

PACIFIC LINGUISTICS

Series C - No. 78

LANGUAGES OF SABAH: A SURVEY REPORT

Julie K. King
John Wayne King
eds



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.W. Bender

University of Hawaii

David Bradley

La Trobe University

A. Capell

University of Sydney

Michael G. Clyne

Monash University

S.H. Elbert

University of Hawaii

K.J. Franklin

Summer Institute of Linguistics

W.W. Glover

Summer Institute of Linguistics

G.W. Grace

University of Hawaii

M.A.K. Halliday

University of Sydney

E. Haugen

Harvard University

A. Healey

Summer Institute of Linguistics

L.A. Hercus

Australian National University

Nguyễn Đăng Liêm

University of Hawaii

John Lynch

University of Papua New Guinea

K.A. McElhanon

University of Texas

H.P. McKaughan

University of Hawaii

P. Mühlhäusler

Linacre College, Oxford

G.N. O'Grady

University of Victoria, B.C.

A.K. Pawley

University of Auckland

K.L. Pike University of Michigan;

Summer Institute of Linguistics

E.C. Polomé

University of Texas

Malcolm Ross

University of Papua New Guinea

Gillian Sankoff

University of Pennsylvania

W.A.L. Stokhof National Center for

Language Development, Jakarta;

University of Leiden

B.K. T'sou

Murdoch University;

University of Hong Kong

E.M. Uhlenbeck

University of Leiden

J.W.M. Verhaar

Gonzaga University, Spokane

All correspondence concerning PACIFIC LINGUISTICS, including
orders and subscriptions, should be addressed to:

The Secretary
PACIFIC LINGUISTICS
Department of Linguistics
Research School of Pacific Studies
The Australian National University
G.P.O. Box 4, Canberra, A.C.T. 2601
Australia.

Copyright © The Authors

First Published 1984

Typeset by Sue Tys

Printed by A.N.U. Printing Service

Bound by Adriatic Bookbinders Pty Ltd

The editors are indebted to the Australian National University for assistance 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 0 85883 297 6

TABLE OF CONTENTS

		page
PART I - The languages of Sabah: a tentative lexicostatistical classification	Kenneth D. SMITH	1-49
PART II - The languages of Sabah: intelligibility testing		
Introduction	Carolyn P. MILLER	51-57
The Lundayeh language	David C. MOODY	59-65
The Illanun language	John E. BANKER	67-74
The Suluk (Tausug) language	David C. MOODY	75-84
The Ida'an language	John E. BANKER	85-90
The Malayic language family	David C. MOODY	91-100
The West Coast Bajau language	Elizabeth F. BANKER	101-112
The East Coast Bajau languages	Janice WALTON and David C. MOODY	113-123
The Tidong language	David C. MOODY	125-137
The Paitanic language family	Julie K. KING	139-153
The Murutic language family	John A. SPITZACK	155-223
The Papar language	Phyllis A. DUNN	225-229
The Dumpas language	Julie K. KING	231-235
The Lotud language	John E. BANKER	237-243
The Bisaya language	Phyllis A. DUNN	245-250
The Tatana language	Phyllis A. DUNN	251-256
The Kuijau language	John A. SPITZACK	257-270
The Eastern Kadazan language	John A. SPITZACK	271-282
The Rungus language	Julie K. KING	283-296

	page
The Kadazan/Dusun language	John and Elizabeth BANKER 297-324
Conclusion	David C. MOODY 325-337
BIBLIOGRAPHY	339-353
LANGUAGE INDEX	355-359
MAPS	
Lundayeh language map	60
Illanun language map	68
Suluk language map	76
Ida'an language map	86
Malayic language map	92
West Coast Bajau language map	102
West Coast Bajau language map	103
East Coast Bajau language map	115
Tidong language map	126
Paitanic language family map	140
Murutic language area map	156
Murutic dialects map	188
Gana language map	190
Keningau Murut language map	192
Beaufort Murut language map	196
Timugon language map	197
Baukan language map	199
Paluan language map	202
Sumambu/Tagal language map	204
Eastern Murutic languages map	209
Papar language map	226
Dumpas language map	232
Lotud language map	238
Bisaya language map	246
Tatana language map	252
Kuijau language map	258
Eastern Kadazan language map	272
Rungus language map	284
Kadazan/Dusun dialects map	298
Kadazan/Dusun language area map	302
Central Dusun dialect map	303
Coastal Kadazan dialect map	307
Sugut Kadazan dialect map	310
Central Dusun dialect map	312
Klias River Kadazan language map	316
Kimaragang language map	317
Garo language map	319
Tebilung language map	320
Languages of Sabah map	329

EDITORS' FOREWORD

A significant amount of research has previously been done on the languages of Sabah (cf. Bibliography). However, a comprehensive study of the relationships of all major Sabah languages has been lacking. Prentice (1970), who himself has contributed a great deal to the understanding of these languages - especially the Murutic languages - stated clearly that basic and comprehensive research was necessary in order to make an overall classification possible.

A massive amount of data was collected by members of the Summer Institute of Linguistics (SIL) during their survey of the languages of Sabah (1978-1981). As the project progressed it became apparent that by reporting the results, a significant contribution could be made to the understanding of the interrelationships of these languages.

Dr Kenneth D. Smith, the Institute's director in Malaysia, took the lead in analysing and reporting the data using lexicostatistics. The results of his study are presented in Part I of this volume.

Part II goes on from there to present the perspective gained after intelligibility testing was conducted throughout Sabah.

Part II begins with an introduction by Carolyn P. Miller which serves to familiarise the reader with the methodology used in the intelligibility testing phase of the survey. Following the introduction are 19 reports dealing separately with each language or language family where intelligibility testing was conducted. The format and progression of each report was standardised to aid the reader in comparing the different languages.

Among the articles in Part II, the reader will find those dealing with languages which are more well-known, such as The Kadazan/Dusun language by John and Elizabeth Banker, The Murutic language family by John A. Spitzack, The Suluk (Tausug) language by David C. Moody, and The Rungus language by Julie K. King.

But others, based on new findings, present language breakdowns and designations which are less familiar, such as in the reports: The East Coast Bajau language by Janice Walton and David C. Moody, The West Coast Bajau language by Elizabeth F. Banker, The Paitanic language family by Julie K. King and The Ida'an language by John E. Banker.

In his conclusion to Part II, David C. Moody attempts to draw the results of the entire survey together to give the reader a more concise picture of the overall language situation, based on the SIL survey. He also attempts to draw some comparisons between the SIL classification and those of Dyen (1965) and Prentice (1970).

Though the articles presented here represent a giant step towards a comprehensive classification of the languages of Sabah, the individual writers suggest areas where delineations remain vague and further research is needed.

On behalf of the Malaysia Branch of the Summer Institute of Linguistics, we wish to express our appreciation to officials of the Sabah State Government including those at the district and village levels, for their great help in providing Institute members with letters of introduction, guides and transportation.

Their own understanding of the languages and peoples in their areas was also very helpful.

We are indebted to all of the people in the many villages of Sabah who gave their time to work with us and answer our questions. Without their help and co-operation this project could not have been undertaken.

It is our hope that this volume will indeed contribute to the understanding of Bornean languages and that it will be a stimulus to further research.

