

PACIFIC LINGUISTICS

Series C - No. 45

SOUTH-EAST ASIAN LINGUISTIC STUDIES

VOL. 3

Nguyen Dang Liem, ed.



Department of Linguistics  
Research School of Pacific Studies  
THE AUSTRALIAN NATIONAL UNIVERSITY

PACIFIC LINGUISTICS is issued through the *Linguistic Circle of Canberra* and consists of four series:

- SERIES A - OCCASIONAL PAPERS
- SERIES B - MONOGRAPHS
- SERIES C - BOOKS
- SERIES D - SPECIAL PUBLICATIONS

EDITOR: S.A. Wurm.

ASSOCIATE EDITORS: D.C. Laycock, C.L. Voorhoeve, D.T. Tryon, T.E. Dutton.

EDITORIAL ADVISERS:

B. Bender, University of Hawaii  
D. Bradley, Australian National University  
A. Capell, University of Sydney  
S. Elbert, University of Hawaii  
K. Franklin, Summer Institute of Linguistics  
W.W. Glover, Summer Institute of Linguistics  
G. Grace, University of Hawaii  
M.A.K. Halliday, University of Sydney  
A. Healey, Summer Institute of Linguistics  
L. Hercus, Australian National University  
N.D. Liem, University of Hawaii

J. Lynch, University of Papua New Guinea  
K.A. McElhanon, University of Texas  
H. McKaughan, University of Hawaii  
P. Mühlhäusler, Technische Universität Berlin  
G.N. O'Grady, University of Victoria, B.C.  
A.K. Pawley, University of Hawaii  
K. Pike, University of Michigan; Summer Institute of Linguistics  
E.C. Polomé, University of Texas  
G. Sankoff, Université de Montréal  
E. Uhlenbeck, University of Leiden  
J.W.M. Verhaar, University of Indonesia, Jakarta

ALL CORRESPONDENCE concerning *PACIFIC LINGUISTICS*, including orders and subscriptions, should be addressed to:

The Secretary,  
*PACIFIC LINGUISTICS*,  
Department of Linguistics,  
School of Pacific Studies,  
The Australian National University,  
Box 4, P.O.,  
Canberra, A.C.T. 2600.  
Australia.

Copyright © The Authors.  
First published 1979.

The editors are indebted to the Australian National University for help in the production of this series.

This publication was made possible by an initial grant from the Hunter Douglas Fund.

National Library of Australia Card Number and ISBN for complete set of volumes: 0 85883 144 9

National Library of Australia Card Number and ISBN for this volume: 0 85883 177 5

TABLE OF CONTENTS

	<i>Page</i>
FURTHER TYPOLOGICAL STUDIES IN SOUTHEAST ASIAN LANGUAGES, by A. Capell	1
<i>Map 1: South-East Asian Languages, and Map 2: Ethnic Minorities in South Vietnam + List of Vietnam Language Groups - preceding page 1</i>	
1. INTRODUCTION	1
1.1. General remarks	1
1.2. Languages and materials used	3
1.3. Scope of the paper	4
2. THE NOUN PHRASE	5
3. THE VERB PHRASE	16
3.1. Introduction	16
3.2. The fully inflected languages	18
3.3. Languages with less inflection	22
3.4. The uninflected languages	27
4. CONCLUSION: TYPOLOGIES OF NEUTRAL DOMINATION	36
<i>References</i>	40
CASES IN ENGLISH AND SOUTHEAST ASIAN LANGUAGES, AND TRANSLATION, by Nguyen Dang Liem	43
0. INTRODUCTION	43
1. CASE RELATIONS IN ENGLISH AND SOUTHEAST ASIAN LANGUAGES	44
2. CASE RELATIONS AND CASE FORMS IN ENGLISH	46
3. CASE RELATIONS AND CASE FORMS IN SOUTHEAST ASIAN LANGUAGES	52
4. TRANSLATION	59
<i>Notes</i>	64
<i>References</i>	65

	<i>Page</i>
THE ROLE OF DE-ETHNISATION AND ATTITUDE IN THE USE OF PILIPINO: A FACTOR ANALYTIC STUDY, by Alejandrino Q. Perez	67
INTRODUCTION	67
RESULTS AND DISCUSSIONS	70
Biographical sketch	77
<i>Notes</i>	78
<i>Appendix "A"</i>	79
<i>References</i>	83
MIXED BILINGUALISM: A PRELUDE TO INCIPIENT CREOLISATION OF PILIPINO?, by Alfonso O. Santiago	85
0. BACKGROUND	85
1. THE PROBLEM	86
2. CREOLISATION AND STANDARDISATION	87
3. THE BILINGUAL SITUATION	87
4. CREOLISATION: EXTENSIONS OF MEANING	90
4.1 Traditional meaning	90
4.2 Contemporary meaning	90
4.3 Meaning of creolisation in this study	91
5. DATA SOURCE AND ANALYSIS	91
5.1 Phonology	91
5.2 Morphology	95
6. A TEST ON CREOLISATION OF PILIPINO	98
7. CONCLUSIONS	101
About the author	103
<i>Notes</i>	104
<i>Appendix A: Taperecorded questions</i>	108
<i>Bibliography</i>	110
THE PEPET IN SAMA-BAJAW, by A. Kemp Pallesen	115
Map: <i>Distribution of Major Subdivis</i> preceding page 115	
1. TOPIC OF DISCUSSION	115

	<i>Page</i>
2. SAMA-BAJAW AS A DISTINCT SUB-GROUPING	115
3. THE DATA	116
4. NORMAL PROTO SAMA-BAJAW REFLEXES OF PAN *ə	116
5. IRREGULAR REFLEXES	119
6. CONDITIONING FACTORS FOR THE PSB REFLEXES OF PAN *ə	124
7. THE DEVELOPMENT OF GEMINATION IN PSB	127
8. PHONEMIC STATUS OF THE PSB REFLEXES *ə AND *ɯ	128
9. PROCESSES BY WHICH ə OR ɯ IS INNOVATED IN SB	128
10. REFLEXES OF PSB *ə AND *ɯ IN DAUGHTER LANGUAGES	130
11. CONCLUSION	133
<i>Notes</i>	134
<i>References</i>	141
ACRONYMIC PATTERNS IN INDONESIAN, by Soenjono Dardjowidjojo	143
1. INTRODUCTION	143
2. DESCRIPTION	144
2.1 Acronymic syllables	145
2.2 Acrotactic	148
2.3 The syllable structure of the acronyms	150
3. DEGREE OF ACCEPTABILITY	152
4. DIRECTIONALITY AND PREDICTABILITY	153
5. COGNITIVE REASONING	154
<i>Notes</i>	158
<i>References</i>	159
KORKU SYLLABLES AND SYLLABLE STRESS, by Norman H. Zide	161
Afterword	181
<i>References</i>	186
MORPHOPHONEMICS OF VERB SUFFIXES IN TSOU, by Patricia Stanley	187
1. INTRODUCTION	187
2. SOME SUGGESTED PHONOLOGICAL ANALYSES	188

	<i>Page</i>
3. VERB INFLECTION	189
3.1. Verbs without stem change	189
3.2. Verbs with suppletive suffixes	190
3.3. Verbs with final vowel deletion	190
3.4. Some residual problems	193
4. SUMMARY	195
<i>Notes</i>	196
<i>Bibliography</i>	197
CONSTRUCTION MARKERS AND SUBGROUPING OF FORMOSAN LANGUAGES, by Raleigh Ferrell	199
1. TOPICALISATION IN FORMOSAN LANGUAGES	199
2. TOPICALISATIONAL VERB INFLECTION	200
3. TOPICALISATIONAL MARKING OF NOUN PHRASES	202
3.1. 3-CM systems	202
3.2. 2-CM systems	203
3.3. $\emptyset$ -CM systems	203
3.4. General and specificational CMs	204
4. SYNTACTIC, LEXICAL AND PHONOLOGICAL SUBGROUPING OF FORMOSAN LANGUAGES	205
5. IMPLICATIONS FOR PROTO-FORMOSAN SYNTAX	208
<i>Notes</i>	209
<i>References</i>	211
FOCUS, TOPIC, AND CASE IN THE PHILIPPINE VERBAL PARADIGM, by Joseph F. Kess	213
<i>Bibliography</i>	235
ON THE DEVELOPMENT OF CONTRASTIVE WORD ACCENT: PANGASINAN, A CASE IN POINT, by R. David Zorc	241
1. PROTO-PHILIPPINE ACCENT	241
2. THE POSITION OF PANGASINAN	243
3. THE LOSS OF PPH ACCENT IN PRE-PANGASINAN	245

	<i>Page</i>
4. THE DEVELOPMENT OF CONTRASTIVE LENGTH IN PANGASINAN	248
4.1. Compensatory length derived from the reduction of *qC clusters	248
4.2. Compensatory length derived from the reduction of geminate clusters	249
4.3. The evidence in doublets for borrowing	250
5. RESIDUAL PROBLEMS AND CONSIDERATIONS	250
5.1. The plural of the word for 'carabao'	250
5.2. The methathesis of *Cq clusters	251
5.3. Innovative developments	251
5.4. Recent borrowings	252
<i>Notes</i>	254
<i>References</i>	257
TOWARDS A RECONSTRUCTION OF THE PRONOMINAL SYSTEMS OF PROTO-CORDILLERAN, PHILIPPINES, by Lawrence A. Reid	259
1. INTRODUCTION	259
2. PRONOUN SETS	260
3. PC PRONOUNS: INTERNAL EVIDENCE	261
4. PC PRONOUNS: EXTERNAL AND PNC EVIDENCE	265
4.1 First person singular (1 s.)	265
4.2 Second person singular (2 s.)	266
4.3 Dual person (1+2 s.)	267
4.4 Third person singular (3 s.)	267
4.5 First person exclusive plural (1 p.)	268
4.6 Second person plural (2 p.)	268
4.7 First person inclusive plural (1+2 p.)	270
4.8 Third person plural (3 p.)	271
5. THE OBLIQUE PRONOUNS	271
6. CONCLUSION	271
<i>Notes</i>	273
<i>Bibliography</i>	274
PERCEPTUAL DIMENSIONS OF TONE: THAI, by Jackson T. Gandour	277
I. INTRODUCTION	277

	<i>Page</i>
II. METHOD AND PROCEDURE	278
A. Language	278
B. Stimuli	280
C. Subjects	282
D. Experimental task	282
E. Method of analysis	284
III. RESULTS	285
A. Interpretation of dimensions	285
B. Subgroup differences in weights for dimensions	293
IV. DISCUSSION AND CONCLUSIONS	295
<i>References</i>	298
 BWE KAREN AS A TWO-TONE LANGUAGE?, by Eugénie J.A. Henderson	 301
<i>Notes</i>	324
<i>Bibliography</i>	325

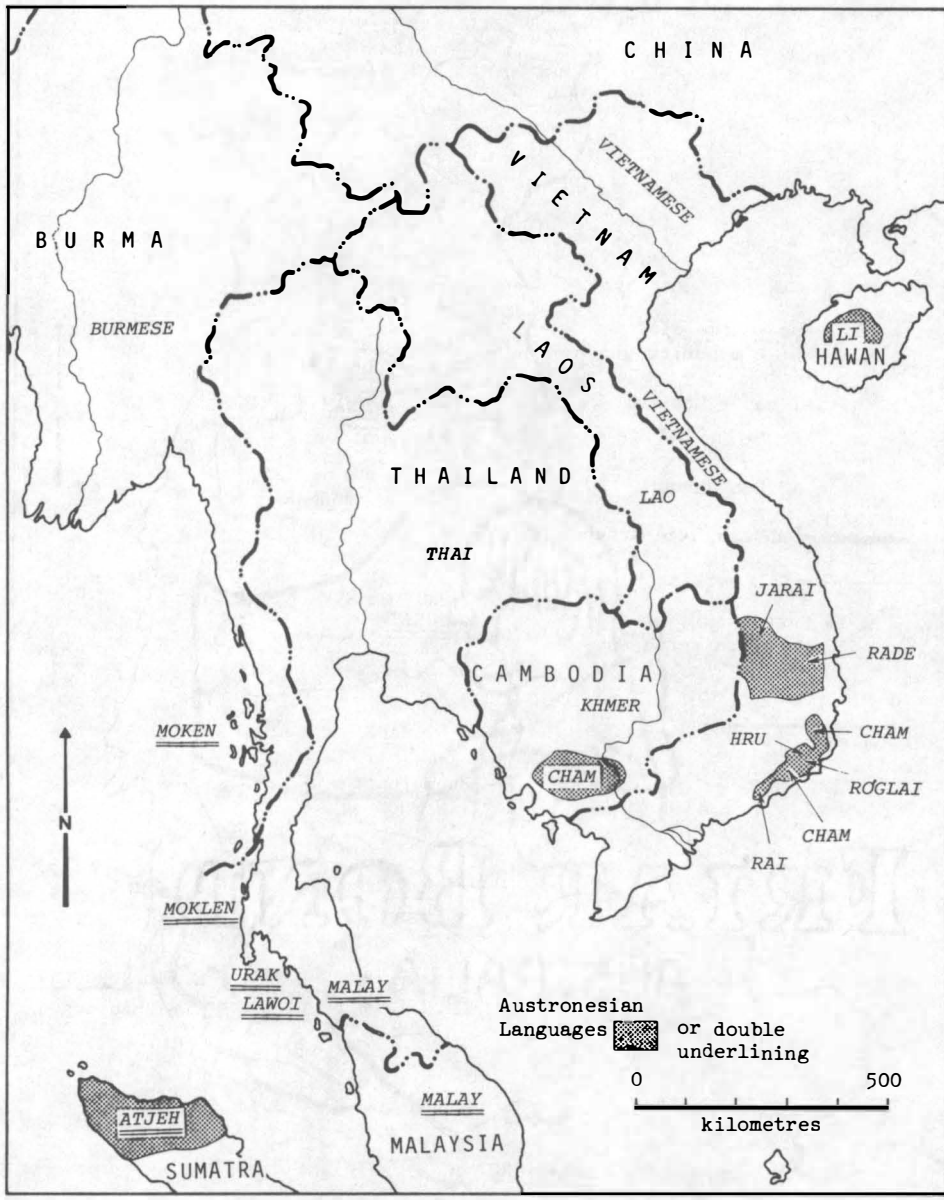
---

EDITORIAL NOTE

Most of the articles included in this volume were received in 1975 or 1977, and theoretical views expressed in them may be at variance with the authors' present views on the subject. This note has been inserted at the specific request of the Editor for this volume on behalf of the authors.

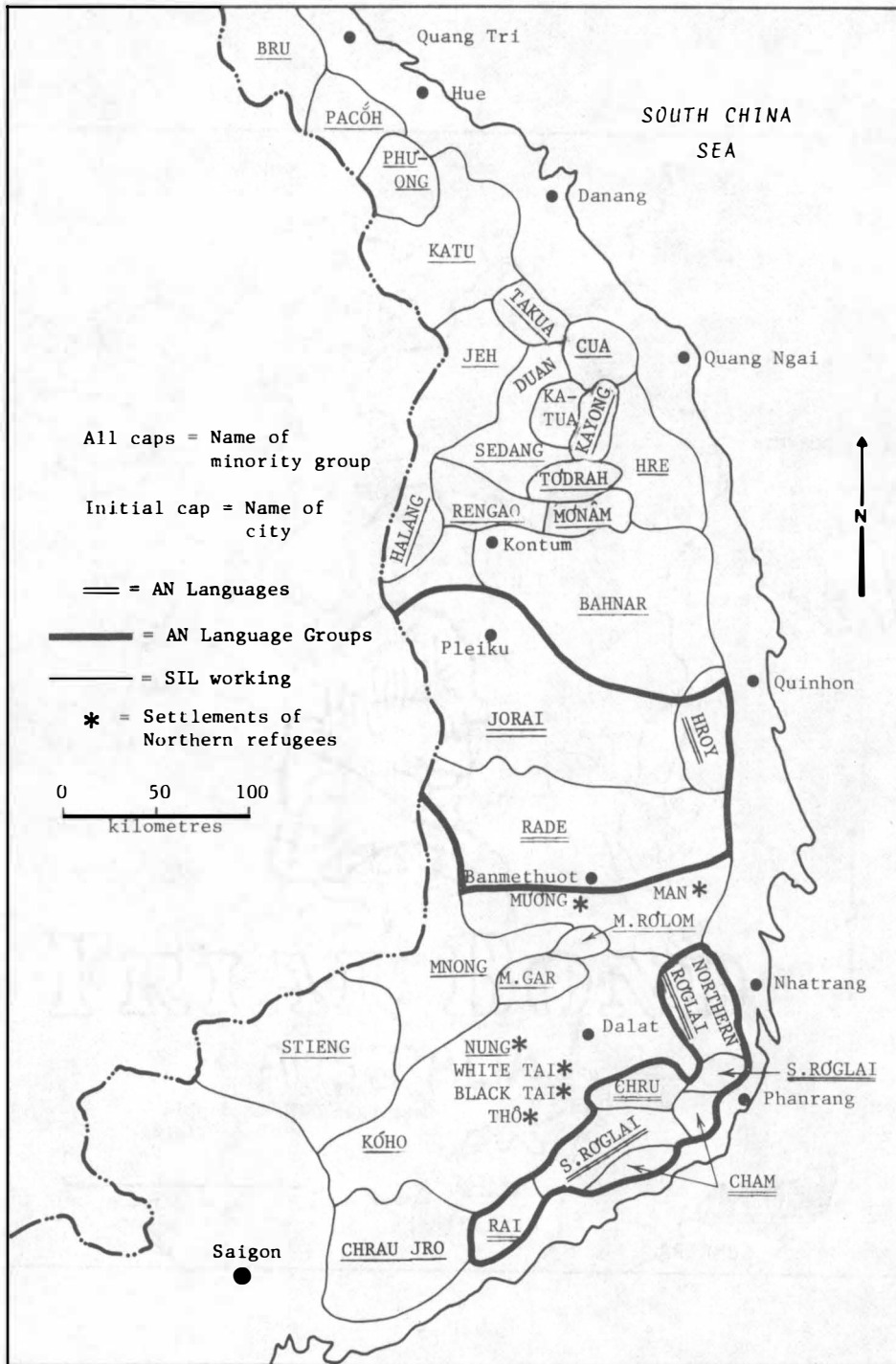
---





Map 1

SOUTH-EAST ASIAN LANGUAGES



Map 2

ETHNIC MINORITIES IN SOUTH VIETNAM

VIETNAM LANGUAGE GROUPS (CÁC NGŨ-TỘC Ở VIỆT NAM)

Việt-Mường (Việtnamuong)

Vietnamese

Mường (Bắc Việt)

Tai

White Tai (Bắc Việt)

Black Tai (Bắc Việt)

Thổ (Bắc Việt)

Nùng (Bắc Việt)

Miao-Yao

Man (Bắc Việt)

Malayo-Polynesian

Chem (Bình Thuận, Ninh Thuận, Châu Đốc)

Rai (Bình Tuy)

Rơglai (Bình Thuận, Khánh Hòa)

Chru (Tuyên Đức)

Radê (Darlac, Khánh Hòa)

Jorai (Pleiku, Phú Bổn, Kontum)

Hroy (Bình Định, Phú Yên)

Mon-Khmer: Katuic

Brū (Quảng Trị)

Pacốh (Thừa Thiên)

Phuong (Thừa Thiên)

Katu (Quảng Nam)

Mon-Khmer: Bahnaric

Takua (Quảng Tín)

Jeh (Quảng Tín, Kontum)

Cua (Quảng Ngãi)

Duan

Katua

Kayong (Quảng Ngãi)

Sedang (Kontum)

Halang (Kontum)

Todrah (Kontum)

Rengao (Kontum)

Monâm (Kontum)

Hrê (Quảng Ngãi)

Bahnar (Kontum, Pleiku, Bình Định)

Mnong (Darlac, Quảng Đức)

Mnong Gar

Mnong Rôlom

Stieng (Phước Long, Bình Long)

Kofo (Tuyên Đức, Lâm Đồng,  
Long Khánh)

Chrau Jro (Long Khánh, Phước Tuy,  
Bình Tuy)



# FURTHER TYPOLOGICAL STUDIES IN SOUTHEAST ASIAN LANGUAGES

A. CAPELL

## 1. INTRODUCTION

### 1.1. GENERAL REMARKS

This paper is called "further" typological studies, because it is a sort of long-delayed sequel to a paper presented to the London Conference of 1961 on Southeast Asian linguistics by F.J. Honey and E.H.S. Simmonds, which embraced Thai and Vietnamese: 'Some elements of nominal structure compared'. To this first paper may then be added one read to the similar conference of 1965 by E.J.A. Henderson under the title 'The Typography of Certain Phonetic and Morphological Characteristics of South-East Asian Languages'.

The present paper extends the study to the verb phrase and brings in other languages, as will be described below.

In the original paper, after a very helpful comparison of the structure of noun phrases in Thai, Vietnamese and Chinese, the authors added: "It would be interesting to extend the work (a) by comparing descriptive systems of the verbal complexes of the three languages, and (b) attempting to apply the method of description of the nominal systems of, say, Khmer and Malay." (p.77). In this paper such an attempt is made, with still other languages added.

The paper has further aims as well, which are in a way by-products of the main aim of the descriptive comparison. In a paper presented to the 1965 London Conference, I outlined a typological approach to language description which I called "concept domination". This, in brief, works on the recognition of a certain bias in languages in many parts of the world, towards elaboration of either the noun phrase or the verb phrase in an utterance. Some languages appear to be more interested in

the details of how an event took place than in who or what performed it. In such languages, the VP may be morphologically elaborate, including complex systems of aspects, moods and tenses, inclusion of subject and object markers in the verbal complex, often along with indicators of direction and other details of the action. This may be called *event domination*, and in the S.E. Asian region Santali provides an example. Other languages seem to be more interested in the actor and/or the goal, and the NP then shows class or gender, number, case, and often the verb marks neither person nor number, and aspect, tense and mood or marked syntactically and not morphologically. The interest is in the NP, and these languages may be called *object-dominated*. The abbreviations ED and OD are used. Other languages, again, appear to be equally interested in both event and object - the African Bantu languages have elaborate noun class systems with concord, and these are involved as much in the verbal complex as in the nominal. These are *double dominated* (DD). They do not occur in the Southeast Asian region. Still others, such as English and most of the Austronesian languages, show little bias in either direction, and the morphology of such languages is comparatively simple. In the extreme case of Chinese and similar languages, there is no overt morphology at all. Most of the southeast Asian languages belong to this type, which is called *neutral domination* (ND).

In 1969 I produced a book entitled *A Survey of New Guinea Languages*, in which this typology was used as a basis of arrangement. A reviewer wrote that the author seemed to have lumped together all languages he could not account for under other headings as "neutral domination", with no further analysis. To a degree that is true, and the author was not forgetful of it even while he did it. In New Guinea ND is comparatively rare, and there was no need to subcategorise the group, any more than to account for the verbal classification of the Athapascan languages of North America, which have no counterpart in New Guinea.

The languages of Southeast Asia, however, provide an opportunity for refining one's thinking about subdivisions within the ND class. It is therefore part of the aim of this paper to do this, in addition to taking up the challenge of Honey and Simmond's paper. It seeks to show that although ND would by implication be a homogeneous group, it is not really so, and subdivisions within it can be made similar to those made within the other classes, both in the original paper and in the *Survey*. English and Thai, for instance, show differences amid common qualities. The paper therefore seeks to examine languages classifiable as ND, but differing among themselves so that ND can be further analysed as the other groups have been.

The 1965 paper by Henderson was concerned with just this type of language that in the "domination" arrangement would be classified as ND. The paper was a phonological examination, and studied the distribution of eight phonetic features spread over sixty languages and India to Oceania. The presence of tone is, of course, a differentiating feature, but it is not relevant in concept domination typology, which functions rather on the morphological and syntactic levels. Even so, the materials used here suggest further avenues of research rather than attain positive conclusions, and they are meant to do so. Although the primary purpose is to provide the necessary subdivisions of the ND type of language, the present study should also lead on to other matters which are not part of this paper, such as the question of what type of deep structure lies behind languages so different in structure as Bantu, Latin and Chinese. It has been suggested that only noun, verb and adjective are represented (as one category) in the base, but it seems doubtful whether adjective can be regarded as primary anywhere. Again it has been suggested that perhaps prepositions derive from verbs. These languages of southeast Asia, like those of the ND group in West Africa, have evidence to offer in this matter. However, it may be put forward as an idea, that the search for universals in language is not so simple as some of the searchers appear to think, and that there is probably no "deep structure" of language as such, traceable through a process of differential development in time and space. These aspects of linguistic research cannot be pursued here, but the paper, if developed to its logical conclusion, could well open up research also in this decidedly different field.

## 1.2. LANGUAGES AND MATERIALS USED

The languages used in the paper on which the present one is based are Thai, Vietnamese and Chinese. Thai has been added as an obvious extension. Korean and Japanese are also obvious extensions as representing north-east Asia; in the south, Khmer is equally called for. Outside this area again, there is the extensive Austronesian family, whose western members impinge on, and indeed intermingle with, the other families. An attempt has been made therefore to give as fair coverage as possible of the languages of this family which are in touch with the other families employed. This means, of course, Malay, but nearer to the field of the other languages, Jarai and Rhade information has been employed, and a missionary among the "Sea Gypsies" of the islands west of Thailand kindly provided material in Urak Lawoi'. Mr Hogan's work here is good, but he has produced his material in a modified Thai script, because the peoples form part of Thailand and it seemed more practical

to produce literacy materials for them in the script employed in the national language - although this could only be done with some modification for certain letters of the Thai alphabet. Moken material was unfortunately unavailable except for a few odd notes. On the southern side of this western Austronesian, the Atjeh language and the Cham dialects offered themselves as near neighbours: and it must not be forgotten that there is a Cham language on the island of Hainan which is very little known to the outside world. This Cham has developed a tonal system based on that of the local Chinese.

### 1.3. SCOPE OF THE PAPER

In the paper presented in 1961, study was confined to the noun phrase. It is desirable therefore to begin from that point here also, to enable a clear study to be made, embodying the results achieved in the previous paper. Towards the end of their paper, Honey and Simmonds stated, "It would be premature to draw any firm conclusions from such facts (as they had shown in the course of the study), but it would be interesting to extend the work by (a) comparing descriptive systems of the verbal complexes of the three languages, and (b) by attempting to apply the method of description to the nominal forms of, say, Khmer and Malay." It was this suggestion that gave the lead into the present paper, which attempts to cover just these fields: to describe the verb phrase and to extend the scope of languages in the way that has been mentioned above.

Burmese was not originally included in the paper, but on second thoughts it has been added. It was omitted because it steps outside the strict SEA area: it belongs to the Tibeto-Burman group of languages (Wolfenden:1929; Pring (1963) and other authors on general linguistic classification) and thus lies apart from the Mon-Khmer (MK) and Austronesian (AN) languages which form the basis of the paper. But it is in geographical company with them and is as worthy of inclusion as Chinese, Korean and Japanese. It serves in a number of cases to point up contrasts between the structures of the various types, perhaps more in the NP than in the VP. There is much argument about the interpretation of VP in Burmese, too much to be even summarised here. Allott's paper on 'The verbal syntagma in Burmese' (Allott 1965:283-308) is the latest and probably the fullest, and is taken here as the guide - but it should be studied in full. In the VP also, tense, voice and mood rather than aspect seem to be dominant, more so than in the MK languages and even the AN languages used here.

Readers of Russian can also use the *Birmanskiy Jazyk* in the Asiatic series of the Moscow Academy of Sciences (Maun Maun Njun etc., 1963)



and some articles in the volume *Jazyki Jugo-Vostočnoj Azii* (Moscow 1967) dealing with various aspects of the language. As in so many of the languages that use local alphabets, transcription in Roman letters varies tremendously among different authors, and it can probably not be claimed that the forms used here are always mutually consistent or even self-consistent, but for the readers intended this will not provide a serious difficulty. Books by W.S. Cornin (1946) and J.A. Stewart (1955) have also been used here.

2. THE NOUN PHRASE

Seeing that there is in most of the languages agreement about the absence of morphological indication of number, gender or case relationships, the first subject of treatment in the noun phrase is the presence of adjuncts in the NP. A simple adjunct such as 'that' or 'good' may be either preposed or postposed, so that mere presence of D or A in the phrase provides nothing diagnostic either typologically or genetically. Where, however, more than one adjunct is present in the NP, more variety of arrangement is possible and the complex NP consisting of the D, N and A may be of importance.

The adjective-noun (AN) phrase comes first in the present treatment. The Chinese, Korean and Japanese order is A + N, the other languages all have N + A, and this includes Khmer and the Austronesian languages.

Examples:

English 'a good man'

Group A	A + N	Chinese Korean Japanese Burmese	xǎo rén chohŭn saram yoi hito \leauŋ té lu
Group B	N + A	Thai Vietnam Jarai Khmer Malay Ur. Law. Moken Atjeh Burmese	phûuchaaj dii gù'o'i tót mnuih 'o'i menùs l?aa orang (jaŋ) baik urak badji? mnut amon urö'əŋ gèt lū \kaŋ

Numeral adjectives may in some of the languages precede the noun, with or without coefficients. Coefficients will be mentioned more fully

below. Jarai shows *lu mnuih*, 'many people', in which the adjective *lu*, 'many' is treated as a numeral and precedes the noun,\* and Malay *oraṅ banjak*, 'people many', as a rule without the ligative *jaṅ*, which is often optional and may make a slight difference in meaning which is not of concern here.

If the adjunct is demonstrative (D), the same dichotomy of languages is found: 'that man', Chinese *nèige rén*; Korean, *ku saram*; Japanese *sono hito*; as against Thai *phûuchaaj nán*; Vietnam *gù'o'i áy*; Khmer *menùs nuh*; Jarai *mnuih ?anǝn*; Malay *oraṅ itu*; Ur. Law. *urak itu*; Moken *mnut idup*. Temiar elaborates its demonstratives to three positions, but the order is still N + A.

When both D and A are present, as in 'that good man', there is more scope for variation, but in point of fact once again two divisions only appear: in languages where A precedes, so does D, so that there is a direct parallel to English 'that good man' (D + A + N), or an arrangement N + A + D, and in the latter case N + D + A is to be construed as N + D is A, i.e. 'that man is good'. The phrase 'that good man' becomes therefore:

Group 1	Chinese Korean Japanese	<i>nèige xǎo ren</i> <i>ku chohun saram</i> <i>sono yoi hito</i>
Group 2	Thai Vietnam Khmer Jarai Malay Ur. Law. Moken Atjeh Burmese	<i>phûuchaaj dii nán</i> <i>gù'o'i tǒt áy</i> <i>menùs l?ah nuh</i> <i>mnuih ?o'i ~?anǎn</i> <i>oraṅ (jaṅ) baik itu</i> <i>urak badji itu</i> <i>mnut amon idup</i> <i>uro'əṅ gèt nan</i> <i>lu ʎcauṅ</i>

If the adjective becomes predicative, as in 'that man is good', a change of order takes place, in such a way that the demonstrative shows the end of the noun phrase and the beginning of the predicate. This rearrangement is found in other languages in other parts of the world also. Here, each language group proceeds in the same way: it places the adjective in the predicate position at the end of the sentence, but according to its original pattern, a difference is still preserved between the two groups:

\*But a recent Scripture translation has *mónuih lu*.

	<i>Topic</i>	<i>Comment</i>
Chinese	nèige rén	xǎo
Khmer	mənùs nuh	lʔaa
Jarai	mnuih ʔanǎn	ʔo'i
Malay	oraŋ itu	baik
Ur. Law.	urak itu	badjiʔ
Burmese	di lu	`kauŋ tɛ́

At this stage, however, some differences of practice begin to appear. Korean and Japanese may verbalise the adjective and so produce a morphological change in its shape: Korean *ku saram chosso*. Japanese has also its subject markers *wa* and *ga* (が): *sono hito ga yoi (des')*. In Malay there is an alternative *baik-lah oraŋ itu*, where *-lah* serves to disjoin the adjective from the noun and to emphasise it.

The next subject of comparison is the expression of ownership. There are several patterns in the languages concerned: 1. the owner precedes the object owned and is linked with it by a possessive particle; 2. the object precedes and is connected with the owner by a particle in the style of English '*house of (my) father*'; 3. the object precedes, followed immediately by the possessor without particles of any kind. These three methods are not absolutely exclusive; they may occur together in a language with semantic differences or with different usages. In Ur. Law. for instance, '*my book*' is either *surac naʔku* or *surac ku*.

1. In Chinese the particle *di* links possessor and possessed: *wǒ di šū*, '*I of book*', '*my book*'; *rén di šū*, '*man of book*', '*man's book*'. The same pattern reappears in Japanese: *watakusi no hon*, and in Korean, *na ūi ch'aek*, all in the same relative order. These languages are again to be classed together, typologically, though of course not morphologically.

2. The second group is formed by the northern members of the southern group: Thai, *nǎŋsŷy khǒŋ phǒm*, '*book possession I*', '*my book*', and Vietnam, *sách cu'a tôi*, analysed the same way.

3. The southern languages of the southern group, including the Austronesian, dispense with a connective particle altogether, and the structure is adjectival: '*book I*': Khmer, *siəvphəu khǒm*, Jarai *mo'nu khaŋ*, '*bird I*', and Malay *buku saja*, '*book I*'. Similarly Temiar, *deeg Aloŋ*, '*Along's house*'; *deeg yeeʔ*, '*house I*', '*my house*', Atjeh *kitab lōn*, '*book I*', and Ur. Law. either *surac-ku* or *surac naʔku*.

If the possessive phrase is expanded by attributive words in the noun phrase, the arrangement is determined by the general rules as to word-order in the given language, e.g.

Khmer: nənùs thom bəy knèak nìh  
man tall three person this

'these three tall men', without possessive, and with possessive:

phtèak thmay khñom nùh,  
house new I this

'this new house of mine'. In the first instance, Urak Lawoi has

səmiya tiŋi tiga urak ini  
man tall three person this

and in the second it has

rumah baru na?ku ini  
house new my this

There are differences of order as against Khmer in these. But there is also for 'my five grandchildren' -

Khmer: cau khñom pram nèak  
grandson I five person

Urak Lawoi: cucu lima urak na?ku  
grandchild five person I

Malay does not use numeral coefficient in its

ana? limaŋ lōn  
child five I

In Malay a similar phrase is kedua anak saya ini, 'two son I this', 'these two sons of mine', for which Urak Lawoi has rumah baru na?ku ini, 'this new house of mine', as above, keeping ini to the end as phrase closer. In both Khmer and Malay there is a degree of flexibility in the arrangement. In Japanese the possessor takes precedence of the possessed in the phrase

watakusi no kono futari no musuko  
I of this two of child

'these two sons of mine', to which Chinese wǒ de èr ge ér zǐ cor-responds.

In the languages in general, predicative forms of possessives usually require repetition of the noun, as in Urak Lawoi rumah ini rumah na?k, 'house this house my', i.e. 'this house is mine'.

The following Table presents a summary of the possessive constructions in the various groups:

Group 1	$N_1 + p + N_2$	Chinese Korean Japanese
Group 2	$N_2 + p + N_1$	Thai Vietnamese
Group 3	$N_2 + N_1$	Khmer Jarai Malay Atjeh Urak Lawoi

where  $N_1$  is the possessor,  $N_2$  the object possessed and  $p$  the connecting particle, of whatever nature. A more detailed subdivision could be made by distinguishing the character of the particle involved in Groups A and B, for these, in point of fact, are not quite the same.

A point is brought out in Honey and Simmond's article, that in Chinese the possessive marker is used after a clause to make it adjectival, i.e. what in European grammars is called a relative clause: *'the book which you bought yesterday' + 'you yesterday buy's book'*, where *-s* stands for Chinese *di*. The construction rank shifts the clause (to use a London School's expression) downward to the status of an adjective, even though a phrasal one, and it precedes the noun to which it belongs because an adjective normally does this in Chinese. The other languages might be expected to show differences in their procedures parallel to their differences from Chinese, and this is actually the case. At the same time, the northern languages, which have hitherto agreed with Chinese (as seen in the preceding matrices) do not do so in this case. Perhaps the fact that Korean and Japanese are inflectional, as against Chinese, may account for the differences. Taking the given example, *'where is the book that you bought yesterday?'* -

Chinese shows:

(nǐ tswó tien maǐ)-di šū dzà y nǎli?  
 (you yesterday buy) 's book at where?

None of the northern languages agrees with this construction. Thus:

Korean. tangsin ōje san ch'aek i ōdi isso?  
 you yesterday buy book where is?

Japanese: anata ga kinō katta hon wa doko desu ka?  
 you yesterday bought book where is?

In these two languages, the whole adjectival clause precedes the noun to which reference is made, as a single adjective would do. As the languages have verbal inflection, however, and markers of subject and object, these are all involved in their normal places. The principle is the same as in Chinese; the form varies by reason of the different structures of the languages.

When the southern languages are studied, an abrupt difference appears: an equivalent to the English relative pronoun is found, at least in its general usage, though the various particles are not truly translation equivalents. They have been referred to as "attributive linking particles" (ALP), and can be shown in the examples from Thai and Vietnamese:

Thai:            nán̄syỳ lêm thǎi khun sỳy mỳa wan nǐi jù thǎi-nǎj?  
                   *book item ALP you buy yesterday stay where?*

Vietnamese: sách mà anh mua hêm-quyen ở đâu?  
                   *book ALP you buy yesterday stay where?*

In Khmer, there is again a departure, and most authors seem to have no difficulty in translating *dael* by the English 'who, which':

siəvphəu dæl lo:k msəl-ten n̄y-u-ʔae na?  
                   *book which you yesterday buy where?*

Amongst the languages that possess a relative pronoun that - whatever its ultimate origin - can be ranked functionally as such is Malay and a large number of the Western Austronesian languages, most of which are outside the scope of the present paper. In Malay, for instance, the type sentence stands:

di-mana-kah buku itu yaŋ saudara sudah beli kelmarin?  
                   *at-where -? book that wh. (you) perf. buy yesterday?*

By way of contrast, however, the mainland AN languages tend not to have any expressed relative particle at all, e.g.

Jarai    mo' nu ih boŋ hotai  
           *bird you eat-raw liver,*  
           '*the bird whose liver you eat raw*' and

Ur. Law. surac kaw beli kemari tet baji?  
           *book you buy yesterday not good,*  
           '*the book you bought yesterday is not good*'

          surac kaw beli kemari dudo? pe??  
           *book you buy yesterday sit where?*  
           '*where is the book you bought yesterday?*'

answering to the Malay example above.

In Malay, however, *yaŋ* can also be used as a kind of nominaliser, without an expressed noun, like English 'one', substituting for the noun: *yaŋ baik* 'the good one', etc. This would suggest that historically *yaŋ* is, like equivalents in many other languages, a demonstrative rather than a true relative.

In Burmese the treatment of the relative is governed by the fact that a relative clause is logically an adjective, and as the adjective precedes the noun, the relative clause precedes the entire subject of the sentence just as in Japanese. There is, however, no equivalent to the Japanese *no* (or for that matter to the Chinese *di*). The order of elements is governed by the SOV ordering of the sentence, and the fact that NP = A - N. The examples that follow are culled from various parts of Stewart (1955:55ff), retaining his transliterations.

1. Relative clause as subject:

*mə-`hma-te' `she shəya*  
*not mistake-makes*  
*'a doctor who makes no mistakes'*

*mə-θe-te' `she shəya*  
*not -die-s doctor*  
*'a doctor who does not die'*

There are also certain markers which can replace an unexpressed noun, chiefly *tha* as in *mə-hou?-tha*, 'what is not true'.

2. Relative clause as object: as might be expected, such a clause is set bodily before the verb, in the grammatical position of O of SOV - and, as in most SEA languages, S as a pronoun may not be expressed, so that we get, e.g.

*`miŋ `pyɔ-that mə-youŋ`-phu*  
*you say-what not-believe*  
*'I don't believe what you say.'*

The Subject 'I' is not expressed; the negative is a discontinuous *mə...phu*. A longer example (Stewart, p.58) is:

*mya''myolwiŋ `yetε' ]wele?kau? shotε' wu?thus''la*  
*Mya Myo Lwin write Golden-Bangle call story-?*  
*'What about the story "The Golden Bangle", by Mya Myo Lwin?'*

3. Relative clause as oblique case in sentence:

*θu `θwa-ði' myo'-go cun-dɔ `gaun-swa θi'-ba-ði.*  
*he going town-Obj. I well know*  
*'I am quite aware that he is going to town.'*

Here the entire clause which is the object of 'know' precedes the main verb, as an object should. In other cases a similar result is obtained with oblique phrases, as in

θwa-jin-de          neiyagou  
*he-wants-to-go place-to*  
*'the place where he wants to go'* (Cornyn:130)

It is desirable to say something more about numeral coefficients at this point in the treatment of NP constructions in the various languages. The use of such definers is common even as far east as Polynesia, but they seem to have their centre on the Asian mainland. Yet the picture that has been establishing itself throughout the earlier part of this paper breaks down at this point. As much has been written on the subject, it is not proposed to illustrate NCs here in any quantity. They are described in parts of Honey and Simmonds' paper. All that is necessary here is to extend the discussion to cover the new languages introduced here. The abbreviation used will be NC for 'numeral coefficient'. Two divisions of NC occurrence must be distinguished: some languages use the NC apart from the numeral construction, along with adjectives where no N is present (this is a connection with Malay usage of *yaŋ*, although of course not historically so), e.g. Thai *lēm yǎy nǐ* 'item large this' for 'this large one', like Malay *yaŋ besar ini*; others do not do this. This type will be labelled AC for 'adjective coefficient'. When a noun is present, the adjective coefficient is not always required. In Khmer the usage of NC is more limited than in some of the other languages. Jacob states that "In certain kinds of counting the NC follows the numeral in close conjunction; in other cases the numeral follows the noun also in close junction." She also adds (p.330) that "any word that occurs following a numeral is held to belong to the category of NC. Many word forms belong to both categories of N and NC".

In the Jarai group, Radé shows NC or its absence: *namboh ceh*, 'six jars' (Malay *tempayan enam buah* (N + num. + NC)); *pluh dua čô mnuih*, 'twelve man', and Jarai, *mo'nu ha dro'i*, 'bird one body'; *sa čô mnuih*, 'one man'; *pă mta mno'ŋ*, 'four face thing' *mta pă*, 'four (such things)'.

In longer phrases the uses in the different languages may be exhibited more briefly:

Chinese	D + [n + c] + A + N
	<i>jè liǎŋ-been dā žū</i>
	<i>this two large book</i>
Thai	N + A + [n + c] + D
	<i>nǎŋ-sǎy yàŋ sǎŋ lēm nǐ</i>
	<i>book large two-item this</i>



Vietnamese	[n + c] + N + A + D hai quyên sách lớn này
Jaray	N + [n + ca] mo'nu ha dro'i
Malay	[n + c] + N + A + D dua buah buku besar ini
Khmer	N + [n + c] neəkniɔn pi ru:p <i>writer two person</i>
	N + [n + c] chə: bæy daəm <i>tree three trunk</i>
Urok Lawoi?	N + [n + c] burok sa ikol <i>bird one tail</i>

There are similar uses in the other mainland Austronesian languages, e.g. Radé: *phuŋ pluh čô mníe êra*, 'ten young women'; *êma čô hlám phuŋ*, 'five of the young women'; *êma čô thao min*, 'five of them were foolish'.

In summary, the occurrences of numeral and adjective coefficients can be set out in the diagram:

	NC	AC
Chinese	+	+
Korean	+	-
Japanese	+	-
Thai	+	+
Vietnam	+	+
Khmer	(+)	+
Jarai	+	+
Malay	+	(+)
Burmese	+	-

The AC is bracketed for Malay because there the ALP *yaŋ* takes the place before the adjective which functions as N; otherwise there is no connective between N and A.

The only other NP construction (from the western point of view) of which mention will be made here is one that in eastern - and African - languages is often regarded as a VP construction. This is the expression of locality whether place or movement. In English the category of the

preposition, 'ON the table', 'IN the room', 'give it TO me', etc. comes into play here. In the languages here under study, some use what are really verbs, 'speak give him', for 'say to him'. The so-called isolating languages seem to have a tendency to this usage, and in the Oceanic languages prepositions not unusually prove to be disguised verbs, one in one language, the other in another, or even capable of indicating tense. Not all these languages actually do this; Urak Lawoi? is one that does not - but then this is Austronesian. Here we find, e.g. kami melaw jaŋan kunja, 'I say to him'. The vocabulary glosses jaŋan as 'with', 'and'. This is still not 'to', which is de?, as in kami tete na? pi de? patay, 'we are not going to the bush'.

In Chinese these usages have been classed in several ways - first, as a coverb, "an occasional function of certain functive verbs": nŷ gěi ta bān jwōdz, 'you give him move table', i.e. 'move the table for him'; wó yùŋ kwaìdz chŕfan, 'I use chopsticks eat food', i.e. 'I eat with chopsticks'. In some cases the word order suggests that the verb is being used adverbially and so becomes final: tā dàu jàŋhǎi chy, 'he reach Shanghai go', i.e. 'he is going to Shanghai'. In other cases the second verb becomes a coverb immediately following the first and these have been called "postverbs", as in tā sùŋ gei wǒ yìběn fū, 'he send give me one book'. i.e. 'he sent me a book'. Again, a resultative verb may be used: tā chŕ wán le fàn le, 'he eat finish past rice finish', i.e. 'he has finished his meal'.

Another language in which this type of category transfer (again from a western viewpoint) commonly takes place is Khmer, and some examples as given by Jacob (op. cit. p.77) follow. She, however, distinguishes major and minor verbs, and in the present section she says, "in past sentences (i.e. earlier examples in the book) it has been possible to pause and begin a new phrase after an independent noun construct or an adverbial construct. In sentences having two verb constructs a phrase may end immediately before the second verb, e.g.

vì:ə tèñ siəv phōu nì (-i) ?aoy khñom, 'he bought these books for me ('to give me')" \*

Other examples on the following page include:

khñom yò:k sɔmbot(r) tǔu pos(te)

I take letter go post

'I am taking the letter to the post-office.'

vì:ə trələp pì: tì: nìh tǔu phtəh

he turn from place this go house

'He is returning home from here.'

\*Some modifications of Jacob's spelling are used wherever the book is quoted here.

khñom nðəm ko:n cěñ pì: sa:la: rìən  
*I lead person child from room - learn*  
 'I led the child out of the school.'

It is not necessary to multiply examples of this usage in which the action is analysed differently from the European languages, and indeed differently from any language in which case relationships are formally recognised.

<i>Feature</i> <i>Language</i>	AN	NA	DN	ND	diS	ALP	POS-	COEFF
Korean	+	-	+	-	+	-	-	+
Japanese	+	-	+	-	+	-	-	+
Chinese	+	-	+	-	+	-	-	+
Vietnamese	-	+	-	+	-	+	-	+
Thai	-	+	-	+	-	+	-	+
Lao	-	+	-	+	-	+	-	+
Khmer	-	+	-	+	-	-	+	+
Malay	-	+	-	+	-	-	+	+
Rade etc.	-	+	-	+	-	-	+	+
Atjeh	-	+	+	-	+	-	+	+
Urak Lawoi?	-	+	-	+	-	-	+	+
Burmese	+	+	+	-	-	-	+	+

SUMMARY TABLE I

KEY:

- AN = adjective before noun
- NA = noun before adjective
- DN = demonstrative before noun
- ND = noun before demonstrative
- diS = di (= 'of') segment, or other language equivalent
- ALP = attributive linking particle
- POS- = no possessive connecting link
- COEFF = numeral coefficients present

### 3. THE VERB PHRASE

#### 3.1. INTRODUCTION

Honey and Simmonds did not include a study of the verb phrase in their original paper, but suggested that such a study would be a useful widening of the typology. The addition is therefore made in the present paper. As the number of languages has been increased, the extension of the themes covered is likewise desirable. Certain parts of the themes were treated by Glazova (1967:257-75). She, however, used fewer languages than are envisaged here, viz., Vietnamese, Thai, Lao and Chinese only. Moreover, her treatment is confined to the perfective aspect of the verb, positive and negative.

In all cases except the formation of derived forms such as causatives, the functional indicators in those languages are free forms. This would follow from the isolating nature of the languages. The only possibilities of variation in isolating languages is the order of the indicators and how many features are included in a given language. The subgroupings within ND will depend on these variables. The available arrangements seem to be:

1. S + i + V
2. S + V + i

where S = subject, V = verbs, and i = indicator; and in some cases a third, with repetition of indicators,

3. S + i + V + i.

The examination made will arrange the languages into what proves to be two contrasting groups or possibly three. In the scope available at present it is impossible to treat the variations in all the details Glazova has given, but it appears that order (1) is found in Malay, Rade and other AN languages, Thai and Vietnamese; order (2) in Chinese, which also in certain verb types shows order (3). The position of the negative is also important. If this is indicated by n, the orders will be formulated as

1. S + n + i + v
2. S + n + V + i
3. S + i + n + V + i.

Practically all the languages seem to share the situation in which a verbal expression unmarked morphologically for time can be construed as either past, present or future according to context, unless and until otherwise defined, and this whether aspect-tense differentiation is available or not.

The treatment of VP will be arranged under the following heads:

1. **Inflecting Languages**, represented by Korean and Japanese in the north and including Burmese in the south, for this language allows of agglutinative compounding under certain conditions. None of these languages belong to either the Mon-Khmer (MK) or Austronesian (AN) groups. Korean and Japanese are both to be ranked as Altaic languages, and Burmese belongs to the Sino-Tibetan family.

2. **Languages with Minimal Inflection**, occupying a sort of middle position between the inflecting and the isolating languages. The AN group belong here; Malay and the north-eastern relatives: Jarai, Rade, Cham, etc. and on the far west Urak Lawoi?, Moklen and Moken, with Atjeh in the south-west. The position of Atjeh has been much disputed: it is accepted here as a member of AN. Apart from AN, there is also Khmer (MK) to be fitted into this group, and by presumption other MK languages which are not included in the paper. For some of these see papers by Robins and Jacob in *Lingua*, 1965, Part II. Khmer and others of these languages have some possibility of inflection, chiefly through infixation, as Jacob's paper brings out.

3. **Isolating or Uninflecting Languages**. Chinese is probably the best known example of these languages, although it has historical connections with Sino-Tibetan as a whole. The Tai group, of which Thai and Lao are the representatives here, must also be included.

Within the subject, "Verb Phrase", the elements to be considered are tense, aspect, mood and voice. Not all these are distinguishable in all the languages: in Chinese, for instance, tense seems to be a redundant term, and the same is true of other languages also. An attempt will be made to collect the evidence for each element under the one heading.

In a full treatment, all the MK languages would need to be studied individually and the results tabulated. This is not feasible at present, especially as some, such as Sakai and Senoi, are imperfectly recorded.

In spite of what is said here about arrangement, the inflecting languages stand so isolated that it is convenient to take them together as a geographical group and deal with all the structural elements in them in sequence. They can then assume their proper place in the subsequent tabulation.

One feature that is common to all the groups is the absence of person indicators in the verb, even where suffixation is present. If the actor is represented by a pronoun it is usually unexpressed unless emphasised. This happens regardless of the linguistic group concerned. Another common feature is the indication of social rank or politeness, which may show itself in the verbal inflection or in a pronoun where the verb is uninflected. In Korean and Japanese it will appear twice if the

subject is expressed, and may appear a third time if there is an expressed pronoun object also.

### 3.2. THE FULLY INFLECTED LANGUAGES

The system of social ranks common to the northern countries enters into the conjugation of the verb. Similar social ranking systems are found in Malay, Javanese and other societies, and to a degree they affect the languages in various ways, but as the two northern languages are the only ones which express social facts in the morphology of the verb by suffixes of a rather elaborate nature, they are also the only ones in which these social facts are also involved in the grammar of the languages. Nor is it only a matter of social rank: there is the matter of politeness apart from social rank. People will use the polite forms across ranks, it is true, but also between themselves within the same social rank so that these simply become polite forms of the verbs, not rank forms as such - social rather than political. If the two overlap, that is incidental. Korean is rather more elaborate than Japanese in this instance, and both are more elaborate than Javanese or Bali. Korean can express up to five degrees of politeness, whereas Javanese does not go beyond three. In both instances the politeness is expressed through the use of special vocabulary, but in Korean and Japanese it is expressed chiefly through grammatical forms of the verb, though there may also be special nouns or pronouns. It is, then, the grammar of politeness that sets the two northern languages farthest apart, rather than the socio-linguistic fact of rank and its expression.

In the present paper, these elaborated honorific forms are involved only so far as they appear overtly in the verbal systems. In both languages the system operates by (1) replacement of a "common" verb by an honorific or humble form, and (2) by a grammatical change in the verb ending - sometimes both at once. Korean has both honorific and polite or rank forms. The operations to be distinguished are, then:

1. replacement of word by word;
2. grammatical indication with or without (1).

In the other languages to be considered the tendency is to idiosyncratic (syntactic and lexical) expression rather than grammatical forms. Thus the factors to be considered are:

1. rank v. politeness - forms and special words;
  - ii. moods and tenses in each type of verb;
  - iii. methods of person indication.

As in these languages person is not marked in the termination of the verb, it is necessary to look at pronominal forms and other ways of

expressing person as acting or addressed. This can be done only very partially in this paper.

(a) *Korean*

Korean is the most northerly in geographical order. It has a number of verb classes, marked by vowels which serve to indicate moods. As a rule person is not marked. To the verb stem or base is added an honorific if called for, then tense marker, mood (or class) marker, and finally the appropriate politeness marker. The result is a set of four possible additions to a verb stem:

Stem + Honorific + Mood + Politeness.

Mood is here being interpreted to include certain sentence medial connective markers as in Japanese. These are equivalent to some at least of the sentence-medial (SM) forms of the Papuan languages in many regions of New Guinea. The morphemes are:

honorifics - here {-*usi*-}

tense - present: -*ə*-

past: -*əss*-

future: -*kess*-

as with *mæg-*, 'eat': *məgəs(i)*- 'someone honorable eats'; present *mægə-*, past *mægəsə-*, future *məgəss-*. On the past stem may be built by combination with other suffixes a pluperfect, *mægəsəsə-*, and a future perfect *mægəsəkəsə-*, which are literary rather than colloquial forms. A useful Table of combinations is found in S.E. Martin, *Korean in a Hurry*, p.129. The various combinations of this type build up some 500 different forms.

There are also verbs honorific by nature, not used except to or about one's superiors, and there are compound verbs, parallel in each case to similar Japanese usages.

These rank forms must be distinguished from polite forms, which are used in rather different circumstances. As in Japanese, there are honorific nouns as well as verbs, and even honorific case endings. The polite suffixes, which come as finals in the VP, are independents of the honorifics, which come immediately after the stem, so that 'he laughs' will be *usə* as a verbal form, *usəyo* as a polite form, and *usəsəyo* as an honorific polite form, involving both -*ssə-* and -*yo*. On the other hand, while the ordinary verb 'give' is *tʃuə-* (polite form), the honorific verb is *turya*, polite *turyəyo*. This system also is paralleled in Japanese.

In fact there are five styles of speech available for the Korean who is perfectly enculturated; the forms differ as between statement, question, common and proposition: for statement, -*sumnida*, -*umnida* (formal),

-nunda, -nda or -da/-ta (plain), -ne (familiar); authoritative -so, -uo and intimate in -u, -a, -e - and to this latter may be added -yo to soften them in terms of politeness.

The expression of the negative is peculiar to the language, and differs thus from Japanese. There are two forms: a prefix an(i)- or a suffix -ji-anso. Thus '*does not eat*' may be an-məgəo or məg ji-anso. Japanese lacks any equivalent to the prefix form, which is stronger than the suffix and has the overtone of dislike or lack of habit, which Japanese does not distinguish. The an- root is really '*not be*'; -ji is a gerundial ending, so that ji-anh- is '*...ing not be*', and it is here that tense is marked. There is also the potential mot followed by ha- "causative", so that capsu-, '*eat*' > mot-capsu- (>moccapsu-), '*is not able to eat*'.

In addition, there are linking and disjoining forms of the verb which cannot be expanded here, but something must be said about verb compounds, which are much as in Japanese but differ from the Chinese methods of construction.

### (c) Japanese

In Japanese the same general typology is seen as in Korean. Honorifics are present, and they are either nouns, verbs or verb-endings. The first is not considered here (although it includes forms used pronominally); the other two are relevant.

The Japanese verb does not mark person: in some honorific verbs, person reference is inherent in the verb itself. Passives, causatives, passive-causative, potential alternatives and desideratives are marked by suffixes which are independent of tense and aspect indication. There are three conjugation classes, distinguishable by four stems applicable to root form, negative base, certain present, and conditional base respectively. These will be found set out in any Japanese grammar. What will be called here "themes", marked "certainty" and "probability", along with the tense scheme of the two may be set out as follows:

THEME: CERTAINTY		THEME: PROBABILITY	
kas-u 'lend'	Present	kasau kasō 'probably lends or will lend'	
kas-ita 'lent'	Past	kasi-te-rō 'probably lent'	

These are indicative forms. There is also an imperative: kase, '*lend*'; a conditional, kase-ba, '*if...lends*'; kasi-ta-ra(ba), '*if...had lent*'; and a concessive, kasi-ta-redo, '*although...has lent*' - also kasi-ta-keredo.



All these themes and tenses can be negated by means of a negative verb, *naka-*, becoming *-nai* in the present. The only exceptional form is *-mai* in the probable present-future. Alternative is marked by a suffix *-tari*: the inflection of a neighbouring verb enables tense and other features to be made clear, e.g. *nai-tari warat-tari site orimas*, '*sometimes crying sometimes laugh-doing (I) am*', i.e. '*sometimes I am crying, sometimes laughing*'. These forms are historically derivable from *-te ari*, the participle *-te*, combined with an older past tense of *aru*, '*be*'.

Further historical analysis of these forms is possible, but not relevant to the present purpose in outlining typologies of present-day SEA languages.

Honorifics are indicated in a number of ways in Japanese -

(i) by means of prefixes: *go-*, *o-*, *mi-*, whose usages are relatively fixed. They are normally applicable to nouns, but *o-* is sometimes prefixed to a verb to provide a polite imperative: *o mati nasai*, '*please wait*'; *o mati asobase*, '*vouchsafe to wait*' is also possible.

(ii) by means of a special set of honorific or humble nouns: these are not relevant here.

(iii) by means of a set of verbal suffixes: (1) *-mas(u)*, which is conjugable, e.g. *ki- mařō*, '*will probably come*', in ordinary polite conversation. There is a corresponding negative, e.g. certain present: *mase-nu*, past *masen desita*: some forms are not used; (2) a somewhat stronger auxiliary *gozarimas(u)* is also in use.

(iv) a set of honorific or humble words is found: an ordinary verb is replaced by one of these, or the causative or passive (potential) verb is used in place of the simple form:

	GENERAL	HONORIFIC	HUMBLE
'give'	yaru	kudasaru; tamau	ageru
'go'	iku	o ide nasaru	mairu
'say'	iu	ossayaru	mōsu

The passive potential variety is shown in *kikasite kudasai*, '*causing to hear condescend*', '*please tell me*'; *nani to ossyaimasita*, '*what did you (hon.) say?*'.

It is clear that these conjugation types in Korean and Japanese are related to each other, and that their roots lie in sociolinguistic causes, which are more elaborate here than in the other languages-areas. Their differences from those of the other SEA languages (Austronesian as well as Mon-Khmer) will appear in due course.

### 3.3. LANGUAGES WITH LESS INFLECTION

Korean and Japanese obviously stand apart from the other languages here treated. At the same time, however, certain chords are found in each - or most of the languages. These are the expression of either social rank or politeness or both. Such factors operate in Chinese, and indeed in most of the SEA languages no matter to what linguistic group they belong. The same social factors seem to have been operative throughout. It does not matter whether a language is of one type or another: any type may possess means of indicating honorific situations or their opposite. Throughout SEA pronouns as a rule are hardly to be regarded as a separate speech category from the noun. They are in general nouns, as are such expressions as '*Your Highness*' in English and '*Exzellenz*' in German. This is true of Thai, Vietnam and related languages. Neither does it matter whether the language is inflectional or not. It seems therefore suitable at this point to resume the method adopted in Section 2, of setting out the occurrences of similar or related structures in the languages *seriatim*, rather than taking each language separately. This was done for Korean and Japanese because in them the differences from general SEA are qualitative, not merely lexical.

In the Austronesian (AN) languages Malay is taken first. It is the best known, but it will appear that the lesser AN languages do not necessarily agree with Malay practice or even structure. They have all in greater or lesser degree been influenced by the MK languages. An unpublished paper by Ernest W. Lee on 'Southeast Asian Areal Features in Austronesian Strata of the Chamic Languages', issued by the Summer Institute of Linguistics and the University of Texas at Arlington is one of a number that set out these divergences very clearly.

In Malay the chief stress is on aspect rather than on tense. Tense indication plays a very minor role and is made only when needed: to this extent the language agrees with Vietnam and other non-AN languages. Either time is not formally marked (except by any required adverb) or an auxiliary verb is used which in an independent form has another meaning. Thus *nanti*, '*wait*' is used colloquially in Peninsular Malay at least, for the future: '*I wait to do it*' means '*I intend, want to or shall, do it*'. In Indonesia *akan* is general, in Malay more literary, and this is a preposition, '*to, for, concerning*'. There is also *hendak, dak*, '*wish*'.

The aspect markers are divided into two main groups: perfective and imperfective -

PERFECTIVE	IMPERFECTIVE
sudah	sedang
telah	tejah
habis	lagi, masih

All these precede the verb to which they refer: *saya sedang makan*, 'I am eating'; *saya sudah makan*, 'I have eaten'. The differences between the various words in the same column are stylistic rather than semantic: *sudah* and *telah* are interchangeable as far as meaning goes, but there are differences in style and level of speech. *Sedang* and *masih* are commoner than *tejah* and *lagi* - but again this is not a strictly linguistic fact in the normal sense of "linguistic".

In these languages, a certain amount of inflection is possible. A set of prefixes and suffixes - in some also limited infixation - is possible, especially in the formation of the causative. In the AN languages this provides not only a link between them and PAN but also a stepping stone to certain of the MK languages. The line of demarcation between methods of indicating the cause of an action is determined by the presence or absence of inflection in this group of languages. Inflection on any large scale usually lifts the language out of the ND class and sets it into one of the other classes, which for this region is always the EV class. Yet within the ND class itself there can be more than one way of indicating causation. Thus in Jarai:

ih pŭ bro'i prŏŋ  
you raise make good

a verb *bro'i*, 'make' occurs. But as such it is a causative, not a "manufacturing" verb; its literal meaning is 'give', as in

bro'i ko' go'mo'i bŏŋ huă djŏp rim hro'i  
give to us food enough for today

In Jarai *bro'i* is used in a manner parallel to Khmer *qaoy*, 'give', which also has a causative function. This does not happen in Malay or in other PAN languages. The form *bro'i* represents PAN\* *bəyay*, 'give'.

In Ur. Law. the verb 'make' appears also as a causative: *buwac* (Malay *buwat*, also PAN), as in *buwac bri badji?*, 'improve it, make it good'; *buwac bri pañak*, 'lengthen it'. The remarkable point here is that 'give' (*bri*) is actually combined in the phrase with *buwac*; 'make it give length', etc. seems to be the idea. Two types of causative formation have been joined. The other auxiliaries of mood and aspect agree fairly well with Malay: *ini kami buwac baleh*, 'this I can do' (Ml. *ini*

saya boleh buwat) with slightly different word order. The negative tet goes with the auxiliary: ini kami buwac tet baleh. Tense markers are more definite than in Malay, and precede the verb: ña teŋah mareh, 'he is coming'; ña na? mareh, 'he will come' (in which na? represents Malay hendak, 'wish' and mareh is mari, PAN \*mayi). But the past marker follows the verb, i.e. it is treated as an aspect, not a tense: mareh da?, 'he has come'. In lapal da?, 'is hungry' the marker indicates quite definitely a "state assumed" force, while kami pi da?, 'I went' shows past action: both ideas combined. There is an alternative method of expression: buwac suc da?: 'he has done it', where suc appears to answer to Malay sudah. Although tet is the normal negative, the future has timaw, as in ña timaw buwac, 'he will not do it'; bilarak timaw bri kañaw makat buah puhok kayu itu, 'you must not eat the fruit of that tree'. Yet timaw is not dehortative: jaŋan is shared with Malay in jaŋan (-na?) buwac, 'don't do this'. This jaŋan, however, has a homonym meaning 'to': ami melaw jaŋan kuña, 'I said to him'. A full analysis on this point is not available.

In the other languages yet to be considered, the tendency is to find both aspect and tense expressible separately and in some cases a difference of arrangement is found: tense particles come before the verb, aspect particles (=A) follow it. Thus in Radé,

Aspect phrases: S - V - A    ñu    ɣà    leh  
   he do perf. = 'he has done it'.

Tense phrases: S - T - V    ñu    so'raŋ    ɣà  
   he fut. do = 'he will do it'.

In Radé as in Ur. Law. negation takes precedence over tense: it is more important to know if something did not or will not happen.

The remainder of the Radé forms may be given at this point, as diagnostic of the general type of this subbranch of AN. Radé shares the placing of the perfective after the verb, while the others precede it, as in Urak Lawoi?. There is no marker for specifically present time or indefinite time; perfective, negative, imperative and permissive are markedly particles following the verb, thus

Phrase patterns: S + neg<sub>1</sub> + t + V + neg<sub>2</sub>  
   ñu    amào    so'raŋ    čih    ôh  
   he not shall write not  
   'he will not write it'.

or a substitute negative:

nũ amão so'raŋ alo' mao  
*he not shall again get*  
*'he will not get any more'.*

Present (general) is unmarked in all the languages.

Past + leh: ñu nà leh, *'he has done it', 'he did it'.*

Future so'raŋ + : ñu so'raŋ nà, *'he will do it'.*

Mood imperative - + bẽ' (cf. Atjeh): nà bẽ', *'do it'.*

potential - du'i + V: ñu nà du'i, *'he can do it'.*

permissive - + boh: ih ðǒ'ŋ boh, *'you may eat it'.* Malay boleh.

Purpose clause is joined to main verb by či'aŋ ñu *'want'*, used as a main verb in, e.g. ñu ci'aŋ ðǒ'ŋ, *'he wants to eat (ðǒ'ŋ)'*. A further example of subordination appears in nà či'aŋ d'un, *'cause that you enter, make you enter'*. A reciprocal form is marked by bi: ya di ih bi lač hdǒ'ŋ diih? *'what do you say of each other?'*, while the reflexive is marked by pô: ih amão du'i bi mtlaih ih pô oh, *'you cannot save yourself'*.

The Atjeh language fits the AN pattern in general, although there has been much discussion as to the real history of the language. In regard to this some notes from Uhlenbeck (1967:876) may be of use in summary: "(Cowan) thought it possible that Atjeh and Cham perhaps together with Selong, a language spoken in the Mergui Archipelago, constitute one subgroup related to the Malayo-Polynesian language", and earlier (p.859), "Since 1940 no new contributions of any great importance have been made to our knowledge of Achehnese after the fundamental work of Snouck Hurgronje (1893, 1900 and 1906)". Cowan's article (1948) compares Atjeh with the MK languages as well as with Cham.

In point of fact, the language uses an infix which rather suggests MK practice; but its verbal system is not based on aspect but quite clearly on tense. The common AN causative prefix \*pa appears as pö-, e.g. lömah, *'see, look at'*, caus. pölömah, *'cause to see'*, but there are also derivatives in -mö- and -öm- (Cowan 1948:439). There is a perfective marked by postposition, however, as in

oh nō-pö-djö-et laŋet ŋon bömu le Potö Alah  
*when create heaven and earth PERF Lord God*  
*'when the Lord God had finished creating heaven and earth...'*

Similarly, the status of na, *'present time'* is somewhat uncertain - it corresponds to Malay ada, *'be, exist'*: na dua musim, *'there are two seasons'*; it appears in djinoe na manton, *'there are still (such and such things)'*, of which the past form is djinoe hana le, *'now there are no more'*; *'in a short time'* is hana treb le, *'not long PERF'*.

Person can be marked in the verb by short forms of the pronouns, which in the singular become really double expressions: kee ka ku dja?, 'I PAST I go', i.e. 'I went'; and these can appear without markers: ku tob, 'I prick'; lo:n plueŋ, 'I go away' - kee and lo:n are both different rank forms of first person singular pronouns. Future time is marked by tem: lo:n tem lo:n dja?, 'I will go', and past time by ka: ku ka lo:n dji, 'I saw him'; udjöen ka dji-toh, 'rain has fallen'. Negation is marked by hana: lanet hana lömah le, 'the sky is no more visible'. This marker precedes the tense markers when both are used: han tem, 'will I not', lit. 'not will'. The past marker ka can also be a marker of condition attained: pöe töŋku ka nö mökawən?

? sir PAST RESPECT married?  
'Are you married?'

to which the answer is simply: ka, 'Yes, I am'. Where there is no value in the time article as such, it can be omitted: lo:n tingai di... 'I am staying at'.

The only marking of mood seems to be between indicative and imperative. In the latter, no marking is necessary in the positive: preh, 'wait!'; in the negative the marker is be?, which is probably an MK element, already mentioned in Radé, Jarai.

Although this account is brief, it does seem to show the uncertain status which Cowan ascribed to the language: there is evidently a strong MK element in it, and the problem is linked with the general position of the mainland AN languages. As this is not a paper on genetic connections, the question cannot be examined further here.

Outside the AN languages, there remain a certain number of languages in which remnants of inflection are found. These are discussed by Shorto (1963:52ff) and include Palaung, Riang-Liang, and Praok amongst the northern MK languages. An article by Jacob in the same volume (Jacob 1963:62ff) discusses similar phenomena in Old Mon and Old and Modern Khmer. In the modern languages there are such derivatives as so:m, 'ask' > smo:m, 'beggar'; chù:əñ, 'conduct business' > chmù:əñ, 'businessman': here the infix -m- serves to produce nouns of agent. There are also -n-, 'making a utensil', another -m- which is causative (slap, 'die' > somlap 'kill'), a noun-forming -mn- (kaet, 'be born' > komlaet, 'birth') and certain less common infixes (Jacob 1968:183-4). Khmer can therefore be regarded as a language in which limited inflection still takes place, as in the languages of the northern area, though indeed still more limited than theirs.

The Khmer verb may be treated in this intermediate section of the paper, before the isolating languages in the strict sense of the term. It is difficult to decide whether to class a marker as a tense or an

aspect sign: usually it is safer to look at the resulting state than at the action itself, for the time of the action is often determined simply by an accompanying adverb as in *vì:ə tə:w phsa:r*, 'he go market', which may be either in process, intended, or accomplished. Once a context is given to an utterance there is no need to indicate details of time. The bulk of the precision markers precede the verb; a few follow it and there are some discontinuous forms, although in these the simple forms can also be used. A second element such as *tè:* usually adds emphasis. In the following lists Jacob's transliterations are in general preferred:

	PRECEDING	FOLLOWING
<i>ba:n</i>	'get': perfective	<i>tè:</i> emphasis, especially <i>mìn...tè:</i> .
<i>dael</i>	'already': perfective	<i>haəy</i> complete
<i>niw</i>	'remain': permanent state	<i>rù:əc</i> 'finished': often
<i>nɨŋ</i>	futurity: state unchanged	<i>ru:əc haəy</i> .
<i>mìn</i>	negation + <i>tè:</i> .	<i>laəy</i> emphasis after
<i>kompun</i>	continuation (Thai <i>kamləŋ</i> )	negative: 'not at all'
<i>cop</i>	'desire', near future (Thai <i>ca</i> )	
<i>tro:v</i>	'necessity'; as 2nd verb, 'rightly'	
<i>?aoy</i>	'give, let, allow, cause'	
<i>kaət</i>	'be born': ability	

Actually there is quite a long list of other such markers set out in the grammars, many of which can also lead an independent life: it is a matter of phraseology rather than of single words. A few can be briefly illustrated in contexts drawn from various sources. Thus: *vè:lì:ə dəl haəy*, 'the time has arrived'; *khñom ba:n mak*, 'I have arrived'; *khñom mìn tə:w tè:*, 'I am not going', 'I did not go'. *Ba:n* is one of those which can act as a full verb. Gorgoniev gives the example *nè:ək ba:n prak khæ ponma:n*, 'you get money month how-much?', i.e. 'how much a month do you get?'. There are also *o?:puk nìw knoŋ bəntùp*, 'father is in his room'; *khñom coŋ tə:w pos(te)*, 'I want to go to the post-office'; *k̀əət tro:v niw bəntùp k̀əət*, 'she has to stay in her room'; *l̀ə:k prəñap t̀əw na:*, 'you rush (go) where?'; *khñom ba:n mə:l kom nùh*, 'I have seen that film'.

### 3.4. THE UNINFLECTED LANGUAGES

For these languages the Chinese verbal system provides the most suitable beginning. It is actually somewhat apart from some of the others and, of course, widely apart from the inflected Korean and Japanese. Although the features listed earlier are present - all languages must be

able to express human requirements - the usual divisions of Chinese verbs are into different categories, which are really semantic-syntactic groupings, in which aspect is expressed rather than time.

With the Chinese system of aspect markings, however, another combines which may be spoken of as phrasal verbs, and these are syntactically ordered arrangements of independent verbs whose mutual relationships within sentences determine the actual value of each of them in a given case. Thus one may speak of "stative verbs" which include items which in English are classified as adjectives - like '*large*' in '*this house is large*' as against '*I want a large house*', and '*always*' in '*he is always talking*'. As against these are found "functive verbs" which function independently, denoting an act or event (including '*having*' a thing). Then there is a class of "coverbs", which are not usually paralleled in European languages but are sometimes found in African and Oceanic languages, e.g. instrumental '*with*' in '*I eat with chopsticks*', which becomes '*I use chopsticks eat rice*'; '*he is going to Peking*', which becomes '*he reach Peking go*'. There are also other special groups such as "post-verbs", including *dzaì*, usually translated into English as '*at*', but really '*to be at (a place)*'; "resultative verbs" (as will appear in other languages in this paper also), e.g. '*I can't understand this book*', becoming '*I look not understand this book*'. There are auxiliary verbs, just as in other languages. For the moment, attention will be limited to functive verbs, to indicate how aspect/time relationships are marked in this language.

In time relationships, present time is not marked unless attention is being drawn to some exact time. This is common in all these SEA languages of the uninflected type, and has appeared already in some which do have inflection. So there is

jeì rén dzwò syà  
this man sit down

which may be stative, '*this man is seated*', or active, '*this man sits down (after doing something else)*', or it may be regarded as a present or past act. In this setting the negative is *bú: rén bú dzwò syà*, '*the man does not sit down, is not seated*'. There is also a negative *méi* which is used in certain grammatical situations such as questions, and with the verb '*have*'.

For the very brief treatment of the aspects, which is all that can be attempted here, Yuen Ren Chao's *Grammar of Spoken Chinese* serves as the basis. Some of his examples are used, but the system of transliteration differs from his. The aspects in his account of the language are chiefly:



1. Perfective, marked by -le, as in wǒ hō-le, 'I have drunk', i.e. 'I did drink and now have finished doing so'; nǐ gǎn chē-le, 'you drove a carriage'. It will be noticed that this marker is a toneless enclitic, and Yuen Ren Chao holds that it does not derive from lái 'come', but from liǎo, 'finish', and this seems to be quite correct. This verb, as tonic, finite form, may itself be followed by its atonic derivative, as in nǐ liǎo-le ǎ, 'you have finished the business'.

2. Progressive, marked by a verb which is radically jaú, 'cause, send, put on', but may be colloquially abbreviated as far as -je or even -j, e.g. tà-yaŋ hái syé-j, 'the sun is still rising'; wǒ syǎŋ-j nǐ ne, 'I am thinking of you' - strengthened by means of the particle ne (see his pp. 801-2). A compound verb in the progressive is broken: wǒ kàn dièn-yǐŋ lái-j, 'I was watching movies'.

3. Inchoative, based on chǐ lái, 'start', as in tā-men wǎn wǎn hū jǎn ku chǐ lái-le, 'they were playing and playing and suddenly began to cry'.

4. An indefinite time past may be marked by gwǒ, 'have you ever (eaten)?', being answered by chī gwò, '(I) have eaten (it, sometime in the past)'.

All these examples show a basic arrangement subject, verb, aspect marker (SVA). This is the order also in the Austronesian Radé, Malay (with occasional exceptions, which have been pointed out). In the MK languages the preferred order is SAV, as Glazova notes in her first set of examples:

Vietnamese	tôi đã đọc quyển sách này
Thai	phôm đã an nǎŋsǎy lèm nǐi
Lao	khoi đã an nǎŋsǎy lèm nǐi
Khmer	khñom ba:n mǎel siəwphiw nih

all meaning 'I have read this book'. In all of these the order is aspect-verb, with no tense marker (in agreement with Chinese but in reverse order), while the AN languages follow a different pattern, e.g. Malay saja sudah membaca kitab ini or buku ini (Indonesia).

Moving across for a moment to the matter of negation, this makes a difference in the order of elements in perfective utterances. In Thai, Lao and Khmer the negative precedes the perfective, in Vietnamese it follows, and in Thai and Lao the tense or aspect marker may be omitted in order to show that there is no intention of completing the action - not that it just has not been completed. In Chinese the negatives precede the verb immediately in any type of sentence, so that SNVA is the logical order under all circumstances.

Some languages also present a continuative-perfective or "completive". Glazova's example will again serve: as against Chinese *wǒ kàn-le chè ben syūla*, 'I have finished reading the book', in which the perfective *le* or *la* occurs, the other languages have a different particle and a different order.

In fact, a distinction is possible between a perfective and what may be called a completive aspect. In English such a distinction is made by the use - generally - of 'up' after adverb: 'I ate it up' is something more than 'I ate it' or 'I have eaten it'. The sentence 'Cut this tree down', 'Alright, now cut it up' highlights the distinction very clearly, and also shows that a completive is not limited to a past tense: it may occur with a present ('I am cutting it up') or a future ('I shall cut it up later'). The use of the adverbial 'up' in English is contradicted by the use of 'down' in Ancient Greek - *kat-esthís*: 'I eat it up', lit. 'I eat it down (till there is nothing left)'. In Melanesian Pidgin, 'Drink up your medicine' is similarly *daunim marasin*. This is a possible subject of a separate essay (Capell 1978); the point at the moment is to show that a completive as against a perfective is possible also in at least some of the SEA languages, though methods of expression vary from language to language. The examples above show its applicability - and so does Glazova's Russian *ja dočital knigu*, *do-* indicating the completiveness. This is really the construction Glazova calls "resultative", for which Russian usually supplies equivalents by prefixation to the verbal stem. In the present group of languages a marker of result is added after the verb. In Thai and Lao *cop* or *cop léw*; in Khmer the form is *V + cop + N + haəy*. In fact she shows that two formulae are possible, one answering to Russian *do-čitat*, 'read through' and the other *pro-čitat*, 'spend time reading'. For the first set she gives:

Vietnamese	<i>dọc hết quyển sách này</i>
Thai	<i>an nǎngsǎy lĕm nĭi cop léw</i>
Lao	<i>an nǎngsǎy lĕm nĭi cop léw</i>
Khmer	<i>mǎel cop siəwphĭw mĭl nih haəy</i>
Chinese	<i>kàn wǎn-la che ben fū la</i>

For the second set she gives 'I spent time reading this book':

Vietnamese	<i>tôi đọc quyển sách này rô'i</i>
Thai	<i>phǎm an nǎngsǎy lĕm nĭi léw</i>
Lao	<i>khoi an nǎngsǎy lĕm nĭi léw</i>
Khmer	<i>khñom mǎel siəwphĭw mĭl nih haəy</i>
Chinese	<i>wǒ kàn-la che ben fū la</i>

For details, including negativisation, reference may be made to the original article of Glazova.

Some of the languages have a fairly lengthy set of particles, either before or after the verb, showing types of action: a mixture of tense and aspect forms results. In Vietnam, for instance, one finds the simple verb used without particle if a context makes clear what is intended - present, past or future event, but greater definition is always possible. The hay indicates custom: ɲu'o'i An-nam hay uõŋ nu'ó'c chè, '*Annamese are accustomed to drink tea*'. This hay is to be distinguished from hã y used with some imperatives. Process can be indicated by đạŋ or du'o'ŋ; on đạŋ (or du'o'ŋ) đọc quyền sách, '*you are reading the book*'. Past action is shown by đã preceding the verb: tôi đã ọŋ quyền sách, '*I did read the book*', and full completion (perfect tense) by có, '*have*' as in English: tôi có ọŋ quyền sách, '*I have read the book*'. An immediate past can be shown by mó'i: nó mó'i đi, '*he has just gone*', while other particles of very similar meaning are also available: rôi before the verb or xon or xon rôi after it. Future time can be left, like the others, to context or be marked by sê before the verb: nó sê đi, '*he will go*'. More immediate action can be expressed as with the past by mó'i or here also by sắp, as in nó sắp đi, '*he is just about to go*'.

Mood as known in European languages - the mental aspect of the utterance - is not marked in Vietnamese: a conjunction nếu, '*if*', can be used, but there is no prescribed construction with it. There is, however, a set of auxiliary verbs expressing various modifications of the action, that do not need to be treated here. The Vietnamese grammars will provide examples.

As in Khmer, so in Thai and Lao there are preposed and postposed particles, and both aspect and tense are present. In fact, the two can be combined in a unit, as in Thai khuan lé:w, '*should perfective*', i.e. '*should have done something*'. This form occurs even in an abbreviated response to, e.g. '*Should he have gone?*' - Khuan lé:w, '*Yes, he should*', which is very much like the English corresponding, and English also is a ND language.

In Thai markers of tense are clear, and simple markers precede the verb. The contrast is between past and future action: if a present time is involved, it is morphemically indicated only as a progressive or habitual, to be mentioned below. A verb with no tense marker is taken within a context. A folk-tale beginning Khrâŋ nỳn jaŋ mii laa tua nỳn, '*time one be donkey body one*' means '*There was once a donkey*'; then it goes on kamlạŋ daan hã ?aahăn, '*walking seek food*', and here kamlạŋ is the marker of progressive aspect, independently of time.

Future time is marked in Thai by cà? as in khun cà? sýy ?àraj, '*you will buy what?*', '*what will you buy?*'; khắw cà? sýy khỏcŋ thii nắj? '*where will they shop?*'. Past time, if marked at all, is shown by dậ

as in *phôm dâj faŋ* 'I listened' or *sŷaŋ khoŋ khou dâj phró?*, 'her voice was beautiful'. These two markers are used in simple statements where time needs to be expressed.

Aspect is marked in rather different ways: here the progressive or continuative is marked in ways not entirely synonymous. The continuation of an action is marked by *kamlaŋ* (Khmer *kompun*) as in the story text above, and in *nám khâaŋ kamlaŋ tòg*, 'the dew is falling'. The other marker can be superadded: in the actual text from which the preceding example was taken the phrase is *námkhâaŋ tòg juu bon bajmáj bajjâa*, lit. 'dew fall stay on leaves grass', i.e. 'the dew fell-and-stayed on the leaves of grass', and that is the literal meaning of *jùu* - 'stay': *khăw phuud jùu*, 'he was (or is) talking'.

Perfective state is marked by *lé:w* which follows at the end of the phrase and indicates that an action is past and completed. It is stronger than a simple past tense. It may be preceded or even replaced by *maa*, 'come', i.e. the state has 'arrived': *phôm paj maa lé:w*, 'I have been', lit. 'I go come complete'. T.G.H. Strehlow in his notes *Thai for Beginners* remarks that "although you know the word *ma* as the verb 'come', it can occur after a verb, and then it puts the first verb into the perfect tense: *paj năj maa?* 'where have you been?' - and this is lit. 'go where come'. The two elements can be separated: *raw daan maa naan lé:w*, 'we have been walking a long time.'" It is, however, possible to omit the marker altogether: Strehlow quotes Haas, *Spoken Thai*, Vol.I, p. 108, No.13:

*phôm rian pha:săa thaj nínòj mŷa jaŋ jùu naj A.*  
*I study language Thai a-little-while yet stay in America*

The SEA omission of pronouns and tense markers appears very clearly here.

Other aspect markers are possible: ability - *dâj*, homonymous with *dâj* 'past time' is one of these markers; there is also a *dâj* = 'get', which would seem to identify with the ability marker:

*dâj nən mâj dâj*  
*get money not can*  
 'Can't you get any money?'

