikgui tak bagi kita tanya (diai).* (reflexive-causative)  
 teacher NEG LET us ask him/her  
 'The teacher won't let us ask him/her (any questions).'

d. *Kereta kita habis bagi dia rosak.* (passive)  
 car us completely PASS he wreck  
 'Our car was completely wrecked by him.'

<sup>5</sup> Our field reports indicate considerable variation in the use of passive *bagi* among native speakers of Malay. Those familiar with the northern dialects of peninsular Malaysia (e.g. Kedah and Perak), for example, generally accepted constructions showing permissive/passive uses of *bagi*, while some of our consultants from the southern part of the country (e.g. Johor) tend to reject such constructions, preferring instead to recast the examples shown to them with alternative passive constructions.

<sup>6</sup> The morpheme *bagi* is also used to express other functions, such as dative, benefactive, purposive, and stance marking; we discuss these uses elsewhere (e.g. Yap & Iwasaki 1998; Yap 1999; 2001, 2002).

A passive interpretation readily emerges in (10d) above when a non-agentive (and non-volitional) argument appears in subject position. An erstwhile causer gets reinterpreted as the demoted agent, while the erstwhile (and semantically bleached) causative verb *bagi* gets reanalyzed as a case marker of a demoted agent (typically referred to as a passive marker in many languages, including the Chinese dialects discussed earlier).

It should be noted here that Malay, like many other languages, has a number of ways to express passive meaning. The prototypical (and most productive) passive in Malay, both in the standard written and colloquial registers, is the passive prefix *di-*. Another fairly productive passive marker in colloquial speech is the periphrastic passive marker *kena* '(be) hit, come in contact with', equivalent to Thai passive marker *thùuk* (Iwasaki & Ingkaphirom 2005). Passive meaning is also occasionally conveyed through the prefixes *ter-* and *ber-*, and the circumfix *ke-...-an*.<sup>7</sup> A noticeable difference between passive *bagi* and these other means of expressing passive meaning is that passive *bagi* requires the overt expression of an agent, as shown in (11). In this regard, passive *bagi* is semantically more similar to the English *get*-passive rather than the *be* + V-en passive, in that both have a reflexive sense in which the affected patient (especially if animate) is often viewed to be to some extent responsible for his/her/its adversative consequence.

(11) Malay					
Ayah	{	<i>bagi</i>	*( <i>orang</i> )	}	<i>tipu</i> .
		<i>di-</i>	(* <i>orang</i> )		
		<i>kena</i>	(* <i>orang</i> )		
		<i>ter-</i>	(* <i>orang</i> )		
father	PASSIVE	people		cheat	
'Father was deceived.' / 'Father got cheated.'					

This constraint against agent deletion applies not only to passive *bagi* in colloquial Malay, but also to passive *béi* in colloquial Cantonese (e.g. Hashimoto 1972; Matthews & Yip 1994), as well as passives in Chaozhou such as Jieyang *k'e?* (Matthews et al. 2005). Passive *gěi* in colloquial Mandarin, on the other hand, permit agent deletion in certain

<sup>7</sup> The prefix *ber-* and the circumfix *ke-...-an* have other functions, and only occasionally convey passive meaning, as illustrated in the examples below:

- (i) *Baju itu masih belum ber-jahit*  
dress that still not BER-sew  
'That dress is still not sewn yet.'
- (ii) *Kalau kita tak tahu bahasa Inggeris, kita akan ke-tinggal-an.*  
if we not know language English we will KE-leave.behind-AN  
'If we do not know English, we will be left behind.'

It is also important to note that the prefix *ter-* is primarily a marker of unintentionality or lack of controllability, and a passive interpretation emerges when the subject is an affected patient instead of a volitional agent, as illustrated in (iii) below.

- (iii) *Basikal-nya ter-humban ke tepi jalan.*  
bicycle-GEN TER-throw DIR side road  
'His/her bicycle was thrown off to the side of the road.'

constructions, as illustrated in (12). This may be an indication that passive *gěi* has grammaticalized further as a passive marker.

(12) Mandarin  
 Fángzì gěi shāo-le. (agentless passive)  
 house GEI burn-ASP  
 'The house was burned down.'

It should also be pointed out that passive *bagi* constructions occur rather infrequently in colloquial Malay, and some native speakers of Malay have pointed out to us that the use of passive *bagi* appears to be more widespread in speech to non-native speakers of Malay. This would suggest that the passive uses of *bagi* may to some extent be influenced by language contact, particularly from the Chinese dialects. Nevertheless, there is sufficient evidence to indicate that the interpretations of certain reflexive *bagi* constructions are ambiguous between a permissive-causative meaning and a passive reading. Much depends on whether the subject is interpreted as a volitional agent or an affected patient. This is highlighted in (13) below.