Julie K. King and John Wayne King, eds
Kota Kinabalu, Sabah
May 1984

PART I

THE LANGUAGES OF SABAH: A TENTATIVE LEXICOSTATISTICAL CLASSIFICATION

Kenneth D. Smith

0. INTRODUCTION	1
1. NON-AUSTRONESIAN LANGUAGES OF SABAH	5
2. AUSTRONESIAN LANGUAGES OF SABAH	5
2.1 Butung language	5
2.2 Javanese stock	6
2.3 North-western Austronesian superstock	6
2.3.1 Lundayeh language	7
2.3.2 Banggi language	8
2.3.3 Illanun language	8
2.3.4 Suluk language	8
2.3.5 Bugis subfamily	9
2.3.6 Ida'an/Begahak subfamily	9
2.3.7 Malayic family	10
2.3.8 Bajau family	12
2.3.9 Bornean stock	17
3. SUMMARY	40
NOTES	44
APPENDIX I: Ethnonyms: an alphabetical listing	46
APPENDIX II: Serial language abbreviation	49

0. INTRODUCTION

Nestled along the coast and tucked into the mountains of Sabah are Sabahans representing various cultural groups and speaking various languages and dialects. The cultural centre of some of these groups is in Sabah whereas for others it is in the surrounding area - including Sarawak, Kalimantan, Sulawesi and the southern Philippines as well as some more distant islands of Indonesia. The purpose of this paper is to present a tentative classification of these languages and dialects of Sabah based upon a lexicostatistical analysis of 344 wordlists representing 325 villages of Sabah.¹

The data used for this classification was gathered by a team of seven trained field linguists between October 1978 and November 1979.² The linguists visited each of the 23 districts of the State of Sabah and interviewed speakers from as many villages within each district as the local district officials

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 1-49. *Pacific Linguistics*, C-78, 1984.

© Kenneth D. Smith

identified as representative of the different languages and dialects occurring within the districts.³ Some 331 representatives of 325 villages were thus interviewed. (Including Bahasa Malaysia and several languages from Sarawak and the Philippines, altogether 344 language samples were included in this study.)

From each representative a standard 367-item wordlist of basic vocabulary was elicited.⁴ Bahasa Malaysia was generally the language of communication. The linguists then compiled a master wordbook in which they listed for each item (or meaning) each different word root found in any wordlist.⁵ Because most word roots occurred with a variety of affixes or sound changes in different wordlists, the various forms of each word root became a set of 'cognate' forms. Each cognate set was assigned a number: using these numbers each wordlist was converted from a string of phonetic forms to a string of numbers. Using a somewhat shorter 327-item wordlist⁶ each wordlist was then compared mechanically with every other wordlist to obtain the percentage of shared cognates for each pairing of wordlists - almost 60,000 pairings. The percentage of shared cognates (PSC) for every pair of wordlists thus obtained are the basis for the following classification of the languages of Sabah.

The classification given here is based upon seven degrees of distinction:

- (1) Between 0 and 15 PSC separates basic linguistic phyla. Different phyla do not share any known historical or genetic origin. The minimal similarity of a few vocabulary items is coincidence or the result of borrowing between languages brought into contact for some reason. Phyla are divided here into superstocks.
- (2) Between 15 and 25 PSC separates linguistic superstocks, the most broad divisions of the phylum. Superstocks are divided into stocks.
- (3) Between 25 and 45 PSC separates linguistic stocks, the broad divisions of the superstock. Stocks are divided into families.
- (4) Between 45 and 60 PSC separates linguistic families, the divisions of the linguistic stock. Families are divided into subfamilies.
- (5) Between 60 and 75 PSC separates linguistic subfamilies, the divisions of the linguistic family. Subfamilies are divided into languages. In this classification it is assumed that there can be little communication between speakers whose languages or dialects are in different subfamilies (unless, of course, the speakers have had occasion to learn the other language through contact with its speakers.)
- (6) Between 75 and 80 PSC separates linguistic languages, the divisions of the linguistic subfamily.⁷ Languages are divided into dialects. In this classification it is not clear how much oral communication is possible between speakers whose different languages are in the same subfamily grouping. Dialect intelligibility testing is underway (1980-1981) to clarify this question (see below).
- (7) Between 80 and 85 PSC separates linguistic dialects, the divisions of the linguistic language. In this classification it is assumed that there is full communication among all speakers of a given dialect; they may or may not fully understand speakers of another dialect within the same language. The greater the number of dialects within a language, the greater potential for loss of communication between some dialects. For any necessary linguistic distinction finer than 'dialect', the term 'subdialect' is used.

In the process of establishing dialects within languages two types of situation have been encountered:

Situation 1: Villages (wordlists) group consistently into dialects, and dialects group consistently into languages as shown in the hypothetical chart Figure 1. In Figure 1 villages A and B are of the same Dialect i, having greater than 85 PSC; Villages C, D, and E are of the same Dialect j, also having greater than 85 PSC; Dialects i and j are of the same Language X since every village of Dialect i has a relation with every village of Dialect j of 80-85 PSC; Village F is of a different Language Y since Village F has a relation with every village of Language X of 75-80 PSC.

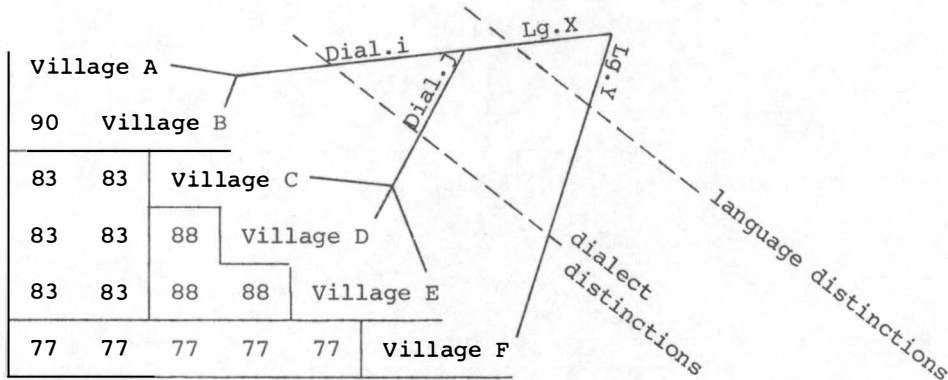


Figure 1: Hypothetical PSC of six villages forming consistent grouping of villages into dialects and dialects into languages. (In this and the following figures the intersection of a column and a row indicates the PSC between the two villages; for example, Village A and Village C are 83% cognate. In this and some of the following charts a language tree is sketched to the right of the chart. The language tree is directly derivable from the boxes drawn in the chart; the boxes enclose PSC within set ranges; the higher nodes of the tree represent the more remote linguistic affiliation; the dotted lines labelled 'language distinctions', etc. indicate the percentage level at which the distinction is made.)

Situation 2: Villages (wordlists) have a graduated relation with other villages forming a 'chain' of dialects as shown in the hypothetical chart Figure 2. What are usually adjacent dialects (pictorially and geographically) have higher relations than non-contiguous dialects. Groupings of villages into dialects is not consistent but somewhat artificial. In Figure 2 Villages A, B, and C form a Dialect i, having greater than 85 PSC; Villages B, C, and D form a Dialect j, having greater than 85 PSC; etc., as shown in Figure 3. Dialects i, j, k, and l form a chain of interconnected dialects. Within any one dialect, intelligibility between villages is assumed; but loss of intelligibility is assumed between the most remote villages, as between Villages A and F. Though intelligibility is lost between Villages A and F they are nevertheless within the same language group because they are connected by a chain of dialects. The overlapping of adjacent links of the dialect chain assures intelligibility between adjacent dialects or possibly even between near but non-adjacent dialects. Intelligibility breaks down as the distance down the 'chain' increases.⁸

In this paper the situation (1) above is assumed unless the term 'chain' is used, in which case the situation (2) pertains.