The negative *mâj* would be linked with Chinese *méi*, so that Thai *mâj dâj* = Chinese *méi yǒu*, 'not have'. Duty is expressed by *khuan*, as mentioned earlier, when it was noted that this can become perfectivised: *khuan paj lé:w*, 'should have gone'.

In Lao the general picture is precisely as in Thai, and very few examples are needed. It is possible here to express a progressive or immediate future by the phrase *kamlaŋ ca?* - not *cà?*: Lao tones often

differ from Thai -- as in *khôôj kamləŋ ca? paj sýkhšɔŋ*, 'I'm going shopping'; *kamləŋ* and *jūu* can be combined, as in *phəən kamləŋ hian nəŋsýy jūu*, 'he is studying'. Here a continuous process rather than a short act still uncompleted is the theme. The two languages agree in using *léew* as perfective sign: Lao has *phəən paj báan léew*, 'he has already gone home'; *kaafee jen léew* 'the coffee has got cold'.

The Burmese verbal system stands rather apart from the others, as may be expected from the different affiliation of the language. There are more than ten analyses of it; that followed here, in matters of doubt and discussion, is usually Alliot's (Alliot 1965:283-309).

As in the normal cases of an uninflecting language, the verb stem itself, essentially monosyllabic, is unchangeable, but combinations of verb stems as in Chinese is possible. There are various syntactic ways of distinguishing combination of stem + stem from stem + auxiliary. Wolfenden (1929:199) says: "direct evidence of the former use by Burmese of pronominal prefixes of the old (i.e. Tibeto-Sinitic) order with verbs is lacking, but the universal occurrence of aspirated initials in verbs having transitive or causative senses, which points to the former existence of directive elements in the language, makes it fairly safe to conclude that subjective prefixes also formerly occurred, since the original and normal Tibeto-Burman verb form could never appear with directive element only". He then goes on to offer evidence for the former existence of directive or objective infixes -- which again sets Burmese apart from the SEA languages hitherto dealt with in this paper.

The basic divisions within the class "verb" is between functive and stative, as in Chinese. The reference is to action or process against state, including adjective constructions (in western terms). In Alliot's setting out, functives include "punctative sentence, with one exponent, the final verb particle *pi*, the negative sentence, marked by the final particle *phu* (but with a preverbal particle *mə* simultaneous), and affirmative sentence marked by finals *te* and *me*. Stewart (1955:29) spells these *tɛ* and *mɛ*, the former present or past time, a narrative particle, and the latter future; *pi* marks completion. Alliot illustrates by *θwa: pi*, 'he has set out, he has gone'; *mə θwa: mɛ*, 'he didn't go; (I)'m not going, (I) won't go'; *θwa: tɛ*, 'he goes, went'; *θwa: mɛ*, 'he will go'. Other students treat the particles rather differently, and there is especially uncertainty about the true value of *pi*. Alliot contains a discussion of previous definitions and why she takes her own stand (p.296ff.), but this is not needed here.

The category of auxiliary verbs is as important in Burmese as in the other languages studied here, and serves to link Burmese more closely to them than its other, rather separative, features. Alliot defines

the situation well: The name auxiliary "is applied to a class of some 25 items whose function is to characterise the degree, likelihood, suitability, possibility etc. of the verb. The members of this word-class are established because, in a negative verbal syntagma, the negative prefix always precedes the verb head and not the auxiliary verb. Auxiliary verbs are distinguished from particles because, in most cases, a homonymous main verb of the same or similar meaning exists, and because they co-occur with certain particles but not with each other." (op. cit. p.287). Stewart calls these "enclitic verbs" (p.42ff.).

The stative verb forms, which include "adjectives" were mentioned in section 2 of the paper. Cornyn (1945:I/131) exemplifies (using a different orthography) by: *lú kâundê*, 'the person is good'; *lûgâun*, 'good person', and *kâundê lú*, 'good person, or person who is good', and goes on: "The latter type is used where the modifying verb is itself modified: *lú θei? kâundé*, 'the person is very good'; *θwe? kâundê lú*, 'a very good person, or a person who is very good'.

Moods and aspects are distinguishable from tenses, unlike Chinese. There is an indicative and an imperative mood, with markers, the latter with zero final particle in the positive and *ne* in the negative. Alliot distinguishes two aspects which she calls culminative and cumulative, marked by *to* for the former, *oun* and *θwei* for the latter. Her explanatory examples include *mə pyo:ne to*, 'don't tell him, then'. In that case, 'don't talk' which is culminative, as against *mə pyo:ne oun*, 'don't speak (yet), don't say any more', which is cumulative. Stewart's definitions are different, and he does not speak of aspects: *tó*, 'imminence, acceptance of the inevitable': *θwa: tó' mə*, 'I am just going'; '*hiq məji' tó^phu*, 'there's no curry left (so it's no good asking for more)' (p.32) and *oun* which he transliterates as 'ouŋ, 'further action in the future': *ci^ouŋmə*, 'I will take a further look; I will look into the matter again' (p.31). In his lists there are 34 particles called "subsidiary verb particles" (pp. 31ff. and 43ff.).

This outline is sufficient to show that Burmese pursues a course in the verb that in general does not coincide with those of the other SEA languages, whether MK or AN.

	PRE-POSED	PERFECTIVE		NEGATIVE POSTPOSED	P A S T		P R E S E N T		F U T U R E		CON-TINU-OUS	REPETITIVE	
		NEGATIVE PREPOSED	POST-POSED		POSITIVE	NEGATIVE	POSITIVE	NEGATIVE	POSITIVE	NEGATIVE		POSITIVE	NEGATIVE
CH.			+la, le	mei'+V+la, le	V+pst(la, le)	mei+V+pst	V	bù+V	+V+	bù+V		V+Rep	mei'+V+Rep
VIET.		P+Neg+V				chu'+a+Pst+V						V+R	Neg.V <sub>1</sub> V <sub>2</sub> , Neg.VR
THAI												V <sub>1</sub> V <sub>2</sub>	
LAO	P+V	Neg+P+V	V+O+P	P+V+P	Pst+V	Neg+Pst+V	V	Neg+V	f+V	f+Neg+V	C+V		VNR
KHM.												V+R	NV <sub>1</sub> V <sub>2</sub> , NVR
MLY.	P+V				(pst)+V	Neg+Pst+V	(Pres)+V				V+C	V+	N+V+ i
RADE					V+P	Neg+V+P					C+V		
ATJ.					Pst+V	Neg+Pst+V	V	Neg+V	f+V	neg+F+V		V+R	NVR
UR.L.			V+P		V(only)	Neg+V							
BURM.					V+P	mə+V+phu	V+tε	mə+V+phu	V+f(mε)	mə+V+phu	V+C		

SUMMARY TABLE II: VERB PHRASE FORMATIONS IN THE LANGUAGES OF SOUTHEAST ASIA

(Based on translation of Glazova's article (op. cit. p. 273) with addition of other languages treated in this paper. Some items of information were not available.)

#### 4. CONCLUSION: TYPOLOGIES OF NEUTRAL DOMINATION

This paper began as an attempt to test what subdivisions may appear among languages designated as "neutrally dominated", ND being defined as the lack of bias towards either object or event. It is now time to reach some conclusion about this.

Typology contrasts with genetic classification. This fact was stressed in the original paper by Honey and Simmonds (pp.71-2). The genetic groups already established on the evidence of comparative historical linguistics are taken as given: Tibeto-Burman-Sinitic, Mon-Khmer, Austronesian. The problem of the exact classification of Atjeh in this system still remains. A structural typology may to a degree overlap genetic classification (Greenberg 1957:34-45; 1963:58-90, especially 83-90). This fact has sometimes led to conclusions undervaluing typology, but this is wrong. What is the relation between ethnic origins and linguistic type still remains unknown, but this does not mean that such a relationship should be denied. In the present study the first contrast is between inflection and its absence. The inflected languages here treated do not only differ in the degree of morphological complication, as they must, but also in conceptualisation of the verbal action, in ways possibly unexpected from those which are uninflected or minimally inflected, such as Malay - both types of neutral domination. They agree with the other languages in expressing social rank, but do it in ways differing from those of the other languages. Korean and Japanese express a different set of verbal circumstances in ways which differ from those of the isolating languages.

There is a tendency for certain differences to appear between the languages as members of a class: the AN languages depart from the MK languages chiefly in the VP. They do not follow the Sinitic pattern at all. In NP they pattern as noun+adjective or noun+connective+adjective (ALP above). This appears not only in the Austronesian languages but also in a few of the MK languages, such as Thai and Lao. In having no "link" for possessives they agree with Khmer, but some coefficients are present as N+number+coefficient.

The majority of the AN languages are not ND but ED - event-dominated. The inclusion of Khmer and Atjeh is justified perhaps by the partial inflection shown in both, but it is also possible to set these languages on the upper end of the ND scale.

The crossing of language families in this paper is deliberate. Typology is not bound up with a language family. As Honey and Simmonds point out, there may also be Sprachbund processes at work. Moreover, any cross-influences (apart from the most likely one of word borrowing) might become apparent in that way.



Taking now the languages that can be classified as ND: how do these differ structurally among themselves? What subclasses of ND suggest themselves? One such might be on the basis of SVO, SOV or some other arrangement on the level of syntax. In point of fact, this subgrouping does not produce results in this instance. Korean and Japanese are the only SOV languages among them, and these are both ED. The only arrangement is SVO, and this is common to MK and AN languages. Hence this type of subgrouping can be overlooked - it appears also on Greenber's demonstration in *Universals of Language* (1963:86), where the three sentence types mentioned appear to be fairly haphazard. Sentence formation, then, is irrelevant.

What, then, of phrase level formation? This has been the chief subject of this paper, and the two Tables produced (as well as the smaller constituent Tables) are the results. Table I, in which the features of NP are set in matrix form provides a résumé of Section 2; those of VP are similarly gathered together in Table II.

In the forms of the NP and VP in the ND languages (and in the first case also the ED languages) the following four sets of possibilities seem to present themselves:

1. Governing elements and particles follow N; they precede V.
2. Governing elements and particles follow both N and V.
3. Governing elements and particles precede N and V.
4. Governing elements and particles precede N, follow V.

The next step towards subdividing ND is to test out these possibilities and determine which of them actually occurs in SEA languages as here treated. The data referred to in *Universals of Language* can also be compared in order to test any wider validity of the propositions made above for ND languages. Korean and Japanese may be omitted here as they are not ND languages - or else the word endings (suffixes) may be regarded, as historically they perhaps were, independent particles which have lost status. In this case these languages will fit into subgroup 2: both types of particles follow both N and V. This is fairly normal in ED languages. The formal implications of the domination concept have yet to be worked out: the author's previous treatment has been primarily semantic rather than formal.

Of the four types given above, Proposition 1 is seen in Vietnamese, where A and D follow N, but verbal markers precede the verb; Malay provides another example. Chinese and Burmese both illustrate No.4. There is no example of proposition 3 - not that there could not be, for English provides one from another area. Of proposition 2 the only examples are the two ED languages just mentioned. Again, there is no logical reason for their absence, and they may well occur elsewhere, especially,

perhaps, in West Africa. The coefficients of numeration are not diagnostic (those of adjectives are!), as they occur in all the languages, of whatever type. As stated, those for ALP constructions are assessable.

The present study of subgroupings with ND results in the following scheme:

- ND 1. (N+; +V) Vietnam; Malay and AN generally; Thai, Lao, Khmer.
- ND 2. (N+; V+) none: Korean and Japanese are ED.
- ND 3. (+N; +V) none - but English is an example.
- ND 4. (+N; V+) Chinese; Burmese.

There is a strong tendency for noun-adjective and noun-demonstrative to cooccur, while adjective-noun and demonstrative-noun similarly cooccur. This is understandable. It means that all noun qualifiers are thought out before the noun is expressed, but strangely enough, the verbal action tends to be thought out before any qualifications of it. This is a matter of semantics in deep structure, outside the scope of the present study, but not without value in their own sphere of deep structure grammar. This phenomenon again is independent of general sentence structure, whether SVO or not - and suggests that deep structure may be based on phrases as much as complete utterances. In the SEA area, SVO is by far the commonest type: it is only the two ED languages and Burmese that depart from it.

Although numeral coefficients are not diagnostic, because they are found in all the types of language treated here, their arrangement in the phrase may provide a ground for subgrouping. Number + classifier is the order in Burmese, classifier + number in Japanese; either order may be followed in Malay and AN as here represented. This may - or may not - indicate that the use of coefficients is a borrowing in AN languages, but this raises difficulties in the wider patterning, for they are marked in Micronesia and present, though less marked in parts of Melanesia and present on a lesser scale again in parts of Polynesia. If they are borrowings, they are ancient, probably of AN period, before there was much differentiation of PAN.

Another contrast is provided by the position of the perfective marker, whether before or after the verb. In Malay and AN generally there is no agreement. In Malay the marker precedes; in the other AN languages it follows the verb, as it does also in Burmese and Chinese. The Radé *leh*, Atjeh *le*, could be related to Chinese *le*, *la*, but it is not likely, if the latter is an abbreviation of *liǎo*; and in the Thai-MK languages treated here there is a double perfective: P + V + P.

Glazova, in summing up, mentions the identity of aspect-tense constructions as a sure sign of mutual dependence within the area. Vietnam, Thai, Lao, and Khmer have structural identity of verbal constructions in

many instances, but Vietnam differs from the others in four cases. Chinese, however, has many peculiarities - she does not include Burmese, but it can be noticed that some of the Chinese peculiarities can be seen also in that language. This is comprehensible in terms of the Tibeto-Burman-Sinitic family relationships, rather than to the fact that Glazova mentions, that Chinese is isolated geographically from the other languages. She lists four points of contrast in Chinese from the other languages, but adds that "notwithstanding the fact that Chinese stands apart in relation to the languages of the Indochinese linguistic area, we should note the agreement of a whole series of the indicators themselves: Chinese *la/liǎo* - Thai, Lao, *lé:w*; Chinese *méi* - Thai *mâj*; Chinese *bù* - Lao *bo*; Chinese *tsen* - Vietnamese *tu'ŋ*, of which she says: "all this bears witness to the strengthening of morphological influence, and the work on historical monuments gives possibility to investigators to determine the question in connection with what period such influence might be practical." These questions, of course, lie outside the present study, but Glazova has done right to draw attention to them. The main purpose of this paper has been to follow up the suggestion originally made by Honey and Simmonds. The second is to provide grounds for subdividing the admittedly rather heterogeneous contents of the "neutral domination" concept. Both have been carried through within the limits of a reasonable length for such a paper, and it remains perhaps to include the languages of the extreme north of the Indochinese regions, such as Palaung, Riang and others - but this must be left till later.

REFERENCES

ALLIOTT, A.J.

- 1965 'Categories for the Description of the Verbal Syntagma in Burmese'. *Lingua*, Part II:283-309.

CAPELL, A.

- 1969 *A survey of New Guinea Languages*. University of Sydney Press.

1978 forthcoming 'Ups and Downs'. *Talanya* vol. 6.

CORNYN, W.S.

- 1946 *Spoken Burmese*. 2 vols. Henry Holt & Company.

COWAN, H.K.J.

- 1948 'Aanteekeningen Betreffende de Verhouding van het Atjesch tot de Mon-Khmer-Talen'. *Bijdragen TLV* 104:429-514.

GLAZOVA, M.G.

- 1967 'Glagolnye Konstrukcii v. Jazykax Indokitajskogo Areala'. *Jazyki Jugo-Vostočnoj Azii*. 257-75. Moscow: Akademia Nauk SSSR, Institut Narodov Azii.

GORGONIEV, J.A.

- 1963 *Kategorija Glagola v. Sovremennom Khmerskom Jazyke*. Moscow.

GREENBERG, J.H.

- 1957 *Essays in Linguistics*. University of Chicago Press.

1963 (edited) *Universals of Language*. Cambridge, Mass.: MIT Press.

HAAS, M. and H.R. SUBHANKA

- 1945 *Spoken Thai*. 2 vols. Henry Holt and Company.

HOGAN, D.W.

- 1972 'Men of the Sea: Coastal Tribes of South Thailand's West Coast'. *Journal of the Siam Society* 60/1:205-34. Also MSS. notes prepared for present writer.

HONEY, P.J. and E.H.S. HENDERSON

- 1963 'Thai and Vietnamese: Some Elements of Nominal Structure Compared'. In: H.L. Shorto, ed. 1963:71-8.

JACOB, M.

- 1968 *Introduction to Cambodian*.

MARTIN, S.E.

- 1954 *Korean in a Hurry*. Charles E. Tuttle Company.

ROLFE, G.E. and T.W.

- 1956 *Spoken Lao*. Book One. American Council of Learned Societies.

SHORTO, H.L., ed.

- 1963a *Linguistic Comparison in South East Asia and the Pacific*. London: School of Oriental and African Studies, University of London.
- 1963b 'The Structural Patterns of Northern Mon-Khmer Languages'. In: Shorto, ed. 1963:45-61.

STEWART, J.A.

- 1955 *Manual of Colloquial Burmese*. London: Luzac & Company.

UHLENBECK, E.M.

- 1967 'Indonesia and Malaysia'. In: T.A. Sebeok, ed. *Current Trends in Linguistics*. Vol.2. 847-98. Mouton.

WOLFENDEN, S.N.

- 1929 *Outlines of Tibeto-Burman Linguistic Morphology*. London: Royal Asiatic Society.



# CASES IN ENGLISH AND SOUTHEAST ASIAN LANGUAGES, AND TRANSLATION

NGUYEN DANG LIEM

0. Introduction
1. Cases Relations in English and Southeast Asian Languages
2. Case Relations and Case Forms in English
3. Case Relations and Case Forms in Southeast Asian Languages
4. Translation

## 0. INTRODUCTION

This paper aims at advocating some grammatical principles of translation from Southeast Asian languages (Cantonese, Khmer, Lao, Thai, and Vietnamese) into English and vice versa. These principles are based upon a case grammar model where both case relations and case forms are taken into account (Starosta 1973a and 1973b), and upon the system of translating set forth by Nida and Taber in these terms:

"Basically there are two systems for translating...

The second system of translation consists of a more elaborate procedure comprising three stages: (1) analysis, in which the surface structure (i.e., the message as given in language A) is analyzed in terms of (a) the grammatical relationships and (b) the meanings of the words and combinations of words, (2) transfer, in which the analyzed material is transferred in the mind of the translator from language A to language B, and (3) restructuring, in which the transferred material is restructured in order to make the final message fully acceptable in the receptor language. This approach may be diagrammed as in Figure 6.

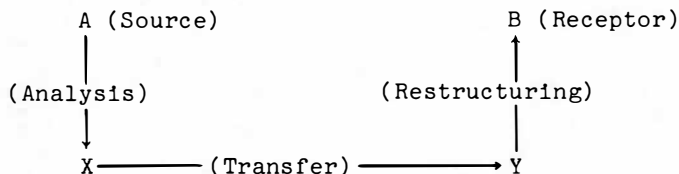


Figure 6.

(Nida and Taber 1969:33).

## 1. CASE RELATIONS IN ENGLISH AND SOUTHEAST ASIAN LANGUAGES

The analysis of English and Southeast Asian languages presented here assumes to be correct Fillmore's claim that there is a finite and universal set of relationships holding between a predicate and its hosted nominals such that every nominal constituent of a proposition is related in a specific way to the predicative verb;

"...for the predicates provided in natural languages, the roles that their arguments play are taken from an inventory of role types fixed by grammatical theory" (Fillmore 1971:376).

Case relationships, hereafter referred to as case relations, may be realised overtly in a variety of ways. These ways include (1) affixation of the noun or pronoun, (2) presence of prepositions or postpositions, (3) word ordering, or (4) marking on verbs (Fillmore 1968:32). Thus, a single case relation can quite normally be realised in different ways, or case forms. Likewise, a single case form can host more than one case relation.

The following twelve universal case relations can<sup>1</sup> be posited. And as they are universals, they should be adequate and sufficient for an analysis of English and the Southeast Asian languages studied:

The AGENTIVE (AGT) case:

The AGT actant is the "instigator of the action identified by the verb" (Fillmore 1968:24). It is assumed here that the agent is not necessarily equated with "intent", for recent research on Tagalog (Ramos 1973) has indicated that certain generalities cannot be formulated unless we allow for non-intentional agents as well as intentional non-agents.

The OBJECTIVE (OBJ) case:

The OBJ actant is the "semantically most neutral case, the case anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself" (Fillmore 1968:25). In general, it will be the element which is acted upon, or whose state or existence is predicated. This



relation subsumes several that have been treated as distinct in other case grammars including Experience, and Result/Factitive. These two types are treated as the interpretations given to the neutral Objective case when it appears with psychological and creative verbs respectively.

The DATIVE (DAT) case:

The case of the "animated being affected by the state of, or experiencing the action of the verb" (Fillmore 1968:24).

The BENEFACTIVE (BEN) case:

The BEN actant receives the benefit of the action identified by the verb.

The COMITATIVE (COM) case:

The COM actant accompanies another actant in the verbal activity or state described.

The INSTRUMENT (INS) case:

The INS actant is "the case of the inanimate force or object causally involved in the action or state identified by the verb" (Fillmore 1968:24).

The LOCATIVE (LOC) case:

The LOC actant indicates "the location or spatial orientation of the state or action identified by the verb" (Fillmore 1968:25).

The DIRECTION (DIR) case:

The DIR actant indicates the orientation of the state or action identified by the verb.

The SOURCE (SRC) case:

The SRC actant indicates the location or time from which action has begun.

The GOAL (GOL) case:

The GOL actant indicates the space or time toward which the action or state identified by the verb has occurred.

The EXTENT (EXT) case:

The EXT actant indicates the space or time through which the action or state identified by the verb has occurred.

The TIME (TIM) case:

The TIM actant identifies the time-setting of the state or action identified by the verb.

Out of the twelve cases, only the AGENTIVE and OBJECTIVE are nuclear in the clause, the DATIVE, BENEFACTIVE, and INSTRUMENTAL cases are semi-nuclear in that they can be hosted only by certain verb classes, and the

rest of the cases, namely, COMITATIVE, LOCATIVE, DIRECTIONAL, SOURCE, GOAL, EXTENT, and TIME are satellites in that they occur with most verbs except those otherwise marked.

The twelve universal case relations are pigeon-holed in language-specific case forms that differ from one language to another. The covert case relations and overt case forms could be set up in bidimensional charts yielding pigeon-holes for possible correspondences of possible, i.e. case relations and case forms of a language.

## 2. CASE RELATIONS AND CASE FORMS IN ENGLISH

In English, the twelve universal case relations are pigeon-holed in twelve overt case forms of which some are marked by their positions *vis-à-vis* the predicative verb, and some are marked by prepositions. The twelve case forms are:

- NM the Nominative position immediately precedes the verb, and there are no prepositions.
- O the Objective position immediately follows the verb, and there are no prepositions.
- A the Agentive position follows the verb, and its objects if any, and it is marked by the preposition *by*.
- B the Benefactive position follows the verb, and its objects if any, and it is marked by the preposition *for*.
- C the Comitative position follows the verb, and its objects if any, and it is marked by the preposition *with*.
- I the Instrumental position follows the verb, and its objects if any, and it is marked by the preposition *of*.
- L the Locative position follows the verb, and its objects if any, and it is marked by the preposition *at*.
- Di the Directional position follows the verb, and its objects if any, and it is marked by the preposition *to*.
- Sr the Source position follows the verbs, and its objects if any, and it is marked by the preposition *from*.
- G1 the Goal position follows the verb and its objects if any, and it is marked by the preposition *until*.
- Ex the Extent position follows the verb and its objects if any, and it is marked by the preposition *in*.
- T the Time position is variant. It can be at the beginning or at the end of the clause, and it is not marked by a preposition.

Naturally, besides the listed prepositions, there are also other prepositions that usually mark one, or at most two, peripheral case relations like LOCATIVE or TIME. Examples of these prepositions are *after, before, against, on, above, under, beneath, while, during, towards,* etc. Being specific case form markers, they do not interest us in this paper. Similarly, the choice of the prepositions *by, for, with, of, to, from, until,* and *in* as case form markers to be discussed could be contested, at least for some of the cases. But that, also, is a question beyond the scope of this paper.

The bidimensional chart of case relations and case forms in English yields 144 possible pigeon-holes of correspondence between case relations and case forms, and the language makes use of thirty two of them as in Chart I below. (If one tries out all the verbs in English, one will most likely find some more correspondences of case relations and case forms. But that, too, will not challenge the validity of the system of translation proposed here.)

Chart 1

## CASE RELATIONS AND CASE FORMS IN ENGLISH

		1	2	3	4	5	6	7	8	9	10	11	12
		NM	O	A	B	C	I	L	Di	Sr	Gl	Ex	T
1	AGT	1		11									
2	OBJ	2	7		15		20	22					
3	DAT	3	8	12					26				
4	BEN	4	9		16								
5	COM					18							
6	INS	5	10	13		19	21						
7	LOC	6		14				23				30	
8	DIR								27				
9	SRC									28			
10	GOL							24			29		
11	EXT				17							31	
12	TIM							25					32

The examples of the correspondences between the case relations and case forms in English are as follows.

1. [+NM, +AGT] is hosted by transitive agentive verbs:<sup>2</sup>

*He bought a book.*

[+NM]                    [+O]

[+AGT]                    [+OBJ]

2. [+NM, +OBJ] is hosted by stative, intransitive, and passive transitive verbs:

*He went away.*

[+NM]

[+OBJ]

3. [+NM, +DAT] is hosted by transitive dative, and passive transitive dative verbs:

*He knew her.*

[+NM]                    [+O]

[+DAT]                    [+OBJ]

4. [+NM, +BEN] is hosted by passive transitive benefactive verbs:

*He was given a book.*

[+NM]                    [+O]

[+BEN]                    [+OBJ]

5. [+NM, +INS] is hosted by transitive instrumental verbs:

*This knife cut the meat well.*

[+NM]                    [+O]

[+INS]                    [+OBJ]

6. [+NM, +LOC] is hosted by stative verbs:

*This room is warm.*

[+NM]

[+LOC]

7. [+O, +OBJ] is hosted by transitive verbs:

*The President bought the book.*

[+NM]                    [+O]

[+AGT]                    [+OBJ]

8. [+O, +DAT] is hosted by transitive dative verbs:

*John sent Mary a book.*

[+NM]                    [+O]                    [+O]

[+AGT]                    [+DAT]                    [+OBJ]

9. [+O, +BEN] is hosted by transitive benefactive verbs:

*John did Mary a favour.*

[+NM]            [+O]    [+O]

[+AGT]           [+BEN] [+OBJ]

10. [+O, +INS] is hosted by transitive instrumental verbs (it is noted that the case relation here could be considered as that of OBJECTIVE rather than INSTRUMENTAL):

*John used the knife to cut meat.*

[+NM]            [+O]

[+AGT]           [+INS]

11. [+A, +AGT] is hosted by passive transitive agentive verbs:

*The book was bought by the President.*

[+NM]                    [+A]

[+OBJ]                   [+AGT]

12. [+A, +DAT] is hosted by passive transitive dative verbs:

*The book was seen by the President.*

[+NM]                    [+A]

[+OBJ]                   [+DAT]

13. [+A, +INS] is hosted by transitive passive verbs marked with [+INS]:

*The glass was broken by the broom.*

[+NM]                    [+A]

[+OBJ]                   [+INS]

14. [+A, +LOC] is apparently hosted by all verbs:

*John was standing by the door.*

[+NM]                    [+A]

[+OBJ]                   [+LOC]

15. [+B, +OBJ] is hosted by transitive verbs that always have *for* with them:

*John was looking for the book.*

[+NM]                    [+B]

[+AGT]                   [+OBJ]

Another way of analyzing the same construction would be to consider *look for* as a two-word verb, hence, *the book* would be [+O, +OBJ]:

*John was looking for the book.*

[+NM]                    [+O]  
[+AGT]                    [+OBJ]

The second solution might be theoretically more sound, because the equivalent passive construction still has the preposition attached to the verb:

*The book was being looked for by John. (?)*

[+NM]                    [+A]  
[+OBJ]                    [+AGT]

16. [+B, +BEN] is hosted by all verbs except those otherwise marked:

*John bought the book for Mary.*

[+NM]                    [+O]                    [+B]  
[+AGT]                    [+OBJ]                    [+BEN]

17. [+B, +EXT] is hosted by all verbs, except those otherwise marked. The EXT can be only in time, not space:

*John walked for two hours.*

[+NM]                    [+B]  
[+OBJ]                    [+EXT]  
                                  [+time]

18. [+C, +COM] is hosted by verbs other than statives:

*John walked with Mary.*

[+NM]                    [+C]  
[+OBJ]                    [+COM]

19. [+C, +INS] is hosted by verbs other than statives:

*John cut the meat with a knife.*

[+NM]                    [+O]                    [+C]  
[+AGT]                    [+OBJ]                    [+INS]

20. [+I, +OBJ] is hosted by transitive verbs that always have *of* with them:

*John thought of Mary.*

[+NM]                    [+I]  
[+DAT]                    [+OBJ]

As in No. 15 above, another way of analyzing the same construction would be to consider *think of* as a two-word verb, hence, *Mary* would be [+O, +OBJ]:

*John thought of Mary.*

[+NM]                    [+O]  
[+DAT]                    [+OBJ]

21. [+I, +INS] is hosted by passive transitive verbs of the *to be made* type. INS might be a misnomer; it might be more accurate to posit a MATERIAL case relation here:

*The airplane was made of wood.*

[+NM]                    [+I]  
[+OBJ]                    [+INS]

22. [+L, +OBJ] is hosted by transitive verbs that always have *at* with them:

*John looked at Mary.*

[+NM]                    [+L]  
[+DAT]                    [+OBJ]

As in No. 15 and No. 20 above, another way of analyzing the same construction would be to consider *look at* as a two-word verb, hence, *Mary* would be [+O, +OBJ]:

*John looked at Mary.*

[+NM]                    [+O]  
[+DAT]                    [+OBJ]

23. [+L, +LOC] is apparently hosted by all verbs (it is to be noted that one could posit more than one LOCATIVE case relation, Platt (1970) posits three LOCATIVES. Such a solution would be capable of indicating which LOCATIVE could be accepted by all verbs, and which ones would not be):

*John stayed at the hotel.*

[+NM]                    [+L]  
[+OBJ]                    [+LOC]

24. [+L, +GOL] is hosted by verbs marked with [+goal]:

*John arrived at the hotel.*

[+NM]                    [+L]  
[+OBJ]                    [+GOL]

25. [+L, +TIM] is apparently hosted by all verbs:

*John arrived at two o'clock.*

[+NM]                    [+L]  
[+OBJ]                    [+TIM]

26. [+Di, +DAT] is hosted by transitive dative verbs:

*John gave the book to Mary.*  
 [+NM]            [+O]            [+Di]  
 [+AGT]           [+OBJ]           [+DAT]

27. [+Di, +DIR] is hosted by verbs marked with [+dir]:

*John went to the hotel.*  
 [+NM]            [+Di]  
 [+OBJ]           [+DIR]

28. [+Sr, +SRC] is hosted by verbs marked with [+source]:

*John went from school.*  
 [+NM]            [+Sr]  
 [+OBJ]           [+SRC]

29. [+G1, +GOL] is hosted by verbs marked with [+goal]:

*John worked until 5 o'clock.*  
 [+NM]            [+G1]  
 [+OBJ]           [+GOL]  
                   [+time]

30. [+Ex, +LOC] is hosted by all verbs except those marked otherwise:

*John worked in his room.*  
 [+NM]            [+Ex]  
 [+OBJ]           [+LOC]

31. [+Ex, +EXT] is hosted by all verbs except those marked otherwise:

*John finished his work in two hours.*  
 [+NM]            [+O]            [+Ex]  
 [+AGT]           [+OBJ]           [+EXT]

32. [+T, +TIM] is hosted by all verbs:

*John went at two o'clock.*  
 [+NM]            [+T]  
 [+OBJ]           [+TIM]

### 3. CASE RELATIONS AND CASE FORMS IN SOUTHEAST ASIAN LANGUAGES

In the Southeast Asian languages considered (Cantonese, Khmer, Lao, Thai, and Vietnamese), the twelve universal case relations are pigeon-holed in twelve overt case forms of which some are marked by their positions *vis-à-vis* the predicative verb, and some by coverbs (for a summary



of analyses of coverbs in some Asian languages, see Clark 1975). The twelve case forms are:

- NM the Nominative position immediately precedes the verb, and there are no prepositions.
- O the Objective position immediately follows the verb, and there are no prepositions.
- D the Dative case form is marked by a coverb meaning '*to give*' or '*to*'. The position follows the Objective position in Cantonese, Khmer, and Vietnamese, but precedes the latter position in Lao and Thai.
- B the Benefactive case form is marked by a coverb meaning '*to help*' or '*for*'. The position follows the Objective position, if any, in all the languages.
- C the Comitative case form is marked by a preposition meaning '*with*'.
- I the Instrumental case form is marked by a preposition meaning '*by means of*'.
- L the Locative case form is marked by a preposition or coverb meaning '*to be at*'.
- Di the Directional case form is marked by coverbs having such meanings as '*to go up*', '*to go down*', etc.
- Sr the Source case form is marked by a preposition or coverb meaning '*to quit*' or '*from*'.
- G1 the Goal case form is marked by a coverb meaning '*to arrive*'.
- Ex the Extent case form is marked by a coverb meaning '*to gain*' or '*in*'.
- T the Time position is a variable one, but it is usually placed at the beginning or the end of the clause.

The twelve covert case relations and the twelve overt case forms can be charted in a bidimensional chart yielding 144 possible pigeon-holes of correspondence. It can be generalised that all the five Southeast Asian languages make use of twenty five correspondences between case relations and case forms as in Chart 2 below (this is a generalisation because one or two of those correspondences might be considered strained, hence not quite grammatical, by native speakers of a language in the group).

Chart 2

## CASE RELATIONS AND CASE FORMS IN S.E.A. LANGUAGES

		1	2	3	4	5	6	7	8	9	10	11	12
		NM	O	D	B	C	I	L	Di	Sr	Gl	Ex	T
1	AGT	1											
2	OBJ	2	6										
3	DAT	3	7	14									
4	BEN		8	15	16								
5	COM					17							
6	INS	4	9				18						
7	LOC	5	10					19					
8	DIR		11						21				
9	SRC							20		22			
10	GOL		12								23		
11	EXT		13									24	
12	TIM												25

The examples of the correspondences between the case relations and case forms are as follows (for the sake of economy, each correspondence will be exemplified in only one language):<sup>3</sup>

1. [+NM, +AGT] is hosted by transitive agentive verbs:

l<sup>á</sup>aw s<sup>ỳ</sup>y kh<sup>ḥ</sup>ḥoŋ (Lao)

he buys things.

[+NM] [+O]

[+AGT] [+OBJ]

2. [+NM, +OBJ] is hosted by stative, and intransitive verbs:

koət t<sup>á</sup>əw (Khmer)

he went.

[+NM]

[+OBJ]

3. [+NM, +DAT] is hosted by transitive dative, and transitive submissive verbs:

khăw	thùuk	tii	(Thai)
he	undergo	beat	
He	was	beaten.	
[+NM]	[+submissive]	[+O]	
[+DAT]		[+OBJ]	

Notice, incidentally, that the correspondence [+O, +OBJ] is filled, as is the case here, by a verb standing for clause when the host verb is marked [+submissive]. All the languages considered have only one submissive verb each, meaning 'to experience something bad', except Vietnamese which has two submissive verbs, được 'to experience something good', and bị 'to experience something bad'.

4. [+NM, +INS] is hosted by transitive instrumental verbs:

dao này	cắt	thịt	(Vietnamese)
this knife	cut	the meat.	
[+NM]		[+O]	
[+INS]		[+OBJ]	

5. [+NM, +LOC] is hosted by stative verbs:

hóoŋ	nìi	hốoŋ	(Lao)
this room	(is)	hot.	
[+NM]			
[+LOC]			

6. [+O, +OBJ] is hosted by transitive verbs:

keúih	maai	syù	(Cantonese)
he	bought	books.	
[+NM]		[+O]	
[+AGT]		[+OBJ]	

7. [+O, +DAT] is hosted by transitive dative verbs:

Ông ấy	bán	tôi	sách.	(Vietnamese)
He	sold	me	books.	
[+NM]		[+O]	[+O]	
[+AGT]		[+DAT]	[+OBJ]	

It is noted that in Lao and Thai, the word order is [+O, +OBJ] before [+O, +DAT].

8. [+O, +BEN] is hosted by transitive benefactive verbs:

Ông ấy mua tôi cuốn này. (Vietnamese)  
*He bought me this volume.*  
 [+NM] [+O] [+O]  
 [+AGT] [+BEN] [+OBJ]  
 [+DAT]

It is noted that the example is ambiguous in that *tôi* 'me' can be either BEN or DAT.

9. [+O, +INS] is hosted by transitive instrumental verbs:

keúih yuhng dou (Cantonese)  
*He used a knife.*  
 [+NM] [+O]  
 [+AGT] [+OBJ]

10. [+O, +LOC] is hosted by intransitive locative verbs:

khǎw yùu haawaay (Thai)  
*He lives in Hawaii.*  
 [+NM] [+O]  
 [+OBJ] [+LOC]

11. [+O, +DIR] is hosted by intransitive directional verbs:

láaw paj hóonhéem (Lao)  
*He went to the hotel.*  
 [+NM] [+O]  
 [+OBJ] [+OBJ]

12. [+O, +GOL] is hosted by intransitive [+goal] verbs:

láaw máa hòot myaη láaw (Lao)  
 [+NM] [+O]  
 [+OBJ] [+GOL]

13. [+O, +EXT] is hosted by all verbs except those otherwise marked:

láaw paj sǎam dyan (Lao)  
*He went for three months.*  
 [+NM] [+O]  
 [+OBJ] [+EXT]  
 [+time]

14. [+D, +DAT] is hosted by transitive agentive verbs:

Ông ấy bán sách cho tôi. (Vietnamese)  
*He sold books to me.*  
 [+NM] [+O] [+D]  
 [+AGT] [+OBJ] [+DAT]  
 [+BEN]

Notice that the example is ambiguous. It could mean '*He sold books to me*' or '*He sold books for me*', i.e. *cho tôi* could be [+D, +DAT] or [+D, +BEN].

15. [+D, +BEN] is hosted by transitive agentive and intransitive actional verbs:

láaw paj háj khôj (Lao)  
*He went for me.*  
 [+NM] [+D]  
 [+OBJ] [+DAT]

16. [+B, +BEN] is hosted by transitive agentive and intransitive actional verbs:

khǎw tham thúk-yàaŋ phǎa lûuk (Thai)  
*He does everything for his child.*  
 [+NM] [+O] [+B]  
 [+AGT] [+OBJ] [+BEN]

17. [+C, +COM] is hosted by verbs other than statives:

khǎw paj kàp phǒm (Thai)  
*He went with me.*  
 [+NM] [+C]  
 [+OBJ] [+COM]

18. [+I, +INS] is hosted by transitive and intransitive actional verbs:

láaw tát paa duàj mǐit (Lao)  
*He cut the fish with a knife.*  
 [+NM] [+O] [+I]  
 [+AGT] [+OBJ] [+INS]

19. [+L, +LOC] is apparently hosted by all verbs (re. note about the same correspondence in English, No.23, p. 51):

khǎw tham ɲaan yùu haawaay (Thai)  
*He works in Hawaii.*  
 [+NM] [+L]  
 [+OBJ] [+LOC]

20. [+L, +SRC] is hosted by verbs marked with [+dir] or [+goal]:

Ông ấy ở Pháp đến (Vietnamese)  
*He from France arrived.*  
 [+NM] [+L]  
 [+OBJ] [+SRC]

Notice that the sentence could be analysed as having two independent clauses and meaning '*He stayed in France, (he) arrived*'.

21. [+Di, +DIR] is hosted by transitive and intransitive verbs marked with [+dir]:

láaw khÿn paj phúu (Lao)  
*(he ascend go mountain)*  
*He went up the mountain.*  
 [+NM] [+Di]  
 [+OBJ] [+DIR]

22. [+Sr, +SRC] is hosted by transitive and intransitive verbs marked with [+source]:

kaa mɔɔk pii haawaj (Cambodian)  
*A comes from Hawaii.*  
 [+NM] [+Sr]  
 [+OBJ] [+SRC]

23. [+G1, +GOL] is hosted by transitive verbs marked with [+goal]:

khǎw maa thŋ haawaay cǎak san fransiskôo (Thai)  
*He came to Hawaii from San Francisco*  
 [+NM] [+Di] [+Sr]  
 [+OBJ] [+DIR] [+SRC]

24. [+Ex, +EXT] is hosted by all verbs except those marked otherwise:

láaw jūu nīi dǎj sǔɔŋ dyan (Lao)  
*He stayed here for two months.*  
 [+NM] [+O] [+Ex]  
 [+OBJ] [+LOC] [+EXT]

25. [+T, +TIM] is hosted by all verbs except those marked otherwise:

láaw si paj taláat mÿy-ÿyn (Lao)  
*He will go to the market tomorrow.*  
 [+NM] [+O] [+T]  
 [+OBJ] [+LOC] [+TIM]

#### 4. TRANSLATION

The system of translation advocated here is that set forth by Nida and Taber (1969:33) whose terms have been quoted in the introduction of this paper, and are repeated below:

"The second system of translation consists of a more elaborate procedure comprising three stages: (1) analysis, in which the surface structure (i.e. the message as given in language A) is analyzed in terms of (a) the grammatical relationships and (b) the meanings of the words and combinations of words, (2) transfer, in which the analyzed material is transferred in the mind of the translator from language A to language B, and (3) restructuring, in which the transferred material is restructured in order to make the final message fully acceptable in the receptor language." (Nida and Taber 1969:33).

The first stage of the translation process is then "analysis". In terms of the grammatical model proposed herein analysis is decoding (re. Chart 3). A message appears in the source language in the surface structure, i.e. it is the clause level that concerns us here, its noun phrases show overt case forms to their predicative verbs. To decode is then to find out the covert case relations that are pigeon-holed in the overt case forms. The process of decoding is indicated by either arrow downwards in Chart 3, and is exemplified as follows.

The surface structure of the message shows the case forms:

*John bought Mary a new dress.*  
[+NM]                      [+O] [+O]                      (OVERT CASE FORMS)

† To decode the message is to find out the covert case relations:

*John bought Mary a new dress.*  
[+AGT]                      [+BEN] [+OBJ]                      (COVERT CASE RELATIONS)

The second half of the analytical process given by Nida and Taber, namely, the analysis of the "meanings of the words and combinations of words", is also very important in translation. However, it is not discussed here because the analysis of the "meanings of the words" is semantics, and the "combination of words" lies in the internal structures of phrases, and hence, both of these aspects lie beyond the realm of this paper which aims specifically at demonstrating the importance of case realisations and case relations in translation.

The second stage of the translation process, namely "transfer, in which the analysed material is transferred in the mind of the translator from language A to language B" (Nida and Taber 1969:33) is, in this proposed model, the conceptualisation of the covert case relations of noun phrases to their hosting predicative verb phrases.

The third stage, namely "restructuring, in which the transferred material is restructured in order to make the final message fully acceptable in the receptor language" (Nida and Taber 1969:33) is, in terms of the proposed grammatical model, encoding (Chart 3, on page 61). To encode a message at the clause level is to pigeon-hole the covert case relations of the noun phrases to their predicative verb phrases in the system of overt case forms of the receptor language. Its process is indicated by either arrow upwards in Chart 3, and is exemplified as follows:

	<i>John</i>	<i>bought</i>	<i>Mary</i>	<i>a book.</i>	(Source)
	[+NM]		[+O]	[+O]	(OVERT CASE FORMS IN SOURCE)
Decoding ↓	[+AGT]		[+BEN]	[+OBJ]	(COVERT CASE RELATIONS)
Encoding ↑	[+NM]		[+D]	[+O]	(OVERT CASE FORMS IN RECEPTOR)
	John	mua	cho Mary	một cuốn sách.	(Receptor, Vietnamese)

The above example of the decoding and encoding process shows that (1) the AGENTIVE and OBJECTIVE case relations appear in the Nominative and Objective case forms respectively in both the source language and the receptor language, and (2) the BENEFACTIVE case relation appears as an Objective case form in English, but as a Dative case form in Vietnamese. The change of case form in the receptor language is necessary in order "to make the final message fully acceptable" (Nida and Taber 1969:33), and is done according to the surface structure requirements of the receptor language. This leads us to the deductions about the priorities in the translating process set forth by Nida and Taber in these terms:

"In conclusion, let us remind ourselves of the priorities in the process of transfer:

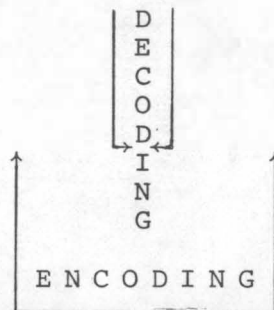
1. At all costs, the content of the message must be transferred with as little loss or distortion as possible. It is the referential, conceptual burden of the message that has the highest priority.
2. It is very important to convey as well as possible the connotation, the emotional flavor and impact, of the message. This is harder to describe than the first, and even harder to accomplish, but it is very important.
3. If, in transferring from one language to another the content and connotation of the message, one can also carry over something of the form, one should do so. But under no circumstances should the form be given priority over the other aspects of the message." (Nida and Taber 1969:118-19).

In the process of translating a message from English to the Southeast Asian languages under consideration, or vice versa, one must bear in mind that the pigeon-holed correlations of case forms and case relations that are circled in Chart 3 are particular to either English or the Southeast Asian languages considered. These case relations, once encoded in the receptor language, will have a different case realisation



Chart 3  
CASES AND TRANSLATION

	12	11	10	9	8	7	6	5	4	3	2	1		1	2	3	4	5	6	7	8	9	10	11	12
	T	Ex	Gl	Sr	Di	L	I	C	B	(D)	O	NM		NM	O	(A)	B	C	I	L	Di	Sr	Gl	Ex	T
1												✕	AGT	✕		(✕)									
2											✕	✕	OBJ	✕	✕		(✕)		(✕)	(✕)					
3										(✕)	✕	✕	DAT	✕	✕	(✕)					(✕)				
4									✕	(✕)	✕		BEN	(✕)	✕		✕								
5								✕					COM					✕							
6							✕				✕	✕	INS	✕	✕	(✕)		(✕)	✕						
7						✕					(✕)	✕	LOC	✕		(✕)				✕				(✕)	
8					✕						(✕)		DIR								✕				
9				✕		(✕)							SRC									✕			
10			✕								(✕)		GOL							(✕)			✕		
11		✕									(✕)		EXT											✕	
12	✕												TIM			(✕)	(✕)			(✕)					✕



from the one in the source language.

	Ông ấy	ở Canberra	đến.	(Source, Vietnamese)
	[+NM]	[+L]		(OVERT CASE FORMS IN SOURCE)
Decoding ↓	[+OBJ]	[+SRC]	↘	(COVERT CASE RELATIONS)
Encoding ↓	[+NM]	[+Sr]		(OVERT CASE FORMS IN RECEPTOR)
	He	arrived	from Canberra.	(Receptor)

The example above shows that sometimes not only do overt case forms differ from one language to another in the translation process but they also occupy different positions *vis-à-vis* other case forms in the clauses.

	John	was given	a book	by Mary.	(Source)
	[+NM]	[+O]	[+A]		(OVERT CASE FORMS IN SOURCE)
Decoding ↓	[+DAT]	[+OBJ]	[+AGT]		(COVERT CASE RELATIONS)
Encoding ↓	[+NM]	[+O]			(OVERT CASE FORMS IN RECEPTOR)
	John	được	Mary cho một cuốn	sách.	(Receptor, Vietnamese)

The example above shows that "It is very important to convey as well as possible the connotation, the emotional flavor and impact, of the message" (Nida and Taber 1969:119). Indeed, the fact that John was given a book by Mary is something benefactory to John. That benefactory connotation is translated by the verb *được* literally meaning '*to experience something good and beneficiary*'. The passive form in English, but having an unfortunate connotation will have to be translated differently as follows:

	John	was beaten	by Mary.	(Source)
	[+NM]	[+A]		(OVERT CASE FORMS)
Decoding ↓	[+DAT]	[+AGT]		(COVERT CASE RELATIONS)
Encoding ↓	[+NM]	[+O]		(OVERT CASE FORMS)
	John	bị	Mary đánh.	(Receptor, Vietnamese)

In the above example, the fact that John was beaten by Mary was an unfortunate matter for him, hence the verb *bị* literally meaning '*to experience something bad and unfortunate*' is used.

All the other correlations of case forms and case relations (i.e. those which are not encircled in Chart 3) could be literally carried over from one language to the other, and would be grammatical. However, such a literal translation might run the risk of not being dynamic. There are semantic/sememic factors which lead any case relation to be expressed by alternative case forms. Given any correlation of case form/case relation in Chart 1 and 2, (1) one needs to determine these semantic/sememic factors which call for the overt case form in question

in the source language, and (2) armed with that and the knowledge of the covert case relations plus the knowledge of how these semantic/sememic factors are realised by one of the possible case forms in the receptor language which express the same case relations, one then makes the right choice of the overt case realisations.

Finally, it is the belief of the author that language is closely related to culture, and that a good translation work must necessarily take that language and culture relationship into account.

NOTES

1. The number of case relations could naturally be reduced; for example, Marybeth Clark (1975) grouped the LOCATIVE, DIRECTIONAL, GOAL, and EXTENT case relations into only one case relation, namely LOCATIVE. Such a reduction of case relations could very well be theoretically preferred, but would only make longer the arguments in favour of the proposed translation theory.

2. As this paper does not intend to classify the verbs in the languages, the information on their characteristics, i.e. whether they are transitive or intransitive, is not exhaustive, and is meant only to second the exemplified correspondences between case relations and case forms. For a classification of verbs in this theoretical model, see Liem 1975, and Liem forthcoming. For classification of verbs in English, see Liem 1969:75-108.

3. The transcriptions follow Huang and Kok (1973) for Cantonese, Brown (1967, 1968, and 1969) for Thai, Crisfield (1970) for Lao, Huffman (1970) for Khmer, and the Vietnamese script.

REFERENCES

BROWN, J.M.

- 1967 *AUA Center Thai Course*, Book I. Bangkok.
- 1968 *AUA Center Thai Course*, Book II. Bangkok.
- 1969 *AUA Center Thai Course*, Book III. Bangkok.

CLARK, Marybeth

- 1975 Ph.D. dissertation, University of Hawaii.

CRISFIELD, Arthur

- 1970 *Speak Lao*. Honolulu: University of Hawaii. (Mimeo.)

FILLMORE, Charles J.

- 1968 'The Case for Case'. In: E. Bach and R.T. Harms, *Universals in Linguistic Theory*. 1-88. New York: Holt, Rinehart, and Winston, Inc.
- 1971 'Types of Lexical Information'. In: D. Steinberg and L. Jakobovits, eds, *Semantics*. 370-92. Cambridge: University Press.

HUANG, Parker Po-fei and Gerald P. KOK

- 1973 *Speak Cantonese*, Book I. Revised edn. New Haven: Yale University Press.

HUFFMAN, Franklin E.

- 1970 *Modern Spoken Cambodian*. New Haven: Yale University Press.

LIEM, Nguyen-Dang

- 1969 *English Grammar, A Combined Tagmemic and Transformational Approach*. PL, C3. 75-108.

- 1975 *Cases, Clauses and Sentences in Vietnamese*. PL, B37.
- forth- Une Analyse Contrastive du Français et des Langues du Sud-Est  
coming Asiatiques (le Birman, le Cantonais, le Laotien, le Thai-  
landais et le Vietnamien). Thèse du Doctorat-ès-Lettres,  
thèse d'Etat. Paris: University of Paris III.
- NIDA, Eugene A. and Charles R. TABER  
1969 *The Theory and Practice of Translation*. Laden: E.J. Brill,  
The United Bible Societies.
- PLATT, John T.  
1971 *Grammatical Form and Grammatical Meaning*. Amsterdam: North-  
Holland Publishing Co.
- RAMOS, Teresita V.  
1974 *The Case System of Tagalog Verbs*. PL, B27.
- STAROSTA, Stanley  
1973a 'The Faces of Case'. *Language Sciences* 25/1-14.  
1973b Case Forms and Case Relations in Sora. Paper presented at  
the First International Conference on Austroasiatic Linguis-  
tics, Honolulu.

# THE ROLE OF DE-ETHNISATION AND ATTITUDE IN THE USE OF PILIPINO: A FACTOR ANALYTIC STUDY<sup>1</sup>

ALEJANDRINO Q. PEREZ<sup>2</sup>

## INTRODUCTION

The Tagalog-based national language of the Philippines, Pilipino (Pil.), began formally to be taught in the schools on June 19, 1940. In the beginning it was taught as a subject in the secondary schools and in college education courses, then the teaching was gradually included at the elementary level.

The formal teaching of Pil. in the schools contributed much to the development and propagation of the language. The mass media, e.g. newspapers, magazines, radio, television, and movies, helped greatly in this development of Pil.

The National Board of Education (NBE) - formerly, Board of National Education - which is the highest policy-making body on education in the Philippines, signed a resolution on February 17, 1970, for the gradual implementation of the use of Pil. as a medium of instruction from Grade I at the elementary level up to college level (Resolution No.70-5, NBE, 1970). However, this resolution was not fully implemented. Then on August 18, 1970, the NBE approved Resolution No.70-21 which paved the way for the use of Pil. as a medium of instruction in the Rizal Course as well as in Philippine Government and Philippine History at the college and university level. This resolution also permitted the use of Pil. in the other courses with the qualification that there be competent teachers, available teaching materials, and readiness on the part of the students. In this connection, the NBE "believed that an educated Filipino should be bilingual in Pilipino and in English". (General Policies on Education, 1967-1972, Board of National Education.)

The recommendation of the Presidential Commission to Survey Philippine Education (PCSPE) is inclined toward the use of two languages, Pilipino and English, as media of instruction in the schools. (PCSPE Education Survey Report, Dec. 1970.)

On August 7, 1973, the NBE approved the new language policy in the Philippines as stated in its Resolution No.73-7, quoted below:

Resolution No.73-7 of the National Board of Education

August 7, 1973

That English and Pilipino serve as media of instruction and be taught as subjects in the curriculum from Grade I to the university level in all schools, public and private; and

Resolved further, that the Department of Education and Culture be requested to prepare and submit for consideration and approval by the Board an implementing scheme or policy guidelines which should include:

- a. subject areas to be taught in English, Pilipino, or in both languages;
- b. schedule of implementation;
- c. preparation of instructional materials.

On the basis of the status of Pil. as described above, research on the role of de-ethnisation and attitude in the use of the language was conducted by the researcher which availed of the factor analytic approach.

#### PROBLEM OF THE STUDY

It is the purpose of this research to make a factor analytic study of the role of de-ethnisation and attitudes on the use of Pilipino (Pil.)<sup>3</sup> in elementary education, which includes the role of language as a medium of instruction as well as a tool for writing textbooks and other social situations in the Philippines.

The study will attempt to answer the following specific questions:

1. What is the role of de-ethnisation in the use of Pil. at the elementary education level?
2. What is the attitude of de-ethnised non-Tagalog subjects (Ss) with respect to the use of Pil. at the elementary education level, in government and in business and trade?

#### SIGNIFICANCE OF THE STUDY

This research will try to give an appraisal of the role of de-ethnisation and attitude of non-Tagalog Ss toward the use of Pil. at the elementary education level and in other social situations related to the use of the language. The findings in this study will serve as guide to language planners, language scholars and administrators on the



course they will take regarding the development and propagation of the language and its role in nation building.

#### Ss OF THE STUDY

A group of sixth grade pupils from the Philippine Normal College (PNC) and from other schools in the Division of City Schools of Manila were used as Ss of this research. The PNC is a state college and the Division of City Schools is a part of the Bureau of Public Schools (BPS).

The Ss were selected on the basis of their being children of non-Tagalog parents, e.g. the father/mother is a Bicolano, the mother/father is a Cebuano, or a member of any other ethno-linguistic group. In other words, the Ss were de-ethnicised, and therefore were no longer speaking the language of their mother or of their father as a first language. Instead, they communicated with their parents in Pil.

The rationale for having chosen the sixth grade Ss is that the researcher assumed that these students had no illusion of travelling abroad. And the language used in the instrument of the research was English in order to minimise if not preclude any prejudice of the Ss themselves.

The male Ss numbered 114 and the female Ss, 235, or a total of 349.

#### MEASURING DEVICE

The Ss were asked to answer 23 items of an aptitude questionnaire with *yes-no-don't know* choices. This aspect of the questionnaire is the second part of the Language Inventory which the researcher prepared. (Please see Appendix A.) The qualification of the data is as follows: *yes* = 2, *no* = 1, and *don't know* = 0.

The rationale behind the three distractors/choices is that being in the sixth grade, the Ss could easily discriminate in their reactions to each of the variables or attitudes.

#### PROCEDURE

The Language Inventory which consists of an attitude list and situations in the use of Pil. as a medium of instruction and as a tool for writing textbooks in elementary education, as well as in the use of Pil. in Philippine social situations, was given to six elementary schools, utilising the grade six pupils. On the basis of the data gathered, 349 Ss were selected, 114 being male and 235 being female. These Ss were all children of non-Tagalog parents, e.g. the father or mother or both were Ilocanos or Pampangos, or came from other ethno-linguistic groups as the case may be. The pupils were all enrolled in the public schools,

that is, the schools were run by the government.

The language used in the instrument, the Language Inventory, was English in order to eliminate or minimise the possibility of bias in the use of Pil.

## RESULTS AND DISCUSSIONS

Table 1 presents the matrix of correlation among the measures of attitude on the use of the Pil. language by the Ss in the different areas, e.g. the role of the language in the different course subjects in elementary education, its role as a medium of instruction as well as its role in the different social situations in Philippine life. This matrix was factor analysed using a Principal Axis solution, with the highest absolute correlation serving as the communality estimate. Three factors were extracted and were rotated by means of a normalised varimax solution. Table 2 presents the rotated factor matrix.

Factor I obtained high loadings from nine measures based on the attitude on the use of Pil. in the schools, in the government and in other Philippine social situations. Two of them (Variables 10, 13) were measures on the use of Pil. in Social Studies and in school offices, respectively; three pertained to the use of the language in government (Variables 14, 15, 16), and four involved the use of the language as a symbolising factor to the country (Variables 17, 18, 19, 21).

This factor seems to suggest that the Ss who tend to prefer the use of Pil. as a medium of instruction in the teaching of Social Studies also desire that the language be used as a communication tool in school offices. This would imply that while the language is employed as a medium of instruction, its use should be extended to other aspects of school activities such as in the functions of school offices. This finding confirms the logical use of the language not only within the four walls of the classroom but also in the offices of the schools which even up to the present have not adapted Pil. in official communications possibly because of the effects of the disparity in the use of Pil. and English (Eng.) in the educational programme of the country, with Eng. still predominating.<sup>4</sup>

The factor suggests also that the Ss are inclined to use Pil. as a communication instrument in government transactions as well as in other occasions, e.g. marketing, in social gatherings, in transportation, etc. This finding confirms further the logical/natural use of the language not only in school offices but also outside the school.

Table 1  
CORRELATION MATRIX\*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	x	04	04	-02	-04	-03	-12	-10	-02	-07	-10	-09	-08	01	10	05	-04	-05	-01	-02	05	-04	-05
2		x	-17	04	02	09	10	08	31	-03	06	12	-04	01	-07	07	05	01	07	10	05	09	05
3			x	00	-05	01	-03	01	-02	-02	-05	-08	16	-01	02	03	-09	12	10	07	-09	-03	23
4				x	06	-03	-03	00	-01	-08	-08	-05	-04	02	-02	09	-03	-01	07	10	04	-03	05
5					x	07	02	04	17	-10	-06	-10	-11	-02	01	-03	-05	-02	07	00	-06	-02	12
6						x	15	14	21	14	26	16	02	08	-02	00	13	06	-02	-03	-03	-02	03
7							x	10	15	-05	-05	01	05	-01	06	10	01	-02	-02	-03	-03	-02	03
8								x	12	24	25	26	16	18	15	30	08	19	12	18	08	18	07
9									x	03	02	05	00	06	10	22	20	11	17	03	01	16	-03
10										x	77	66	41	22	22	26	16	15	14	21	17	12	-03
11											x	76	39	13	13	24	15	15	05	07	05	15	02
12												x	39	14	16	27	21	22	10	11	15	18	04
13													x	25	23	26	14	14	10	25	12	08	03
14														x	55	33	31	22	06	34	19	10	00
15															x	44	42	28	13	28	17	02	-08
16																x	18	29	17	23	16	10	12
17																	x	40	00	06	04	16	09
18																		x	05	21	12	05	12
19																			x	07	-03	-02	-05
20																				x	29	05	-00
21																					x	-03	09
22																						x	04
23																							x

\*The variables are presented in the same order as they are presented in Table 2. The decimal points have been omitted for convenience.

Table 2  
ROTATED FACTOR MATRIX

	Variables/Factors	I	II	III	IV
1	Speaking Pil. is nationalistic	-.03	-.13	-.04	.02
2	Speaking Pil. is shameful	.01	.03	.47	.22
3	I can speak Pil.	-.09	-.05	-.16	.04
4	I am proud when I speak Pil.	-.04	-.10	.03	.01
5	I like the Pil. language	.04	-.12	.23	.07
6	The Pil. language is beautiful	-.02	.20	.34	.16
7	The Pil. language is rich	-.09	-.03	.34	.06
8	Speaking Pil. is likeable	-.28	.22	.35	.19
9	I will understand Science better if it will be taught in Pil.	-.14	-.03	.57	.34
10	I will understand Social Studies better if it will be taught in Pil.	-.33	.78	-.07	.73
11	I will understand Mathematics better if it will be taught in Pil.	-.19	.87	.07	.80
12	The books that I used should be written in Pil.	-.26	.80	.12	.72
13	Pil. should be used in school offices	-.37	.39	-.12	.31
14	Pil. should be used in government offices	-.68	-.02	-.00	.46
15	Pil. should be used on all occasions	-.73	-.05	-.01	.54
16	Pil. should be used in conducting meetings	-.57	.12	.19	.37
17	Pil. should be used in business and trade	-.45	.05	.23	.26
18	The Pil. language is a symbol for the Philippines	-.47	.06	.13	.24
19	The Pil. language should be used in teaching the Filipino children	-.15	.06	.13	.04
20	The Pil. language is an effective tool for national unity	-.47	.03	-.03	.22
21	The Pil. language is effective in understanding our fellow Filipinos	-.31	.05	-.10	.11
22	The Filipino leaders should speak Pil.	-.10	.16	.34	.09
23	We should change the name of the Pil. language	-.07	-.04	.11	.05

In addition, the Ss suggest that Pil. should be used in parliamentary procedures such as in conducting meetings. This implies a wider role or scope in the use of the language, a condition which may be gleaned from the popularity gained by, and the development and enrichment of,

the language brought about by the mass media, the schools, and most of all by the people themselves who use the language in their day to day contacts.

This factor, moreover, seems to suggest that the Ss are inclined to believe/agree that Pil. is an effective tool in the promotion of business transactions and trade (Variable 17). It also reflects the significant use of the language as a symbolising factor to the country, i.e. the Pil. language is a symbol for the Filipino nation, implying that the national language is one of the symbols of the Philippines and it is an effective tool in fostering national unity and an effective instrument in understanding the Filipinos (Variables 18, 20, 21) for it eliminates or minimises the communication gap among the people regarding the dissemination of information on the various government programmes. This finding is in consonance with the finding of Otanes and Sibayan (1969:143-53) that "Pilipino is necessary for good citizenship and participation in the nation's affairs which are expressed in such reasons as to be patriotic and to understand one's heritage".

It is significant to note that Pil. is in the process of modernisation. While this modernisation process is in the making, the language, as part of this process, is moving towards the direction of standardisation. And one of the functions of a standard language (Garvin, 1969) is the unifying function. This function is intricately tied up with nationalism. Such finding in this factor indicates that Pilipino is fulfilling its unifying function among the Filipinos.

This factor, therefore, seems to describe the use of the Pil. language in the schools as well as its role as a communication instrument for the promotion of understanding in government and non-government activities alike.

Four measures obtained high loadings on Factor II, all of which pertain to the attitude of the Ss on the use of Pil. in the schools in general (Variables 10, 11, 12, 13).

This factor seems to suggest that the Ss believe that they would understand the instruction in Social Studies better if it would be taught in Pil. (Variable 10), as in Factor I. It suggests further that they would comprehend better the teaching of Mathematics (Variable 11) if it would be done in the same language. At this point, it seems there is a complementary effect in the use of Pil. as a medium of instruction in such academic areas as Social Studies and Mathematics in the elementary level. This finding is divergent from the report of Otanes and Sibayan (143-53) that "Pilipino is not yet good enough for the use in teaching the more rigorous and technical subjects like arithmetic and science". The issuance of the Policy on Bilingual Education (Dept.

Order No.25, s. 1974), mandates that Pil. should be used in teaching Social Studies/Social Science while English should be used in teaching Science and Mathematics.<sup>5</sup> Such mandate in the teaching of the latter disciplines is incongruent with the finding of the present study.

The Ss would suggest further that Pil. be used in the writing of their books, that is, the books they used in the elementary school. And as in Factor I, they also prefer that the language should be used in the school offices (Variable 13), signalling a confirmation of the wider use of the national language in the school campus which was not practised before.

This factor, therefore, seems to suggest a dimension on the use of Pil. in the different academic disciplines in the school as well as in the writing of books for use in elementary education classes.

Factor III obtained appreciable loadings on six measures. One is about the attitude with respect to speaking Pil. (Variable 2), three are descriptions about Pil., and the remaining two pertain to the use of the language - one in the school and the other by Filipino leaders (Variables 9, 22).

This factor seems to suggest that the Ss believe at the moment that speaking Pil. is shameful (Variable 2), which may be attributed to some sort of paradoxical outlook on the use of the language as reported in Factors I and II. However, looking closely at this finding, it would mean that the other function of a standard language which is prestige as suggested by Garvin, has yet to be attained by the Pil. language. It could be inferred that such finding is still a part of that deep-rooted colonial mentality of the Filipino in using English. The aspect on how to develop in the Filipino a sense of pride in being able to speak his own language must be an interesting aspect of research (Sibayan, 1973).

This factor seems to suggest further that the Ss who believe that Pil. is a beautiful and rich language (Variables 6, 7) are inclined to believe also that speaking it is also likeable (Variable 8). Take note of the paradoxical reaction that speaking Pil. is shameful. It is interesting to remember at this juncture that the Ss are children of non-Tagalog parents and that they communicate with them in Pil., hence this would imply that they are already de-ethnicised as far as the use of language is concerned.

Furthermore, this factor seems to suggest that the Ss prefer that the instruction in Science be done in Pil. because they feel they would understand it better (Variable 9) in this language. This would imply that comprehension in the learning process, in an academic subject like Science, would be more functional if the medium of instruction is the

very language to which the Ss are already oriented and with which they are more familiar because it is the language of their homes. In other words, learning is more functional if it goes side by side with understanding via the language. However, this is incongruent with the finding of Otanes and Sibayan (143-53) whose Ss believed that Pil. could not yet handle the teaching of Science. This is also incongruent with the mandate of the Policy on Bilingual Education that Science and Mathematics should be taught in English (Department Order No.25, s. 1974).

In addition, the Ss suggest that the Pil. language should be used by Filipino leaders in the exercise of their official functions, to eliminate or minimise the communication gap between these leaders and the people whom they serve. This would imply further a wider role of Pil. in the state of affairs in the Philippines. Such expansion of the role of Pil. was vividly expressed by President Ferdinand E. Marcos when he said:

"The Philippines, under the New Society, is reaffirming more strongly than before our commitment to national unity through a national language. It is my desire that the national language now known as Pilipino, but which will develop further to become Filipino, should now be firmly incorporated in all the college entrance examinations and in all civil examinations, as well as in the management development courses of the Development Academy of the Philippines. It is also my desire that the important documents of the States, from now on, should be published in two languages - both English and Pilipino. And it is my desire to see Pilipino rapidly established as a medium of instruction together with English in appropriate courses in our higher institutions of learning. I direct the Department of Education and Culture and the Institute of National Language to take strong and immediate measures to implement these policies." (Ferdinand E. Marcos, 'National Language Unity')

Therefore, this factor seems to describe a dimension on the description of Pil. and its role among Filipino leaders.

This generality on the subordination of the ethnic language in favour of Pil., bringing about the de-ethnisation process, has been reported in other studies (see Balagot, 1972; Barrios, 1972; Galang, 1972; Racho, 1972). This would indicate that one index to the wider acceptance as well as faster development and propagation of Pil. is the process of de-ethnisation.

A general study (Villamin, *et al.*, 1971) on the attitudes on the use of Pil. as a medium of instruction by the undergraduate students, graduate students, and faculty at the Philippine Normal College, revealed that the undergraduate students had a positive attitude towards the use of the language as a communication tool for instruction. But

the graduate students considered it laborious to learn and use Pil. and the faculty members considered it not useful to learn Pil. because most of them had not been oriented to it. Having finished their education through English, they were indifferent to Pil. At this point, this phenomenon gives justification to Sutan Takdir Alisjahbana's pinning hopes on the schools and the young in the development of Bahasa Indonesia, a condition which is taking place with respect to Pil.

In another study (Castillo and Yap, 1971) which aimed to determine the attitudes of the elite, as represented by the students of the Ateneo de Manila University, towards the use of Pil. as a medium of instruction, it was found that the Ss favoured Pil. instead of English. The expectation that the Ss would favour English because they belonged to the elite group was rejected. The researchers concluded that "On the whole, the Ateneo sample expressed a desire for a change which is probably inspired by nationalism...". I would add that if the Atenean Ss had not expressed a desire for change in favour of Pil., they would be left behind clinging to that deep-rooted colonial mentality.

In general, de-ethnisation and attitudes are playing important roles for a wider and faster development and propagation of Pil. Hence, they are also significant indices to the faster acceptance of the language as a tool for national affairs, thereby fulfilling the function of a national standard language.



*Biographical Sketch*

ALEJANDRINO Q. PEREZ (b. 1937, Concepcion, Tarlac, Philippines) is an assistant professor at the Philippine Normal College teaching language and literature courses in Pilipino in the Graduate School Department. He is concurrent Head of the Language and Literature Specialization in Pilipino of the same Department. He obtained his college education, undergraduate and graduate, at the PNC: B.S.E.Ed. (1960), M.A. (1964). He finished his Ph.D. at the University of Sto. Tomas, the oldest institution of higher learning not only in the Philippines but also in Asia.

AQP is the founder-president of the Pambansang Samahan sa Linggwistikang Pilipino, Ink. (PSLP) (National Association for Pilipino Linguistics, Inc.). He is also the founder-president of the Asian Association on National Languages (ASANAL) which was organized on December 20, 1972 during the First Conference on Asian Languages held in Manila, Philippines. He has been the conference director of the First and Second Conferences on Asian Languages (Dec. 18-22, 1972; Dec. 16-21, 1974, respectively). He is also a member of the Board of the Institute of National Language, representing the Pampango Language.

AQP was a lecturer at the University of the Philippines for three academic years in 1971-4.

Books written and edited: *Mga A! Ng Panahon* (The Ah's of Time), 1970; *Pilipinyana*, a modest project towards the direction of encyclopedia-making in Pilipino produced in mimeoscript. The first four volumes are: Tomo I *Wika* (Language), Tomo II *Edukasyon* (Education), Tomo III *Panunuring Pampanitikan* (Literary Criticism), and Tomo IV *Panitikan* (Literature); *Language Policy and Language Development of Asian Countries* (co-editor with Alfonso O. Santiago).

NOTES

1. I would like to express my thanks to my friends and colleagues who helped me in the preparation of this paper: Professor Emma S. Castillo for making and running the computer programme, Att. Benjamin M. Pascual for editing, and Professor Alfonso O. Santiago for his comments and suggestions.

2. Head, Language and Literature Specialization in Pilipino, Graduate School, Philippine Normal College; Member, Board of the Institute of National Language, representing the Pampango language which is one of eight major languages of the country; President, Pambansang Samahan Samahan sa Linggwistikang Pilipino, Ink. (National Association for Pilipino Linguistics, Inc.) and Asian Association on National Languages (ASANAL).

3. The term Pilipino refers to the National Language of the Philippines, while Filipino refers to the people.

4. For a long time since the teaching of Tagalog-based national language called Pilipino in June, 1940, the teaching of Pil. in the schools had been confined only to the teaching of Pil. as a content subject. It was only in the later part of the sixties that Pil. was allowed to be used as medium of instruction in other subjects, such as Social Studies, Mathematics, etc. In 1974, the issuance of Dept. of Education and Culture Order No.24, s. 1974, known as the Implementation Guidelines of the Education Bilingual Policy, which states that Pil. should be used in the teaching of Social Studies/Social Science, Health Education, Work Education, and Physical Education, while English should be used in teaching the English course, science, and mathematics.

5. For further information about the implementation of the bilingual policy on education, see Dept. Order No.25, s. 1974 of the Department of Education and Culture.

APPENDIX "A"

Form No. 1

\_\_\_\_\_  
Leave this blank

LANGUAGE INVENTORY

DIRECTION: Please read this form carefully. Then kindly fill up every blank on this sheet. If you have any question, ask the teacher.

\_\_\_\_\_  
Name                      Grade              Section              Sex

\_\_\_\_\_  
School                      Date of birth              Place of birth

Place where you are living: \_\_\_\_\_  
    Town              Province              City

Birthplace of father: \_\_\_\_\_  
    Town              Province              City

Birthplace of mother: \_\_\_\_\_  
    Town              Province              City

The following are some of the languages of the Philippines. Select among these languages the answer which you should give to the questions below. If the language that you should answer is not in this list, write it on the blank provided for.

Aklanon	Cuyunon	Ilocano	Kinaray-a	Pangasinan
Antiqueno	Gadaang	Ifugao	Magindanao	Romblonon
Bicol	Hiligaynon	Isinay	Masbateño	Sambal
Bolinao	Ibaloy	Isneg	Pampango	Tagalog
Cebuano	Ibanag	Itawes	Maranao	Waray
Chavacano	Igorot	Ivatan		

Here are the questions. Answer them carefully.

1. The language of your father is: \_\_\_\_\_.
2. The language of your mother is: \_\_\_\_\_.
3. Language used by father and mother when conversing: \_\_\_\_\_.
4. Language used at home: \_\_\_\_\_.
5. Language used by father when talking to you: \_\_\_\_\_.
6. Language used by mother when talking to you: \_\_\_\_\_.
7. Language used when talking to your brother/sister: \_\_\_\_\_.
8. Language used in talking with your neighbours: \_\_\_\_\_.
9. Language used in talking with your friends: \_\_\_\_\_.
10. Language used in expressing yourself: \_\_\_\_\_.
11. Language used in buying things in the market: \_\_\_\_\_.
12. Language used in the school campus, that is outside the school room: \_\_\_\_\_.
13. Language used in playing: \_\_\_\_\_.
14. Language used in answering the telephone: \_\_\_\_\_.
15. Language used in taking a jeep/bus/taxi ride: \_\_\_\_\_.

The following questions are about the use of PILIPINO, the national language of the Philippines. Please read them carefully and answer each item. Use a check mark (✓) on your answer. If you have any question, ask the teacher.

1. Speaking Pilipino is nationalistic. 
 Yes  
 No  
 Don't know
   


---
2. Speaking Pilipino is shameful. 
 Yes  
 No  
 Don't know
   


---
3. I can speak Pilipino. 
 Yes  
 No  
 Don't know
   


---

- |   |  |
|---|--|
| 4. I am proud when I speak Pilipino.  | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 5. I like the Pilipino language.  | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 6. The Pilipino language is beautiful.  | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 7. The Pilipino language is rich.   | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 8. Speaking Pilipino is likeable.   | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 9. I will understand Science better if it will be taught in Pilipino.         | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 10. I will understand Social Studies better if it will be taught in Pilipino. | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 11. I will understand Mathematics better if it will be taught in Pilipino.    | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 12. The books that I used should be written in Pilipino.                      | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 13. Pilipino should be used in the school offices.                            | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 14. Pilipino should be used in government offices.                            | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 15. Pilipino should be used on all occasions.                                 | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 16. Pilipino should be used in conducting meetings.                           | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 17. Pilipino should be used in business and trade.                            | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |
| 18. The Pilipino language is a symbol for the Philippines.                    | <input type="checkbox"/> Yes<br><input type="checkbox"/> No<br><input type="checkbox"/> Don't know |

19. The Filipino language should be used in teaching the Filipino children.  Yes  
 No  
 Don't know
- 
20. The Filipino leaders should speak Pilipino.  Yes  
 No  
 Don't know
- 
21. The Filipino language is effective in understanding our fellow Filipinos.  Yes  
 No  
 Don't know
- 
22. The Filipino language is an effective tool for national unity.  Yes  
 No  
 Don't know
- 
23. We should change the name of the Pilipino language.  Yes  
 No  
 Don't know
- If your answer is yes, check the name of the language you want or you can suggest another one:
- Filipino
- Philippino
- Pfilipino
- Maharlika
- \_\_\_\_\_ Other name
-

REFERENCES

- BALAGOT, Lydia R.  
1972 The Ilocano in Manila: His De-ethnization. Typescript.  
25 pp.
- BARRIOS, Sister Mary Angela  
1972 The De-ethnization of the Non-Tagalog Towards Tagalization.  
Typescript. 27 pp.
- CASTILLO, Emma S. and Gloria YAP  
1971 An Attitudinal Survey on the Use of Pilipino as Medium of  
Instruction. Typescript. 23 pp.
- FISHMAN, Joshua A.  
1971 'The Impact of Nationalism on Language Planning: Some Com-  
parisons between Early Twentieth-Century Europe and More  
Recent Years in South and Southeast Asia'. In: Rubin and  
Jernudd, 1971:3-20.
- GALANG, Rosita G.  
1972 De-ethnization of the Kapampangans. Typescript. 9 pp.
- GARVIN, Paul L.  
1964 'The Standard Language Problem - Concepts and Methods'. In:  
Dell Hymes, ed. *Language in Culture and Society*, 521-6.  
New York: Harper and Row.
- MARCOS, Ferdinand E.  
1974 National Language and Unity. Speech read at the Second  
Conference on Asian Language, held at the NSDB Conference  
Hall, December 16-21, 1974, jointly sponsored by the  
Pambansang Samahan sa Linggwistikang Pilipino, Ink.,  
Institute of National Language, and the Asian Association on  
National Languages (ASANAL).

**MINDO, Dinnah F., Gemma P. Orquieza C. RAFAEL**

- 1971 *Factors Affecting the Acquisition of Pilipino*. Manila: Philippine Normal College Language Study Center. Typescript.

**OTANES, Fe T. and Bonifacio P. SIBAYAN**

- 1969 *Language Policy Survey of the Philippines*. Manila: Philippine Normal College Language Study Center.

**PHILIPPINES: BUREAU OF PUBLIC SCHOOLS**

- 1971 *Bureau of Public Schools Memorandum No.20, s. 1971: Pambansang Kalagayan ng Pilipino sa mga Paaralang Bayan*. Hunyo.

**PHILIPPINES: DEPARTMENT OF EDUCATION AND CULTURE**

- 1974 *Department Order No.25, s. 1974 (June 19, 1974). Implementing Guidelines for the Policy on Bilingual Education*.

**PHILIPPINES: NATIONAL BOARD OF EDUCATION**

- 1970 *Resolution No.70-5, February 17, 1970*. Manila, Philippines.  
1970 *Resolution No.70-21, August 18, 1970*. Manila, Philippines.

**PHILIPPINES: PRESIDENTIAL COMMISSION TO SURVEY PHILIPPINE EDUCATION (PCSPE)**

- 1970 *Report*. December, 1970. Manila, Philippines.

**PINEDA, Ponciano B.P.**

- 1974 'The National Language Policy of the Philippines'. In: Al Q. Perez and A.O. Santiago, eds *Language Policy and Language Development of Asian Countries*, 1-10. Manila: Pambansang Samahan sa Linggwistikang Pilipino, Ink.

**RACHO, Albert L.**

- 1972 *Who Speaks What Language to Whom and When? - A Pampango Example*. Typescript. 14 pp.

**RAY, Punya Sloka**

- 1963 *Language Standardization: Studies in Prescriptive Linguistics*. The Hague: Mouton.

**SIBAYAN, Bonifacio P.**

- 1974 'The Development of Pilipino and Sociolinguistics'. In: Al Q. Perez and A.O. Santiago, eds *Language Policy and Language Development of Asian Countries*, 28-37. Manila: Pambansang Samahan sa Linggwistikang Pilipino, Ink.

**VILLAMIN, Araceli M., Paz M. BELVEZ, and Velinda A. SAMONTE**

- 1972 *A Study of Attitudes Towards Language of the Philippine Normal College Faculty, Undergraduate, and Graduate Students*. Typescript.



# MIXED BILINGUALISM: A PRELUDE TO INCIPIENT CREOLISATION OF PILIPINO?<sup>1</sup>

ALFONSO O. SANTIAGO

## 0. BACKGROUND

From a radio broadcast, PULUNG-PULONG SA KAUN-LARAN (English loan-words italicised for easy identification):<sup>2</sup>

Kung pag-uusa-an no natin ang *irregularities*, ano?  
Kung sabagay, kung ang isang estudyante ay nakapasa rito dahil lamang sa isang *leakage* wika nga, o iyong nakakuha siya ng mga *question* na kanyang nasagot kaagad, palagay ko *this will be to the advantage of the student concerned*. Kung saka-sakaling siya'y makalusot, makarating ng kolehiyo, e baka maging *very embarrassing on his part* naman, ano? Kung saka-sakali na kung nasa kolehiyo na siya ay hindi siya makaangkop doon sa *level* na dapat niyang kalagyan. *I think this has something to do now with the pattern of education* na sinusunod sa mga kolehiyo... *Aside from this*, sa *NCEE examination* ay naghihigpit na rin sila ngayon, sapagkat *they won't worry about anymore* dito sa sinasabi nilang *decrease in enrolment* sapagkat magkakaroon na rin sila ng *technological, vocational*, at saka *occupational courses* so that they cannot afford anymore to get in people who are not fit for college. Hindi po ba, Miss Sangalang?

And, also, during a conference on linguistics and bilingualism, the master of ceremony<sup>3</sup> blurted out over the microphone:

Tayo'y magkakaroon ng *fifteen minutes break* at *afterwards* tayo'y babalik dito to *resume the conference*. Merong *sandwiches* at *soft drinks* diyan sa *corridor* para sa mga gustong mag-*refreshment*. Ang *next speaker* na isang kilalang *linguist* ay ipakikilala sa atin *after the break*.

The "mixing" of Pil(ipino) and Eng(lish), as shown in the above quotations, is fast becoming the normal acceptable style these days among the Tag(alog) bilinguals, especially in urban centres, like the Greater

Manila Area (GMA).<sup>4</sup> Any GMA Tag bilingual will accept the above manner of code switching<sup>5</sup> as typical and prevalent.

Goulet (1971:83-6) gives the following extralingual factors influencing language mixing, as well as the motivations for its use and the functions it serves: (1) for precision, i.e. Eng words give the exact meaning the speaker wants to convey; (2) for comic effect, i.e. mixing is very effective in creating humour; (3) for transition, i.e. a shift in language may mark a transition in thought; (4) for atmosphere, i.e. Pil heavily laced with Eng expressions conveys a "stateside" effect; (5) for creating social distance, i.e. "distance" is created between two interlocutors when one starts speaking purely in Eng; (6) for snob appeal, i.e. parents may try to set off their children from those of their neighbours by teaching them Eng as a first language; (7) for secrecy, i.e. parents who do not want their small children to understand the conversation at a particular moment resort to mixing of Eng with the vernacular.

## 1. THE PROBLEM

The Department of Education and Culture (DEC), in consonance with what is embodied in Article IV, Section 3 of the 1972 Revised Constitution<sup>6</sup> and with Resolution No.73-7 of the National Board of Education,<sup>7</sup> came up with an operational definition of bilingualism in Philippine education - that Pil and Eng shall be used as separate media of instruction in definite subject areas. This one-subject-one-language policy is aimed, it is presumed, to produce highly diglossic Filipinos who will be able to function adequately in the separate use of Eng and Pil in any language domains.<sup>8</sup>

It is being hypothesised in this study, however, that in spite of the policy of the DEC on the mutually exclusive use of Eng and Pil in the different school subjects, there will still be a random mixing of Eng and Pil which will eventually lead to the creolisation of the latter. Languages in contact usually start linguistic change with simple borrowing; then the borrowing becomes complex and breeds linguistic convergence, which in turn eventually breeds a creolised variety of either language.