(13) Malay  
 Aku tak suka **bagi** orang marah.  
 I NEG like give people scold  
 'I don't like to let people scold (me).' (permissive-reflexive reading)  
 'I don't like to be scolded.' (passive reading)

In a sense, one could argue that passive interpretations of *bagi* are highly context-dependent, unlike the case for the prototypical prefix *di-*. Nevertheless, the evidence from colloquial Malay shows that the language is structurally equipped to facilitate the emergence of passive uses of *bagi*. Moreover, the evidence further indicates that the causative-passive relationship involving the 'give' morpheme is linked via the reflexive. Based on our observations of *bagi* constructions in colloquial Malay, we obtain the following slightly modified grammaticalization path as follows:

lexical 'give' > causative 'give' > reflexive-causative 'give' >  
 (reflexive-)passive 'give'

The reflexive link becomes less transparent as the passive uses of 'give' morphemes become more productive and extend to more contexts where the affected subject no longer need to be animate and volitional. In the case of colloquial Malay, the use of passive *bagi* is still fairly restricted, and the constraint that the subject be animate is still fairly strong, such that passive constructions with inanimate subjects are often not acceptable, though not always, as illustrated in (14a) and (14b). In colloquial Cantonese and Mandarin, on the other hand, this constraint appears not to hold, and affected patient subjects need not be volitional, as illustrated in (15) and (16) respectively.

(14) Malay

- a. *Rumah kita HABIS bagi api jilat!*<sup>8</sup>  
house we finish give fire lick  
'Our house was completely licked by the fire!'
- b. *?Duit tu bagi orang ambe?* (=ambil)!  
money that give people take  
'The money was taken (i.e. stolen) by someone!'

(15) Cantonese

*kéuih béi lousi fat lau tong*  
s/he PASS teacher punish stay.back class  
'S/he was punished (by the teacher) and made to stay back after class.'

(16) Mandarin

*Fángzì gěi (huǒ) shaō-le.*  
house give fire burn-ASP  
'The house was burned down (by the fire).'

In addition to animacy and agent volitionality (hence controllability), aspectual and modal cues also play an important role in disambiguating between causative and passive interpretations (see also Matthews et al. 2005). Consider (14a) and (14b) above. In (14a), the availability of completive aspect marker *habis* ('finish') yields a telic and change-of-state reading that is highly compatible with the resultative state interpretation of passive constructions. In (14b), further expanded as (14b') below, a passive interpretation emerges in the context of a realis interpretation (i.e. entailment of loss of money), while a 'let'-type permissive/causative interpretation emerges in the context of an irrealis interpretation (i.e. there is no entailment that someone has taken the money).

(14b') (i) *Duit 'tu bagi orang ambe?* (=ambil)!  
money DEF give people take  
'The money was taken (i.e. stolen) by someone!'

(ii) *Duit 'tu, bagi orang ambe?* (=ambil)!  
money that give people take  
'That money, let someone have it!'

Note that there is also a more prominent pause (indicated by a comma) in the causative construction in (14b':ii) above, compared to the passive counterpart in (14b':i). This pause indicates that the theme *duit 'tu* ('that money') is more loosely associated with the rest of the sentence, functioning more as a topic NP. It is thus possible to construe of an implicit causer preceding the verb *bagi* ('let') in the main clause. For example, we can obtain the reading: 'That money, (we/you) (just) let someone have it!' In the passive *bagi* construction, on the other hand, the theme *duit 'tu* ('that money') is much more tightly

<sup>8</sup> (14b) would be unacceptable, or far less acceptable, without the adverbial *habis* 'completely'.

integrated into the sentence, appearing in subject position, and conveying a strong sense of affectedness as a result of the action expressed by the verb.