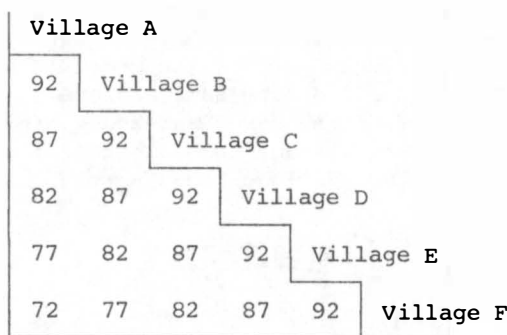


Figure 2: Hypothetical PSC for six villages forming a dialect chain within a single language group

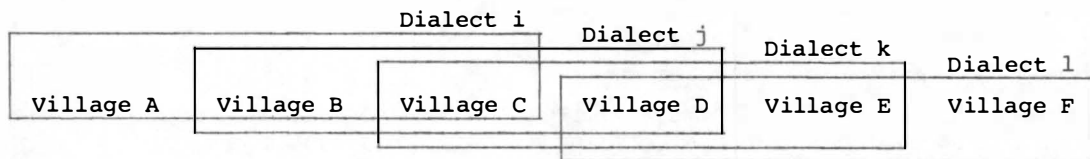


Figure 3: Hypothetical dialect chain derived from Figure 2

The classification of superstock, stock, and family at the 15-25, 25-45, 45-60 PSC ranges, respectively, is based upon a mass of comparisons within the stated range. A single isolated high comparison is usually disregarded, probably evidence of a special case of borrowing.

This classification is tentative. A lexicostatistical classification gives consideration only to the retention (or, from the opposite viewpoint, the replacement) of basic vocabulary items over a span of time. As language groups of the past have split and migrated apart from each other their original speech has changed. The sound systems have changed so that subsequently the two groups which once had spoken alike begin to speak the same words with slightly altered sounds. And the grammatical features of their languages - features like word order, affixes, particles, etc. - also change. Consequently the degree of one's comprehension of a dialect or language related to one's own dialect or language is affected adversely by the increasing amount of replaced vocabulary items, of sound changes and of grammatical changes. Lexicostatistics measures only the first of these three aspects of language change.

The linguistic team which gathered the wordlists for this study are revisiting each district of Sabah to test the speakers of the languages for their comprehension of related languages and dialects using tape-recorded stories. The results of this dialect intelligibility testing will be used to alter the assignment of the labels of 'language' and 'dialect'. It is expected that some 'languages' distinguished here will prove to be only 'dialects' as the speakers of such 'languages' show that they comprehend some other 'language'. On the other hand, some 'dialects' distinguished here will prove to be 'languages' as the speakers of such 'dialects' show that they do not comprehend some other related 'dialect'.

1. NON-AUSTRONESIAN LANGUAGES OF SABAH

It is not within the scope of this paper to discuss the Chinese, Indian, English or other non-South-east Asian languages spoken in Sabah; neither have mother-tongue speakers of Bahasa Malaysia been interviewed (howbeit the dialect intelligibility testing survey includes a Bahasa Malaysia test tape to determine the degree of comprehension of Bahasa Malaysia throughout the state by the speakers of the local languages).

The only non-Austronesian language spoken by residents of Sabah whose cultural area is near Sabah is Chabacano (Chavacano), an Indo-European creole of Tagalog, Cebuano and Spanish origin. Though most speakers of Chabacano live in Zamboanga, Philippines, there is a community of Chabacano speakers in Kg Air SA (Semporna; for district abbreviations see Note 3). From this one wordlist it is shown that Chabacano has only 5-14 PSC with any of the languages included in Section 2 below.

2. AUSTRONESIAN LANGUAGES OF SABAH

The Austronesian languages of Sabah represent three superstocks on the basis of 15 to 25 PSC between them. One superstock is represented by the Butung language (abbreviated in the accompanying chart as BU), another by Javanese ("Jawa" locally) (JA), and the third by North-western Austronesian to which most of the languages of Sabah belong.⁹ Representative cognate percentages between these superstocks are shown in Figure 4.

BU (Sapadulang LD)		
18	JA (Kubuta TU)	
21	21	DN (Bongkud RU)

Figure 4: PSC relations of three representative languages of the three superstocks of the Austronesian phylum found in Sabah. (In most figures giving PSC figures the columns are headed by a 2-letter language/dialect abbreviation followed by specific village name and 2-letter district abbreviation; the language/dialect abbreviations are identified below the chart. BU = Butung language; JA = Jawa, representing Javanese stock; DN = Dusun, representing North-western Austronesian superstock.)

2.1 Butung language

The only sample of the Butung language obtained in Sabah was collected in Sapadulang LD. The Butung people, numbering about 200,000, traditionally inhabit Butung Island and the Tukangbesi Islands off south-east Sulawesi, Indonesia (Grimes 1974). The PSC relation of the single Butung wordlist with two other wordlists is given in Figure 4.

2.2 Javanese stock

Several varieties of Javanese are spoken in Sabah. The Javanese people numbering 60,000,000 inhabit the Sunda Islands from Java to Timor. Javanese as spoken in Tenom appears to be from a different linguistic family than the other varieties which form a Javanese family. Within the latter Javanese family three subfamilies are postulated, one of which includes two languages. The five samples (wordlists) of Javanese encountered in Sabah thus represent five languages. The PSC relations of these languages are shown in Figure 5.

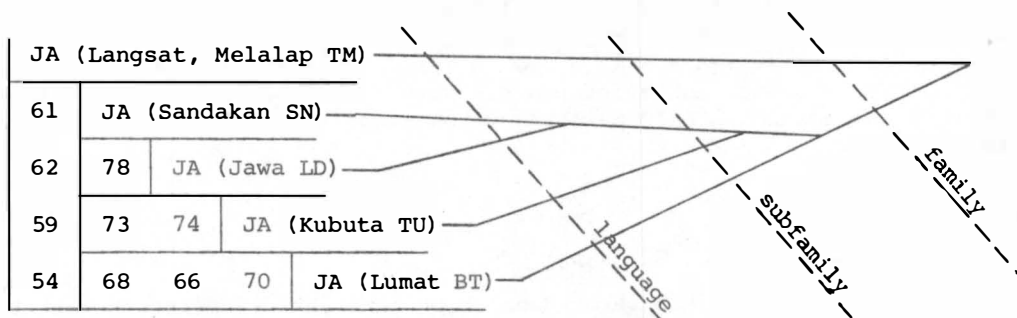


Figure 5: PSC relations of five language samples of the Javanese stock in Sabah. (JA = Jawa.)

The villages, districts and probable homeland of these languages are:

- | | | |
|-------------|------------------------------|----------------|
| (1) Jawa TM | Langsat, Melalap TM | Kapan, Java |
| (2) Jawa SN | Sandakan SN | Tasa, Java |
| (3) Jawa LD | Jawa LD | Central Java |
| (4) Jawa TU | Kubuta TU (established 1917) | Makalang, Java |
| (5) Jawa BT | Lumat BT | Java |

Unfortunately the correspondence of these Javanese languages in Sabah with known Javanese dialects in Indonesia has not yet been established. "Jawa" in Sabah is probably generally thought to represent a single language. Consequently only one wordlist was obtained in each of the districts where "Jawa" was reported. Since the five samples proved to be five distinct languages there may yet be more Javanese languages represented in Sabah among the "Jawa" speakers.

2.3 North-western Austronesian superstock

On the basis of 25 to 45 PSC the languages of Sabah represent nine linguistic stocks within the North-western Austronesian superstock. These are represented by the Lundayeh, Banggi, Illanun and Suluk languages, the Bugis and Ida'an subfamilies, the Malayic and Bajau families, and the Bornean stock (Paitanic, Murutic and Dusunic). Representative cognate percentages of these stocks are given in Figure 6. Lundayeh has a notably lower relation (25-29 PSC) with each of the other stocks than the others have with each other (30-46 PSC). Within the latter group, the Bornean stock, Banggi language and Ida'an subfamily form a subgroup (41-45 PSC) and due to borrowing, the relations of the Bajau family with both the Malayic stock and the Suluk language are elevated (46, 43 PSC). (For additional discussion on the Suluk-Bajau-Malayic relation see Section 2.3.8 Bajau family.)