Specifically, it is being hypothesised that the random mixing of Eng and Pil now obtaining among students and professionals of the GMA is a linguistic phenomenon which can be considered as incipient creolisation of the latter.

## 2. CREOLISATION AND STANDARDISATION

Dell Hymes (1971:84) defines creolisation as that "complex process of sociolinguistic change comprising expansion in inner form, with convergence, in the context of extension in use". Language standardisation, on the other hand, is defined by Ferguson (1968:31) as the "process of one variety of a language becoming widely accepted throughout the speech community as a supradialectal norm... rated above regional and social dialects...".

Whether these two processes impinge upon each other is not the subject of this paper. Offhand, however, this researcher strongly believes that creolisation works both ways in the standardisation of a language, i.e. some of the universal dimensions expected of a standard language are enhanced while the others are hindered.

With the present stage of development of Pil, one can easily deduce that it can not yet measure up with the universally accepted norms for language standardisation. And if Pil is really being creolised (something to be proven in this study) because of its continued contact with Eng, the more it becomes premature to talk about its possible standardisation. It would be like an over-eager father wanting his child to be born even during the gestation period.

At this stage, perhaps one relevant thing that language scholars can do is to pinpoint the location of Pil in the whole network of language standardisation, identify the changes it is undergoing, study where it is heading to while it coexists with Eng in the tongues of the bilingual Filipinos.<sup>9</sup>

This study is one of its kind. It focuses on the possible symptoms or manifestations of incipient creolisation of spoken Pil among students and professionals of the GMA. Of course, this may appear to be a long shot toward standardisation but, surely, it is spotlighting Pil where it is now.

Perhaps one good study subsequent to this is something about the possible standardisation of Pil in the face of its creolisation.

## 3. THE BILINGUAL SITUATION

The effect of the Filipinos' linguistic and cultural contacts with Spa(nish) and Eng is mirrored in both the spoken and written Tag prevalent especially in the GMA. In fact, to an ordinary Spanish or American listener, Tag, with all its peculiar intonation and staccato rhythm, will not sound altogether foreign because he will be able to retrieve a hodgepodge of Spa or Eng words woven in its intricate system of affixation. And if the listener is uninitiated, he might suspect that Tag

is an Indo-European language, belonging to the same family where Spa or Eng belong.

A little knowledge of Philippine history, however, will make one understand that the Spa and Eng words interspersed in Tag utterances are actually loanwords from the two foreign languages; that such is the result of our contact with Spa for almost 400 years and with Eng for more than half a century.

Theoretically, the longer the period of contact, the greater would be the linguistic influence of the coloniser's language on that of the colonised. The almost four centuries of Spa rule in the Philippines could have completely hispanised the Filipinos, i.e. the Spa language could have completely nativised and replaced the native languages. This did not take place, however. Frake (In: Hymes, op. cit., p. 223), in tracing the origins of the Spa creoles in the Philippines, says that the consequences of hispanisation in the New World and in South-East Asia differed:

In the Philippines, in spite of rapid Spa conquest, almost total conversion to Christianity, and over three hundred years of occupation, the Spa language failed to establish itself. Spa replaced no indigenous Philippine language, and its role as an auxiliary language was sufficiently tenuous that it was quickly supplanted by Eng after the American occupation. Today, apart from the many Spa loanwords in Philippine languages and a few speakers of Spa in the upper echelons of society, the linguistic legacy of Spain in the Philippines is limited to the existence of several communities that speak a Spa creole language as their mother tongue.

This is in contrast with Eng, which became more widespread only after two decades of American rule in the Philippines:

By 1918 in the Philippine Islands, 49.2% were literate, 26.4% being males and 22.8% being females. Of the literate native population ten years of age and over, the census of 1918 found that 33.9% of the males and 22.4% of the females spoke English, while only 30.4% of the males and 16.9% of the females spoke Spa; 32.1% of the males and 21.5% of the females were able to read and write Eng while only 27.0% of the males and 14.5% of the females were able to read and write Spa. The larger proportion of Filipinos with a knowledge of Eng shows the progress made since implantation of the American educational system.<sup>10</sup>

The above data and discussion can be summed up as follows:

<i>Colonisers' Language</i>	<i>Percentage of Fil. Who Can Speak In</i>	<i>Percentage of Fil. Who Can Read &amp; Write In</i>	<i>Years of Colonisation</i>
Spanish	47.30	41.50	after almost 400 years
English	56.30	53.60	after only 20 years

A study of the differences between the Spanish and the American colonial philosophies, in general, and educational and language policies, in particular, may perhaps help account for the difference in impact of the two languages on the Filipinos. The Spanish era in the Philippines may be characterised simply by "raising the cross and thrusting with the sword"<sup>11</sup> and preserving Spa as an aristocratic language available only to the few elites and not to the "Indios". On the other hand, the first thing that the Americans did, in sharp contrast with the Spaniards' lackadaisical policy, when they colonised the Philippines, was to educate the Filipinos and teach them the Eng language side by side with the principles of democracy on a massive scale.<sup>12</sup>

There were other factors, of course, that influenced the nature of Spanish and American colonisation in the Philippines, besides the differences in policies and attitudes toward language. One of them was the nature of contact itself, i.e. the incentive to learn, and therefore the impact of the Eng language and culture was greater. Another factor was the instructional materials, i.e. there was a dearth of materials in Spa; on the other hand, there was a deluge of Eng materials (Phelan 1959:132).

Presently, after only more than half a century of contact with Eng, and in spite of the fact that the Philippines is no longer under American domination, Eng remains as one of the two official languages of the country. This can be attributed to two principal reasons: (1) Eng continues to be an international language - the language of education, science and technology, diplomacy and foreign relations - serving as the Filipinos' link with the outside world, and (2) unlike the Spaniards, the Americans left no legacy of hate among the Filipinos. As such, the Filipinos continue to look up to their former colonial master's language as a source of knowledge and advancement.

It is in the light of the above that the incipience of creolisation in Pil will be described. Considerable attention has been devoted to the description of the contact situation to show that all the factors discussed have affected or are affecting the spoken Pil of the educated Tag bilingual in the GMA.

#### 4. CREOLISATION: EXTENSIONS OF MEANING

##### 4.1 TRADITIONAL MEANING

Creole, believed to have originated from Portuguese crioulo, via Spanish and French, originally meant a white man of European descent born and raised in a tropical or semi-tropical colony. The term was then applied to certain languages spoken by creoles in and around the Caribbean in West Africa and was later extended to other languages of similar types. (DeCamp, in: Hymes, op. cit., p. 15.)

Like pidgins,<sup>13</sup> creoles before were considered marginal, deviant languages, in the circumstances of their origin... a "barbarous corruption of the standard language" (DeCamp, op. cit., p. 26) ... a "degradation of the great languages of culture" (Hall 1966:40). If mestizos were born during wars and colonisations, creoles were also born with the same circumstances - bastard languages appropriate only for the stigmatised natives, existing largely at the margins of historical consciousness - "on trading ships, on plantations, in mines and colonial armies, often under the most limiting or harshest of conditions" (Hymes, op. cit., p. 5).

So much so that studies along this line were often considered as a whimsical and useless hobby... not to be taken very seriously (DeCamp, op. cit., p. 33).

##### 4.2 CONTEMPORARY MEANING

The above traditional meaning of creole has already undergone further modifications. Lately, creole studies have become a respectable academic field and the research has intensified and their significance to general linguistics, anthropology, and sociology has now been established. No longer are young linguists advised not to "waste their time on such peripheral subjects if they wish to get on in the academic world" (Hymes, op. cit., p. 3). Born as a separate discipline in 1959 (the First International Conference on Creole Language Studies was held this year in Jamaica), it came of age in 1968 (the Second International Conference on Pidgin and Creole Languages was held this year, also in Jamaica). Presently, it has become obvious that creole studies has already been given its rightful standing as the equal of the studies of other languages.

Moreover, it is no longer the case that creoles remain in the realm of pejorative denotations. In South-East Asia, Bazaar Malay, which is still widely used as a contact and trade language, has served as a basis for a new national language. Rebaptised Bahasa Indonesia, it is now the official language of the Indonesian Republic. Here we have an example

of a pidgin being deliberately creolised and made into the vehicle of a national culture in which latter function it is highly respected throughout the world (Hall, op. cit., p. 18).

#### 4.3 MEANING OF CREOLISATION IN THIS STUDY

Traditionally, it is the coloniser's language or the "upper" language which is being creolised or nativised. Very seldom do we find the "lower" language being the one creolised. In the Philippines, between Eng and Pil, it is just normal to expect that it is Eng which will be creolised. In this study, however, it will be shown that it is Tag of the GMA instead which shows signs of incipient creolisation.

Status of language contacts have now changed. Almost every nation these days which was formerly under foreign domination appears to be nurturing bilingualism - maintaining the language of its former colonial master while developing simultaneously one of its indigenous languages as the national language, irrespective of whether it is being creolised or not. In fact, at the rate things are shaping up, one cannot even be sure of what Pil will "look like", linguistically, after it has passed its nebular or evolutionary stage.

This kind of linguistic change, resulting from the creolisation of the "lower" language - in this case, Pil - has never received serious attention before among local as well as foreign language scholars.

### 5. DATA SOURCE AND ANALYSIS

#### 5.1 PHONOLOGY

The principal references used in this section for data purposes were the following: *Balarila Ng Wikang Pambansa* (1941) of the Institute of National Language, representing the contact with Spa, and the *Tagalog Reference Grammar* (1972) of Schachter and Otones, representing the contact with Eng. It might be mentioned that although the Spa rule in the Philippines ended at the turn of the 19th century, the *Balarila* which was written some forty years hence still is an embodiment of the influence of Spa on Tag.

Following are some types of phonological modifications undergone by Tag as a result of its contact with Spa and Eng. In the transcription of words, the /' / marks the stressed syllable indicating loudness (and its correlates) in the case of Eng, and length (and its correlates) in the case of Spa and Tag.

## 5.1.1 Contact with Spanish

<i>SPANISH</i>	<i>TAGALOG</i>
(/f/ > /p/)	
telefono /teléfonó/	telepono /teléponó/
(/v/ > /b/)	
corcovado /korkovádo/	korkobado /korkobádo/
(/ð/ > /d/)	
seda /séða/	seda /séda/
(/z/ > /s/)	
mismo /mízmo/	mismo /mismó/ ~ /mísmo/
(/ç/ > /s/)	
chile /çile/	sile /sile/
(/ç/ > /ts/)	
plancha /plánça/	plantsa /plántsa/ ~ /plánsa/
(/θ/ > /s/)	
cabeza /kabéθa/	kabesa /kabésa/
(/h/ > /s/; /r/ > /l/)	
jugar /hugár/	sugal /sugál/
(/rr/ > /r/)	
guitarra /gitárra/	gitara /gitára/
Loss of initial syllable:	
demasiado /demasyádo/	masyado /masyádo/
hermano /ermáno/	manong /mánong/
Loss of consonant or vowel:	
comadre /komádre/	kumare /kumáre/ ~ /kumádre/
cuestion /kwestyón/	kustyon /kustyón/
puesta /pwésta/	pusta /pustá/
globo /glóbo/	lobo /lóbo/
Addition of morpheme:	
ojal /ohál/	ohales /uháles/ ~ /oháles/
perla /pérlla/	perlas /pérlas/
Metathesis:	
pared /paréd/	pader /padér/
Shift of stress:	
Calixto /kalíksto/	Kalisto /kalistó/
"De-clusterisation":	
Francisco /fransísko/	Paransisko /paransísko/ ~ /pransísko/



## 5.1.2 Contact with English

## ENGLISH

## TAGALOG

(/f/ &gt; /p/ ~ /f/)

telephone /téléfon/

phonograph /fónográf/

teléponó, telepón, telefón, telefónó

ponógrapó, pónograp, pónográf,  
fónograp, fónográf, ponógrafó,  
fonógrapó, fonógrafó

(/v/ &gt; /b/ ~ /v/)

victory /víkturí/

víkturí, bíkturí

(/θ/ &gt; /t/ ~ /θ/; /ʃ/ &gt; /s/ ~ /ʃ/)

toothbrush /túθbraš/

tútbras, túθbras, túθbraš, tutbraš

(/ð/ &gt; /d/ ~ /ð/)

brother /bráðer/

bráder, bráðer

(/z/ &gt; /s/ ~ /z/)

zipper /zíper/

síper, zíper

(/č/ &gt; /ts/ ~ /č/)

chocolate /čókljet/

tsókoléyt, tsókolét, čókolét,  
čókljeyt, čokoléyt

(/ey/ &gt; /e/ ~ /ey/)

baby /béybi/

bébi, béybi

(/ow/ &gt; /o/ ~ /ow/)

Coke /kówk/

kók, kówk

(/æ/ &gt; /a/)

bag /bág/

bág

Shift of stress:

restaurant /réstoránt/

restáwran, restáwrant, restáurán,  
réstorán, réstoránt

## DISCUSSION

Southworth (In: Hymes, op. cit., pp. 260-1), in his study of the prior creolisation of Marathi, mentions a number of phonological changes that took place in that language, together with the other Indo-Aryan languages of the Indo-Pakistan subcontinent during its contact with the Dravidian languages. Some of the modifications are as follows: assimilation of the many and varied consonant clusters, loss of final consonants, followed later by the loss of all final short vowels, with disastrous morphological consequences, loss of the variable pitch accent, phoneme mergins, spirantisation and/or loss of most single intervocalic stops, loss of medial h in further positions, additional phonemic mergers, loss of unaccented /i/, /u/, and short /a/.

Unlike the Indo-Aryan languages, the phonemic system of Tag did not undergo drastic changes as a result of its contact with Spa and Eng. Some changes did occur, like substitution, simplification, addition, subtraction and change in stress, but in general, the Tag phonemic system has remained virtually stable.

This reaffirms Sapir's position (1921:210-15) that:

The borrowing of foreign words always entails their phonetic modification. There are sure to be foreign sounds or accentual peculiarities that do not fit the native phonetic habits... (but) the highly significant thing about such phonetic inter-influencing is the strong tendency of each language to keep its phonetic pattern intact.

There is predictability in the phonological changes resulting from the influence of Spa on Pil. Notice, however, that the change resulting from the influence of Eng appears confusing at this stage of contact. For every loanword, there is usually a set of variants acceptable to Tag bilinguals. And these variants are not being used in a mutually exclusive manner, either. Entry No. 2, p. 93 - *phonograph* - for instance, has at least eight possible variants: *ponograpo*, *ponograp*, *ponograf*, *fonograp*, *fonograf*, *ponografo*, *fonograpo*, *fonografo*. This is obviously a result of a confusion in borrowings from Spa and Eng. Lopez (1965:503) classifies them into two: *sophisticated* and *folk* pronunciations, although at this stage his dichotomous classification is already inadequate.

To the original 17 phonemes of the Old Tagalog (Marcilla 1895:32) - A, I, U, KA, GA, NGA, TA, DA, NA, PA, BA, MA, YA, LA, WA, SA, HA - only three were added to them - E, O, R - after almost four centuries of contact with Spa.<sup>14</sup> Also, a number of consonant clusters from Spa have found their way into the old syllabary of Tag, although Lope K. Santos<sup>15</sup> tried hard to "de-clusterize" them in his *Balarila*; e.g. *kurus*, instead of *krus*, *baraso*, instead of *braso*, *boruha*, instead of *bruha*, etc.

The Tag bilingual of the GMA still fluctuates between the use of Eng sounds /f, v, θ, ð, ʒ, z/ and their respective substitutions /p, b, t, d, sy, s/. Notice, however, that these Eng sounds are fast becoming part of the phonemic inventory of the Tag bilingual. In fact, Schachter & Otones, op. cit., pp. 22-4, already consider /f/ and /ts/ as a recent addition to the EMT's (Educated Manila Tagalog) stock of phonemes, although they do not yet consider them as integral part of the phonemic system of all Tag speakers. But one really very noticeable addition to the phonological repertoire of the GMA Tag bilingual is his ability to pronounce clusters in all positions, evidently as a result of his exposure to the Eng language.

## 5.2 MORPHOLOGY

Bloomfield (1933:453) maintains that a loanword is usually subjected to the system of the borrowing language, both as to syntax... and as to the indispensable inflection... and word formation. This claim is reaffirmed by Haugen (1950:217) by stating that

... loanwords... incorporated into the utterances of a new language, must be fitted into its grammatical structure. This means that they must be assigned by the borrower to the various grammatical classes which are distinguished by his own language.

The reverse, however, also happens. The massive incorporation of lexical items from the "upper" language (coloniser's language) into the "lower" language (colonised language) brings with them derivational and inflectional affixes which even become productive in the latter aside from causing its morphological system to undergo restructuring. (Alleyne, in: Hymes, op. cit., p. 171.)

Polomé (also in Hymes, op. cit., pp. 57-9), in his study of the morphological changes that took place in the formation of the Katanga (Lubumbashi) Swahili Creole, gives the following: changes in shapes of class prefixes, loss of locatives, disruption of the concord system, entailing (a) replacement of complex rules of class agreement for possessives in noun phrases with animate nouns with one stereotyped prefix, (b) invariance of adjectives, (c) replacement of adverbials with adjectives, (d) invariance of numerals, (e) invariance of demonstratives, (f) replacement of all pronominal affixes for "things" in verbs by one stereotyped affix, and (g) reduction of conjugation of the indicative verbs to three main tenses.

How much has Tagalog morphology been affected by its contacts with Spa and Eng shall be discussed very briefly below. (For purposes of systematic discussions, Tag morphological borrowings are grouped into two: roots and affixes.)

In addition to the principal references used for phonology - *Balarila Ng Wikang Pambansa* 1941 and the *Tagalog Reference Grammar* 1972 - Goulet's *English, Spanish and Tagalog: A Study of Grammatical, Lexical and Cultural Interference* is also used here for data source.

5.2.1 Spa plural	Tag plural/singular
veces 'times'	beses 'times/time'
horas 'times'	oras 'times'time'
voces 'voices'	bozes 'voices/voice'
alhajas 'jewellery'	alahas 'jewellery' (plural) 'jewellery' (singular)

5.2.2 Definite articles *el/la* and preposition *a* borrowed as part of nouns, resulting in double noun marking:

<i>el toro</i>	<i>'the bull'</i>	and <i>el toro</i>	<i>'the the bull'</i>
<i>la mesa</i>	<i>'the table'</i>	ang <i>lamesa</i>	<i>'the the table'</i>
<i>a veinte</i>	<i>'on the 20th'</i>	<i>sa abeinte</i>	<i>'on on the 20th'</i>
<i>a la una</i>	<i>'at 1 o'clock'</i>	<i>sa alauna</i>	<i>'at at 1 o'clock'</i>

Notice that this is beginning to happen in Eng *the*:

<i>the end</i>	ang <i>the end</i>	<i>'the the end'</i>
	<i>nag-the end</i>	<i>'it the ended too soon'</i>
	<i>agad</i>	

5.2.3 Confusion in Spa gender:

<i>cenicero</i>	<i>'ash tray'</i>	<i>sinisera</i>
<i>clara</i>	<i>'white of egg'</i>	<i>klaro</i>
<i>tiro</i>	<i>'throw, shot'</i>	<i>tira</i>
<i>misma</i>	<i>'myself'</i>	<i>mismo</i>

5.2.4 Eng verbs borrowed in their base and gerund forms are inflected not for tense but aspect:

<i>hayk</i>	- <i>hayking</i>	<i>'hike'</i>	- <i>'hiking'</i>
<i>maghayk</i>	- <i>maghayking</i>	<i>'go for a hike'</i>	
<i>naghayk</i>	- <i>naghayking</i>	<i>'went for a hike'</i>	
<i>naghahayk</i>	- <i>naghahayking</i>	<i>'is/are hiking'</i>	
<i>maghahayk</i>	- <i>maghahayking</i>	<i>'will go hiking'</i>	

5.2.5 Borrowed word or phrases inflected conforming to Tag system of affixation:

Spa <i>guwapo, korona</i>	
<i>ipinakikipagpaguwapuhan</i>	<i>'comparing the handsomeness of somebody with others'</i>
<i>kinokoronahan</i>	<i>'(somebody) being crowned'</i>
Eng <i>lektyur, bending exercise, via Europe</i>	
ang <i>pagkakalektyur</i>	<i>'the manner the lecture was conducted'</i>
<i>Nagbebending exercise kami araw-araw.</i>	<i>'We do our bending exercises every-day.'</i>
<i>Nag-via Europe sila.</i>	<i>'They travelled via Europe.'</i>

5.2.6 Borrowing of Spa affixes: *-ero/-era, -ador/-adora, -ista, -ante, -aryo*, etc., most of them now productive in Tag

basagulero/basagulera	'trouble maker (female or male)'
karatista	'karate expert (female or male)'
senador/senadora	'senator (male or female)'
kometryante	'comedian (male or female)'
pormularyo	'formula'
-ito/ita, -ilyo/ilya	
pobresito/pobresita	'poor little thing (male/female)'
Mestisilyo/Mestisilya	'hybrid, of European features (male/female)'
-ano/-ana; -eño/-eña	
probinsyano/probinsyana	'from the province (male/female)'
Batanggenyo/Batanggenya	'from Batangas (male/female)'
-eryo/-erya	
sementeryo	'cemetery'
pansiterya	'restaurant'

5.2.7 Eng diminutives or morphemes of endearment have started replacing the local and Spa:

Tag	Spa	Eng
Berting	Bertito	Bobby
Saling	Charito	Sally, Rosie

5.2.8 Eng suffixes have found their way into Tag as bound morphemes:

*-less, -let, -er/-ess, 's, -able, -ation*

islibles	'sleeveless'
bakles	'backless'
taples	'topless'
istarlet	'starlet'
aylet	'eyelet'
singer	'singer'
weyter	'waiter'
weytres	'waitress'
Nena's Sari-Sari Store (instead of Tindahan Sari-Sari ni Nena)	

For comic effect:

tanggalabol	'detachable' (from tanggal 'to remove' + -able)
bakunesyon	'vaccination' (from bakuna 'to vaccinate' + -ation)

#### DISCUSSION

It is obvious from the above data that not a few Spa and Eng roots and affixes have already become part of Tag morphology. Sapir (1921: 215), however, would probably consider these morphological borrowings

as only slightly "different in kind from the mere borrowing of words". In a sense, he is correct because actually *retratista*, for instance, comes from a Spa root *retrato* + a suffix *-ista* with loss of *-o*. Notice, however, the words *karatista*, *daldalero*, *basagulero*, *sabungero* where the Spa suffixes *-ista* and *-ero* are attached to a native or to borrowed Eng term, giving evidence that the Tag bilingual is aware that he is borrowing Spa affixes which can be used productively into the Tag language.

The fact cannot be denied that these loanwords from Spa and Eng have added (and are adding in the case of Eng) new grammatical category to Tag morphology, e.g. the concept of gender shown in such contrasts as *-ero/-era*, *-ador/-adora*, *-ito/-ita*, *-ilyo/-ilya*, *-enyo/-enya*, and Eng *-er/-ess*. These coexist or are in alternation with the kilometric Tag system of marking gender. Below are some example:

<i>Spanish loans</i>		<i>Tag equivalents</i>
<i>kusinero/-era</i>	' <i>cook</i> '	tagapaglutong lalaki/babe
<i>tindero/-era</i>	' <i>shopkeeper</i> '	tagapagtindang lalaki/babae
<i>Bikolano/-ana</i>	' <i>from Bicol</i> '	lalaking/babaing taga-Bikol
<i>kondenado/-ada</i>	' <i>condemned, sentenced</i> '	nahatulang lalaki/babae

The above data confirms Jespersen's (1922:213-14) position that the grammar of a language can be affected by borrowing since "the vocabulary of a language and its 'grammatical apparatus' cannot be nicely separated". Derivative endings certainly belong to the grammatical apparatus of a language; yet Jespersen claims that many such endings have been taken over into another language as parts of borrowed words and have then been freely combined with native speech-material".

On the whole, it can be said that Spa and Eng borrowings have not greatly affected the Tag morphological system. The inherent system of word formation of Tag has remained intact. Rather, on the whole, it is the borrowings which are made to conform to its morphological system.

At this juncture, it can be surmised that instead of effecting morphological changes in Tag, the borrowings from Spa and Eng often simply present the Tag bilingual speaker with alternate shapes for certain morphemes. Truly, there is no denying the fact that the Tag lexicon has been deluged with borrowings both from Spa and Eng but as Sapir claims, these are "but superficial additions on the morphological kernel of the language" (op. cit., p. 220).

## 6. A TEST ON CREOLISATION OF PILIPINO

Marfil and Paigna (1970:4), paraphrasing Hoenigswald (1960:1), claims that "language is systematic - more specifically, that segments

and sentences of a text occur in a unique order or arrangement - and whatever innovations may evolve from it will likewise show systematic patterns of arrangement. These two scholars, using Chomsky's 1957 and 1965 transformational-generative models, analysed the occurrences of Eng shifts in Tag sentences as observed in the *Taliba* daily and posit predictability in the switching of codes by a Filipino Tag-Eng bilingual.<sup>16</sup>

Since this paper is a study on creolisation, and does not intend to establish any patterns of code switching different from or similar to those established by Marfil and Pasiona, the approach will be essentially different. A simple test (see Appendix A) on whether Pilipino is being creolised or not was devised by this researcher, the procedure of which is as follows:

1. Ten conversational questions were taperecorded. These questions were constructed with the following guidelines in mind:
  - a. The question should be basically or structurally Pil with Eng words and phrases interspersed in them.
  - b. The first question, for psychological reason, should be purely Pil with no Eng loanwords whatsoever.
  - c. The mixing should not be random. It should be in accordance with the code switching rules formulated by Marfil and Pasiona (op. cit., p. 11 *passim*).
2. Three monolingual Tagalog speakers were chosen:<sup>17</sup> (A) Martin Reyes, 60; (B) Eufrosina Dayao, 68; and (C) Amado Cabading, 63 - all can read and write in Tag but illiterate in Eng as they were not formally schooled. They learned how to read and write in Tag either through the each-one-teach-one campaign of the Adult Education of the Bureau of Public Schools or through their respective children acting as "little teachers".
3. The participation of the three respondents in the interview were as follows:
  - a. They listened to the replay of the taperecorded questions which were repeated twice (to be sure that their inability to comprehend is not due to some kind of extralingual interference);
  - b. They then responded to the questions according to how they understood each of them;
  - c. Their individual responses to each question were taken down for analysis;
  - d. If a question was not understood by anyone of the three respondents, he/she was asked what word or words he/she did not understand;

e. Again, the responses were taken down for analysis.

4. Interpretation of results.

DISCUSSION

It turned out that, on the whole, the three respondents - A, B, and C - could not respond to most of the questions because the sentences were not comprehensible to them. And when asked why, their usual response was also a question.

In Q. 2, for instance, two of them - B and C - asked: "Anong /eds/?" (age). Respondent A was able to answer the question correctly obviously through context.

Notice, however, that they were all able to respond intelligently to Q. 3 in spite of the use of school and teachers. The reason is that these words, although Eng loanwords, have already become part of their vocabulary.

Q. 4 was totally incomprehensible to them because of agrarian reforms and implement.

Q. 5 was understood by all of them, obviously, through context. Since paninda nowadays is usually associated with "high prices". It is also possible that they were able to retrieve the meaning of prices via presyo.

Q. 6 was not understood by them because of word cheaper and the phrase in other countries. The meaning of the word economist was obviously extracted via ekonomista.

Q. 7 evoked confusion in Respondents A and C because they understood peace to mean fish. And since the meaning they associate with order is 'to ask for something', they thought I was 'ordering some fish' through them. So the response of A was "Hindi panahon ng isda ngayon dito sa atin." Respondent B asked: "Anong /pisenorden/?" (peace and order).

Q. 8, also, was not comprehensible to all of them, obviously because of the phrase cattle rustlers. They only know the Tag phrase, magnanakaw ng kalabaw.

Q. 9 was understood by two of them - B and C - obviously through context again, since it is possible that the word enroll is already part of their vocabulary. Respondent A thought I was asking for a townmate who likes to teach in Manila.

Q. 10 was comprehensible to all of them in spite of the use of Just in case, address, and seven-one-eight because even though they do not understand just in case, they can still get the meaning through context. The word address is obviously already part of their vocabulary, as with the Eng numerals, such as seven-one-eight.



Below is the chart showing the result of the creolisation test.

Respondents	QUESTIONS									
	1	2	3	4	5	6	7	8	9	10
(A) Martin Reyes	/	/	/	x	/	x	x	x	x	/
(B) Eufrosina Dayao	/	x	/	x	/	x	x	x	/	/
(C) Amado Cabading	/	x	/	x	/	x	x	x	/	/

Legend: / = correct response; x = incorrect response or could not respond

### 7. CONCLUSIONS

This researcher posited earlier in this paper that instead of Eng, it is Pil which is undergoing changes or being creolised simply because the brand of English by ordinary Tag bilinguals in the GMA is becoming more and more comprehensible to the American native speakers of Eng while the brand of Pil being spoken by the same group is becoming less and less comprehensible to a Tag monolingual. This position has been positively proven in this study.

Following are four criteria given by Weinreich (1970:69-70) for deciding whether or not a new language has been formed out of the crossing of two others:

1. a form palpably different from either stock language;
2. a certain stability of form after initial fluctuations;
3. functions other than those of a workaday vernacular (e.g. use in the family, in formalised communications, etc.);
4. a rating among the speakers themselves as a separate language.

Criterion No. 1 has been proven positive, i.e. if Eng is Language 1 and Tag is Language 2, there is a forming a new language, Language 3, which is becoming more and more incomprehensible to the monolingual native speakers of Language 1 and Language 2.

Criterion No. 2 - stability of form - has not yet been attained since creolisation of Pil is yet in its incipient stage. As has been shown in the data, there were fluctuations and confusions going on in the phonological, morphological and even in the semantics of Pil as a result of the influence of Eng. Normally, this results in linguistic instability which is characteristic of any dynamic acculturative process.

Criterion No. 3 - breadth of function - is assumed in this study, i.e. mix-mix Pil is used in different domains, as shown in the studies of Marfil and Passigna, Barrios, *et al.*, and Bautista.

Criterion No. 4 - speakers attitudes - is not covered in this study, since it is believed that at this stage of creolisation of Pil, it has not yet generated sufficiently separatistic attitudes among its speakers to recognise and claim it as a new language.

Let it be made clear at this point, however, that this study does not prove nor claim that Pil is already a creole. What this study aims to point out is that Pil may be said to be in the stage of incipient creolisation. And if the present contact situation between Pil and Eng persists, creolisation of Pil will just be a matter of time.

Of course, Pil will not suffer the fate of the many European immigrant languages in America. Pil is spoken in the home country and, therefore, the pressures that were brought to bear on the European immigrants to America do not exist in the Philippines.

At any rate, if Pil really becomes creolised in the future, it will be a kind of creole which is homegrown and, therefore, its characteristics will be uniquely different from those of the other classical creoles of the world. Neither will its characteristics be similar to those of the Chavacano Creoles of Cavite, Ermita, Ternate, and Zamboanga, considering the Whinnom theory (In: Hymes, *op. cit.*, p. 224) that they had all diverged from a common source in the Moluccas and, therefore, not homegrown. Pil creole will still be recognisable as a Philippine language but with unusually heavy Eng lexical influence, considerably exceeding the influence of Spa.

Pil creole, then, will be a language that is virile and synamic, as it will be used and proudly owned not only by the Tagalogs but by all Filipinos, Tagalogs or non-Tagalogs alike, a language which will mirror the Filipinos as a distinct race who, as a consequence of fate and history, belongs to a nation of mixed tongues and mixed cultures, no longer parochial in outlook and disposition.

*About the Author*

ALFONSO O. SANTIAGO (1930 -; Nueva Ecija, Philippines) is Assistant Professor at the Philippine Normal College teaching linguistics and literature courses in Pilipino in the Graduate School Department; concurrently Head of the Materials Development Unit of the same department. He obtained his college education all at the PNC: Elementary Teacher's Certificate (1952), Bachelor of Science in Elementary Education (1965), Master of Arts in Education (1967). He is presently pursuing his Ph.D. in linguistics also at PNC (in consortium with Ateneo de Manila University).

AOS is co-founder of the Pambansang Samahan Sa Linggwistikang Pilipino of which he is the present Vice-President and Editor of its journal - *Linggwistikang Pilipino* - published semi-annually. He is also Associate Editor of the *Philippine Journal for Language Teaching*, another journal published annually by the Philippine Association for Language Teaching.

Awards and Scholarships: Journalism Award, PNC, 1965; BPS-PNC/NDEA-UCLA (Ford Foundation) Scholarship (1965-7); Asia Foundation Scholarship (1971-4); PNC-Language Study Center Faculty Development Scholarship (1974-present).

Books written and edited: *Panimulang Linggwistika* (Introductory Linguistics, author), *Sining Ng Pagsasalingwika* (Art of Translating, author), *Language Policy and Language Development of Asian Countries* (co-author with Dr A.Q. Perez), *Panimulang Pagbasa Sa Pilipino* (Beginning Reading in Pilipino, editor), *Patuloy Na Pag-Unlad Sa Pagbasa Sa Pilipino* (Continuous Progression in Reading in Pilipino, editor).

NOTES

1. Pilipino, in this paper, is being distinguished from its basis, Tagalog. There are many varieties of Tagalog - Bulacan Tagalog, Laguna-Tagalog, Batangas-Tagalog, Nueva Ecija-Tagalog, Bataan-Tagalog, Quezon-Tagalog, etc. which are mutually intelligible but each has its own linguistic peculiarities. Other varieties of Tagalog are found in non-Tagalog regions - Bisayan-Tagalog, Ilocano-Tagalog, Pampango-Tagalog, etc., differing from each other according to the influence of the linguistic peculiarities of the region. All of these Tagalog varieties, aside from English, may be said to be now having their own impacts on Manila-Tagalog (Manila being the nerve centre of the country's civilisation - culturally, educationally, technologically, economically, etc.) which may rightfully be called *the* Pilipino language. Pilipino then, in this study, is the language that is now being formed in the Greater Manila Area where the natural amalgamation process in the formation of a true *national* language is now taking place.

2. Taken from the September 1973 taperecorded data of Miss Ma. Lourdes Bautista, an Ateneo-PNC Consortium scholar for a Ph.D. in Linguistics, who has defended her dissertation just very recently on 'The Filipino Bilingual's Linguistic Competence: A Model Based on An Analysis of Tagalog-English Code Switching'. Miss Bautista's tapes are made up of 1508 utterances distributed among 564 turns of speaking and 22 speakers, 66.31% of which constitute or contain some kind of code switching.

3. Taken from the taperecorded proceedings of the 5th Annual Convention of the Linguistic Society of the Philippines on 'Linguistics and Bilingualism in the Philippines', July 26-27, 1974, held at the Philippine Normal College, co-sponsored by the Institute of National Language, the Bureau of Public Schools, and the Bureau of Private Schools. The master of ceremony being quoted in this paper was Dr Fe Aldave Yap, Assistant Director of the Institute of National Language.

4. Greater Manila Area is the geographic boundary covering the cities of Manila, Caloocan, Pasay and Quezon, and the municipalities of Makati, Mandaluyong, Pasig and Marikina.

5. Code switching, in this study, is the alternate use or systematic mixing of two languages - Pilipino and English - in a given communication process. It differs from *language shift* as defined by Weinreich (1970: 68) wherein there is a change from the habitual use of one language to that of another, a very common phenomenon in a highly diglossic speech community.

6. Article XV, Section 3 of the 1972 Constitution states that: "Unless otherwise provided by law, English and Pilipino shall be the official languages".

7. Resolution No. 73-7 of the National Board of Education, dated August 7, 1973, reads: "That English and Pilipino serve as media of instruction and be taught as subjects in the curriculum from Grade I to the university level in all schools, public and private".

8. The Department Order No. 26, s. 1974 reads in part:

In consonance with the provisions of the 1972 Constitution and a declared policy of the National Board of Education on bilingualism in the schools, in order to develop a bilingual nation competent in the use of both English and Pilipino, the Department of Education and Culture hereby promulgates the following guidelines for the implementation of the policy:

- a. Bilingual education is defined, operationally, as the separate use of Pilipino and English as media of instruction in definite subject areas, provided that additionally, Arabic shall be used in the areas where it is necessary.
- b. The use of English and Pilipino as media of instruction shall begin in Grade I in all schools. In Grades I and II, the vernacular used in the locality or place where the school is located shall be the auxiliary medium of instruction; this use of the vernacular shall be resorted to only when necessary to facilitate understanding of the concepts being taught through the prescribed medium for the subject, English, Pilipino or Arabic, as the case may be.
- c. English and Pilipino shall be taught as language subjects in all grades in the elementary and secondary schools to achieve the goal of bilingualism.
- d. Pilipino shall be used as medium of instruction in the following subject areas: social studies/social science, character education, work education, health education and physical education.

9. Filipino refers to the people; Pilipino to the language. Manila-Tagalog and Pilipino are used interchangeably since they are considered essentially synonymous by the writer.

10. *Census of the Philippines, 1918*, II, 60-2, quoted by W. Cameron Forbes, *The Philippine Islands* (New York: Houghton Mifflin Company, 1928), I, 416, n. 2.

11. Forbes, op. cit., I, 49, citing a translation of S. Vidal y Soler, *Viajes por Filipinas de F. Jagor* (Madrid: 1875), p. 305.

12. Forbes, op. cit., II. Appendix VII. Following is the pertinent part of President McKinley's instructions to the members of the Civil Commission leaving for the Philippines, which was actually prepared by Elihu Root, Secretary of War, with the help of William H. Taft, Chairman of the Commission:

It will be the duty of the Commission to promote and extend and, as they find occasion, to improve the system of education already inaugurated by the military authorities. In doing this they should regard as of first importance the extension of a system of primary education which shall be free for all, and which shall tend to fit the people for the duties of citizenship and for the ordinary avocations of a civilized community. This instruction should be given, in the first instance, in every part of the Islands in the language of the people. In view of the great number of language spoken by the different tribes, it is especially important to the prosperity of the Islands that a common medium of communication may be established, and it is obviously desirable that this medium should be the English language. Especial attention should at once be given to affording full opportunity to all the people of the Islands to acquire the use of the English language.

13. A language, in order to be considered a pidgin, should meet two conditions: (1) its grammatical structure and its vocabulary are sharply reduced, and (2) the resultant language is native to none of those who use it. (Hymes 1971:3; Hall 1966:xii.)

14. The total number of phonemes of Tagalog will be 21 if the glottal stop /ʔ/ is added to it.

15. Lope K. Santos, the author, in fact, rejected the consonant clusters, to wit:

- ... Di maaring kakitaan ang wagas na pananagalog ng mga pantig... na gaya ng alinman sa mga sumusunod:
- (a) Na, sa isang pantig ay magkaroon ng hihigit pa sa dalawang katinig.
  - (b) Nam ang dalawang katinig ay magkasama sa unahan o sa hulihan ng isang pantig.

**TRANSLATION:**

- ... *The following do not occur in puristic Tagalog:*
- (a) *More than two consonants in one syllable.*
  - (b) *Two consonants (clusters) in syllable initial or syllable final positions.*

16. A corollary study by Bautista (on-going), also using Chomsky's 1965 grammatical model, attempts to construct a model of the Filipino bilingual's linguistic competence, after typologising Tag-Eng switching of codes. Bautista also expects to be able to establish predictability of code switching among Filipino Tag-Eng bilinguals.

17. The three respondents were chosen from among the elder relatives of the researcher in Cabaio, Nueva Ecija, a Tag-speaking community. Incidentally, the three monolingual Eng-speaking counterparts were no longer sought to act as respondents to the test since it is very obvious that said questions which are basically Tagalog will not be comprehensible to an American.

APPENDIX A  
TAPERECORDED QUESTIONS

Directions to the respondent before the replay of the taperecorded questions:

Interviewer: Mayroon po akong dalang taperecorder dito. Kapag pinaandar ko ito ay may maririnig kayong mga tanong. Ang bawat tanong ay dalawang beses na ulitin upang marinig ninyong mabuti. Kaya't bawat tanong ay hihintayin muna ninyong dalawang beses na ulitin bago ninyo sagutin. Halimbawa:

Magadang araw ho. Kumusta kayo?

Magadang araw ho. Kumusta kayo?

Magadang araw ho. Kumusta kayo?

Ano, halimbawa, ang sagot ninyo sa ganitong tanong?

Interviewee: Mabuti.

Tama. Ngayon po ay magsisimula na tayo.

(Turn on the taperecorder.) Note: Every question is repeated twice actually.

- Q. 1 - Sa tingin ko ay malakas na malakas pa kayo. Wala pa ba naman kayong nararamdaman sa inyong katawan?
- Q. 2 - Ano na ba ang *age* ninyo ngayon?
- Q. 3 - Kumusta naman ang *school* natin? Sinu-sino na ba ang mga bagong *teachers* diyan?
- Q. 4 - Maiba po ako ng usapan. Balita ko ay maayos na raw ang lagay ng mga *farmers* ngayon dito sa atin dahil sa mga *agrarian reforms* na ini-*implement* ng ating gobyerno. Ano po ang masasabi ninyo tungkol dito?
- Q. 5 - Kumusta po naman ang *prices* ng mga paninda ngayon?



- Q. 6 - Kung sabagay, ang halaga ng lahat ng bagay sa buong daigdig ay talagang pataas nang pataas. Hindi lang dito sa atin. Ayon sa mga *economist*, mas *cheaper* pa ang *prices* ng mga bilingin sa Pilipinas kung ico-*compare* natin sa *prices* ng mga bilingin in *other countries*. Ano po ang masasabi ninyo tungkol dito?
- Q. 7 - Kumusta po naman ang *peace and order* dito ngayon?
- Q. 8 - Wala na po bang nahuhuling mga *cattle rustlers* ngayon?
- Q. 9 - Ako nga po pala ay sa Maynila na nagpipirmi ngayon. Sa *Philippine Normal College* po ako nagtuturo. Wala po ba tayong *relatives* dito na gustong doon mag-*enrol*?
- Q.10 - Buweno, salamat po sa inyo nang marami. *Just in case* po na maluluwas kayo ng Maynila. Magpunta naman kayo sa amin. Matutuhan po kaya ninyong puntahan ang *address* na *seven-one-eight* (718), Mercedes, Ermita? Paalam na po.

BIBLIOGRAPHY

- ALLEYNE, Mervyn C.  
1971 'Acculturation and the Cultural Matrix of Creolization'. In:  
Hymes, ed. 1971:169-86.
- BARRIOS, Mary Angela, *et al.*  
1974 The Greater Manila Speech Community: Bilingual or Diglossic?  
Term paper, Linguistics 685 - Seminar on Multilingualism  
Ateneo-PNC Consortium for a Ph.D. in Linguistics.
- BAUTISTA, Ma. Lourdes  
On- The Filipino Bilingual's Linguistic Competence: A Model Based  
going on an Analysis of Tagalog-English Code Switching. Ateneo-PNC  
Consortium for a Ph.D. in Linguistics.
- BLOOMFIELD, Leonard  
1933 *Language*. New York: Henry Holt.
- DECAMP, David  
1971 'Introduction: The Study of Pidgin and Creole Languages'. In:  
Hymes, ed. 1971:13-39.
- FERGUSON, Charles A.  
1968 'Language Development'. In: J.A. Fishman, C.A. Ferguson and  
J. Das Gupta, eds *Language Problems of Developing Nations*,  
27-35. New York: John Wiley.
- FISHMAN, Joshua A.  
1970 *Sociolinguistics: A Brief Introduction*. Rowley, Mass.:  
Newbury House.
- FORBES, W. Cameron  
1928 *The Philippine Islands*. 2 vols. New York: Houghton Mifflin.

FRAKE, Charles O.

- 1971 'Lexical Origins and Semantic Structure in Philippine Creole Spanish'. In: Hymes, ed. 1971:223-42.

GOULET, Rosina Morales

- 1971 *English, Spanish and Tagalog: A Study of Grammatical, Lexical and Cultural Interference*. Special Monograph Issue No.1. Loyola Heights: Linguistic Society of the Philippines, Ateneo de Manila University.

GUMPERZ, John J.

- 1971 'Convergence and Creolization: A Case for the Indo-Aryan/ Dravidian Border in India'. In: Anwar S. Dil, ed. *Language in Social Groups: Essays by Gumperz*. California: Stanford University Press.

HALL, Robert A., Jr

- 1966 *Pidgin and Creole Languages*. Ithaca: Cornell University Press.

HAUGEN, Einar

- 1950 'The Analysis of Linguistic Borrowing'. *Language* 26:210-31.  
 'The Confusion of Tongues'. In: Anwar S. Dil, ed. *Ecology of Language: Essays by Haugen*. California: Stanford University Press.

HOENIGSWALD, Henry M.

- 1960 *Language Change and Linguistic Reconstruction*. Chicago: University of Chicago Press.

HYMES, Dell, ed.

- 1971 *Pidginization and Creolization of Languages*. Proceedings of a Conference Held at the University of the West Indies, Mona, Jamaica, April 1968. London: Cambridge University Press.

JESPERSEN, Otto

- 1922 *Language: Its Nature, Development and Origin*. London: George Allen & Unwin.

LOPEZ, Cecilio

- 1965 'The Spanish Overlay in Tagalog'. *Lingua* 14:467-504.

MARCILLA, Cipriano

- 1895 *Estudio de las Antiguas Alfabetos de Filipinas*. Malabon, Rizal.

**MARFIL, Alice E. and Aida L. PASIGNA**

- 1970 An Analysis of Shifts from Tagalog to English in Printed Materials. Master's thesis, Philippine Normal College, Manila. [!]

**OTANES, Fe T. and Bonifacio P. SIBAYAN**

- 1969 *Language Policy Survey of the Philippines*. Manila: Philippine Normal College Language Study Center.

**PHELAN, John Leddy**

- 1959 *The Hispanization of the Philippines: Spanish Aims and Filipino Responses, 1565-1700*. Madison: University of Wisconsin Press.

**PHILIPPINES: SURIAN NG WIKANG PAMBANSA (Institute of National Language)**

- 1941 *Balarila ng Wikang Pambansa*. Manila: Bureau of Printing. (Written by Lope K. Santos.)

**POLOMÉ, Edgar C.**

- 1971 'The Katanga (Lubumbashi) Swahili Creole'. In: Hymes, ed. 1971:57-9.

**RAY, Punya Sloka**

- 1963 *Language Standardization: Studies in Prescriptive Linguistics*. The Hague: Mouton.

**SAMARIN, William J.**

- 1966 'Self-annulling Prestige Factors Among Speakers of a Creole Language'. In: William Bright, ed. *Sociolinguistics*, 188-213. The Hague: Mouton.

**SAPIR, Edward**

- 1921 *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.

**SCHACHTER, Paul and Fe T. OTANES**

- 1972 *Tagalog Reference Grammar*. Berkeley: University of California Press.

**SIBAYAN, Bonifacio P.**

- 1970 The Standardization of Pilipino. Paper read before the 1st Annual Convention of the Linguistic Society of the Philippines, Ateneo de Manila University, Padre Faura, Manila.
- 1973 'The Development of Pilipino and Sociolinguistic Studies'. *Philippine Association for Language Teaching Journal* 5 & 6.

SOUTHWORTH, Franklin C.

- 1971 'Detecting Prior Creolization: An Analysis of the Historical Origins of Marathi'. In: Hymes, ed. 1971:255-73.

STEWART, William A.

- 1962 'Creole Languages in the Caribbean'. In: Frank A. Rice, ed. *Study of the Role of Second Languages in Asia, Africa, and Latin America*, 34-53. Washington, D.C.: Center for Applied Linguistics of the Modern Language Association of America.

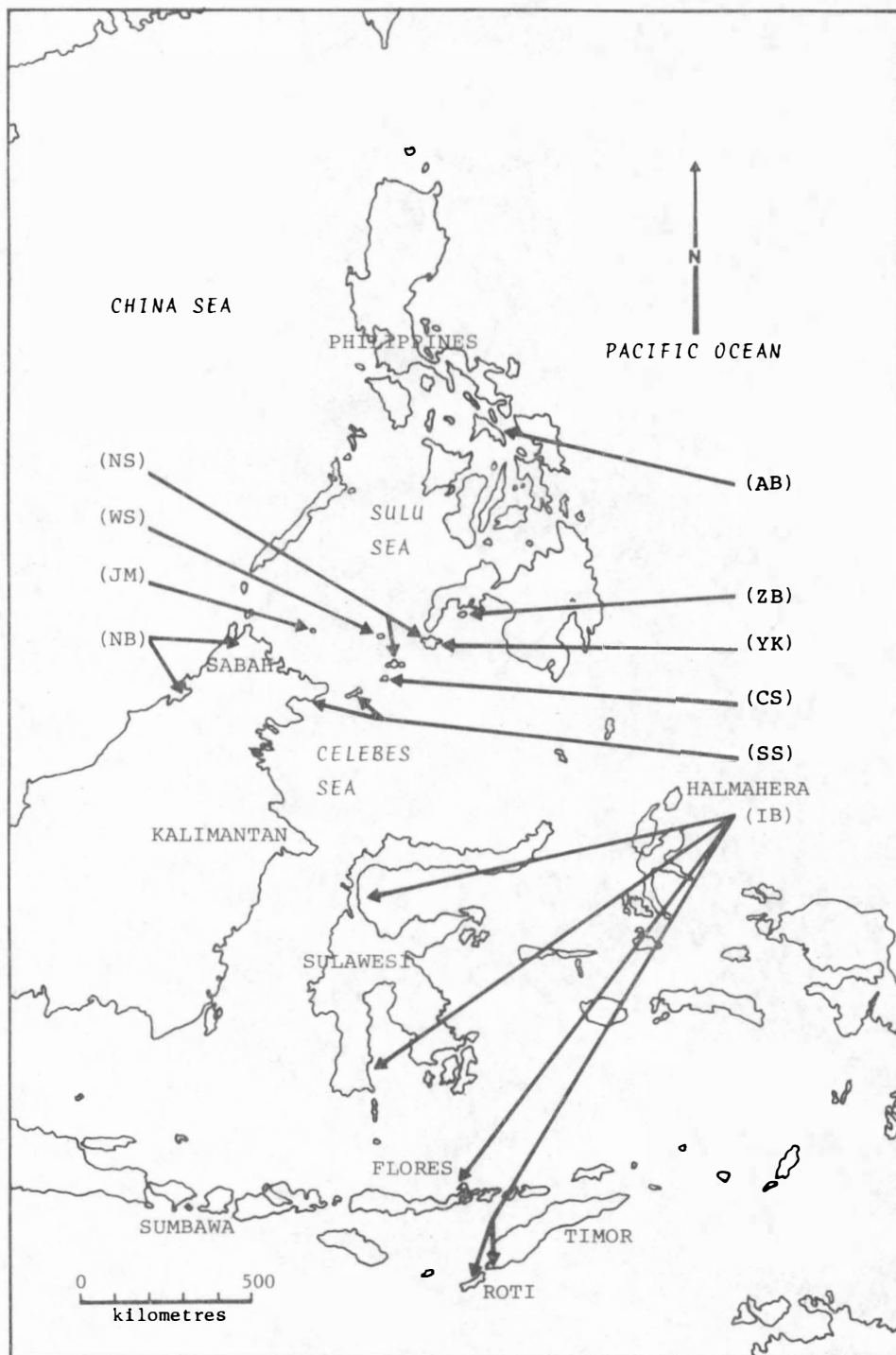
WEINREICH, Uriel

- 1970 *Languages in Contact*. The Hague: Mouton. (1st edn 1953.)

WHINNOM, Keith

- 1956 *Spanish Contact Vernaculars in the Philippine Islands*. Hong Kong: Hong Kong University Press.





DISTRIBUTION OF MAJOR SUBDIVISIONS OF SAMA-BAJAW





# THE PEPET IN SAMA-BAJAW

A. KEMP PALLESEN

1. Topic of discussion
2. Sama-Bajaw as a distinct sub-grouping
3. The data
4. Normal Proto Sama-Bajaw reflexes of PAN \*ə
5. Irregular reflexes
6. Conditioning factors for the PSB reflexes of PAN \*ə
7. The development of gemination in PSB
8. Phonemic status of the PSB reflexes \*ə and \*ɯ
9. Processes by which ə or ɯ is innovated in Sama-Bajaw
10. Reflexes of PSB \*ə and \*ɯ in daughter languages
11. Conclusion

## 1. TOPIC OF DISCUSSION

PAN \*ə, the so-called 'pepet' vowel of Proto-Austronesian, described by Conant (1912) as the "original indifferent vowel... colorless and indefinite in pronunciation", is notable for the variety of its realisations in the daughter languages. This variety is described for Sama-Bajaw<sup>1</sup>, a sub-grouping of Austronesian, where it finds especially rich expression. Some explanatory hypotheses are proposed for the splitting of PAN \*ə into three Proto Sama-Bajaw phonemes.

## 2. SAMA-BAJAW AS A DISTINCT SUB-GROUPING

Sama-Bajaw languages occur in greatest number and diversity along the Sulu Archipelago (Philippines) and up into Sibuguey Gulf on the western end of Mindanao. Other languages of the group are Abaknon (AB)<sup>2</sup> in the

Central Philippines; Sabah Bajau (NB) on the West Coast of Sabah, Malaysia; and a number of closely related but geographically dispersed groups (IB) in Indonesia - in Sulawesi, the Moluccas, and the Lesser Sunda Islands.

The position of the Sama-Bajaw subgrouping within Austronesian has not yet been satisfactorily determined. It is a Hesperonesian sub-group, but the data do not clearly support the generally accepted assignment to the Philippine Hesion<sup>3</sup>. Central Sulu (CS) languages share 31% of cognates with both Malay and Cebuano on a 372 word list<sup>4</sup>, and 38% with Siocon Subanon, a Meso-Philippine language of western Mindanao. Sabah Bajau (NB) shares 37% with Malay and 32% with Siocon Subanon. (The 54% which Central Sulu languages share with Tausug, a displaced Southern Bisayan language spoken in Central Sulu, is due to considerable linguistic convergence.) The sub-group is atypical, though perhaps not unique (if Sangir is a member of the Philippine Hesion), in having \*h as the regular reflex of PAN \*R.

Within the sub-grouping Sama-Bajaw is divided into six sections on the basis of cognacity, syntactic typology, and reflexes of PSB phonemes. The dividing lines of these sections correspond to discontinuities in chains of mutual intelligibility. Since it is not important for the topic of this paper, this internal subdivision is not discussed in further detail, beyond noting that apart from relationships with (AB), which shows a considerable degree of convergence with Central Philippine languages, the lowest internal Sama-Bajaw cognate relationship is between (CS) and Indonesian Bajaw (IB), with 54%.

### 3. THE DATA

Accurately recorded data based on the Elkins word list, and considerably augmented in several cases, are available for all sections of Sama-Bajaw and form the basis of this study. The notation is essentially phonemic unless otherwise indicated. The Proto Sama-Bajaw forms cited are from a corpus of about 2500 reconstructions currently in preparation.

### 4. NORMAL PROTO SAMA-BAJAW REFLEXES OF PAN \*ə

Sama-Bajaw languages exemplify Conant's<sup>5</sup> 'law of gemination'. Gemination in SB is the result of two processes: (1) the gemination of a single medial consonant, and (2) the assimilation of the first consonant of a heterorganic cluster formed by the reduplication of a monosyllable. Non-geminate homorganic clusters are formed by the assimilation of an infix \*ŋ<sup>6</sup> to the point of articulation of the following consonant.

A following homorganic cluster, geminate or otherwise, is the essential environment for the reflection of PAN \*ə as \*ə<sup>7</sup>. Elsewhere, i.e. where this environmental condition is not met, the reflex of PAN \*ə is \*u<sup>8</sup> or \*a. PSB is thus like Malagasy in having diverse reflexes of PAN \*ə in penult and ultima, but it does not conform to Conant's<sup>5</sup> observation that "the Philippines languages, wherever they show this double vocalism, have a in the penult". The phonemic status of \*ə and \*u in PSB is discussed in Section 8.

4.1 The three PSB reflexes of PAN \*ə are now presented and exemplified for the penultimate and ultimate syllables, i.e. for the dissyllabic words which comprise the great proportion of the Proto-Austronesian corpus. The proposal of this paper is that the differentiation of PAN \*ə into the PSB reflexes \*ə, \*u and \*a is most satisfactorily explained by positing contrastive stress in the dialect of PAN<sup>9</sup> from which PSB most immediately derives, at least in those words which contain PAN \*ə. This stress hypothesis is discussed in Section 6.

\*ə is the reflex of unstressed PAN \*ə in the penult; \*a is the reflex of unstressed PAN \*ə in the ultima, when the vowel of the penult is other than PAN \*ə; \*u is the reflex of PAN \*ə elsewhere, i.e. of stressed PAN \*ə in penult and ultima, and of unstressed PAN \*ə in the ultima when the vowel of the penult is also PAN \*ə:

PAN \*Rəbáh 'to throw down, destroy' > \*həbbaq 'to topple'; PAN \*səpsəp 'to suck' > \*səssəp; PAN \*lípət 'to fold' > \*lipat 'to fold in two'; PAN \*línəh 'kind of plant (sesame)' > (\*)lɪŋə<sup>10</sup> 'Sesamum orientale'; PAN \*kəmkəm 'to keep close' > (\*)kumkum 'to hold in closed fist'; PAN(B) \*SəRə(Ct) 'constriction, to tighten' > (\*)hugut<sup>10</sup> 'to be secure'; PAN \*hajəŋ 'charcoal' > \*arun; PAN \*lə(ŋ)bəŋ 'grave, ditch' > \*ləbbun 'to bury'.

4.2 Lack of data makes it impossible to give an adequate description of the PSB reflexes of PAN \*ə in the antepenultimate syllable. The one PAN example which has a PSB reflex is unsatisfactory because of the ambiguity of the Ja. and Ml. evidence for PAN \*ə<sup>11</sup>:

PAN \*sərampaŋ 'sprout, tine' (the citations in VLAW suggest 'fish spear' as a more satisfactory gloss) > \*sahapaŋ 'multi-tined fish spear'. There are numerous comparisons, however, between Malay and Sama-Bajaw which show that SB regularly has \*a where Ml. has ə in the antepenult (the possible loan status of some of the following forms does not affect the point of the relationship between the vocalisms):

PSB \*jambatan 'bridge, pier' : Ml. jəmbatan, jambatan 'bridge'; \*kəranjaŋ 'open-weave basket' < PAN \*kəranzaŋ : Ml. kəranjaŋ 'basket';

(\*salassay < \*(saləssay) 'to settle a dispute' : Ml. sələsay 'to solve, settle, complete'. It seems probable that the correspondence of PSB \*a and Ml. ə in the antepenult reflects PAN \*a.

Although the PAN > PSB evidence is sparse, the internal evidence of Sama-Bajaw supports the reconstruction of five vowels in the antepenultimate syllable: \*a, \*i, \*u, \*ə, \*ʉ. Three of these are of interest here<sup>12</sup>; \*a is included because of the partial overlap of its reflexes with those of \*ə and \*ʉ.

4.21 \*a is posited for the correspondence: e (occasionally a in (YK), ə (occasionally a) in (NB)<sup>13</sup>, a in other languages:

(\*bagunbun 'dust'. (ZB, NS.2-3, CS.5, SS.4) bagumbun, (YK) begumbun, (NS.4-9, WS, CS.2, SS.1-4, Tsg.) bagunbun.

\*saruduŋ 'roofed area'. (YK) seuduŋ 'extension of roof overhang', (CS.2) sauduŋ 'roofed, unwallled part of house', (NB.1) səruduŋ 'cooking area'.

4.22 \*ə is posited for the correspondence: a in (AB), i (with a few exceptions) in (ZB) and (NS-SS), e or i in (YK)<sup>14</sup>, ə in (NB), and a or i in (IB)<sup>15</sup>:

\*dəbuhiq 'last night'. (AB) dabuqi, (YK) debuhiq, (NS-SS) dibuhiq, (NB) dəbuiq, (IB) dabuiq.

\*ənsəllan 'oil'. (YK) isəllan, insəllan, (CS.2,5) ənsəllan, (CS.3, JM) insəllan, (NB) ənsəlan, [IB.9] ansəlan.

4.23 \*ʉ is posited for the correspondence: a in (AB), u in (ZB, YK, NS-SS and JM), u or ə in (NB), a or i in (IB). Only the (AB) and (IB) reflexes distinguish this set from \*u.

\*kʉhəpoq 'grouper, rock cod'. (YK, NS.3, WS.1, CS.2) kuhəpoq, (SS.4) kuhəpuq, (IB.1) kiəpuq.

\*tʉqolan 'bone'. (AB) taqulan, (ZB.3, CS.2-3, SS) toqolan (\*tuqolan).

The internal PSB evidence thus suggests that PAN \*ə in the antepenult split into the reflexes \*ə and \*ʉ, a split analogous to that described for the penult. The primary conditioning factor in the antepenult, however, does not appear to be contrastive stress, as is posited for the penult (in Section 6). The small number of examples permits only the following tentative conclusion at this stage:

In the antepenultimate syllable \*ʉ is the reflex of PAN \*ə before the consonants \*h, \*w and \*l; \*ə is the reflex of PAN \*ə otherwise.

\*h and \*w are consonants which do not geminate, and this limitation inhibits the reflex of PAN \*ə as \*ə in the penult (discussed in

Section 6.1). It is reasonable to suppose that the same inhibiting function is exercised in the antepenult also. \*l follows \*u in two of the six examples where \*u is reconstructible in the antepenult, but there is no equivalent restriction on the gemination of \*l. The inclusion of \*l in the above distribution statement remains unexplained at present.

No gemination takes place following \*ə in the antepenult, contrary to what might be expected from the constraints on the realisation of PAN \*ə as \*ə in the penult. There is apparently an overriding constraint which may be informally stated as "no word may have two sets of geminates". Such a constraint is current in (CS.2), a language which reflects penultimate \*ə as ə: buhat 'to serve refreshments' becomes (with stative prefix mag- and initial CV reduplication) magbubuhat 'one who serves refreshments'. təttaq 'to chop firewood' becomes, by the same affixation process, magtatəttaq 'one who chops firewood'. Where a homorganic cluster (nasal plus stop) follows \*ə in the antepenult the reflex ə does occur in (NS-SS), but not otherwise:

\*əntəllə 'egg'. (AB) antallu, (NS.1-2, CS.2-3,6, SS.1,4-5) əntəllə, (IB.1-2,4) antillo.

\*dəqilaw 'yesterday'. (AB) daqillaw, (ZB.1,3, NS, WS.2, CS, SS.1,3-4) diqilaw, (IB.1,5) dilaw.

## 5. IRREGULAR REFLEXES

Three hypotheses are considered as possible explanations for reflexes of proto-phonemes which are aberrant insofar as they do not conform to the rules stated in Section 4<sup>15</sup>: (1) that the words in question are loans, (2) that the reconstructed proto-forms are incorrect, or incomplete insofar as they fail to indicate phonological variants of the proto-form (doublets), and (3) that there are further developments within PSB, either regular processes of sound changes for which the conditioning factors must be discovered and described, or innovations (apparently unconditioned variants of protoforms). Each of these hypotheses has its legitimate function<sup>16</sup>, but of the three borrowing is the least rigorous and the most open to misuse. Dyen's cautionary word (Dyen 1953) in this respect is worth keeping in mind:

"If phonetically similar phonemes occur in the words of related languages in contact, and their correspondence does not agree with any formula that has been otherwise established, it is not safe to conclude that the words are related only by borrowing. ...it may be discovered that this (borrowing) explanation is insufficient"

5.1 Almost all the exceptions to the rules given in Section 4 are those where PAN \*ə > \*a, a situation similar to that noted by Llamzon (1966) for Tagalog. In this section the general problem of \*a reflexes of PAN \*ə is considered, for ultima as well as penult, since there are plainly possible alternatives to the interpretation that \*a in the ultima is the reflex of unstressed PAN \*ə.

5.11 Malay is an obvious possible source for SB loans; it has exerted considerable influence on Sulu languages for many centuries, and it reflects PAN \*ə as a in the penult. There is, furthermore, evidence that other Philippine languages replace the penultimate or antepenultimate ə of Ml. loans with a (a replacement noted by Dyen (1953) for Ngadju Dayak loans from the same source)<sup>17</sup>. In the following examples from non-Sama-Bajaw languages the supposition that we are dealing with loan words is supported by the occurrence of non-inherited features in addition to the replacement of PAN \*ə with a, i.e. by the presence of a Malay prefix or of a second irregular prefix of a PAN phoneme:

Mar. parampuan 'woman' : Ml. pəməpuan; Tsg. jarat 'noose' : Ml. jarat (the expected Tsg. reflex of PAN \*zərat is (dj)ulat). SB languages show the same pattern:

CS.3 balbaŋsa 'of noble birth' : Ml. bərbəŋsa.

The foregoing evidence justifies consideration of the possibility of borrowing from Malay as an explanation for PAN \*ə > \*a irregularities. It also suggests an alternative source for Tg. katám 'carpenter's plane' (if it is in fact a borrowing), for which Llamzon<sup>18</sup> proposed Kapampangan as a plausible source<sup>19</sup>. A Malay source would also explain the presence of katam in Sulu languages.

Such a hypothesis would be strengthened if a Philippine language were to exhibit the expected regular reflex of PAN \*ə in PAN \*kətəm (e.g. Tg. kitím, CB kutum or PSB (\*kətəm/kəttəm)). This is the case with PSB (\*)ganap and \*gənnap 'to become complete' (Tg. ganap and Ml. gənap) for which VLAW gives only PAN \*gənəp as a possible source. Alongside the PSB doublet, which may be a borrowing, we have the CS.2 form kənnop 'to add to' < \*(kənnəp). Apart from the devoicing of the initial consonant this is a regular reflex of PAN \*gənəp.

There is, however, no proof of the borrowing hypothesis for such problem cases. The apparently irregular form may be a regularly reflected doublet of the proto-language or of some meso-language. As more data show the same 'irregular' reflex occurring in languages not in contact, and as such forms prove to be unambiguously reconstructible for the proto-languages of distinct sub-groupings, a doublet hypothesis gains strength. A borrowing hypothesis, while not ruled out, requires more substantive evidence - or greater ingenuity.

5.12 The problem of PAN \*ə > \*a is one for which alternatives (2) and (3) must be thoroughly investigated before a borrowing explanation is entertained. Alternative (2) - doublets in a proto-language - is supported by a significant number of PAN doublets which show an alternation between PAN \*ə and PAN \*a in both penult and ultima:

PAN \*basuR/\*bəsuR 'satisfied'; PAN \*basih/\*bəsih 'iron'; PAN \*dampul/\*dəmpul 'adhesive'; PAN \*kaRat 'to bite' / \*kəRət 'to cut off'; PAN \*lankep 'to be complete' / \*rankep 'to be combined'; PAN \*tanaq/\*tanəq 'ground, earth'; PAN \*zagzag/\*zəgəg 'to gain a firm footing'.

Of the corpus of PSB words which reflect a PAN form with an infix \*ŋ a remarkably high proportion - over half - show the reflex \*a for PAN \*ə in the penult. Among them are:

PAN \*gəndaŋ 'kettledrum' > \*gandaŋ 'drum; to beat time'; PAN \*qəntih 'to stop, finish' > (\*)hantiq 'to break a journey' (Ml. mənanti 'to wait for', hənti 'to stop'); PAN \*gənDit 'a girdle' > \*kan(dj)it 'scarf; sash'; PAN(B) \*ləŋTik 'graceful curve' > (\*)lantik (Tg., Mar. lantik); PAN \*təmpət<sup>20</sup> 'domicile' > (\*)təmpat 'site where psychic power is concentrated' (Mar. təmpat 'tomb'); PAN \*tə(n)tu 'certainly' > \*tantu (Mar. tanto, CB tantu 'to an extreme degree'; (cf. also PKr. tantu 'certainly' and mutu tantu 'too, overly', and the suggested Port. origin for this disputed reconstruction).

It can hardly be the case that there is any greater inherent probability that such forms will be borrowed from Malay; rather it is evidence of the facility with which \*ə and \*a alternate in this environment, at some stage in the development of PSB.

The \*ə/\*a alternation before an infix \*ŋ was evidently an established pattern at the PAN level. Examples given below are those which have a Sama-Bajaw reflex:

PAN \*lamba(r)/\*ləmba(r) 'material, "piece" as a numerical classifier' > \*lamba 'strand, piece'; PAN \*ma(n)tah/\*məntah 'unripe, raw' > \*mataq; PAN \*sambah 'respect' / \*səmbah 'reverence' > (CS.2) səmba 'to revere'.

5.13 One further set of data remains to be considered, where PAN \*ə > \*a in the final syllable:

PAN \*li(ŋ)kər 'to circle, wind' > (\*)ka)liŋkar 'to surround'; PAN \*lipət 'to fold' > \*lipat 'to fold in two'; PAN \*rapət 'joined together' > (\*)lrapat 'to be joined without any gap' (Mar. lapat 'airtight', Tg. lapat 'fitted in'); PAN \*saŋət 'excessive' > \*saŋat 'intensity; severity'; PAN \*tabəq 'to be fat' > (\*)tabaq 'fat which adheres to flesh' (CB tabaq 'pork fat', Mwa tabaq, Tg. tabáq 'to be fat').

It is proposed in Section 6 that PSB \*a, when it corresponds to PAN \*ə in the ultima, is the regular reflex of PAN [ʌ], the unstressed allophone

of PAN \*ə in the ultima. Such a hypothesis must contend with explanations in terms of borrowing and doublets (the alternatives (1) and (2) above). For the borrowing hypothesis it may be argued that the SB forms in the above examples are exactly what would be expected if they were in fact loans from Malay. One reflex at least, however, cannot be explained as a borrowing from Ml. (nor from a Philippine language which reflects PAN \*R as g):

PAN \*bárəq 'ulcer, abscess' > \*bahaq 'swelling, a boil'.

As has been shown in 5.11 there is ample precedent from PAN for positing the existence of doublets which exhibit the a/ə alternation in the ultima. Examples like PAN \*rapət above, for which diverse Philippine languages reflect a in the ultima, also suggest the possibility of doublets, at least at the level of a Proto-Philippine meso-language.

The conclusions indicated at this stage are: (a) that the presence of undoubted Malay loans in SB and in other Philippine languages, and the pattern of ə to a replacement in such words, support the possibility of borrowing as an explanation; (b) that there is evidence from PAN in support of the hypothesis that some PAN \*ə > \*a forms are the regular reflex of previously unidentified PAN doublets; (c) that in view of the high frequency of forms showing the a/ə alternation before a homorganic nasal-stop cluster some at least of these must be inherited from a language ancestral to PSB, and (d) that the data suggest that some of the a forms are either PSB innovations, or the result of regular phonological processes in the ultima. The arguments for the latter proposal are presented in Section 6.3 and are not repeated here.

The choice between the competing hypotheses is not resolved. Each case must be considered separately; but any borrowing hypothesis can be sustained only by good evidence.

5.2 Irregularities other than PAN \*ə > \*a make up a small residue, all of which is now considered. In the following material the expected reflex, indicated by [...], follows the actual PSB or daughter language form:

(\*)bəkəs \*[bak(a)əs], (CS.2) bekos [bak(ao)s] 'to be marked by transverse bands' : PAN(B) \*bakəs 'belt'. This is probably a PSB innovation, with \*ə replacing the vowel of the penult<sup>21</sup>. The (CS.2) form is a secondary development, perhaps via \*(bekəs).

(C) (YK) dellət \*[d(ei)llet] 'noose' : PAN \*dirət 'rope' (cf. PAN \*zərat 'loop'). Probably a PSB innovation with vowel replacement and stress shift, from \*(dərrət).

\*hadən \*[həddan/hədan] 'ladder' : PAN \*həzan. A PSB innovation by vowel metathesis.



\*intaq \*[əntaq] 'to eat raw' : PAN \*qə(n)taq 'to be raw'. A PSB innovation which is unusual with regard to the replacement of PAN \*ə by \*i.

\*kehet \*[kəhət] 'to cut up; to slice' : PAN \*kəRət 'to cut off'. Probably a PSB innovation, without parallel other than the next example.

\*keket \*[kəkət/kəkkət] 'to bite' : PAN \*kəkət 'to hold fast'. Ho. hehitrah (VLAW) 'grip, bite' confirms the semantic connection. This and the preceding example display the regular (YK) reflex of the posited \*(kəhət) and \*(kəkət), but it is unlikely that the (YK) form was the loan source for other PSB languages, in view of the wide spread of unambiguous reflexes of the reconstructions given.

\*kijət \*[kədut/kəddut] 'small muscular movement; twitch' : PAN \*kədut 'to smart' (VLAW gives Ja. kədut 'muscular jerkings'). It is not clear that the PAN and PSB forms are directly related, though a doublet may be involved. Cf. PSB \*kuddət/kəddut 'startled movement', and PAN(B) \*kə(z)u(Ct) 'pinch' (with Singhi kujet noted as metath.).

(NS, CS) leṅog [dɔŋog/dəŋnog] 'noisy, disturbed' : PAN \*dəŋəg 'to hear'. Probably not inherited from this PAN form. The CB data, liṅug-liṅug 'to ignore by pretending not to hear' and duṅug 'to hear', suggest a doublet, PAN \*(liṅəg).

(CS.2) lipaq [lɔpaq/ləppaq] 'to apply cosmetic paste' : PAN \*ləpah 'paste'. Probably a reflex of a doublet with PAN \*(lipaq) which VLAW suggests somewhat uncertainly (cf. Ml. lepa 'plaster, mortar').

(YK) luḡuṅ [ləḡuṅ/ləḡḡuṅ] 'thunder' : PAN(B) \*ləḡuṅ 'booming sound'. Probably the reflex of a doublet \*(luḡuṅ). Cf. Mar. ruḡuṅ and WBM ruḡuṅ 'thunder', and the same vocalism in CB huḡuṅ 'produce a steady humming sound' and in Tg. huḡoṅ 'noise like thunder'.

\*pəssil \*[pəsəl/pəssəl] 'to apply pressure with the fingers; to massage' : PAN(B) \*pəcəl 'to squeeze in the hands'. Possibly a PSB innovation, but cf. the final vowel of CB písil, písil 'to squeeze s.t. with the tips of the fingers wide apart'.

(\*)sande(lr) \*[sənd(əw)] 'to derive support from' : PAN \*səndəR 'lean against'. Probably the reflex of a doublet. Cf. Batak sandir 'to lean against a wall', and CB sandig 'to lean against'.

(\*)sumbaliq \*[səmbəlliḡ] 'to slaughter by draining blood' : PAN \*səmbəlih 'to slaughter according to Moslem rites'. Probably a Philippines innovation. Cf. Mar. sombaliq, WBM sumbaliq.

(CS.2) tanog [tonog/tənnog] 'resonant' : PAN \*tənəR 'voice'. Probably a PSB or CS innovation. Cf. CB tunug 'widely known, resonant'; Tg. tinig 'voice'; which are regular reflexes of the PAN form.

\*timbak \*[təmbək] 'to shoot a firearm' : PAN \*təmbək. Either a borrowing from Tg. (timbang) (unrecorded), or the reflex of a Philippines innovation. Cf. Mar. timbak.

\*tiŋur \*[təŋur/təŋŋur] 'tanbark mangrove (*Ceriops* sp.)' : PAN \*təŋer. Probably a PSB innovation.

(\*)tuman \*[tuman/təmman] 'certain; to fulfil' : PAN \*təman 'to be accustomed'. Possibly a borrowing from CB tuman 'extreme', or from Mar. toman 'true; to fulfil' (in which the expected Mar. reflex would be (təman)). Possibly also the reflex of a doublet \*(tuman). Cf. the Ja. doublet: təmən 'to be faithful'; tuman 'accustomed'.

## 6. CONDITIONING FACTORS FOR THE PSB REFLEXES OF PAN \*ə

Dempwolff<sup>22</sup> noted the occurrence of stress in the three primary languages of his study, but concluded that it was not possible to compare stress in a useful way. Zorc (1972), in a paper entitled 'Current and Proto Tagalic Stress', raises the question:

"If... one posits stress as predictable and hence non-phonemic for Proto Malayo-Polynesian, how is it that it does become phonemic in so many daughter languages in the Philippines?"

In a concluding paragraph, having noted the need to posit both penultimate and ultimate stress for Proto Tagalic, and the existence of at least one minimal pair, he raises the further question:

"Did Proto Tagalic inherit or innovate its ultimate stresses. A careful investigation of the lexicons of Tagalic and other Philippine languages is called for. As time progresses and areas of research increase we should be able to discover just how far back in time we can postulate phonemic stress."

It is here proposed that differential stress in PAN<sup>9</sup> was the primary conditioning factor for the splitting of PAN \*ə into the PSB reflexes \*ə, \*u and \*a. According to this hypothesis, allophonic variations of PAN \*ə existed as a result of contrastive stress, such that unstressed PAN \*ə in the penult was manifested by the normative allophone [ə]; unstressed PAN \*ə in the ultima was manifested by the lowered allophone [ʌ] when the vowel of the penult was other than PAN \*ə, and by the raised allophone [i]<sup>23</sup> when the vowel of the penult was also PAN \*ə. Stressed PAN \*ə in both penult and ultima was manifested by the raised allophone [i]. The distribution of the posited allophones corresponds to the distribution of the PSB reflexes as stated in Section 4.1, so that PAN [ə] > \*ə; PAN [ʌ] > \*a, and PAN [i] > \*u, as in the following:

PAN		PSB	
*lɔŋah [lɔŋah]	'sesame'	> *lɔŋa	'sesame'
*gətáh [gətah]	'sap'	> *gəttəq	'sticky sap'
*tÉRəb [tɪRɪb]	'quantity, mass'	> (*)təgəb <sup>24</sup>	'heavily loaded'
*ləŋén [ləŋɪn]	'forearm'	> *ləŋŋən	'upper arm'
*púter [putʌr]	'to turn around'	> *putar	'to revolve; to unwind'

6.1 Stress is indeterminate for words where other factors supercede stress as the primary conditioning factor. Two such cases are suggested by the data.

6.11 Stress is indeterminate when PAN \*ə in the penult is followed by a consonant which does not geminate in PSB, e.g. PSB \*h. The reflex of PAN \*ə as \*ə is inhibited, since a following homorganic cluster is an essential condition for this reflex in the penult:

PAN \*bəRas 'husked rice' > \*bəhas  
 PAN \*bəRsay<sup>25</sup> 'canoe paddle' > \*(bəhsay) > \*busay

There are thus two processes by which \*u could have been derived from PAN \*ə in the above examples. If stress occurred on the penult the posited PAN [ɨ] allophone of PAN \*ə would have been reflected as \*u; had stress occurred on the ultima the unstressed PAN [ə] allophone would likewise have been reflected as \*u, due to the limitation just described.

6.12 Stress is also indeterminate for forms with an infix PAN \*ŋ following penultimate PAN \*ə. The present data provides no evidence of contrast between \*ə and \*u in this environment. This appears to be more than a data limitation, and the possible phonetic explanation is suggested in Section 7.2. No stress has been assigned to examples such as those given in 6.11 above, or to the following:

PAN \*kə(m)bar 'twin' > \*kəmar, (\*)kambar; PAN \*gəntiŋ 'thin' > \*gəntiŋ 'slim-waisted'.

6.2 The differentiation of stress carries a low semantic load:

PAN			
*təŋáh	'middle'	> *təŋaŋ	'middle (range)'
*təŋah	"	> *təŋaŋ	'middle (point)'
*lí(ŋ)kər	'to circle, wind'	> (*) (ka)liŋkaɪ	'to surround'
*li(ŋ)kér	" " "	> *(liŋkəɪ)	(CS.2) lenkol 'halo about sun'
*bə(n)təŋ	'stomach'	> *bətəŋ	'stomach, bulge, pregnant'
*bá(n)təŋ	"	> (*) bətəŋ <sup>26</sup>	'mature, green coconut'
*ləpás	'loose, free'	> *ləppas	'emerge; pierce through'
*ləpas	" "	> (*)ləpas	'to be lost'
*(təgəl)	('persistent')	> *təggəl	'long period of time'
*(təgəl)	( " )	> *(təgəl)	(CS.2) togoɪ 'persist'

In some cases the reflexes of stress-differentiated PAN pairs show no current semantic contrast in SB languages:

*kəbál 'invulnerable' >	*(kəbbəl) (YK)	kebbel
*kəbəl "	> *(kəbəl) (CS.2)	kobel
*huláj 'maggot' >	*(ulət) (YK)	olet
*húləj "	> *(ulat) (NS, CS)	ulat

6.3 As has been mentioned in Section 5.13 there are competing hypotheses for the proposal that PAN \*ə in the ultima had stressed and unstressed allophones [ɨ] and [ʌ] which were reflected in PSB as \*u and \*a respectively. The following arguments are not conclusive, but they warrant serious consideration of the proposal.

6.31 The parallelism between the split in the penult and the posited split in the ultima is phonetically plausible. In both syllables the phonetically higher reflex is that of the posited stressed allophone of PAN \*ə. The unstressed allophones and their PSB reflexes differ in phonetic height, with the lower occurring in the ultimate syllable.

6.32 If two languages, not closely related and not in contact, share a distinctive phonological phenomenon there is reason to seek an explanation in similar phonological processes. Tagalog, for which the regular reflex of PAN \*ə is i, and Sama-Bajaw both show significant numbers of words which exhibit ultimate a instead of the expected reflex of PAN \*ə. Llamzon<sup>27</sup> cites some doublets which parallel the SB data:

batís, batás 'to ford a creek' < PAN \*batəs 'boundary'; hilís, hilás 'cut'; lapit<sup>28</sup> 'close, draw near', lapat 'fitted in' < PAN \*rapət 'joined together'; sapit 'arrived', sapat 'attached'.

One would expect, if doublets are the correct explanation for these a reflexes in both languages, to find evidence of the same doublets in other languages of the Central and Southern Philippines. This evidence does not appear to be abundant in Cebuano Bisayan, but more data are needed, both from CB and other languages such as Bikol, Maranao, Subanon and Sangir.

If \*(baRaq) 'ulcer, abscess' is reconstructed as a doublet (alongside PAN \*baRəq) on the evidence of Tg. bagáq 'abscess' and PSB \*bahaq 'swelling; boil', we must suppose the persistence of both members of the doublet down to the Proto Tagalic stage at least. Cf. CB baguq 'disease characterised by... swelling in the region of the stomach' (and also CDu bagəq 'to be swollen').

The persistence of such doublets is not unprecedented, to be sure, but it is significant that the two languages which justify the posited doublet are languages which exhibit similar alternations between a and the regular reflex of PAN \*ə. It thus seems reasonable to consider the

possibility that we are not dealing here with a doublet but with a distinctive, shared phonological development.

6.33 The hypothesis of contrastive stress in the ultima is consistent with the alternations displayed within PSB, not all of which have an established PAN source. In the first example below the shift of stress to the ultima is accompanied by reduction of original PAN \*i in the penult to \*ə<sup>29</sup>:

PAN	*lípət	'to fold'	>	*lipət	>	'to fold in two'	
	*(ləpət) <sup>30</sup>			>	*ləppət	>	'to fold up'
	*(táhəp)	'to winnow'	>	*tahap			
	*(taháp)			>	*(tahəp)	>	(AB) tahup

## 7. THE DEVELOPMENT OF GEMINATION IN PSB

Universal penultimate stress is reconstructible for PSB. It is therefore proposed that the posited contrastive stress of PAN became non-contrastive in PSB. In all the PSB daughter languages which reflect it ə is a short vowel. It does not occur with the vowel lengthening which is the usual concomitant of stress in the daughter languages (and presumably reconstructible as a feature of stress in PSB also). It has been proposed in Section 6 that PSB \*ə only occurred as the reflex of unstressed PAN \*ə in the penult. It is further proposed that with the shift of stress from the ultima to the penult the short vowel was retained and that compensatory length was carried by the following consonant, thus causing gemination. Thus PAN \*bəkás 'inheritance' becomes, with the regularisation of stress, PSB \*bək:as 'trace of former occupancy'.