It is interesting to note that while passive uses of ‘give’ morphemes are attested in many Chinese dialects and some of the Tai and Austroasiatic languages, as well as to some degree in colloquial Malay as it is spoken in parts of West Malaysia, the passive use of the ‘give’ morpheme does not show up in Thai, Khmer and Vietnamese.<sup>9</sup> This is significant because the ‘give’ morpheme in these languages is used to express a variety of functions, including the causative-permissive (see Table 2 for some examples). In terms of a causative > passive development, we thus see the following grammaticalization path for these languages:

lexical ‘give’ > permissive-causative ‘give’ > \* passive ‘give’

The absence of passive uses of the ‘give’ morpheme in these languages permit us to examine the effects of language-specific constraints on the ‘give’ constructions of areally contiguous languages in East and Southeast Asia. In a recent crosslinguistic comparison of Malay, Thai and Chinese ‘give’ constructions, Yap and Iwasaki (1998) highlighted how the ‘give’ morpheme *hây* in Thai is semantically more restricted in the range of causer and causee arguments that it can take. More specifically, in a causative construction, *hây* is restricted to a volitional causer and a volitional causee; hence it is strongly associated with an agentive subject. This then prevents *hây* from developing into an agent marker to convey passive meaning. That is, since *hây* does not accept non-volitional (and thus adversely affected patient) subjects, an adversative passive interpretation cannot emerge. The contrast in (17a) and (17b) below highlight this volitional causer (i.e. agentive subject) constraint on causative *hây* constructions which blocks the development of passive uses of *hây*.

**Table 2: Absence of passive ‘give’ constructions in Thai, Khmer, and Vietnamese**

Lexical GIVE	Permissive-causative GIVE	Passive GIVE
<b>Thai</b>		
<i>kháw hây n̩en chán</i> 3SG GIVE money 1SG 'S/he gave me money.'	<i>m̩ēe hây l̩uuk pay duu n̩ay</i> mother GIVE child go watch movie 'Mother let her child go and see a movie.' (Newman 1996:189)	NONE
<b>Khmer</b>		

<sup>9</sup> Ruetaivan Kessakul (personal communication Aug 31, 1998), however, provides the following sentence as an example of a passive use of *hây* ‘give’ in Thai.

*chán m̩ây yàak hây kháw duu thùuk*  
I NEG want give him look down  
'I don't want to let him insult (me).'  
'I don't want to be insulted by him.'

Note that the above sentence also shows ambiguity between a causative and passive interpretation. It is possible that the causative > passive development is very restricted in Thai. Further work is needed to examine the extent of this development.

koat baan aoy luy khñom 3SG PAST GIVE money me 'He gave me money.'	kom aoy koat dyg ey NEG GIVE 3SG know PRT 'Don't let him know.'	NONE
<b>Vietnamese</b>		
Ông áy va cho con gái chiéc xe HON that just GIVE daughter CL car 'He just gave his daughter a car.' (Matisoff 1991:429)	Ông áy khñong cho tôi thôî HON that NEG GIVE 1SG resign 'He wouldn't let me resign.' (Matisoff 1991: 429)	NONE

(17) Thai

a. *dèk khon níi hây dèk khon hùn yík tua kháw*  
child CL this LET children CL other pinch body her/him  
'The child let the other children pinch her/him.' (volitional causer)

b. *dèk khon níi hây dèk khon hùn yík*  
child CL this LET children CL other pinch  
'The child let the other children pinch (someone).' (volitional causer)  
\* 'The child was pinched by the other children.' (non-volitional causer)

(17a) and (17b) also illustrate that reflexive-causative uses of *hây* require that the subject be a volitional (i.e. intentional) causer. This also accounts for the absence of (reflexive-)passive uses of *hây*.

The picture we get is that in Thai, *hây* is not generally used to convey reflexive meaning. This, however, would be perfectly consistent with what we know about the semantic constraint on *hây*, namely, that the subject associated with the morpheme *hây* is required to rank high in volitionality and agentivity. Reflexive actions involve a situation in which an agent performs an action that affects (or would affect) the agent itself; consequently, from a semantic perspective, the agentive argument is at the same time the affected patient. This means that the agent in a reflexive construction is not fully agentive (i.e. not a prototypical agent), and particularly in adversative contexts, where the subject acquires a strong "affected patient" role and hence invites a passive interpretation, morphemes other than *hây* are recruited instead, among these being *thùuk* '(be) hit, come in contact with' and *doon* 'hit forcefully', with (*dây*) *ráp* '(get) receive' being used for passive constructions with felicitous interpretations. We see then how the semantic constraints on *hây* block its emergence as a (reflexive-) passive morpheme.

A similar semantic constraint appears to account for the absence of the causative > passive development of 'give' morphemes in Khmer and Vietnamese, though the extent of the constraint for each language remains to be more closely examined. More detailed analysis of the 'give' constructions of individual languages will be necessary to help us identify the range of arguments the 'give' morpheme in a given language can take, as well as to identify other possible factors that may contribute to the grammaticalization patterns of 'give' constructions within the language.