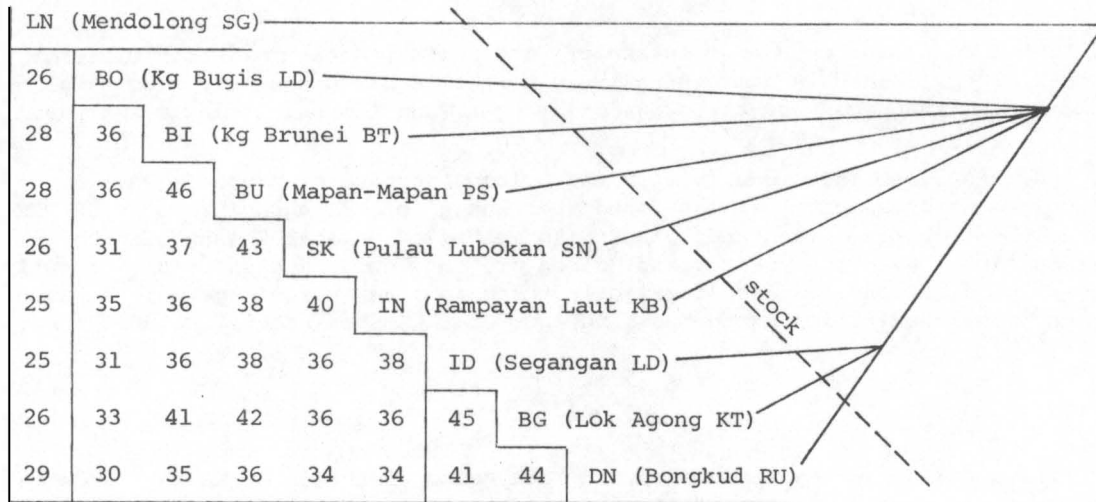


Figure 6: PSC of languages representing the nine linguistic stocks of the North-western Austronesian superstock in Sabah. (LN = Lundayeh language; BO = Bugis language, representing Bugis subfamily; BI = Brunei language, representing the Malayic family; BU = Bajau, West Coast, representing the Bajau family; SK = Suluk language; IN = Illanun language; ID = Ida'an language, representing the Ida'an subfamily; BG = Banggi language; DN = Dusun language, representing the Bornean stock.)

2.3.1 Lundayeh language

Lundayeh (LN) is spoken along the south-western border of Sabah and in neighbouring Sarawak. Some Lundayeh people have recently migrated north into Keningau District. Four samples (wordlists) of Lundayeh from Keningau, Tenom and Sipitang districts are compared in Figure 7. All four samples are very similar (87-89 PSC) and are considered subdialects of the Lundayeh language. The speakers in Kawang SG refer to their language as Lun Lod but it is clearly identified linguistically as Lundayeh.



Figure 7: PSC of four subdialects of the Lundayeh language. (LN = Lundayeh; LL = Lun Lod.)

2.3.2 Banggi language

Banggi is spoken on Banggi Island off the northern coast of Pitas District (howbeit Banggi Island is administratively part of Kudat District). Two samples of the Banggi language were taken in Lok Agong KT and Limbuak Darat KT and have 88 PSC.

Though a close relationship with the Molbog language of southern Palawan, Philippines was suspected, it was found that Banggi of Lok Agong has only 52 PSC with Molbog of Ramos Island and 50 PSC with Molbog of Balabak Island off the southern tip of Palawan. Banggi and Molbog are thus quite distant linguistically (members of different language families); they are a few percentage points closer to each other than either is to the languages of the other stocks included here.

2.3.3 Illanun language

Illanun is spoken in Lahad Datu and Kota Belud districts of Sabah. In each district a distinct dialect of Illanun is spoken having 77-81 PSC. Figure 8 shows the PSC of two samples of each dialect.

The Maranao language of Lanao, Mindanao, Philippines, has 68 PSC with Illanun of Kulambai KB placing Illanun and Maranao within the same linguistic family.

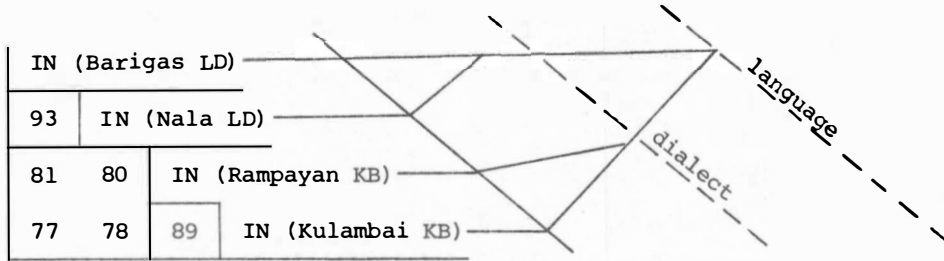


Figure 8: PSC of four samples, two dialects of Illanun. (IN = Illanun.)

2.3.4 Suluk language

Suluk (SK) is the name given in Sabah to Tausug (TS), the trade language of the southern Philippines, spoken by approximately 250,000 people from Jolo south through the Sulu Archipelago. Seven samples of Suluk were obtained from Sabah communities representing each of the east coast districts except Kinabatangan District. Figure 9 indicates the very close relationships among these samples (82-92 PSC) and their relationship with Tausug of Jolo (77-87 PSC). These clearly represent a single language.

SK (Istimewa SN)							
88	SK (Tabanak LD)						
89	91	SK (Pulau Lubokan SN)					
87	89	92	SK (Titingan TU)				
84	86	91	88	SK (Silam LD)			
82	84	89	89	86	SK (Semporna SA)		
82	83	87	84	84	82	SK (Kolapis LS)	
77	81	84	84	82	87	79	TS (Jolo PHL)

Figure 9: PSC of seven samples of the Suluk language compared to Tausug of Jolo, Philippines (PHL). (SK = Suluk; TS = Tausug.)

2.3.5 Bugis subfamily

Bugis (BO) with 2,500,000 speakers, is traditionally spoken in South Sulawesi, the western part of northern and central Sulawesi and in Kutai, Pasir and Pegatan in the south-east (Grimes 1974). Three samples of the Bugis language have been obtained from Bugis communities in the Sandakan, Tawau and Lahad Datu districts of Sabah. These three samples differ from each other near the 80 PSC mark; whether these constitute one or two languages because of a possible breakdown in communication with such differing vocabularies must be determined from intelligibility testing; see Figure 10.

BO (Kg Bugis LD)		BO (Parpadoam SN)		BO (Muhibbah Raya TU)	
81					
76	79				

Figure 10: PSC of three samples of the Bugis subfamily. (BO = Bugis.)

2.3.6 Ida'an subfamily

Ida'an (ID) and Begahak (BE) are ethnonyms representing communities of speakers of the Ida'an/Begahak language of the Ida'an linguistic subfamily. "Sungai" (SI) is also sometimes used by the Ida'an speakers to describe their language.

Four samples of Ida'an and one of Begahak, all in Lahad Datu, show that they are closely related to each other (87-95 PSC). Four samples of Sungai (i.e. Ida'an) from Kinabatangan and Sandakan districts show a divergence to dialect distinctions (78-84 PSC). These two groups are related below the 80 PSC

level (70-79 PSC) suggesting that Ida'an/Begahak of Lahad Datu and Ida'an Sungai of Kinabatangan and Sandakan are different languages. Intelligibility testing must clarify this point. See Figure 11.