7.1 In (NB), the only languages which do not have universal penultimate stress<sup>31</sup>, stress falls on the ultima only when the vowel of the penult is ə. In (NB), furthermore, the length of the medial geminate is substantially reduced, so that it no longer carries the length feature of the stressed syllable. Stress thus reverts to the ultima. In the following examples the raised period indicates a very slight lengthening of the consonant; rather more when the ə is initial:

PSB \*əmmaq 'father' > (NB) əmaq<sup>32</sup> [m·áq] and > (CS) əmmaq [ḿmaq]; PAN \*bə(n)tis 'shin-bone' > \*bəttis 'leg' > (NB) bətis [bət·ís] and > (CS) bəttis [béttis].

(CS) languages display an extreme degree of transference of syllabicity from stressed ə to the following consonant. The phonetics below are simplified with respect to the allophones of the other segments:

PSB \*ənnuṃ 'six' > (CS.2) ənnom [ṇnom]; PSB \*əmpit 'to hold a child (on the lap)' > (CS) əmpit [ṃpit]; PSB (\*)əggok 'to make a gulping motion' > (CS.2) əgguk [ˈg̊guk] (the initial g is slightly imploded); PSB (\*)əttus 'disease which pocks the skin' > (CS) əttus [ˈt̪tus] (tongue tip is held in the stop position for the normal duration of a stressed vowel).

7.2 Where PSB reflects a PAN word in which PAN \*ə in the penult occurred before an infix \*ŋ, only \*ə occurred as the reflex of PAN \*ə. This is not surprising; the nasal carried the compensatory length when stress occurred on the penult, and the normative vocalism of PAN \*ə was maintained. This environment, it is presumed, conditioned the allophones of PAN \*ə so that [ə] (rather than [ɨ]) was the allophone of stressed PAN \*ə in the penult before a nasal-stop cluster. Since both stressed and unstressed PAN \*ə in this environment have the reflex \*ə it is not possible to determine stress.

#### 8. PHONEMIC STATUS OF THE PSB REFLEXES \*ə AND \*u

We now consider the question of the phonemic status in PSB of the PSB reflexes \*ə and \*u of PAN \*ə. It has been proposed in the previous section that contrastive PAN stress became non-contrastive in PSB. The contrastive function of stress in PAN was thus taken over by phonemically contrastive vowels, as witnessed by the phonemic status of the reflexes of \*ə and \*u in all daughter languages except (YK)<sup>33</sup>. PSB thus represents the completion of a transitional period during which contrastive stress plus sub-phonemic vocalic contrast gave place to non-contrastive stress and phonemic vowel contrast. The gemination of the medial consonants was a by-product of the change in the pattern of stress placement, and played only a secondary role in marking phonemic contrast.

#### 9. PROCESSES BY WHICH ə OR u IS INNOVATED IN SB

There are four processes by which ə or u is developed from vowels other than PAN \*ə. In some cases the development is reconstructible for PSB; in others it appears to have taken place subsequent to the subdivision of PSB.

9.1 The first of these innovations is based on the analogy of stress-conditioned reflexes of PAN \*ə in a reduplicated monosyllable. The effect of contrastive stress on the PSB reflexes is displayed in the

following examples; with ultimate stress the heterorganic cluster of reduplicated monosyllables is assimilated to a geminate and the vowel of the penult is reflected as \*ə:

PAN \*kémkəm 'to keep close' > (\*)kəmkməm 'to hold in closed fist';  
 PAN \*səpsép 'to suck' > \*səssəp; PAN \*kəpkəp 'to grasp' > (\*)kəpkəp,  
 but PAN \*kəpkép > (\*)kəkkəp.

A similar process based on this pattern may operate on a reduplicated monosyllable with a vowel other than \*ə. The medial cluster becomes a geminate, and the vowel of the penult is sometimes innovated as ə:

PSB \*kəkbab 'to fan' > \*(kəkkab) > (YK) kekkab; (\*)kutkut 'to scratch' > \*(kəkkut) > (ZB.3) kukkut, and (CS.2) kakkut (where the vowel of the penult in both cases is a secondary development); \*səssək 'narrow gap' and (CS.2) soksok < \*(səksək) suggest a back formation on the same analogy, although the (CS.2) form may be original, lost in other languages to the innovating modification.

Assimilation of a heterorganic cluster does not, however, obligatorily require this vowel replacement:

(\*)bugbug 'starch food cooked till mushy' > (YK) bubbug 'rice gruel'.

The alternation also occurs within a single daughter language, marking minor semantic modification:

(CS.2) loklok 'to withdraw into a hole in order to avoid detection';  
 (CS.2) ləllək 'to be just below surface, as a nail driven into wood'.  
 (The CS.2 vowel o is a merger of two PSB proto-vowels, \*ʉ and \*o, so that the above forms may be reflections of PSB \*(lʉklʉk).)

9.2 The second innovation is based on the alternation between \*ə and some other vowel (not a) in the penult, with the regular gemination of the medial consonant after the innovated ə. There are precedents for this pattern in PAN:

PAN \*kumuR, \*kəmuR 'to move to and fro in the mouth'.

In some cases the innovation is of PSB provenance:

(\*)bəttad 'to stretch out; to set s.t. out' and (NS, CS) bitad 'to pull taut'. Cf. PAN(B) \*bi(ŋ)CtT)ad 'spread out'.

(\*)bulud 'mountain' (in NS.7-10, WS, CS, SS.1-3) and \*bəllud 'mountain' (in ZB, NS.2-3, SS.4-5, NB, IB).

The alternation is also exploited to mark semantic modifications within a single daughter language, apparently as a fairly recent innovation. A few examples are presented from CS.2, where the alternation is abundant although no longer an active process:

babas 'to abate, as wind'	bəbbas 'to abate, of swelling'
bakat 'cut in flesh'	bəkkat 'to break a line'
kilat 'lightning; to flash'	kəllat 'to show light, of sun'

kulit 'outer layer, skin'	kəllit 'leathery'
tigaŋ 'dried out'	təggaŋ 'barren womb'
tubud 'to well up, as water'	təbbud 'to billow up, as smoke'

9.3 PAN \*a in the ultima before final PAN \*h is regularly reflected as PSB \*u<sup>34</sup>:

PAN \*kimah 'giant-shell' > \*kimu 'giant clam'; PAN \*duwah 'two' > \*duwu; PAN \*taliŋah 'ear' > \*taliŋu.

9.4 PAN \*a in the penult sometimes becomes \*u when the vowel of the ultima is the \*u reflex of PAN \*ə, and the medial consonant is \*l or \*r:

PAN \*balés 'to retaliate' > \*balus or \*bəlus 'to reciprocate an action'; PAN \*tazəm 'sharp' > \*tarum or \*tərum.

#### 10. REFLEXES OF PSB \*ə AND \*u IN DAUGHTER LANGUAGES

PSB had seven phonemic vowels after the regularisation of stress placement on the penultimate syllable:

i		u
e	ə	o
	a	

Two daughter languages, (ZB.1, WS.1) have retained this seven-vowel system. (ZB.2-3, NS, WS.2, CS, SS, JM, NB, IB) have all reduced the PSB proto-system to a six-vowel system by merging \*u with another phonemic vowel (with either \*a, \*o or \*u). Some (NS) and (IB) languages are in a stage of transition between six and five vowels, as ə, a phonetically unstable segment, merges diversely with other phonemic vowels. (YK) has a five-vowel system resulting from the merger of both \*ə and \*u with \*e, and (AB) has a three-vowel system resulting from the merger of \*o and \*u with \*u, of \*ə with \*a, and of \*e with \*i.

10.1 The reflexes of PSB \*ə and \*u are now presented for each of the daughter languages (V<sub>1</sub> refers to the vowel of the penult; V<sub>2</sub> to the vowel of the ultima):

10.11 In the antepenult \*ə > a in (AB); > i in (ZB, NS-SS, JM), usually; > e or i in (YK); > u or ə in (NB); > a or i in (IB):



\*təmboroq 'finger'. (AB) tamburuq 'hand', (ZB) tomboq < \*(tomboloq) < \*(təmboloq), (YK) timbooq 'index finger', (NB) təmboro 'index finger'.

10.12 In the penult \*ə > a in (AB); > ə in (ZB, NS-SS, JM, NB, IB), with secondary modifications as indicated for (ZB, NS, IB); > e in (YK):

\*bəttuŋ 'stomach; pregnant'. (AB) battuŋ, (ZB.3, NS.2, CS, SS.2, JM) bəttuŋ, (WS.1) bəttuŋ, (WS.2) bəttuŋ, (NB.1) bətuŋ, (IB.8) bətah.

\*əmmaq 'father'. (ZB.1, NS-SS, JM) əmmaq, (NB) əmaq, (IB.1) əmmaq 'mother'.

10.121 In (ZB.2, NS, IB) ə is an unstable segment, and secondary processes are still continuing, by which ə is replaced or merged with an existing phoneme. In (ZB.2) and (NS) two processes operate and compete with each other and with the inherited reflex ə. ə usually persists when it is initial in a word; elsewhere ə is sometimes (apparently unpredictably) replaced by e (especially in NS.6), or ə assimilates to V<sub>2</sub> (especially in ZB.2, NS.1,9-10):

\*ləŋŋuŋ 'upper arm'. (ZB.1, WS.1) ləŋŋuŋ, (ZB.2, NS.2-3,7-10, CS.4, SS.4) loŋŋon, (ZB.3, CS.1-3,6, SS.1,3) ləŋŋon, (YK) ləŋŋen, (WS.2) ləŋŋun.

\*bəlla 'to cook'. (ZB.1, NS.3,6-7, WS.2, CS.2-6, SS.1,4, JM) bəlla, (ZB.2-3, NS.10) balla, (YK, NS.3,5,8) bella.

10.122 In (IB) ə usually persists when it is initial in a word; elsewhere it is usually replaced by i when V<sub>2</sub> is \*u or \*a, and sometimes replaced by i or u in other environments:

\*bəlla 'to cook'. (IB.1) billa.

\*bəlli 'to buy'. (IB.1) bəlli, (IB.2) billi.

\*bənnəd 'muscular cramp'. (IB.1) binnaq.

\*bərrat 'heavy'. (IB.1) birraq, (IB.4) bərraq.

\*dəppe 'fathom'. (IB.1) duppa, [IB.9] dəpe.

\*əllum 'to be alive'. (IB.1) əllumaŋ, əlluŋ.

\*kəlluŋ 'neck'. (IB.1) killoh, (IB.2) kulloq, [IB.9] kəlo.

10.21 In the antepenult \*u > a in (AB); > u in (ZB, YK, NS-SS, JM); > u or ə in (NB); > a or i in (IB):

\*dəwaqi 'to descend'. (AB) dawaqi, duwaqi, (ZB, YK, NS-SS, JM, NB) duaqi.

\*kəhəpoq 'grouper; rock cod'. (YK, NS.3, CS.2) kuhəpoq, (IB.1) kiəpuq.

10.22 In the penult \*u > u in (AB, WS.2); > u in (ZB.1, WS.1); > o in (ZB.2-3, NS, CS, SS, JM, NB); > e in (YK); > o in (IB) when both V<sub>1</sub> and V<sub>2</sub> are \*u (there are no data for \*u in penult and a different vowel in the ultima):

\*butaŋ 'to place; to install'. (AB) butaŋ, (WS.1) butaŋ, (CS.2-3) botan.

\*duhuŋ 'to desist from doing s.t.'. (YK) deheŋ, (NS.1, CS) dohoŋ, (IB.1) doqoh 'to abate, of a storm'.

\*bulus 'to reciprocate an action'. (AB) bulus 'to take revenge', (NS.1, NB.1) boos 'to borrow', (WS.1) buus 'to borrow'. (NB.4) bohos < \*(boos) 'to take revenge'.

10.23 In the ultima, in an open syllable, \*u > a in (AB, ZB, NS-SS, JM, IB); > e in (YK); > o in (NB):

\*haya 'big'. (AB, ZB.1-2) haya, (YK) haje, (CS, SS) hea, (NB) oyo.

\*sawa 'snake'. (AB) sawa, (ZB, NS-SS, IB.1) soa, (YK) sawe, (JM, NB) soo.

In the ultima, in a closed syllable, \*u > u in (AB, WS.2); > u in (ZB.1, WS.1); > o in (ZB.2-3, NS-SS, JM); > e in (YK); > a in (IB):

\*anum 'to weave a mat'. (ZB.1) anum, (ZB.3, NS.3, CS.2-3, SS.2-4) anom, (ZB.2, WS.2) anum, (NB.1-2) ŋonom, (IB.1,4-5) anaŋ.

\*dakup 'to seize'. (CS.3) dakop 'to abduct a woman', (NB.1-2) .dəkop, dokop, (IB.9) dakap.

\*ənnum 'six'. (AB) annum, (ZB.1, WS.1) ənnum, (ZB.2, NS.3,5,10) onnom, (ZB.3, NS.2,7-9, WS.2, CS, SS, JM) ənnom, (NB.1,3-4) ənom, (IB.1,5) ənnaŋ.

10.24 In (NB) a secondary replacement of the o reflex of \*u is taking place, such that o > a.<sup>34</sup> There is considerable alternation between a and o in (NB), in both directions, possibly under the influence of cognates with Malay which show a regular correspondence between a and o in the ultima.

\*kəttup 'to slice, as meat'. (NB.1) ŋətap 'to chop small'.

\*ənnum 'six'. (NB.1,3-4) ənom, (NB.1-2) ənam, Ml. ənam.

\*(h)inŋət 'to tie'. (NB.2) eŋkot, (NB.1) inŋkaq, əŋkaq, Ml. ikat.

\*sərrət 'tightly wedged'. (NB.1) sorot, (NB.2) sərət, Ml. sərət 'jammed'.

\*səssər 'to compel s.t. to move'. (NB.1) səsar, səsor 'to drive away', Ml. səsar 'to push s.t. aside'. Cf. also Ml. pərtama 'first', borrowed as (NB.1) pərtomo.

## 11. CONCLUSION

In order to facilitate the description of the Proto Sama-Bajaw reflexes of PAN \*ə - an interesting and rather complex set of data - we have proposed that contrastive stress was a feature of a language ancestral to PSB, though not necessarily of Proto-Austronesian. Most of the SB data are explicable in terms of this hypothesis; a borrowing or doublet hypothesis is indicated for only a small proportion of the corpus.

However, satisfactory though the hypothesis appears to be for the PSB data, it needs to be tested against the data of other PAN subgroupings. Of particular value, it seems, are the data of those languages in which PAN \*ə splits into two or more phonemes, and those languages which exhibit secondary developments such as gemination of the medial consonant after penultimate PAN \*ə. Another useful set of data may be provided by languages which, like Sama-Bajaw, have phonemic contrast between u and o, and between i and e. Such splits may be due in part to contrastive reflexes of stress-conditioned allophonic variations of PAN \*u and PAN \*i.

## NOTES

1. Sama is a widespread autonym (from Sulu, Sabah and Indonesia), and is reconstructible as the proto-form by which speakers of the language referred to themselves. Bajaw is the Malay designation used throughout Malaysia and Indonesia, and for one group in the Southern Philippines. The compound name of the title has been chosen to facilitate identification, and at the same time to retain the autonym.

Other names which occur currently and in the literature are Orang Laut, Samal (the usual external designation for Sama in the Philippines), Luwaqan (nomadic Philippines Sama) and "Sea Gypsies".

2. The following conventions have been followed: Proto-Austronesian reconstructions have been written in Dyen's orthography with the exception that the pepet is written as \*ə, and the symbol e retained to indicate the mid-front vowel [e] which is phonemic in Sama-Bajaw. ʷ indicates a high back-central unrounded vocoid; ɨ indicates a high central unrounded vocoid. I have followed Blust in writing the PAN infixed nasal as \*ŋ in his reconstructions, but have not modified the VLAW material. Malay citations have been modified by writing the non-syllabic high vocoids as -w and -y, the digraph ng as ŋ, and e as ə where it represents the mid central vowel.

Since we are dealing with two levels of reconstruction the convention is followed of prefixing all Proto-Austronesian forms with the abbreviation PAN, and only prefixing Proto Sama-Bajaw forms with PSB when there is a possibility of confusion. PSB forms well-attested in Sulu languages, but unattested outside them, are indicated by (\*), and hypothetical forms by \*(...); thus PAN \*bélah 'to split' > \*(bɨlɨq) > (CS.2) bɨlɨq 'bamboo lath'. Primary stress is marked by a grave accent over the vowel of the stressed syllable. [...] enclosing a language code indicates that there is some uncertainty about the phonetic interpretation of some part of the following citation. Language citations appear

bold type and are identified by a preceding letter code in the case of Sama-Bajaw languages, or by an abbreviation.

Sources and abbreviations are as follows:

PAN: Proto-Austronesian (Dempwolff); PAN(B): Proto-Austronesian (Blust); PSB: Proto Sama-Bajaw; Batak (Rodda); CB: Cebuano Bisayan (Wolff); CDu: Casiguran Dumagat (Headland & Headland); Ho.: Hova (VLAW); Ja.: Javanese; Mar.: Maranao (McKaughan and Macaraya); Ml.: Malay, Indonesian (Echols and Shadily); Mwa.: Mamanwa (Miller and Miller); Minahasa and Togian Bajaw (Adriani); PKr.: Papia Kristang (Hancock); Proto East Mindanao - a subgrouping of Southern Bisayan which includes Mamanwa and Tausug (author's fieldnotes); SB: Sama-Bajaw daughter languages, collectively considered; Skr.: Sanskrit (Gonda); Tg.: Tagalog (English); Tsg.: Tausug (Taqusuug) (Ashley); VLAW: *Vergleichende Lautlehre des Austronesischen Wortschatzes*; WBM: Western Bukidnon Manobo (Elkins).

Sama-Bajaw daughter languages have been assigned letter codes, with following letters indicating dialect or sub-dialect variations within each coded section. The expansion of the code is indicated as a mnemonic aid: (AB-Abaknon): Abaknon or Capulenyu (Capul Island, Central Philippines, between Masbate and Northern Samar); (ZB-Zamboanga): Batuan, Lutangan and Sibukuq Sama (Sibuguey Gulf and Zamboanga Peninsula, Mindanao); (YK-Yakan): Yakan (Basilan Island); (NS-Northern Sulu): Baangiq and related dialects (northeastern Sulu, with migrant groups in other parts of the Philippines); (WS-Western Sulu): Pangutaran and Soang Bunaq (islands about 35 miles north-west of Jolo Island in Sulu); (CS-Central Sulu): Sama Dilaut (nomadic and semi-nomadic, from eastern Sabah to Zamboanga) and several related dialects spoken by sedentary groups (central Sulu); (SS-Southern Sulu): Tandug Baas, Simunul and Sibutuq (southwestern Sulu and eastern Sabah); (JM-Jama Mapun): Mapun (Cagayan Sulu Island, north of Sandakan, Sabah); (NB-North Borneo): West Coast Sabah Bajaw (Labuan Island and the coastal fringe from Papar to Banggi); (IB-Indonesian Bajaw): Indonesian Bajaw (Indonesian Timor, Roti Island, Kajoa in the western Halmaheras, and various places around the Sulawesi coast). (CS.2) indicates Sama Dilaut of the Siasi Lagoon area, Siasi, Central Sulu, the language with which the writer is best acquainted. (NS-SS) includes (NS, WS, CS, SS), a single section characterised by a chain of mutual intelligibility. Sulu languages include (ZB, YK, NS-SS, JM).

The following acknowledgements are made of unpublished sources: James Fox (Roti data (IB.3)), Charles Frake and Carol Molony (Kajoa, Halmahera data (IB.2)), Gerard Rixhon (Sulawesi data (IB.4-6)), and various members of the Summer Institute of Linguistics, Philippines, especially Dietlinde Behrens and Janet Pack (YK), Jeanne and Helen Miller

(Mamanwa), Charles Walton and William Hall (various Sama word lists) and Siocon Subanon (Hall).

3. Dyen (1965) assigns one language of the sub-group, Yakan, to the Philippine Hesion as a first order sub-grouping, and indicates the problematic nature of the assignment by a double asterisk.

4. The word list is a list adapted by Richard E. Elkins from the Swadesh list for the use of the field workers of the Summer Institute of Linguistics in the Philippines. A comparison of cognate counts between the Swadesh and the Elkins list indicates a difference of  $\pm 4\%$ . The Elkins list forms the basis of *Philippine Minor Languages: Word Lists and Phonologies* (Reid, ed.).

5. Conant op. cit.

6. Blust's (1970) assumption has been followed that the data are best explained by positing a single segment \*ŋ which was sometimes infixated preconsonantly in PAN, but Dempwolff's reconstructions have not been rewritten in accordance with this assumption.

7. A minor exception to this statement occurs in the antepenultimate syllable, where PSB \*ə occurs without a following homorganic cluster. This case is discussed in Section 4.2.

8. Since the daughter languages show a wide diversity of reflexes of PSB \*ɯ (e, a, o, u) there was no clear phonetic identity for this proto-form. ɯ was chosen for pragmatic reasons: (1) it is a unique symbol for the correspondence set, and (2) in the only two languages which maintain the seven-vowel phonemic system of PSB the current phonetic realisation of this proto-phoneme is [ɯ], a high back-central unrounded vocoid. In view of the other reflexes of PSB \*ɯ no assumption is made about the rounding of the proto-phoneme.

9. Since Dempwolff's and Blust's reconstructions are based essentially on languages of Dyen's Hesperonesian Linkage, the proto-language of the forms indicated as PAN in this paper are more correctly of Proto-Hesperonesian provenance. The ancestor of PSB for which contrastive stress is posited is thus Proto-Hesperonesian or some lower order proto-language. More work is needed before such mid-level groupings can be satisfactorily identified; the reconstruction of reliable forms for low level proto-languages is a step in this direction. The label PAN in this paper, when it refers to the stress hypothesis, should be read as shorthand for "PAN or some lower order proto-language of Proto-Hesperonesian".

10. These examples, because of their relatively limited range of occurrence, are susceptible to interpretation as loans from a language which has ə or ɯ reflexes of PAN \*ə in both penult and ultima, i.e. a language such as Proto East Mindanao and its daughter language Tausug. Upland Tausug maintains the contrast between ɯ and u, lost in Lowland Tausug. There are, however, other data showing this vocalism which have good PSB credentials though without an identified PAN source: \*duhɯŋ 'to desist from doing something'.

11. Ml. and Ja. ə in the antepenult may derive from either PAN \*ə or \*a, an ambiguity which Blust indicates where appropriate, in his Adenda. Malay regularly, though not invariably, replaces original a with ə in the antepenult: mərdeka 'free' (Skr. maradeka); kəpala 'head' (Skr. kapāla 'skull'); fatwa (Ar.) 'instructions' and pətua '(religious) advice', where the Arabic word has been modified in the second form to conform with Ml. phonology.

12. \*u and \*i do not constitute a problem since their reflexes are invariable in all PSB languages (except (NB), which has extensive vowel reduction in the antepenult).

13. (YK) and (NB) are similar to Malay in having a rule by which phonemic a is modified in the antepenult: (YK) saqi 'mother' and seqinen 'his mother'; (NB.1) papag 'to beat' and pinəpagan 'beaten'. Cf. (CS.2) saqina 'his mother' and binabagan 'provided with a cross-beam' (from babag 'cross-beam'). The above illustrations reflect only part of a more complex process in both languages. There are other constraints for (YK) on the reduction of a to e, and for (NB) other vowels and other environments are implicated.

14. (YK) reflexes are about evenly divided between e and i, and (IB) reflexes between a and i. There is, however, no correlation between these sets of variations, and insufficient data in either case to suggest any conditioning hypothesis. It should be noted that in (IB) \*ə has multiple and largely unpredictable reflex potential in the penult as well. In some cases there are currently competing forms: (IB.1) limbaq, ləmbaq, lumbaq 'to repeat'.

15. I have not included analogic replacement of phonemes since, except for those cases where innovations can be shown to have spread by some storable process (Zorc's (1972) morphological regularisation of stress to indicate morphemic contrast between segmentally identical forms is a good example), this explanation is frequently a less satisfactory treatment of the data than simply noting that the problem is unexplained.

16. The correct solution may, of course, require a combination of these hypotheses. Doublets may not be of proto-language provenance, but may be the result of early inter-dialect or inter-language borrowing. Once borrowed they become inherited forms of the daughter languages, subject to native laws of sound change. Again, since sound change processes lose their impetus or change phonological direction, the phoneme X of a loan word may undergo different (though regular and describable) processes from those which operated earlier on a phonemically identical phoneme in an inherited word.

17. An alternative explanation is that the Malay source was a dialect of M1. which realised standard Malay ə as a, although this is not the case for current contacts between Sulu languages and Malay.

18. Llamzon, *op. cit.*

19. It is not necessary to suppose that a source language form was borrowed in the form in which it currently appears. Lowland Tausug, which reflects PAN \*ə as u in all positions, and once assimilated the ə of borrowed Sama forms in the same way (Tsg. kuddum 'to frown' probably from PSB \*kəddəm 'to close the eyes'), now assimilates borrowings from Sama languages in a different way, generally by harmonising Sama ə with the final vowel of the word:

Sama dənda 'woman' becomes danda; Sama pəssi 'fishhook' becomes pissi; Sama əmboq 'ancestor' becomes umbuq when spoken by Tausug who do not control the phonology of Sama.

20. PAN \*təmpət 'domicile' is poorly attested in VLAW.

21. Innovated ə is discussed in Section 9. This is a fairly typical example of the process described in 9.2.

22. Dempwolff (1934-8), *op. cit.*

23. The harmonising of the vowels in this context may have been a PSB process; the evidence is ambiguous.

24. This word is possibly a loan from a Philippine language in view of the \*g reflex instead of the expected \*h reflex of PAN \*R. Medial PAN \*R is reflected as PSB \*g more frequently than initial and final PAN \*R, and there may be conditioning factors. Cf. also such forms as \*pəggaq < PAN \*pəRaq 'to squeeze out juice', and the doublet \*taguq/ta(h)uq 'to keep', both well attested for PSB. The doublet probably reflects, respectively, PAN(B) \*(CtT)aguq 'hidden' and PAN \*taRuq 'to set down'.



25. Dempwolff (1925) cites this form, but only \*bəsay is given in VLAW. The \*R infix is supported by such Philippine reflexes as Proto East Mindanao \*bəgsay 'canoe paddle'.
26. Both PSB forms appear to derive from PAN \*bətəŋ. The form (\*)bətəŋ 'mature, green coconut' though an innovation with this meaning, may be an example of the exploitation of stress contrast. The semantic connection is at least plausible, with a shared component of swelling and fruition.
27. Llamzon, op. cit.
28. Blust (1970) proposes an alternative etymology for Tg. lapit: PAN(B) \*ra(ŋ)pi(Ct) 'approach'.
29. See Section 9.2 for a discussion of this innovation pattern.
30. The change in stress here may be an innovation of early PSB provenance, and the form in parentheses here is included only to facilitate comparison with the established PAN form. No claim for PAN status is intended.
31. I have no direct data for (AB) stress, though the material was recorded by an experienced field worker who would, I expect, have indicated contrastive stress. In any case there is no evidence in the corpus of some six hundred (AB) words of homophonous pairs for which one might expect stress to be the contrastive feature, and it is assumed with reasonable confidence that the above statement is equally applicable to (AB).
32. The phonemic status of the single versus geminate consonant contrast in (NB) is uncertain. It carries a low contrastive load in any case, even when the vowel of the penult is other than ə.
33. (YK) reflects PSB \*ə and \*u as e in all environments (merging these both with \*e), and \*a as e under predictable, stress-governed conditions.
34. The PSB reconstruction of this phoneme is based on the correspondence set: -e in (YK), -o in (NB), -a elsewhere. The (YK) form is crucial for the identification of this set, since the (NB) evidence is somewhat ambiguous. In (NB) there is a secondary development whereby a > o in an open final syllable, and penultimate a > o when the final vowel is o. Thus:
- \*mata 'eye' > (NB) moto; \*hadən 'ladder' > (NB) odon.

A similar centralising process takes place with PAN \*u and PAN \*i before final PAN \*h or PAN \*q, with these variants reflected in PSB as \*o and \*e.

REFERENCES

ADRIANI, N.

- 1900 *Tijdschrift voor Indische Taal-, Land- en Volkenkunde*.  
Deel XII/460-90. Batavia, Albrecht and Co.; 's Hage: M.  
Nijhoff.

ASHLEY, S.

- 1975 *Tausug-English Dictionary*. In press.

BLUST, Robert A.

- 1970 Proto-Austronesian Addenda. *Oceanic Linguistics* 9(2):104-62.

CONANT, Carlos Everett

- 1912 'The pepet law in Philippine languages'. *Anthropos* 5:920-47.

DEMPWOLFF, Otto

- 1925 'Die L-, R- und D-Laute in austronesischen Sprachen'. *Zeitschrift für Eingeborenen-Sprachen* 15:19-50, 116-38, 223-38, 273-319. Berlin: Dietrich Reimer.

1934- *Vergleichende Lautlehre des Austronesischen Wortschatzes*.

1938 Band I-III. Berlin: Dietrich Reimer.

DYEN, Isidore A.

- 1953 'Dempwolff's \*R'. *Language* 29:359-66.

1965 *A Lexicostatistical Classification of the Austronesian Languages*. Indiana University Publications in Anthropology and Linguistics, Memoir 19 of the *International Journal of American Linguistics*. Baltimore: Waverly Press, Inc.

ECHOLS, John M., and Hassan SHADILY

- 1961 *An Indonesian-English Dictionary*. New York: Cornell University Press.

ELKINS, Richard E.

- 1968 *Manobo-English Dictionary. Oceanic Linguistics Special Publication No.3.* Honolulu: University of Hawaii Press.

ENGLISH, Leo James

- 1965 *English-Tagalog Dictionary.* Department of Education, Republic of the Philippines. Clayton, Victoria: Wilke and Co., Ltd.

GONDA, J.

- 1952 *Sanskrit in Indonesia.* Nagpur: The International Academy of Indian Culture.

HANCOCK, Ian F.

- 1973 'Malacca Creole Portuguese: a brief transformational outline'. *Te Reo* 16:23-44.

HEADLAND, Thomas N., and Janet D. HEADLAND

- 1974 *A Dumagat (Casiguran) - English Dictionary.* PL, C28. The Australian National University.

LLAMZON, Teodoro A.

- 1966 'Tagalog reflexes of PMP \*ə'. *Anthropological Linguistics* 8/3:13-23.

McKAUGHAN, Howard P., and Batua A. MACARAYA

- 1967 *A Maranao Dictionary.* Honolulu: University of Hawaii Press.

MILLER, Jeanne, and Helen MILLER

- n.d. *Mamanwa-English Dictionary.* Summer Institute of Linguistics. Typescript.

REID, Lawrence A., ed.

- 1971 *Philippine Minor Languages: Word Lists and Phonologies. Oceanic Linguistics Special Publication No.8.* Honolulu: University of Hawaii Press.

RODDA, Rosemary, and Audrey MAYER

- n.d. *Batak-English Dictionary.* Summer Institute of Linguistics. Typescript.

ZORC, David R.

- 1972 'Current and Proto Tagalic Stress'. *Philippine Journal of Linguistics* 3/1:43-57.

# ACRONYMIC PATTERNS IN INDONESIAN<sup>1</sup>

SOENJONO DARDJOWIDJOJO

## 1. INTRODUCTION

If we look at the shaping of a language such as Indonesian, we can readily see that this language arose out of circumstances under which a vehicle of communication was desperately needed among people with different language backgrounds (Alisjahbana, 1957; Dardjowidjojo, 1967; Halim, 1972). Despite the fact that this situation fits fairly well with Hockett's definition of a pidgin (Hockett, 1958, p.422), no one has ever volunteered to label Indonesian as a pidgin language, even when we know that some foreign languages, notably Sanskrit (Gonda, 1952), "helped shape" the language substantially.

When the term *Melayu* was changed into *Bahasa Indonesia* (*Indonesian language*) in 1928 and then adopted as our national language in 1945, Indonesian assumed a function much heavier than it had ever shouldered before. The use of Dutch was soon banned, and the Japanese occupation of the country could not change the fate of the course. Indonesian soon became *the* language to be used among different ethnic groups and in official communications. In academic circles, Indonesian had to progress as fast as the scientific endeavours demanded it to.

While at the moment we cannot say that Indonesian has achieved a standard by which we can measure all and any deviations as being non-Indonesian, we can certainly say that we are in the process of standardising and modernising our national language. In order to cope with the multi-directional demands, Indonesian not only has to borrow new terms for new concepts from other languages - *data*, *sensus*, *fonim*, *bisnis* to cite but a few - but it also has to intensify internal creations.

This paper is to look into one type of internal creation which I have called acronymisation. While I am fully aware that acronymic processes are found in virtually all languages, it seems to have a unique role and significance in the development of Indonesian. It is true that the acronymic phenomena have been in the language for a long time, but they did not become popular and productive until the 60s. In military academies, for instance, cadets are taught these acronyms as part of their courses (Departemen Angkatan Darat, 1968). The increase of military role after the 1965 abortive coup d'etat - and therefore the increase of acronyms found in mass media - only added fuel to the flame. People began to acronymise anything acronymisable and "play" with this new "in-thing". Puns began to appear. Among the Javanese, for instance, the acronym *pentilkecakot* '*telephone inspector for the city sub-district*' was coined from the full Indonesian forms *penilik tilpun kecamatan kota*, despite the fact that there is no such a position existing. It is apparent that this acronym was made merely in fun since the forms *pentil kecakot* do constitute real Javanese words meaning '*nipples unintentionally sucked*'.

While quite a number of these acronyms is found only in written forms, many are used orally as well, especially in cases where they show congruities with the phonotactic rules of the language. Thus, forms such as *pangkopkamtib* '*Commander of the Operation Command of the Restoration of Safety and Order*', which is derived from *panglima komando operasi pemulihan keamanan dan ketertiban*, are very much on the written side of the language, whereas *berdikari* '*to stand on one's own feet*', which is derived from *berdiri di atas kaki sendiri*, is very common both in its oral and written forms.

## 2. DESCRIPTION

Indonesians do not make a distinction between acronyms and abbreviations. The term *singkatan* '*shortened form*' is used to refer to both.<sup>2</sup> In this paper I will use the terms acronym and abbreviation interchangeably to represent the sense of *singkatan*.

In terms of familiarity and well-establishedness, we can classify acronyms into two major categories: (i) those acronyms which have been used in the language for a long time, and (ii) those which have been recently coined. Members of the first category are known virtually to every literate Indonesian and they have been used consistently by everyone in the country. Forms such as *kpd*, *tsb*, *a.l.*, *AURI*, *yth*, for instance, are not only known by Indonesians, but each form represents the same full form and the same semantic concept, namely, *kepada* '*to(ward)*',

tersebut 'previously mentioned', antara lain 'among others', Angkatan Udara Republik Indonesia 'Air Force of the Republic of Indonesia', and yang terhormat 'Dear (So and So)' respectively.

The second category is very interesting to observe, because, while it is productive in its own right, it also bewilders even native speakers living in the country. This paper will be limited only to this second category.

## 2.1 ACRONYMIC SYLLABLES

Basically there are two ways to coin an acronymic syllable: (i) by observing some kind of syllabification of the words to be abbreviated, and (ii) by taking into account the graphemic representation of these original words.

There are several subtypes which belong to the first type above. Perhaps the most common of all is that the acronymic syllable is derived from the first syllable of the full form. Thus, *or-*, *mu-*, and *bi-* of *orba* 'new order', *muker* 'work conference', and *Bima* 'Blue coloured night train' are derived from the full forms orde, musyawarah, and biru respectively.

A second subtype of an acronymic syllable consists of the last syllable of the full form. The acronymic syllables *-dan*, *-pur*, and *-yon*, for instance, are used to represent the full forms komandan, tempur, and bataliyon to form the acronymic words *dandim* (komandan distrik militer) 'commandant of the military district', *zipur* (zeni, tempur) 'combat troop', and *danyon* (komandan bataliyon) 'battalion commandant', respectively.

The third subtype requires that the original full form ends in a consonant. If the first syllable of the full form has a CV, the acronymic syllable is formed by taking this CV plus the last C of the full form. This brings about the existence of forms such as *dit-*, *dir-* and *ban-* of *ditjen* 'director<sup>ate</sup> general', *dirjen* 'director general', and *banser* 'multi purpose troop' where *dit-*, *dir-* and *ban-* are derived from the underlined parts of the full forms direktorat, direktor, and barisan respectively. *Ditjen*, *dirjen*, and *banser* come from *direktorat jendral*, *direktor jendral* and *barisan serba guna*.

Although there are not many examples found, there seems to be a tendency to apply this rule where the first syllable of the full form begins with a vowel. In this case the acronymic syllable is formed by taking the first V and the last C of the full form. Thus, '*inspector*' and '*engineer*' are abbreviated as *ir* as in *Irjen* (Inspektur Jendral) '*Inspector General*' and *Ir. Soekarno* (Insinyur Soekarno) '*Engineer Soekarno*'.

The fourth subtype involves cases where the last letter of the full form is a vowel. If the first letter of the first syllable of the full form is a consonant, the acronymic syllable is coined by juxtaposing the first consonant and the last vowel. Thus, kepala '*chief*', dua '*two*', and tertinggi '*highest*' are respectively abbreviated as ka as in KASAD (Kepala Staff Angkatan Darat) '*Army Chief of Staff*', da as in Letda (Letnan Dua) '*Second Lieutenant*', and ti as in Koti (Komando Operasi Tertinggi) '*the Highest Operation Commando*'.

While we have seen cases where the initial letter(s) of a full form is used to make an acronymic syllable with or without any additional letter(s), we have not seen cases where the last letter of the full form is used to represent the whole full form. I have so far found only one example where the last letter is used to represent the whole. The full form cepat '*fast*' is abbreviated into t as in Kopasgat '*Commando of the Fast Moving Troop*' which stands for Komando Pasukan Gerak Cepat.<sup>3</sup>

As we have seen from the above examples, virtually all of what Fries called content words (Fries, 1945), which are abbreviated, are represented in the acronyms one way or another. There are cases, however, where a content word is deleted from the acronym. In the case of Menlu '*Minister of Foreign Affairs*', where men = menteri '*minister*', the use of lu to represent luar '*outside*' and negeri '*country*' can perhaps be explained on the basis of redundancy. The collocation of menteri and luar forces people with no choice but to add the word negeri obligatorily since menteri luar by itself does not occur in the language and that the only possible menteri luar is menteri luar negeri.

In most cases, however, there does not seem to be any explanation available. The use of ser as in banser given above to represent the words serba '*various*' and guna '*use*' is a case in point. The morpheme serba in the context of banser can collocate with several other possible words such as neka '*type*', lengkap '*complete*' etc. and would still produce meaningful acronyms.

Still in some cases, not only is the deleted word a content word, but that that content word happens to be very crucial. Thus the acronym konjeran '*Commando of the Mine Sweepers*' lacks the acronymic form for penyapu '*sweepers*' because konjeran is supposed to represent komando jenis penyapu ranjau.

The role of the root whose derived form becomes the source of an acronymic syllable also seems important. We have cases where the acronymic syllable is not derived from the full form per se but from the root underlying the full form. The word for money, uang, for instance, is often abbreviated as u irrespective of the actual full derivative form. Thus the Academy of Finance and Banking is abbreviated as Akubang



where ak = akademi, u = keuangan, and bang = bang. And Ekubang '*Economics, Finance, and Development*' comes from Ekonomi, Keuangan, dan Pembangunan.

All of the acronyms given above are based on segmental features. There are cases where the determining factor seems to be the location of the stress, which normally falls on the penultimate syllable. The daily word pertahánan '*defence*' is abbreviated into han as in hankam (pertahanan dan keamanan) '*defence and security*', hansip (pertahanan sipil) '*civil defence*' and hanra (pertahanan rakyat) '*people's defence*'.

Virtually all of the examples that we have had so far look and/or sound real Indonesian words, that is, they fit very nicely with the word structure of the language. The graphemically oriented acronymic syllables seem to be based on the wish - unconscious as it may be - on the part of the inventor to give people some hint so that they could probably guess what the acronym may have been derived from. Some of these acronyms still conform to the phonotactic rules of the language. Thus for '*General Election*' and '*(Some kind of) Prime Minister*' people use pemilu (pemilihan umum) and Menutama (Menteri Utama) instead of any other possible acronyms such as \*pemu and \*Menut - the latter two conforming fully also to the Indonesian phonotactic rules.

In our attempt to give people hints, we occasionally run into problems. Some of the acronyms come in conflict with the well-established norms. The coining of brig for brigadir in brigjen (brigadir jendral) '*brigadier general*', may for mayor in mayjen (mayor jendral) '*major general*', bant for bantuan in kojebant (komando jenis bantuan) '*supporting commands*' must have been based on the wish of the inventor to "help" people out. In our attempt to give people hints, we coined the acronyms brigjen, mayjen, and kojebant, but these forms violate Indonesian phonotactics - the juxtaposition of gj, yj, and nt as a final consonant cluster is not phonotactically justified.

From the foregoing analysis we can see that while the acronymic phenomena in Indonesian are rather hectic, there are "guidelines" - however inconsistent they may be as we will see later on - which people say they follow. However, there are a few cases where the acronymisation does not follow any of the patterns we have established so far. An example of this "deviation" is the acronym jubir '*spokesman*', which is derived from juru '*expert*' and bicara '*speak*'. While the ju of juru follows the regular acronymic rule, the bir of bicara is unique in that it takes the first CV bi and a consonant, r, from somewhere among the rest of the elements in the original word.

The same phenomenon also occurs in the previously cited Kopasgat. While the ko and the pas (plus the t of gat as discussed before) are

normal, the use of *ga* to represent *gerak* 'move' is unique.

A similar situation is also found in the use of *kam* to represent *keamanan* 'safety'. Here the first *C* is taken and then followed by the first syllable of the root *aman* - which is *a* - and closed with the first consonant of the second syllable of the root - which is *m*.

Finally, there is a unique case where each acronymic element represents a rather complex concept. Graphically, the element represents a phrase or a sentence. The acronym *USDEK*, for instance, represents

U = Undang2 Dasar 1945	'the 1945 Constitution'
S = Socialisme Indonesia	'Indonesian Socialism'
D = Demokrasi Terpimpin	'Guided Democracy'
E = Ekonomi Terpimpin	'Guided Economy'
K = Kepribadian Indonesia	'Indonesian Identity'

## 2.2 ACROTACTIC

The term *acrotactic* is used here to refer to the ways in which acronymic syllables are combined to form acronymic words. Since the degree of acceptability of an acronymic word is related to the similarity or dissimilarity between it and the Indonesian word structure, it is necessary to sketch very briefly some aspects of the word structure of Indonesian relevant to our present discussion.

Basically Indonesian has a relatively simple syllable structure: (i) CV, (ii) CVC, (iii) VC, and (iv) V. Early and recent contacts with other non-Indonesian languages have made the language acquire other syllable structures such as CCV, CCVC, CCCV, and CCCVC. We notice here that no consonant clusters occur at the end of a syllable. No voiced stops,<sup>4</sup> voiced or voiceless affricates occur in syllable final positions. And finally, there are not many cases where two vowels, especially if they are the same vowels, occur one after the other.

There are several generalisations which we can make regarding the shapes of the acronymic words. First, the relatively simple but inherent canonical forms of Indonesian definitely exert a structural pressure on the shapes of the acronymic words. The bulk of the acronyms in Indonesian today results from the combination of two of these: CV, CVC, VC and V. Some examples,

CV + CVC	: muker from musyawarah kerja 'work conference'
	dubes from duta besar 'ambassador = great envoy'
	caper from calon perwira 'candidate for officer'

- CV + CV : kotl from komando operasi tertinggi '*highest operation commando*'  
 Bima from Biru Malam '*Blue coloured Night Train*'  
 pati from perwira tinggi '*high ranking officer*'
- CVC + CVC : parpol from partai politik '*political party*'  
 Golkar from Golongan Karya '*technocrat group*'  
 cerpen from cerita pendek '*short story*'
- CVC + CV : turba from turun ke bawah '*fact finding*'  
 hanra from pertahanan rakyat '*people's defence*'  
 letda from letnan dua '*second lieutenant*'
- VC + CVC : ormas from organisasi massa '*mass organisation*'  
 orpol from organisasi politik '*political organisation*'  
 atmil from atase militer '*military attaché*', etc.

It is obvious that the most important factor in coining an acronym is what the end result will sound or look like and not what particular element or elements from the original full forms should be taken. Thus forms such as muker, caper, Golkar not only follow the Indonesian phonotactic rules, but that each of the acronymic syllables happens also to be a real Indonesian syllable. However, this is not true for dubes and cerpen where the original words besar and cerita, which are normally cut into be-sar and ce-ri-ta, are abbreviated as bes and cer respectively, thus leaving at less than a syllable and ita one and a half syllables.

Another example which is rather extreme is the acronymic word kostrad '*Commando of Strategy of the Army*'. While ko, a, and d are normal, representing komando, angkatan, and darat respectively, the str is just a mere string of letters, unpronounceable and foreign. And yet kostrad is a very well-known and well-used word, orally as well as in written form.

The second generalisation involves the juxtaposition of two vowels. The fact that two same vowels very rarely occur consecutively compels the language users to avoid as much as possible an acronym with a  $V_1$  and  $V_1$ . Thus for Atase Angkatan Laut '*Naval Attaché*', and Akademi Angkatan Bersenjata Republik Indonesia '*Academy of the Armed Forces of the Republic of Indonesia*' people use Atal and Akabri where the t and the k may have been used as separators of the two as.

If the two vowels are dissimilar, the above rule may not be followed. Thus acronyms such as Koarsa (Komando Armada Siaga) '*Active Fleet Commando*', Aip (Ajun Inspektur Polisi) '*Adjunct Police Officer*', Koanda (Komando Antar Daerah) '*Inter-regional Commando*' are found.

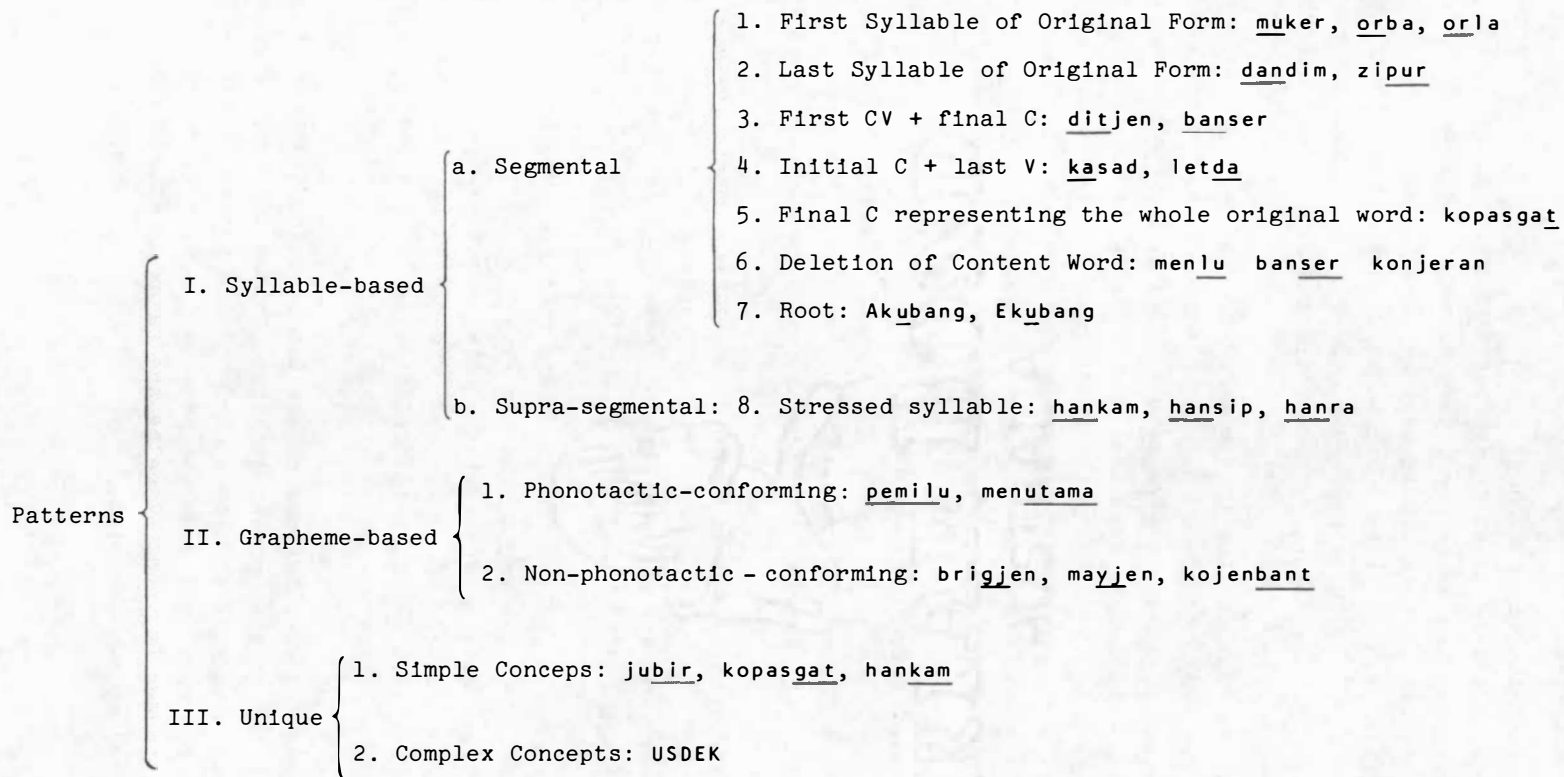
Another generalisation that we can make pertains to the number of syllables in the acronyms. Just as is the case with the Indonesian words, the bulk of the acronymic words in Indonesian also has either two or three syllables. A one syllable acronym is extremely rare. Acronyms with more than three syllables are also found occasionally, and they usually follow the basic syllable structure of the language. So we have acronyms such as kojarsena (korps pelajar serba guna) '*Students' Multipurpose Corps*', Menabungka (Menara Bung Karno) '*Bung Karno's monument*', sendratari (seni drama dan tari) '*Indonesian ballet*', etc.

In some cases the acronyms are not only long but they also deviate very much from the normal Indonesian word structure. The acronym pangkopkamtib which we gave earlier, in addition to having four syllables, also looks and sounds as if it were a foreign form which can be cut into four words pang, kop, kam, and tib. The acronym lfalpolekrochsosbud reported by De Vries (De Vries, p.341) must be not only the longest acronym but also one of the most "un-Indonesian".

### 2.3 THE SYLLABLE STRUCTURE OF THE ACRONYMS

We can summarise the syllable structure of the acronyms as follows:

(see chart on following page)



### 3. DEGREE OF ACCEPTABILITY

Irrespective of the actual shapes of the acronyms, be they normal or deviant, Indonesians tend to at least tolerate their influx. Mass media, in particular the newspapers, take advantage of this situation since it gives them practicality as well as brevity.

How acceptable an acronym is seems to be determined by the nature of the acronym itself. If the acronym deviates very much from the normal Indonesian form, people tend to reject it on the oral aspect but tolerate it in written form.

If the acronym "sounds nice to the ears" people tend to treat it as if it were a real non-acronymic Indonesian word. The acronyms pemilu, repelita (rencana pembangunan lima tahun) '*five year development plan*', Golkar and many others have been used regularly in speech.

The willingness of the people to treat these acronyms as real words brings about a rather interesting syntactic phenomenon. Semantic aspects permitting, many of these acronyms are subjected to normal syntactic rules of the language. The acronym berdikari cited earlier is now used as a verb as in

- (1) Indonesia harus bisa berdikari. '*Indonesia must be able to stand on her own feet.*'

Some of the acronyms are subjected to morphological processes. At the time when former Foreign Minister Soebandrio was about to be tried, slogans such as

- (2) Soebandrio harus dimahmilubkan. '*Soebandrio must be court-martialed.*'

were seen, where the acronym mahmilub (Mahkamah militer luar biasa) '*special military court*' was used as a verb base and affixed with di- and -kan.

A chairman of an organisation who feels that his organisation is being infiltrated by the communists can say

- (3) Organisasi kami digerpol oleh orang komunis. '*Our organisation is politically sabotaged by the communists.*'

where gerpol (gerilya politik) '*political guerilla*' is used as a verb with the passive prefix di- and a slightly different meaning from the original.

For space-saving purposes newspapers use acronyms that sometimes lead to the point of incomprehensibility. Unless one keeps abreast with the continuous acronymic processes, he - even a native speaker - will find that he can read, but does not necessarily understand what he is reading. Sentences (4) and (5) below are taken from newspapers, but (6) and (7) are my own creations.

- (4) Pangdak VII/Jaya Irjen Pol. Drs. Soekahar sekali lagi menegaskan bahwa Komdak VII/Jaya tidak akan segan2 untuk ... (from *Merdeka*, June 16, 1970).
- (5) Dalam rangka penyelesaian tapol G.30.S/PKI di daerah Jawa Barat, dalam waktu dekat Laksus Pangkopkamtib Jabar akan segera ... (from *Merdeka*, June 16, 1970).
- (6) Tapol2 G.30.S/PKI golongan Lekra yang sudah dimahmilubkan dan dibebaskan boleh masuk orpol2 atau ormas2 baru untuk ikut pemilu tahun depan.
- (7) KAMI, KAPI, KASI dan Kojarsena harus bisa berdikari dalam pelita, demikian kata Pangdam IV Jateng Mayjen Gombloh Surodirjo.

All the underlined words above are acronyms.

#### 4. DIRECTIONALITY AND PREDICTABILITY

From the foregoing analysis we can see that the directionality and the predictability of the acronymic phenomena, both from the receptive and productive sides, cannot be easily determined, to say the least. It is true that, due to their frequency of occurrence or some kind of regularity, some acronyms can be "deciphered" or coined relatively easily. The frequent usage of forms such as *han*, *dit*, *dir*, and *kam*, for instance, enable people to know what these acronyms stand for in combination with other acronymic forms, despite the fact that they are derived in rather unique ways. Forms such as *bimas*, *tapol*, and *Golkar* are easy to understand because, in addition to their frequent usage, they are formed on the basis of the syllables of the original words.

The fact that some acronymic syllables have been used rather consistently enables also people - with some luck - to coin new acronyms. Thus, if the term for '*political pressure*', which is *tekanan politik*, becomes popular, perhaps the acronym coined would be *tepol*, where *pol* is already "accepted". If for any reason an acronym is needed for *pertahanan kota* '*city defence*', the coined term would probably be either *hanko* or *hankot*.

In most cases, however, the matter is not very simple at all. On the "decipherisation" side, several problems can be readily seen. To begin with, given an acronymic form - be it a word or a syllable - we cannot tell if this form stands for one word or a string of words. Given the forms *dan*, *mil*, *jen* we are told that each stands for *komandan*, *militer*, and *jenderal* respectively. But the forms *dim*, *rem*, and *kop*

stand for more than one word each, namely *distrik militer*, *resot militer*, and *komando operasi pemulihan*. The problem increases when these acronyms are combined with other acronyms to form still new acronyms.

As mentioned before, another problem that we have involves the inconsistency of the coining process. It often happens that a single semantic concept is represented by more than one acronymic form. Thus, *militer* is abbreviated into *mil* as in *koramil* (*komando rayon militer*) '*commando of a military sub-subdistrict*', into *m* as in *kodam* (*komando daerah militer*) '*commando of a military region*', and into *mi* as in *mahmilub* (*mahkamah militer luar biasa*) '*special military court*'.

We have also noticed earlier that *serba guna* '*multi purpose*' was abbreviated into *ser* in *banser* and *sena* in *kojarsena*. Another example is the acronym for *komando*. While in virtually all cases it is abbreviated as *ko*, it becomes *kon* in *konjeran* - perhaps for homorganic reasons.

The reverse of the above situation is also true, that is, two or more different semantic concepts being represented by one and the same acronymic form. We recall that *ti* of *koti* '*highest commando of operation*' stands for the full form *tertinggi*, but *ti* is also used for *tingkat* as in *Daswati I* (*Daerah Swatantra Tingkat I*) '*Autonomous Region Level I*'. The form *rem* mentioned earlier stands for *resot militer*, but this same acronym is also used for *resimen* '*regiment*'.

The active coining aspect of the acronyms is also problematic. I may have given the impression that the patterns for acronyms sketched above are definitive rules. In a way they are. But the problem is that we do not know exactly what or which particular words or phrases are to be subjected to which rule(s). We recall, for instance, that there is a pattern which says that an acronymic syllable can be formed by having the first CV plus the C which closes the original word. Thus *direktur*, *direktorat*, and *barisan* are abbreviated into *dir*, *dit*, and *ban* respectively. This rule, however, is not followed all the way. The acronym for *komandan*, for instance, could have been *\*kon*, instead of *dan*, and *\*kondim* sounds as homorganic as the accepted *dandim* (*komandan distrik militer*) '*commandant of the military district*'.

## 5. COGNITIVE REASONING

There is no doubt that the reason why people acronymise forms is well rooted in their instinctive desire to follow what Zipf has called "principle of least effort" (Zipf, 1949). While this principle is not a basic requirement for human survival, it is definitely a universal path that every human being chooses when faced with a problem to solve. Acronymisation is only a very minor sample of this human instinct. It is found in any language of wider communication.



The question that interests us, then, is not why people acronymise, but rather why they acronymise the way they do? I believe there is a possible answer for this question, although I must admit that it only answers partially.

As we have seen in Section 2.2 the creation of acronyms seems to be based almost exclusively on the norms which inherently exist in the language and, therefore, shared by members of the speech community. We must hold this responsible for the fact that the bulk of the acronymic word structure and the number of syllables in the acronym conform very much to the Indonesian counterparts. This is also the factor which makes native speakers say "sounds nice to the ears" when asked why a particular word is acronymised in a particular way. Acronyms such as *bimas* (*bimbingan massal*) '*mass guidance*', *turba*, *pemilu*, *menutama* etc. must have been based on this principle.

In some cases the acronyms are coined in such a way that they also constitute real Indonesian words - of course, with different meanings. The choice of *pelita* (*pembangunan lima tahun*) '*five year development*', *Jaya* (*Jakarta Raya*) '*Greater Jakarta*', *KAMI* (*Kesatuan Aksi Mahasiswa Indonesia*) '*Indonesian Students' Association*', must have been based on the fact that *pelita*, *jaya*, and *kami* are indeed also Indonesian words meaning '*light*', '*victorious*', and '*we*' respectively.

Still in some other cases the shape of the acronym may have been dictated not only by the existence of real Indonesian words but also by the cultural values prevalent within the society. The luxurious blue coloured train that runs at night from Jakarta to Surabaya, *Biru Malam*, could have been called \**Bilam*, \**Rulam*, or \**Ruma*, all of which follow the phonotactic rules of the language. Yet the official name is *Bima* '(literally) *Blue at Night*', because, I presume, this word happens to be the name of the most physically powerful hero in the Javanese version of the Mahabhārata.

When former President Soekarno was transferring most of his power to General Soeharto on March 11, 1966, to restore peace and order after the abortive coup d'etat, the letter of authority was referred to as *Super Semar*. The first word, *Super*, which is derived from *surat perintah* '*letter of order*', is obviously inspired by the English word *super* which also carries a powerful connotation in Indonesian. The second word, *Semar*, which is derived from *sebelas Maret* '*March 11*', is an Indonesian word, used mostly by Javanese, and refers to a godly character from the Javanese Mahabhārata who is to live on earth to make sure that things are run properly by human beings as well as gods. The acronym *Super Semar*, therefore, carries the spirit very well.

We know that what non-linguist native speakers call "nice to the ears" is in fact a very basic linguistic principle which has recently been referred to as competence. It is this competence which enables the Indonesian people to generate nice sounding/looking acronyms.

While "nice to the ears" is certainly a solid ground for acronymisation, we have seen inconsistencies where a full form is abbreviated into several different acronyms: militer into mil, m, and mi as in koramil, kodam and mahmilub respectively. While the choice of mi instead of mil in mahmilub may have been influenced by the otherwise presence of double l, which is not totally foreign but extremely rare, there is no reason why militer in koramil and kodam should take two different forms, especially when these terms were invented by the same source - the military office in Jakarta. The acronyms could have been koram (after all we have also the acronym korem!) and kodam, or koramil and kodamil - all of which follow the phonotactic rules of the language, and are as nice to the ears as the existing acronyms.

One thing which is rather disturbing is that if competence is a unique and inherent property of human beings, there should not be many cases where acronymic forms deviate, in some cases very much, from what is inherent in the language. I am not saying that language is, or should be, fully logical. I am saying that language is systematic and that trends of development revolve around the network within the system. The three cognitive reasons I have just mentioned are well within this network.

There is a substantial number of acronyms, however, which I would venture to say "lie outside the network". The previously mentioned cases such as pangkopkamtib, konjenbant, and Ifalpolekrochsohub and other forms such as ditaj (direktorat ajudan jendral) '*Directorate of the Adjutant General*', urhibjah (urusan hiburan dan kesejahteraan) '*Section on Entertainment and Welfare*', depdag (departemen perdagangan) obviously do not sound nice to the Indonesian ears. Since this is the case, is it possible that native speakers in this particular instance are what Chomsky calls "not aware of their internalized grammar" (Chomsky, 1970, p.194), or, in fact are they following what Humboldt accurately expressed 138 years ago, that is, "no matter how innate language is in its entirety, it still possesses at the same time an independent external existence, exerting a power against man himself" (p.6)?

Since native speakers by definition possess a linguistic competence, and yet in our present case they create surface forms which are not traceable to their internalised grammar, it is clear that Humboldt's "independent external existence" must be a factor, if not *the* factor,

that can explain why people generate deviant acronyms. As linguists we should be concerned with this phenomenon, because if this is to continue - which seems to be the case in Indonesia - we are in fact witnessing a language development from two opposing polarisations. I am not saying that this is unfortunate, but I am saying that it is extremely unique, to say the least, and that the long range ramifications, especially in the phonological structure, should be watched very closely.

NOTES

1. This is a slightly revised version of my paper "Acronymization as an Alternative for Linguistic Borrowing: A Case in Indonesian", read in absentia at the Third Annual Meeting of the American Council of Teachers of Uncommonly-Taught Asian Languages, Denver, Colorado, November, 1974.
2. Anton M. Moeliono uses the term *kata pancung* to refer to acronyms.
3. The old spelling of *cepat* is *tjepat*. So the *t* of *kopasgat* could have been derived from the first two letters, *tj*, which constitute a single phoneme, rather than from the last letter *t* as assumed here. It is, however, unlikely, because the abbreviation would have now been changed into *kopasgac*, if this had been the case.
4. Very few words do end in *b*, *d*, or *g*, but they are pronounced by most speakers as voiceless stops.

REFERENCES

ALISJAHBANA, S. Takdir

- 1957 *Dari Perdjuaan dan Pertumbuhan Bahasa Indonesia*. Djakarta:  
P. T. Pustaka Rakjat.

CHOMSKY, Noam

- 1970 'Remarks on Nominalization'. In: Roderick A. Jacobs and  
Peter S. Rosenbaum, *Readings in English Transformational  
Grammar*. Massachusetts: Ginn and Company.

DARDJOWIDJOJO, Soenjono

- 1967 *Indonesian Syntax*. Ph.D. thesis. Washington, D.C.

DEPARTEMEN ANGKATAN DARAT

- 1968 *Buku Petunjuk Administrasi Umum*. Djakarta.

DE VRIES, J. W.

- 1970 'Indonesian Abbreviations and Acronyms'. *Bijdragen tot de  
Taal-, Land- en Volkenkunde*, vol. 126. Holland.

FERGUSON, Charles E.

- 1968 'Language Development'. In: Joshua A. Fishman *et al.*,  
*Language Problems on Developing Nations*. New York: John  
Wiley & Sons, Inc.

FRIES, Charles C.

- 1945 *Teaching and Learning English as a Foreign Language*. Ann  
Arbor: The University of Michigan Press.

HALIM, Amran

- 1971 'Multilingualism in Relation to the Development of Bahasa  
Indonesia'. *RELIC Journal*, vol. 2, No. 2.

**HOCKETT, Charles F.**

1958 *A Course in Modern Linguistics*. New York: The Macmillan Company.

**HUMBOLDT, Wilhelm von**

1836 Translated into English by George C. Buck and Frithjof A. Raven under the title *Linguistic Variability and Intellectual Development*. 1972. Philadelphia: University of Pennsylvania Press.

**MAKKAI, Adam**

1972 Acronymy in English. A Stratificational Re-examination. Paper read at the XI International Congress of Linguists, Italy.

**MOELIONO, Anton M.**

1969 'Kata Pancung dalam Koran'. *Kompas*, January 12, 1969.

**WELLS, Rulon**

1956 'Acronymy'. In: Morris Hale, *For Roman Jakobson*. The Hague: Mouton Publishing Company.

**ZIPF, George K.**

1965 *Human Behavior and the Principle of Least Effort*. New York: Hafner Publishing Company.

## KORKU SYLLABLES AND SYLLABLE STRESS

NORMAN H. ZIDE

The syllable<sup>1</sup> is the unit of stress in Korku. It is assumed that 'stress', though difficult to define, is a feature, or some complex of features, validly and reliably identifiable by any linguist working on Korku as significantly characterising the syllables of that language, and that any complete description of Korku phonology needs some notion very similar to that we refer to as "stress". A Korku syllable is heard as being either "stressed" or "unstressed",<sup>1</sup> and a syllable can be assigned a certain degree of "strength"<sup>2</sup> - there are four degrees represented in "normally stressed" phonological words (PWs) - as a function of its stress relative to the syllables and/or junctures immediately preceding and following it within the PW. A set of rules is offered which derives the strength of a syllable from its consonant and vowel composition<sup>3</sup> and gives the expected stress markings of syllables of any

---

<sup>1</sup>An implicit - if messy - definition of the syllable along the lines of Trim and O'Connor (J. O'Connor and J. Trim, 'Vowel, Consonant, and Syllable - a Phonological Definition', *Word*, 1953, 103-22) can be obtained from the data given in this section and the next. The phonological importance of the syllable can be seen throughout the following discussion.

<sup>2</sup>"Strength" - the term "rank" may be preferable in having no confusing connotations - is not a phonetic term; it is a derived construct characterising syllable types distinguished by their consonant-vowel-semivowel shapes. The actual stress of any syllable can be determined given the strength of the syllable type it belongs to, the strengths of the other syllable types tokens of which occur within the same PW, and the relative position of this PW's syllables. A set of rules which characterises every syllable of every Korku PW as either "stressed" or "unstressed" on the basis of this position and strength data is discussed at length in this chapter.

<sup>3</sup>In a very few cases, e.g. in the form /kei/, it is not clear whether a syllable - here the ultima - is stressed or not. Where this happens, the decision is made on the basis of considerations of simplicity and utility elsewhere.

given strength in a PW as a function of these strengths and the positions of the syllables with reference to each other and the PW-bounding junctures of the form. Where a word's stress does not fit the predictions of these rules, it is supernormally - or phonemically - stressed on one of its syllables. The phonemic stress is then included in the strength assignment machinery along with the C and V component information - which now recognises syllables of five degrees of strength - and the rules are applied again, and yield results which adequately describe the syllable patterns of all forms of the language previously wrongly characterised.

Syllables are "marked" by application of a set of four rules in a fixed order; and result of the application of the set is a "final marking" consisting of pluses and minuses, these indicating stressed and unstressed syllables respectively (e.g. <+--+>, <-+--+>, <+--+>). These symbols do not represent phonemes in any usual sense of that term.

#### *The Korku Phonemes*

Consonants (C): p, b, m, t, d, n, c, j, k, g, N, q, l, r, R, s.

Semivowels (W): y, w.

Vowels (V): i, e, (i), a, o, u.

Accompaniments (A)<sup>1</sup> are /~/ (nasalisation),<sup>2</sup> /\_/ (voiceless aspiration-low tone), /\_/ (voiced aspiration-low tone).<sup>3</sup>

"A" have no effect on stress weights and will not be mentioned further in this connection.<sup>4</sup>

The symbol X<sup>5</sup> will be used for "syllable" where no further specification as to syllable type is wanted; S indicates a CVC syllable, s a VC syllable; S a closed syllable: either S or s. Z indicates a CV syllable, z a V syllable; Z an open syllable: either Z or z. Phonemic stress is indicated by /'/ over the syllable vowel.<sup>6</sup> X1, X2, ... indicate the

<sup>1</sup>/~/ accompanies both W and V; /\_/ and /\_/ V only. The consonant symbol N represents the velar nasal; the palatal nasal is here written ñ; q represents glottal stop; and R a retroflex flap. b, d and j are preglottalised before consonants in close juncture, and pause. d and t are postalveolar ('retroflex') consonants.

<sup>2</sup>'A' are written above and below V and W, e.g. /mū<sup>ˈ</sup>dæj/, /ipti<sup>ˈ</sup>yaten/.

<sup>3</sup>Phonemic stress /'/ and the junctures /+,#,=/ can be considered to constitute a fifth subclass of phonemes.

<sup>4</sup>Note, however, that /ñ/, but not /y/ occurs as C<sup>wf</sup>.

<sup>5</sup>To be consistent, X should have been used to represent "any syllable", x for any onset-less syllable (s or z), and X for any onset-possessing syllable (S and Z). I use X instead of X since it is typographically simple and is the only one of these three class indicators at all frequently used.

<sup>6</sup>But see also the discussion of ambisyllabic consonants.



first, second, etc., syllables of PW;  $x^n$  the PW-final syllable,  $x^{n-1}$ , its penultimate syllable, etc.,  $x^i x^j$  indicates any two adjacent syllables. The first consonant (onset) of a syllable is symbolised by  $x^i(C1)$ , the second (coda),  $x^i(C2)$ ;  $x^i(V)$  indicates the vowel (nucleus); each syllable has one and only one vowel. The bar / is used to indicate that the thing indicated is both the symbols the bar separates, thus the symbol C can be defined as  $x^i(C2)/x^j(C1)$ , i.e. it is ambisyllabic. A PW syllabic formula is indicated by the use of these symbols enclosed in angular brackets < > with hyphens written between syllables, e.g. <Z-Z-z-Z>, <Z-S-s>, <Z-z-s>, <Z-Z-z-s>. All the PWs indicated by this sort of syllabic formula are bounded by one or another of the three open junctures of Korku: "within-word juncture", /+/, e.g. in /ketej+ketejba/;<sup>2</sup> "word-juncture", /#/ , e.g. in /dija#anteq/,<sup>3</sup> and "phonological phrase juncture", /=/ , e.g. in /=dija#bateq#heqen=/.<sup>4</sup> Any phoneme sequence between open junctures is a PW; the stress pattern of any PW is determinable by means of the system to be described here. No further open juncture indications will be written in this section; all forms to be discussed are PWs unless they are clearly designated as something else; all forms between "/ -brackets" are PWs, if they contain no medial open junctures.

The syllabification of Korku PWs whose syllables are of the types listed above - we will discuss "ambisyllabic consonants", and other non-basic syllable types in later paragraphs - can be done in only one way, i.e. in a PW of any consonant-vowel composition, there is only one permissible analysis of the phoneme sequence into syllables.<sup>5</sup>

This syllabification can be obtained as follows: (1) count backward from the end of the PW until either a second vowel, or a non-PW-final consonant not immediately preceded by another consonant - whichever of the two comes first - is reached; indicate a syllable boundary at that point (i.e. before the second vowel, or before the second (immediately preceding) consonant); thus, /katkomku/<sup>6</sup> is katkom-ku, /tipiej/<sup>7</sup> is

<sup>1</sup>I use  $C^1$  and  $C^2$  for onset and coda, i.e.  $C^2$  does not mean the second consonant of a syllable - the syllable may have only a  $C^2$  - but that it follows the syllable vowel.

<sup>2</sup>/ketej+ketejba/ 'clatters'.

<sup>3</sup>/dija#anteq/ 'his mother'.

<sup>4</sup>/=dija#bateq#heqen=/ 'his father came'.

<sup>5</sup>A non-basic syllable type is one which occurs only when accompanied by phonemic stress; "accompaniment" here, means is placed coincident with onset, nucleus, or coda of the syllable.

<sup>6</sup>/katkomku/ 'crabs' (plural).

<sup>7</sup>/tipiej/ 'tell him (or her)'.

tipi-ej, etc. (2) Continue to move toward the front of the word marking syllable boundaries between all sequences of two vowels and of two consonants; thus, ti-pi-ej, kat-kom-ku. Any sequence of phonemes bounded by hyphens or by a hyphen and an open juncture is a syllable.

Two kinds of ambisyllabicity<sup>1</sup> are distinguished here; they are represented by separate symbols and in part, discussed separately because one must be considered phonemic (this is indicated by a consonant with a (phonemic) stress mark over it (e.g. in /kimiñ<sup>en</sup>/<sup>2</sup>) and referred to as  $\overset{\cdot}{C}$ ; the other by no (necessary) stress mark, and referred to as cc. The two are in complementary distribution within the PW:  $X^{cc}$  always occurs as  $X^1(C2)/X^2(C1)$ , while  $\overset{\cdot}{C}$  occurs as  $X^i(C2)/X^j(C1)$  where  $X^i \neq X^1$ .

$\overset{\cdot}{C}$  and  $^{cc}$  can be indicated in our syllabic formulae by special symbols,<sup>3</sup> but this unnecessarily complicates the formulae and masks certain useful distributional information.

The representation here preferred for ambisyllabic syllables is this: S with stressed coda ( $\overset{\cdot}{VC}$ ,  $CVC$ ) are represented as they would be without the /'/: s, S; X with stressed onset ( $\overset{\cdot}{CVC}$ ,  $\overset{\cdot}{CV}$ ) are written as onsetless, i.e. as VC or <s>, and V or <z>. Thus, /lemedejba/ would be <z-S-z-z>, /kimiñaten/<sup>4</sup> <z-S-z-S>, etc. If a transcription representing  $\overset{\cdot}{CV}$  and  $CVC$  as Z and S respectively were used, the resulting formula would be homonymous with those for PW, different in their consonant and vowel structure, e.g. /pulumkiba/<sup>5</sup> and /lemedeba/<sup>6</sup> would be identical in terms

<sup>1</sup>'Ambisyllabicity' is a phonetic term used to indicate that the ambisyllabic (consonant) is not heard as clearly belonging to only one of its neighbouring syllables (as either onset or coda), but as belonging less clearly to both: as coda of its 'predecessor', and as onset of its 'successor'. The occlusion of the ambisyllabic is usually but not necessarily longer than that of a non-ambisyllabic; the syllable-timed rhythm indicates ambisyllabicity when it does not indicate a C as belonging prosodically to one and only one syllable. (The syllable it would normally belong to is that whose nucleus is the vowel following the ambisyllabic.) Ambisyllabics contrast with both geminates and unambisyllabic single consonants. The "strong" ambisyllabics are prosodically ambisyllabic and have longer occlusions than "weak" ambisyllabics. The 'kinds of ambisyllabicity' mentioned above are distinguished not phonetically (although they could be distinguished phonetically, i.e. the  $^{cc}$  are all voiceless stops, the  $\overset{\cdot}{C}$  never are, etc.) but on the grounds of their phonemic status.

<sup>2</sup>/kimiñ<sup>en</sup>/ 'to the daughter-in-law'.

<sup>3</sup>For instance, by the following symbols (the ambisyllabic syllable types that occur are  $\overset{\cdot}{CV}$ ,  $\overset{\cdot}{CVC}$ ,  $VC$ ,  $CVC$ ,  $^{cc}VC$  and  $^{cc}VC$ ):  $Z^b$  for  $\overset{\cdot}{CV}$  and  $^{cc}V$ ;  $S^b$  for  $\overset{\cdot}{CVC}$  and  $^{cc}VC$ ;  $s^a$  for  $VC$ ;  $S^a$  for  $CVC$  and  $^{cc}VC$ . /lemedejba/ 'massage him', 'rubs (something into) him' would be represented as <z-S<sup>a</sup>-S<sup>b</sup>-z>. An  $X^a$  is always followed by an  $X^b$  and an  $X^b$  is necessarily preceded by an  $X^a$ . A less redundant, more useful transcription is given above.

<sup>4</sup>/kimiñaten/ 'from the daughter-in-law'.

<sup>5</sup>/pulumkiba/ 'bleaches it' (intensive).

<sup>6</sup>/lemedeba/ 'rubs it (in)'.

of these syllabic formulae. Z and z have different strengths and thus have different implications within the stress system although the distribution of  $\overset{\cdot}{C}V$  is sufficiently limited that no occurring forms test these implications where they differ.  $^{CC}$ , with which there are a number of reasons for equating  $\overset{\cdot}{C}$  in strength, does however occur in relation to Y (in //kokoyokiba//, for instance) where it is not stressed, and thus, by our rules, can be no stronger than Y; strength is a transitive relation so that if  $^{CC}V$  is no stronger than Y, it must be weaker than Z.

Such formulae<sup>1</sup> (like all those composed entirely of basic syllables in Korku) are simply analysable into their component syllable types. In other words, one can indicate the phonemic stress of non-basic syllables as a resyllabification:<sup>2</sup> <Z-S-s-Z> (e.g. /koloŋiŷba/<sup>3</sup> can only represent -CVCVC- as its second and third syllables since the ordinary -CVCVC- sequence does not permit the breakdown into -CVC-VC, and would have been syllabified as CV-CVC, i.e. as <-Z-S->.

Besides the stressed syllables we have talked of ( $\overset{\cdot}{C}$  and  $^{CC}$ ), there remain the basic syllables that occur phonemically stressed: CVC, VC, CV, and A. These are indicated as  $\overset{\cdot}{S}$ ,  $\underset{\cdot}{s}$ ,  $\underset{\cdot}{z}$ , and  $\overset{\cdot}{A}$  respectively; they are different from simple S, s, Z, and A and must be distinguished from them in these formulae since the strength differential cannot be indicated in any simpler way.

<sup>1</sup>One can also use a formulation in which each syllable is represented by a strength level indicator, e.g. /titipiba/ 'to tell' as <1-4-4-3>, and /simkiŷeq/ 'hens (dual), accusative' as <2-1-2> where the numerals stand for the degrees of strength. The distribution of these strength degree indicators within the PW does not permit automatic analysis into syllable types since there are PWs where a strength degree indicator can be interpreted in more than one way, i.e. as more than one syllable type - and correspond to occurring forms. Thus, either CV or CVC could occur as an instance of <3-2-3-3>. Those syllables that are analysable only down to the strength level (but not further into syllable type; within the level) are: at 2° level - CVC, VC, and C $\overset{\cdot}{V}$ . At 3° level - CV, #V, and A. (One can consider /#/ - here used to represent any open juncture - as a C and thus obviate the need to distinguish CV from #V, and CVC from #V.)

Note that C $\overset{\cdot}{V}$  (unlike C $\overset{\cdot}{V}$ ) is distinguishable from the other syllables on its strength level in these formulae: it is always surrounded by 2° syllables (e.g. in the example used above, /simkiŷeq/); the other 1° syllables require one 1° neighbour or one 4° neighbour, or open juncture as one neighbour. The  $^{CC}$  examples have the /#/ neighbour; the -CVCVC (-) the 1° neighbour, and the -CVCV the 4° neighbour.

<sup>2</sup>We write, that is, S and s rather than the  $\overset{\cdot}{S}$  and  $\underset{\cdot}{s}$  needed to distinguish phonemically stressed basic syllables. A more crucial example than /kalohiŷba/ is one with strong syllables besides the phonemically stressed ones: /ninijenbi/ and /katkohenkaq/ 'in opening too' and 'to the crab-emph.'. These are CVCCVCVCVC and CVCCVCVCV respectively, these being uniquely transformable to the original C, V, C, CC sequences. The former, for instance, - <S-s-s-Z> - : working from the end of the word to the beginning, Z must be CV; s must be VC. This is preceded by another s, thus another VC giving, so far, -VCVCCV, such a sequence so syllabified is possible only if  $X^{n-2}(C2)$  is  $X^1(C2)/X^3(C1)$  which, since  $X^1 \neq X^1$ , must be  $\overset{\cdot}{C}$ . The sequence is, then, -VCVCCV. The first syllable, S, a closed syllable can only be followed by an open syllable symbol <s>, if it has an ambisyllabic coda: CC. Thus, the original C, V sequence must be CVCCVCVCV.

<sup>3</sup>/kalohiŷba/ 'chooses me'.

Ẑ is found in peculiarly limited environments: as X2 before a non-final Z, i.e. in X-Ẑ-Z-; the more interesting limitation is in the selection of X3(C1), which is usually C<sup>sf</sup> (k, t are the only two C<sup>sf</sup>, that are found here). This is noteworthy because Ċ never occurs as Ċ<sup>sf</sup>, and C<sup>c</sup> which does, never occurs in this syllable position (the X2-X3 border). The suggestion, then, can be made that the "placement" of the stress /' / within the syllable is automatic, and is a function of the consonant selection of X3(C1) and/or X2(C2)/X3(C1). If this C is C<sup>sf</sup> (i.e. /k, t/), then the form is -Ẑ-Z-; if it is C<sup>wf</sup>, then it is to be read -S-z- (i.e. CVĊV, not CṼCV). In which case we could say that the placement of phonemic stress was - or could be - on the syllable; that further assignment within the syllable was automatic and a function of consonant X3(C1) selection. The syllables resulting from that selection would, in any case, be of different strengths.<sup>1</sup>

Placement of stress "within the syllable" is not, however, automatic since with the phonemes /l/ and /w/ there is contrast between, e.g. -CVlĊ- and -CṼlv-, and one must in such forms phonemically<sup>2,3</sup> contrast the

<sup>1</sup>e.g. Z is of 3° strength, s of 2°.

<sup>2</sup>One might prefer in any case to distinguish the two since their (relevant) syllable strengths differ.

<sup>3</sup>One can write Ċ-containing sequences with two phonemic stresses, e.g. for what we have written elsewhere /lemədəba/ and /lemədəjba/, and /katkōhən/ we can write /lemədəba/, /lemədəjba/, and /katkōmən/; if we keep the earlier (normal) syllabification analysis rules and it seems simpler to do so - we have to make one change in the syllable strengths to derive the proper marking for /katkōmən/ (i.e. for Ẑ syllables preceded by CVC syllables where the latter are positionally stressed (i.e. stressed by rule 3)). If the old weights were kept, the form S-Ẑ-s would be <+→, which in fact it is not. But, if Ẑ were called primary (1°) in strength rather than 2°, the proper formula, <-++>, would result from the application of the rules. Ẑ in the earlier (-Ċ-) transcription can be called 1° without changing any of the final markings; this can be done because CṼ occurs in very limited environments (always before weaker syllables than itself), and it never occurs in critical relation to 1° syllables, i.e. in a position where an indication of relative strength would be forced. Is there any advantage, then, to calling it secondary rather than primary (this in addition to the distinct advantages of the single stress working as the morphophonemics)? It seems that there is or can be under certain assumptions. The distribution of CṼ (i.e. Ẑ) is peculiar in being limited to X2, and to pre-Z syllables (and, less usefully in this connection, to pre-low tones syllables). The morphophonemics of the Ẑ-forms suggest that there are forms where the Ẑ syllables might be expected to precede S syllables; but where this would be expected, Z is found instead, e.g. /müdəkeq/ - morphemically - {müdə}{ki}{eq} '(somebody) hit (past tense) it', but /müdəkenej/ - morphemically - {müdə}{ki}{eq}{ej} '(somebody) hit him'. One might expect <sup>̄</sup>/müdəkeq/ if CṼ were of primary strength and primary syllables were characterised - as they seem to be - by stressed (allophonic) representations in all their occurrences. In the light of a simpler characterisation of the morphophonemic rules of Korcu, then, CṼ as of 2° stress - and the one - stress interpretation of Ċ forms - seems simpler.

I have made no attempt to justify the use of "non-occurrence" as evidence; I think it valid as such, but the assumptions on which one should build a case for the validity of such evidence are too complex to be gone into here. I presuppose some such acceptable set of assumptions, and I suggest that it is usable as evidence in this connection, and, roughly, how.

stress placement, e.g. /sɪm|lēba/<sup>1</sup> and /tɪp|lēle/<sup>2</sup>.