### 3 Conclusion

In this paper we have seen that a number of languages employ the same morpheme to convey both causative and passive meaning. In some of these languages (e.g. Manchu-Tungusic, Chinese, Malay), the common morpheme is derived from a verb meaning 'give'. Evidence from these languages suggests that passive meaning could evolve from permissive-causative 'give' constructions via the reflexive context. With examples from languages such as Thai, which show many other similar uses of the 'give' morpheme *and yet lack passive uses of 'give'*, we were also able to identify some language-specific constraints that restrict the further grammaticalization of causative 'give' constructions into (reflexive-) passive contexts. A closer examination of other East and Southeast Asian languages will permit us to see more clearly the extent of the causative-passive relationship (or lack thereof) in the 'give' constructions in these areally contiguous languages, and to further ascertain if, and to what extent, the causative > passive development of 'give' morphemes is more than just an isolated phenomenon.

Another interesting area for further investigation is a broader typological comparison of languages that use morphological 'give' constructions and those that use periphrastic 'give' constructions. In this paper, we have discussed some Manchu-Tungusic languages as examples of the former, and some Chinese dialects as well as Malay and Thai as examples of the latter. For future research, it would be very useful to investigate the causative > passive development (along with other grammaticalization patterns) of 'give' constructions in many more languages.

### Acknowledgment

An earlier version of this paper was originally presented at the Eighth Annual Meeting of the Southeast Asian Linguistic Society in Kuala Lumpur, Malaysia, July 20-22, 1998. Parts of this paper was further expanded (crosslinguistically) into a paper presented at the International Conference of Cognitive Linguistics at Stockholm University, July 10-16, 1999, and published as Yap & Iwasaki (2002). We gratefully thank participants at these two conferences, and the editors and reviewers of this present volume, for their valuable feedback. We would also like to especially thank Phensri Jejantra, Ruetaivan Kessakul, and Amy Meepoe-Baron for sharing with us their native speaker intuitions on Thai; to Xiao-Xin Sun, Hong-Yin Tao and Yue Wang, on Mandarin Chinese; to Peggy Wing-sze Ng, Gloria Sau-chun Poon, Joanne Wing-shan Wong, Virginia Yip, on Cantonese; and to Kartini Ahmad, Aini Rozita Ahmad Anuar, Norhaida Aman, Mazlin Mohd Asary, Jim Colins, Ruziyanti Mokhtar, Norizan Rajak, Rogayah Razak, Pamela Thomas Joseph, Teo Kok Seong and Wuey-Ping Tye, for discussions on passive 'give' constructions in Malay. We are also very grateful to Stephen Matthews, Weera Ostapirat and Kingkarn Thepkarnchana for helpful suggestions and comments on this paper. We alone are responsible for any errors.

### Abbreviations

ACC	Accusative	NOM	Nominative
ASP	Aspect marker	NON-FUT	Non-future tense
CAUS	Causative	PASS	Passive marker
CL	Classifier	PAST	Past tense
DAT	Dative	PL	Plural
DECL	Declarative	SFP	Sentence-final particle
DEF	Definiteness marker	SG	Singular

NEG              Negation morpheme

## References

Gabelentz, Hans von der. 1861. *Über das Passivum. Eine sprachvergleichende Abhandlung*. Abhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften, 8, 450-546. Cited in Haspelmath (1990) and I. Nedjalkov (1993).

Hashimoto, Mantaro J. 1988. The structure and typology of the Chinese passive construction. In Masayoshi Shibatani (Ed.), *Passive and voice. Typological Studies in Language*, volume 16. Amsterdam/Philadelphia: John Benjamins.

Hashimoto, Oi-kan Yue. 1972. *Studies in Yue dialects 1: Phonology of Cantonese*. Cambridge, UK: Cambridge University Press.

Haspelmath, Martin. 1990. The grammaticalization of passive morphology. *Studies in Language*, 14, 25-72.

Iwasaki, Shoichi & Preeda Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge, UK: Cambridge University Press.

Keenan, Edward L. 1985. Passive in the world's languages. In Timothy Shopen (Ed.), *Language typology and syntactic description, Volume 1: Clause structure*, pp. 243-281. Cambridge: Cambridge University Press.

Knott, Judith. 1995. The causative-passive correlation. In David C. Bennett, Theodora Bynon & B. George Hewitt (Eds.), *Subject, voice, and ergativity: Selected essays*, pp. 53-59. London: University of London, School of Oriental and African Studies.