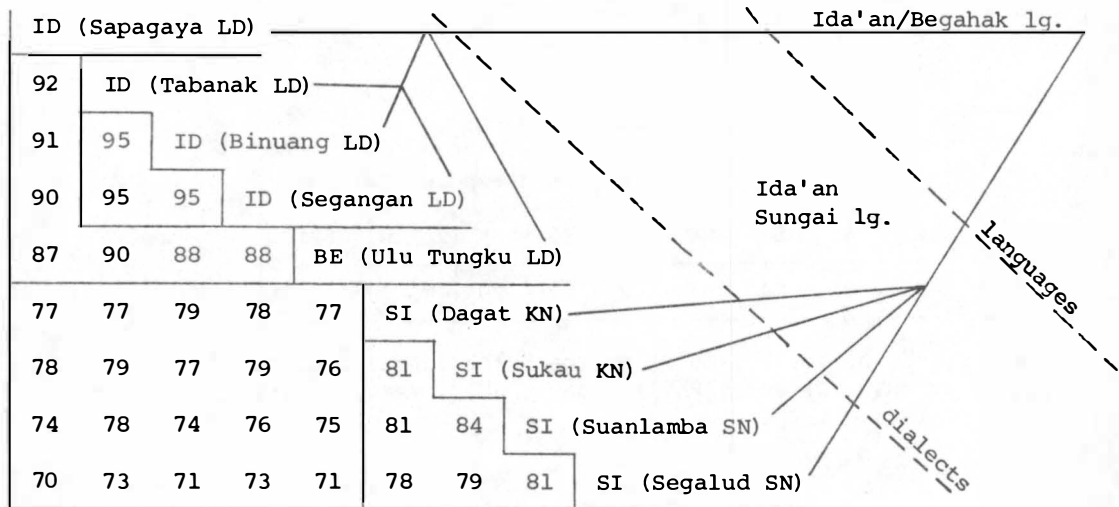


Figure 11: PSC of languages and dialects within the Ida'an subfamily. (ID = Ida'an; BE = Begahak; SI = Sungai.)

2.3.7 Malayic family

The Malayic family of languages has two distinct divisions within Sabah at the subfamily level (i.e. 60-75 PSC). One of these is represented only by the Iban language; the other is a group of languages of the Malayic subfamily including Bahasa Malaysia/Cocos Malay and Brunei/Kedayan. See Figure 12.

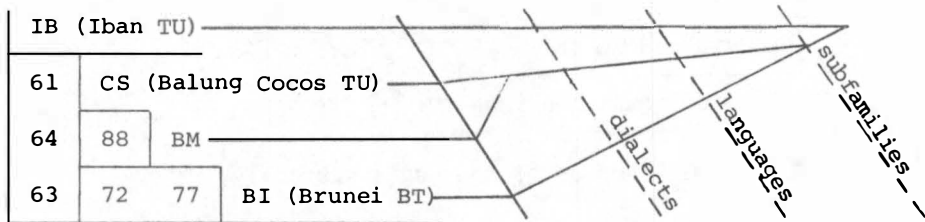


Figure 12: PSC relations of sample Sabah languages within the Malayic family. (IB = Iban; CS = Cocos Malay; BM = Bahasa Malaysia; BI = Brunei.)

2.3.7.1 Iban language

The Iban (IB) people, sometimes called Sea Dayaks, numbering approximately 238,000, inhabit portions of Sarawak, Brunei and Northern Kalimantan (Grimes 1974). Kg Iban in Tawau is a community of Iban speakers who came to Sabah from Sibuan, Sarawak, in 1954.

2.3.7.2 Malayic subfamily

The Malayic subfamily as represented in Sabah consists of two languages: Brunei/Kedayan and Bahasa Malaysia/Cocos Malay.

This study included wordlists from 21 villages in eight districts on both sides of the state whose languages were identified as Brunei (BI), Kedayan (KY), or Brunei-Kedayan (BK). Their mutual relationships range from 77-94 PSC. The wordlist of the "Brunei" village of Seladan-Tarap in Papar is at least 85% cognate with all the other wordlists of this group. Clearly Brunei, Kedayan and Brunei-Kedayan are a single homogeneous language without notable dialects. The villages studied are:

"Brunei":	Patau-Patau LN	Buang Sayang PR
	Menumbuk KP	Benoni PR
	Palu-Palu KP	Kg Sipitang SG
	Weston BT	Melalap TM
	Brunei BT	Kolapis LS
	Kota Klias BT	Berhala Darat SN
	Mawau BT	Tanjung Aru SN
	Seladan-Tarap PR	
"Kedayan":	Layang-Layangan LN	Lingkungan BT
	Lambidan KP	Pantai SG
	Lembaga Lumadan BT	
"Brunei-Kedayan":	Kalanahan PR	

Cocos Malay (CS) speakers were encountered in Tawau and Lahad Datu. Their wordlists are 91% cognate with each other. Furthermore, they are 82% and 88% cognate with Bahasa Malaysia (BM). The speakers of this dialect came from the Cocos Islands (also called Keeling Islands; area nine square miles) in the Indian Ocean south of Sumatra.

The Bahasa Malaysia wordlist used for comparison does not specifically represent the speech of a single village as other wordlists in this study do. But where there was a choice of more than one form for a given entry, the form more commonly used in Sabah was chosen for comparison.

Figure 13 shows the PSC of representative subdialects of the Brunei/Kedayan language, of Bahasa Malaysia and of the two Cocos Malay subdialects. The relation is probably that of distinct languages. (The very high and inconsistent 93% cognate relation between Bahasa Malaysia and the Brunei language of Berhala Darat SN is dismissed as probably due to a disproportionate number of loans from Bahasa Malaysia; among the Brunei-Kedayan subdialects its relation is correspondingly 3-5% lower than the others.)

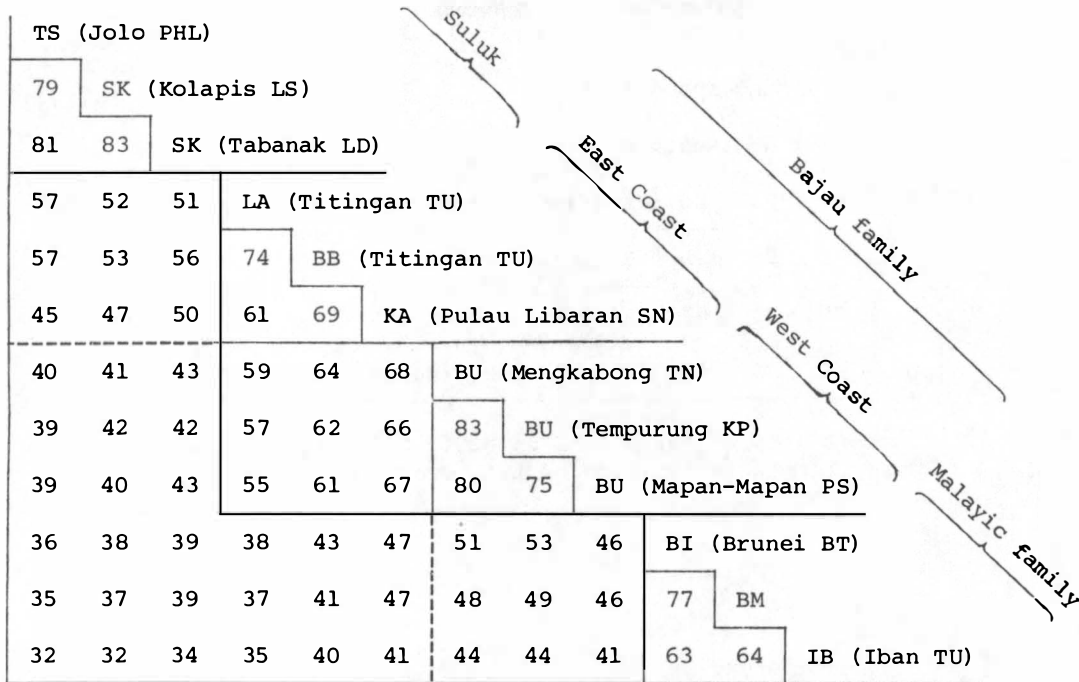


Figure 14: PSC relations of representative dialects of the linguistic stocks represented in Sabah by the Suluk language and Bajau and Malayic families showing borrowing by East Coast Bajau from Suluk and borrowing by West Coast Bajau from the Malayic family. (TS = Tausug; SK = Suluk; LA = Laminusa; BB = Bajau Banaran; KA = Kagayan; BU = Bajau; BI = Brunei; BM = Bahasa Malaysia; IB = Iban.)