<sup>cc</sup> occurs only as X1(C2)/X2(C1). A contrast in stress pattern occurs between CV<sup>cc</sup>VCV(-) forms and CVCVCV(-) forms when X2(C1) and/or X1(C2)/X2(C1) is C only where X1(C1) = X1(C2)/X2(C1), e.g. in such cases as (ke<sup>kk</sup>erked) and (ko<sup>kk</sup>oyokiba). Thus the forms /titipiba/<sup>3</sup> and /tipikiba/<sup>4</sup> - if <sup>cc</sup> be considered non-phonemic (and it can be) - would be, in syllabic formulae, <S-z-z-z> (from CV<sup>cc</sup>VCVCV) and <Z-z-z-z> respectively; these would be identical in stress pattern <+--+> but other forms similarly contrasting in stress would not be. They can be considered non-phonemic if one makes use of the X1(C1) selection in defining the allophones of its X1(C2)/X2(C1) since wherever these are identical the latter has distinctive <sup>cc</sup> allophones, and wherever they are different it does not. For consonants other than C<sup>sf</sup><sup>5</sup> (=p, t, c, k, s) as X1(C1): whenever X2(C1) is C<sup>wf</sup> it is ambisyllabic, i.e. is X1(C2)/X2(C1); where it is not (i.e. is C<sup>sf</sup>), it is X2(C1) only, and not ambisyllabic, thus C<sup>a</sup>VC<sup>b</sup>VCV(-) has an ambisyllabic C<sup>b</sup> (a) where C<sup>b</sup> = C<sup>wf</sup>, and (b) where C<sup>b</sup> = C<sup>sf1</sup> where C<sup>a</sup> = C<sup>sf1</sup>. Thus, in a phonemic transcription before one can assign a strength to an X it must be compared with respect to its onset with X2(C1). This sort of environment has been used in allophonic statement for vowel accompaniments here, but not otherwise for consonants.<sup>6</sup>

<sup>1</sup>/sɪm|lēba/ 'sweetens it'.

<sup>2</sup>/tɪp|lēle/ 'informs us (cislocative)'.

<sup>3</sup>/titipiba/ 'tells (customary)'.

<sup>4</sup>/tipikiba/ 'tells, will tell (intensive)'.

<sup>5</sup>The distributional classification of consonants into C<sup>sf</sup>, C<sup>wf</sup>, etc. is given below:

C<sup>wf</sup> (consonants that occur in PW-final position) b,m,d,n,g,N,q,l,r,R

V<sup>wf</sup> a,e,i,o,u W<sup>wf</sup>y (but only when y<sup>1</sup>, i.e. when accompanied by /~/)

C<sup>sf</sup> (consonants that occur syllable-finally but NOT word-finally) p,t,c,k,(s)

V<sup>sf</sup> (i) W<sup>sf</sup> NONE

C<sup>wi</sup> (consonants which occur word-initially) p,b,m,t,d,n,c,j,k,g,l,r,s

V<sup>wi</sup> a,e,i,o,u W<sup>wi</sup> NONE

C<sup>si</sup> (consonants which occur syllable- but not word-initially) q,R

V<sup>wi</sup> NONE W<sup>wi</sup> w,y

<sup>6</sup>The morphological boundaries of the forms /titipiba/ and /tipikiba/ are: (ti/tipi/ba v. tipi/ki/ba) and if one wrote /titipiba/ for the former the position of stress would lend itself in the morphophonemics to convenient generalisation about //'; if stress be considered nonphonemic - which it will be - an additional statement is needed defining the ambisyllabics where C is C<sup>sf</sup>, and C<sup>a</sup> = C<sup>b</sup> as "morphophonemically stressed", and then the generalisations about the morphophoneme //''' can be made as before.

"Short syllables" are of the same degree of strength as  $^cV$  and  $^iCV$ : quaternary. They are defined as those having "short vowels" as nuclei; a short vowel -  $V^x$  in the sequence  $(-)^v^w - C^x V^x - C^y V^y (-)$  - is recognised where  $C^x$  is  $C^{sf}$ ,  $C^b$  or  $/m/$ ,  $C^y$  is  $C^r$ ,  $n$ , or (rarely)  $m$ , where  $C^x \neq C^y$ , and where  $V^x$  is the same as either  $v^w$  or  $v^y$  or of the same vowel height as either. There are three vowel heights; the vowels at these heights are: low,  $/a/$ ; mid,  $/e, o/$ ; and high,  $/i, u/$ . Thus,  $/d\underline{i}kuni/$ <sup>1</sup>,  $/ruk\underline{u}ni/$ <sup>2</sup> and  $/c\underline{i}tere/$ <sup>3</sup> have short vowels as X2 nuclei;  $/supari/$ <sup>4</sup>,  $/sikari/$ <sup>5</sup>, and  $/g\underline{o}taRi/$ <sup>6</sup> do not. The former are  $\langle 3-4-3 \rangle$ , the latter are  $\langle 3-3-3 \rangle$ . Some short syllables have been characterised already: as  $^cV$ , e.g.  $/g\underline{i}g\underline{i}Ri/$ <sup>7</sup>,  $\langle S-z-z \rangle$ , or  $\langle 1-4-3 \rangle$ ; the remainder will be represented by the symbol  $v$ , e.g. in  $/bukala/$ <sup>8</sup>: as  $\langle z-v-z \rangle$ , or  $\langle 3-4-3 \rangle$ .<sup>9</sup>

From the morphophonemics of Korku one can reconstruct what was very likely the situation in recent Korku but is not in present-day-Korku,

<sup>1</sup> $/d\underline{i}kuni/$  'bedbug'.

<sup>2</sup> $/ruk\underline{u}ni/$  'a species of fish'.

<sup>3</sup> $/c\underline{i}tere/$  'partridge'.

<sup>4</sup> $/supari/$  'areca nut'.

<sup>5</sup> $/sikari/$  'hunter'.

<sup>6</sup> $/g\underline{o}taRi/$  'antelope'.

<sup>7</sup> $/g\underline{i}g\underline{i}Ri/$  'to catch fish with hook and line'.

<sup>8</sup> $/bukala/$  'caterpillar'.

<sup>9</sup>The differential stress implications of  $\langle 3-3-3 \rangle$  and  $\langle 3-4-3 \rangle$  forms are difficult to distinguish, if, indeed, they are distinct in some common PW, i.e. in those of five syllables since a  $\langle 3-3-3-3-3 \rangle \langle +--+ \rangle$  and a  $\langle 3-4-3-3-3 \rangle \langle +---+ \rangle$  where the 4 is a short syllable are quite similar. The similarity is to be expected by the nature of our definition of stress, and that of the phonetic character of short syllables since an unstressed syllable preceded by a short syllable automatically receives a modicum of stress as the successor of a short syllable. This "modicum" seems to be indistinguishable in degree from the degree of stress a syllable would receive as a stressed X3 in word-medial - - position. A six-syllable word with X2 short (e.g. *capinikuteten*) would force a stressed syllable into greater prominence (if S3 were stressed) since it necessarily is the case that where a syllable precedes two - rather than one - unstressed syllables it gets more stress. Thus, for a 6Z word a clear differentiation - of a stressed X3 which is not predicted by our rules from a stressed X4, which is - would be obtained. Unfortunately, these words are extremely rare and not appreciating their crucial nature when collecting the field data, I have not such forms in my lists (they may turn up in texts since all the texts have not yet been examined for such information). My guess as to the result - for whatever my Korku Sprachgef\u00fchl is worth - is indicated by the transcriptions above (which are consistent with the behaviour of short syllables elsewhere, K. morphophonemics, etc.), e.g. that they are stressed on X4 as the rules would predict. (If they were not, X3 would have to be phonemically stressed or, less probably, the rules would be modified.) In fast speech, some PW with non-short vowels have the non-short replaced by short, e.g.  $/sikari/$  becomes  $/sikiri/$ .

where a further difference in syllable strength was wholly predictable from consonant selection. Here we deal with a three-way classification of consonants: the  $C^{sf}$  (as above)  $C^b$  (b,d,j,g), and  $C^r$  (r,R,l,m(?),n). The details given here are approximate; a fair amount of careful reconstruction would be needed to justify this scheme in full detail. The situation seems to have been that where a present  $CVCVC^{sf}V-$  occurred, an earlier  $CVCVC^{sf}V-$  with <Z-Z-Z-> stress was found; where present  $CVCVC^bV-$  occurred, an earlier  $CVCVC^bV-$ , with <Z-S-z-> stress; where we now have  $CVCVC^rV-$ , an earlier  $CVCC^rV-$  was found. (X2(V) was short, but nonphonemic.) The grounds for these assumptions are given in the section on "short vowels", and in that on the clitic-like suffixes: -dan, -dun, +da- (the permissive auxiliary verb), -do, -ja, -ba, -bi, and -gon .

I give below a table of syllable strengths followed by a discussion of some of its limitations. Represented in our syllabic formulae by:

1°	$C\dot{V}C, \dot{C}VC, V\dot{C}, CVC, CV^C$	$\dot{S}, s, s, S, S$
2°	$CVC, C\dot{V}, VC$	$S, \dot{Z}, x$
3°	$CV, \#V$ (post-open juncture V), a	$Z, Z, A$
4°	$WV, {}^{cc}V, \dot{C}V, C^V$	$Y, z, z, v$
5°	$V(\neq a)$	$z$

Rules (unamended version):

I. Mark  $x^n$ .

II. (a) Mark every primary syllable; (b) mark every 2° bordered by at least one syllable weaker than itself. (Consider  $x^n$  to be no weaker than  $x^{n-1}$ . Consider (preceding) open juncture to be weaker than X1.) (c) Mark every weak (3°, 4°) syllable preceding a syllable weaker than itself.

(PRELIMINARY NOTE:) Sequences of consecutive unmarked syllables - referred to hereafter as unmarked syllable stretches (USS) - are either initial or medial. A minimal initial USS consists of one syllable. A minimal homogeneous medial USS consists of two syllables (homogeneous here means "of the same degree of strength"). A minimal heterogeneous medial USS consists of three syllables.

III. Each supra-minimal USS must be reduced to minimal stretches by marking one or more of its syllables. Mark the first syllable of an initial supra-minimal USS and every odd-numbered syllable following it until no supra-minimal USS remains; mark the second syllable of a homogeneous medial supra-minimal USS and every odd-numbered syllable following it until no non-minimal USS remains; reduce all supra-minimal

heterogeneous USS to homogeneous-equivalent (HE) USS in the following manner: count as one syllable-equivalent every syllable followed by a syllable of like strength, and every syllable preceded by a counted syllable. Thus, -Y-Y-Z- counts as three (HE) syllables, -z-Y-Z- as no HE syllables. Treat the HE syllable as homogeneous syllables. Thus, if a supra-minimal number - say, three - occurs, the second is stressed.

*Amendments:*

I. Mark every  $X^n$  unless that  $X^n$  be a lengthening or a diphthongisation, in which case mark  $X^{n-1}$ .

IV. Stress the initial syllable of any PW whose final syllable is not lengthening or diphthongisation (i.e. whose final syllable is stressed) which has only one stress.

Examples:<sup>1</sup> (1.) /kokosomoroden/; by I <-----+>; by IIa <+----++>; by III <+--+-->.

Final markings can be indicated by a period after the syllabic formula, e.g. <+--+-->. (2.) /kokoyoba/; by I <---+>; by II <---+>. (3.) /mudakekukiba/; by I <-----+>; by II <+----+>; by III <+--+-->. (4.) /gada/; by I <--+>; by IV <+--+>. The syllable types and syllable formulae for these are: (1) CV<sup>CC</sup>VVCVCVCVCVC, S-z-Z-Z-S-s, 1-4-3-3-1-1; (2) CV<sup>CC</sup>VYVCV, S-z-Y-Z, 1-4-4-3; (3) CVCV<sup>C</sup>CVCVCVCV, z-Z-Z-Z-Z, 3-2-3-3-3-3; (4) CVCV, z-Z, 3-3.

There are two related problems concerning this system (strength assignments plus rules): "wrong results" yielded by the system for PW containing (-)CV-V(-) which can be and should be corrected within the system, i.e. by amending it; and cases of "diphthongisation" in (-)CV-VC(-) sequences for which no simple revisions within the system as presented above are possible.<sup>2</sup>

<sup>1</sup>/kokosomorod/ 'a species of tuber'; /kokosomoroden/ 'in the k.'; /kokoyoba/ 'shaves (someone)', 'cuts (someone's) hair'; /mudakekukiba/ 'must have beaten them'; /gada/ 'river'.

<sup>2</sup>In this exposition the system is now offered as a useful heuristic tool, i.e. as something accounting satisfactorily (and necessarily) for most K. forms; for the remainder it is suggested that although the system is, of course, subject to revision of the data required by it, in a number of cases the data themselves are reinterpretable, and that such reinterpretations - suggested by the inconsistency of the earlier interpretation with the tentative stress-weighting system - are better than the original interpretations were.

I am not suggesting that there is any particular significance in the order in which I arrived at various results, i.e. if it had been different, the later causes for revision would have been built into the system earlier, etc. I do make use of problems found, and revisions made (in the actual order I found them and made them) because such a presentation suggests the use and authority such a system was shown to have, and that that may be of some interest along with the final set of rules which could, of course, have been presented immediately, i.e. without "earlier versions, inadequacies, later version, etc."



By diphthongisation, I refer to the "fusion" of two adjacent vowels - these, by definition, belonging to separate syllables - which cannot be simply accounted for by our formulae; such forms as /mūdaejba/<sup>1</sup>, /koleiȳba/<sup>2</sup>, and /tupuejba/<sup>3</sup> would in our formulae (all) be represented as <+--+>, and sound something like <-++> where the two medial syllables act as a single syllable of the type CVC. The degree of fusion varies with the homo- or heterotonicity of the vowels (the vowels are more fused if homotonic), the speed of speech (the greater the speed, the greater the fusion), the strength of the preceding (X1 in the above examples) syllable (the stronger it is, the greater the fusion), and with the particular v<sup>i</sup> and v<sup>j</sup> involved. Neither transcription - <+--+>, <-++> - nor any other using this sort of plus and minus combination describes the forms heard, and we have arbitrarily preferred as an "ideal type" representation - and one useful elsewhere in the grammar - the former, which leaves our rules as they are, and adds more "allophonic"<sup>4</sup> data of the sort we have just given as additional information to be applied in going from the formulae to the sounds of the language.<sup>5</sup>

The second group of diphthongal forms to be discussed is clearly stressed, and its stress can be adequately represented by the system used here, but the representations generated by our (unamended) rules are wrong. These forms are PW which have final v<sup>i</sup>v<sup>j</sup> (but not all PW with final v<sup>i</sup>v<sup>j</sup> are in this group) and medial and final -a-a(-).<sup>6</sup> Where, in a -CV<sup>i</sup>v<sup>j</sup>(-) sequence v<sup>i</sup> = v<sup>j</sup> v<sup>i</sup> is stressed and v<sup>j</sup> is not regardless of PW position. Thus, /sasaapo/<sup>7</sup>, /golaa/<sup>8</sup>, /doo/<sup>9</sup>,

<sup>1</sup>/mūdaejba/ 'hits him'.

<sup>2</sup>/koleiȳba/ 'peers at me.'

<sup>3</sup>/tupuejba/ 'drenches him'.

<sup>4</sup>The use of the term "allophonic" might be justified by stating that these are characteristic of the phoneme /'/; they are also characteristic of stressed syllables other than those containing /'/ . The "allophones" of "unstressed" CV preceding VC characterise no other unstressed syllables but these, and are more like strength syllables allophones in their phonetic properties.

<sup>5</sup>Another example of fusion not well representable by some combination of pluses and minuses is found in -C<sup>s</sup>f<sub>v</sub>-C(r,R)V#, and -C<sup>s</sup>f<sub>v</sub>-C(r,R)VC- forms where the two syllables are often given a single chest pulse, the first of the two being quite short but phonetically more stressed than unstressed syllables characteristically are, e.g. /citere/ 'partridge', /jujukirij/ 'to sweep'.

<sup>6</sup>Medial -v<sup>i</sup>v<sup>j</sup>- has been adequately accounted for by Rule II.

<sup>7</sup>/sasaapo/ 'to purify'.

<sup>8</sup>/golaa/ 'gather! (translocative)'

<sup>9</sup>/doo/ 'to put'.

/jee/<sup>1</sup>, /tii/<sup>2</sup> should be <+--+>, <-+-->, <+-->, and <+--> respectively. Where in -CV<sup>1</sup>V<sup>j</sup>(-) sequences V<sup>1</sup> = V<sup>j</sup>, X<sup>j</sup> can be defined as a "lengthening" of X<sup>1</sup>. Note that length is not phonemic, and that the tone of the lengthening (X<sup>j</sup>) here is independent of that of X<sup>1</sup>. If one has a phoneme "length" in the phoneme inventory, the phoneme /i/ (which in one earlier analysis, was shown to contrast with /e/ only before a medial /e/, e.g. in such pairs as /kolieba/<sup>3</sup> and /haveeli/<sup>4</sup>) can be written as /e/ in this position and the two forms given could be phonemicised /koleeba/ and /have:li/. There is a better interpretation of /kolieba/ available (see the section on WV), and length as a phoneme is unnecessary.

Where in final -CV<sup>1</sup>-V<sup>j</sup> sequences, V<sup>1</sup> ≠ V<sup>j</sup> the following stress patterns occur (I put in parentheses those sequences satisfactorily accounted for previously: (a) as lengthenings, and (b) those with final /a/ not as lengthening which are stressed as per the rules; I write the stressed vowel with a capital letter):

<i>Dissyllables</i>							<i>Polysyllables</i>						
	/a/	/e/	/i/	/i/	/o/	/u/		/a/	/e/	/i/	/i/	/o/	/u/
a/	(Aa)	Ae	*	aI	Ao	aU	a/	(Aa)	Ae	*	{Ai aI	Ao	*
e/	(eA)	(Ee)	*	Ei	Eo	*	e/	(eA)	eE	*	EI	*	*
i/	*	*	*	*	*	*	i/	*	*	*	*	*	*
i/	(iA)	(iE)	*	(Ii)	*	iU	i/	(iA)	iE	*	*	*	*
o/	*	{oE Oe	*	Oi	(Oo)	Ou	o/	(oA)	Oe	*	Oi	*	Ou
u/	(uA)	uE	*	uI	*	(Uu)	u/	(uA)	uE	*	Ui	*	*

One would like to define a relationship, "diphthongisation", between pairs of syllables analogous to that of lengthening, and for certain of the V<sup>1</sup> V<sup>j</sup> pairs one finds similarly simple describable relationships between the V<sup>1</sup>s and V<sup>j</sup>s; thus, V<sup>1</sup> is stressed (and V<sup>j</sup> is not) where V<sup>1</sup> is /a/, and V<sup>j</sup> is a vowel of mid height: Ao, Ae. A mid vowel is stressed when followed by its high vowel (i.e. a front mid followed by a front high; a back mid followed by a back high): Ei, Ou. Vowels retain their X<sup>n</sup> stress, i.e. are not diphthongisations - where V<sup>1</sup> is /a/ and

<sup>1</sup>/jee/ 'who'.

<sup>2</sup>/tii/ 'hand'.

<sup>3</sup>/koleeba/ 'peers at it'.

<sup>4</sup>/haveeli/ 'mansion'.

$v^j$  is high: aU, aI. (Note that aU occurs only in (two)<sup>1</sup> dissyllables, and that in polysyllables both Ai and aI occur<sup>2</sup> (but we have only one example of each). Oi occurs and all Oi forms (in dissyllables and polysyllables) have doublet forms with Oe (and vice versa). Thus, the two do not contrast and can be represented by a single phonemic form as a simple diphthongisation. iU and uI occur but the former only in monosyllabic forms, and the latter sequence is in the one<sup>3</sup> polysyllabic form it occurs in Ui.

It does not seem desirable to distinguish  $v^i v^j$  sequences in dissyllables from those in polysyllables; for the former we characterise the relationship of diphthongisation between a  $v^i$  and a  $v^j$  as one where  $v^j$ , though in  $x^n$ , is not stressed, and  $v^{n-1}$  is stressed. The  $v^j$ , the diphthongisation, is either a front vowel or a back vowel, and there is no example of contrast of a  $v^i$  being followed by two "front diphthongisations", i.e. if Ao occurs, Au will not. (Lengthenings do contrast with diphthongisations, e.g. /Cee/ v. /Cei/. A single semivowel symbol could be used for front diphthongisation, and one for back diphthongisation, e.g. the possible  $v^i v^j$  combinations could be written /ay/, /ai/, /aw/, /au/, /ey/, /ew/, /ie/, /iw/, /oy/, /ow/, /ui/, /ue/. For none of these except for oE is the diphthongisation of the same vowel quality as a non-diphthongised sequence, and here although Oi occurs, we have doublet forms with Oe for all Oi forms so that we can represent both as /oy/, and oE as /Oe/. In the polysyllabic forms we find a diphthongal Ai that must be distinguished from Ae by phonemic stress /ap<sub>i</sub>ai/ (the form occurs in some dialects as /apei/, which would not require the /'/), and the sequence Ui which, if we do not want to make a distinction between polysyllabics and dissyllabics (since the dissyllabic final was uI), we would treat as we did /ap<sub>i</sub>ai/, i.e. as /tur<sub>i</sub>ui/, or by the use of the symbol /y/ (in both): /ap<sub>y</sub>y/, /tur<sub>y</sub>y/.<sup>4</sup>

We shall here not use the "diphthong transcription" of any of these  $v^i v^j$  sequences, but define those sequences listed on the previous page where a particular  $v^j$  following a particular  $v^i$  is transcribable - and would be transcribed in our list as a diphthong - as diphthongisations.

<sup>1</sup>/jau/ 'barley'; /tau/ 'behind'.

<sup>2</sup>/ap<sub>i</sub>ai/ 'three'; /sip<sub>i</sub>ai/ 'soldier'.

<sup>3</sup>/turui/ 'six'.

<sup>4</sup>Since /y/ occurs nowhere else in Korku as a coda - /ỹ/, however, does - and there is no advantage in interpreting (turUi) as /turuy/ rather than /tur<sub>i</sub>ui/, we have preferred the latter interpretation. We can call the /<sub>i</sub>ui/ and /<sub>i</sub>ai/ phonemic diphthongisations, as opposed to automatic. They are morphophonemically similar to the other diphthongisations in the morphology of numeral stems and that of the verb suffix (-ya-) after verb stems ending in final /ar/, /or/, and /ur/.

The two forms not yet mentioned are /ea/ (phonetically (īa), and /ee/ (phonetically (īe)) which, like /ue/ and /ie/ are stressed on the ultima and can be listed as such and transcribed phonemically as such. They will be alluded to in the section on WV. Thus, in short, we need to amend to Rule I to state that  $X^n$  is marked except where it constitutes a lengthening or a diphthongisation, in which case it is not marked and  $X^{n-1}$  is marked instead, and we add a Rule IV to account for initial stress in dissyllables where the ultima is not stressed.

A set of diphthongisations is characterised by an automatic stress on the  $V^j$  of a  $(- )CV^i-V^j-CV(- )$  sequence;  $V^j$  is /a/ in all these cases. Presumably, at an earlier stage of the language, such  $CV-a-CV(- )$ , when initial, were automatically <-++>; this is no longer the case; we now have such contrasting forms as /kuabi/<sup>1</sup>, /kuagej/<sup>2</sup>, and /siɔli/<sup>3</sup>, phonemicised as indicated. The phonemicisation is consistent with that for  $\overset{\cdot}{C}$ - since morphophonemic convenience requires the two to be treated alike. The shift of (systematically) "normal" stress from /kuagej/ - presumably, earlier /kuagej/ but pronounced the same way - to the present phonemically stressed form resulted from the same factors as the shift to  $\overset{\cdot}{C}$ - from -c-: from the loss of a phonologically distinct clisis of the "clitic-like suffixes" mentioned in connection with  $\overset{\cdot}{C}$ -. In all  $-CV^iV^j-$  sequences where  $Z^i \neq Xl$ , and  $V^j$  is /a/, the stress is automatically on the /a/ if it is positionally stressable, i.e. there are no examples where the  $CV(X^i)$  syllable would be positionally stressed with  $X^j$ , A, unstressed. One might expect a rule to be needed to shift the stress to the /a/ where  $-X^i-a-$  occurs positionally stressed with the  $X^i$  syllable as "shortened" or "diphthongised", but there are no examples where the positional stress, if any, has not been already preempted by the A. Thus nonphonemic "weak gemination", i.e.  $^{cc}$ , lengthening, and diphthongisation are terms here used as characterising the relationships of certain pairs adjacent syllables; if  $X^j$  is a lengthening or a diphthongisation of  $X^i$ , or if it is S by virtue of a coda that is a weak geminate then the stress of the pair of syllables follows from this relationship of the two and need not be otherwise indicated in the phonemic transcription though no interpretation of these features as phonemic can be made.

WV syllables: /ɪ/. /ɪ/ contrasts with /e/ only in  $-CV^i-e-$  position,

<sup>1</sup>/kuabi/ 'the well too'.

<sup>2</sup>/kuagej/ 'spank him'.

<sup>3</sup>/siɔli/ 'finish it (cislocative)'.

e.g. in the two forms /kolieba/<sup>1</sup> and /pepeeda/<sup>2</sup>. Although the phoneme sequence Vye does not occur in this dialect of Korku, Vya and Vyu are common with /e/ as V, and the pre-y allophones of /e/ in these cases are much like what they are of /i/ before /e/. A phonemic interpretation of /kolieba/ as /koleyeba/ yields an expected stress pattern that corresponds to the forms as heard which is what is wanted here; the morphophonemics does not much suffer from the change. Writing /koleyeba/ here suggests writing /eye/ elsewhere where similar allophones of the /e/s are found, and doing so in the locative forms of nouns with final ei stem forms proves to make the statement of noun morphology slightly more regular, e.g. /konjeyen/ for what would have been written earlier as /konjeen/. We will, therefore, retain the transcription /koleyeba/ and drop the phoneme /i/ from our inventory. WV syllables and the choice of possible interpretations regarding their strengths will be discussed in some detail here, and an attempt will be made to make explicit in this case precisely what "considerations of morphophonemic simplicity" - mentioned in other cases as decisive but not explicitly justified - can be. The immediately relevant morphophonemic considerations are discussed; others, i.e. the characterisation of the morphophonemic rules involving /i/, and that of the morphophonemic rules in general will not be gone into here since I know of no advantage accruing to either of the interpretations given below to be gained by recourse to a detailed examination of these rules.

The strength of WV syllables can be interpreted in two ways: one interpretation (Int. I) calls yV syllables 4°, and wV 3°; the other (Int. II) calls them both 4°.<sup>3</sup> Others were found to have no comparable

<sup>1</sup>/kolieba/ 'peers at it/them (inan.)'.

<sup>2</sup>/pepeeda/ 'to produce'; this is a reduplicated form from the stem peeda - borrowed from the Hindustani /paidaa/. The i-e contrast in this position has probably come about very recently with the introduction of morpheme-internal /ee/ sequences in loanwords from the Hindi, /i/ occurring (and /e/ non-occurring) before morpheme boundary.

<sup>3</sup>There is also data questionably interpretable as evidence supporting a weaker interpretation of wV to be obtained from the phonology of loanwords from Hindi (where H. /wa/ is reflected as /o/ in Korku), and in the distribution of wV (and it resembles that of yV in this respect). Its distribution is such that apart from the examples of {ya} and {wa} forms referred to above, and in spite of the fact that yV and wV syllables are not infrequent they are not found in positionally stressable positions. The assumptions on which the validity of such evidence would need to be argued are complex and will not be presented or justified here. I think such data can usefully be put forward here but only as evidence of an earlier, weaker /wV/. The whole matter of what sort of a construct "strength" is pertinent to such arguments and the question of what is or might be meant by "an earlier, weaker wV".

advantages,<sup>1</sup> but one of these - interpreting both as tertiary - is represented in a set of sample forms below.

Under these two interpretations, the following statements would occur in the morphophonemics.

First, note that *yV* and *wV* do not otherwise occur in decisive (decisive with reference to stress assignment) positions so that our argument is concerned entirely with cases of the forms - verb forms - containing the verb mode suffixes {*ya*} and {*wa*}.<sup>2</sup> The verb mode suffix representation /-y-/ is interpreted morphemically as {*ya*}{uq}; if it were not it would have to be added to {*ya*} and {*wa*}, but would not much affect our argument.

Int. I = *wV* 3°; *yV* 4°

1) No mode suffix is phonemically stressed in present tense verb forms. One can replace "in present tense verb forms" with "in syllables not immediately preceding tense suffix-containing syllables" in this statement, and in the corresponding one in the *WV-as-4°* interpretation given below. The statement is not, in either case, purely phonological.

2) In past tense forms all verb stem final syllables are stressed, phonemically if necessary. Phonemic stress is necessary, i.e. the syllables in question are not stressed automatically, where such syllables are open and would not be positionally stressed, or stressed by Rule 2. All such cases are of dissyllabic verb stems with open second syllables (e.g. /gata-/ 'to find, obtain'; /tipi-/ 'to tell') where X3 is of CV or *wV* shape; the occurring X3 are /le/, /te/, /ke/, and /ve/.

These are, morphophonemically, {li}{eq}, {ta}{eq}, {ki}{eq} and {wa}{eq} respectively. {*ya*}{eq}, which is the only other mode suffix + {eq} combination found, occurs in this X3 position as /e/, i.e. as *V*, and is therefore automatically stressed.

<sup>1</sup>To apply, roughly, the distinction of MacCorquodale and Meehl (K. MacCorquodale and P. Meehl, 'Hypothetical Constructs and Intervening Variables', *Psychological Review* (55), 1948), the term "strength" is used generally and explicitly as an intervening variable, but implicitly and occasionally as a hypothetical construct; the implications of the latter usage extend but do not contradict those of the earlier, which are basic to our treatment of the data.

<sup>2</sup>{*ya*} (and {*ya*}{uq}), if we wish to consider it separately, is very common; {*wa*} is quite rare: with most object suffixes it is obsolescent, with a few it is, I think, obsolete. It seems used comfortably only with third person singular objects. Drake (J. Drake, *A Grammar of the Kurku Language* (Calcutta, 1903)) omits the suffix entirely from verb forms of this sort, and there may be idiolects or dialects in the Dharni area - Drake probably is describing a Chikalda dialect (he doesn't say) - that wholly lack the morpheme in this position. For such dialects, if the rest of the data relevant is comparable with that for this dialect, there is no problem. The statement that *WV* syllables are 4° would seem to be simplest for such systems.

Int. II - wV, yV (i.e. WV) 4°

1) No mode suffix is stressed in present tense forms except {wa} in -CV-{wa}-CV- position where the V immediately preceding it is /a/, /e/, /i/. Where that vowel is /o/ or /u/, the morpheme {wa} has the allomorph {a}, which is automatically (positionally) stressed.

2) In past tense forms all final syllables of verb stems are stressed, phonemically, if necessary. Phonemic stress is necessary where syllables are open and would not be positionally stressed, or stressed by Rule 2. If wV is 4°, it, like V will not be stressed. If wV is 4°, it, like V, will not be stressed as an X3; in such a case X2, which is stronger than X3, will be automatically stressed by Rule 2.

Sample forms:

Int. I	golayaba	tipikenej	tipiwenej	gatawakuba	gataenej
Int. II	golayaba	tipikenej	tipiwenej	gatawakuba	gataenej
MP	{gola}{ya}{ba}		{tipi}{wa}{eq}{ej}		{gata}{ya}{eq}{ej}
		{tipi}{ki}{eq}{ej}		{gata}{wa}{ku}{ba}	
both as 3°	golayaba	tipikenej	tipiwenej	gatawakuba	gataenej

I think Int. I is preferable to Int. II because it does not give us a stressed wV, -wa-, where, on grounds of simplicity of mode suffix morpho-phonemic description, we do not want it and this at the cost of stressing X2(CV) preceding -we- where all the analogous pre-Ce- or Ce- (i.e. syllables preceding the mode suffix plus past tense-containing syllables) are also stressed, so that one statement can be made about this X2(CV) being stressed before -C (and w) V-. (The analogous suffixes are -ke-, -te-, -le-; the only other mode suffix plus past tense suffix found after verb stems is {ya}{eq}, which is phonemically represented by the V syllable /-e-/ which by the rules will not be stressed; its predecessor X2(CV) will be.) The similarity in syllable shape of the {ya}{eq} forms and the {wa}{eq} forms has been lost by the reduction of yV to V thus increasing overall similarity of wV to CV syllables. In Int. I, statement 1) is simpler, and 2) is similar, "where necessary" referring to a slightly different class of forms in the two cases. I think that so far as any other considerations hold, there is nothing to choose between the necessarily phonemically stressed class in Int. I and that in Int. II. Since Statement 1) is simpler in Int. I, I have considered that

advantage decisive.<sup>1,2</sup>

## MORPHOPHONEMICS

The most basic morphophonemic problems in Korku concern stress, tone, and vowel quality. These will be taken up here. We interpret tone - our earlier "tone-cum-aspiration" - and vowel quality as a single variable, "vowel quality redefined", and take up the limitations on vowel selection within morphemes in terms of sets of "harmonic" vowels and the rules of vowel harmony obtaining among these.

## STRESS

Given the morpheme boundaries of a word with phonemic stress, some information on morpheme selection, and a set of morphophonemic rules, phonemic stress is wholly predictable.

<sup>1</sup>Note that we are still considering wV as Wv, not as CV although we attribute a characteristic tertiary strength to it on the basis of the considerations argued above. The WV-CV distinction in describing the distribution of phonemes seems a useful one; therefore, we keep it.

<sup>2</sup>There is no *a priori* reason to favour treating yV and wV as functionally identical; the distributions of the two differ, after C for instance. Other pertinent data on the instability of yV forms in positionally stressable positions are: the preference for more stable - 5° or 3° - alternants in these positions, and the existence of such alternants. (Neither is the case for wV syllables.) Thus, /a/ for /ya/ - these are allomorphs of {ya}, /y/ for /yy/ - these are morphemically not identical - /y/ is the morpheme {y}, /yy/ is {ya}{y}; here the two forms in the affirmative indicative verbs are wholly interchangeable. In the negative (indicative and imperative) forms only {y} containing suffixes are usually passive in meaning, and the imperative forms are uncommon, particularly the positive imperative. Another development found in this (Dharni) dialect, but unknown in Lahi Korku is the replacement of /yy/ after vowels by /i/, in some idiolects the distribution is completely complementary and /i/ → /iq/ before open juncture -- contains the morphemes {ya} and {y}. Where both /yy/ and /i/ occur they can be said both to be free variants of the (same) suffix {ya}{y}. In the idiolects I am most familiar with /i/ replaced /yy/ after V in almost all cases but occasional /yy/ in free variation with /i/.

V and CV are in Korku morphophonemics more stable than V. The /i/ and /y/ alternants - and the /y/ cases are not allomorphic alternants - of /yy/ and the /a/ alternant of /ya/ when it is a lengthening are 5° in strength and therefore weaker than yV is; the /a/ alternant of /ya/ when it is not a lengthening, i.e. when it follows /e/, /i/, /o/, or /u/ is 3° and therefore stronger than the yV syllables are.

{ya} alternants rarely occur in positionally stressable position; they do so only before the uncommon probabilistic suffix {ki}. (This {ki} is not to be confused with a homonymous morpheme, the mode suffix which figures in some of the examples in this section.) Verb forms with {ya} take only two object suffixes (and are unlike all other verb stem plus mode suffix forms in this respect): -kiỹ, and -kej. These are both strong syllables and are stressed by Rule 2 when following weaker (e.g. CV or WV) syllables; thus, /ya/ is never positionally stressable when followed by an object suffix. {wa} was positionally stressable - and stressed - in just these environments, e.g. the form /gatawakuba/ has no parallel \*/gatayakuba/ although for all the other mode suffixes parallel forms (gatakekuba, etc.) exist. (These all mean '(someone) finds them' (plus some modal suffix-related meaning)').



1) Any stressed consonant /C/ is immediately followed by a morpheme boundary. Ex.: /karube/, /tarag*i*ŷba/; enclosing morphemes in { } parentheses, these forms are {karub}{e} and {tarag}{iŷ}{ba}. Given the morpheme boundaries, any morpheme-final intervocalic consonant not preceding X2(V) will be stressed. These will now be indicated only by the morphemic parentheses.

2) Any final syllable of a verb stem not automatically stressed when preceding a syllable containing the past tense morpheme {eq} will be phonemically stressed. Ex.: /gatakenej/, /tipilele/. Given the recognition of the past tense suffix {eq}, what precedes it is definable as a verb stem and as such will have its final syllable stressed. Given the rules for determination of automatic stress and the morphophonemic rules<sup>1</sup> describing the allomorphic representation of the morphemes in past tense forms, the forms that are derived from the morpheme sequences {gata}{ki}{eq}{ej} and {tipi}{li}{eq}{le} have their verb stem final syllables as X2 where X3 is CV, and are therefore not stressed automatically. They must occur by this rule, with phonemic stress: as /gatakenej/, and /tipilele/.

3) CVC syllables of specified morphemic content when preceding the suffixes {keq}, {dan} and {ba} are phonemically stressed if not otherwise stressed. The suffixes are verb tense suffixes, but {keq}<sup>2</sup> has a homonymous nominal suffix, the accusative marker, which has the same stress-related properties. The syllables preceding {keq}, etc. are themselves morphologically delimited: only CVC syllables containing a person suffix<sup>3</sup> carry the stress in this position. The only person suffix preceding the nominal accusative {keq} is the dual {kiŷ} which is identical with the third person dual person suffix.

Only two cases were noted where /'/ was either morpheme-medial or its morpheme content was questionable. These are /pipijito/<sup>4</sup> and /suluruj/. /pipijito/ is the reduplicated form of /pijito/ and gives

<sup>1</sup>The morpheme combination rule operating in these forms is the following (all mode suffixes have the shape CV): A mode suffix, {CVi} followed by {eq} yields a form /Ceq/ before /#/ , /Ce/ before {C-} and /Cen/ before {V-}. The tone of the resulting morpheme is /\_ / if its C1 is not P, /\_ / if it is. Thus, {ki}{eq}>/keq/, not \*/keq/.

<sup>2</sup>They could be considered not as homonyms, but as a single morpheme.

<sup>3</sup>The "person suffixes" or "animate object suffixes" that can occur in this position are {iŷ} 1st sg.; {ej} 3rd sg.; {liŷ} 1st du. exclusive; {lom} 1st du. inclusive; {piŷ} and dual; {kiŷ} 3rd dual; and {buŷ} 1st plural inclusive. The two with initial V, {iŷ} and {ej} combine with preceding mode suffixes to form CVC person suffix-containing syllables which take phonemic stress in the environments.

<sup>4</sup>/pipijito/, /pipijito/ 'to annoy'; /suluruj/ 'a species of lizard'; /dikili-/ 'to push'; /sapana-/ 'to dream'; /solor/ /silir/ 'to slip, slide'.

the last three syllables the stress pattern of the stem itself where the /j/ is automatically ambisyllabic in X1(C1)/X2(C1) position. Verb stems of three syllables are rare in Korku and /pijito/ and the other two examples noted, /dikili-/ and /sapana-/, are all loans from the Hindi. The other two have P as X2(C1) and X2 is a short syllable so that a stress shift analogous to that in /pipijito/ would be less likely. The other example, /suluruj/, seems to be {sulur} plus {uj}. Both {sulur} and {silir} meaning 'to slip, slide' are found in Korku, and a {sulur} with similar meaning is assumed. The suffix {uj} is unknown, but {rvj} is a very common noun ending. In almost all cases there is no morphemic identification of what precedes the {rvj} but a case for its morphemic status can be made. We assume a {sulur}{ruj} with -rr<sup>1</sup> becoming -ṛ- as the likeliest morphemic analysis of the form.

/kuali/ 'rabbit' is a single morpheme and - like the few other morphemes of /CV<sup>(i,u)</sup>aCV/ shape in the language - is phonemically stressed on the second consonant; formerly, in such forms /a/ was treated as if it were V1, the C following it then being (automatically) stressed as X1(C2)/X2(C1).

The verb stem {sia} 'to finish' takes phonemic stress on the /a/ when it receives no automatic stress; it receives automatic stress only when preceding vowels.

---

<sup>1</sup>-rr- is not found in Dharni Korku; this cluster does exist in Pachmarhi Korku.

## AFTERWORD

The preceding paper consists of unaltered versions of sections of a dissertation on Korku submitted in 1960 (*Korku Phonology and Morphophonemics*, University of Pennsylvania, 1960); the dissertation was written up in final form in 1959 and 1960 after fieldwork and preliminary analysis in India in 1956-8. Typographical errors have been corrected.<sup>1</sup> The morphophonemic transcription (examples of which turn up in this paper) in the dissertation which reduces the two phonemes "voiced low tone/aspiration" (written v) and "voiceless low tone/aspiration" (written v̥) to one (written with a single underlining) has been rejected since the discussion of low tone/aspiration which introduce it are not included here in favour of the phonemic - two phoneme - transcription used elsewhere in this paper. Attempts to rewrite the paper were given up. The assumptions and formalisms and informalisms of its time and place have changed, and were I to write a paper using these data today, it would be different, e.g. there would be more phonetic detail about stress, less separation of "phonology" from morphophonemics, more systematic treatment of loan phonology, and a complete list of the forms providing the data for the section on diphthongs, etc. However, the earlier presentation and "its" data still seem worth publishing. The problems of describing the stress systems of other Munda languages (and other Indian languages) remain and the data and analyses for Korku offered seem useful, e.g. for Gutob (a South Munda language, also largely stress-timed, but lacking all the suffixal and enclitic morphology of Korku and the other North Munda languages), and even for Gta? (another South Munda language, but one which in its syllable structure looks very different). Some areal similarities (with

---

<sup>1</sup>At least one error of fact should be corrected: the statement in footnote 2, on page 178, on {uq} and {ya}{uq} as being semantically indistinguishable is wrong.

Hindi<sup>1</sup> and Marathi) have been noted but will not be discussed here.

The writer's interests in the comparative linguistics of Munda also have benefited by an exposure to the problems of Korku stress, and such matters - to be touched on briefly below - as reduplication, automatic vowels, the loss of pre-tonic automatic (and some non-automatic) vowels in Gta?, the questions of fusional morphology and the possible archaic character - for Munda and for Austroasiatic - of the suffixes and enclitics in Korku (and North Munda) - all of these in Munda - can be discussed in the light of what we know about Korku stress.

### *Reduplication*

The peculiar stress of reduplicative initial syllables<sup>2</sup> in Korku is seen in the example given /titipi/, i.e. (CV<sup>CC</sup>VCV), from *tipi-* 'to show, inform' this as compared with /tipiki/ (a regular CVCVCV). Reduplication<sup>3</sup> elsewhere in Munda has problems, i.e. is not a simple CV prefix (or infix) with the ordinary intra-morpheme stress pattern. In Gutob the automatic vowel selection is different for reduplicatives, e.g. *be-be?*, not the expected *bi-be?* (which Remo has), from *bed-* 'to give'. Remo has unexplained CVC, as well as "regular" CV reduplicatives. Gta? rapid speech reduces the CV reduplicative prefix (which can be elicited in slow, careful speech) to a preglottalised consonant prefix, e.g. 'bbi? (*bi?*- 'to give'), 'wwe (*we* - 'to go'), etc. The Gutob automatic V selection in reduplication, for instance, might be explained as purely morphological conditioning, but I think there is more than that reconstructible responsible for the peculiar automatic v selection in modern Gutob reduplicatives. One might very tentatively reconstruct a reduplicative (prefix) of the form *CV̄* or *CV* for Proto-Munda.

### *Automatic V, and Gta? pretonic syllable reduction*

That the initial vowels of disyllabic morphemes in the Munda languages are frequently automatic, i.e. depend on the following tonic, morpheme-

<sup>1</sup>See Aryendra Sharma, *Hindi Word-Accent*, *Indian Linguistics*, 30, 1969, pp. 115-18, and Ashok R. Kelkar, *The Phonology and Morphology of Marathi*, Cornell University unpublished Ph.D. dissertation, 1958.

<sup>2</sup>In the preceding paper reduplication was analysed (in a footnote) as, in effect, a prefix. It could also be treated as an infix (infixes are slightly more productive in Korku), i.e. the reduplicative -ti- (C<sub>1</sub>V<sub>1</sub>) infixed into *tipi*, after C<sub>1</sub>, or - more plausibly - after V<sub>1</sub>.

<sup>3</sup>Pinnow identifies two types of reduplication as probably going back to Proto-Munda. I would agree. The type discussed above - the iterative or customary reduplication - which is general in Munda reduplicated only the first syllable, or perhaps the first CV (see above). It may - as in modern South Munda - produce overt reduplicatives only with monosyllabic stems. The other reduplication is the expressive - onomatopoeic - reduplication, and repeats the whole stem with /+/ juncture between the repetitions, e.g. (in the paper above) *ketej+ketejba*. These paired forms (e.g. echo words, other conjoined words of the same class) with /+/ juncture are areally common.

final, syllables for their vowel qualities is well known.<sup>1</sup> This non-distinctive pretonic vowel is lost in Gta?, resulting in an unMunda but "Southeast Asian" syllabic patterns (e.g. such words as gta?, cognate with Gutob gutob 'ethnic name', plæg, Gutob pi:rig 'bird' etc.). If we did not know with some certainty what Gutob-Remo and Gutob-Remo-Gta? were like, we might suspect - looking at just Gta? and some Mon-Khmer languages - that Gta? was extremely archaic in some of its phonology. Knowing that it is not leads us to the more interesting problems of how Gta?<sup>2</sup> (rapidly) became phonologically "Southeast Asianised", i.e. acquired some or all of the following properties which it now possesses:<sup>3</sup>

- 1) Rare pretonic vowels (and few of these interpretable as "automatic") with resulting initial consonant clusters.
- 2) Reduction of CV reduplicative prefix to 'C-', thus GR \*no-non, Gta? 'n-nwa (and with syllabic first person pronominal prefix {N} (homorganic syllabic nasal), n'nwa 'I chase' (\*non- 'to chase').
- 3) Loss of most final obstruents, i.e. GRG l > Ø; m, n, ñ, ŋ > Ø or ŋ (no details given here on selection), b, d, j, g > ? or g.
- 4) Syllabic stress with higher pitch in a few non-stem morphemes, developed from v<sup>i</sup>-v<sup>i</sup> > v<sup>i</sup>. The most common exemplification of this is in the negative prefix á-. á-coŋ-ke (from a-a-coŋ-ke) 'he did not eat', contrasts with a-coŋ-ke 'he ate'.
- 5) The development of diphthongs from simple (tonic) vowels in greater (i.e. tonic ultima) stress environments. The original - monophthongal - vowels are preserved in (non-reduced) less stressed (non-tonic) syllable position.
- 6) Perhaps the enclitic nominal combining forms (CFs) and their grammar.

<sup>1</sup>The vowel is usually a copy of the tonic vowel, or one of a smaller set of automatic vowels (e.g. i and u for Gutob, i, u and a for Remo) naturally selected. In Korku with a few exceptions (e.g. where C<sub>1</sub> is k, kol<sub>a</sub> 'yesterday') v<sup>aut</sup> copies the following-tonic-vowel. Korku loan phonology provides some interesting examples of automatic vowel selection. In Dharni Korku where a word from Hindi (the local variety) of (C)əCəC shape is borrowed, the tonic, ultimate V is o, but the pretonic is e, e.g. Hindi gerəm 'warm, hot', Dh.K. gerom [gerəm], Lahi Korku has garam. In such Kharia forms as selhob 'antelope' (where one reconstructs a laryngealised tonic vowel) presumably the V<sub>1</sub> e is also an automatic vowel reflecting an earlier similar (or identical) vowel to the tonic. One reconstructs, perhaps, SM \*sv<sub>1</sub>hoXb or \*sv<sub>1</sub>hEXb.

<sup>2</sup>One characteristic of Gta? - probably archaic, but not found in NM - is the unstressed initial syllabic nasal vowels, e.g. nta?, 'egg' (Gutob utob/itob). These are either lost or stressed with some vocalic adjustments (in certain environments) in cognates in NM. Similar things can be observed in IA, e.g. in Eastern Indo-Aryan: thus Hindi-Urdu əmir (with stress on the long i syllable), Bengali amir. The only three languages preserving what looks to be an old set of pronominal prefixes in the verbal system are the SM languages Gta?, Gorum and Juang.

<sup>3</sup>The rules accounting for Gta? vowel loss have been summarised by N. Zide (see N. Zide, 'A Note on the Historical Phonology of Gta?, Combining Form Derivation', *Indian Linguistics*, 33, 1972, pp. 184-90). Such things as ambisyllabic consonants work in a way reminiscent of Korku. Similar features seem relevant to an explanation of the more complex combining form (CF) derivation rules for Sora and Sora-Gorum (SG). (See A.R.K. Zide's paper in the proceedings of the Honolulu Austroasiatic conference for more on this.)

The stress system found in Korku as applicable to the numerous suffixes and enclitics, particular in the Korku verb lacks much fusional morphology - elaborate internal sandhi - and any distinctive (to particular functional sets of morphemes) stress machinery. What morphophonemic variation there is (see the dissertation on the developments from the verbal suffix {ya}) is fairly transparent, and seems quite recent. More fused and less transparent is the demonstrative stem derivational morphology, and that of the numerals. This fact would tend to support the notion that the Korku verb (and the North Munda verb more generally) suffix system is comparatively recent, and certainly is not to be thought of as Proto-Austroasiatic. Certain affixes with initial vowels, e.g. the (in Korku) intransitive imperative -e, the transitive past -e?, and perhaps the passive-potential -uq go back to PM. Similarly, few of the Korku nominal postpositions are old in Munda. Korku lacks the verbal pronominal prefixes, the causative prefix and infix (except, perhaps for a- in a-jom 'to feed, give-to-eat', a-nu 'to give-to-drink') and the nominal prefixes preserved in SM (except for V- in some pronouns and the numbers 'three' and 'four').

Something should be said about diphthongs and semivowels treated at some length with regard to their stress in the preceding paper. Most of them came in with loanwords, but the borrowing must have been complex and over a long period of time.

Presumably we do not reconstruct y or w, but some y come from earlier (non-borrowed) ñ. In some cases perhaps y (or post-vocalic i) comes from the laryngeal X (as presumably the i in Mundari hai 'fish' from NM \*kaX; there are other examples of this in SM). There are a few examples of intramorphemic diphthongs, but most of them come via recent loans. The exceptions, e.g. siu- 'to cultivate, plough' probably reflect verbal suffixes, e.g. si()-uq-? The morphology of the numerals is complex, but among other developments (e.g. of X, say, in turui 'six') it looks as if a final suffix -i is found in some of them. It is clearly present in bar-i 'two'. The final -ia, and -ea in many names are probably from borrowed suffixes, and in fact - although I cannot begin to prove this - it looks as if almost all the diphthongs and

semivowels are simply derivable from full vowels occur in loan forms, forms whose complex history is largely obscure.<sup>1</sup>

<sup>1</sup>Something more (the dissertation, *Korku Phonology and Morphophonemics*, University of Pennsylvania, 1960, can be consulted for fuller information) on tone-aspiration and tone sandhi may be helpful here. Tone - and whether Korku "tone" is properly so labelled will be discussed elsewhere, i.e. low tone is associated with aspiration in Korku. A low-toned vowel that immediately follows (with no intervening morpheme boundary) an aspirable non-initial consonant (i.e. p, t, k; c can be added since s can be interpreted as aspirated c, but there are problems in doing this) is automatically aspirated, and the reverse holds, i.e. where any aspirated consonant (of the above set) preceding a non-initial-syllable vowel (with no intervening morpheme boundary between them), that vowel is necessarily low-toned. The two are perfectly correlated, i.e. where occurrence restrictions permit, one cannot occur without the other. These restrictions are: 1. (Overt) low tone occurs only in non-word-initial syllables immediately preceded by high-toned syllables. 2. Voiced aspirates occur (overtly) only word-initially, and thus necessarily in high-toned syllables, where word high tone is not overridden by a preceding low tone in the phonological phrase. 3. The restrictions on the occurrence of voiceless aspirates require within-word morphological information: a voiceless aspirate occurs in a morpheme only when there is no preceding aspirate in the morpheme. More than one (two I think is the limit) - overt - voiced aspirates can occur in a simple word (i.e. a phonological unit with no + or = junctures), necessarily in different morphemes, the second voiced aspirate (usually kh), usually (always?) initially in a monosyllabic verbal or nominal suffix, e.g. [ukhùkhènéj], {uku}{ki}{eg}{ej}, 'hide-intens.-past-3sg.an.obj.', 'hid him/her (intens.)'.

The dissociation of aspiration from low tone in environments where both can occur occurs only where there are reduced or contracted vowels separated from their preceding consonants (in the same syllables) by morpheme boundaries. Thus, /koÿkiÿba/ (i.e. [koÿkiÿbà] - from {koÿ-kiÿ-ba} 'call-3an.dual-predicator', i.e. 'calls them (dual)') contrasts with /koÿkiÿba/, i.e. [koÿkhiÿbà] - from {koÿ-ki-iÿ-ba} 'calls me (intensive)'. There are several sources of aspiration-low tone in Korku that seem independent of any "underlying low tone-aspiration", one being in certain (specified) associations with preglottalised consonants (and glottal stop), e.g. the {-ki-iÿ} above, presumably modelled on {-ki-ej} > -khèj (i.e. /-kej/), and, most generally and productively, in reduplication of CVC<sup>8-</sup> (C<sup>8</sup>: b, d, j, q) verbstems, e.g. kab-, Rdup-kab > kakhàb (i.e. /kaka**b**/), 'to bite'. There is no reason to claim that the stem should be \*k**ab**-, or that the V is "low-toned" (although this could be directly shown only with the stem in non-initial position (where it never occurs)), since all vowels in all CVC<sup>8-</sup> verbstems (and these are very common) would then have to be considered low-toned, which - for several reasons - is ill-advised.

REFERENCES

DRAKE, J.

1903 *A Grammar of the Korku Language*. Calcutta.

KELKAR, Ashok R.

1958 *The Phonology and Morphology of Marathi*. Ph.D. dissertation, Cornell University.

MacCORQUODALE, K. and P. MEEHL

1948 'Hypothetical Constructs and Intervening Variables'. *Psychological Review* 55.

O'CONNOR, J.D. and J.L.M. TRIM

1953 'Vowel, Consonant, and Syllable - a Phonological Definition'. *Word* 9:103-22.

SHARMA, Aryendra

1969 'Hindi Word-accent'. *Indian Linguistics* 30:115-18.

ZIDE, Norman H.

1960 *Korku Phonology and Morphophonemics*. Ph.D. dissertation, University of Pennsylvania.

1972 'A Note on the Historical Phonology of Gta', Combining Form Derivation. *Indian Linguistics* 33:184-90.



## MORPHOPHONEMICS OF VERB SUFFIXES IN TSOU

PATRICIA STANLEY

1. Introduction
2. Some Suggested Phonological Analyses
3. Verb Inflection
  - 3.1. Verbs without Stem Change
  - 3.2. Verbs with Suppletive Suffixes
  - 3.3. Verbs with Final Vowel Deletion
  - 3.4. Some Residual Problems
4. Summary

### 1. INTRODUCTION

Tsou, an Austronesian language of Taiwan, has commonly been analysed without final consonants in underlying forms. At the same time, there are not only medial consonant clusters of a fair degree of complexity, but also such clusters initially. In addition, final vowels fall phonetically into three separate groups, characterised by devoicing, an [h] offglide, and glottalisation respectively (Lin 1952:210). Along with the unusual contour of its base forms and the variation in treatment of final segments, Tsou is alleged to have a high degree of irregularity in verb inflection. Before inflectional suffixes, some of which have three allomorphs, vowels sometimes change, sometimes drop, and occasionally remain unchanged.

It is the contention of this paper that the underlying structure of this language is far different from that implied superficially by the phonetic data sketched above. The supposedly irregular verb inflection is in fact quite regular; moreover, the underlying morphemes of the language are of a quite different configuration than indicated by an

analysis which demands that there be no morpheme-final consonants. The argument will be that both the treatment of final vowels and the phonological changes involved in suffixing can be accounted for most economically if underlying forms may have, in any syllable, final consonants.

## 2. SOME SUGGESTED PHONOLOGICAL ANALYSES

With minor difference, previous workers (Ogawa 1935, Lin 1952, Tung 1964, and Tsuchida 1969) consider the segmental phonemes of Tsou to be the following:<sup>1</sup>

p	t	c	k	ʔ	i	u
b	d				e	ə
m	n		ŋ		a	
f	s		h			
v	z	r				

Both Tung and Ogawa agree that syllabic vowels may occur without intervening consonants. Two-consonant clusters are quite common, both initially and intervocalically. Tung's phonetic characterisation of consonant clusters will be pertinent to the discussion to follow. He points out that while in general a voiceless stop is unaspirated, "Only when /p t c k/ are followed by /f/ or /s/ a 'puff of the air' is always heard between the stop and the spirant." (Tung 1964:10). This puff of air is so distinct that, as he points out, it is often transcribed as a vowel, either an echo of the preceding vowel or a neutral or low vowel ([ə] or [a]).

While there is general agreement on the status of consonants and vowels, glides are more of a problem. The close phonetic transcription of Ogawa indicates that there is some free variation between vowels and glides. This variation is found within the speech of a single speaker, as, for example, the following set of alternations within a single text (Ogawa 1935):

jaintsa iaintsa 'say'

It does not appear from this data that glides are phonemically distinct from vowels.

Tung never records glides separately in the orthography. Rather, he states a rule for determining whether in context a segment is to be interpreted as syllabic (i.e. a vowel) or non-syllabic (i.e. a glide):

"/e/ is syllabic immediately before another stressed vowel or after it but not in the final position. /o/ is non-syllabic between two other vowels of which the one preceding it is stressed. Otherwise they are syllabic." (Tung 1964:20)

Tung indicates another variation, one between vowels and continuant consonants:

"Whereas Tfuea and Luhtu [Tsou villages] have /pz, nz, hz/, etc., Tapangu [a third village] has the vowel /i/ in place of /z/,... It goes without saying that the /z:i/ correspondence does not apply to /z/ as a simple consonant in Tfuea and Luhtu. In that case /z/ in Tfuea and Luhtu is also /z/ in Tapangu." (Tung 1964:18)

"In the speech of the Luhtu people, another apical vowel we symbolize with 'r' is found in the place of /e/ in the other two dialects on many occasions in certain positions, leaving /e/ totally uneffected [sic] in those positions only in relatively few cases ... However, the situation is not so simple in regard to the occurrence of /r/. It is in fact not consistently distinguished from /e/ with different persons." (Tung 1964:20)

These descriptions of vowels and non-syllabic alternations make the nature and form of these non-syllabics far from clear. For example, the segment transcribed as e may either be syllabic, non-syllabic (but presumably not a variant of r), or varying dialectally with r. o may be syllabic or non-syllabic, but presumably distinct from the segment v which is recorded in the phonological inventory.

These variations are attributed, then, sometimes to free variation, and at other times to dialect differences. In any event, their phonemic status is in doubt. As we shall see, proper analysis of these segments is central to the question of canonical form in Tsou. A great deal about their nature and behaviour may be learned through the analysis of verb forms, the problem to which we now turn.

### 3. VERB INFLECTION

Inflected verbs have both suffixed and unsuffixed forms.<sup>2</sup> While the unsuffixed verbs may have other affixes, for our purposes we shall be concerned only with the difference between suffixed and unsuffixed forms. These suffixed forms will, furthermore, be limited to two categories: those in which the suffix a (or its putative allomorphs va or za) appears, and those in which the suffix i (with its putative allomorphs vi or zi) appears.

#### 3.1. VERBS WITHOUT STEM CHANGE

Regular verbs are those in which the suffixation of a or i apparently causes no changes in the stem.<sup>3</sup> These are such verbs as:

<i>Unaffixed</i>	<i>Affixed</i>	
etokə	etokəa	'strike with cudgel'
əmnə	əmnəa	'be good'
maaseu	paaseua	'fish with a net'
coŋo	coŋoa	'hurt'
tufŋi	tufŋia	'fish by torchlight'
tmələ	taləi	'hear'
eofŋu	eofŋui	'headhunt'
pofuŋu	pofuŋua	'carry on head'
zonso	zonsoi	'shoot deer by a stream'
sume	sumea	'be sweet (of wine)'

Clearly, these forms may be analysed as the simple addition of a suffix to a stem, without morphophonemic change. If all stems are in fact vowel-final, then most of the verbs should be in this category (allowing for some irregular forms). In fact, according to Tung, most verbs are irregular in one way or another. It is these supposedly irregular forms which will in fact demonstrate regular morphophonemic changes.

### 3.2. VERBS WITH SUPPLETIVE SUFFIXES

A large group of verbs are analysed by Tung as dropping the stem-final vowel of the unaffixed form before adding the suffix. The suffix will then itself be irregular, either *za* or *va* as an allomorph of *a*, or *vi* or *zi* as an allomorph of *i*. For example:

<i>Unaffixed</i>	<i>Affixed</i>	
ahoi	ahoza	'begin'
toʔtohəŋə	toʔtohəŋva	'think'
sifkou	sifkova	'flay'
eansou	eansovi	'breathe'
toniou	toniovi	'wash'
tiavai	taivaza	'hold with two hands'
eiʔmi	eiʔmzi	'come from'

To consider these forms as irregular misses a clear phonological generality. Where the final vowel in the unaffixed form is *i*, the putative suffix begins with *z*; where it is *u* or *ə*, the suffix is said to begin with *v*. It is not the suffixes which vary, it is the stems.

The question then is whether a stem-final vowel changes to a consonant before a suffix; or whether on the other hand the consonant becomes a vowel (or glide) in word-final position. Since in the preceding examples we have vowels which have not changed before suffixes, these

must be instances of word-final continuant consonants becoming glides. Only before suffixes are they unchanged.

Additional evidence that these forms should have final consonants in the stem is provided by forms of the following type:

<i>Unsuffixe d</i>	<i>Suffixe d</i>	
aasoe	aasoeza	'peep'
aaŋae	aaŋaeza	'share'
tatae	tataeza	'admire'
ŋøe	noevi	'carry a burden'

We now see that a final *v* or *z* is deleted if word-final following *e*; and this is not the case with word-final vowels, as can be seen from examining the forms previously cited.

### 3.3. VERBS WITH FINAL VOWEL DELETION

The largest group of verbs falls into yet another category. In this group, verb-final vowels drop without a trace before the suffixes. Examples of verbs of this type are:

<i>Unsuffixe d</i>	<i>Suffixe d</i>	
tiemucu	tiemuca	'hold hands'
efeutu	efeuta	'harvest'
euʔpici	euʔpica	'divide'
sochipi	sochipa	'look after'
tuefisi	tuefisa	'pull out'
aezuhu	aezuha	'change'
miusnu	miusni	'walk to'
iupu	iupi	'be together'
suputu	suputi	'meet'

It was indicated earlier that a consonant is always released, even if it is the first member of a cluster. Tung indicates this release as occurring only before *s* or *h*, but Ogawa transcribes the release vowel generally for all consonant clusters. If the supposed final vowel in the forms above is an echo vowel, then its non-occurrence before a vowel is quite regular. In addition, the consonant release phenomenon is not limited to the environment before another consonant, but word-finally as well. In short, the forms above are to be analysed with stem-final consonants. This final release follows glottal stop as well, and at least one other, previously unanalysed phoneme, as we shall see later.

While the consonant release phenomenon is most often heard as an epenthetic vowel, Ogawa sometimes transcribes a neutral vowel ([ə]) in

this environment. Accordingly, possibly some instances of final schwa are to be included as instances of consonant release, for example:

<i>Unaffixed</i>	<i>Suffixed</i>	
koicə	koica	'scold'
coecəŋə	coeconə	'walk'

While most echo vowels follow consonants, some follow a segment transcribed as e:

<i>Unaffixed</i>	<i>Suffixed</i>	
eahŋəəə	eahŋəea	'request marriage'
ehkueu	ehkuei	'encircle'

It is obvious from the assessment of the data so far, as well as from the remarks of Tung quoted earlier, that a large degree of uncertainty exists in the transcription of vowels *vis-à-vis* voiced continuants. In the case of e, for example, the same informants are interpreted as frequently varying between [r] and [e]. On the other hand, e, according to Tung, has a non-syllabic variant which is not equivalent to [r]. Clearly, at least two phonemes, and possibly more, are being confused. The present analysis provides a partial disambiguation. Segments which are followed by an echo vowel are consonants: hence the e preceding the epenthetic vowel above is a consonant, not a vowel or resonant.

The assignment of the epenthetic vowel-forming e to a consonantal phoneme has advantages other than the regularisation of paradigms. One is that the phoneme /r/, which had been restricted to the Luhtu dialect, now must be part of the inventory in all dialects. This is in accord with the notion that dialects are much more likely to differ phonetically than phonemically. It can now be stated that r is much more lax in the Tapangu and Tfuea dialects than in Luhtu. Laxing of voiced continuants in general is characteristic of dialect differences in Tsou. Tung's statements regarding the distribution of r and z *vis-à-vis* e and i may be tabulated as follows:

<i>Luhtu</i>	<i>Tfuea</i>	<i>Tapangu</i>
[..pr..]	[..pe..]	[..pe..]
[..pz..]	[..pz..]	[..pi..]

It is apparent that no laxing rule is present in Luhtu in these examples. The rule is general in Tapangu, and more restricted in Tfuea. One might expect from the fact that Tfuea is intermediate linguistically between Tapangu and Luhtu, in regard to this, that it might also be in the intermediate position geographically. This is the case.

Possibly another apparent irregularity is to be explained by such a continuant laxing phenomenon: this is the absence of final l. It could

well be that this segment is phonetically laxed, if final, in all dialects. Further research is needed to make such a determination.

In addition to the environments described above, some epenthetic vowels appear with apparently no conditioning factor at all. An epenthetic vowel seems to follow directly a final phonemic vowel:

<i>Unaffixed</i>	<i>Suffixed</i>	
zotpuu	zotpui	'pound'
aəiəiəə	aəiəiəə	'take good care of'

Note that a number of instances have already been given with stem-final vowels. In these instances, there is no doubling or lengthening of the vowel in the unaffixed forms. These forms should have some segment between the phonemic vowel and the epenthesis.

Published data on Tsou are particularly inconsistent in regard to the transcription of glottal stop. This suggests that phonemic glottal stop is often phonetically not distinct from a syllable boundary between contiguous vowels. From the viewpoint of systematic phonology, however, the distinction between a syllable break and a glottal stop is critical. If these forms are analysed without final underlying glottal stop, then they are irregular forms. Such an analysis also entails a very unusual distribution for the glottal stop: unlike other consonants, it never occurs in morpheme-final position.<sup>4</sup> Hence, it is proposed here that these forms do contain a morpheme-final glottal stop, and undergo the regular process of acquiring an epenthetic vowel if the glottal stop is also word-final.

#### 3.4. SOME RESIDUAL PROBLEMS

As was pointed out earlier, verbs which do not change the final vowel before the suffix are far less frequent than those which do. The following are examples of the latter type:

<i>Unaffixed</i>	<i>Suffixed</i>	
toʔtoħəŋə	toʔtoħəŋva	'think'
eəsŋece	eəsŋəcva	'stride away'
toaluŋu	toaluŋva	'angle'
eoəsao	eoəsava	'play'
toesoso	toesosvi	'fly'

In these forms, any final non-front vowel (except a) alternates with a pre-suffix v. The situation with front vowels is more ambiguous, because the transcription of Tung does not differentiate between i and y. Nevertheless, the same pattern does clearly occur when the front vowel is e. Forms with e-y alternation are:

<i>Unaffixed</i>	<i>Suffixed</i>	
eepe	eepia	'raise'
psoepepe	psoeepia	'fly up'
soepe	soepia	'stick in'

If the forms which have *v* immediately preceding the suffix are analysed without that segment in the underlying form, then it is impossible to predict the height of the vowel in word-final position. Likewise, the change from *y* to *i* or *e* is unpredictable (quite aside from the fact that there is no independent justification for positing an underlying /*y*/). These forms must have final vowels in the base.

We now have two classes of verbs with final vowels: those where the final vowel may become a glide before the suffix, and those where it does not. The question now is whether these classes are in fact phonologically conditioned; or whether there is simply free variation which is not indicated by the corpus. Another possibility is that these represent morphological classes.

An indication that the difference is phonological rather than simply morphological is Lin's observation that there are audible phonetic differences among final vowels which are otherwise homorganic. The "devoiced vowel" has already been analysed as a final epenthetic segment. The glide-forming and non-glide-forming classes of vowels may possibly coincide with the *h*-offglide and glottal-offglide classes (though not necessarily in that order). A hypothesis for a phonological difference between these two classes should account also for the phonetic difference.

The assumption has been made that vowels optionally become non-syllabic when in contact with other vowels. The last group of forms considered apparently show this type of variation. The verbs which never undergo a final vowel change must then be the exceptional forms.

One possible analysis is that these forms are not vowel-final, but have an additional segment which separates the last vowel from the suffix, a consonantal segment which prevents desyllabification. Such an analysis then raises problems about the nature of the possible final consonant. In this situation, one would expect the consonant to be followed by an epenthetic vowel, or, like *v* and *z*, to become a glide if word-final. In fact, a third word-final glide is described by Lin: the *h*-offglide. If this last offglide is also phonemic, and if it is velar, as seems plausible, then the reason that it is also not heard between vowels is obvious: like the glottal stop, *it*, too, is heard only as a syllable boundary between vowels.

The positing of a velar continuant, /*ɣ*/, as an underlying phoneme has several arguments to recommend it. Considerations of symmetry, for



example, should call for a voiced velar continuant in a language which has velar voiced stops, nasals, and voiceless continuants, as well as a series of labial, dental, and palatal voiced non-nasals.

Secondly, words which are transcribed with initial vowel seem to be phonologically idiosyncratic when / prefixed. It appears likely that some of these apparent idiosyncracies can be resolved by positing an underlying / $\gamma$ / (realised perhaps as some sort of "smooth onset" contrasting with glottal stop) in initial position in some forms.

It is also the case that related Formosan languages, Paiwan for example, show velars in final position in cognates where / $\gamma$ / is proposed for Tsou (for example, Ts. /ʔŋayo/, Pai. /ʔŋaŋʔ/, 'mouth'; Ts. /fʔuhuy/, Pai. /uquʔ/, 'back'). Comparative data is of course not evidence for synchronic phonology. What it does argue, however, is that there once was such a phoneme in Tsou. The alternative here is that there was independent parallel development of a phoneme in Paiwan and a morphological class in Tsou, a most unlikely prospect. Whether there is now such a phoneme in Tsou or only a morphological class as the remains of an earlier phoneme can only be determined by internal considerations. The evidence cited earlier indicates that there is still such a phoneme.

Another problem which requires further research is that behaviour of voiced continuants. It appears that the demarcation between final segments which take epenthetic vowels and those which become glides is still not clear. In some instances, e.g. tpuzu : tpuza 'put into fire', apparently final z receives epenthesis. In like manner, nasals sometimes receive epenthesis, sometimes remain in final position, and sometimes drop. Much more work is needed to understand the complete conditioning for these phenomena.

#### 4. SUMMARY

The description of Tsou without final consonants and with a great deal of irregularity in verb inflection is a description of surface phonetic facts only. From the point of view of systematic phonology, Tsou has morphemes capable of ending in consonants as well as vowels. Further, the inflection is quite regular, with predictable morphophonemic changes.

It is clear that new research in the field is needed to resolve the phonemic and phonetic questions which the available data raises.

NOTES

1. *d* is transcribed by Tung as *l*, and the segment which we have transcribed as *ə* as *ɱ*. Tung regards the bilabial voiced continuant as an allophone of *b*, but Ogawa considers it an allophone of *v*.
2. The suffixes given are for focus. The appearance and form of focus inflection appears to be dependent on aspect as well, however, as are the focus particles of Paiwan, a closely related language. A description of the various forms of focus and aspect particles is given in Ferrell 1970, p. 16. The Tsou forms seem to be comparable.
3. All examples are taken from Tung 1964.
4. Current analyses are very weak with respect to the glottal stop. For example, the only word-initial glottal stops listed in Tung are those in numbers, a highly dubious analysis. A rigorous analysis of distribution would also bring into question the putative pre-glottalised segments *b* and *l*. Such an analysis is now in process, and seems to indicate that these segments are phonemically clusters.

## BIBLIOGRAPHY

## FERRELL, Raleigh

- 1969 *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. Monograph No.17. Taipei: Institute of Ethnology, Academia Sinica.
- 1970 *Paiwan Stem Vocabulary* (draft). Taipei. Mimeo. 180 pp.
- 1971 Ergative, Passive or Focus-marking? Elicitation of Syntax in Indonesian Languages. Paper given at Colloques Internationaux du Centre National de la Recherche Scientifique sur Langues sans Tradition Écrite. Paris.
- forth- 'Le Paiwan'. In: *Les Langues du Monde* (revised edition).  
coming Paris: Centre National de la Recherche Scientifique. (In press.)

## FILLMORE, Charles J.

- 1968 'The Case for Case'. In: Emmon Bach and Robert T. Harms, eds *Universals in Linguistic Theory*, 1-88. New York: Holt, Rinehart and Winston.

## LIN, Hen-li

- 1952 'Tsou Language'. In: Wei, H.L., Yu, C.C. and Lin, H.L. *Ethnography of the Ethnic Group Tsou* [in Chinese], 209-53. Taipei: Commission of Historical Research of Taiwan.

## McKAUGHAN, Howard P.

- 1962 'Overt Relation Markers in Maranao'. *Language* 38/1:47-51.
- 1970 'Topicalization in Maranao - an Addendum'. In: S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic Studies in Honour of Arthur Capell*, 291-300. PL, C-13.

OGAWA, Naoyoshi and Erin ASAI

1935 *Taiwan Takasagozoku Densetsu-shu*. [*Myths and Traditions of the Formosan Native Tribes.*] Taipei: Yaihoku Imperial University.

TSUCHIDA, Shigeru

1969 Tsou Wordlist. MS.

TUNG T'ung-ho

1964 *A Description Study of the Tsou Language, Formosa*. Special Publication No.48. Taipei: Institute of History and Philology, Academia Sinica.

# CONSTRUCTION MARKERS AND SUBGROUPING OF FORMOSAN LANGUAGES<sup>1</sup>

RALEIGH FERRELL

1. Topicalisation in Formosan Languages
2. Verb Inflection
3. Marking of Noun Phrases
  - 3.1. 3-CM Systems
  - 3.2. 2-CM Systems
  - 3.3.  $\emptyset$ -CM Systems
  - 3.4. General and Specificational CMs
4. Syntactic, Lexical and Phonological Subgrouping
5. Proto-Formosan Syntax

## 1. TOPICALISATION IN FORMOSAN LANGUAGES

Obligatory sentence focus (topicalisation) is basic to all Formosan aboriginal languages. The verb in all instances is inflected to show the role of the topicalised Noun Phrase (such as agent, object/goal, locus, instrument, cause, beneficiary). The topicalised NP itself is then marked by either (1) an over phrase-introducing particle (Construction Marker), or (2) a fixed-order function slot for the NP, or (3) both of these syntactic devices. These devices, whether overt CMs or constituent NP order, are equational: the same device which identifies a NP as being in primary relationship with the Verb Phrase also equates NPs in non-verbal (equational) sentences.<sup>2</sup>

Non-topicalised NPs are also marked by either CMs or constituent NP order. Typically, all non-topicalised NPs may be marked by identical,

non-equational CMs, except that in languages having a discreet genitive/partitive marker the latter is used to mark the agent/actor in non-Agent Focus sentences.<sup>3</sup>

In certain Formosan languages, the topicalisational CMs perform a dual role by being differentiated to indicate proximity or specificity as well as the focus/non-focus roles outlined above. In at least one language, focus/non-focus roles are marked strictly by fixed NP order; and CMs, which obligatorily introduce each constituent NP, function solely as articles indicating relative proximity or specificity. These unusual developments distinguish these languages from other western Austronesian ones, and heighten the interest of Formosan languages to comparative Austronesian syntactic studies.

## 2. TOPICALISATIONAL VERB INFLECTION

The primary importance of the verb inflections in Formosan focus (topicalisation) system is attested to by the widespread preservation of cognate forms for these inflections. This is in contrast to the variability and relative instability of overt CMs and other NP-marking devices connected with topicalisation.

In Formosan languages generally the various roles which may be filled by topicalised NPs are subsumed structurally under four verbal inflections, herein referred to as AF (Agent Focus), OF (Object Focus), RF (Referent Focus), and IF (Instrument Focus).<sup>4</sup> Figure 1 shows topicalisational verb inflections in various Formosan languages.

	<i>Amis</i>	<i>Atayal</i>	<i>Bunun</i>	<i>Favorlang</i>	<i>Paiwan</i>	<i>Pazeh</i>	<i>Saisiat</i>	<i>Sediq</i>	<i>Siraya</i>	<i>Thao</i>	<i>Yami</i>	<i>Tsou</i>	<i>Saaroa</i>	<i>Rukai</i>
AF	/m/	/m/	m-	/m/	/m/	mV-	/um/	/m/	/m/	/m/	/m/	/m/	/um/	ʷ-
OF	-ən	-un	-un	-en	-ən	-un	-Vn	-un	-en	-in	-ən	-i(?)	(?)	ki-(?)
RF	/m/+an	-an	-an	-an	-an	-an	-an	-an	-an	-an	-an	-a(?)	-a(na)	ta-+a(n)
IF	sa-	sə-	-is	i-	si-	si-+i	si-	sə-	(?)	sa-	i-	-(n)eni	-(a)ta	sa-+a(n)

Figure 1. TOPICALISATIONAL VERB FOCUS INFLECTIONS IN FORMOSAN LANGUAGES

	3-CM Systems						2-CM Systems			
	<i>Amis</i>	<i>Atayal</i>	<i>Kuvalan</i>	<i>Paiwan</i>	<i>Siraya</i>	<i>Yami</i>	<i>Puyuma</i>	<i>Sediq</i>	<i>Rukai</i>	<i>Saaroa</i>
CM=	o	∅	a	a	ta	u	a	ka	ka	ka
CM≠	to	sa	ta	tua	tou	su	} da	∅	sa	na
CMgen	no	na	na	nua	ki	nu				

Figure 2. GENERAL FOCUS CMs IN SOME FORMOSAN LANGUAGES

## 3. TOPICALISATIONAL MARKING OF NOUN PHRASES

## 3.1. 3-CM SYSTEMS

Amis, Kuvalan, Paiwan, Siraya, Yami, Pazeh, Thao, Favorlang and Ciuli-Atayal each have three topicalisational NP Construction Markers. In these 3-CM systems, the topicalised NP is introduced by an equational CM (CM=), the agent of non-Agent Focus sentences is introduced by the genitive/partitive CM (CMgen), and all other non-topicalised NPs are marked by a non-equational CM (CM≠).

The following examples from Paiwan show how 3-CM systems typically work. Note that in non-Referent Focus sentences, locative NPs may be introduced by secondary, CM-like markers<sup>5</sup> such as the Paiwan *i* 'at, in', which may either co-occur with the CM≠ or may result in deletion of the latter.<sup>6</sup>

'(The) man hunts wild-pigs in the mountains with a spear':

AF	q/m/alup hunts	a CM=	caucau man	tua CM≠	vavuy pig	$\left\{ \begin{array}{l} \text{tua} \\ \text{i tua} \\ \text{i} \end{array} \right\}$ CM≠	gadu mtn	tua CM≠	vu uq spear
OF	qalup-ən	nua CMgen	caucau	a CM=	vavuy	$\left\{ \begin{array}{l} \text{tua} \\ \text{i tua} \\ \text{i} \end{array} \right\}$ CM≠	gadu	tua CM≠	vu uq
RF	qalup-an	nua CMgen	caucau	tua CM≠	vavuy	a CM=	gadu CM≠	tua	vu uq
IF	si-qalup	nua CMgen	caucau	tua CM≠	vavuy	$\left\{ \begin{array}{l} \text{tua} \\ \text{i tua} \\ \text{i} \end{array} \right\}$ CM≠	gadu	a CM=	vu uq

Saaroa must also be included among the languages having 3-CM systems. Although Saaroa has only two overt topicalisational CMs (CM= and CM≠), structurally the equivalent of a 3-CM marking system is maintained by assigning the function of a CMgen to a  $\emptyset$ -marker. The following examples are illustrative:

m-aci?i ka caciji  
die CM= one person  
'one person died'

k/um/ita na a|a|iamu  
see[AF] CM≠ fly  
'(they) saw a fly'



sa|a?a takolu  
*trail wildcat*  
*'the wildcat's trail'*

Squliq-Atayal, on the other hand, has only CM $\neq$  and CMgen. In this language, the role of CM= is filled by lack of an overt CM before the topicalised NP:<sup>7</sup>

?malup squliq sa bizuak  
*hunt man CM $\neq$  pig*  
*'the man hunts the pig'*

nasal na squliq  
*house CMgen man*  
*'the man's house'*

### 3.2. 2-CM SYSTEMS

Bunun, Puyuma, Saisiat, Rukai and Sediq each distinguish by overt marking only two categories of NP. The 2-CM systems lack a separate genitive/partitive CM but differ from Squliq-Atayal in that all non-topicalised NPs, including the agentive NP in non-AF sentences, are marked by identical non-equational devices (see Figure 2). The following examples are from Puyuma:

t/əm/bəl a marəwadi ɟa ʔau  
*bury[AF] CM= brothers CM $\neq$  people*  
*'the brothers buried the people'*

romaʔ ɟa ʔau  
*house CM $\neq$  man*  
*'the man's house'*

Thus whereas Saaroa and Squliq-Atayal each have only two overt CMs but mark by  $\emptyset$  the categories of NP introduced in Paiwan respectively by CM= and CMgen, conversely in 2-CM systems two separate categories of NP are overtly marked by identical CMs. Constituent NP order then must play a more important role in 2-CM systems than in 3-CM ones to indicate the case-like roles of the various non-topicalised NPs.<sup>8</sup>

Tsou also appears to have a general 2-CM system. But as will be discussed in 3.4, Tsou has separate CMs in each of the two categories to show several degrees of proximity or specificity and warrants separate treatment.

### 3.3. $\emptyset$ -CM SYSTEMS

In Kanakanabu, there are no overt topicalisational CMs; all topicalisational NP marking is accomplished by constituent NP order:

ni-macay cau  
 died man  
 'the people died'

ivatu vavuru k/um/aəne vuruna  
 come pig eat[AF] eel  
 'the pig came and ate the eel'

Aside from fixed NP order, one way in which topicalisational ambiguity is avoided in Kanakanabu is by a preference similar to that of Squliq-Atayal (see 3.1) for non-AF sentence constructions where non-agent NPs are present in the sentence.

Maga and the other two so-called "Lower Three Villages Rukai" languages<sup>9</sup> also achieve topicalisational NP marking strictly by NP order. However, as will be seen in 3.4, these languages do have obligatory NP-introducing particles which fulfil quite different functions from topicalisational CMs.

### 3.4. GENERAL AND SPECIFICATIONAL CMs

Besides the general topicalisational CMs discussed thus far, two additional types of specificational CMs are found in Formosan languages. These two types of specificational CMs indicate respectively (1) proximity or definiteness, and (2) named individuals.