Lai, Huei-ling. 2001. On Hakka BUN: A case of polygrammaticalization. *Language and Linguistics*, 2 (2): 137-153.

Li, Fengxiang & Lindsay Whaley. (forthcoming). Birth-death-resurrection: The grammaticalization cycle of causatives in Oroqen dialects. In A. Malchukov & L. Whaley (eds.), *Recent advances in Tungusic linguistics*. Wiesbaden: Otto Harrassowitz.

Malchukov, Andrey L. 1993. Adversative constructions in Even in relation to passive and permissive. In Bernard Comrie & Maria Polinsky (Eds.), *Causatives and transitivity*, pp. 369-384. Amsterdam/Philadelphia: John Benjamins.

Malchukov, Andrei L. 1995. *Even*. Languages of the World, Materials 12. Munchen/Newcastle: LINCOM EUROPA.

Matisoff, James A. 1991. Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth Closs Traugott & Bernd Heine (Eds.), *Approaches to grammaticalization, Vol. 2: Focus on types of grammatical markers*, pp. 383-454. Amsterdam/Philadelphia: Benjamins.

Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A comprehensive grammar*. London/New York: Routledge.

Matthews, Stephen, Huiling Xu, & Virginia Yip. 2005. Passive and unaccusative in the Jieyang dialect of Chaozhou. *Journal of East Asian Linguistics*, 14, 267-298.

Nedjalkov, Igor V. 1978. Passivnye konstrukcii v tunguso-man'czurskix jazykax. In S.D. Kacnel'son (Ed.), *Lingvisticheskie issledovaniya 1978. Problemy grammaticheskogo stroja jazyka*. Moscow: Akademiya Nauk SSSR Institut Jazykoznanija, 66-75. Cited in Knott (1995).

Nedjalkov, Igor V. 1993. Causative-passive polysemy of the Manchu-Tungusic -bu/-v(u). *Linguistica Antverpiensa*, 27, 193-202.

Nedjalkov, Igor. 1997. *Evenki*. London/New York: Routledge.

Nedjalkov, Vladimir P. 1964. O svjazi kauzativnosyi i passivnosti. In *Voprosy obscego i romano-germanskogo jazykoznanija*, 301-310. Ulfa. Cited in I. Nedjalkov (1993).

Newman, John. 1996. *Give: A cognitive linguistic study*. Berlin: Mouton de Gruyter.

Norman, Jerry. 1982. Four notes on Chinese-Altaic linguistic contacts. *Tsing Hua Journal of Chinese Studies*, 14, 243-247.

Novikova, K. A. 1968. *Evenskij jazyk*. In P. Ja Skorik (Ed.), *Jazyki narodov SSSR* 5, 88-108. Leningrad: Nauka. Cited in Knott (1995).

Sunik, O. P. 1962. *Glagol v tunguso-man'czurskix jazykax*. Moscow/Leningrad: Nauka. Cited in Knott (1995).

Xu, Dan. 1994. The status of marker gei in Mandarin Chinese. *Journal of Chinese Linguistics*, 22, 363-394.

Yap, Foong Ha & Shoichi Iwasaki. 1998. 'Give' constructions in Malay, Thai and Chinese: A polygrammaticization perspective. *CLS 34: Papers from the Main Session*. Chicago: Chicago Linguistic Society.

Yap, Foong-Ha. 1999. *The polysemy of 'give' constructions in Malay and other languages: A grammaticalization perspective*. Ph.D. Dissertation, University of California, Los Angeles.

Yap, Foong Ha. (2001). *A diachronic analysis of bagi constructions in Malay: Implications for the grammaticalization of 'give' constructions in serializing languages*. Paper presented at the 6th International Conference on Languages of the Far East, Southeast Asia, and West Africa, at the University of St. Petersburg, September 25-28.

Yap, Foong Ha. (2002). *The polysemy of 'give' in Malay discourse: Diachronic evidence from 16th to 20th century texts*. Paper presented at the 9th International Conference of Austronesian Linguistics. Australian National University, Canberra, January 8-11.

Yuan, Jiahua. 1983. *Hanyu fangyan gaiyao* (Essentials of Chinese dialects), 2nd edition. Beijing: Wenzi Gaige Chubanshe.

Zaxarov, Ivan. 1879. *Grammatika man'czurskogo jazyka*. St. Petersburg. Cited in Knott (1995).

Zhan, Bohui. 1982. *Xiandai hanyu fangyan* (Contemporary Chinese dialects). Wuhan: Hubei Renmin Chubanshe.