2.3.8.1 West Coast Bajau language

The West Coast Bajau language (BU) is represented by 14 wordlists taken in ten predominantly west coast districts though the language includes dialects found now in Sandakan and Labuk-Sugut on the east coast. One wordlist was collected from Tenom, Kuala Penyu, Penampang, Kota Kinabalu, Pitas, Labuk-Sugut and Sandakan, two wordlists from Tuaran and Kota Belud, and three wordlists from Papar. Only "Bajau" is used as the ethnonym in the samples of this dialect chain and no Philippine-based Sama wordlist is linked to this chain.

There is a central network of seven wordlists linked by relationships of 90 PSC or higher as shown in Figure 15.

Connected to the central network by relationships of between 73-86 PSC are the remaining seven non-central wordlists, as shown in Figure 16 which repeats the Bajau wordlist from Mengkabong TN as representative of those cited in Figure 15.

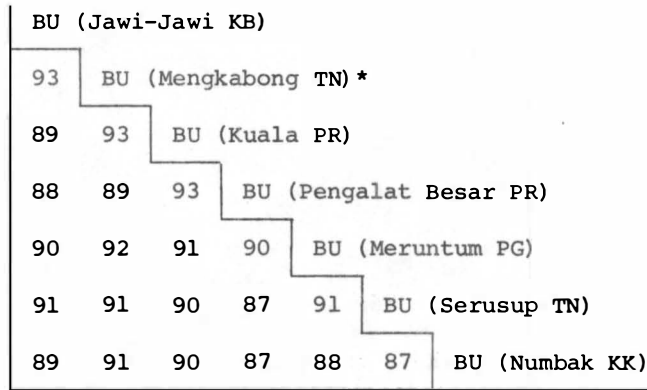


Figure 15: PSC relations of seven central West Coast Bajau villages. (* representative village of this group in Figure 16; BU = Bajau.)

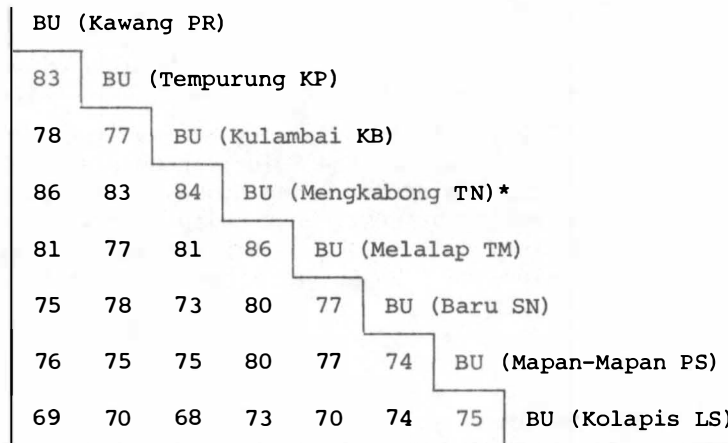


Figure 16: PSC relations of seven non-central West Coast Bajau villages and one central West Coast Bajau village. (* representative village from central West Coast Bajau, see Figure 15.)

2.3.8.1 East Coast Bajau language

The East Coast Bajau dialect chain is represented by 25 wordlists that are interlocked by links of at least 75 PSC. The variety of ethnonyms given by speakers of this dialect chain, the respective districts where these people live, and the number of wordlists taken are given below. Six Philippine wordlists here noted by abbreviation PHL are included in order to compare the Sabah Bajau languages with the Philippine Sama languages.¹⁰ Only Ubian is also found on the Sabah West Coast in Kota Belud, Kota Kinabalu and Kuala Penyu.

- BB Bajau Banaran TU (1)
- BD Bajau Darat LD (1)
- BL Bajau Laut SA (1)

- BS Bajau Semporna SA (1)
- BY Bajau Balangingi LD (1)
- JM Jama Mapun PHL (1)
- KA Kagayan SN (2)
- LA Laminusa TU (1)
- SB Sibutu SA (1), LD (1)
- SM Simunul SA (1), SN (1)
- SN Sikubung SA (1), LD (1)
- SX Sama PHL (4)
- UN Ubian KB (1), KK (1), KP (1), SA (1), LD (1), PHL (1)

This dialect chain has a strongly interlocked central network of dialects linked by relationships of 85 PSC or higher represented here by 15 wordlists. Except for five wordlists taken from Philippine sources, the PSC of these dialects are shown in Figure 17. It is not evident in this figure that every dialect is related to another by no less than 85 PSC; the Philippine-based wordlists are the missing links. The dialects included in Figure 17 which do not have evident dialect links in the figure of 85 PSC or higher are:

- (1) Bajau Banaran (BB), Titingan TU and Bajau Darat (BD), Dasar LD, which are both 86 PSC with Sama (SX), Sitangkay PHL, which is 89 PSC with Bajau Laut (BL), Bangau-Bangau SA, but which is, reflecting back, only
 - (a) 77 PSC with Bajau Banaran (BB), Titingan TU above, and
 - (b) 81 PSC with Bajau Darat (BD), Dasar LD above.

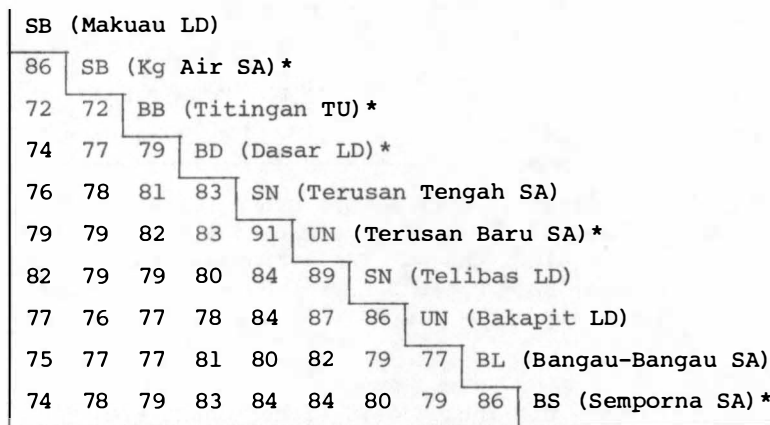


Figure 17: PSC of ten samples of a core group of the East Coast Bajau dialect chain. (* identifies villages repeated in Figure 18; SB = Sibutu; BB = Bajau Banaran; BD = Bajau Darat; SN = Sikubung; UN = Ubian; BL = Bajau Laut; BS = Bajau Semporna.)

Connected to the central network by relationships of between 75-84 PSC are the remaining ten wordlists (one of which is from a Philippine source). Seven of these are shown in Figure 18 together with five representative wordlists repeated from Figure 17. The dialects included in Figure 18 which do not have evident dialect links in the figure of 75 PSC or higher are:

- (1) Kagayan (KA), Pulau Libaran SN, which is
 87 PSC with Kagayan (KA), Sibuga Besar SN, which is
 83 PSC with Jama Mapun (JM), Cagayan de Sulu PHL, which is
 80 PSC with Sama (SX), Sitangkay PHL, which is
 89 PSC with Bajau Laut (BL), Bangau-Bangau SA, but which is, reflecting
 back, only
 68 PSC with Kagayan (KA), Pulau Libaran SN.
- (2) Laminusa (LA), Titingan TU, which is
 80 PSC with Sama (SX), Siasi PHL, which is
 84 PSC with Bajau Banaran (BB), Titingan TU, but which is, reflecting
 back, only
 74 PSC with Laminusa (LA), Titingan TU.