Atayal, Puyuma and Rukai have parallel sets of general and definite (or proximal) CMs. The latter indicate either that the NP is nearby or gives it a specificity similar to that expressed by the definite article in European languages:

	<i>Squliq-Atayal</i>		<i>Puyuma</i>		<i>Rukai</i>	
	General	Specific	General	Specific	General	Specific
CM=	∅	qo·	a	na	ka	ku
CM≠	sa	s-qo·	} da	ka-na	} sa	ki
CMgen	na	na-qo·				

Tsou, on the other hand, is said to have a complex system of topicalisational CMs which simultaneously indicate focus relationships and relative proximity for each NP:<sup>10</sup>

	<i>In Sight</i>			<i>Out-of-sight</i>			
	Near	Away	Remote	Known	Hearsay	Speaker sees, Hearer not	Hearer sees, Speaker not
CM=	e	si	ta	o	na	yao	eyo
CM <del>≠</del>		ta		to	no		

Maga and the other two "Lower Three Villages Rukai" languages have CM-like specificational particles which, although they have nothing to do with topicalisation, are obviously related to the specificational topicalisational CMs of other Formosan languages. These obligatory NP-introducing particles in Maga are:

na (visible)

ku (out-of-sight)

Thus, in the following Maga examples topicalisational NP marking is effected by NP order, and the obligatory CM-like particles serve roughly as definite/indefinite articles:

latbi na v|akə

weep child

'the child weeps' (and the child is visible)

latbi ku v|akə

weep child

'the child weeps' (and the child is out-of-sight)

The second type of specificational CMs, found in Amis, Atayal, Bunun, Kuvalan, Paiwan, Puyuma, Siraya, Yami, Saisiat and Favorlang, indicate that the NP represents a named individual. Cognate CMs fulfilling the same functions are common among Philippine languages. Examples from Paiwan and Squliq-Atayal are:

	<i>Paiwan</i>	<i>Squliq-Atayal</i>
CM=(pers)	ti	i
CM <del>≠</del> (pers)	tʔai	∅, -an
CMgen(pers)	ni	ni

#### 4. SYNTACTIC, LEXICAL AND PHONOLOGICAL SUBGROUPING OF FORMOSAN LANGUAGES

A tripartite subgrouping of Formosan languages (Atayalic, Paiwanic, Tsouic) has been suggested by lexicostatistical studies (see Dyen 1971). On both lexical and phonological grounds there is little doubt concerning the close genetic relationship between Squliq-Atayal, Ciuli-Atayal and

Sedıq (the Atayalic subgroup). Tsou, Kanakanabu and Saaroa also cluster well to form the Tsouic subgroup; it is likely that Rukai and the "Lower Three Villages Rukai" are also Tsouic languages.<sup>11</sup> All the other Formosan languages are thus far considered to fall into the catch-all Paiwanic grouping; comparative phonological work has not yet progressed to the point that further subgrouping can be done with any degree of confidence. Nor have the relationships between the three major subgroups been determined.

I have noted (Ferrell 1972) apparent confirmation of the three-way subgrouping of Formosan languages from the standpoint of relatively superficial differences in transitive verb syntax. It appears, however, that some syntactic processes such as conjunct verb constructions may link Tsouic and Paiwanic somewhat more closely than either of these relates to Atayalic.

The present study has indicated the relative instability of topicalisational NP-marking devices. It will be seen from Figure 3 that considerable flexibility is possible in the actual marking mechanism for NPs in focus constructions (3-CM, 2-CM and  $\emptyset$ -CM systems) without serious consequences to the basic focus system itself, and the choice of one or the other of these systems cuts across all subgroupings established by lexical and phonological comparison. Nor does geographical distribution indicate any strong area influences.

Thus the fact that Thao and Saaroa both have 3-CM systems and lack person name CMs, whereas each of these languages is surrounded by respectively more closely related languages which do have the latter, suggests that the similarities between Thao and Saaroa are probably due to independent simplification, rather than to genetic or area influences.

The number and form of general topicalisational CMs in Formosan languages then are of little value in establishing genetic subgroupings. The two types of specificational CMs (see 3.4), on the other hand, may have more important implications in that their occurrence in various languages can not be the result of parallel simplification, and independent invention is not particularly likely. The non-occurrence of these specificational CMs in given languages may be due to independent simplification; but their occurrence most likely indicates either genetic or area relationships.

Separate topicalisational CMs for use with personal names are widespread in western Austronesian languages. In Formosa they are found in Atayal and all the Paiwanic languages except Pazeh and Thao. They are conspicuously absent from the Tsouic languages (see Figure 3).

Discreet topicalisational CMs indicating proximity or definiteness are found in all three major Formosan subgroups (Atayalic, Paiwanic,

Tsouic), although in Paiwanic they are found in only a single language (Puyuma).

	Personal Name CMs	General Focus CMs			Specificational (Definite) CMs
		3-CM System	2-CM System	∅-CM System	
<i>Amis</i>	+	+			
<i>Bunun</i>	+		+		
<i>Favorlang</i>	+	+			
<i>Kuvalan</i>	+	+			
<i>Paiwan</i>	+	+			
<i>Pazeh</i>	+	+			
<i>Saisiat</i>	+		+		
<i>Siraya</i>	+	+			
<i>Thao</i>	(?)	+			
<i>Yami</i>	+	+			
<i>Puyuma</i>	+		+		+
<i>Atayal</i>	+	+			+
<i>Sediq</i>			+		
<i>Saaroa</i>		+			
<i>Kanakanabu</i>				+	
<i>Maga-Rukai</i>				+	+
<i>Rukai</i>			+		+
<i>Tsou</i>			+		+

Figure 3. DISTRIBUTION OF CM-TYPES IN FORMOSAN LANGUAGES

There appears to be little likelihood of area influence explaining the occurrence of these CMs in both Atayalic and Tsouic languages. Parallel

innovation is a possibility, but the simplest assumption is that they are a common inheritance from a proto-language. Their loss in some languages from each group, viz. (Atayalic) Sediq and (Tsouic) Saaroa and Kanakanabu, could then be due to independent simplification. This could also be the case for the Paiwanic languages individually, although their near-universal absence in Paiwanic languages might rather indicate that their absence represents a shared innovation of the Paiwanic languages as a group. Their occurrence in (Paiwanic ?) Puyuma, on the other hand, either (1) strengthens the case for the existence of these CMs in Proto-Formosan, or (2) must be attributed to independent invention (if Puyuma is in fact a Paiwanic language), possibly as a result of area influence.

In summary, the occurrence and specific number of general focus CMs in Formosan languages can tell us little regarding either genetic relationships or language contacts. The occurrence of special (definite/specificational and Personal Name) CMs, on the other hand, may be considered to support the three-way subgrouping of Formosan languages in that (1) the Tsouic subgroup is marked by the absence of Personal Name CMs, and (2) the Paiwanic subgroup [with the unexplained exception of Puyuma] is marked by the absence of definite/specificational CMs. Meaningful genetic subgrouping of present-day Formosan languages on the basis of CMs alone would not be feasible, however, since some languages (Atayal and Puyuma) have all three types of CMs whereas other languages from different subgroups (e.g. Sediq, Saaroa) lack both specificational and Personal Name CMs.

##### 5. IMPLICATIONS FOR PROTO-FORMOSAN SYNTAX

The universal distribution of obligatory sentence topicalisation in present-day Formosan languages suggests that topicalisation by focus inflection was characteristic of whatever proto-language was ancestral to all these languages.<sup>12</sup> In this Proto-Formosan focus system, as in the modern languages, verb focus inflections indicating the case-like roles of topicalised NPs were of primary importance.

As in the modern languages generally, topicalised NPs were probably marked by overt CMs. Present data suggest the possibility that CMs in the proto-language either served the dual function of topicalisational NP marking and indicating general/specific or near/remote distinctions, or that there were separate sets of CMs performing these functions.

## NOTES

1. This paper was presented at the First International Conference on Comparative Austronesian Linguistics, Honolulu, 1974. Grateful acknowledgement is given to the Centre National de la Recherche Scientifique, Paris, under whose auspices I carried out ethnographic and linguistic fieldwork in Taiwan in 1968-70, and to the National Science Foundation for grant #GS-28818 for linguistic fieldwork in summer 1971.

2. Compare examples from Paiwan, in which this equational relationship is indicated by a CM, and Atayal in which it is shown by phrase order:

*Paiwan*

- S VP + NP k/m/an a alak (*eats CM child*) '(the) child eats'  
 S NP + NP alak a vavaian (*child CM woman*) 'daughter; (the) child is female'

*Atayal*

- S VP + NP m-aniq laqi (*eats child*) '(the) child eats'  
 S NP + NP laqi knairil (*child woman*) 'daughter; (the) child is female'

3. Other types of CMs, such as those introducing relative, coordinate or subordinate clauses, those marking verb/topic transposition (and thus additional emphasis on the topic), those showing genitive/partitive relationships, etc., are mentioned here only where they bear upon topicalisation.

4. In Paiwan, for example, inflection of the verb for AF shows that the topicalised NP is agent or actor; OF may indicate direct object, patient or goal; RF may be locus, beneficiary or indirect object; and IF may be instrument or cause. For obvious semantic reasons individual verbs may be "defective", i.e. not susceptible to inflection for topicalisation for one or more of these potential roles.

5. These prepositional CM-like markers should probably be considered to be conjunct verbs; cf. my communication to the First International Conference on Comparative Austronesian Linguistics, Honolulu, 1974, "Conjunct Verbs and Verb-Object Incorporation in Formosan Languages".
6. In such cases, inclusion of the CM~~≠~~ gives relative specificity to the locational NP, whereas deletion of the CM~~≠~~ makes the locative NP more general; in other words, when the CM~~≠~~ co-occurs with the "prepositional" particle, it is semantically comparable to the definite article in English.
7. Note that in Squaliq-Atayal, although the general CM~~≠~~ (sa) exists its use is in fact rare due to an apparent preference for non-AF constructions when object, referent or instrument NPs are included in the sentence.
8. For instance, to differentiate agent from beneficiary in non-AF sentences.
9. The relationship between Rukai "proper" and the so-called "Lower Three Villages Rukai" is problematic. Whether the latter, i.e. "Maga" (Tardkanuu), "Tona" (Koṅadavanu) and "Mantauran" (Opunoho) are actually three separate languages or dialects of a single language is also undecided.
10. This follows Ogawa and Asai's (1935) analysis.
11. Lexicostatistically Rukai appears to be closer to Paiwan. However, the propinquity and close cultural ties between the Paiwan and Rukai suggest heavy borrowings; and structural considerations appear at this point to indicate closer genetic ties between Rukai and Tsouic.
12. "Proto-Formosan" in this paper is not meant to infer that all Formosan languages form a single subgroup of Austronesian. Comparative research has not yet provided a decisive answer to this question. Proto-Formosan here means simply whatever ancestral language all Formosan languages did share in common. Thus Proto-Formosan might conceivably turn out to mean, for example, Proto-Formosan-Philippine, Proto-Western Austronesian, or even Proto-Austronesian.



REFERENCES

DYEN, Isidore

- 1971 'The Austronesian Languages of Formosa'. In: Thomas A. Sebeok, ed. *Current Trends in Linguistics*, vol. 8, *Linguistics in Oceania*. The Hague.

FERRELL, Raleigh

- 1972 'Verb Systems in Formosan Languages'. In: Jacqueline M.D. Thomas and Lucien Bernot, eds *Langues et Techniques, Nature et Société*. Paris.

OGAWA Naoyoshi and Erin ASAI

- 1935 *Taiwan Takasagozoku Densetsu-shū* (Myths and Traditions of the Formosan Native Tribes). Taipei: Taihoku Imperial University.



## FOCUS, TOPIC, AND CASE IN THE PHILIPPINE VERBAL PARADIGM

JOSEPH F. KESS

The one salient feature which sets Philippine languages off from other languages and which has consequently received a great deal of attention in linguistic descriptions has been the verb morphology and the relationship of the affixed verb to other parts of the sentence. The relationship of the affixed verb to the grammatical subject, or topic, of the sentence is an area which has been elaborated on by most descriptions of Philippine languages and an area which has labelled and re-labelled. The relationship between the sentence constructions so produced have been variously termed *voice*, active and passive (with three or more passive construction types being indicated), or *focus* construction types. The affixed verbs themselves have been referred to as having been marked by *focus-affixes*, and the relationship of the focus-affixed verb to the *subject* or *topic* complement has been taken to constitute a case relationship. In fact, it has been suggested that the case relationship is one indicated by these focus or case-marking affixes on the verb, such that the affixed verb indicates whether the topic is agent or actor, object or goal, location or referent, or instrument of the action indicated by the semantic content of the verb. Some descriptions, it might be added, have elaborated the basic set of four into more, including, for example, a beneficiary focus construction, an aptative construction, and so forth.

Furthermore, this basic schema has been suggested as implying a relationship between sentence types with the various sentential complements in one sentence construction able to be realigned with another sentence construction merely by a change in the focus- or case-marking affixes on the verb and a shift in the sentential complements with the

now-to-be focused sentential nominal complement moved to the topic position and marked by the topic-marking particle *ang*. For a more complete treatment of this type of description in the history of Philippine studies, see Constantino (1971) and McKaughan (1971), and for an appraisal of the manner of description, see Kess (1967, 1972, 1975, and 1976).

Certainly one typical illustration of this point of view and the way in which focus has been taken to operate by some can be seen in the following excerpt from Thomas' (1958) discussion of Mansaka sentence structure. "One of the most striking and important features of Mansaka and many other Malayo-Polynesian languages is the ability to put in the limelight a noun in any of the major sentence spots. The whole sentence polarizes toward that noun. This feature we are calling 'focus'. The form of the verb indicates which of the noun spots is being focused, and the noun occupying the spot is marked by" the topic-marking particle. The implication clearly is that the topic is far from just another surface structure noun phrase. It bears a special relationship to the verb. Moreover, one expects that any substantive phrase can become the topic of a sentence, according to shift in focus affix markers in the verb and the case-marking particles.

Some, like McKaughan (1958), instead chose to stress the syntactic aspects of the verb-topic relationship. Nevertheless, the relationship was singularly unique in some respect, and this state of affairs was noted in numerous Philippine languages and was taken as the cornerstone to ensuing descriptions of many Philippine languages. As a result, the description of many Philippine languages are easy to translate one to the other, perhaps as much a result of the fact of the relatively close relationships of the languages involved as well as the relatively close parallelism in the format of the description of the languages concerned. For example, note McKaughan's classic and highly influential description of Maranao. McKaughan (1958:18) notes that "The case-marking particles indicate the syntactic relations between any substantive phrase and the verb. The ... syntactic relations between the topic (always introduced by the particle *so*) and the verb are marked by verb inflection ...". One could simply replace the Maranao particle with the appropriate Tagalog particle and the statement could stand largely unamended in terms of the way in which such descriptions have been worded and elaborated.

The notion of focus in the description of Tagalog, and of Philippine languages in general, has been an interesting linguistic tradition. In fact, one must certainly agree with Constantino's (1971:118) observation that "the history of Philippine linguistics is largely the history of

the study of the major Philippine languages, especially Tagalog". The earliest descriptions were largely of Tagalog and many subsequent theoretical formulations in linguistics in one form or another were applied to Tagalog by reason of its status as one of the major languages of the Philippines as well as its status as one of the more accessible ones. By and large, my concern in this article has been with the past history and future treatment of the notions of topic, focus, and case in the Tagalog verbal paradigm, but by analogy many of the considerations can be extended to similar problems in other languages of the same verbal paradigm typology. This interplay has been a characteristic of past discussions, and there is little reason to discontinue this tradition - it is the nature of other traditions that are argued for or against in the following pages.

The earliest linguistic works on Tagalog, not including the work of the Spanish grammarians of several centuries past who recorded their observations on Tagalog after the Spanish conquest of the Philippines, date from the turn of the century with the work of Frank R. Blake and Leonard Bloomfield. Obviously, Bloomfield's name is familiar enough to most linguists, regardless of generation or intellectual inclinations. Blake's is not, and in some ways this is unfortunate, for Blake published articles on Tagalog and Philippine studies in general from the turn of the century until only several decades ago. Blake's use of familiar terminology in the description of the particular verbal construction network that characterises Tagalog and other languages of the area make Tagalog seem less exotic and its grammatical features not quite as dramatically different from other languages of familiar acquaintance. Thirdly, Blake's notion of the relationship between specific verbal affixes and construction types and the underlying case relationships of items in the sentence was often both more tolerant and more accurate than his contemporaries'.

It is true, however, that Blake was somewhat ambiguous in his treatment of the concept of case relationships and their expression in the surface structure of sentences. On the one hand, in an early article on the expression of case by the verb in Tagalog, Blake gives the impression that the "case-indicating function of the verb is developed to a high degree" in Philippine languages (Blake, 1906:183). Although Blake noticed the overlap between some forms and their corresponding verbal construction types, as for example, the fact that for him "the locative and ablative are expressed by the *an* passive" (Blake, 1960:185) and the triple overlap in the use of the *i*- construction (1906:188), this article leaves in general a somewhat different impression. The impression is that case relationships are expressed in the verbal inflections employed

in the verbal construction. Secondly, that the noun complement which appears as subject (*topic* is the more common term in current linguistic parlance for Philippine studies) appears as such because of its having been singled out for emphasis or as a focus of attention on the part of the speaker, thus underlining for the listener that nominal complement in the sentence which might be considered uppermost in the mind of the speaker.

Thirdly, one is enticed to the unwarranted conclusion that any given verb may be inflected for the various case relationships and the corresponding syntactic construction types by merely employing the appropriate verbal inflections and then re-arranging the nominal complements of the sentence to correspond with the particular verbal construction. There is a fourth implication, but one which is not overtly stated. This is that the verbal affixes in question only and always represent a single case relationship of the verb to the subject (*topic*) and a single syntactic construction type.

Here it may be best to let Blake speak for himself on the expression of case by Tagalog verbs.

"In Tagalog in a verbal sentence, that adjunct of the verb which is of most importance in the eyes of the speaker or writer is made the subject of the sentence, and the rest of the sentence is conformed to the character of this subject, the other adjuncts of the verb, which for the time being are of minor importance, having their case relations expressed by means of inflection. The verb might thus be said to express the case with emphasis; the various inflected forms, without emphasis. The sentence 'he looked for the book with the light in the room,' may be expressed in four different ways according as the agent, the object, the instrument, or the place, are specially emphasized.

If the idea is 'he, and no one else, was the one that did the looking,' the active of the verb would be used with the agent as subject, e.g., *siyá ang hungmánap nang libro nitó-ng ilaw sa silid.*

If the book is uppermost in the mind of the speaker or writer, the book, the object of the action, is made the subject of the *in* passive, e.g., *ang libro ay hinánap niyá nitó-ng ilaw sa silid.*

If the idea is that 'this light, and no other' was used, the light, the instrument of the search, stands as the subject of the *i* passive, e.g., *itó-ng ilaw ay ihinánap niyá nang libro sa silid.*

If the idea is that 'the room and no other place' is where the search was made, the room is made the subject of the *an* passive, e.g., *ang silid ay hinanápan niyá nang libro nitó-ng ilaw."*

One notes in passing Blake's use of *hungmanap* instead of *humanap*. Where Bloomfield used a single speaker for his 1917 work, Blake instead relied heavily on the work of the earlier Spanish grammarians, checking

out his assumptions with Tagalog texts and more current data. As a result, some of his lexical items and grammatical usage citations are rather dated, and Constantino (1971:125) has justly criticised him for this. On the other hand, since Blake's terms are often those used by the early Spanish grammarians, obviously extensions of European traditional grammar, the grammatical treatment is cast in a less exotic and more familiar mould. The point is that conflicting terminology is one sure way to insure that languages will look different and allow of little or no comparison. At least this feature is relatively absent in Blake's work, while in Bloomfield's description it was the keystone of the description, and in fact, part of the underlying philosophy of science Bloomfield and later descriptions brought to their investigations. The corollary was that each language was to be described in terms as unique entities, without reference to traditional models, European or otherwise.

Blake's assessment of case was more accurate in another article (1930) in which his analysis of the notion of case and the expression of case relationships in surface structure is rather akin to more modern versions of case grammar in syntactico-semantic discussions. In fact, Fillmore (1968) points out this article of Blake's as noteworthy and comments that some of his observations are inspired by or parallel to those of Blake. It is all the more poignant a commentary by Blake, for this point of view is opposed to some of the impressions which may have been generated by his own earlier work and Bloomfield's work in Tagalog. It does, however, present a better picture of what Blake must have had in mind, and raises considerations which are crucial to the consideration of focus, case, topic, and related topics in Philippine linguistic studies.

Curiously, in this publication, Blake provides a perfectly clear explanation of the notion of underlying case. The implication is that case as an underlying universal set of grammatical concepts is one thing, and the manner in which languages express various forms of surface relationships (also often called grammatical case in Indo-European) are different and need to be considered separately. Thus, Fillmore (1968:21) adopts the usage first provided by Blake in this (1930) discussion and uses:

"the term *case* to identify the underlying syntactic-semantic relationship, and the term *case form* to mean the expression of a case relationship in a particular language - whether through affixation, suppletion, use of clitic particles, or constraints on word order."

Even in earlier discussions of the case relations of the verb to the topic complement, Blake (1906:188) was careful to note that while such

cases did "not conform exactly in their scope to any of the cases commonly recognized in Indo-European grammar; sometimes two forms are used to express what is ordinarily considered one case, sometimes one form expresses two or more cases". Moreover, Blake was careful to point out that surface case expression was not limited to verbal forms alone. After all, the Tagalog sentential particles *ang*, *nang*, and *sa* are case-marking particles too. Blake's terms for their case functions are exemplified in his christening of them by the traditional terms *nominative*, *genitive*, and *oblique*. Compare Bloomfield's terms *subjective*, *disjunctive*, and *local*. The important thing to note here is that in assigning unique and unmistakable relations between the affixed verb and the topic complement, insufficient attention was paid to the obvious fact of ambiguity in the use of the sentential particles *ang*, *nang*, and *sa*. Certainly the function of *nang* and *sa* is highly ambiguous; indeed, Blake's choice of the term *oblique* for the particle *sa* attests to this. Thus, one rarely notes in the literature an attempt to unambiguously assign one and only one case function to *sa*, but one finds over and over again the attempt to assign, say, an instrumental case focus to the *i-* affix or an actor focus to the *-um-* affix.

Others, as for example, Scheerer's early (1905) description of the Nabaloi "dialect", had also noticed these characteristics of the Philippine verbal paradigm, but it was Bloomfield who contributed greatly to the differentiation of terminology and underlined the uniqueness of the syntax of Philippine languages. As has been pointed out elsewhere, it may have been that the underlining was too forceful. Blake (1919) in fact criticised Bloomfield for just this in his review of the latter's 1917 monograph *Tagalog Texts with Grammatical Analysis*. In at least five instances Blake criticises Bloomfield's introduction of "unfamiliar" and "unusual" terms and the disappointment of "obscure" explanations. At one point, an apparently exasperated Blake could not resist translating one of the longer and less transparent explanations by a single word, "verb". Note Bloomfield's definition followed by Blake's single word substitution in parentheses.

"These last categories are the following, viz., I) static and transient words, a transient being word expressing "an element of experience viewed as impermanent, i.e., belonging to some limited portion of time" (or somewhat more simply, a verb-F.R.B.)" (Blake, 1919:90).

Incidentally, Capell (1964) has also called attention to the difficulty "found in the delineation of verbal systems in the languages of the Philippines and similarly structured languages in neighboring areas". However, Capell's re-analysis of the verbal paradigm in Philippine languages is for quite different reasons and does not clear up the confusion



regarding the nature of focus, though he does call attention to the fact that there are varying degrees of focus and that these can be expressed in other ways.

Indeed, as McKaughan (1970:291) observes, Bloomfield found Philippine languages "so different that he avoided the use of familiar grammatical terminology to explain them in favor of new designations" in his analysis of Tagalog (1917) and Ilokano (1942). But it may be more to the point to say that he made them so very different by his avoiding the use of familiar terminology. McKaughan's other conclusion is indeed true to fact; in this latter feature, Bloomfield has been followed by most Philippine linguists up to the present time.

The net result of Bloomfield's influence was that his description inevitably became, as Constantino (1971:127) points out, the "model of the starting point of almost every modern descriptive study of any of the Philippine languages and dialects ... As a result of Bloomfield's influence, Philippine linguistics has become thoroughly descriptive and classificatory or taxonomic, and to a certain extent non-intuitive". If one may add to this last observation, non-intuitive to the degree that certain obvious considerations have been overlooked, in fact neglected, in the drive to make verbal constructions fit into a symmetrical network of focus or voice types, which unambiguously and uniformly comprises the verbal system of the language. However, as every student of Tagalog has learned, as soon as one moves from quoted examples into attempting to generate random sentences with verbal predicates, the process quickly moves from the realm of the systematic to trial and error.

Once again, one wonders how this could have been overlooked by those descriptions following Bloomfield's work. As Constantino (1971:139) notes, "many Philippine languages distinguish in the morphological shape of their verbs more than three passives". Ilokano, for example, which Bloomfield also worked on and published an outline (1942) of has at least five passives shown by the different affixes -en, i-, -an, i-an, and pag-. The singling out of only three passive constructions (signaled by -in-, i-, and -an) in Tagalog seems unreasonable in view of the fact that the closely related and highly parallel grammatical systems of the Philippine languages differ in the number and manner of their so-called passives.

The reason is obviously to be found in Bloomfield's position within the development of modern descriptive linguistics in the sense in which Kuhn (1970) outlines the development of theoretical paradigms in scientific disciplines. One need only speculate momentarily on the effect of this work on the course of Tagalog studies, and Philippine studies in general during the ensuing decades. It is only when the paradigm changes

and the new avenues of research are opened that new concerns in the description of Philippine languages begin to emerge. Actually, two separate and distinct paradigms, generative theory and case grammar, may be considered at this juncture as calling for different approaches to the treatment of Tagalog syntax.

Obviously, in line with the generative transformational press for an understanding of linguistic universals in the light of a universal theory of language, the uniqueness of the Philippine verbal paradigm no longer need be underscored, for we are told that languages will invariably differ in their surface manifestations. It is in their deep structure origins that we are assured that we can expect to find similarities. Methodologically, the nature of the syntactic theory of language that we bring to bear in our linguistic investigations will also posit parallels and universals in all the languages of the world, Philippine languages included.

Moreover, there is a crucial difference here, in that the switch from taxonomic observations to generation of grammatical sentences provides the impetus, in fact, necessity, for considering the Philippine verbal paradigm in new ways. One can no longer entertain the notion that the verbal construction types in Tagalog alternate with one another in completely predictable ways, as does the passive for active transitive verb sentences in English. Some descriptions have tried to coerce Tagalog verbal constructions into this mould, indicating that perhaps the active sentence (the actor-focus) verbal construction was the kernel from which other sentences of the goal-, locative-, or instrumental-focus type could be derived. This, of course, has been an implicit assumption, either pedagogical or descriptive, on the part of many descriptions ever since Blake and Bloomfield. However, as soon as the task turns from one of listing to rule formulation with the express concern of generating grammatical sentences and avoiding ungrammatical sentences, it becomes obvious that this position is untenable. However, it would appear that some descriptions cling to this notion, in spite of the obvious evidence to the contrary that this is simply not so (Buenaventura-Naylor, personal communication).

It may be that what has been termed focus is the syntactic manner in which Philippine languages keep track of new information introduced into the conversation or narrative. Buenaventura-Naylor (1973) hints at this in her dissertation, and it would appear that this line of investigation promises to be most rewarding.

What can focus be taken to mean then? One may disagree with Hidalgo's (1970:27) claims that focus is such that "a constituent is brought into sharp perspective so that the attention of the listener is drawn closer

to that constituent which is presumably in the speaker's mind. This element which is in sharp perspective, or which is in focus, we call "topic". However, one is reluctant to deny that focus equates with the question of emphasis entirely, for there are obviously ways of indicating emphasis which are not co-extensive with this syntactic phenomenon. One of these is the use of intonational features to call attention to or place emphasis upon a particular sentential complement. Certainly another method is the pre-position of the item singled out for such emphasis. For example, compare the following pairs:

Mayroon akong pera.	<i>'I have some money.'</i>
Ako'y may pera.	<i>'I'm the one who has some money.'</i>
Wala akong anak.	<i>'I don't have a child.'</i>
Ako'y walang anak.	<i>'I'm the one who doesn't have any children.'</i>
Nasa bahay ang dalaga.	<i>'The girl is in the house.'</i>
Ang dalaga ang nasa bahay.	<i>'The one in the house is the girl.'</i>

Here one is tempted to speculate on how such obviously mentalistic considerations could have possibly entered into a tradition conceived and engendered in a period of anti-mentalistic methodology and description. Such descriptions employing concepts like emphasis, focus of attention, and so forth, are patently mentalistic, and the labelling of the verbal construction types by the focus labels clearly relates to the attempt to label the semantic content of topic complements and their functions.

In the past, most analyses have treated this nominal complement as standing in a special relationship with the verbal predicate and have termed this nominal complement *topic* in place of *subject*. Moreover, those verbal constructions which were the result of the limited number of verbal affixes which were thought to correspond to particular syntactic arrangements and specially related topics were consequently termed focus construction types and the corresponding affixes focus affixes. The traditional view, of course, had attempted to portray that special relationship in largely semantic terms, and this is reflected in the names given to the construction types. For example, actor-focus constructions were constructions in which the actor or agent was the topic; in goal-focus constructions, it was the goal or object which was taken as topic; in locative constructions the location, direction, or person to or from which the action was oriented; in instrument-focus constructions, the instrument or means by which the action was performed; in benefactive-constructions, the beneficiary for whom the action is performed. When even such broad labels did not seem to cover the semantic

range of the items covered in the topic positions, some resorted to broader labels which would attempt to cover the category of lexical items which fit into the topic position more fairly, if not as neatly. Such considerations obviously underlie the use of terms like referent-focus, accessory-focus, and implicative-focus by some descriptions.

It would have seemed that a simple cross-classificational listing along the lines suggested by Chomsky's *Aspects* (1965) treatment would have been one way to proceed with the problem at hand. This approach would have required that verbs would have been marked for some kind of a *focus index*, as suggested by Kess (1967) and Buenaventura-Naylor (1973), which gave the information appropriate to triggering transformations which would have led to the formulation of grammatical strings for each verb in the lexicon. Actually, this focus index was to have included three kinds of features: focus affix, subcategorisational restrictions, and semantic features. Thus, each verb would have listed for it those verbal focus affixes which can co-occur with it. Secondly, it is necessary to include restrictions similar to the subcategorisational restrictions suggested by Chomsky (1965), for it is obvious that verbal affixes are not an entirely reliable clue to the privileges of occurrence of various nominal complements in the sentence. For example, while the -um- infix has been taken to indicate the so-called actor-focus, it does not unambiguously provide *a priori* information as the number, type, and semantic relationship of the various nominal complements which may share the sentence with it. For example, compare the roots *ulan* 'to rain' and *bili* 'to buy' in their appearance with -um-.

Umulan sa Maynila.

*'It rained in Manila.'*

Bumili siya nang bigas sa tindahan  
para sa nanay.

*'He bought rice at the store for  
his mother.'*

Other semantic counter-evidence to the case-marking abilities of such verbal affixes have been noted in stative or ambient sentences of a meteorological nature. For example, in sentences like the following, there is no suggestion of what could possibly be the actor of such actor-focus sentences (the -um- infix has been taken as one of the primary markers of actor-focus sentences).

Umulan kahapon sa Maynila.

*'It rained yesterday in Manila.'*

Humangin nang malakas.

*'The wind blew strongly.'*

Lumindol.

*'There was an earthquake.'*

Moreover, in a sentence like *Uuwi sana ang binata pero umulan* 'The bachelow was about to go home but it rained' the two verbs *uwi* 'to go home' and *ulan* 'to rain' do not have the same topic complement although

they occur in the same sentence construction and are marked by a form of the -um- infix. One intuitively knows that no deletion rules removing identical subjects here has applied, for *ulan* has none.

It is intriguing that a school of linguistic thought which had its parallels in the behaviorism which overtook the psychology of the period never made any attempt to ascertain whether focus was in fact emphasis, focus of attention, or whatever, on the basis of behavioral evidence from experimental investigation. It is true that a fair number of more recent psycho-linguistic investigations have pointed out that the relationship between the active and passive in English is not meaning-preserving and that the active is not equated with the passive in terms of performance variables by speakers of the language. Moreover, some of the studies suggest that the function of the passive is in fact that of highlighting or emphasising the logical object, now moved to grammatical subject position and initial in the sentence.

There is some indirect support for this interpretation of focus in psycholinguistic investigations of differences for the active and passive sentence types in English. Such investigations have shown that sentence types like passive may be more appropriate to some contexts than others, being used in performance to call attention to or to emphasise certain topics rather than others. Turner and Rommetveit (1968) discovered that the function of the English passive is to place emphasis on the object of the action, thereby taking similar emphasis away from the logical (but not grammatical) subject of the sentence. As recall prompts they employed pictures of situations, and found that passives were elicited by showing pictures of the objects in such situations. Conversely, active sentences would be elicited by pictures of the agentive subject, as would pictures in which the entire situation was represented. This is congruent with Johnson-Laird's (1968) suggestion that the choice of the passive voice is directly related to placing emphasis on the underlying object by the overt means of having it appear in the sentence-first position, as happens with logical objects in the passive frames.

Other studies have reported similar findings. Clark (1965) found that subjects provided different responses to active and passive sentence types when required to fill in words missing in the two syntactic patterns. Animate nouns apparently figure more importantly in passive sentences in which the grammatical subject (but logical object) position is the slot to be filled. These responses were almost double the number given for transitive verb active sentences in which the grammatical object (still logical object) was to be filled in. Similarly, Johnson (1967) also found differences for the active and passive sentence types

in an experiment on syntactic position and meaning rated according to the semantic differential technique. Subject and object functions in passive sentence types were evaluated as more similar in terms of their semantic differential ratings than the corresponding subject and object function slots in active sentence types. Olson and Filby (1972) found differences in the comprehension of active and passive sentences under a variety of circumstances. These experiments unfortunately do not tell exactly what the differences between active and passive sentence types are, but they do graphically underline the fact that differences between the two sentence types do exist and that they cannot be taken as equivalent and meaning-preserving in any real sense. What then of the various Tagalog passives?

There is also striking evidence from Wason's (1961, 1965) psycholinguistic work with the negative in which he finds the "context of plausible denial" the most likely and appropriate function of negative constructions in English. It is likely that some such similar function may turn out to be true for Tagalog passives, or non-actor-focus-constructions, but to my knowledge this has not been adequately surveyed, and certainly was not surveyed during this period, in spite of claims that this was how speakers operated and that this was how such sentences functioned.

Under the previous kind of approach, it would have even made perfectly good sense to investigate the selectional feature restrictions of the topic complement to the verbal predicate. However, with the introduction of a fresh new approach of case grammar, entirely new perspectives are possible, and they do in fact present the researcher with perspectives that are more likely to reflect the case-marking situation in Tagalog verbal constructions. Actually, some of this was foreshadowed in the presentation of the dichotomy between deep structure and surface structure, but not with sufficient departure from the notion of syntax as central. It is only with Fillmore (1968, 1971), Chafe (1970), and Bever (1970, 1971) and their approach to the concept of meaning as the central point of departure in language that competent description of case becomes possible. This paper restricts itself to Fillmore's discussion of case, for it touches in essence most closely to the problem at hand. Moreover, it also touches terminologically closely to the problem at hand and the way in which it has been treated and labelled since the turn of the century.

Thus, focus is now construed as entirely a surface structure phenomenon. The relationship of the topic complement is a special one, but not one defined in unambiguous semantic terms or in unambiguous surface case grammar terms. The relationship is as special as the fact that

sentences in Tagalog have a subject or topic complement, and different kinds of verbal constructions correlate highly with particular syntactic arrangements. But nothing more can be made of this fact. One must, nevertheless, continue to mark focus potential (we may as well keep the term focus, and now use it in the sense of a particular set of verbal affixes which correlates significantly with syntactic arrangements for verb roots). Why? Because these affixes as marked for verbs will give some clue as to which transformational rules will apply in the grammar and which will not. For example, a verb marked for -um-, mag-, -in, and i- will trigger just those transformational rules which will provide strings generated by those rules. Or if one operates on the insertion principle, then verbs listed with appropriate features in the lexicon may appear in certain syntactic slots.

The key difference now is that such verbs must also be marked for their case propositions. In specific, one must mark verbs for the particular case relationship that the topic complement does bear to the verb in question. The case functions will obviously overlap with the surface manifestations in some sentences, and differ in others; this is no longer troublesome, for we have already dismissed these as being totally reliable guides as to which case relationships verbs do in fact take as their constellation of case relationships. A complete analysis would thus include both the surface structure information discussed above as well as a complete semantic description of the possible cases that a given verb will admit. Obviously, the two are separate concerns.

Various strategies have been taken to remedy the problem posed by either verb classification or by a generative treatment of the Philippine verb. The most common strategy has been some overlap of the taxonomic aspects of listing and verb classification and the problem of generation. Some descriptions have taken the verbal affixes as the basis of the verbal classification, acknowledging that verbs may admit of different behaviours in respect to their occurrence with the various affixes chosen by each verb. However, the notion often implicit in such descriptions is that the verb itself, with its voice affix, is in itself the key to the remainder of the syntactic construction. The work done by Miller (1964) for Mamanwa, Kerr (1965) for Cotobato Manobo, and Llamzon (1966) for Tagalog center around investigation of such verbal affixes.

Miller's and Kerr's results both provide the basis of verb classification, but are relatively weak generational devices and do not offer clear statements of the case relationships of the case structure of the verb. Llamzon's investigation of the verbal affixes in Tagalog is, like Kerr's and Miller's results, most instructive of the *focus potential* of

the verb. Llamzon took Bloomfield's 397 transients (verbs), listed in his lexical index, and tried them out for a battery of seven such affixes (-um-, mag-, man-, -in, i-, -an, and ma-).

There are several interesting things to note about Llamzon's findings for the verb and the affixes. First, one notes that there is overlap between the affixes in several ways. There are several actor-focus affixes, -um-, mag-, and man-, with no apparent single clear-cut boundary able to be drawn between them. The ma- affix would appear to go both ways, being at times what looks like an actor-focus (*Natutulog ang bata 'The child is sleeping'*) and at other times what looks like a goal-focus (*Nakuha niya ang bata 'He took the child.'*).

Secondly, it was hard to find clear-cut unambiguous semantic characterisations for the noun phrases which occurred as topic complements for the verbal constructions which were the result of such verbal affixation. This would appear to confirm the fact that the notion of focus discussed previously as having been commonly used was a handy mnemonic device, by no means covering the entire array of phrases which occurred as topic complement with a specific verbal affix. Thirdly, if there is a single concept which can be construed as actor-focus, why more than one actor-focus construction? One could at least have hoped for some kind of complementary distribution between the various actor-focus affixes, -um-, mag-, and man- (or mang-), such that where one occurred the other two did not, but such is far from the case.

It is obvious that the 397 verbs vary greatly in the co-occurrence with just these seven affixes. There is no *a priori* way of knowing exactly and accurately which verbs will co-occur with which affixes. The enterprise is one in which *a posteriori* information appropriately labelled and marked in the lexicon entry for each verb is the only guarantee to the generation of grammatical sentences in Tagalog or other Philippine languages. For example, Llamzon found that of the 397 verbs, only 305 could occur with -um-, 51 with man- (mang-), 331 with ma-, and so forth.

On the other hand, another popular strategy of the last decade has been to investigate what Chomsky (1965) called sub-categorisational restriction, that is, which noun phrase complements are obligatory for certain verbs. Some work has also been done on which noun phrase complements are optional to certain verbs, thus providing a configuration of sentential complements which are then to be marked as optional or obligatory for the verbs in the lexicon. Thus, for example, see Kess (1967) and Constantino (1965). Forster and Barnard (1968) have provided a classification of Dibatawon verbs on the basis of their occurrence with obligatory situational slots like actor, goal, instrument, and



site. Reid (1966) employs the tagmemic criteria of potential clause expansion and of nuclear grammatical slots with the possibility of clause transformation. However, it should be noted that any attempt at verb classification in a purely taxonomic fashion is no guarantee of generativity beyond the sample class, and can only be valuable when a by-product of the information *a posteriori* cross-classificationally marked for verbs is in the lexicon entries. The sequence of information is clear, the latter is a primary consideration and is first and foremost.

Constantino (1965) has provided just such a valuable constituent analysis and transformational analysis of the major sentence patterns of twenty-six Philippine languages. Interestingly, his approach includes the specification of complements which may occur with verbs in sentences containing verbal predicates. For example, Constantino notes that active verbs are to be subdivided into six subclasses on the basis according to which complements occur with them obligatorily or optionally. For active verbs alone, Constantino considered seven different kinds of complement: the indefinite goal complement, the definite goal complement, the locative complement, the benefactive complement, the instrumental complement, the reciprocal actor complement, and the agentive complement.

Here it should be pointed out that an amalgamation of the two concerns is essential. Just as a listing of what has been termed focus potential is insufficient, so also is a simple notational listing of the various nominal complements that various verb roots may take. It is possible to link the two up, and in fact, this is likely the best approach to employ. Obviously, one cannot expect a focus affix by itself to be predictive of the syntactic construction's optional or obligatory nominal complements accompanying the verb. For example, we have already noted the syntactic constructional possibilities the actor-focus *-um-* infix may participate in. For example, contrast the following sentences:

Umulan.	<i>It rained.</i>
Bumili siya nang bigas.	<i>She sold rice.</i>
Tumaba si Maria.	<i>Mary got chubby.</i>

Or for example, contrast the following sentences with the actor-focus *nag-* prefix (from *mag-*):

Naglinis siya nang sapatos.	<i>He cleaned the shoes.</i>
Nagtiis sila.	<i>They suffered.</i>
Naglagay siya nang tubig sa baso. (Obligatory locative complement)	<i>He put the water in the glass.</i>

Thus, verbs will have to be marked for focus affixes as well as for those nominal complements which appear obligatorily or optionally in the syntactic construction in which the resultant affixed verbal construction appears. See Kess (1967), Buenaventura-Naylor (1973), or Ramos (1973) for suggestions as to how various aspects of the problem may be projected.

This still does not provide us with a sufficient basis to freely generate grammatical sentences containing verbal predicates beyond those accounted for in the lexicon accompanying the description. This is still a basic inescapable problem and one that is crucial to the description of Philippine languages, if we are to move beyond the stage we now find ourselves at. McKaughan's (1971) rhetorical question as to whether the verbs can be grouped by semantic criteria is one possible line of approach. However, this line of approach must be clearly defined as to what is semantic and what is not? What is grammatical and what is not? What is focus and what is not? What is the relationship of the topic to the verb and what is not? First, it must be assumed that Tagalog verbs must be marked for a variety of surface syntactic features in the manner described in the preceding pages. Secondly, one must now incorporate semantic considerations in the description of the Philippine verbal paradigm, and more than likely, in a way which is not necessarily directly related to surface features in the syntax of Philippine verbs. At least several approaches have suggested themselves as being capable of providing both the descriptive machinery and philosophical basis to handle the problem effectively.

Here it may be best to survey two of these approaches and what their contributions, potential and actual, may be construed as their resolution of the stalemate described in the previous paragraphs. The first of these is the method of incorporating semantics into generative theory presented by Katz and Fodor (1963), Katz and Postal (1964), and institutionalised by Chomsky (1965). The second of these approaches is case grammar as a means of answering fundamental questions of the case relationships inherent and unchanging in related sentences containing verbal constructions as well as to a description of verbs which allows both generativity and predictability of both syntactic and semantic relationships. Other similar approaches will be bypassed for obvious reasons of space and their lack of overt clearly defined representatives in Philippine verbal formulations as yet. It should be sufficient to say that the parallel thread which runs through Fillmore's argument as well as Chafe's, Bever's, and others', is the primacy of semantics or whatever one wishes to call that level of language which underlies the conceptual framework which language is but one reflection of.

The third kind of approach which has been taken up to this point has been one in which selectional features have been charted for the several part of speech categories. By and large, this has been restricted to the part of speech categories of noun and verb in Philippine studies, and in fact, usually even more restricted to the selectional restrictions which operate between the noun head in the nominal complement taken as topic in the grammatical sense and the verb of the major verbal predicate of the sentence. For example, one notes the collocational privileges apparent in sentences like the first pair and the collocational privileges apparently violated in the second pair of sentences.

Namutla ang pasyente.

*The patient got pale.*

Nalungkot ang dalaga.

*The girl became sad.*

\*Namutla ang hangin.

*\*The wind got pale.*

\*Nalungkot ang basura.

*\*The rubbish became lonely.*

The question of the place of metaphor, puns, modern prose, beat poetry of the Ginsberg and Ferlinghetti type, and modern rock music genres notwithstanding, the obvious way to treat such facts are to assign certain features to one or the other part of speech class, noun or verb, and then assign commensurate, potentially matching features to the other part of speech class. What has most often been done is to assign such semantic or lexical features to the noun and have the collocational relationships of other parts of speech classes, like verbs and adjectives, be defined in respect to these inherent noun features. Though this approach borders on the notion of underlying semantic properties, in essence it is purely an attempt to account for the surface properties of certain classes of grammatical sentences as opposed to other kinds of ungrammatical sentences. In the development of generative theory in the last decade, it was of course the way in which the semantic level was both introduced into a theory of language and the way in which the domain of semantics was defined to make it amenable to the kind of linguistic theory then available.

It would appear that to the degree that case grammar formulations, the fourth approach, are fruitful in their application to the Philippine verbal paradigm, the information provided by the case relationship of a nominal complement to the verb may in effect answer any necessary questions about the semantic feature surface relationship of the noun and verb involved as well. If, as Ramos (1973:30) observes in her discussion of those cases conceptually inherent to the basic sense of the Tagalog verb, "the roles actants may perform are grouped in Tagalog into five inherent case relationships in the deep structure: the agentive case (A), the objective case (O), the directional case (Dir), the

locative case (L) and the instrumental case (I)", then one can expect some degree of semantic predictability as to the semantic status of the nouns which fill a given role with a given verb. Indications from Ramos' work and others point to this, but this is only offered here as a suggestion of possible advantageous by-products of a case-like approach to the Tagalog verb.

This fourth and last approach is derived from Fillmore's recent arguments (1968, 1971) for the inclusion of the grammatical notion case in the base component of the grammar of every language. Fillmore sees such case relationships as primitive terms in the grammatical theory of language even though such categories may in fact be covert categories, not necessarily expressed in the surface structure. In fact, he notes that any attempt to restrict the notion of case to surface structure alone must fail. Here it may be added that the attempt to handle surface structure in Tagalog as the only kind of possible case grammar has resulted in the most counter-intuitive descriptions and a convoluted type of logical taxonomy which has coerced constructions into a tidy set of classes, but at great intuitive cost. The question of case in Tagalog is no different than that in Latin; in its underlying conceptualisation, the various cases should be universal relationships possibly expressed in different ways by the surface structures of various languages. Thus, it is not at all surprising to find that Latin employs inflectional endings (traditionally referred to as case, but different from Fillmore's underlying cases), nor is it surprising to find that Tagalog has verbal predicates with syntactic complements, with verbal predicates exhibiting certain surface relationships between the verb and its topic complement (traditionally referred to as focus-marking or case-marking relationships, but again different from Fillmore's sense of underlying cases).

Fillmore's analysis (1968:21) suggests that "the sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relationship". The implication is that the various permitted arrays of distinct cases occurring in simple sentences may express a notion of sentence type that may possibly have universal validity, with the result a classification of verbs in a language which again may have some universal validity. The crucial consideration is that one is here considering underlying cases, not the various types of mechanisms by which surface case is, was, or may be shown in languages. As soon as this is the important consideration, languages may be allowed to become as different as is conceivably possible by the surface mechanisms available to language in the sense noted by Greenberg's (1966) statistical inventories.

Such semantic primacy positions, like those of Fillmore and Chafe, have adopted the position of the verb being in some way central or primary; this is in contrast to selectional feature notations where nouns were described by means of a small number of inherent features and other part of speech classes, like verbs and adjectives, were defined in terms of their collocational possibilities. This shift in position is reflected in Ramos' (1973:23) case grammar treatment of the Tagalog verb in which it is assumed that "the verb is central to the Tagalog sentence. Nouns are peripheral and tied to the verb center by relations such as agentive, objective, instrumental, directional, locative, and the like".

Fillmore (1968:24-5) has recognised at least the following case notions as a set of possible universal concepts to be designated for inclusion in the grammar. The cases are as follows: agentive (A), the animate instigator of the action identified by the verb; instrumental (I), the inanimate force or object casually involved in the action or state identified by the verb; dative (D), the animate being affected by the state or action identified by the verb; factitive (F), the object or being resulting from the action or state identified by the verb, or understood as a part of the meaning of the verb; locative (L), the location or spatial orientation of the state or action identified by the verb; and the objective (O), the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself. Other cases, Fillmore implies, may have to be added, and this is in fact what does happen in later expansions of case grammar. See, for example, Fillmore's later discussion cum expansion of the same topic (1971), in which several new cases are added and several terminological changes are to be noted.

What Fillmore's approach suggests is that the focus paradigm of Philippine verbal constructions is best treated as a surface structure manifestation which may or may not provide clues as to the underlying case relationships. In some cases, it apparently does, as for example, in Fillmore's illustration from one of McKaughan's Maranao examples in which the focus construction types and their corresponding verbs in this particular instance do appear to overlap nicely with the cases in question. In this instance, the traditional focus or voice or case-marking labels given the several verbal constructions appear to fit the actual underlying cases nicely, and better than that, appear to give a correct indication of what the case relationship of topic to verb happens to be. However, it may be asking too much to ask this of all such verbal construction types, as the following examples from Ramos' (1973) Tagalog discussion indicate:

Itago mo ang bola.

*Keep the ball.*

Ibili mo siya nang kotse.

*Buy a car for him.*

Ipunas mo ang trapo.

*Use the rag to wipe it.*

Here is an instance of a single verbal affix, a so-called focus-marking affix in some descriptions and a so-called case-marking affix in other descriptions, marking three separate and distinct grammatical relations between the verbal construction and the nominal complement in the topic position. The *i-* prefix has been used in sentences in which the object, the beneficiary, and the instrument respectively appear at the topic complement.

Schachter (1961) noticed this early on, pointing out that the same sentence *Binayaran ng lalaki ang alipin* had two different readings, stemming from two different derivational histories. On the one hand, the sentence can read as '*The man paid for the slave*'; on the other, the sentence reads as '*The man paid the slave*'. The first has *alipin* as the object in the topic position, the second reading has *alipin* as indirect-object in the topic position. Thus, while it is true that underlying every topic phrase in verbal predicate sentences there must be a case distinction of some sort, one cannot agree with McKaughan and others in saying that "that case distinction is overtly carried by an affix in the verb" (McKaughan, 1970:295).

Moreover, there is also the question of sentences without verbal predicates. Such sentences still have grammatical subjects marked by *ang* and sentential syntactic position. Such sentences offer a variety of non-topic sentential complements, none of which can be said to overtly and unambiguously mark the topic. But such sentences nevertheless do have a topic which may stand in a particular case relationship the rest of the sentence. Consider for example the following:

Ang dalaga ang maganda.

*It is the girl who is pretty.*

Ang bulaklak ang para sa dalaga.

*It is the flower which is for the girl.*

Sa Maynila ang bahay.

*The house is in Manila.*

Binata ang abogado.

*The lawyer is a bachelor.*

May pera ba ang titser?

*Does the teacher have any money?*

Here it should be noted that in some quarters the characterisation of the verbal paradigm is changing in just this direction, the direction of case grammar formulations. As has been pointed out by Kess (1975), one cannot help but be struck by the changing semantics of focus, such that previous characterisations of the concept of focus seem to be in the process of being rethought and more and more influence of the case grammar and/or primacy of semantics approach is to be seen. For example, see the recent work published by Miller and Miller (1973) for

Mamanwa (in contrast to Miller (1964)), West (1973) for Amganad Ifugao, and others.

McKaughan's (1971) rhetorical question as to whether there may yet be some key to the Philippine verbal paradigm may in some ways be answered by the application of case grammar notions to the problem. However, the answer is of a larger order of magnitude, for if we answer such questions we really answer such questions for the human verbal paradigm repertoire. Take, for example, Ramos' (1973:25) observation that "the role types are themselves unanalyzables corresponding to elementary perceptions on the part of human beings concerning matters relative to an action. The roles include relations to actions such as who did it, who experienced it, where it happened, what the result was, and a few others. A number of these role notions may be universal, and can be grouped in each language into a limited number of cases, namely, agent, instrument, object, directional, location, and so forth". I would disagree with the observation that such role notions may be universal; rather, they *must* be construed as universal if the theory is to be meaningful in any linguistically universal sense. The point is that if we solve the problem of roles and verbal case-argument constellations for a given language or closely related set of languages, we should have solved the core of the problem for all languages. The problem is now one of validating what we have found, or rather what we have hypothesised, for languages. The evaluation and validation problem now hinges on the seemingly unanswerable question which underlies the basic rift between the current revival of rationalism in certain sciences like linguistics and empiricism. The very means by which results of the former can be proved is ruled out by the philosophical foundations of the former and by the chasm separating what is taken as legitimate concerns and the data exhibiting or supporting those concerns by rationalistic speculation and empiricistic verification. For a lengthier discussion of the basic psycholinguistic problem of the relationship between linguistic theory and psycholinguistic investigations directed at testing the psychological validity of such linguistic constructs, see Kess (1976).

Thus, McKaughan's (1971) suggestion that there may yet be a key to the Philippine verb's full paradigm, something like the four principal parts for Latin verbs in classical studies, has not to date been fulfilled and does not appear to be immediately answerable by the approaches we have taken up to this time. However, what is essential is that what we have accomplished by way of investigation of the verbs is appropriately charted. This paper suggests using the terms *focus* only to refer to the type of focus-marking affix seen in the discussion up to this

point and *case* to relate to the actual underlying case relationship of complements to the verbal predicate. This is a necessary departure from the previous view of focus often held, but essential to an adequate description of the case-relationship of verb to topic. Moreover, the use of the affixes as listed in the lexicon entry for each of the verbs may be taken as only suggestive of, but not entirely reliable and predictive of the appearance of syntactic complements that may appear in the same sentence frame with the verbal predicate and its topic complement. Such other sentential complements, like the locative complement, the benefactive complement, the objective complement (specific and non-specific), and so forth, would have to be specified as to their obligatory or optional status with each verb in the lexicon.

The question is now one of what we have learned and we shall do with the information we have amassed in respect to the nature of the verbal paradigm. It is apparent that a complete syntactico-semantic description of the Tagalog verb must include each of the four types of information discussed in the paper. Each type of information, incidentally, almost coincides with different periods of theory development in linguistic science theory, and each is valuable and essential to a complete understanding of how the Tagalog verb operates and how it may be generated and how it fits in with linguistic universal constructs.

It should suffice to say that a complete description must at least include generative information regarding the co-occurrence potential of such so-called focus affixes, the co-occurrence potential of the various syntactic complements in sentences with each of the verbal constructions resulting from the affixation of such verbal affixes, the selectional feature restrictions between the verb and the topic complement (as well as the other syntactic complements), and lastly, a statement of the actual underlying case relationship between the noun phrase standing as topic in relation to a specific verbal predicate.

Moreover, it might be best if descriptions of the verbal paradigm restrict themselves to the psycholinguistic evidence at hand, before phrasing descriptions which have profound ramifications which cannot be supported by the present state of our knowledge. This also implies that continuing psycholinguistic inquiry into such problems is much to be desired and that this field of endeavour is one which will likely provide us with a metric device for evaluating some aspects of linguistic description and possibly for choosing one portrait of the verbal paradigm over another.



BIBLIOGRAPHY

BEVER, T.

- 1970 'The Cognitive Basis for Linguistic Structures'. In: J.R. Hayes, ed. *Cognition and the Development of Language*. New York: John Wiley.
- 1971 'The Integrated Study of Language'. In: J. Morton, ed. *Biological and Social Factors in Psycholinguistics*. London: Logos Press.

BLAKE, Frank R.

- 1906 'Expression of Case by the Verb in Tagalog'. *Journal of the American Oriental Society* 27:183-9.
- 1919 'Review of *Tagalog Texts with Grammatical Analysis*, by L. Bloomfield'. *AJPh* 4:86-93.
- 1930 'A Semantic Analysis of Case'. *Curme Volume of Linguistic Studies*, ed. by James T. Hatfield, et al. 34-49. Baltimore: Waverly Press.

BLOOMFIELD, Leonard

- 1917 *Tagalog Texts with Grammatical Analysis*. University of Illinois, Studies in Language and Literature 3:2-4. Urbana, Illinois: University of Illinois.
- 1942 'Outline of Ilocano Syntax'. *Lang.* 193-200.

BUENAVENTURA-NAYLOR, P.

- 1973 Topic, Focus, and Emphasis in the Tagalog Verbal Clause. Unpublished Ph.D. dissertation, University of Michigan.

## CAPELL, Arthur

- 1964 'Verbal Systems in Philippine Languages'. *Philippine Journal of Science* 93:231-49.

## CHAFE, Wallace L.

- 1970 *Meaning and the Structure of Language*. Chicago: The University of Chicago Press.

## CHOMSKY, Noam

- 1965 *Aspects of the Theory of Syntax*. Cambridge: M.I.T. Press.

## CLARK, H.H.

- 1965 'Some Structural Properties of Simple Active and Passive Sentences'. *J. Verbal Learn. Verbal Behav.* 4:365-70.

## CONSTANTINO, Ernesto

- 1965 'The Sentence Patterns of Twenty-six Philippine Languages'. *Lingua* 15:71-124.
- 1971 'Tagalog and Other Major Languages of the Philippines'. *Current Trends in Linguistics*, ed. by Thomas A. Sebeok, *Linguistics in Oceania*, vol. VIII. The Hague: Mouton.

## FILLMORE, Charles J.

- 1968 'The Case for Case'. *Universals in Linguistic Theory*. Edited by Emmon Bach and Robert T. Harms. 1-88. New York: Holt, Rinehart and Winston, Inc.
- 1971 'Types of Lexical Information'. In: D.D. Steinberg and L.A. Jakobovits, eds *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge: Cambridge University Press.

## FORSTER, J., and M.L. BARNARD

- 1968 'A Classification of Dibabawon Active Verbs'. *Lingua* 20:265-278.

## GREENBERG, J., ed.

- 1966 *Universals of Language*. 2nd edn. Cambridge, Mass.: M.I.T. Press.

## HIDALGO, A.C.

- 1970 'Focus in Philippine Languages'. *Philippine J. Linguistics* 1:25-32.

JOHNSON, M.G.

- 1967 'Syntactic Position and Rated Meaning'. In: L.A. Jakobovits and M.S. Miron, eds *Readings in the Psychology of Language*. Englewood Cliffs, N.J.: Prentice-Hall.

JOHNSON-LAIRD, P.

- 1968 'The Choice of the Passive Voice in a Communicative Task'. *Brit. J. Psychol.* 59:7-15.

KATZ, J.J., and J.A. FODOR

- 1963 'The Structure of a Semantic Theory'. *Lang.* 39:170-210.

KATZ, J.J., and P. POSTAL

- 1964 *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: M.I.T. Press.

KERR, Harland B.

- 1965 'The Case-Marking and Classifying Function of Cotabato Manobo Voice Affixes'. *Oceanic Linguistics* 4:15-47.

KESS, Joseph Francis

- 1967 Syntactic Features of Tagalog Verbs. Unpublished Ph.D. dissertation, University of Hawaii.
- 1972 'The Utility of Focus'. *Kivung* 5/3:184-91.
- 1975 'On the Semantics of Focus'. *Anthropological Linguistics* 17:353-62.
- 1976 *Psycholinguistics*. New York: Academic Press.
- 1976 'Reconsidering the Notion of Focus in the Description of Tagalog'. In: N.D. Liem, ed. *South-East Asian Linguistic Studies*, vol. 2. PL, C42.

KUHN, T.S.

- 1970 *The Structure of Scientific Revolutions*. 2nd edn. Chicago: University of Chicago Press.

LLAMZON, T.A.

- 1966 'Main Transient Formations in Tagalog'. *Phil. J. Sci.* 143-57.

MILLER, H.

- 1973 'Thematization in Mamanwa'. *Linguistics* 110:55-74.

## MILLER, J.

- 1964 'The Role of Verb Stems in the Mamanwa Kernel Verbal'.  
*Oceanic Linguistics* 3:87-100.
- 1973 'Semantic Structure of Mamanwa Verbs'. *Linguistics* 110:74-82.

## McKAUGHAN, Howard P.

- 1958 *The Inflection and Syntax of Maranao Verbs*. (Publications of the Institute of National Language.) Manila: Bureau of Printing.
- 1962 'Overt Relation Markers in Maranao'. *Lang.* 38/1:47-51.
- 1970 'Topicalization in Maranao - an Addendum'. In: S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic Studies in Honour of Arthur Capell*. PL, C13. 291-300.
- 1971 'Minor Languages of the Philippines'. In: Thomas A. Sebeok, ed. *Current Trends in Linguistics*. Vol.VIII. *Linguistics in Oceania*. The Hague: Mouton.

## OLSON, D., and N. FILBY

- 1972 'On the Comprehension of Active and Passive Sentences'.  
*Cogn. Psychol.* 3:361-81.

## RAMOS, T.V.

- 1974 *The Case System of Tagalog Verbs*. PL, B27.

## REID, Lawrence A.

- 1966 *An Ivatan Syntax*. (*Oceanic Linguistics* Special Publication No.2.) University of Hawaii Press.

## SCHACHTER, Paul

- 1961 'Structural Ambiguity in Tagalog'. *Language Learning* 11:135-145.

## SCHEERER, O.

- 1905 *The Nabaloi Dialect*. Manila: Bureau of Public Printing.

## THOMAS, D.D.

- 1958 'Mansaka Sentence and Sub-sentence Structures'. *Philippine Social Sciences and Humanities Review* 23:330-58.

TURNER, E., and R. ROMMETVEIT

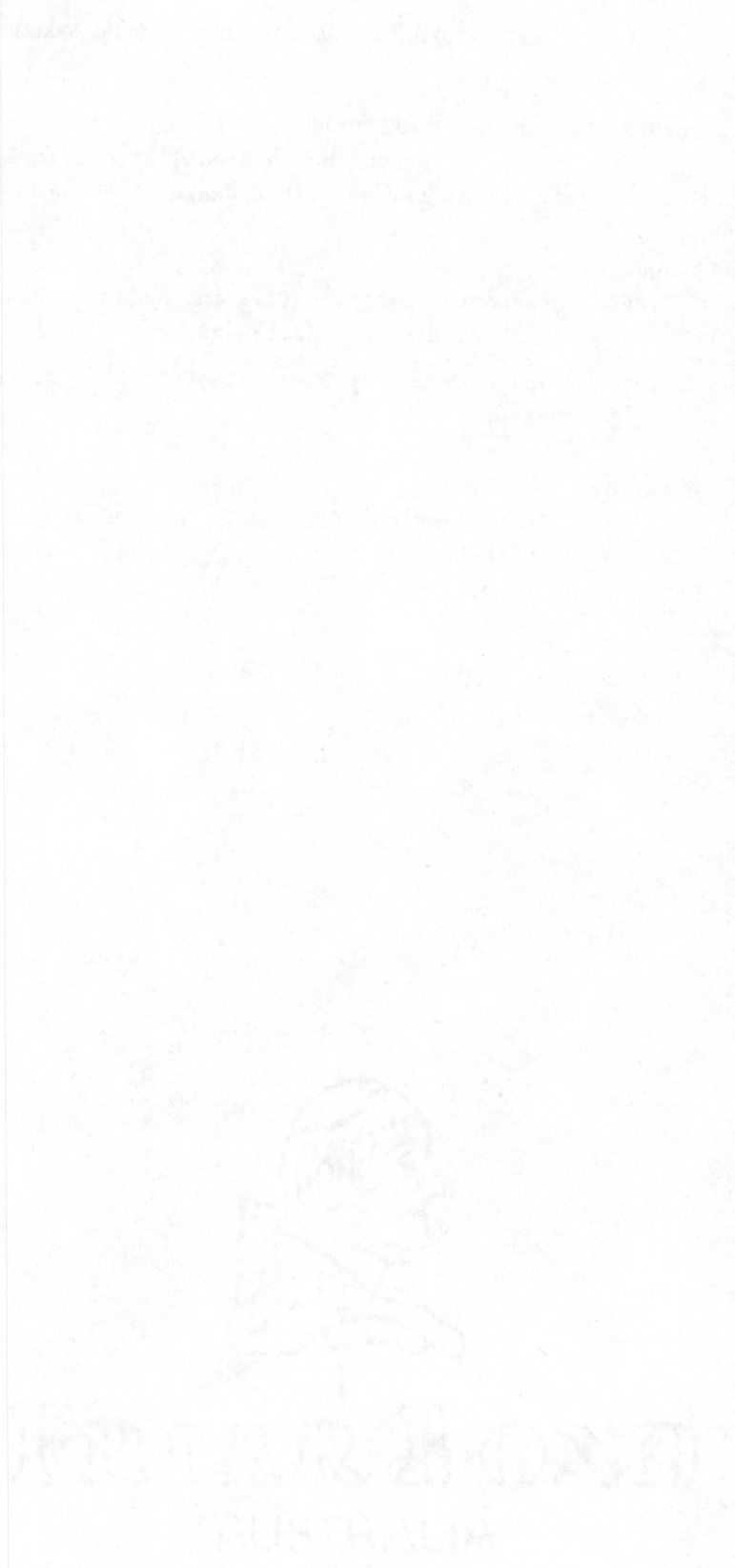
- 1968 'Focus of Attention in Recall of Active and Passive Sentences'.  
*J. Verbal Learn. Verbal Behav.* 7:543-8.

WASON, P.

- 1961 'Response to Affirmative and Negative Binary Statements'.  
*Brit. J. Psychol.* 52:133-42.
- 1965 'The Contexts of Plausible Denial'. *J. Verbal Learn. Verbal Behav.* 4:7-11.

WEST, A.

- 1973 'The Semantics of Focus in Amganad Ifugao'. *Linguistics*  
110:98-122.



## ON THE DEVELOPMENT OF CONTRASTIVE WORD ACCENT: PANGASINAN, A CASE IN POINT

R. DAVID ZORC

### 1. PROTO-PHILIPPINE ACCENT

Contrastive accent, manifested as vowel-length or its absence on the penult, is a phenomenon found in a large number of Philippine languages, e.g. Aklanon (Akl), Balangao (Blw), Bikol (Bik), Cebuano (Ceb), Hanunoo (Han), Ibanag (Ibg), Ifugao (Ifg), Ilokano (Ilk), Isneg (Isg), Kalinga (Kla), Kapampangan (Pam), Sambal (Sbl), and Tagalog (Tag). Research in progress<sup>1</sup> on the status of an impressively large number of cognate forms that agree on the placement of such accent (e.g. penult length as in Tag *di:laq* [dí:laq] 'tongue' (example #6 below), or penult shortness as in Tag *mata* [mátá]<sup>2</sup> 'eye' (example #25 below)) in these genetically diverse languages suggests that the phenomenon may be attributed to their common parent language, herein called Proto-Philippine (PPH).<sup>3</sup>

Forms reconstructable with penult length include:<sup>4</sup>

(1) Akl, Ceb *qaba:gah-*, Bik, Ilk, Isg *qaba:ga*, Kla *qaba:la* 'shoulder', Kalamian *kabalaq* 'arm' < PPH \*qaba:Ra[h].

(2) Akl, Ceb *ba:gah-*, Tag, Bik, Han *ba:ga*, Pam, Sbl *ba:ya*, Ilk *ba:ra*, Ifg *ba:la* 'glowing embers' < PPH \*ba:Ra[h].

(3) Ceb, Bik, Ibg, Ilk, Ifg, Isg, Pam *ba:lu* 'widow' < PPH \*ba:lu.

(4) Isg *da:ga*, Ilk *da:ra*, Ifg *da:la*, Pam, Sbl *da:yaq* 'blood' < PPH \*da:Raq.

(5) Bik, Ceb, Blw, Han, Ibg, Ilk, Isg, Pam *da:lan* 'path, trail' < PPH \*da:lan.

(6) Akl, Ceb, Bik, Han, Pam, Tag, Sbl *di:laq*, Ifg, Ilk, Isg *di:la* 'tongue' < PPH \*di:laq.

(7) Akl, Ceb, Bik, Han hi:lut, Tag hi:lot, Pam, Sbl, Ilk, Isg qi:lut  
'to massage, rub' < PPH \*hi:lut.

(8) Akl, Ceb, Bik, Han qi:kug, Sbl qi:kuy, Kankanay qi:ko 'tail' <  
PPH \*i:kuR.

(9) Akl, Ceb ku:tuh-, Bik, Han, Isg, Ilk, Pam, Sbl ku:tu, Tag ku:to  
'(head) louse' < PPH \*ku:tu[h].

(10) Ceb, Pam, Sbl, Tag la:la, Bik ra:ra, Ifg, Ilk, Isg la:ga 'to  
weave, braid (mats)' < PPH \*la:ja.

(11) Bik, Ceb, Han, Ilk, Sbl, Tag la:na, Pam la:ña 'coconut oil' <  
PPH \*la:ña.

(12) Akl, Ceb, Bik, Han qi:pun, Tag qi:pin, Ibg, Isg qi:pan, Ilk,  
Sbl qi:pən 'tooth' < PPH \*qi:pən.

(13) Akl, Ceb, Bik, Han pu:sud, Tag pu:sod, Ilk pu:səg, Isg pu:sag,  
Pam pu:sad 'navel' < PPH \*pu:səj.

(14) Akl, Ceb, Bik, Han, Ilk, Isg, Pam si:ku, Tag si:ko, Ifg hi:qu,  
Sbl hi:ku 'elbow' < PPH \*si:ku.

(15) Akl, Ceb, Bik, Han, Pam, Sbl tu:buq, Tag tu:boq, Ilk, Isg, Ifg  
tu:bu, Ibg tu:vu 'to grow, sprout' < PPH \*tu:buq.<sup>5</sup>

Forms reconstructable with a short penult vowel include:

(16) Ceb, Han, Ilk, Pam, Sbl, Tag qanak, Ibg, Isg qanaq ~ qanak-  
'child' < PPH \*anak.

(17) Akl, Ceb qapuh-, Tag qapo, Ilk, Isg qapu, Ibg qafu 'grand-  
child' < PPH \*apu[h].

(18) Akl, Ceb, Bik, Han, Ibg, Ilk, Isg, Pam, Kla, Tag qasin,  
Kalamian kasin 'salt' < PPH \*qasin.

(19) Bik, Pam bagaq, Sbl bayaq, Ilk bara, Ifg, Kla bala 'lungs' <  
PPH \*baRaQ. (There is a doublet \*ba:RaQ among Southern Philippine lan-  
guages, e.g. Akl, Ceb, Han, Tag, Kalagan ba:gaq 'lungs'.)

(20) Akl, Ceb, Bik, Pam, Tag basaq, Ilk, Isg basa, Sbl bahaq 'wet'  
< PPH \*basaq.

(21) Akl, Ceb, Han bugas, Tag bigas, Sbl bayah, Isg baggat 'husked  
rice' < PPH \*bəRas.

(22) Han, Ibg, Ilk, Isg, Pam danum, Sbl lanum 'water' < PPH \*danum.

(23) Ilk (da)kayu, Kla (di)qayu, Pam (qi)kayu, Tag kayo 'ye' < PPH  
\*kayu.

(24) Bik, Ceb, Han, Ilk, Isg, Tag laḡuy 'to swim' < PPH \*laḡuy.



(25) Akl, Ceb, Bik, Han, Ifg, Isg, Ilk, Pam, Sbl, Tag mata 'eye' < PPH \*mata.

(26) Ceb puluq, Bik, Han puruq, Isg pugu, Ibg fugu 'island' < PPH \*pujuq.<sup>6</sup>

## 2. THE POSITION OF PANGASINAN

In most instances where forms in Pangasinan (Png) are clearly cognate (and least likely to be borrowings), such forms do not agree in reflecting penult length: Png qabala 'shoulder' (#1), balu 'widow' (#3), dala 'blood' (#4), dalan 'road, path' (#5), dila 'tongue' (#6), qilut 'to massage' (#7), qikul 'tail' (#8), kutu 'louse' (#9), laga 'to weave mats' (#10), lana 'oil' (#11), qipən 'tooth' (#12), puseg 'navel' (#13), siku 'elbow' (#14), tubu 'to grow, sprout' (#15). On the other hand, forms cognate with those having a short penult do agree: Png qanak 'child' (#16), qapu 'grandchild' (#17), qasin 'salt' (#18), bala 'lungs' (#19), basa 'wet' (#20), balas 'husked rice' (#21), danum 'water' (#22), kayu 'ye' (#23), lanuy 'to swim' (#24), mata 'eye' (#25), pugu 'island, islet' (#26).

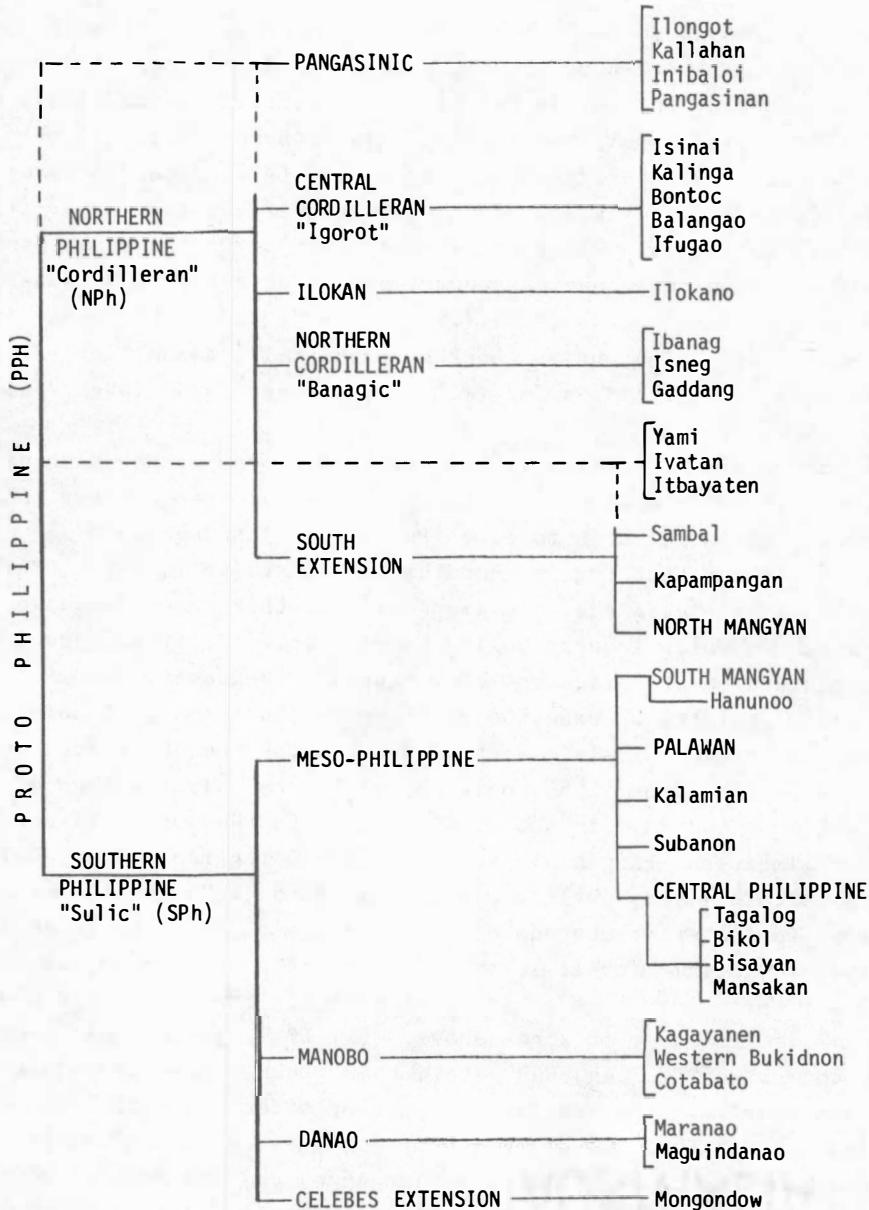
That this state of affairs is the result of loss, rather than of retention, can be determined by having a sound hypothesis about the subgrouping of Png in relation to other Philippine languages.

It is now clear that Png is immediately related to Inibaloi, Kallahan (Kayapa and Keley-1), Karaw, and Ilongot;<sup>7</sup> together these languages form a subgroup which is here called Pangasinic.<sup>8</sup> Previous surveys have not agreed on the placement of Pangasinic languages. Thomas and Healey (1962) put Png between the Northern and Southern Philippine families, but group Inibaloi independently as one member of Northern Philippine (NPh). Dyen (1965) puts Inibaloi (the only language of the group used in his study) in the Cordilleran (NPh) Hesion. Llamzon (1974) clearly delineates a Pangasinic subgroup (including Pangasinan, Kallahan, and Inibaloi), but puts it within what Reid (1974) calls the Igorot or Central Cordilleran subgroup of NPh languages. Note that each study identifies Pangasinic (or at least some members of the group) as NPh.

At least three solutions could be offered for the absence of penult length in the first set of words above. (1) If Pangasinic had separated earlier than Ilk, Ifg, Tag, Han, etc., these latter languages together might have developed contrastive word accent after the split with Pangasinic, but before the NPh and SPh languages began to diverge.

(2) Since there are other Philippine languages that do not reflect PPH accent, e.g. Kuyonon and Tausug within Bisayan (Zorc 1977), Buhid within South Mangyan, Gaddang within Banagic, etc., it is therefore

plausible that Png independently lost contrastive accent in cognate forms. (3) The development of contrastive accent could have been a dialectal feature of PPH, and languages like Ivatan, Tausug, Pangasinan, and Gaddang are all survivors of dialects in which such contrasts were not made, while Ilokano, Ibanag, Ifugao, Tagalog, and Hanunoo are all survivors of dialects where contrasts were made.



Of these proposals, #3 is the least intuitively satisfying and involves complex reasoning, recourse to which is not necessary in the case of Png (as will be demonstrated below); #1, while plausible, contradicts the present state of our knowledge which unequivocally puts Pangasinic within NPh.

An eclectic view of the subgrouping of a number of Philippine languages is presented in the Tree Diagram (on facing page). The positions of the Pangasinic and Batanic (Ivatan or Bashiic) subgroups have been approximated; the ambivalence is indicated by the broken lines. A large number of known Philippine languages have been excluded here, but those given represent a good sampling of the major Ph groups (excluding Borneo). Data for the grouping of languages within NPh or SPh are drawn from Thomas and Healey (1962), Dyen (1965), and Llamzon (1974); the major subgroup nodes: Pangasinic (Llamzon 1974), Central Cordilleran (Reid 1974), Ilokan and Banagic (Dyen 1965), North Extension and Meso-Philippine (Zorc 1974 and 1977), Manobo (Elkins 1974), Danao (Allison 1974), Celebes Extension (Charles, dissertation in progress).

For the present, #2 is taken to be the most attractive and simple explanation of the status of length in pre-Png, i.e. Pangasinan lost contrastive word accent, but developed or subsequently redeveloped the phenomenon in a unique way.

### 3. THE LOSS OF PPH ACCENT IN PRE-PANGASINAN

Zorc (1972) treats the provenance of Tagalic (Central Philippine) accent. It was noted that slightly more than half (57.3%) of the Tag and Akl forms from the basic vocabulary (reflected by a modified Swadesh 200-meaning list) were accented on the ultima (i.e. had short penult vowels) when a choice of penult length or shortness on forms with an open penult (CVCV(C)) was potentially available. Several phonological and morphological reasons for this were put forward, a number of which also apply to Png. Other important factors have come to light since that article was written.

Benton (1971b:257-60) gives the Swadesh 200-meaning list for Png. Of the 165 forms with a single consonant after the penult vowel, only 27 have penult length, all of which can be explained as secondary developments, borrowings, or recent innovations (4.-5. below). Of the remaining 138 forms with a short penult, all of those that can be reconstructed for PPH with penult length appear to indicate the loss of such length in pre-Png. A large portion of basic vocabulary (from outside the Swadesh lists) also reflects the loss of PPH penult length:<sup>9</sup>

(27) PPH \*qali:ma[h] > Akl, Ceb qali:mah-, Blw li:ma, Ilk, Isg, Kla qi:ma (metanalysed as if it had an \*-al- infix), Png lima, Kalamian kalimaq 'hand'.

(28) PPH \*[]a:min 'all' > Ilk qa:min 'all', Akl qa:min 'to use up entirely', Png qamin 'all'.

(29) PPH \*qa:ñud > Ilk qa:nud, Pam qa:ñud, Tag qa:nod, Png qanur 'to drift, flow'.

(30) PPH \*qasa:wa > Ilk, Kla, Akl, Ceb qasa:wa, Png qasawa 'spouse'.

(31) PPH \*[]a:su > Ilk, Kla, Pam qa:su, Tag qa:so, Png qasu 'dog'.

(32) PPH \*ha:ɲəs 'to gasp, pant, breathe' > Ilk qa:ɲəs 'to breathe', Akl, Ceb ha:ɲus 'to breathe heavily, gasp', Png qaɲəs 'to breathe'.

(33) PPH \*ha:səŋ > Pam, Ilk, Isg qa:səŋ, Akl, Bik, Tag ha:səŋ, Png qasəŋ 'gills'.

(34) PPH \*ba:buy > Ilk, Blw, Akl, Ceb ba:buy, Png babuy 'pig'.

(35) PPH \*ba:ləs > Ilk ba:ləs, Ceb ba:lus, Png baləs 'to revenge, repay'.

(36) PPH \*ba:ɲun > Ilk, Isg, Akl, Ceb, Han ba:ɲun, Png baɲun 'to rise, get up'.

(37) PPH \*bi:laŋ > Ilk, Blw, Isg, Akl, Han, Tag bi:laŋ, Png bilan 'to count'.

(38) PPH \*bu:lan > Ilk, Isg, Han, Ceb, Bik bu:lan, Png bulan 'moon, month'.

(39) PPH \*bu:ɲa[h] > Akl, Ceb bu:ɲah-, Ilk, Isg, Tag bu:ɲa, Png buɲa 'fruit'.

(40) PPH \*da:ləm 'deep' > Blw da:līm, Kla qa-da:lom, Akl ma-da:tum 'deep', Png daləm 'inside, depth'.

(41) PPH \*du:ma 'other, different, to accompany' > Ilk du:ma 'to differ', Dibabawon Manobo du:ma 'companion, other', Png duma 'different'.

(42) PPH \*[]i:ɲaR 'noise, trouble' > Pam qi:ɲay- ~ qi:ɲe, Png qiɲal; note Malay iɲar 'noise'; possibly cognate is Ilk qi:ɲar 'to defy'.

(43) PPH \*kawa:yan > Ilk, Ibg, Isg, Akl, Ceb, Bik, Tag kawa:yan, Png kawayan '(spiny) bamboo'.

(44) PPH \*la:ki 'male, man' > Ilk, Isg, Tag la-la:ki 'man', Akl ta:ki 'man, male', Png laki 'male, boy'.

(45) PPH \*lu:haq > Akl, Ceb, Bik, Tag lu:haq, Ifg, Kla lu:wa, Png lua 'tears (from crying)'.

- (46) PPH \*lu:bid > Bik, Ceb, Tag, Ilk, Isg, Kla lu:bid, Png lubir '*string*'.
- (47) PPH \*lu:jan > Ilk lu:gan, Isg lu:gan, Sbl lu:lan, Bik lu:nad (metathesis), Png lugan '*to ride (vehicle, canoe)*'.
- (48) PPH \*Ru:suk 'rib' > Ilk ru:suk '*epigastrium*', Akl, Ceb, Bik gu:suk 'rib', Png lusuk '*abdomen*' (for original meaning note Malay rusok 'rib').
- (49) PPH \*lu:tuq > Bik, Ceb lu:tuq, Tag lu:toq, Ilk, Isg lu:tu, Png lutu '*to cook*'.
- (50) PPH \*mu:Riŋ > Ilk, Itneg mu:giŋ '*forehead*', Isg mu:giŋ '*face*', Itbayaten muyiñ '*face*', Png muliŋ '*forehead*'.
- (51) PPH \*ŋa:jan > Isg ŋa:gan, Bik, Han ŋa:ran, Png ŋaran '*name*'.
- (52) PPH \*qu:ban > Ilk, Pam, Akl, Ceb, Bik, Tag qu:ban, Png quban '*grey-hair*'.
- (53) PPH \*[]u:pak > Akl, Ceb, Tag, Pam, Sbl, Ibg qu:pak '*bark (of tree), peeling*', with doublet PPH \*[]u:bak > Bik, Pam qu:bak, Png qubak.
- (54) PPH \*qu:ləj '*snake, worm (creeping creature)*' > Ilk qu:ləg, Ifg qu:log, Png quləg '*snake*', Bik, Ceb qu:lud, Pam qu:lad, Sbl qu:əl '*worm*'.
- (55) PPH \*qu:lu > Blw, Ilk, Ifg, Isg, Han, Ceb, Sbl qu:lu, Png qulu '*head*'.
- (56) PPH \*[]u:taq > Blw, Ifg, Ilk, Isg, Kla qu:ta, Png quta '*to vomit*'. The NPh languages give evidence of penult length, while forms like Sangil m-úta, Malay m-untah, attest the antiquity of the form; SPH languages generally reflect a different etymon (innovation?), PSP \*su:ka.
- (57) PPH \*pa:jay > Ilk pa:gay, Ifg pa:ge, Han pa:ray, Ceb, Tag pa:laj, Sbl pa:li, Png pagəy '*unhusked rice, the rice plant*'.
- (58) PPH \*pi:liq > Akl, Ceb, Bik, Han, Pam, Sbl, Tag pi:liq, Ifg, Ilk, Isg, Kla pi:li, Png pili '*to choose, select*'.
- (59) PPH \*pu:nas > Ilk, Isg, Tag, Bik pu:nas, Blw, Sbl pu:nah, Png punas '*to wipe (off)*'.
- (60) PPH \*su:su > Akl, Ceb, Han, Ilk, Kla, Pam su:su, Ifg hu:hu, Png susu '*(female) breast*'.
- (61) PPH \*ta:[]u[h] > Akl, Ceb ta:wuh-, Bik, Han ta:wu, Tag ta:qo, Ilk, Sbl, Pam ta:(q)u, Png tuqu '*person*' (see note 7 for assimilation of a to u in Png form).