KA (Pulau Libaran SN)											
66	UN (Pulau Gaya KK)										
61	62	LA (Titingan TU)									
66	76	66	UN (Kuala Abai KB)								
69	70	74	72	BB (Titingan TU)*							
71	73	73	81	82	UN (Terusan Baru SA)*						
69	68	71	73	79	84	BS (Semporna SA)*					
68	70	70	73	79	83	83	BD (Dasar LD)*				
70	70	67	74	75	80	79	76	SM (Simunul SA)			
69	67	67	70	77	79	75	75	74	BY (Telisai LD)		
68	68	64	72	72	79	78	77	78	72	SB (Kg Air SA)*	
69	67	66	71	74	79	74	75	76	71	73	SM (Bokara SN)

Figure 18: PSC of representatives of a core group (marked with *, see Figure 17) and of the more remote dialects of the East Coast Bajau dialect chain. (KA = Kagayan; UN = Ubian; LA = Laminusa; BB = Bajau Banaran; BS = Bajau Semporna; BD = Bajau Darat; SM = Simunul; BY = Bajau Balangingi; SB = Sibutu.)

The Sabah-based wordlists not included in either chart are:

- (1) Ubian (UN), Sangkabok KP, which is 86 PSC with Ubian (UN), Kuala Abai KB;
- (2) Kagayan (KA), Sibuga Besar SN, which is 87 PSC with Kagayan (KA), Pulau Libaran SN.

The closest ties of the six Philippine-based wordlists with Sabah languages are:

Sama (SX), Sitangkay PHL, 89 PSC with Bajau Laut (BL), Bangau-Bangau SA;
 Sama (SX), Sibutu SA, 87 PSC with Sibutu (SB), Kg Air SA;
 Sama (SX), Siasi PHL, 84 PSC with Bajau Banaran (BB), Titingan TU;
 Sama (SX), Pangutaran PHL, 75 PSC with Bajau Banaran (BB), Titingan TU;
 Ubian (UN), Soong Buna PHL, 75 PSC with Bajau Banaran (BB), Titingan TU;
 Jama Mapun (JM), 83 PSC with Kagayan (KA), Sibuga Besar SN.

2.3.9 Bornean stock

Of the 344 wordlists included in this study, 107 wordlists are classified in the preceding sections of this paper. The remaining 237 wordlists, two-thirds of the entire data, are classified in the Bornean stock. Though population figures are not yet available following this linguistic division of the peoples of Sabah, the Bornean stock represents the vast majority of the local people of Sabah.

On the basis of 45-60 PSC the Bornean stock has four divisions: the Tidong language, the Paitanic family, and the Murutic and Dusunic families. Three samples of each of these are compared in Figure 19.¹¹

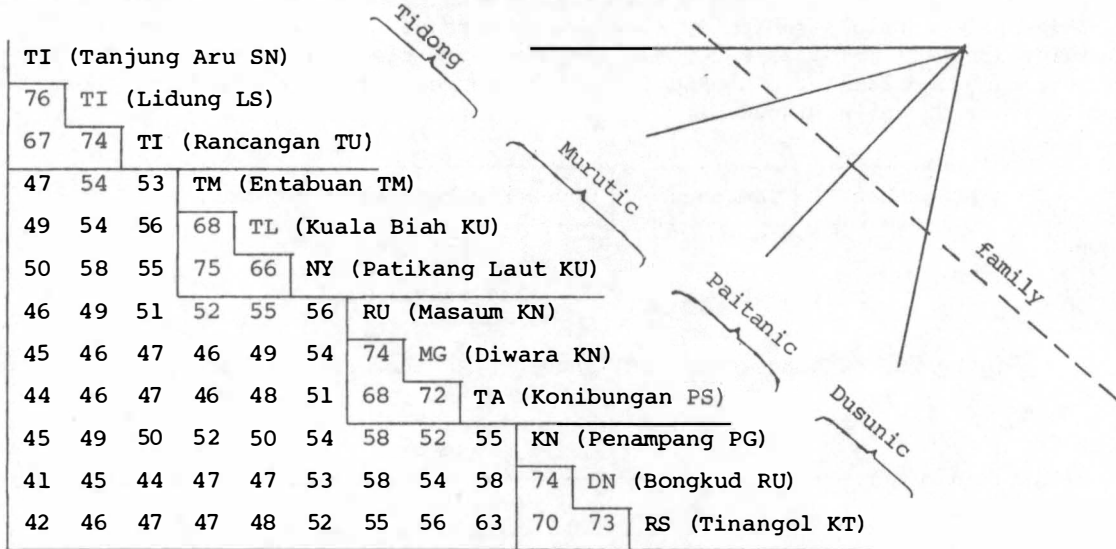


Figure 19: PSC relations of three samples each of the four divisions of the Bornean stock. (TI = Tidong; TM = Timugon; TL = Tagal; NY = Nabay; RU = Rumanau; MG = Makiang; TA = Tambanua; KN = Kadazan; DN = Dusun; RS = Rungus.)

2.3.9.1 Tidong language

The Tidong language (TI) which is classified as a constituent of the Bornean stock is represented here by only three wordlists - one each for Labuk-Sugut, Sandakan and Tawau districts. The Labuk-Sugut communities evidence a recent northern migration of Tidong in that the Tawau community is at the northern end of the traditional area of the Tidong which has been described as "coast-line of Sabah encircling Cowie Harbor" and "along Sembakung and Sibuka rivers of Eastern Kalimantan" (Grimes 1974). The PSC relations of the three samples of Tidong are shown in Figure 19. In terms of language defined by a dialect chain, these three dialects are marginally considered a single language; mutual intelligibility is doubtful and must be tested.

2.3.9.2 Paitanic family

The Paitanic family of languages consists of five languages: Lingkabau, Lobu, Abai Sungai, Tambanua and Upper Kinabatangan. The lexicostatistic data suggests that these five languages form a language (not dialect) chain wherein some languages are much closer to each other than to others in the chain. Their mutual relations do not lend themselves to a language tree relation which implies consistent bifurcation in successive stages. The language chain is shown in Figure 20. The Upper Kinabatangan language appears to be the central language link of the chain. These language links are established by at least one dialect of one language having a relation with at least one dialect of another language of 75-80 PSC, which would allow these languages to be described as forming a language chain within a subfamily - but since most of the inter-language relationships are well below 75 PSC these languages are classified as belonging to a single language family. The PSC figures of representative dialects of these languages are given in Figure 21.

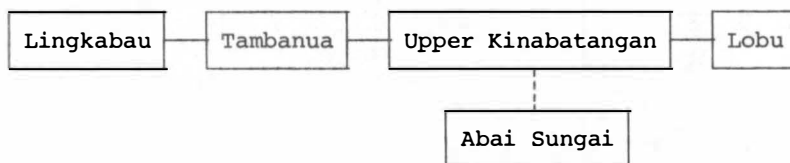


Figure 20: Language chain of languages in the Paitanic family

LU (Lingkabau KM)	_____										Lingkabau language	
70	SI (Pulau Jambongan LS)	} _____										Tambanua language
79	80	TA (Konibungan PS)										
65	70	77	SI (Abai KN)	_____							Abai Sungai language	
64	67	70	73	DA (Bukit Balacan LD)								
71	72	76	79	85	KB (Pinangah KN)						} Upper Kinabatangan language	
71	76	79	73	75	85	SI (Gum-Gum SN)						
68	67	71	67	69	80	74	SU (Liupampang KN)					
66	60	69	58	60	67	65	72	LO (Tampias RU)	_____			Lobu language

Figure 21: PSC relations of representative dialects of five languages in the Paitanic family. (LU = Lingkabau; SI = Sungai; TA = Tambanua; DA = Dusun Segama; KB = Kolobuan; SU = Sinabu; LO = Lobu.)

2.3.9.2.1 Lingkabau language

Lingkabau (LU), represented by only one wordlist, is spoken in Lingkabau village KM. Its highest PSC relation with any other wordlist is only 81% with Tambanua Sungai of Agan village LS (see 2.3.9.2.4 below).

2.3.9.2.2 Lobu language

Lobu (LO), represented by only one wordlist, is spoken in Tampias RU. Whereas the speakers of this language refer to their language as Lobu, they refer to themselves as the Tobu people. The highest relation the single wordlist has with any others is 76 PSC with both of the Rumanau dialect samples of the Upper Kinabatangan languages (see 2.3.9.2.5 below).