(62) PPH \*ta:kut > Bik, Tag, Pam ta:kut, Ifg ta:qot, Png takut 'to fear'.

(63) PPH \*tu:lak > Tag, Pam, Sbl, Ilk (dial) tu:lak, Png tuiak 'to push'.

In addition to the 52 forms cited above (#1-15, 27-63), many more (over a hundred) have thus far been found in Png with short penult vowels, where the etyma reconstructable for PPH have long penult vowels. To save space, we will restrict the citation of forms to those above. The selection is sufficiently large and varied enough to make it difficult to believe that all these instances could have been produced by analogical change. It is thus a simpler hypothesis that the PPH penultimate length was lost in pre-Png.

#### 4. THE DEVELOPMENT OF CONTRASTIVE LENGTH IN PANGASINAN

Although the original PPH length was lost in pre-Png, modern Png clearly has penult and antepenult length: la:ga 'to blanch' (#66) as opposed to laga 'to braid' (#10); qanak 'child' (#16), but qa:nak 'children' (4.2. below); ka:ruman 'yesterday' (#70). Each instance observed in the data must be explained as a later development in the history of Pangasinan. Two ways in which length resulted have been found, both involving compensation for the loss of a consonant; the lost consonant was either: (1) preconsonantal glottal stop (\*qc), or (2) the first member of a geminate cluster.

##### 4.1. COMPENSATORY LENGTH DERIVED FROM THE REDUCTION OF \*qc CLUSTERS

Some examples of the loss of \*q (also observed in Tag) include:

(64) PPH \*baqRu[h] 'new' > Akl, Ceb baqquh- (metathesis), Bik baqqu, Ilk baru, Isg bagu, Ifg balu, Ivatan vaquuq, Kalamian bakluq, Tag ba:go, Png ba:lu.

(65) PPH \*haqlu 'pestle' > Akl haṭqu (metathesis), Ilk qalqu, Bik haqlu, Tag ha:lo, Png qa:lu. There is also a doublet, PPH \*qahlu > Ceb qalhu (metathesis), Itbayaten qahquq, Kalamian kaluq.

(66) PPH \*laqgaq 'to blanch, boil in water' > Ceb laqgaq, Bik gaqgaq (assimilation), Kalamian lakgak, Tag la:gaq, Png la:ga.

(67) Kallahan guqquq, Ilongot guqqu, Png gu:gu 'to scratch (an itch)' < PPG \*guqquq.

(68) Inibaloṭ maqciṃ, Kayapa maqdiṃ, Ilongot madiqdiṃ, Png qa:rəm 'afternoon' < PPG \*maq()dəm (with dissimilation or reanalysis of the

prefix in Png).

(69) Kayapa kuq-kuqgip, Png ku:gip 'to dream' < PPG \*kuqgip.

(70) Inibaloi kaqçiman, Kayapa kaqdiman, Png ka:ruman 'yesterday' < PPG \*kaq-dəman. Note antepenult length in Png.

#### 4.2. COMPENSATORY LENGTH DERIVED FROM THE REDUCTION OF GEMINATE CLUSTERS

The phonotactics of Png prohibit geminate clusters within a morpheme, but permit other kinds of clusters (e.g. Png qiknul 'egg', kətkət 'to bite', kəlpə 'fog', qamta 'to know', bəkləw 'neck', nəŋnəŋ 'to see', makdəm 'sharp', lupda 'to spit', etc.). Nevertheless, geminate clusters probably occurred in pre-Png. For example, in the following we find Png -a:l- corresponding to a Kayapa -all-:

(71) Kayapa qa-ballɨg, Inibaloi qɨ-badɨg, Png ba:ləg < PPG \*balləg 'big'.

In one instance, Proto Pangasinan (PPG) appears to have formed a geminate cluster as the result of assimilation; again, Png shows length where Kallahan has a corresponding geminate cluster:

(72) Kayapa qaggɨw, Keley-1 qaggew, Png qa:gəw < PPG \*qaggəw 'day'; but Kalamian kaldaw, Ifg qalگو, Bik, Ilk qaldaw, Isg qalgaw, Tag qa:raw < PPH \*qaljaw.

Among NPh languages, pluralisation of nouns referring to people, kin, or relationships is accomplished by consonant gemination, sometimes with the addition of CV- or CVC- reduplication. Thus, there is Ilk qubiŋ 'child' ~ qubbiŋ or qub-qubbiŋ 'children', qa:di 'younger sibling' ~ qaddi 'younger siblings', qama 'father' ~ qamma or qamqamma 'fathers', qapu 'grandparent' ~ qappu or qapqappu 'grandparents, gentlemen, ladies'; Ifg quŋa 'child' ~ quŋquŋŋa 'children'. In Png analogous plurals are formed with the addition of penult length: Png qanak 'child, offspring' ~ qa:nak 'children', qugaw 'child' ~ qugu:gaw 'children', qagi 'younger sibling' ~ qaga:gi 'younger siblings', tuqu 'person' ~ tutu:qu 'persons'. These forms can be explained as the result of the loss of the first member of geminate clusters with subsequent compensatory lengthening (as in #71-72 above), i.e. from \*qannak, \*qug-(q)uggaw, \*qag-(q)aggi, \*tu-tuququ respectively. Once introduced, this method of pluralisation then became productive, as in polis 'policeman' ~ popo:lis 'policemen'.

The Png active nonpast prefix man- requires penult length on inflected forms, e.g. tanəm 'to plant', but man-ta:nəm 'will plant'. The length can be explained as compensation for the reduction of the following geminate cluster. Note that an analogous formation in Kayapa

requires gemination of the consonant following the penult vowel, as in Kayapa *man-tanim* 'will plant' (*tanim*), *mam-bayyad* 'will pay' (*bayad*), *man-qabbuŋ* 'will dwell' (*qabuŋ*), *man-taqqi* 'will defecate' (*taqi*), etc.

The Png future passive suffix *-ən* yields derivatives with antepenult length, as in Png *ta:wag-ən* 'will be called' or *pa-u:gip-ən* 'will be put to sleep', regardless of the accent of the base (*ta:wag* 'to call', *pa-ugip* 'put to sleep'). In Kayapa analogous formations also have geminate clusters: *qaddal-in* 'will be studied' (*qadal*), *bunnuq-in* 'will be killed' (*bunu*). Thus, the Png forms again appear to be the result of compensatory lengthening associated with the reduction of a following geminate cluster, i.e. pre-Png *\*tawwag-ən*, *\*pa-uggip-ən*, etc.

#### 4.3. THE EVIDENCE IN DOUBLETS FOR BORROWING

With length thus introduced secondarily, forms from other dialects could be borrowed without undergoing assimilation in accent.<sup>10</sup> Dyen (1956:87) notes: "Where doublets are found whose difference cannot be assigned to the effects of analogy, that member is inherited whose correspondences conform to those found in the basic vocabulary. The other member is a borrowing".

(73) Png *pugu* 'island, islet' is the inherited form (see #26), while Png *pu:lu* and Ilk *pu:ru* are irregular in accent and in the respective reflexes for PPH *\*pujuq* 'island', although a PNP *\*pu:juq* is possible (Kalinga *pu:gu*, Sbl *pu:luq*). Png *pu:lu* is clearly a borrowing.

(74) Png *qusil* 'to run after, pursue' is regularly derived from PPH *\*u:siR*, note Han, Isg *qu:sig* 'to pursue', Tag *qu:sig* 'to persecute', Ilk *qu:sig* 'to investigate', while Png *qu:sig* 'to investigate' is probably a borrowing from Ilk, based on the similarity in semantics and the irregularity in the correspondence for PPH *\*R*.

The two examples illustrate the loss of PPH penult length in pre-Png, but the subsequent development of length as the result of secondary change: first, internal developments (compensatory lengthening), then external influences (borrowing).

## 5. RESIDUAL PROBLEMS AND CONSIDERATIONS

### 5.1. THE PLURAL OF THE WORD FOR 'CARABAO'

The plural form of *duwag* 'carabao' is *dərə:wag*. Since this noun is not readily associable with the class discussed in 4.2. (names of kin and relationships), the length appears to be phonologically irregular. However, the shape of the trisyllabic form suggests strongly that the



dissyllabic form was originally \*dəwəg (duwəg being the result of assimilation of \*ǰ to w); the plural dərə:wəg would then arise from internal consonant doubling with CV- reduplication (i.e. from \*də-dəwəg) by regular change. The importance of the *carabao* within the family unit, as a work animal and as a symbol of wealth and importance, does not make its association with the abovementioned noun class an unlikely hypothesis.

## 5.2. THE METHATHESIS OF \*Cq CLUSTERS

There are a few forms in Png that have penult length which appear to be irregular when compared with cognate forms in other Philippine languages that have short penult vowels or Cq clusters. If independent metathesis of \*q is posited for pre-Png, the length is regularly derived according to the same phenomenon discussed in 4.1.:

(75) Png ta:wən 'sky' may be derived from PPH \*tawqən 'year' > Blw, Ilk tawən, Ifg tawon, Inibaloi, Kayapa, Kankanay tawqin 'year', in that various cycles of time or religious events were probably determined by the positions of celestial bodies. Pre-Png \*taqwən > Png ta:wən. Cognate forms for 'sky' are found in other Pangasinic languages (Ilongot tawin, Inibaloi tabin), but nowhere else in the Philippines.

(76) Png bitə:wən 'star' derives from PPH \*bitəwqən > Tag bituwīn, Buhī Bikol bitəqwən, i.e. pre-Png \*bitəqwən (with metathesis of \*q as in the Buhī Bikol form). The dialectal by-form, Png bitu:wən, is a later development with assimilation of ə to the following w, as posited for \*dəwəg above (5.1.).

(77) Png la:bi 'night' developed from PPH \*Rabi[q]i > Ilk rabiqi, Blw, Kla labi, Tag gabi ~ gabqi, i.e. pre-Png \*laqbi.

## 5.3. INNOVATIVE DEVELOPMENTS

The following forms appear to be innovations in Png, in that the penult length in each instance is difficult to explain.

(78) Png ba:ləy 'town' is somehow associated with PPH \*balay > Ceb, Han, Ilk, Isg balay 'house', all with a short penult vowel. The original word for 'house' has been replaced in Pangasinic by an innovation: Png, Ilongot, Kayapa qabuŋ > PPG \*qabuŋ. Png ba:ləy is necessarily the result of some secondary change. It is possible that the Png word is analogously formed from a verb such as \*man-balləy 'to settle' : \*balləy 'settlement, town', thereby explaining the irregular development of penult length. (Bontok, for example, has ba-balləy 'village'.)

Since resort is here made to analogy, someone might propose that there also was a workable analogy by which non-length or the loss of length was introduced into pre-Png (3.). It is the small number of forms with irregular penult length that justifies recourse to analogical change in forms such as Png ba:ləy; on the other hand, it is the overwhelmingly large number of forms from the basic vocabulary with irregular penult shortness (where PPH had penult length) that rules out analogical change as a likely hypothesis.

(79) Png qasə:wək 'smoke' clearly involves some unusual development from PNP \*[]asuk > Blw, Ilk, Kla qasuk, Ifg qahuq, Sbl qahuk. Its problematic history is regarded as beyond the scope of this paper.

(80) Png da:lin, Ilongot d+:gin, Inibaloi cadin (no accent or length indicated) 'earth' appear only in these Pangasinic languages; no etymon can be assigned outside of this subgroup. Although the length in this form may be the result of secondary changes, e.g. from PPG \*dallin, it is nevertheless possible that the innovation took place after length was introduced into pre-Png and could be productively applied in new formations.

#### 5.4. RECENT BORROWINGS

The majority of the instances of penult length found in the data of Benton (1971a) and Fernandez Cosgaya (1865) are readily explained as borrowings from surrounding Philippine languages such as Ilk, Pam, or more recently from Pilipino, which is based on Tag. For example:

(81) Png da:yat 'sea' is from a language in which PPH \*R > y, such as Pam da:yat; Akl, Ceb, Bik, Tag da:gat, Sbl la:yat < PPH \*da:Rat.

(82) Png pa:lar 'lines in palm of hand' is probably from Pam or Tag pa:lad 'palm' (Ceb, Bik pa:lad, Ibg pa:lag < PPH \*pa:laɟ); since the inherited word for 'palm' is Png dakulap (Ilk, Isg daku:lap, Han dalu:kap < PPH \*daku:lap). The indigenous reflex would have yielded Png \*palag, and the secondary meaning (also found in Pam and Tag) makes the form suspect.

(83) Png, Pam, Ilk, Ibg, Tag ka:yo 'piece of cloth' is an item of trade that could pass freely from culture to culture; the source of the form is unknown.

(84) Png ba:sa 'read' could have been borrowed from any one of a number of languages with words of the same form (cf. Bik, Ilk, Pam, Tag ba:sa). These words could all exemplify a rapid spread of a form throughout the Philippines that is ultimately from Malay baca; its

initial Philippine contact language cannot be determined, although the Tagalog-speaking area is perhaps the most likely one.

(85) Png qa:pu, a term of respect used in address to a priest, has two possible sources: either from Tag qa:poq '*patriarch*', Pam qa:pu '*grandparent, sir*' term of respect for elders, or from a generalised plural respect form, e.g. Ilk qappu '*grandparents, gentlemen, ladies*'. In the former case, it would be a (relatively) recent borrowing; in the latter, it could be a secondarily developed inherited form (i.e. \*qappu > qa:pu). However, its limited use in address to priests favours its historically-recent introduction along with Catholicism.

NOTES

1. This paper is a partial result obtained in the Austronesian Genetic Classification Project directed by Isidore Dyen at Yale University, and supported by the National Science Foundation (Grant #38073X). This paper as well as much of my current research owes much to the advice, efforts, and critiques of Isidore Dyen and Harold Conklin. Errors of judgement, interpretation, and citation are my own responsibility.

2. Stress, taken as a primary feature of accent in Zorc (1972), is here treated as secondary to vowel length or shortness. Long vowels are generally stressed; a short penult vowel is followed by stress on the ultima (as indicated in the phonetic transcriptions given). However, stress is a syntactic feature and may shift in various sentence intonations. Vowel length is a primary feature, and is not lost.

3. Some Indonesianists, such as Blust (personal communication) take issue with the label "Philippine" in that (1) Proto Philippine, as a distinct meso-language of Austronesian, has not been substantiated on qualitative grounds, and (2) Philippine is a political label that excludes Borneo and Celebes, each of which has languages with close genetic affiliations to languages spoken in the Philippine archipelago. The languages discussed herein are clearly in an especially close genetic relationship to one another (see tree diagram), and the second objection is irrelevant here.

4. The following conventions are used in the data and the reconstructions: The spelling of o (where observed arbitrary), has been changed to u; while [o] is a different phoneme from [u] in Tag, and perhaps other languages treated here, it does not affect the reconstructions in any way. Data from languages that do not make contrasts of length are employed only where necessary to establish a particular element of the

reconstruction; such language names, e.g. Kalamian, Kallahan, etc., are spelled in full rather than abbreviated. Morphophonemic final h- in Bisayan (Akl, Ceb) appears to have a high correlation with final PAN \*-S and \*-H, and is therefore tentatively attributed to PPH (Zorc 1977). Evidence for PPH \*q comes from the correspondence Kalamian k, other languages used here q.

5. Over a hundred other reconstructions can be obtained from the cognate forms cited in Reid (1971), e.g. \*bu:lan 'moon', \*la:ŋit 'sky', \*ba:buy 'pig', \*bi:laŋ 'count', \*buqa:ya 'crocodile', etc.

6. Over a hundred similar reconstructions can be obtained from the cognate forms cited in Reid (1971), e.g. \*qabu[h] 'ashes', \*qanup 'hunt', \*[]apuy 'fire', \*buhək 'hair', \*dayuq 'far', \*dəkət 'to stick (to)', etc.

7. This Pangasinic group is substantiated on the basis of a number of exclusively shared innovations. Phonological: the assimilation of \*a in the penult to the vowel in the ultima, e.g. PPH \*ta:[]uh > PPG \*tuqu 'person', PPH \*ba:[h]i > PPG \*biqi 'woman'. Lexical: PPG \*kiyəw 'tree, wood' replacing PPH \*ka:yuh; PPG \*tikəy 'short'; da:lin 'earth'; \*ta:wən 'sky'; \*[q]abuŋ 'house'; etc. Morphological: metathesis of certain prefixes and infixes, e.g. PPH \*mə- > PPG \*[q]əm- adjective, PPH \*-um- > PPG \*[q]un- active nonpast. Functor: PPG \*tu 'his/her'; PPG \*si[q]ka- nominative pronoun formative; PPG \*tan deictic denoting position near addressee; PPG \*man deictic denoting remote position.

8. Since Pangasinan is a major Philippine language and is the best known within the group, the label Pangasinic is convenient. If strong objections would be raised by Inibaloi, Kallahan, Karaw, or Ilongot speakers (as Bisayans object to the label "Tagalic", or Ifugaos to "Igorot"), then some appropriate geographical label might be chosen, e.g. South-Central Cordilleran, when the extent of NPh languages is established.

9. The reconstructions presented here are validated by evidence from diverse and critical Philippine languages (Ilk, Isg, Tag, Han, Ceb, etc.). They will be discussed in greater detail in a forthcoming paper on PPH accent. Space does not allow a full presentation of data, but examples #1-15 clearly indicate the reconstruction of penult length for PPH.

10. During the posited period when pre-Png did not have penult length, borrowings would have been assimilated in that speakers who did not

have a language with penult length would not pronounce loanwords with length. While distinguishing between inherited forms and early borrowings is generally a matter of speculation, the following forms with irregular reflexes as well as short penult vowels may be taken as tentative evidence of assimilated early loans into Png:

PPH \*hu:ras 'to wash' > Akl, Ceb, Bik, Tag hu:gas, Isg qu:gat, Sbl qu:yah, Ilk qu:gas, but Png quras. The Png form shows r for expected l (\*qu:las); it is probably a borrowing from a dialect in which PPH \*R > r, such as Ilk, although the Ilk word shows a g. It is possible that pre-Png borrowed the form from pre-Ilk, but that Ilk subsequently replaced its inherited \*qu:ras with the g-form.

PPH \*[R]u:yud 'to pull' > Sangil mi-ríudiq, Akl, Bik, Ceb, Han, Ifg, Ilk, Isg gu:yud, Sbl gu:luy (metathesis), but Png guyur. The g- reflex of PPH \*R- in Ifg, Ilk, Sbl, and Png is probably indicative of a very early borrowing spread throughout the archipelago, with loss of length in pre-Png.

REFERENCES

- ALLISON, E. Joe  
1974 Proto Danao: A Comparative Study of Maranao, Maguindanao, and Iranon. M.A. dissertation, University of Texas at Arlington.
- BENTON, Richard A.  
1971a *Pangasinan Dictionary*. *PALI Language Texts: Philippines*. Honolulu: University of Hawaii Press.  
1971b *Pangasinan Reference Grammar*. *PALI Language Texts: Philippines*. Honolulu: University of Hawaii Press.
- CONKLIN, Harold C.  
1953 *Hanunoo-English Vocabulary*. Berkeley: University of California Press.
- DYEN, Isidore  
1956 'The Ngaju-Dayak "old speech stratum"'. *Language* 32/1:83-7.  
1965 *A Lexicostatistical Classification of the Austronesian Languages*. *Memoir 19, Research Center in Anthropology, Folklore, Linguistics*. Baltimore: Waverly Press Inc.
- ELKINS, Richard E.  
1974 'A Proto-Manobo Word List'. FICOCAL paper. *Oceanic Linguistics* 13:601-41.
- FERNANDEZ COSGAYA, Lorenzo  
1865 *Diccionario Pangasinan-Español*. Manila: Establecimiento Tipográfico del Colegio de Santo Tomás.
- FICOCAL = First International Conference on Comparative Austronesian Linguistics. Honolulu: January 2-7, 1974.

**LLAMZON, Teodoro A.**

- 1976 'A Subgrouping of 101 Philippine Languages'. FICOCAL paper.  
In: Nguyen Dang Liem, ed. *South-East Asian Linguistic Studies*,  
vol.2. PL, C42.

**REID, Lawrence A.**

- 1971 *Philippine Minor Languages: Word Lists and Phonologies*.  
*Oceanic Linguistics Special Publication No.8*. Honolulu: Uni-  
versity of Hawaii Press.
- 1974 'The Central Cordilleran Subgroup of Philippine Languages'.  
FICOCAL paper. *Oceanic Linguistics* 13:511-60.

**THOMAS, David and Alan HEALEY**

- 1962 'Some Philippine Language Subgroupings: a Lexicostatistical  
Study'. *Anthropological Linguistics* 4/9:21-33.

**TSUCHIDA, Shigeru**

- n.d. Sambal Vocabulary. MS.

**VANOVERBERGH, Rev. Morice**

- 1956 *Iloko-English Dictionary*. Baguio City: Catholic School Press.
- 1972 *Isneg-English Vocabulary*. *Oceanic Linguistics Special Publica-  
tion No.11*. Honolulu: University of Hawaii Press.

**WOLFF, John U.**

- 1972 *A Dictionary of Cebuano Visayan*. *Philippine Journal of Lin-  
guistics Special Monograph No.4*. Manila: Linguistic Society  
of the Philippines.

**ZORC, R. David**

- 1969 *A Study of the Aklanon Dialect*. Vol.2: *Aklanon to English  
Dictionary*. Kalibo: Aklan Printing Center.
- 1972 'Current and Proto Tagalic Stress'. *Philippine Journal of  
Linguistics* 3/1:43-57.
- 1974 'Internal and External Relationships of the Mangyan Languages'.  
FICOCAL paper. *Oceanic Linguistics* 13:561-600.
- 1977 *The Bisayan Dialects of the Philippines: Subgrouping and Re-  
construction*. PL, C44.



# TOWARDS A RECONSTRUCTION OF THE PRONOMINAL SYSTEMS OF PROTO-CORDILLERAN, PHILIPPINES<sup>1</sup>

LAWRENCE A. REID

## 1. INTRODUCTION

In an earlier paper (Reid 1974) evidence was presented for a subgroup of Philippine languages labelled Central Cordilleran. This group consisted of the following languages: Isinai, Ifugao, Balangao, Bontok, Kankanay, Kalinga and Itneg. The Central Cordilleran group (CC) forms part of a much larger subgroup of Philippine languages which is spoken over most of Northern Luzon. This larger subgroup is labelled Cordilleran, and in addition to the Central group mentioned above comprises also a Southern Cordilleran subgroup, and a Northern Cordilleran subgroup. The Southern group (SC) consists of Pangasinan, Inibaloi, Karaw, and the various dialects of Atipulu, Amduntug, Kalanguya, Kallahan, Kayapa and I-wak spoken in the provinces of Ifugao and Nueva Vizcaya and subsumed here under the name Kallahan. SC probably also includes Ilongot.

Northern Cordilleran (NC) consists of at least the following languages: Ibanag, Gaddang, Yogad, Isneg, Malaweg, Itawis (also called Itawit), Ilokano, and the languages of the various Negrito groups of Cagayan, Isabela, and Aurora Subprovince, labelled variously as Agta, Atta and Dumagat.<sup>2</sup>

The internal relations among the NC languages are not yet well understood although preliminary investigation indicates that Ilokano and Casiguran Dumagat probably form separate branches, not having closer connections to any of the other NC languages. An attempt to subgroup these languages on the basis of shared phonological innovations in conjunction with innovations in the case marking particles and pronominal systems has been attempted by James Tharp (1974).

The purpose of this paper is to examine the pronoun systems of these three groups (SC, CC and NC) and to attempt reconstruction of the pronominal systems of the meso-languages and the parent language - Proto-Cordilleran (PC).

## 2. PRONOUN SETS

All of the Cordilleran languages have multiple sets of pronouns carrying differing case functions and other syntactic properties. It is convenient to recognise at least three basic case systems, Nominative, Genitive and Oblique, the Nominative consisting of short and long forms, and the Oblique being usually the combination of a particle otherwise marking Oblique personal noun phrases, and one of the Nominative forms.

The syntax of the pronominal systems will not be discussed in this paper. This was briefly presented for the CC languages in Reid (1974), and by Constantino *et al.* (1967) for Ilokano and Isinai, in addition to Tagalog and Kapampangan, languages which are not generally considered to form part of the Cordilleran subgroup.

Eight pronouns are reconstructable for each set, differing in person and plurality components. First, second and third person singular forms occur with corresponding plurals, including the expected distinction between first person inclusive and exclusive plurals. Although evidence from elsewhere in the Philippines indicates the possibility that Proto-Philippines did not have a distinctive dual form, it is probable that a dual form existed in PC. The dual forms are here labelled 1+2 p., and the exclusive plural forms are 1 p. The use of singular and plural labels are not entirely appropriate, since semantically, dual forms are plural, requiring plural agreement in certain verb and adjectival constructions. Conklin's "minimal" versus "nonminimal" features are descriptively more adequate, however, I have opted to retain the terms singular and plural because of their familiarity.

Long Nominative pronouns consist of two formative segments, an initial segment which is the result of the prefixation of one or more Nominative case-marking particles to the latter segment, which is the pronominal segment.

Short Nominative pronouns consist only of a pronominal segment and this segment is similar to, if not identical to, that of the long forms.

Genitive pronouns are usually either identical to or reduced forms of the short Nominative pronouns.

## 3. PC PRONOUNS: INTERNAL EVIDENCE

3.1 Tharp (1974) reconstructs the following long form Nominative pronouns for Proto-Northern Cordilleran (PNC):

1 s.	*si akən	1 p.	*si kami
2 s.	*si kaw	2 p.	*si kamu, *si kayu
1+2 s.	*si kita	1+2 p.	*si kitam
3 s.	*iya, *V(n)su	3 p.	*ida

3.2 Reid (1974) reconstructs the following long form Nominative pronouns for Proto-Central Cordilleran (PCC):

1 s.	*siyakən, *sakqən <sup>3</sup>	1 p.	*dakami, *dikami
2 s.	*siqika, *sikqa	2 p.	*dakayu,
1+2 s.	*daqita, *dita, *data	1+2 p.	*dataku, *ditaku
3 s.	*siya	3 p.	*daqida, *dida

The data upon which these reconstructions are based, are fully presented and discussed in the papers cited, and will not therefore be repeated here. However, a word is in order about the alternations that were reconstructed for PCC. Some of the alternations still exist in some languages, e.g. Bontok freely alternates *daqita*, *data* '1+2 s.' and *daqida*, *dida* '3 p.'. Other languages reflect a variant with a *da-* formative for some pronouns, but a *di-* formative in others, e.g. Blw *diqni* '1 p.' (from earlier \**daqni*) but *ditaaw* '1+2 p.', Ifg. *daqyu* '2 p.' but *dituqu* '1+2 p.'. In all cases the variation is in the initial, case-marking formative. It is probable that these variants reflect a situation in Pre-CC in which at least the 2 s. pronoun was marked with a case formative \**siqi*, whereas the dual and plural pronouns were marked with \**daqi-*, the *si* and *da* corresponding respectively to the singular and plural personal Nominative case-marking particles. The *qi-* formative is a reflex of an earlier Nominative marker which by this time had lost its function in the language. There is plenty of external evidence to support the reconstruction of a Nominative case-marking particle \**qi* for Proto-Cordilleran.

Pre-CC long Nominative pronouns were probably as follows:

1 s.	*siyakən	1 p.	*daqikami
2 s.	*siqika	2 p.	*daqikayu
1+2 s.	*daqita	1+2 p.	*daqitaku
3 s.	*siya	3 p.	*daqida

The pre-CC \**siyakən* '1 s.' is suggested by Itg. *diyakən*. (Itneg reflected the \**daqi-* formative as *di-*, except in the 3 s., e.g. *dita* '1+2 s.', *dikami* '1 p.', *dikayu* '2 p.', and has generalised the *di-* to

both 1 s. and 2 s. forms as well). It should be noted that \*-yak+n reflects an earlier \*qi-akən. All other CC languages reflect a development of \*siyakən to \*sakən (see note 3).

3.3 Evidence from Ilongot, Kallahan, Inibaloi and Pangasinan (see Chart 1) suggests that these languages descended from a common ancestor, Proto-Southern Cordilleran (PSC) having the following long Nominative pronouns:

1 s.	*siyak	1 p.	*siqikami
2 s.	*siqika	2 p.	*siqikayu
1+2 s.	*siqikita	1+2 p.	*siqikitayu
3 s.	*siya	3 p.	*siqida

It is apparent that Kallahan and Inibaloi share a number of innovative developments. Two of these changes affected the 2 s., 1+2 s., 1 p., 2 p. and 1+2 p. forms. One was the voicing of the velar obstruent in these forms.<sup>4</sup> The second, and probably subsequent change, was the reduction of the unstressed high front vowel preceding the pronominal formative. This change probably also affected the 3 p. pronoun.

Prior to these changes however, other developments occurred which are reflected also in Pangasinan. One was the change of the medial syllable \*-ki- to \*-ka- in \*siqikita '1+2 s.' and \*siqikitayu '1+2 p.' producing respectively \*siqikata and \*siqikatayu by analogy with the 1 p. and 2 p. forms which both begin with the sequence \*siqika-. The analogical change spread also to the 3 p. form, changing \*siqida to \*siqikada.

The inherited 3 s. pronominal formative was replaced by \*tu. This form was originally a demonstrative but became the 3 s. Genitive pronoun in PSC. All the SC languages including Ilongot share this innovation. The displaced \*-ya apparently took the demonstrative function of \*-tu. Note Png. i-ya, Ibl. sa-ja-y and KlnKl. hu-ya 'this', each of which shows a reflex of \*-ya. The 3 s. long Nominative then became \*siqikatu, reflected in Png. sikatu.

In Kallahan and Inibaloi, \*siqga (< \*siqiga- < \*siqika-) was re-analysed as the long form Nominative case formative, and the remaining pronominal segments were equated with the forms in the Genitive pronominal set. Thus \*siyak became \*siqga-k and \*siqika became \*siqga-m. The final \*-k and \*-m being respectively the Genitive post vowel variants for 1 s. and 2 s.

Ilongot and Pangasinan independently reduced the \*siqi- initial formative to si-.

Ilongot shows several developments not shared by any of the other SC languages. \*Siqikitayu '1+2 p.' became sikisi by regular phonological

change (\*siqikitayu > \*sikitayu > \*sikitay > \*sikiti > sikisi). Similar rules reduced \*siqikayu '2 p.' to siki (\*siqikayu > \*sikayu > \*sikay > siki).

One further change peculiar to Ilongot is the change of the 3 p. form from \*siqida to siyay dī by analogy with the 3 s. form (\*siqida > \*siya + qida > \*siyayda > siyay dī).

The distribution of innovations among these languages suggests a sub-grouping as displayed in Figure 1.

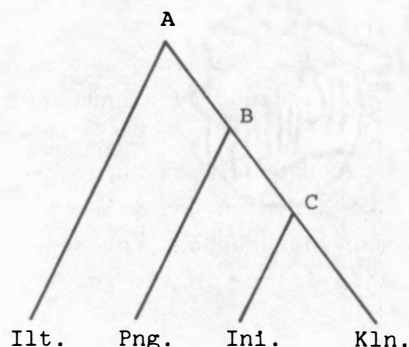


Figure 1

SUBGROUPING OF THE SOUTHERN CORDILLERAN LANGUAGES  
(BASED ON PRONOMINAL INNOVATIONS)

The development of the long Nominative pronominal system in SC was probably as displayed in Chart 1 (overleaf).

3.4 A comparison of the long Nominative pronouns reconstructed for Pre-CC and for PSC, suggests the following reconstructions for PC:

1 s.	*siyakən	1 p.	*siqikami
2 s.	*siqika	2 p.	*siqikayu
1+2 s.	*siqikita	1+2 p.	*siqikitayu, *siqikitaku
3 s.	*siya	3 p.	*siqida

The shape of the final syllable of 1+2 p. is ambiguous since PCC shows a final \*-ku and PSC shows a final \*-yu.

Innovations which characterise Pre-CC then, are the change from \*si- to \*da- on the dual and plural pronouns, and the reduction of the pronominal formative of the 1+2 s. and 1+2 p. pronouns to correspond to the short Nominative pronouns, \*-ta and \*-taku respectively. Ilokano shares both of these innovations (i.e. change from \*si- to \*da- in the dual

and plural pronouns and loss of medial \*-ki- in the 1+2 s. and 1+2 p. pronouns) however on other grounds, phonological as well as morphological, Ilokano seems to group with the NC languages. Vanoverbergh (1955: 73) noted the use in some districts "of sikami, sitayo, sita and sikayo instead of dakami, datayo, data and dakayo". It is possible that Ilokano has independently changed \*si- to da- on the basis of the same analogy that brought about the change in Pre-CC. The Ilokano change however is apparently of more recent origin than that occurring in Pre-CC, since it has not yet replaced entirely the si- initial forms.

The only pronominal innovation characterising PSC was the loss of the two final segments from the 1 s. pronoun, a loss which has since occurred probably independently, in Ilokano, Agta and Isneg in the Northern Cordilleran subgroup.

An evaluation of the reconstructions suggested above, in the light of Tharp's reconstructions of Northern Cordilleran long Nominative pronouns and external evidence, provides support for some of the reconstructions but requires a number of revisions in the shape of others.

	A	B	C	
1 s.	*siyak	>	*siqqak	
2 s.	*siqka	>	*siqqam	
1+2 s.	*siqikita >	*siqikata >	*siqqata	
3 s.	*siya >	*siqikatu >	*siqqatu	
1 p.	*siqikami	>	*siqqami	
2 p.	*siqikayu >	>	*siqqayu	
1+2 p.	*siqikitayu >	*siqikatayu >	*siqqatayu	
3 p.	*siqida	*siqikada >	*siqqada	
1 s.	siqak	siak	siqkaꞤ	hiqqak
2 s.	sika	sika	siqkam	hiqqam
1+2 s.	sikita	sikata	siqkata	hiqqata
3 s.	siya	sikatu	siqkatu	hiqqatu
1 p.	sikami	sikami	siqkami	hiqqami
2 p.	siki	sikayu	siqkayu	hiqqayu
1+2 p.	sikisi	sikatayu	siqkataju	hiqqatayu
3 p.	siyay di	sikara	siqkara	hiqqada
	Ilt.	Png.	Inl.	KlnKy

Chart 1

DEVELOPMENT OF THE SOUTHERN CORDILLERAN LONG NOMINATIVE PRONOUN SYSTEMS  
(The letters A-C represent the nodes shown in Figure 1)

#### 4. PC PRONOUNS: EXTERNAL AND PNC EVIDENCE

The following discussion will focus not only on the evidence for the long Nominative pronouns, but also on the short forms, since it appears that the short forms have in some instances become the analogical base for modifying the shape of the long forms. Discussion of Genitive forms will also be given when relevant.

##### 4.1 FIRST PERSON SINGULAR (1 s.)

On the basis of a widely occurring set of cognates, the form \*aku has been reconstructed for PAN (see Dempwolff 1938:13). Dyen (1962:215) reconstructs the form as \*x<sub>3</sub>aku. Note the following languages: Tagalog qako, Malay aku, Tongan au, Tsou aʔo, Atayal saku?, Ami kako?, etc. An additional form \*a(η)kən was reconstructed by Dempwolff with the meaning 'mine' or 'appropriate, acquire'. This form also has reflexes in Philippine languages, e.g. Tagalog qákin 'mine'.

None of the Cordilleran languages appears to reflect \*aku with a final vowel. Short Nominative pronouns reflect a PC \*-ak, the long forms reflect PC \*-akən (< PAN \*a(η)kən). The only evidence presently available to indicate that PC \*-ak (as a short Nominative pronoun) is a result of loss of the \*-u of PAN \*aku rather than the result of substitution of \*akən for \*aku with subsequent loss of final \*-ən, comes from Ilokano. In this language the completive enclitic corresponding to Tagalog na is Ilk. -en ~ -n. The former variant occurs following consonants (e.g. /naládaw/ + /-en/ → naládawen 'It's already too late'), the latter occurs following vowels (e.g. /naŋankami/ + /-en/ → nangan-kamin 'We have already eaten'). However, following the pronoun -ak the enclitic particle is -on not -en (e.g. /naŋának/ + /-en/ → nangánakon 'I have already eaten. '), apparently from an earlier \*-aku + \*-en > \*-akun.

It is probable that in Proto-Philippines, if not at some earlier date, the contrast between \*-aku as a short Nominative pronoun and \*-akən as the equivalent pronominal formative in the long Nominative set was already well established. In addition to the evidence that PC developed from such a system, Ivatan (which has been shown by lexicostatistics to be a possible first order subgroup within the Philippines (Dyen 1965, Thomas and Healey 1962), as well as Yami (Ivatan's sister language on Botel Tobago Island off the south-eastern coast of Formosa), and a number of Manobo languages in Mindanao all show the contrast between first person formatives in long and short Nominative pronouns (e.g. Ivatan, Yami qaku, yakən; Agusan Manobo, Ilianen Manobo a, siakən; Tasaday a, akən). Evidence for -akən as a long Nominative formative comes also from Subanon, and from Maranao where it appears as the pronominal segment

of the Genitive *qakən*, Oblique *rakən*, long Nominative *sakən*, but not the short Nominative, which is *qaku*.

Of the Cordilleran languages, the full *\*-akən* formative is reflected in the 1 s. long forms of each of the CC languages. It is also reflected as such in Casiguran Dumagat, Atta, Gaddang, Yogad, Itawis, and Ibanag of the NC languages. However, in all of the SC languages (including Ilongot), and sporadically in the NC languages (Ilokano, Isneg and Agta) the pronominal formative on the long Nominative form is *-ak* or *-aq* rather than a full reflex of *\*-akən*. This reduction probably developed by analogy with the short Nominative pronoun *-ak* (< *\*aku*).

The Genitive 1 s. is reconstructed as PC *\*-ku* following consonant final stems, and *\*-k* following vowel final stems. They are reflected as such in all SC and CC languages, and in all the NC languages except Casiguran Dumagat. Evidence from NC and SC languages, as well as from languages outside these groups (even from languages as far south as Timogon Murut in Sabah) show reflexes of *\*ta* for Genitive 1 s. when in combination with one of the short Nominative pronouns in a "passive" sentence, e.g. Timogon

*potoyon takamin 'I will kill you.'*  
*kill I-you*

*\*ta* is therefore reconstructed as one of the PC forms for Genitive 1 s.

#### 4.2 SECOND PERSON SINGULAR (2 s.)

On the basis of a number of forms in Formosa, Dyen (1965:302) reconstructed PAN *\*iXu()*<sup>6</sup> *'thee, thou'*. Dahl (1973:122) reconstructs both *\*ka* and *\*(i)Su* for the same pronoun. In addition, Dahl states that the two forms frequently appear in combination giving rise to the contracted *\*kaw* which Dempwolff (1938:76) assigned to PAN. The Formosan evidence clearly supports Dahl's statement with such forms as Kananabun *iikásu*, Kuvalan *?aisu?*, Ami *kíso'*, Rukai *kəsó:'*, etc. (Ferrell 1969:187). The Kananabun form and Tagalog *qikaw*, if not reflexes of an earlier *\*ikaSu* were both formed on the same structural principal of a case marker *i* plus a pronominal formative. For Proto-Philippines then, it is probable that *\*-kaw* was not a full pronoun but the pronominal formative on the long Nominative 2 s. pronoun, *\*-ka* is widely attested throughout the Philippines, including all of the Cordilleran languages as the short Nominative 2 s. pronoun.

In NC a final *-w* appears on the 2 s. long Nominative in about five languages and appears on Tharp's PNC reconstruction of 2 s. It is necessary then to revise the PC reconstruction to *\*siqikaw*. Both PSC



and PCC are presumed to have lost the final segment by analogy with the corresponding short Nominative pronoun \*-ka.

The Genitive 2 s. is reconstructed as PC \*-mu following consonant final stems, and \*-m following vowel final stems. They are reflected as such in most of the Cordilleran languages. In NC, Gaddang and Yogad have replaced \*-mu with -nu, retaining the -m reflex of the variant following vowel final stems. The same replacement has taken place, apparently independently, and relatively recently, in the CC languages Kalinga and Balangaw. The replacement has not occurred in Itneg, a language which subgroups closely with Kalinga. Tharp (1974) reconstructs \*-nu ~ -m as the PNC forms for 2 s. However the evidence suggests that \*-nu was a 2 p. form which developed from earlier \*niyu (see section 4.6 below). Subsequently the 2 p. form was extended to replace the 2 s. form, a development commonly found in languages of the world. -nu still exists as the 2 p. form in Isneg, Ibanag and Itawis of the NC group.

#### 4.3 DUAL PERSON (1+2 s.)

In PAN there was apparently no distinction between the dual form of a pronoun (1+2 s.) and the first person inclusive plural (1+2 p.) form. Dempwolff (1938:81) reconstructed PAN \*kita with the meaning '*we, inclusive*'. Philippine languages which have developed a distinction have retained a reflex of PAN \*kita with the restricted meaning of '1+2 s.' and have added a third syllable to the form to create a plural form meaning '1+2 p.', e.g. Batak kita '1+2 s.', kitami '1+2 p.', Kalagan kita '1+2 s.', kitadun '1+2 p.'; Western Bukidnon Manobo sikita '1+2 s.', sikitaw '1+2 p.'.

This evidence along with the NC reconstruction supports the postulated PC \*siqikita '1+2 s.'.

The Genitive 1+2 s. is reconstructed as \*-ta. It is reflected as such in all the Cordilleran languages except Ilongot which has si for both 1+2 s. and 1+2 p. This form is a normal phonological development in Ilongot of \*tayu, which is reconstructed for PSC 1+2 p.

#### 4.4 THIRD PERSON SINGULAR (3 s.)

The form reconstructed by Dempwolff (1938:67) for PAN is \*ija. This probably contains a case formative \*i-, and is reflected as such in the Ivatan long Nominative qíya and Mamanwa qiza '3 s.'. The \*-ya formative occurs apart from \*qi- in various other Philippine languages, e.g. Samal ia, Tagbanwa (Aborlan) kan-ya, Tausug s-iah '3 s.'. The reconstruction of PC \*siya appears to be supported. The short Nominative 3 s. pronoun is widely attested as Ø. It is reconstructed as such for PNC, PCC and PSC, and therefore also for PC.

The Genitive 3 s. is reconstructed as \*-na. It is reflected as such in all the CC and NC languages. It was replaced in SC by \*-tu (see section 3.3 above).

#### 4.5 FIRST PERSON EXCLUSIVE PLURAL (1 p.)

Dempwolff (1938:74) reconstructed \*kami '*we, exclusive*'. The form is widely reflected both in the Philippines and elsewhere. It is a short Nominative pronoun as well as the pronominal formative on the long Nominative pronouns. It is reconstructed for PNC and is strongly supported as the PC reconstruction.

The Genitive 1 p. is reconstructed as \*-mi. It is reflected as such in all the Cordilleran languages.

#### 4.6 SECOND PERSON PLURAL (2 p.)

Consideration of the variety of 2 p. forms in NC as well as in languages outside the Cordilleran groups considerably complicates the reconstruction of this pronoun for PC.

Reconstructions for PAN include Dempwolff's \*kamu, Dahl's \*mu, and Dyen's \*mi?, \*mu?, and \*miu?. Dempwolff's reconstruction, with its \*ka-formative, was probably a Nominative form and is fairly widely attested as such in Philippine languages, particularly in languages outside the Cordilleran group (e.g. Bikol, Hiligaynon, Kinaray-a, Cebuano, Samar-Leyte, Tausug, Hanunoo, Batak, Tagbanwa, Mamanwa, Kalagan, Mansaka, etc.). Its presence in the Cordilleran group is restricted to Ibanag in NC which has kamu and sikamu respectively for the short and long Nominative 2 p. forms. However pronouns which apparently derive from such forms are found in Casiguran Dumagat (-kam and sikam), Yogad (-kam and sikam) and Agta (-kam and ikamuy). On the basis of this evidence Tharp (1974) reconstructs \*kamu and \*si kamu for PNC. However he also reconstructs \*kayu and \*si kayu for the same pronouns to account for evidence from the other NC languages, Atta (-kayu and sikayu), Gaddang (-kayu and ikkayu), Ilokano (-kayu and dakayu), Isneg (-kayo and dakayu) and Itawis (-kayu and ikayu).

It is probable that \*-mu was the earliest of the various forms that have been reconstructed for the Genitive PAN 2 p. After \*-mu had extended its meaning to encompass 2 s. a \*-yu formative was added to 2 p. to create a new singular-plural distinction in the second person. This change not only affected the Genitive pronouns but also the Nominative pronouns producing the following:

Genitive	2 p.	*-muyu
Nominative	2 p.	*-kamuyu

This Genitive form is reflected as a *-muyu* in Botolan Sambal and as *muyuh* in Kelabit (Northern Borneo). With loss of the first vowel and syllabification of the glide it appears as *miiyu* in Itbayaten and *-miu* in Baler Dumagat, Tondano (Sulawesi) as well as in Palauan. Chamorro reflects the form as *mizu* (< \**miyu*). Casiguran Dumagat *-moy* and Agta *-muy* reflect it with loss of the final vowel, the result of pressure towards monosyllabism in the Genitive pronouns in these languages (e.g. DgtC: *-ki, -mo, -na, -mi, -moy, -tam, -di*). Languages in the south of the Philippines, such as the Mansakic group (Mansaka, and Kalagan or Tagakaolo) as well as Mamanwa have forms which are probably innovative developments of \**-muyu*. These languages reflect a \**-mayu*, which possibly developed by analogy with the vowel sequence in the reduced Nominative 2 p. \**-kamu*. Note the following Mamanwa Nominative and Genitive pronoun matches: 1 s. *haqu, naqu*; 1+2 s. and p. *kita, nita*; 3 s. *qiza, naqiza*; 2 s. *kamo, mazo*; 3 p. *siran, niran*. Dyen's suggestion (1974) that the Mansakic \**-mayu* possibly reflects a PAN \**-mæyu* is dubious because each of these languages reflects PAN \**ə* as *i*.

The Nominative 2 p. \**-kamuyu* is not known to be reflected directly as such in any language. With loss of the medial vowel it appears as Chamorro *hamzu* (< \**kamyu*), and with syllabification of the glide as Palauan *kəmiu*. The \**-muyu* sequence is also apparent in Timugon Murut *ramuyun* 2 p. long Nominative. The reconstruction of Nominative 2 p. \**kamuyu* is necessary also to account for the divergent developments in PC mentioned above, producing, with various degrees of reduction, *-kamuy, -kamu, -kam*, and with haplology of the medial syllable *-kayu*.

The CC and SC Genitive 2 p. is almost invariably *-yu*, matching the Nominative 2 p. *-kayu* in these languages. The presence of *-kayu* in Tagalog and Kapampangan, languages of Luzon that probably do not belong to the Cordilleran group, is probably best accounted for by borrowing from that group.

A puzzling development in some of the Northern Cordilleran languages, and one which shows a parallel development in the Manobo group as well as sporadically elsewhere is the appearance of *-nu* for the Genitive 2 p. In NC all the non-Negrito languages, except Ilokano (that is excluding the Dumagat languages as well as Agta and Atta) have *-nu* for either the Genitive 2 p. or the 2 s. form. It is assumed that this innovation began in the 2 p. and spread in some languages to 2 s. A possible line of development was from \**-miyu*, reanalysed as \**-niyu* (Genitive marker \**ni* + \**-yu*), a form commonly found in the Philippines. \**-niyu* then became \**-nu*, probably on the same analogical basis that produced the forms \**-niya* (\**ni* + \**-ya*) and \**-na* for 3 s. It is perhaps significant that disyllabic forms such as *-niya* 3 s. and *-niyu* 2 p. occur together in

many languages (e.g. Tagalog, Bikol, Samar-Leyte, Hiligaynon, Tausug, Hanunoo and Maranao) whereas there appear to be no languages having -nu 2 p. which also have -niya 3 p.

The PC Genitive 2 p. is reconstructed as \*-muyu. The PC short Nominative 2 p. is reconstructed as \*-kamuyu, and the corresponding long Nominative as \*siqikamuyu.

#### 4.7 FIRST PERSON INCLUSIVE PLURAL (1+2 p.)

As indicated above (4.3), many Philippine languages appear to have developed a dual pronoun (1+2 s.) by appropriating PAN \*kita '1 incl. pl.' for the dual and forming the inclusive plural by suffixing a "plural" formative to the dual form, -don in Sarangani Manobo and Kalamangan, -yu in SC, Ilokano, Itneg and Tagalog and -da in Isneg and Itawis. Other Philippine languages appear to have suffixed other formatives, e.g. -ku in CC, and -mu or -m in many languages of NC and fairly extensive outside the Cordilleran subgroup, including Kapampangan and the languages of Mindoro. Maranao siktanu probably has a similar source.

Considering reconstructions for PNC, PCC, and PSC, PCC \*-taku appears to be exclusively shared by CC and is eliminated as a possible PC reconstruction. Although PSC \*-kitayu also appears in Ilokano tayú<sup>7</sup> and Tagalog táyu, the lack of similar cognates in languages apparently more closely related to these two languages than the SC languages, raises the possibility that borrowing has taken place with subsequent loss of the ki syllable. If Ilokano did not borrow this form from an SC language (e.g. Inibaloi or Pangasinan, or one of the ancestral stages of these languages) but inherited it, along with PSC from PC, Ilokano would need to be considered a separate first order branch of Cordilleran, a hypothesis which is not strongly supported by other data.

It was noted above that Tagalog had probably borrowed kayu '2 p.' from a Cordilleran language. It is probable that Tagalog táyu is likewise a borrowing, since the languages with which it is most closely related (Bikol, Cebuano, Samar-Leyte, Kinaray-a, etc.) all reflect kita '1 incl. pl.'. These languages do not distinguish a dual pronoun.<sup>8</sup>

The most likely candidate for PC 1+2 p. is \*-kitam, the form reconstructed for PNC. This form has a fairly wide distribution outside of Cordilleran and is therefore supported by external evidence.

Another possibility for PC that cannot be decisively eliminated is that PC did not distinguish a dual pronoun and that the distinction developed after the split into Northern, Central and Southern groups. This would account for the different reconstructions required for the proto-languages of these groups. In fact the distinction may have developed in the NC languages after Ilokano had split off (Tharp makes

Ilokano a first order branch in NC) providing a motivation for the Ilokano borrowing of *-tayu*.

However, because of the occurrence of apparent cognates in Mindoro and elsewhere, *\*siqikitam* is reconstructed for PC in place of the forms discussed above. The Genitive 1+2 p. is reconstructed as *\*-tam*.

#### 4.8 THIRD PERSON PLURAL (3 p.)

Dempwolff (1938:152) reconstructed *\*t'ida* as the PAN third person plural pronoun. Dahl (1973) reconstructs it as *\*t'ida*. Reflexes of this form in Philippine languages as *sila*, *sira*, or *sida* are numerous and function as long Nominatives. They can be considered to consist of a case formative *si* and a 3 p. formative which occurs both as a short Nominative and a Genitive pronoun.

Although the case formative appears in many languages as *si*, in a number of other languages it appears as *qi*, e.g. Kapampangan *ila* and Ilokano *qida*. The reconstruction of PC *\*siqida* is thus supported by the external evidence.

### 5. THE OBLIQUE PRONOUNS

Evidence from many of the Cordilleran languages as well as from languages outside the group clearly indicates that the Oblique pronouns were constructed with a marker for Oblique personal noun phrases, and the Nominative pronominal formative. The Oblique markers are reconstructed for PNC by Tharp (1974) as *\*kani* (singular) and *\*kada* (plural). They are reconstructed for PCC by Reid (1974) as *\*kan(i)* (singular) and *\*kan da* (plural). For PNC, Tharp reconstructs the Oblique forms, singular and plural with the marker *\*kani* plus the reconstructed Nominative pronominal formatives. In PCC, the Oblique was formed with the *\*kan* Oblique marker in combination with the long Nominative pronouns. The Oblique pronouns of PNC are assumed to more closely reflect the PC system than do the PCC pronouns.

### 6. CONCLUSION

The pronominal systems of Proto-Cordilleran are reconstructed as follows:

#### I Long Nominative Pronouns

1 s.	<i>*siyakən</i>	1 p.	<i>*siqikami</i>
2 s.	<i>*siqikaw</i>	2 p.	<i>*siqikamuyu</i>
1+2 s.	<i>*siqikita</i>	1+2 p.	<i>*siqikitam</i>
3 s.	<i>*siya</i>	3 p.	<i>*siqida</i>

## II Short Nominative Pronouns

1 s.	*-ak	1 p.	*-kami
2 s.	*-ka	2 p.	*-kamuyu
1+2 s.	*-kita	1+2 p.	*-kitam
3 s.	*∅	3 p.	*-da

## III Genitive Pronouns

1 s.	*-ku ~ *-k, *-ta	1 p.	*-mi
2 s.	*-mu ~ *-m	2 p.	*-muyu
1+2 s.	*-ta	1+2 p.	*-tam
3 s.	*-na	3 p.	*-da

## IV Oblique Pronouns

1 s.	*kanyaken	1 p.	kanikami
2 s.	*kanikaw	2 p.	kanikamuyu
1+2 s.	*kanikita	1+2 p.	kanikitam
3 s.	*kanya	3 p.	kanida

## NOTES

1. Research for this paper has been supported by a University of Hawaii Intramural Research Grant for which I hereby express my gratitude. I wish to thank Robert Blust for his comments, and also for data from Kelabit.
2. See Fox and Flory 1974 for the most up-to-date linguistic map of the Philippines.
3. Since completion of Reid 1974, I have decided that the reconstruction \*səkqən is not PCC, but is of more recent provenance, probably Proto-Nuclear-Cordilleran, the parent language of Bontok-Kankanay, Balangaw and Ifugao. The form that should be reconstructed for PCC 1 s. pronoun is \*səkən. The glottal stop was apparently introduced into the form by analogy with the 1 s. form \*sikqa.
4. Ini. k̤ (a voiceless, front velar stop) is a regular development of \*g in syllable initial position (see Reid 1974, Sec. 2).
5. The change from ki to ka also appears in Kapampangan (ikata 1+2 s.) a language which has not been shown to be a member of the Cordilleran group. The change probably developed independently in this language.
6. \*X represents a reconstructed hiatus or "non-vowel" with a sibilant reflex in some Formosan languages (Dyen 1965:30).
7. This form also occurs in Itneg, a CC language, but it is probably a borrowing from Ilokano (see Reid 1974).
8. tayuh also appears in Kelabit (Northern Borneo).

BIBLIOGRAPHY

CONSTANTINO, Ernesto, Consuelo J. PAZ, and Marietta N. POSONCUY

- 1967 'The Personal Pronouns of Tagalog, Ilukano, Isinai and Kapampangan'. In: Mario D. Zamora, ed. *Studies in Philippine Anthropology (In Honor of H. Otley Beyer)*, 567-91. Quezon City, Philippines: Alemar Phoenix.

DAHL, Otto Christian

- 1973 *Proto-Austronesian. Scandinavian Institute of Asian Studies Monograph Series*, No. 15. Lund, Sweden: Studentlitteratur.

DEMPWOLFF, Otto

- 1938 *Vergleichende Lautlehre des Austronesischen Wortschatzes*, vol. 3: *Austronesisches Wörterverzeichnis. Beihefte zur Zeitschrift für Eingeborenen-Sprachen*, Beiheft 19. Berlin: Dietrich Reimer.

DYEN, Isidore

- 1962 'Some New Proto-Malayopolynesian Initial Phonemes'. *Journal of the American Oriental Society* 82:214-15.
- 1965 'Formosan Evidence for Some New Proto-Austronesian Phonemes'. *Lingua* 14:285-305.
- 1974 'The Proto-Austronesian Enclitic Genitive Pronouns'. In: *Papers of the First International Conference on Comparative Austronesian Linguistics (Honolulu, January 2-7, 1974)*. *Oceanic Linguistics* 13:17-31.

FERRELL, Raleigh

- 1969 *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification. Academia Sinica Monograph* 17. Nankang, Taipei: Institute of Ethnology.



FOX, Robert B. and Elizabeth H. FLORY, compilers

- 1974 *The Filipino People: Cultural-Linguistic Map of the Philippines*. Manila: National Museum of the Philippines.

REID, Lawrence A.

- 1971 *Philippine Minor Languages: Word Lists and Phonologies*. *Oceanic Linguistics Special Publication No. 8*. Honolulu: University Press of Hawaii.
- 1974 'The Central Cordilleran Subgroup of Philippine Languages'. In: *Papers of the First International Conference on Comparative Austronesian Linguistics* (Honolulu, January 2-7, 1974). *Oceanic Linguistics* 13:511-60.

THARP, James A.

- 1974 'The Northern Cordilleran Subgroup of Philippine Languages'. *Working Papers in Linguistics, University of Hawaii* 6/6:53-114.

THOMAS, David and Alan HEALEY

- 1962 'Some Philippine Language Subgroupings: A Lexicostatistical Study'. *Anthropological Linguistics* 4/9:21-33.

VANOVERBERGH, Morice

- 1955 *Iloko Grammar*. Baguio City, Philippines: Catholic School Press.



# PERCEPTUAL DIMENSIONS OF TONE: THAI

JACKSON T. GANDOUR\*

**ABSTRACT.** One hundred and fourteen Thai subjects made direct ratings of dissimilarity between pairs of pitch patterns superimposed on a synthetic speech-like syllable. An INDSICAL analysis of the dissimilarities data revealed four dimensions which were interpreted as AVERAGE PITCH, LENGTH, DIRECTION and SLOPE. These interpretive labels were supported by results of a multiple linear regression analysis. No significant differences in tonal perception could be attributed to an individual subject's dialect background.

## I. INTRODUCTION

A search for linguistic explanations that account for the nature of tonal systems must converge with fundamental processes associated with the production and perception of tones. Using a multidimensional scaling procedure, the primary aims of this paper are (1) to discover the fundamental dimensions underlying Thai individuals' perception of different kinds of pitch patterns superimposed on a synthetic speech-like syllable and (2) to determine the extent and kinds of individual differences in perception that may be attributed to dialect background.

---

\*Postdoctoral Fellow. The research for this paper was supported in part by Social Science Research Council and National Science Foundation dissertation fellowships, and by NSF G537235X during my doctoral program in linguistics at the University of California, Los Angeles.

I wish to especially thank the following persons and institutions of Thailand for their assistance and cooperation: (Central Institute of English Language) Jimmy Harris and Kanda Sittachita; (Ramkhamhaeng University) Pat Noisaengsri and Udom Warotamasikkhadit; (Chiang Mai University) Tom Scovel; (Phuket Teacher Training College) faculty members of the English and Thai departments. I also wish to thank the Thai students at these various institutions who participated in the experiment.

Requests for reprints should be sent to:

Jackson T. Gandour, Department of Audiology and Speech Sciences, Purdue University, West Lafayette, IN 47907.

The secondary aim of the paper is to determine how close these perceptual dimensions correspond to earlier proposed phonological features of tone.

The multidimensional scaling procedure used in this investigation, INDSCAL (=INDividual Differences SCALing: Carroll and Chang 1970; Carroll and Wish 1974a, 1974b; cf. Harshman 1970, 1972), simultaneously analyses similarity (or dissimilarity) matrices for several individuals. The data values in the matrices generally correspond to subjective distances between stimulus objects based on judgments of similarity of different individuals. The similarities are assumed to be related to distances between stimuli in some latent psychological space. The INDSCAL procedure determines a statistically unique set of dimensions that usually can be interpreted without rotation of the axes. In addition to coordinates for the stimuli on each dimension, INDSCAL also provides information about the relative weights or perceptual saliences of each dimension for every individual. The distances between the stimuli depend on the subjects' dimension weights as well as on the stimulus coordinates. The dimension weights for a particular individual subject indicate approximately how much each dimension should be stretched so that the distances between stimuli will correlate as highly as possible with that subject's similarity (or dissimilarity) ratings. The stimulus coordinates on each dimension for all subjects may be plotted in a "group (composite) stimulus space", the dimension weights for individuals in a "subject space". It is primarily by analysis of the subject space that we may determine to what extent differences in perception, if any, may be attributed to particular individuals or subgroups.

## II. METHOD AND PROCEDURE

### A. LANGUAGE

Thai, a member of the Tai branch of the Austro-Thai language family, is an example of a tone language with five contrastive tones - three level tones and two contour tones, traditionally labelled "high" (´), "mid" ( ), "low" (`), "falling" (^) and "rising" (ˇ), e.g. /khaá/ 'to engage in trade', /khaa/ 'to be stuck', /khaà/ 'a kind of spice', /khaà/ 'to kill', /khaã/ 'leg'. For phonological description of the Thai tones, see Henderson 1949, Abramson 1962 and Gandour 1975. Figure 1 presents average fundamental frequency trajectories of the high, mid, low, falling and rising tones of Thai in word-final position.

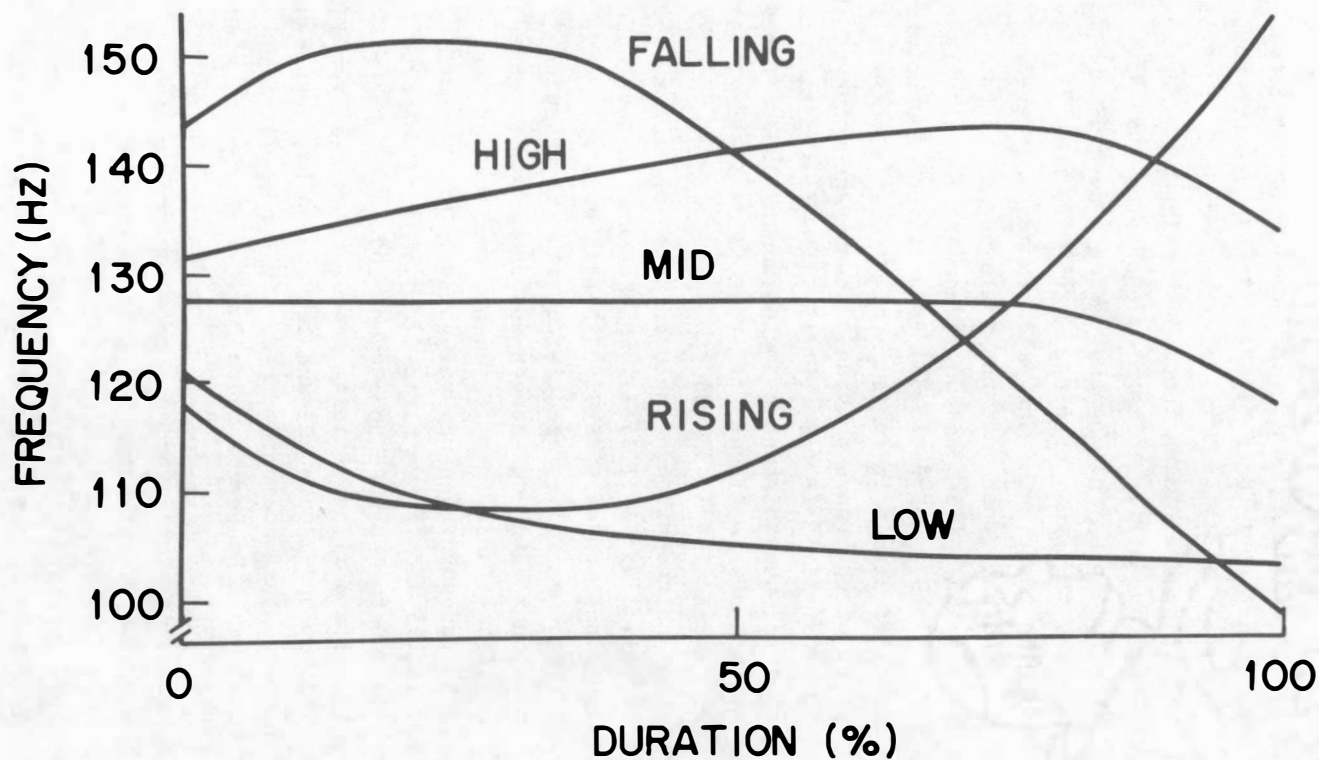


FIGURE 1

Average fundamental frequency contours of Thai tones on double vowels in word-final position (one speaker, adapted from Abramson 1962, by permission of Indiana University Research Center in Anthropology, Folklore and Linguistics).

## B. STIMULI

The stimulus set included thirteen different tonal patterns (see Figure 2). There were three level tones (11 33 55); ten contour tones, five falling tones (53 31 51 53 31) and five rising tones (35 13 15 35 13). Within each subset of falling and rising tones, two pairs of tones traverse the same pitch range but differ in magnitude of slope (53-53̄, 31-31̄:35-35̄, 13-13̄); the remaining falling and rising tones, 51 and 15, have the same direction and magnitude of slope as (53̄ 31̄) and (35̄ 13̄), respectively, but differ in the size of the pitch range. Differences in begin and end point divided the tones into the following six subsets: (11 13 15 13̄), (33 31 35 31̄ 35̄), (55 53 51 53̄) according to begin point; (11 31 53 31̄), (33 53 13 53̄ 13̄), (55 35 15 35̄) according to end point. Differences in the amount of change in fundamental frequency between the begin point and end point of the tones provide a ternary grouping of the tones: (51 15), (53 53̄ 31 31̄ 35 35̄ 13 13̄), (11 33 55). In addition to these pitch characteristics, the thirteen stimulus tones also differed in length. Nine of the tones were "long" (11 33 55 53 31 51 35 13 15), four of the tones were "short" (53̄ 31̄ 35̄ 13̄).

This particular set of thirteen tones included seven of Wang's (1967: 99) phonological tones (11 33 55 53 31 35 13); his other two level tones (22 44) and four bidirectional tones (535 313 353 131) were excluded.

Actual fundamental frequency values associated with the stimulus tones were intended to approximate real-speech measurements of tones (cf. Abramson 1962). All the fundamental frequency trajectories were linear: 11, 33 and 55 had steady frequency values at 100, 125 and 150 Hz, respectively; 35, 13, 15, 35̄ and 13̄ had rising frequency values from 125 to 150, 100 to 125, 100 to 150, 125 to 150 and 100 to 125 Hz, respectively; 53, 31, 51, 53̄ and 31̄ had falling frequency values from 150 to 125, 125 to 100, 150 to 100, 150 to 125 and 125 to 100 Hz, respectively. The rate of change in fundamental frequency for the linear rising and falling tones was 1 Hz per 12 msec for (53 31 35 13), 2 Hz per 12 msec for (53̄ 31̄ 35̄ 13̄ 51 15).

These fundamental frequency trajectories were superimposed on a synthetic speech-like syllable that phonetically approximated [wa], using a line analog speech synthesiser on the PDP-12 computer at the Phonetics Laboratory, University of California, Los Angeles (for description of speech synthesiser, see Rice 1971). In this synthetic syllable, both the first and second formants displayed rising transitions into the vowel; the steady-state portion of the vowel constituted about 62 per cent of the total duration, with spectral peaks at 630, 1130 and 3300 Hz for the first three formants.

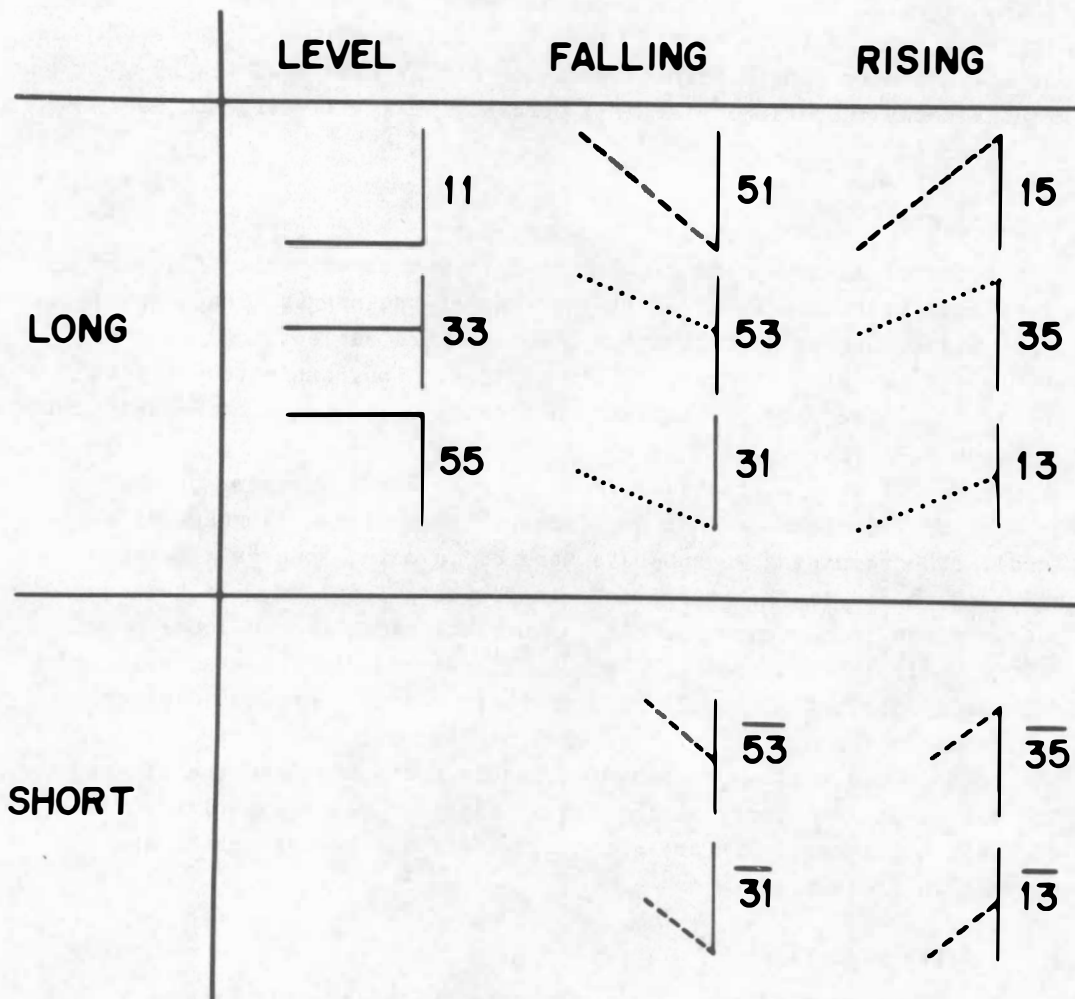


FIGURE 2

Numeric and corresponding graphic representations of the thirteen stimulus tones used in the experiment (cf. Chao 1930). The ordinate represents a 5-step pitch scale ranging from 5 (highest point) to 1 (lowest point), and the abscissa represents time: 55 "high level", 13 "low rising", 53 "high falling", etc. A suprascript horizontal bar identifies the numeric representations of the "short" tones. Solid, dashed and dotted lines along the abscissa indicate magnitude of slope of the thirteen stimulus tones.

The duration of the syllable was 312 msec for the nine "long" tones, 156 msec for the four "short" tones (cf. Abramson 1962). The amplitude curve was controlled to display a rather abrupt rise followed by a gradual decay (duration: rise = 31 percent, peak = 11 percent, decay = 58 percent).

### C. SUBJECTS

A total of 114 subjects participated in the experiment. All were in attendance at either a university or teacher-training college. Subjects were paid for their participation. None of the subjects had received any formal training in music or linguistics. They exhibited varying levels of proficiency in English. Subjects' ages ranged between 18 and 23, mean = 21 years.

Of the 114 Thai subjects, 38 were monodialectal speakers of the Central Thai dialect (= Thai or Siamese, the national language of Thailand). The remaining 76 subjects were bidialectal speakers: 65 were born and raised in Phuket province in southern Thailand, 11 in Chiang Mai province in northern Thailand. They all had attended local provincial schools through the tenth grade. The local dialect was their first "acquired" dialect; Central Thai was their second "learned" dialect. It is the latter dialect that is used as the medium of instruction in the schools and medium of communication in radio, TV, and the cinema. The bidialectal subjects having had at least 10 years experience with the national language in the schools, were fully conversant in the Central Thai dialect.

### D. EXPERIMENTAL TASK

Subjects were told that they were going to hear words from a foreign language, and that these words all had the same sequence of consonant and vowel but different pitch patterns, as in the Thai words /nâ/ 'face' and /nǎ/ 'thick'.

The stimulus set was played twice to acquaint the subjects with the nature and range of pitch variations between the thirteen tones. They were then told that their task was to report their impression of how different the pitch patterns are between pairs of these words, by circling an appropriate number on an 11-point didimilarity scale (see Figure 3). They were also told to ignore any other differences between the words that they might hear.



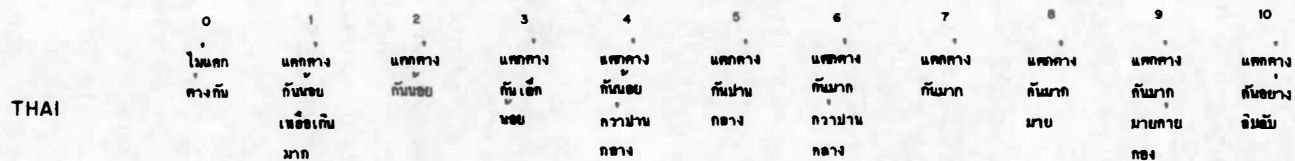


FIGURE 3

11-point scale (0 = no difference, 5 = medium difference, 10 = extreme difference) on which Thai subjects gave ratings of dissimilarity between pairs of stimulus tones.

The paired-stimuli were presented in a natural speech carrier frame / karunaa priàphtiáp siǎŋ khoǒŋ kham tǒò pay ní /  
'Compare the pitch of the following words \_\_\_\_\_'.

This natural speech context encouraged subjects to listen to the synthetic stimuli in a speech mode of processing.

A trial consisted of the carrier frame, the paired-stimuli and the response period. The time-interval between the carrier frame and the paired-stimuli was .56 secs, between stimulus *a* and stimulus *b* of the paired-stimuli (interstimulus interval) 0.2 secs and between the paired-stimuli of one trial and the beginning of the carrier frame for the next trial (intertrial interval) 4.0 secs.

Four blocks of trials were presented, each block consisting of 91 trials, yielding a total of 364 paired-comparison judgments, four judgments for each stimulus-pair. Ten buffer trials were placed at the beginning of block 1, three buffer trials at the beginning of each of the remaining three blocks. The trials were presented under one pseudo-random order in blocks 2 and 4. In blocks 1 and 2, the stimuli for each stimulus-pair type were presented in order *a-b*; in blocks 3 and 4, the order of presentation was reversed. A block of ten practice trials was administered before proceeding into the actual experiment.

Stimuli were presented on a Uher Model IC-4000 tape recorder over a loudspeaker in a conventional classroom setting. This enabled more than one subject to be tested during any one session, although no more than twelve subjects were normally run at a single session. Each subject was seated within a comfortable hearing range of the tape recorder. Brief rest periods were scheduled between each of the four blocks of trials. At the conclusion of the session, subjects were asked to complete a biographical information sheet and questionnaire. The entire experiment lasted approximately 1 hour and 15 minutes.

#### E. METHOD OF ANALYSIS

The input to INDSCAL consisted of 114 (individual subjects) thirteen (stimulus tones) x thirteen (stimulus tones) symmetric data matrices. Each of the 114 individual subject data matrices contained distance estimates for each paired-comparison of the thirteen stimulus tones. The distance estimate was obtained by averaging the four dissimilarity scores for each pair of stimulus tones for each subject, converting these averaged dissimilarity scores into distance estimates, and subsequently into scalar products using standard procedures described in Torgerson (1958). The aim of INDSCAL is to determine, by means of an iterative least squares procedure, the stimulus coordinates and the

subject weights that account for as much total variance in all subjects' data as possible. The output from INDSCAL, consisted of two matrices, a thirteen (stimulus tones)  $\times$   $k$  (dimensions) matrix of coordinates of the thirteen stimulus tones on  $k$  dimensions (plotted in a "group stimulus space"), and a 114  $\times$   $k$  matrix of weights of each of the 114 individual subjects on the same  $k$  dimensions (plotted in a "subject space"). INDSCAL analyses of the subjects' input dissimilarity matrices were performed in several  $k$ -dimensional spaces in order to determine the correct number of dimensions.

### III. RESULTS

#### A. INTERPRETATION OF DIMENSIONS

A 4-dimensional INDSCAL analysis of these dissimilarities data was found to provide the best summary interpretation of the major perceptual structures employed by subjects in making their paired-comparison ratings of dissimilarity between the thirteen stimulus tones. This 4-dimensional solution accounted for 74% of the total variance in the dissimilarities data for all subjects. Since only 4% more variance was accounted for in five dimensions and 9% less was accounted for in three dimensions, it was concluded that four dimensions were both necessary and sufficient. The high correlations (.81 or above) obtained between each individual subject's original dissimilarities data and the distances between the stimulus tones in his associated appropriately-weighted 4-dimensional solution indicate that the private stimulus space for a particular individual may be derived quite reasonably from the group stimulus space by differentially stretching or shrinking the four dimensions.

The four dimensions that emerged from the multidimensional scaling analysis are numbered from 1 to 4 in decreasing order of variance accounted for: dimension 1 = 26.5%, dimension 2 = 20.4%, dimension 3 = 19.8% and dimension 4 = 7.3%. The proportion of variance accounted for by each of the four dimensions indicates that dimension 4 was somewhat weaker in saliency than the other three dimensions. Although dimension 4 accounted for a smaller proportion of the variance, its statistical reliability as well as the reliability of the composite 4-dimensional group stimulus space was confirmed through the computation of correlation coefficients between each of the four dimensions for separate independent split-half (57 subjects) solutions. If the stimulus coordinates between corresponding dimensions of the split-half solutions correlate highly, then we can reasonably conclude that the total group solution is very reliable. The correlations between stimulus coordinates

on the corresponding unrotated dimensions of the split-half solutions were as follows: dimension 1 = .98%, dimension 2 = .99%, dimension 3 = .98% and dimension 4 = .90%. The moderately high replicability of dimension 4 across split-half INDSCAL analyses establishes its statistical reality.

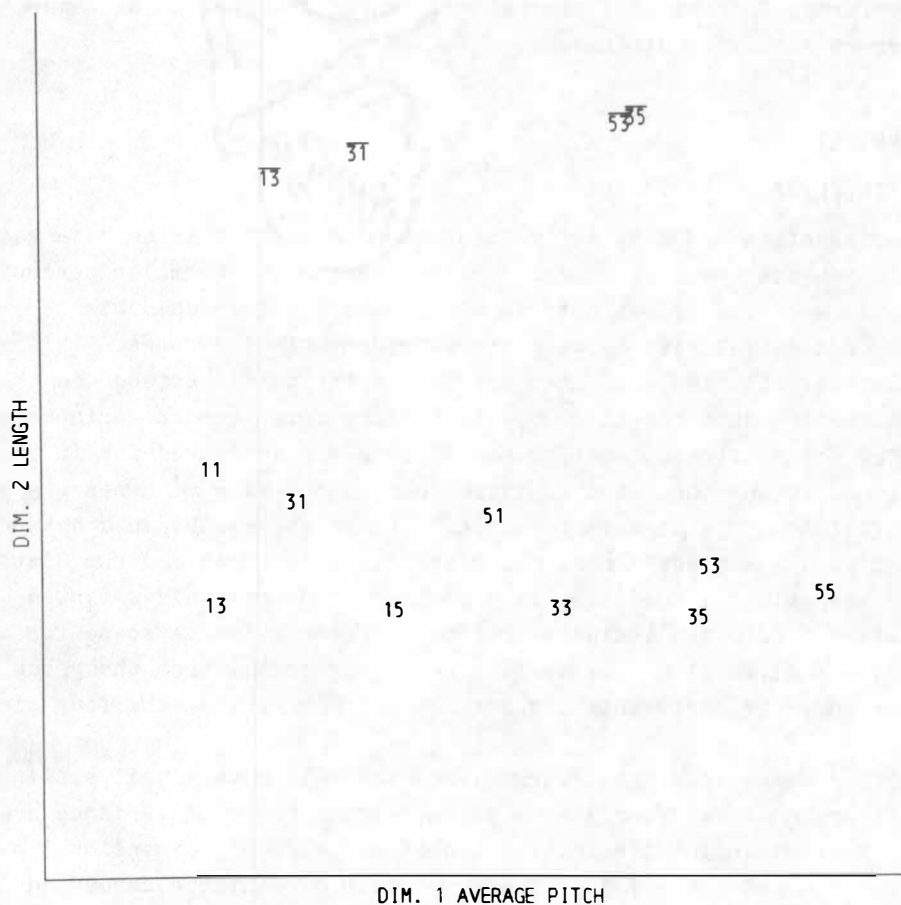


FIGURE 4

Dimensions 1 and 2 from the INDSCAL 4-dimensional group stimulus space.

Figure 4 shows a plot of the coordinates or projections of the stimulus tones on the first and second dimensions in the group stimulus space. Interpretation of the dimensions was guided by inspection of the

order and position of the tones along each dimension, particularly how the tones at one extreme differ from those at the other. It should be pointed out, however, that the locations of the stimulus tones in the 4-dimensional group stimulus space came directly from the INDSICAL analysis, and preceded any particular interpretation assigned to the dimensions.

Dimension 1 is interpreted as AVERAGE PITCH. Stimulus tones 11 and 55 are found at either extreme; 33 near the middle. The clustering of 15 and 51 with 33, and the clustering of (T3 3T 31 13) and (53 35 53 35) lead to the following assignment of average pitch values: 55 = 5, (53 35 53 35) = 4, (15 51 33) = 3, (T3 3T 31 13) = 2, 11 = 1. The second dimension, interpreted as LENGTH, contrasts the nine long tones (11 13 31 15 51 33 35 53 55) with the four short tones (T3 3T 53 35). This dimension is based on the duration, rather than pitch, characteristics of the thirteen synthetic tonal stimuli.

Figure 5 (overleaf) shows a plot of the third and fourth dimensions. Both of these dimensions are based on the pitch characteristics of the stimulus tones. Along the third dimension, interpretively named DIRECTION, rising tones (15 35 13 35 T3) are found at one extreme, falling tones (51 31 3T 53 53) at the other and level tones (33 11 55) near the middle. The fourth dimension, labelled SLOPE, appears to separate the level tones (33 11 55) from the contour tones (51 31 53 3T 53 T3 35 13 35 15). Along this dimension, tones tend to be ordered by the size of the pitch range as determined by the difference in fundamental frequency values between the beginning and ending points of the stimulus tones: (51 15) = 50 Hz, (31 3T 53 53 T3 35 13 35) = 25 Hz, (33 11 55) = 0 Hz. This dimension also appears to be related to the magnitude of slope of the stimulus tones.

The interpretation of the four dimensions was supported by the results of multiple linear regression analysis. Table I (page 287) presents the eight properties that were treated as dependent variables; the stimulus coordinates of the four dimensions in the group stimulus space were treated as independent variables. The regression analysis locates directions in the group stimulus space best corresponding to these properties. Of the eight properties, 1-4 are of a binary nature, 6-8 tertiary and property 5 quinary. Properties 1-3 correspond to Wang's (1967) binary phonological features of tone [ $\pm$ CONTOUR], [ $\pm$ RISING] and [ $\pm$ FALLING], respectively. Property 5 corresponds to duration characteristics of the stimulus tones, properties 5-8 various pitch characteristics (cf. Section II.8).

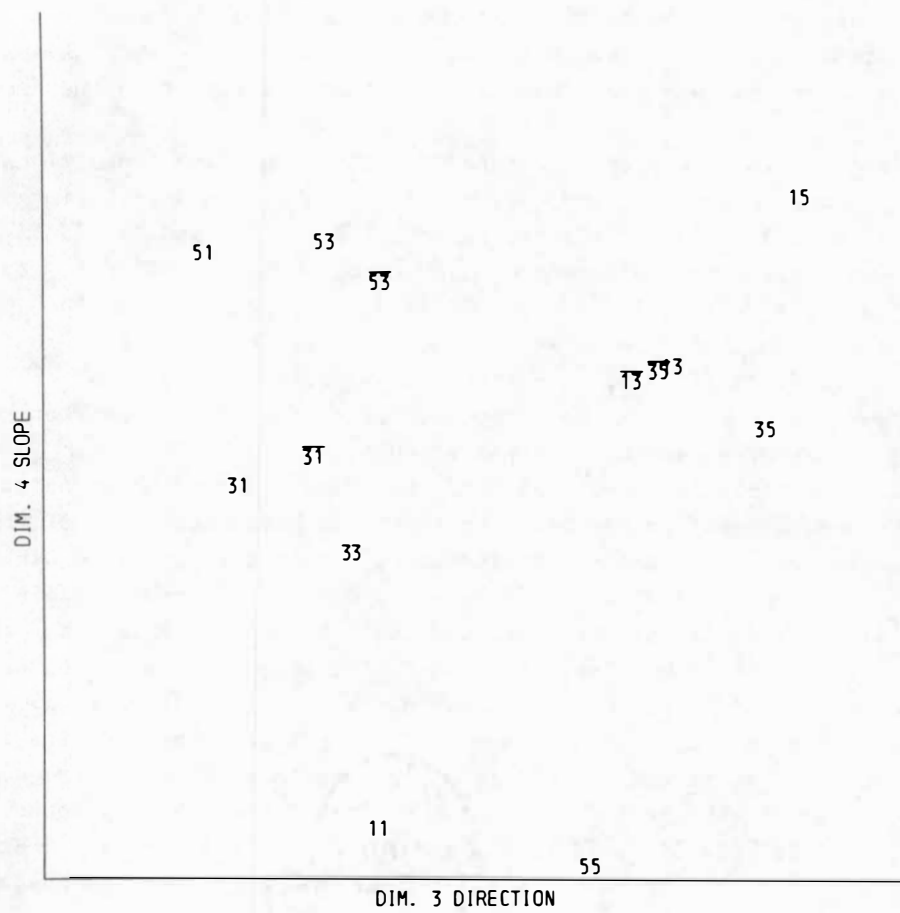


FIGURE 5

Dimensions 3 and 4 from the INDSICAL 4-dimensional group stimulus space.

	<u>11</u>	<u>33</u>	<u>55</u>	<u>53</u>	<u>31</u>	<u>51</u>	<u>35</u>	<u>13</u>	<u>15</u>	<u>53</u>	<u>31</u>	<u>35</u>	<u>13</u>
1. Level v. Contour	0	0	0	1	1	1	1	1	1	1	1	1	1
2. Rising v. Nonrising	0	0	0	0	0	0	1	1	1	0	0	1	1
3. Falling v. Nonfalling	0	0	0	1	1	1	0	0	0	1	1	0	0
4. Long v. Short	0	0	0	0	0	0	0	0	0	1	1	1	1
5. Average Fundamental Frequency	100	125	150	137.5	112.5	125	137.5	112.5	125	137.5	112.5	137.5	112.5
6. Magnitude of Slope	0	0	0	1	1	2	1	1	2	2	2	2	2
7. Amount of Change in Fundamental Frequency (Hz)	0	0	0	25	25	50	25	25	50	25	25	25	25
8. Rising v. Level v. Falling	0	0	0	1	1	1	2	2	2	1	1	2	2

TABLE I

Values associated with the thirteen stimulus tones for each of the eight properties used in the linear regression analysis.

The associated linear regression results are shown in Table II. The multiple correlations in the last column show how well each of the properties can be predicted from the four INDSGAL dimensions. The normalised regression weights show the relative importance of each dimension for predicting order and position of stimulus tones for a particular property.

	<i>Regression Weights</i>				Multiple Correlation
	Dim. 1	Dim. 2	Dim. 3	Dim. 4	
1. Level v. Contour	-.124	-.235	.115	.957*	.882
2. Rising v. Nonrising	-.240	-.070	.925*	.288	.958
3. Falling v. Nonfalling	.151	-.114	-.868 <sup>#</sup>	.459	.901
4. Long v. Short	-.055	.988*	-.135	-.051	.986
5. Average Fundamental Frequency	.981*	-.008	.157	.111	.980
6. Magnitude of Slope	-.086	-.584	.131	.796 <sup>#</sup>	.933
7. Amount of Change in Fundamental Frequency	-.158	.094	.053	.982*	.864
8. Rising v. Level v. Falling	-.224	-.167	.677 <sup>#</sup>	.681 <sup>#</sup>	.936

TABLE II

Linear regression of eight properties on 4-dimensional space for Thai subjects' dissimilarity ratings data. \* = regression weight .925 and above; # = regression weight between .675 and .925.



Figures 6 and 7 show the geometric representation of the linear regression results in Table II. In this representation, the direction and extent of the straight (dashed) line corresponding to a property indicates which of the INDSCAL dimensions most nearly corresponds to that property.

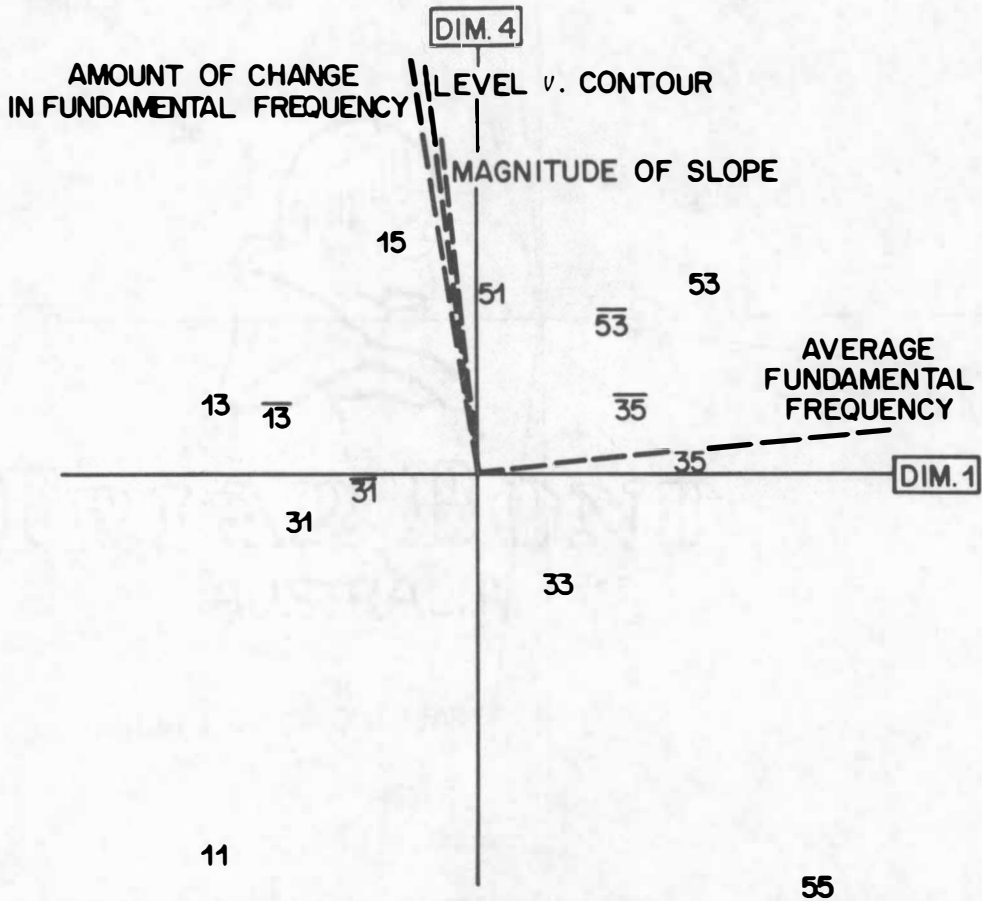


FIGURE 6

The dimension 1 - dimension 4 plane from the INDSCAL 4-dimensional group stimulus space. Projections of straight (dashed) lines optimally corresponding to amount of change in fundamental frequency, level versus contour, magnitude of slope and average fundamental frequency.

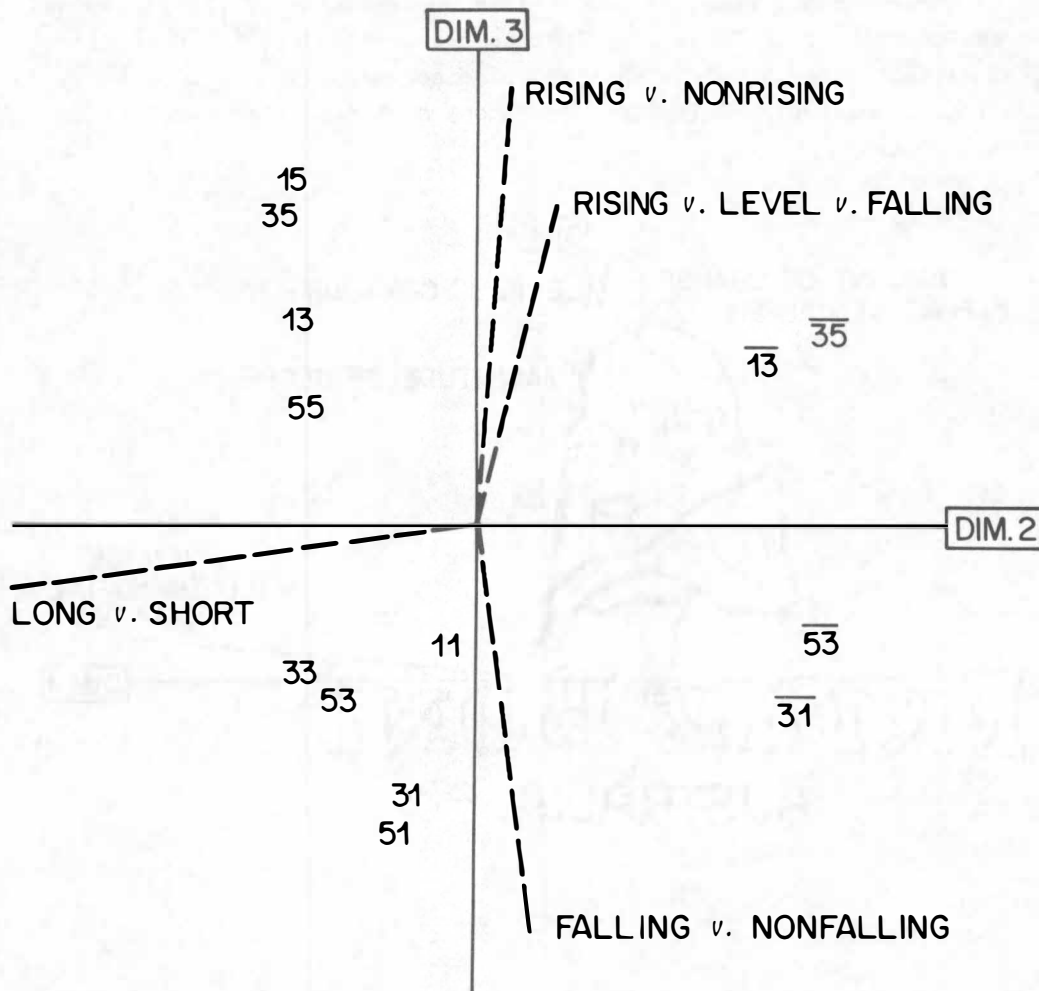


FIGURE 7

The dimension 2 - dimension 3 plane from the INDSCAL 4-dimensional group stimulus space. Projections of straight (dashed) lines optimally corresponding to rising versus nonrising, rising versus falling, long versus short and falling versus nonfalling.

The direction and extent of the straight lines optimally corresponding to these properties in the multidimensional space support our interpretation of the four dimensions. In Figure 6, dimension 1 closely corresponds to average fundamental frequency, dimension 4 to amount of change in fundamental frequency, level versus contour and magnitude of slope; in Figure 7, dimension 2 closely corresponds to long versus short, dimension 3 to rising versus nonrising, falling versus nonfalling and rising versus level versus falling. The third dimension, however, does

suggest that rising (property 2) pitch movement is represented more directly in terms of subjects' perceptual dimensions of tone than falling pitch movement. It is of further interest to note that the clustering of stimulus tones on the third and fourth dimensions lead to a close correspondence to Wang's binary phonological features of tone (properties 1-3).

#### 8. SUBGROUP DIFFERENCES IN WEIGHTS FOR DIMENSIONS

Subjects' weights on the AVERAGE PITCH and LENGTH dimensions are indicated in Figure 8. The relative salience of AVERAGE PITCH is greater than LENGTH for most of individual Thai subjects as illustrated by the tendency for subjects to cluster in the middle-right area of the plot.

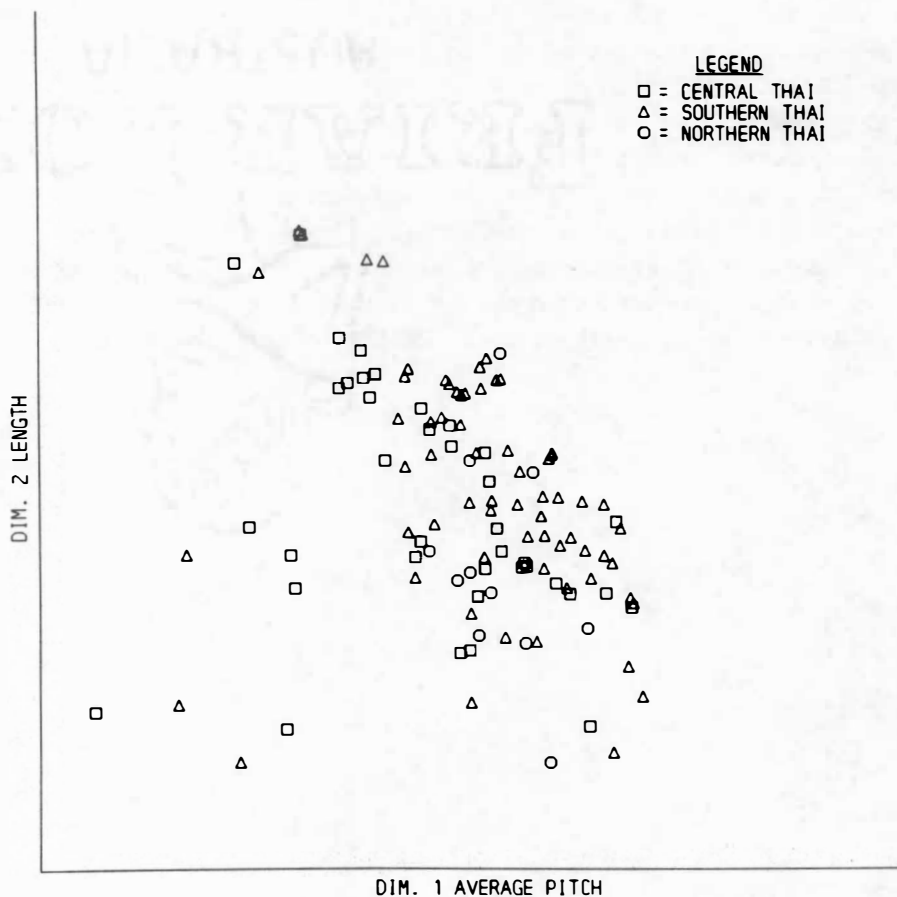


FIGURE 8

Dimensions 1 and 2 from the INDSICAL 4-dimensional subject space.

A 1-way analysis of variance of the *mean* subject weights on each of the dimensions across the three Thai dialect subgroups, however, does indicate a significant difference on dimension 1 - AVERAGE PITCH between the Central Thai and Southern Thai subgroups ( $F = 7.08$ ,  $df = 1$ ,  $p < .01$ ). No significant differences were obtained between Thai dialect subgroups on the second, third and fourth dimensions. Though the mean weights for dimension 1 differ significantly between Central Thai and Southern Thai, the considerable overlap between dialect subgroups calls for caution in interpreting this apparent difference in the perceptual spaces of these two dialect subgroups (cf. Brown 1965:59). The essential homogeneity of the perceptual space for the Thai group can be attributed, at least in part, to the fact that both Northern Thai and Southern Thai subjects were educated bidialectals who had had several years exposure to Central Thai; and the possibility that the differences in the phonological inventory of tones between the three dialect subgroups were simply not great enough to induce differences in the perceptual space (for description of tones of Thai dialects, see Brown 1965). For similar apparent reasons, Fox's (1974) multidimensional scaling analysis of vowel perception failed to demonstrate a difference in the perceptual spaces utilised by two dialect subgroups of American English.

The essential homogeneity of the perceptual space for the Thai group is further shown in Figure 9. Subjects' weights on the third and fourth dimensions tend to cluster in the lower-right area of the plot indicating relatively greater salience for DIRECTION than for SLOPE.

(See FIGURE 9 next page.)

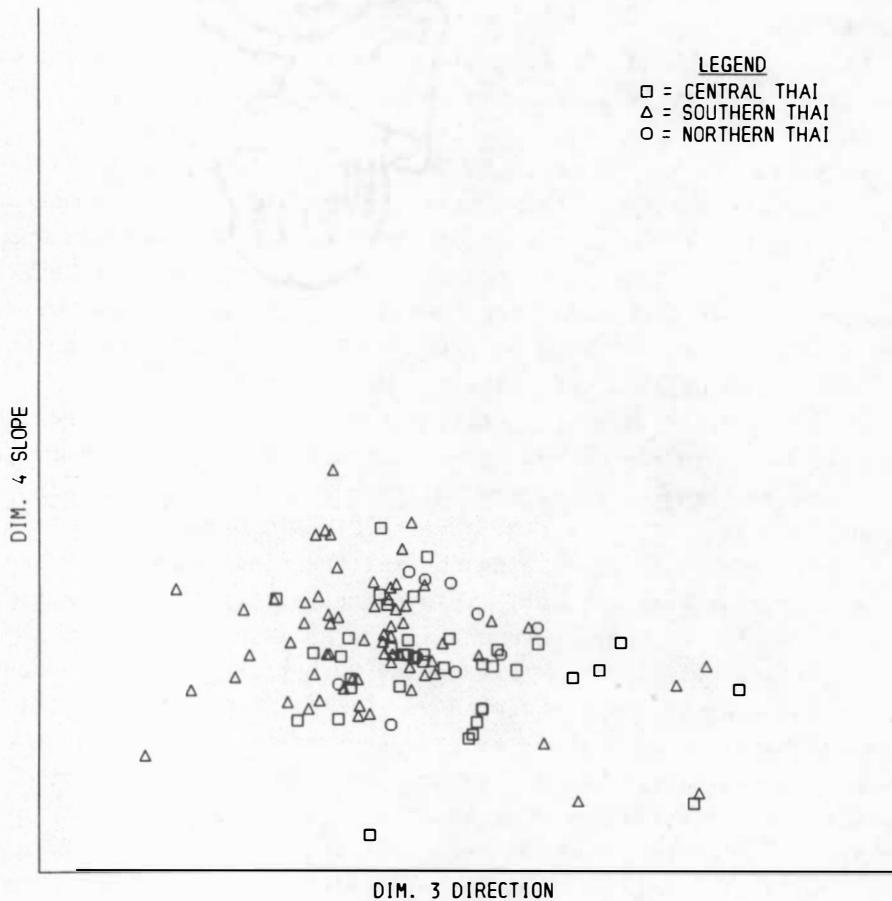


FIGURE 9

Dimensions 3 and 4 from the INDSICAL 4-dimensional subject space.

#### IV. DISCUSSION AND CONCLUSIONS

Our data indicate that four dimensions underlie the subjects' dissimilarity ratings of the thirteen synthetic stimulus tones. Of the four dimensions, AVERAGE PITCH was the most important factor that subjects incorporated into their perceptual judgments. This dimension would appear to be a lower-level auditory dimension; its utilisation at higher levels of linguistic processing is certainly not indicated by the

phonological inventory of Thai tones, or for that matter, the lexical tones of any tone language. Hombert's (1976) multidimensional scaling analysis of the perception of tones of bisyllabic nouns in Yoruba (tone language spoken in West Africa), interestingly enough, also extracted a dimension that was, in part, related to the average fundamental frequency value of the vowel in the second syllable.

The emergence of the nontonal LENGTH dimension as a relatively important factor in Thai subjects' dissimilarity judgments is not at all surprising in view of the structure of the stimulus set, and perhaps more importantly, the linguistic function of vowel duration in Thai. The contrast between long and short vowels (e.g. /hàat/ 'shoal' v. /hàt/ 'to practice') more than likely heightened the Thai subjects' sensitivity to this particular physical property of the stimuli.

The tonal third and fourth dimensions too seem to be of a more linguistic nature. Both may be related to pitch characteristics that are not only used to signal phonological distinctions in Thai, but also other tone languages of the world. On the DIRECTION dimension, it may be observed that neither the rising nor falling tones cluster according to the beginning or ending point. This suggests that the Thai subjects perceived direction of pitch movement in the stimulus tones, not as movement *from* a fixed point A *to* a fixed point B, but instead as movement in a direction *away from* A and *toward* B. On the SLOPE dimension, it may be observed that the level tones tend to separate from the contour tones. This dimension may be related to Abramson's (1962) proposed "static" versus "dynamic" classification of the Thai tones. In his Yoruba study, Hombert too found dimensions that were principally related to the direction and amount of change in fundamental frequency. The convergence in results from these two studies clearly strengthens the claim for the psychological reality of these dimensions.

The results of a linear regression analysis supported our interpretation of the dimensions. Of particular interest is the close correspondence obtained between the third and fourth dimensions and Wang's unit-contour tone features. There is no physical property of the stimulus set that would bias the obtained result; indeed, one might have predicted the emergence of dimensions on which the stimulus tones clustered according to begin point and/or end point. The fact that the decomposition of the contour stimulus tones into begin point and end point apparently had little influence on Thai listeners' perceptual judgments converges nicely with the traditional classification of Thai as a "contour tone language" (Pike 1948, Ladefoged 1975).

Contrary to expectation, individuals did not differ much across dialect subgroups. This, of course, does not mean that dialect member-

ship may not have an influence on one's tone perception. The dominant influence of the national language could be effectively eliminated or restricted through the use of monodialectal speakers of Northern Thai and Southern Thai dialects. With monodialectal speakers, we might then be able to furnish more direct experimental evidence bearing on Brown's (1965:59) rather appealing suggestion that "contour" is more important than "register" for Central Thai and Northern Thai dialects, but that "register" is more important than "contour" for Southern Thai dialects. Individual differences in tone perception have been demonstrated across language groups - Thai, Yoruba and American English (Gandour and Harshman 1978).

The perception of tone deals with how a listener transforms, organises and structures the pitch information arising from the speech signal. The nature of the dimensions that emerged from the INDSCAL analysis are clearly consistent with a proposal that information-bearing aspects of pitch are organised in terms of simple oppositions (binary or n-ary) along a number of independent dimensions rather than in terms of a number of steps along one or a few dimensions. Of the four dimensions, the second, third and fourth are apparently utilised in signalling linguistic categories in Thai, and consequently, show a greater tendency toward binarity. These dimensions, though obviously not exhaustive of all possibilities, represent a beginning toward a definition of a necessary and sufficient set of perceptually-based features of tone.

REFERENCES

ABRAMSON, A.S.

- 1962 'The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments'. *International Journal of American Linguistics* 28/2, Part II.

BROWN, J.M.

- 1965 *From Ancient Thai to Modern Dialects*. Bangkok: Social Science Association Press of Thailand.

CARROLL, J.D. and J.J. CHANG

- 1970 'Analysis of Individual Differences in Multidimensional Scaling via a n-way Generalization of "Eckart-Young" Decomposition'. *Psychometrika* 35/3:283-319.

CARROLL, J.D. and M. WISH

- 1974a 'Multidimensional Perceptual Models and Measurement Methods'. In: E.C. Carterette and M.P. Friedman, eds. *Handbook of Perception*, vol. II: *Psychophysical Judgment and Measurement*. 391-447. New York: Academic Press.
- 1974b 'Models and Methods for Three-way Multidimensional Scaling'. In: D.H. Drantz *et al.*, eds. *Contemporary Developments in Mathematical Psychology*, vol. II. 57-105. New York: W.H. Freeman.

CHANG, J.J. and J.D. CARROLL

- 1972 How to use PROFIT, a computer program for PROperty FITing by optimizing nonlinear or linear correlation. Bell Laboratories, Murray Hill, New Jersey.



CHAO, Y.R.

1930 'A System of Tone Letters'. *La Maitre Phonétique* 45/24-7.

FOX, R.A.

1974 An Experiment in Cross-dialect Vowel Perception. Papers from the 10th Regional Meeting of the Chicago Linguistic Society. 178-85.

GANDOUR, J.

1975 'On the Representation of Tone in Siamese'. In: J.G. Harris and J.R. Chamberlain, eds. *Studies in Tai Linguistics*. 170-95. Bangkok: Central Institute of English Language.

GANDOUR, J. and R. HARSHMAN

1978 'Crosslanguage Differences in Tone Perception: a Multidimensional Scaling Investigation'. *Language and Speech* 21/1-33.

HARSHMAN, R.A.

1970 'Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multidimensional Factor Analysis'. *UCLA Working Papers in Phonetics* 16.

1972 'Determination and Proof of Minimum Uniqueness Conditions for PARAFAC 1'. *UCLA Working Papers in Phonetics* 22/11-117.

HENDERSON, E.J.A.

1949 'Prosodies in Siamese, a Study in Synthesis'. *Asia Major* (new series) 1/189-215. Reprinted in: F.R. Palmer, ed. *Prosodic Analysis*. 1970:27-53. London: Oxford University Press.

HOMBERT, J.M.

1976 'Perception of Tones of Bisyllabic Nouns in Yoruba'. *Studies in African Linguistics*, Supplement 6. 109-21.

LADEFOGED, P.

1975 *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

PIKE, K.L.

1948 *Tone Languages*. Ann Arbor: University of Michigan Press.

RICE, L.

- 1971 'A New Line Analog Speech Synthesizer for the PDP-12'. *UCLA Working Papers in Phonetics* 17/58-75.

TORGERSON, W.S.

- 1958 *Theory and Methods of Scaling*'. New York: Wiley.

WANG, W. S-Y.

- 1967 'Phonological Features of Tone'. *International Journal of American Linguistics* 33/93-105.

WISH, M. and J.D. CARROLL

- 1974 'Applications of Individual Differences Scaling to Studies of Human Perception and Judgment'. In: E.C. Carterette and M.P. Friedman, eds. *Handbook of Perception* vol. II: *Psychophysical Judgment and Measurement*. 449-91. New York: Academic Press.

BWE KAREN AS A TWO-TONE LANGUAGE?  
AN ENQUIRY INTO THE INTERRELATIONS OF PITCH,  
TONE AND INITIAL CONSONANT

EUGÉNIE J.A. HENDERSON

1.1 When Professor Gordon Luce directed my attention to Western Bwe Karen many years ago, he pointed out that the special interest these dialects hold for Sino-Tibetan linguists is the fact that they are a living testimony to one of the principal tenets of Sino-Tibetan linguistics: namely, that the loss of an earlier distinction between voiced and voiceless stops has resulted in the doubling of the original number of tones in many languages in the area. It is assumed that pitch differences correlated with the presence or absence of voice, which were phonologically non-distinctive as long as the voice/voiceless contrast survived, became phonemic themselves once the voice contrast was lost. The better known Sgaw and Pwo dialects of Karen, in which there is no longer a phonemic opposition between voiced and voiceless stops, are commonly described as having six tones (four for Bassein Pwo, see Jones 1961); whereas Western Bwe, which preserves the opposition between voiced and voiceless stops, is described as having only three tones. At the time, the only "modern" publication on Karen was Haudricourt's celebrated 1946 paper, in which he accepted the traditional 6 tone analysis of contemporary Pwo and Sgaw, but divided the 6 tones into a high and low series, with 3 tones in each, systematically associated, as in Chinese and Tai, with features of the initial consonants. By comparing these two dialects, and without benefit of knowledge of Western Bwe, he postulated a 3-tone system for Proto-Karen, and a 3-way system of initial consonants: voiceless unaspirated, voiceless aspirated and voiced. Since Haudricourt's third tone was associated with stopped syllables

only, like the Chinese ch'ü sheng, this left a 2-tone system for un-stopped syllables. Nevertheless, some consternation was expressed when, after working for some time with the Western Bwe dialect which has no stopped syllables to complicate the issue, I suggested in a seminar paper that there are clear indications that in this dialect the underlying (synchronic) phonological contrast is between 2 rather than 3 tones.

1.2 In recent years, the case for the reconstruction of Proto-Karen and of Sino Tibetan itself as 2-tone languages has been further and cogently argued by such scholars as Benedict and Jones.<sup>1</sup> It seems appropriate therefore, to present here the evidence pointing towards a 2-tone interpretation of contemporary Bwe. I must emphasise that the evidence presented is purely synchronic: I leave its historical interpretation to colleagues better versed than I in this field. I have however attempted to correlate the Bwe forms in my material with the forms cited by Benedict in the revised and expanded version of his paper to the Second Sino-Tibetan Conference (Benedict 1969), and with those cited by Jones in his paper to the Fourth Sino-Tibetan Conference (Jones 1971), and with the hypotheses put forward by Haudricourt (1946 and 1953).

1.3 My research and field work were mainly concerned with what is sometimes referred to as the Blimaw dialect of Western Bwe, but I was also able to do a little work on the closely related Geba dialect, and shall draw upon this, and upon Luce's Geba word lists (Luce 1959, and further material at present in the press), when appropriate.

2.1 It is beyond dispute that from the phonetic point of view Blimaw exploits three pitch levels for lexical purposes. These are high level (pitch 1), and mid level (pitch 2), and low level (pitch 3), as in  $l\epsilon^1$  'moon',  $l\epsilon^2$  'leaf' and  $l\epsilon^3$  'to keep, conserve' - an excellent prima facie case for postulating three tones, it would seem. A closer examination of the distribution of these pitches in relation to initial consonants and of other factors affecting pitch in this dialect - such as tone-sandhi, the use of what I have called tone-dissimilation (Henderson 1961 and 1967) in the word-compounding process, and the use of pitch-raising as a syntactic marker - soon demonstrates, however, that the picture is not quite so simple as it at first appears. On the other hand, some of these factors offer explanations of forms which at first glance appear tonally deviant from a historical point of view. I have

described a number of these pitch-determining factors in some detail in Henderson (1961) and so will refer to them only briefly here. I cannot, however, forbear to point out how misleading it can be to select for comparative purposes the odd cognate syllable from a dissyllabic or polysyllabic word without being aware of the possible effect of such factors on the pitch of the syllable concerned. A case in point is the apparent tonal irregularity of the Blimaw words for 'tail', 'tooth', and 'pot' in Jones' lists (see section 6 below). The full Blimaw word for 'tail' is not  $me^3$  but  $ka^1me^3$  ( $ka^1$  = 'bottom'), that for 'tooth' is not  $me^3$  but  $\theta o^1me^3$ . In both instances the low pitch of the final syllable, which Jones seems to find irregular, is the perfectly regular realisation on the second syllable of a disyllabic compound of this kind of a common type of high-low word-intonation pattern, and is thus not necessarily to be equated with the pitch to be expected of the component morphemes in isolation. From the point of view adopted in this paper I do not regard these forms as tonally irregular. See Section 5.1.6. Similarly, the (to Jones) "irregular" high pitch on  $bo^1$  in 'pot' is the realisation of a second dissimilated pattern, low-high - the full form of the Blimaw word for 'pot' being  $g\delta^2bo^1$  or  $g\delta^3bo^1$ .<sup>3</sup>

2.2 Below, at 2.4 will be found a table showing the distribution of the Western Bwe (Blimaw) syllables occurring in my material, arranged according to vowels, consonants, and pitches. All Western Bwe syllables are of the CV type. There are no final consonants, no diphthongs, and no distinctive vowel nasalisation.<sup>4</sup> There may be a glide (r, R or w) between C and V; since the pitch distribution of such syllables does not differ significantly from that of syllables without glides, they are omitted from the table for the sake of simplicity. As observed elsewhere (Henderson 1965), *all* mid and low pitch syllables before a pause may exhibit glottal constriction finally (i.e. a weak glottal stop); high pitch syllables before a pause never do so, but an h-like off-glide is sometimes audible. (The universality of these pre-pause syllable-closing features in Bwe prohibits their interpretation as echoes of formal final consonants, now vanished, but may nevertheless be relevant to the history and development of the tones.)

2.3 The symbols in the Table, where not self-explanatory, are to be interpreted as follows:

✓ marks the occurrence of a syllable as a full, "normal" monosyllabic word, e.g. as a noun, pronoun, or verb, etc.

1 indicates that the syllable is found as the first (bound) syllable

of a disyllabic or polysyllabic word or compound, but is not found as an independent word. For the purposes of this paper, the term "compound" is used to include the very common type of "double" construction in which one of the elements does not occur independently, e.g.  $\theta i^2$ - $ja^2$ , as in  $m\epsilon^2\theta i^2m\epsilon^2ja^2$  'to be able to do s.' in which  $m\epsilon^2$  'to do' and  $ja^2$  'to be able' occur as independent verbs, but  $\theta i^2$  does not. (There are independent words  $\theta i^2$  'comb',  $\theta i^2$  'dental decay' and  $\theta i^2$  'to thread on a string', but it doesn't make sense to attempt to identify any of these with the first element of  $-\theta i^2 - ja^2$ .)

2 indicates that the syllable concerned is found only as the second (bound) element in a disyllabic word or compound. The numerals 1 and 2 together (e.g. 12) indicate that the syllable concerned is found both as the first and as the second element in disyllabic words or compounds, but that it is not found as an independent word.

3 (one instance only) indicates that the syllable is found only as the third element in a polysyllabic word or compound. Information about the position of a syllable within a polysyllabic word is especially important in this language because of the effect polysyllabic word-patterns can have upon the pitch of component syllables.

C means that the syllable is only found as a numeral classifier.

E indicates a syllable found as an exclamation only.

F marks a loan-word (usually from Burmese). There are doubtless many unidentified loan-words in the material.

O indicates an onomatopoeic or phonaesthetic word of some kind.

O1 and O2 indicate that the syllable is found only as the first or second element respectively in an onomatopoeic or phonaesthetic expression.

P marks syllables which only occur as affixes, particles or particle-like words which are highly susceptible in this language to pitch variation that can be associated systematically with syntactic, morphological and emotional factors. P2 means 'only found as a particle OR as the second syllable of a disyllable', and so on.

Q indicates a question-word used (*how much, how many, how narrow, how small, what kind, etc.*) in wh-type questions (and often in the answers to them) which characteristically bears a high pitch as part of an overall interrogative intonation pattern. Q1 means 'only found in question words OR as the first syllable of a disyllable', and so on.

2.4 TABLE SHOWING DISTRIBUTION OF PITCHES

		1. HIGH PITCH						2. MID PITCH						3. LOW PITCH													
		i	ɪ	e	ɛ	a	ɔ	o	u	i	ɪ	e	ɛ	a	ɔ	o	u	i	ɪ	e	ɛ	a	ɔ	o	u		
1.	ʔ	✓	✓	E	✓	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
2.	p	F1	F1	12	F	✓	✓	✓	✓	2	F	✓	✓	✓	2	✓	✓										
3.	t	3	Q	✓	1	1	✓	✓	✓	2	O1	✓	✓	✓	✓	✓											
4.	c	1	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
5.	k	✓	1	✓	✓	✓	✓	✓	1	✓	✓	✓	✓	✓	✓	✓	✓										
6.	β	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
7.	ɗ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
8.	ph	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
9.	th		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓							2				
10.	ch	✓	Q							F	✓					✓											
11.	kh	✓	✓	1	✓	✓	✓	✓	✓	✓	2	✓	✓	✓	✓	✓	✓										
12.	h	1	✓	✓	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
13.	ɟ	✓	✓	F1	✓	✓	1	✓	✓	✓	✓	✓	✓	✓	✓	✓											
14.	θ	✓		✓	✓	F12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓											
15.	x				✓	F1				✓	O1	✓	O2	O2			O	✓									
16.	b	F	P	O	P	Q	✓			1	1	✓	✓	✓	✓	✓	12	2	✓	✓	1	✓	✓	1	1		
17.	d		Q	1	C	Q	C	P	1	✓	2	✓	✓	✓	✓	✓	✓	1	2	✓	✓	1	✓	1	✓	✓	
18.	j			F	1	2	E	✓			✓	✓	✓	✓	✓	✓	✓	✓			✓	1	✓	✓	✓		
19.	g			P	P	2		P	1		✓	✓	✓	✓	✓	2	✓	✓	O1	✓	✓	✓	O	✓	✓		
20.	m	✓	P	✓	✓	✓	✓	1	✓	✓	✓	✓	✓	✓	✓	✓	✓	2	✓	2	✓	✓		✓	1		
21.	n	✓	Q1	✓	✓	Q1	P	P	✓	✓	✓	P2	✓	✓	✓	P	✓				✓	✓	✓	✓	✓		
22.	l	✓	✓	✓		✓	✓	✓	2	✓	12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	C	✓	✓	✓	✓	
23.	r		✓	✓		O2		✓		✓	1	1	✓	✓	2	1	12	✓				✓		✓	O		
24.	R											✓	✓	✓	✓					✓	✓		✓	✓	✓		
25.	?w	✓								✓	2																
26.	w	2	✓	✓	✓					✓	✓	✓	✓	✓		2	1	✓	✓	✓	✓	✓		✓	1		
27.	?y				✓	2					✓	O	F	O													
28.	y			✓	1	✓	✓	✓		F1	✓	✓	✓	✓	✓		✓	12	✓	✓	✓	✓	✓	✓	✓		
29.	#																								P	E	2

3.1 The pattern for rows 1-14 is immediately apparent. The initials concerned comprise voiceless aspirated and unaspirated stops, glottalised stops (including ?) and voiceless "aspirates" or "hissers": h, ʃ and θ. Syllables beginning with these consonants are never found on a low pitch except as the result of a very common and perfectly regular type of tone-sandhi (described in more detail in Henderson 1961) whereby all mid pitch syllables following a low pitch syllable are lowered until the end of the phrase or breath-group, or until the next high pitch syllable.

The one apparent exception to this rule in my material is the syllable  $tha^3$ , which occurs in  $bwe^2tha^3$ , an impolite name used to describe the Geba-speaking Karens living near the great watershed beyond Than-daung. My informant said the English semantic equivalent would be something like *"wild men of the hills"*. His suggested explanation of the lowered pitch of  $tha^3$  was that the name is probably a contraction of the phrase  $bwe^2go^3tha^3$  'the people above' - cp.  $jəbwe^1 ə-go^3tha^3$  'above, over, the table', as contrasted with  $jəbwe^1 ə-kho^2$  'on the table'. In  $go^3tha^3$  'above', the second element is probably to be associated with  $tha^2$  'to rise, go up', its lowered pitch being the predictable outcome of tone-sandhi after  $go^3$ . This seems a reasonably plausible explanation.

It seems clear that for syllables with voiceless and glottalised initials there is only a 2-way pitch contrast, high v. mid. Note that this also holds good for preglottalised ?w and ?y in rows 25 and 27 respectively.

Certain oddities in the distribution of voiceless consonants may be noted. *th* and *ch* appear to be in complementary distribution, *ch* appearing before the close front vowels *i* and *ɪ*. The form  $chu^2$  'dog', which appears to be an exception, is in free variation with the form  $thwi^2$ , and offers an example of that curious fluidity of features within syllables that James Matisoff (1973) has noted as a characteristic of the languages of the area.

Initial *x* (in row 15) is rare and appears to be an exception to the rule for voiceless consonants. I am in some doubt as to whether it can be regarded as a genuinely Bwe Karen consonant. Some of the few words in which it occurs seem to be loans from Sgaw, e.g.  $xu^1$  'six', and  $xo^3$  'eight'.  $xu^1$  and  $xo^3$  were used by my informants in counting series and in certain compound numerals, but the everyday words for 6 and 8 were  $θə^1θv^2$  and  $lwi^1θv^2$ , literally '3 pairs' and '4 pairs'.  $xɪ^2$  'to scream, yell' is perhaps phonaesthetic; similarly  $xε^2xo^2$  'to be clumsy and slow', which is in free variation with  $khe^2kxə^2$ ; and  $xε^3xε^3$  'slowly', which is in free variation with  $kre^2kre^2$ . That leaves us with  $xa^2$  'to measure (rice) in baskets', and  $xa^1$  'the same', still unaccounted for.



3.2 Moving from the voiceless initials to the voiced stops in rows 16-19, one finds syllables spread over all these pitches, but whereas straightforward monosyllabic nouns, and verbs, etc. preponderate in the mid and low pitch columns, under the high pitch column almost all the instances recorded are in forms particularly susceptible to pitch modification as a mark of syntactic or "expressive" function. It is, of course, only too easy to "cheat" when making an analysis of this kind in which one is following one's linguistic intuitions to a large extent. I have, however, honestly tried to apply the same (admittedly subjective) criteria to all entries in the table. Thus the absence of anything but ticks in rows 6 and 7 means that there are clear and incontestable examples of "ordinary" native monosyllabic words for each entry. The assortment of Fs, Ps, Qs etc. for rows 16-19 in the high pitch column means that the only instances of high pitch syllables with these initials were in -

- (a) loanwords (such as bi<sup>1</sup> 'opium', je<sup>1</sup> 'bazaar' < Burmese);
- (b) particles used at the beginning or end of sentences and regularly subject to pitch-raising or lowering as an intonational feature, such as be<sup>1</sup> and ge<sup>1</sup>;
- (c) a class of words used in wh- questions, the intonation of such sentences always comprising a high pitch on the question-word followed by a low falling pitch on the matching sentence-final particle;
- (d) exclamations and vocatives, which may have high pitch among their possible realisations;
- (e) the first or second syllable of disyllabic words, which frequently in Bwe have a word-intonation pattern involving tonal dissimilation, i.e. either low-high or high-low.
  - cp. la<sup>2</sup> 'to descend', but la<sup>1</sup>de<sup>3</sup> 'to fall'
  - ɔ<sup>2</sup> 'to speak', but ɔ<sup>1</sup>ʃa<sup>2</sup> 'to tell'
  - ca<sup>2</sup> 'to see', but ca<sup>1</sup>le<sup>2</sup>
  - ca<sup>1</sup>le<sup>3</sup> 'to search'
  - je<sup>2</sup> 'to be soft', but je<sup>1</sup>pho<sup>2</sup> 'to be young and tender'
  - ni<sup>1</sup> 'to obtain', me<sup>1</sup> 'wife',
  - but ni<sup>2</sup>me<sup>1</sup> 'to marry'
- (f) numeral classifiers or quantifiers which appear sometimes as "raised pitch" versions of mid pitch nouns,
  - cp. de<sup>2</sup> 'thing, object'
  - ə-de<sup>1</sup>ʃi<sup>1</sup> 'ten things'

(g) syntactic markers like  $du^1$  'when, till',  $bo^1$  'until, so that',  $ge^1$  'after all', which, whatever their initial consonant, are regularly pronounced with very high pitch frequently accompanied by lengthening of the vowel and a following pause, all of which features serve to mark off very clearly the major sentence constituents in co-ordinate and subordinate constructions. Such words are also found with initials from the row 1-14;  $do^1$ , for example, is one of the commonest syntactic markers of this kind. The reason there is no entry P under the o column in row 7 is because there are also monosyllabic nouns and verbs like  $do^1$  'village',  $do^1$  'to let, allow',  $do^1$  'to swell'. There are no nouns and verbs homophonous with the forms marked 'P' in rows 16-19.

The two entries marked with / in these rows are (a) the inalienable locative form  $-bu^1$ , 'inside', as in  $nə-cu^2bu^1$  'in your hand', also  $dε^2-bu^1$  'hole'; and (b) the form  $jo^1$  used after verbs to express the notion of 'mutual aid', e.g.  $mε^2jo^1$  (cp.  $mε^2$  'to do') 'to give someone a hand with something',  $a^2jo^1$  (cp.  $a^2$  'to eat') 'to help someone to eat'. A high pitch is not regularly predictable for other words of similar function, so that I feel unable to call upon intonation, syntactic or otherwise, to explain these forms away. And yet there is some indication in the case of  $-bu^1$ , which never stands on its own, and is never first syllable in a word, that a "step-up" word-intonation pattern may be involved. In some Bwe villages the form of the word is  $-bu^2$ , with mid pitch; and it is to be remarked that the quantifier for  $dε^2-bu^1$  'hole' is  $bu^2$ , which must surely be related, e.g.  $dε^2bu^1ki^1bu^2$  'two holes'. Compare also Geba  $-bu^2$  'inside'.

Even allowing for one or two unexplained exceptions, the general pattern for the pitch alternations proper to syllables with initial voiced stops is markedly different from that of voiceless or glottalised consonants. It is, I think, fair to postulate the regular pitch choice for such syllables as being between mid and low.

3.3 Rows 20-22 present yet a third pattern. Nouns and verbs are scattered pretty evenly over all three pitches in Blimaw. If, however, we compare Blimaw words in these rows with their Geba counterparts, in many instances the Geba forms have voiceless  $hm$ ,  $hn$  or  $hl$ , where Bwe has  $m$ ,  $n$  or  $l$  on high or mid pitch. It is clear therefore that the Blimaw high and mid columns represent a merging of voiceless and voiced sonorants, and one may postulate an underlying / $hm$ -  $hn$ -  $hl$ -/, behaving quite regularly as regards pitch. The relation of underlying initials and pitches may be arranged thus:

Pitch	1	2	3
	/hm/	/hm/	
		/m/	/m/
	/hn/	/hn/	
		/n/	/n/
	/hl/	/hl/	
		/l/	/l/

Where we have words beginning with m, n or l in Blimaw on a high pitch we shall expect to find Geba cognates with hm-, hn- and hl-. Where Blimaw has words beginning with m, n or l on a mid pitch we shall expect to find Geba equivalents with either hm-, or m-, hn- or n-, hl- or l-. As far as my limited Geba material goes, this seems generally to be the case, though if the records are to be relied upon, there are exceptions which require further investigation.

3.4 In the remaining rows 23-28 we have a mixed bag of glides, most of them of rather rare occurrence. Preglottalised ?w and ?y have already been seen to conform with the pitch rules for glottalised consonants, i.e. pitch will be either high or mid, never low. High-pitch w and y are matched, as with the nasals and laterals, by voiceless hw and hy in Geba, so that the same solution may be proposed. High pitch y-words in Blimaw are somewhat suspect. Words with ya<sup>1</sup>- as the first syllable are possibly examples of word intonation; one of them, ya<sup>1</sup>do<sup>3</sup>, appears to be a loan from Burmese. ye<sup>1</sup> 'five' is the form used in some counting constructions, but the regular Bwe form for 'five' is ye<sup>3</sup>. yo<sup>1</sup> 'to mix' (e.g. lime with betel) may be a Burmese loan; yu<sup>1</sup> is only used in poetic language; yo<sup>1</sup> is used in much the same contexts as jo<sup>1</sup> (see above).

The glides r and R are difficult to classify. They appear unstable, being frequently found to be in free variation either with each other, or with initial w or h. R is not found with high pitch. -Ro<sup>2</sup> and Rɔ<sup>2</sup> are in free variation with ho<sup>2</sup> 'to clear (taungya)', and ho<sup>2</sup> 'morning, early' respectively; Ru<sup>2</sup> 'snake' varies freely with Ru<sup>3</sup> and with wi<sup>2</sup>. Rɛ<sup>3</sup> 'to tackle, bring down' has a variant we<sup>3</sup>; Rɔ<sup>3</sup> 'to poke' has a variant rɔ<sup>3</sup>.

r seems most firmly established on mid pitch. On high pitch, ru<sup>1</sup>, used to describe buildings (usually of stone) such as offices, law-courts, etc., may be borrowed. re<sup>1</sup> 'to make friends with, be friendly towards' and ri<sup>1</sup> 'to have a stinging acid taste' are the only other instances recorded of high pitch monosyllables with this initial. On

low pitch  $ru^3 ru^3$  is an onomatopoe;  $ro^3$  'to poke' has the variant  $Ro^3$ ;  $ru^3$  'to smelt (lead)' has the variant  $we^3$ . The only example without a variant is  $ru^3$ , a quantifier of seasons or periods of time. A case could be made out for the phonological interpretation of  $r \sim R \sim w$  as variant realisations of a single underlying glide, there being a tendency for the R realisation to appear before open and back vowels on mid and low pitch, and for the r- realisation to appear before high vowels on high pitch, and perhaps also before rounded vowels in general. (Note the absence of  $*wo$ ,  $*wo$  and limited distribution of  $wu$ ,  $wu$ .) For the purposes of this paper, however, these initials will be kept separate, R being classed with the mid-low pitch set, and r with the high-mid set, though with exceptions noted.

3.5 Row 29 shows the rare instances of words or syllables containing a vowel without a preceding consonant or glottal stop. There is a gradual rather breathy onset to these syllables, quite distinct from the sharp attack of the syllables with initial glottal stop. In  $ke^1u^3ke^1ae^1$ , an expression meaning 'What a waste!' (cp.  $ke^1$  'to be bad, spoilt, useless'), the onset to the syllable  $u^3$  is so fricative that it almost becomes  $wu^3$ . It should be noted that there are no monosyllabic words other than particles or exclamations with this kind of beginning.

4. It is argued from the above that if tone-sandhi and intonational influences, whether grammatical or emotional, are abstracted for appropriate separate treatment by phonological or syntactical rules, Western Bwe exploits a two-fold pitch contrast for lexical purposes. Syllables may be classed as having relatively high or relatively low pitch, and it is this alternation that I would regard as manifesting the phonological tones of the language. These tones may, to distinguish them from mere pitches, be labelled A (for the relatively high one) and B (for the relatively low one). (This labelling is not purely arbitrary; it is intended to facilitate comparison with Benedict's tones A and B, which by and large appear to correspond closely with my A and B.) The phonetic *realisation* of these tones in terms of high, mid and low pitch is determined by and predictable from the nature of the (phonological) initial consonant.

4.1 The Table at 2.4 may be simplified and rearranged to show the relations between pitch, tone and initial consonant as illustrated below.

The square bracketed forms are those whose realisations overlap in Blimaw, but not in Geba.

The arrangement into sets is not entirely arbitrary, but is meant to suggest the possibility, which I cannot explore further in this paper, of further phonological generalisations as regards the features functioning in the Bwe initial system.

	<i>1. High Pitch Realisation</i>	<i>2. Mid Pitch Realisation</i>	<i>3. Low Pitch Realisation</i>
I	A p A ph	B p B ph A b	B b
II	A β A hm	B β [B hm] [A m]	B m
III	A t A th ~ ch	B t B th ~ ch A d	B d
IV	A d' A hn	B d' [B hn] [A n]	B n
V	A c A ʃ	B c B ʃ A j	B j
VI	A hl	[B hl] [A l]	B l
VII	A θ	B θ	
VIII	A x	B x	
IX	A ? A h	B ? B h	
X	A ?y A hy (?)	B ?y [B hy] [A y]	B y
XI	A ?w A hw	B ?w [B hw] [A w]	B w
XII	A r	B r A R	B R

Sets I-V illustrate the operation of syllable-initiating features which might be given some such labels as "tight" or "sharp" onset to the vowel (for the voiceless unaspirated and glottalised initials), "loose" or "aspirated" onset (for the aspirates and  $\int$ ), and voiced onset. With some modification of present definitions, it might be possible to use the Halle and Stevens (1971) larynx features *spread/constricted* and *stiff/slack*, or the multivalued Ladefoged (1972) features *glottal stricture* and *voice onset*. A further feature would be required to distinguish Set I from Set II, and Set III from Set IV. What is required is not *nasal/non-nasal* but rather *oral/non-oral* - *oral* applying to Sets I and III, *non-oral* to Sets II and IV in which there is either nasal cavity resonance or glottal modification to supplement the oral articulatory gesture.

Sets VI and VII appear to be without the "tight" onset feature - though Bwe  $\theta$  is sometimes quite strongly glottalised. On the synchronic evidence, without recourse to what is known of the history and origin of Bwe  $\theta$ , one might be tempted to propose a grouping of  $\theta$ ,  $h_1$ , and  $l$ . (This alignment is not perhaps so wildly unlikely from the phonetic point of view as it may appear. Bwe  $\theta$  has quite firm dental contact so that air flows round the sides of the tongue rather than "through" the narrow gap between tongue and incisors, as is frequently stated for English  $\theta$ . There is thus common phonetic ground with the lateral  $l$ .) Set IX might be interpreted as being the "tight" and "loose" versions of zero consonant, while X-XII may contain examples of zero + glide initials. An enquiry is needed into the status of the other glide-initial complexes, viz.  $ph_1-$ ,  $bw-$ ,  $\beta w-$ ,  $\theta r-$  etc. before these sets could be handled with any hope of plausibility.

5. It remains to compare the allocation of phonological tones A and B to Bwe syllables proposed here with the examples cited by Benedict (1969) and Jones (1971).

5.1 Below I have added Bwe examples (where I could identify them) to the lists given by Benedict. Benedict's reconstructed tones \*A and \*B are included as he gives them. Page references are to Benedict (1972). The corresponding Bwe forms are shown with pitch marks (1, 2 or 3), followed by the tone allocation, A or B, suggested in this paper.

## 5.1.1 Benedict's Numerals (p. 28)

	TB	K	W. Bwe	
'two'	*g/ni	*khi <sup>A</sup>	ki <sup>1</sup>	A
	*g/ni/s	*hni <sup>A</sup>	(gi <sup>3</sup> when counting in pairs)	(B)
'three'	*g/sum	*som <sup>A</sup>	θo <sup>1</sup>	A
	*b/sum			
'four'	*bləy	*lwi/t	lwi <sup>3</sup>	B
			lwi <sup>1</sup> ~ lu <sup>1</sup> (in certain contexts)	(A)
'five'	*b/ŋa	*pa/t	ye <sup>3</sup>	B
			(ye <sup>1</sup> in some contexts)	(A)
'nine'	*d/kəw	*khu/t	khwi <sup>1</sup> ~ khu <sup>1</sup>	A
	*d/gəw	*ku/t		
	*kəw/a	*gu/t		

Note the variants for 'two' used in counting in pairs; also Tone A variants for 'four' and 'five' used in some contexts, e.g. some compound numerals, but felt by Bwe informants to be less genuinely "Bwe" than the Tone B forms.

## 5.1.2 Benedict's TK tone \*A = Chinese tone B (p. 29)

	TB	K	W. Bwe	
'elephant'		*tshaŋ <sup>A</sup>	gǎŋa <sup>1</sup>	A
'boil, cook'			(see below)	
'die'		*si <sup>A</sup>	θi <sup>1</sup>	A
'grandmother'	*phəy <sup>A</sup>	*phi <sup>A</sup>	[ə-] phi <sup>1</sup>	A
'rain'	*r/wa		wε <sup>2</sup>	A

Of the many words for various cooking processes, none seems to fit very well with Benedict's cited Burmese form tshu<sup>A</sup>, corresponding to Ancient Chinese t'siwo<sup>B</sup>. Perhaps the nearest is [u<sup>1</sup>bu<sup>1</sup> 'to roast', which has Tone A in both syllables. There is complete agreement here between my Tone A and Benedict's.

## 5.1.3 Benedict's TK tone \*B = Chinese tone A (p. 29)

	TB	K	W. Bwe	
'ginger'		*eŋ <sup>B</sup>	[θě] ?e <sup>1</sup>	A
'cold'			co <sup>2</sup>	B
'carry on shoulder'	(Bur. tham <sup>B</sup> )		?ya <sup>2</sup> (?)	B
'salt'	(Bur. tsha <sup>B</sup> )		[ɪ <sup>2</sup> ]θɛ <sup>2</sup>	B
'old woman'			mu <sup>2</sup> bwe <sup>2</sup>	A
'house'	*kyim <sup>A</sup> ~ *kyum <sup>A</sup>	*hyi[m] <sup>B</sup>	hi <sup>2</sup>	B
'neck'	*liŋ		[gɔ <sup>3</sup> ]	(B)
'hawk, kite'			[le <sup>3</sup> ]	(B)
'near'	*ney ~ *nay		[bu <sup>2</sup> chi <sup>2</sup> ]	(B)

There is disagreement here over 'ginger', which, with a presumed θ- prefix, is tone A in W. Bwe. If mu<sup>2</sup>bwe<sup>2</sup> 'old woman' is correctly identified, this also is Tone A rather than B. It is interesting, in view of Benedict's comments at the bottom of p. 29, to note that the Tone B form mu<sup>3</sup>bwe<sup>3</sup> in Western Bwe means 'parents-in-law'.

## 5.1.4 Benedict (p. 30) (Words not listed elsewhere)

	TB	K	W. Bwe	
'rat'	*(śa-)yəw	*yü <sup>B</sup>	yu <sup>3</sup>	B
'pain'	(Bur. na <sup>A</sup> )		ni <sup>2</sup> na <sup>2</sup> (=be seriously hurt, to suffer severely)	B
'smoke'	*kəw	*khu <sup>B</sup>	khu <sup>2</sup>	B
'elder sibling'	*kəw <sup>A</sup>		[we <sup>1</sup> ] ?	(A)

## 5.1.5 Benedict (p. 31) Reflexes of ST Tone \*A ~ TB \*A ~ Karen A

	TB	K	W. Bwe	
'call, cry out'	*gaw *kaw		ha <sup>2</sup> (?) (cp. ho <sup>1</sup> 'to scold')	B (A)
'body'	*guŋ		co <sup>1</sup> (?) (=corpse)	A
'body'	*(s)kəw		[hi <sup>2</sup> ]kle <sup>1</sup>	A
'red'	*(s/)kyeŋ		[li <sup>2</sup> ]gɔ <sup>2</sup> (?)	A
'ashamed'			[θa <sup>1</sup> wa <sup>3</sup> ]	



	TB	K	W. Bwe	
'water'	*təy	*thi	chi <sup>1</sup> Geba: thi <sup>1</sup>	A
'hair, feather'	{*tsâm *sâm}		ʃu <sup>2</sup>	B
'fat, grease'	*tsil		θo <sup>2</sup>	B
'new, fresh'	*sar		θe <sup>1</sup>	A
'breath, sound'	} *səm		θe <sup>2</sup>	B
'heart'			θa <sup>2</sup>	B
'alive'	} *śriŋ		θu <sup>1</sup> [klu <sup>2</sup> ]	A
'green'			θu <sup>1</sup> [mu <sup>1</sup> ]	A
'white'	*[p]wa·r	*?(b)wa	βu <sup>1</sup> [θa <sup>1</sup> ]	A
'bear'	*d/wam	*tham	the <sup>1</sup>	A
'I, me'	*ŋa	(*ŋə)	ye <sup>2</sup>	A
'silver'	*/d/ŋul		hu <sup>1</sup>	A
'name'	*r/miŋ	*men	mi <sup>2</sup>	A
'thou'	{*naŋ *(na)}	*na	ne <sup>2</sup>	A
'boil, fry'			ga <sup>2</sup>	A
'barking deer'		*(tə)khi	[so <sup>3</sup> ]khi <sup>1</sup>	A
'span'		*tha	the <sup>1</sup>	A
'to be sick, to hurt'		*tsha	ʃe <sup>1</sup>	A
'ten'		*tshi	ʃi <sup>1</sup>	A
'grandfather'		*phu	phu <sup>1</sup>	A
'older sibling'		*phu	we <sup>1</sup>	A
'pus'		*phi	mi <sup>1</sup> Geba: hmi <sup>1</sup>	A
'husks, chaff'		*phe	phi <sup>1</sup>	A
'village'			[ɬo <sup>1</sup> ]θɔ <sup>1</sup> wɔ <sup>1</sup>	A
'buy'			[ʔa <sup>1</sup> ]bwi <sup>2</sup>	A
'sell'			[ʔa <sup>2</sup> ]ʃe <sup>1</sup>	A
'ripe, cooked'		*hmin	mi <sup>1</sup>	A
'sun'		{*ni	[lũmu <sup>2</sup> ]	A
'day'		*ni	[mu <sup>2</sup> ] (ni <sup>2</sup> = classifier of 'days')	A
'laugh'		*ni <sup>A∨B</sup>	ye <sup>3</sup>	B
'smell'		*hnum	nu <sup>1</sup> Geba: hnu <sup>1</sup>	A
'nose'		*hna	ne <sup>2</sup> [khě <sup>1</sup> ɬe <sup>1</sup> ]	A
'fathom'			khli <sup>1</sup>	A
'wind' n.		*(kə)li	[gɻ]li <sup>2</sup>	A

	TB	K	W. Bwe	
'grandchild'		*li	li <sup>3</sup>	B
'dog-flea'		*k(h)li	kle <sup>2</sup>	B
'boat'		*khli	khli <sup>1</sup>	A
'tongue'		*ble	bli <sup>3</sup>	B
'spirit, ghost'		*(ka)la	[gǣ]le <sup>2</sup>	A
'moon'		*hla	le <sup>1</sup> Geba: hle <sup>1</sup>	A
'warm'		*lom	le <sup>2</sup>	A
'hundred'		*rya	gǣye <sup>2</sup>	A

There is considerable disagreement here in the forms with "hissing" initials - ʃ, h and θ - in which Tone B is often found where, if I understand Benedict's table alright, Tone A is to be expected. From 'alive' onwards there is agreement everywhere, except in 'grandchild', 'dog-flea' and 'tongue'. ye<sup>3</sup> 'to laugh' (Tone B) presumably derives from the B variant of the reconstructed form.

Note the example of pitch dissimilation in the words for 'buy' and 'sell', ?a<sup>1</sup>bwi<sup>2</sup> and ?a<sup>2</sup>ʃe<sup>1</sup>, in which we presumably have in each case the same verbal prefix ?a-, with in the first case high pitch to dissimilate from the mid-pitch of bwi<sup>2</sup> (high-low word-intonation), while in the second case ?a- is mid pitch, thus forming a low-high intonation pattern with the following high pitch ʃe<sup>1</sup>. This is the regular behaviour of verbs with this particular prefix.

#### 5.1.6 Benedict's ST Tone \*B ~ TB Tone \*B ~ Karen Tone \*B (p. 32)

	TB	K	W. Bwe	
'bitter'	*ka	*kha	khε <sup>2</sup>	B
'excrement'	*kləy		ɪ <sup>2</sup> (?)	B
'to open mouth/door window/door'	{ *ka * (m)ja * (ga)		khɔ <sup>2</sup> khɔ <sup>1</sup> [gle <sup>3</sup> ]	B (see note below)
'dog'	*kwəy	*thwi	thwi <sup>2</sup> ~ chu <sup>2</sup>	B
'tiger'	(*k/la)		khɪ <sup>1</sup>	A
'dumb'	*m/a		[tə]ʔu <sup>2</sup> (?)	B
'eat'	*əm	*am	a <sup>2</sup>	B
'jaw'	*gam		khε <sup>2</sup> (= chin, lower jaw)	B
'bird'	(*tow)	*tho	tho <sup>2</sup>	B
'youngest child'	*za/doy		θe <sup>1</sup> ɔe <sup>2</sup>	B

	TB	K	W. Bwe	
'wash'	*m/syil		cu <sup>1</sup> (?)	A
'child'	{*tσα *za	*sa	[phu <sup>2</sup> ]θε <sup>2</sup>	B
'carry on back'	*bəw	{*ʔbü *phü/n	phu <sup>2</sup>	B
'ear'	*r/na	*na	nε <sup>2</sup> [ku <sup>1</sup> ]	B
'hear, listen'	*g/na		[ʃɔ <sup>1</sup> ]ne <sup>2</sup>	B
'soft'	*n[ə]m		jε <sup>2</sup> (?)	A
'tail'	*r/may	*me	[ka <sup>1</sup> ]me <sup>3</sup> Geba: [ka <sup>1</sup> ]hmi <sup>2</sup>	B
'female, woman', 'mother'	*mow	*mo ~ *mu	[bo <sup>3</sup> ]mu <sup>1</sup> , mo <sup>2</sup> mo <sup>2</sup>	(see below) B
'close (eyes), sleep'	*myel		[ʃɔ <sup>1</sup> ]mi <sup>1</sup> = 'sleep' Bi <sup>2</sup> = 'to close (eyes)'	A B
'fork'			kwa <sup>2</sup>	B
'congeal'			[θǎ]ga <sup>3</sup>	B
'body dirt'		*khri	wi <sup>2</sup> Geba: hwi <sup>2</sup>	B
'mortar'		*tshom	[ʃi <sup>1</sup> ]ʃo <sup>2</sup>	B
'urine'		*tshi	ʃi <sup>2</sup>	B
'flower'			phɔ <sup>1</sup>	A
'bean'			[βɔ <sup>2</sup> ]βε <sup>2</sup>	B
'bamboo'		*hwa	hu <sup>2</sup>	B
'tooth'		*swa	θo <sup>1</sup> [mε <sup>3</sup> ]	(see below)
'blood'		*swe	θwi <sup>2</sup> ~ θu <sup>2</sup>	B
'get, obtain'		*ne	ni <sup>1</sup>	A
'breast'		*nü	nu <sup>2</sup>	B
'dream'		*maŋ	[ʃɔ <sup>1</sup> mi <sup>1</sup> ]ma <sup>2</sup> Geba: hma <sup>2</sup>	B
'fire'		*hme	mi <sup>2</sup> Geba: hmi <sup>2</sup> , mi <sup>2</sup>	B
'road, way, track'		*lam	le <sup>2</sup> kle <sup>2</sup>	B B
'buffalo horn (used as musical instrument)'			gwe <sup>2</sup> Geba: gwe <sup>3</sup>	A B
'arrow'		*bla	ble <sup>2</sup>	A
'bow'		*khli	khli <sup>2</sup>	B
'snake'		*ru	wi <sup>2</sup> ~ Ru <sup>2</sup> ~ Ru <sup>3</sup>	A ~ B

There is general agreement here, notable exceptions being the words for 'flower', 'tiger', 'obtain' and 'sleep' which have Tone A according to the criteria set out here. Perhaps the true reflex of Benedict's TB

\*myel 'close the eyes, sleep' is not Bwe mi<sup>1</sup> (Tone A) but ɓi<sup>2</sup> (Tone B) 'to close (the eyes)'. cu<sup>1</sup> 'to wash', also Tone A in my reckoning, is perhaps not related to TB \*m/syil. The high pitch of the first syllable of kho<sup>1</sup>gle<sup>3</sup> 'door, window' is not irregular. It is quite clearly the same morpheme as the tonally regular kho<sup>2</sup> 'to open', with the raised pitch appropriate to high-low word-intonation. So also with the high pitch of the first syllable of θo<sup>1</sup>me<sup>3</sup> 'tooth'. A similar process of pitch dissimilation within words has raised the pitch of the second syllable of bo<sup>3</sup>mu<sup>1</sup> 'female, woman'; note the expected mid pitch (Tone B here) of the monosyllabic form mo<sup>2</sup>. In several instances in this set of words Geba forms are useful in determining whether a Blimaw mid pitch syllable with initial nasal or w is to be regarded as a realisation of Tone B rather than Tone A. The aspirated hm in Geba [ka<sup>1</sup>] hmi<sup>2</sup> 'tail' and hma<sup>2</sup> 'dream' confirm that the corresponding unaspirated Blimaw forms must represent the low tone alternants of /hm-/ syllables, not the high tone alternants of /m-/ syllables. The same is true of Blimaw wi<sup>2</sup> 'body dirt', the Geba equivalent of which is hwi<sup>2</sup>. Luce gives Geba variants, both hmi<sup>2</sup> and mi<sup>2</sup>, beside Blimaw mi<sup>2</sup> 'fire', but the fact that there is an aspirated form in Geba, added to Benedict's reconstructed K \*hme, seems to justify the allocation of Tone B here also. I have no explanation for the Tone A of je<sup>2</sup> 'soft'. The Tone B form demanded by Benedict's analysis would be \*je<sup>3</sup> gwe<sup>2</sup> 'buffalo horn' and ble<sup>2</sup> 'arrow' are also Tone A rather than B in Blimaw. However, I recorded a low pitch form in Geba for gwe<sup>3</sup> 'buffalo horn'. Unfortunately my data does not include the Geba form for 'arrow'.

#### 5.1.7 Benedict's TK Tone \*B = Chinese Tone A (p. 33)

	TB	K	W. Bwe	
'sour'	*sin		ʃi <sup>2</sup>	B
'liver'	*m/sin	*sün	θu <sup>2</sup> θu <sup>1</sup> [θa <sup>2</sup> ]	(see below)
'tree, wood'	*siŋ	*sen	θo <sup>2</sup>	B
'flesh, meat'	*sa		hi <sup>2</sup>	B
'kin, aunt'	*sru		θe <sup>1</sup> = 'kin, to be related to'	A
			(to <sup>2</sup> = 'aunt'?)	(B)
'go, walk'	*wa *s/wa		he <sup>2</sup>	B
'fish'	*(s/)ŋya	*h'na	da <sup>3</sup> [pho <sup>3</sup> ]	B
'year'	*(s/)niŋ	*hniŋ	de <sup>2</sup>	B

Almost total agreement with Benedict here. The apparently exceptional Tone A for θe<sup>1</sup> 'kin, to be related to' is noteworthy. This form

in Bwe may be regarded as a prefix of kinship, or perhaps as a verb expressing kinship. It is found in such sentences as:  $y\theta-\theta\epsilon^1 bu^2w\epsilon^2l\epsilon^3l\epsilon^1c\epsilon^2$  'I am related to him';  $y\epsilon^2l\epsilon^1c\epsilon^2(mi^2 \epsilon-)\theta\epsilon^1bu^2w\epsilon^2$  'I and he (are) brothers'. Compare also  $-\theta\epsilon^1d\epsilon^2$  'youngest child, last child of a family', in which the  $\theta\epsilon^1$  may be the "kinship prefix" or perhaps a reflex of  $\theta\epsilon^2$  'child, offspring' (as suggested by Benedict's TB \*za/doy - see 5.1.6).

### 6.1 THE JONES AND HAUDRICOURT HYPOTHESES

In his paper to the Fourth TB Conference, R.B. Jones added Bwe words (taken from Luce's lists) to his own material from other Karen dialects, and arranged his word lists in sets, associated with his reconstructions of two Proto-Karen (PK) "basic initial types" (aspirate and non-aspirate), two PK tones, "high" (') and "low" (`),<sup>5</sup> and laryngeal final elements marked ', h, and q. For "open syllables", this gave a total of six sets which he compared with Haudricourt's classification of similar syllables. Haudricourt's solution was based on 3 basic types of initials (voiced, voiceless non-aspirate, and voiceless aspirate), combined with 2 original tones (for non-stopped syllables). In his Bwe examples Jones does not usually mark the pitches, nor does he always indicate when a cognate syllable has a prefix or is "bound", i.e. not found as an independent word in the dialect concerned. Both these factors have a bearing upon one's assessment of the tonal "regularity" or "irregularity" of Bwe forms. I have therefore added pitch marks to the Blimaw forms cited by Jones, and have commented on some points of interest. Page references are to Jones' paper.

#### 6.1.1 Jones' Set I (p. 4)

Jones' PK	Haudricourt's Common Karen
*aspirated initials	*voiced initial
*low tone	*Tone 1
*final element '	

'tongue' bli<sup>3</sup>, 'buy' bwi<sup>2</sup>, 'pot' (bo<sup>1</sup>), 'horn, trumpet' gwε<sup>2</sup>, 'warm' lε<sup>2</sup>, 'wind' li<sup>2</sup>, 'work, do' mε<sup>2</sup>, 'dizzy, drunk' mū<sup>2</sup> (Geba: hmu).

Most of these forms are Tone A for me. Exceptions are bli<sup>3</sup> 'tongue' (Tone B) and mu<sup>2</sup> 'dizzy, drunk', where the aspirated form in Geba also points to interpretation as Tone B. (Note that the nasalisation noted by Jones for this word is not phonologically relevant.)

The Bwe material seems here to fit the Haudricourt solution well, i.e. voiced initial (with one exception) and the higher of the two

possible tones.

Note that the full word for 'pot' in Blimaw is go<sup>3</sup>bo<sup>1</sup> and that for 'wind' gĩli<sup>2</sup>. The high pitch of bo<sup>1</sup>, which Jones finds irregular, is discussed in Section 2.1. The full form of 'buy' is ʔa<sup>1</sup>bwi<sup>2</sup>.

#### 6.1.2 Jones' Set II (p. 5)

Jones' PK	Haudricourt's Common Karen
*non-aspirate initial	*voiceless non-aspirate initial
*low tone	*Tone 1
*final element '	

'ginger' -ʔe<sup>1</sup>, 'drink' ʔɔ<sup>1</sup>, 'white' ʔu<sup>1</sup>, 'spear' -ʔa<sup>1</sup>, 'bury' ʔu<sup>1</sup>,  
'umbilical cord' ɗi<sup>1</sup>, 'spread' ɗe<sup>1</sup>, 'axe' kɔ<sup>1</sup>, 'moon' lɛ<sup>1</sup>.

Note that ʔe<sup>1</sup> and ʔa<sup>1</sup> are second syllables of disyllables: θǎ<sup>1</sup>ʔe<sup>1</sup>  
'ginger', θǎ<sup>1</sup>ʔa<sup>1</sup> 'spear'. All these words are Tone A in my (synchronic) analysis.

#### 6.1.3 Jones' Set III (p. 6)

Jones' PK	Haudricourt's Common Karen
*aspirate initial	*voiceless aspirate initial
*low tone	*Tone 1
*final element h	

'flower' phɔ<sup>1</sup>, 'water' chi<sup>1</sup>, 'sweet' ʃi<sup>1</sup>, 'chicken' ʃɪ<sup>1</sup>, 'boat' khli<sup>1</sup>,  
'seed' khwi<sup>1</sup>, 'dry' we<sup>1</sup> (Geba: hwe), 'new' θe<sup>1</sup>, 'ripe, cooked' mi<sup>1</sup>  
(Geba: hmi).

All these words agree in tone, and are Tone A in my analysis. Note that phɔ<sup>1</sup>, which is irregular in Benedict's list in 5.1.6, fits in quite regularly here. There is also agreement everywhere over the aspirated initials (provided one takes Geba as a guide for the Western Bwe material). I have been unable to identify Jones' word mu 'spear trap' in my material. Possibly this is because a preceding syllable has been omitted (cp. similar omissions in the words for 'pot', 'wind' and 'buy' in 6.1.1, and 'ginger' and 'spear' in 6.1.2), which makes the form difficult to trace unless one happens to remember it.

#### 6.1.4 Jones' Set IV (p. 7)

Jones' PK	Haudricourt's Common Karen
*aspirate initial	*voiced initial
*high tone	*Tone 2
*final element '	

'father' pa<sup>2</sup>, 'old, mature' bwε<sup>2</sup>, 'arrow' ble<sup>2</sup>, 'wash (face)' bla<sup>2</sup>, 'ant' do<sup>2</sup>, 'hot' go<sup>2</sup>, 'rattan' wi<sup>2</sup>, 'snake' Ru<sup>2</sup> ~ wi<sup>2</sup>, 'sun' -mu<sup>2</sup>, 'tail' (-me<sup>3</sup>), 'stone' lo<sup>2</sup>.

All these words have Tone A except pa<sup>2</sup> 'father' (Tone B), and -me<sup>3</sup>, from ka<sup>1</sup>me<sup>3</sup> 'tail'. The low pitch of the last syllable of ka<sup>1</sup>me<sup>3</sup> may be ascribed to word-intonation (see 2.1); nevertheless the Geba cognate form hme indicates that we are here dealing with the realisation of Tone B and initial /hm-/ rather than with the realisation of Tone A and initial /m-/.

Note: the full Blimaw word for 'sun' is lūmu<sup>2</sup>.

Western Bwe agrees in initial with Haudricourt, except in the word pa<sup>2</sup> 'father', which seems to be irregular in this set. Tonally, however, the words in this set have (with the 2 exceptions mentioned) the *higher* of the two tonal possibilities - and this usually corresponds with Haudricourt's Tone 1. In this respect W. Bwe fits Jones' tonal solution for this set of words better than Haudricourt's.

Note that ble<sup>2</sup> 'arrow', which was an exception in 5.1.6 is here quite regular.

#### 6.1.5 Jones' Set V (p. 8)

Jones' PK	Haudricourt's Common Karen
*non-aspirate initial	*voiceless non-aspirate initial
*low tone	*Tone 2
*final element h	

'many' ?ε<sup>1</sup>, 'blow' ?u<sup>2</sup>, 'rice, paddy' Bu<sup>2</sup>, 'turbid' du<sup>2</sup>, 'flea' kle<sup>2</sup>, 'hand' cu<sup>2</sup>, 'breath' θε<sup>2</sup>, 'fat' θo<sup>2</sup>.

All these words are Tone B except the first, and are thus in agreement with both Jones and Haudricourt. The seemingly irregular high pitch of ?ε<sup>1</sup> 'many' is ascribable to the fact that this is a bound form, occurring in the disyllabic word ?ɔ<sup>2</sup>?ε<sup>1</sup> 'to be many'. Note that the words θε<sup>2</sup> 'breath' and θo<sup>2</sup> 'fat', which appear irregular in Benedict's list at 5.1.5, are regular members of this set.

#### 6.1.6 Jones' Set VA (p. 9)

Jones' PK	Haudricourt's Common Karen
*aspirate initial	*voiceless aspirate initial
? low tone <sup>6</sup>	*Tone 2
? final element h	

'child' -pho<sup>2</sup>, 'sow' phwi<sup>2</sup>, 'bone' khwi<sup>2</sup>, 'right (hand)' thwe<sup>2</sup> (full form dəthwe<sup>2</sup>), 'pungent' he<sup>2</sup>, 'tooth, tusk' (me<sup>3</sup>) in θo<sup>1</sup>me<sup>3</sup>. (I cannot

trace in my material the words cited by Jones for 'orphan' and 'insect'.)

All these words seem to be regularly Tone B, even the  $me^3$  of  $\theta o^1 me^3$ , which Jones marks as irregular. Any irregularity in this form seems to lie in the nasal initial rather than in the tone, since all other initials are aspirates. Apart from this last word, the Bwe material accords well with both solutions.

#### 6.1.7 Jones' Set VI (p. 10)

Jones' PK

\*aspirate initial

\*high tone

\*final q

Haudricourt's Common Karen

\*voiceless aspirate initial

\*Tone 2

'spherical'  $phlo^2$ , 'swing, rock'  $thwa^2$ , 'sour'  $ʃi^2$ , 'bitter'  $khe^2$ , 'body dirt'  $wi^2$  Geba  $hwi^2$ , 'star'  $ʃe^2$ , 'bow'  $khli^2$ , 'flesh'  $hi^2$ , 'leaf'  $le^2$  Geba  $hle^2$ , 'nail, claw'  $mi^2$  (full form  $\theta i mi^2$ ) Geba  $mi^2$ , 'bamboo'  $hu^2$ .

All aspirate initials (except  $[\theta i]mi^2$  'nail, claw', Geba  $mi^2$ , not  $hmi^2$ , apparently); all Tone B. This is in complete agreement with Haudricourt, but differs tonally from Jones.

#### 6.1.8 Jones' Set VIA (p. 11)

Jones' PK

\*non-aspirate initial

\*high tone

\*final q

Haudricourt's Common Karen

\*voiceless non-aspirate

\*Tone 2

'be, have'  $?a^2$ , 'eat'  $?a^2$ , 'dung'  $?i^2$ , 'delicious'  $?wi \sim ?u^2$ , 'crooked'  $ke^2$ , 'grain basket'  $pu^2$ , 'left hand'  $ci^2$  (full form  $t\acute{e}ci^2$ ), 'liver'  $\theta u^2$ ,  $\theta u^1 \theta a^2$ , 'exact'  $\beta e^2$ , 'egg'  $di^2$ .

All the above agree with Haudricourt in having non-aspirate initials (though not always voiceless, e.g.  $\beta$ -), and low tone. They agree with Jones in initial, but not in tone. (Note the raised pitch of  $\theta u^2$  'liver', in the disyllabic form  $\theta u^1 \theta a^2$ , which would be misleading if cited on its own.)

7. Jones notes that Haudricourt classes Set VIA with Set V, and Set VA with Set VI, "without comment on the differences of tonal correspondences". For Haudricourt, Sets V, VA, VI and VIA all have Tone 2; for Jones V and VA have low tone, VI and VIA have high tone. As will be



seen above, my synchronic interpretation of the Western Bwe material appears to support Haudricourt, all these sets having Tone B. Jones, of course, is drawing upon a far wider range of dialects than was accessible to either Haudricourt or myself and therefore has the far more difficult problem of trying to produce a solution which will as far as possible reconcile the differences, tonal and otherwise, between them. Not surprisingly, therefore, his solution is more complicated than Haudricourt's. The satisfying neatness of the latter, however, leads me to hope that further exploration of the factors that may systematically influence the pitch realisations of underlying lexical tones may enable scholars to iron out what at present seem to be tonal anomalies or irregularities in some of the other Karen dialects, and between Karen as a whole and its Sino-Tibetan relations.

NOTES

1. But note doubts expressed on p. 3 of Jones 1971.
2. Misprinted as me<sup>3</sup> in Jones 1971.
3. Compare Geba gǝ<sup>2</sup>bo<sup>2</sup> 'pot'. Geba does not appear to have Blimaw's preference for pitch dissimilation.
4. Jones occasionally marks nasalisation in the Blimaw forms taken from Luce's material. Such nasalisation is, however, a matter of phonetic detail, occurring after nasal consonants. Nasalisation in Bwe is not phonologically distinctive.
5. Since elsewhere in *Karen Linguistic Studies* Jones uses ' to indicate high tone and ` to indicate low tone, I have assumed he means this convention to apply also to Proto-Karen, though I do not recall that he ever explicitly says so. If he merely means to differentiate 2 Proto-tones without prejudging their nature in any way, this would cast rather different light upon certain cases where Jones' "high" corresponds to my "low" (B) and vice versa.
6. I am not clear from Jones' notation which tone and final element he proposes for this set.

BIBLIOGRAPHY

BENEDICT, P.K.

- 1969 'The Sino-Tibetan Tonal System', paper presented to the 2nd Conference on Sino-Tibetan Reconstruction, Columbia University, and subsequently published (1972) with revisions in: Thomas and Bernot, eds *Langues et Techniques: Nature et Société*.

HALLE, M. and K.N. STEVENS

- 1971 'A Note on Laryngeal Features'. *Quarterly Progress Report of the Research Laboratory of Electronics*. 101. M.I.T.

HAUDRICOURT, A.G.

- 1946 'Restitution du Karen Commun'. BSLP 42.  
1953 'A Propos de la Restitution du Karen Commun'. BSLP 49.

HENDERSON, E.J.A.

- 1961 'Tone and Intonation in Western Bwe Karen'. *Burma Research Society Fiftieth Anniversary Publication No.1*.  
1965 'The Topography of Certain Phonetic and Morphological Characteristics of South East Asian Languages'. In: Milner and Henderson, eds. *Indo-Pacific Linguistic Studies*, II.  
1967 'Grammar and Tone in South East Asian Languages'. *Das Tonem und seine Realisierung in den Sprachen Ost- und Südostasiens*, Wissenschaftliche Zeitschrift der Karl-Marx Universität, Leipzig.

JONES, R.B.

- 1961 *Karen Linguistic Studies*. Berkeley and Los Angeles: University of California Press.

## JONES, R.B.

- 1971 Some Problems with Proto-Karen Tones. Paper presented to the 4th Conference on Sino-Tibetan Language and Linguistics, Indiana.

## LADEFOGED, P.

- 1972 'Phonological Features and Their Phonetic Correlates'. *Journal of the International Phonetic Association* 2/1.

## LUCE, G.H.

- 1959 'Introduction to the Comparative Study of Karen Languages'. *JBRIS* 42.

## MATISOFF, J.

- 1973 'Tonogenesis in Southeast Asia'. In: Larry Hyman, ed. *Consonant Types and Tone, Southern California Occasional Papers in Linguistics*, No.1.