2.3.9.2.3 Abai Sungai language

The Sungai language (SI) as spoken at Abai KN has its closest relation at 80 PSC with the Sinarupa dialect spoken at Bulot KN and the Sungai dialect spoken at Kuamut KN - both of the Upper Kinabatangan language.

2.3.9.2.4 Tambanua language

The Tambanua language (TA) is primarily found in Pitas and Labuk-Sugut districts but a migrant group which call themselves Paitan (PN) now live in Beaufort. The Pitas group refer to themselves as Tambanua whereas those in Labuk-Sugut generally refer to themselves as Sungai (SI). Linguistically, as this data shows, the two groups speak a common language with only slight dialectal variations. The 15 wordlists collected from this language group form a dialect chain as indicated in the PSC figures of Figure 22. In this figure the Sungai village Pantai Buring LS stands for seven other Tambanua Sungai villages which are interrelated at 92-93 PSC: Tapat LS, Botition LS, Agan LS, Lingkabau LS (not to be confused with Lingkabau village in Kota Marudu where the Lingkabau language is spoken - see 2.3.9.2.1 above), Sungai-Sungai LS, Kubulu LS and Simpangan LS. The Tambanua village Konibungan PS stands for two other Tambanua villages which are interrelated at 94-95 PSC: Tambilidon PS and Sungai Elloi PS.

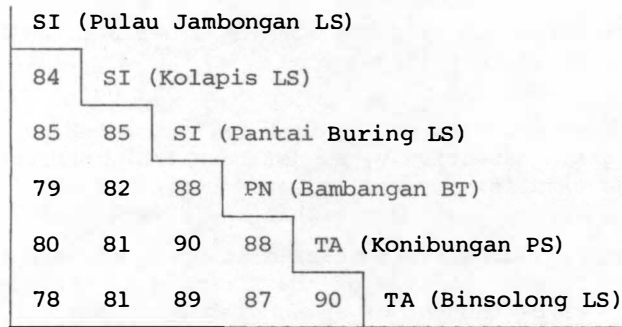


Figure 22: PSC relations of representative villages of the Tambanua language. (SI = Sungai; PN = Paitan; TA = Tambanua.)

2.3.9.2.5 Upper Kinabatangan language

The Upper Kinabatangan language is the language primarily spoken in the upper Kinabatangan River area of Kinabatangan District, though some groups have migrated and are now found in Sandakan and Lahad Datu. The various names by

which the language (or its dialects) is known are Makiang, Kolobuan, Sinarupa, Sinabu, Rumanau, Sungai and Dusun Segama. It is represented here by 15 word-lists.

The Upper Kinabatangan language consists of a linkage of four dialects in which one, the "Makiang" dialect, has a central position to each of the others, as shown in Figure 23. The Gum-Gum Sungai (SI) and Dusun Segama (DA) are related to the "Makiang" group by 82-85 PSC and 78-85 PSC respectively, whereas they are only 75 PSC with each other. Using Pinangah as representative of the "Makiang" group the Sinabu/Rumanau group is related to it by 74-84 PSC, with the Sinabu dialects notably closer (80-84 PSC) than the Rumanau (74-78 PSC).

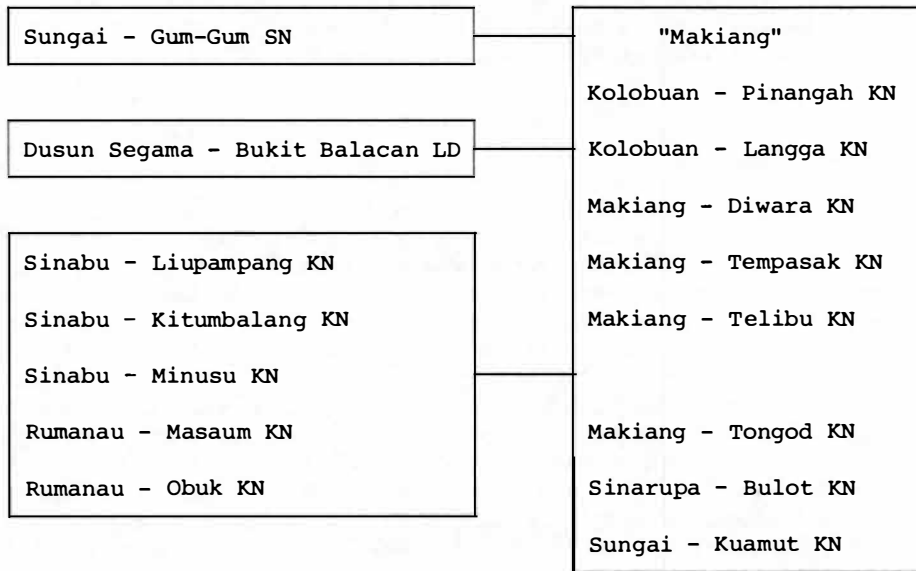


Figure 23: Dialect names and villages forming a dialect chain of the Upper Kinabatangan language

Speakers in the three villages of Pinangah KN, Diwara KN, and Bulot KN referred to their languages, respectively, as Kolobuan (KB), Makiang (MG) and Sinarupa (SA). But the wordlists proved to be 99 PSC. All the PSC figures of the eight "Makiang" dialect villages are within the 87-99% range.

The three Sinabu (SU) villages are related at 93-94 PSC and the two Rumanau (RU) villages have 95 PSC; the relation of the Sinabu with the Rumanau villages is 81-88 PSC - probably to be thought of as subdialects. But their relations with the other dialects of the Upper Kinabatangan language are 74-84 PSC with Pinangah Kolobuan representing "Makiang", 68-77 PSC with the Gum-Gum Sungai, and 62-72 PSC with Dusun Segama. See Figure 24.

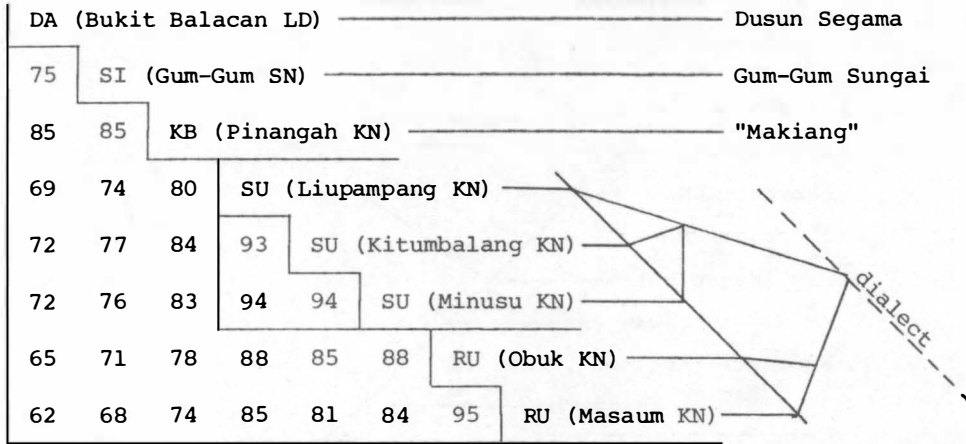


Figure 24: PSC relations of dialects of the Upper Kinabatangan language. (DA = Dusun Segama; SI = Sungai; KB = Kolobuan; SU = Sinabu; RU = Rumanau.)

2.3.9.3 Murutic family

The Murutic family of languages consists of 15 languages based upon the analysis of 36 wordlists. Eight of these languages are predominantly related within the 75-80 PSC range and are thus grouped under the classification of Central Murut subfamily. The other seven languages are all sufficiently different from each other and from Central Murut that they are distinguished at the subfamily level; i.e. predominantly 60-75 PSC. The Central Murut languages are all found in Keningau and Tenom districts except for one which spills over into Kinabatangan and a migrant group in Beaufort. The other Murutic languages are found in the central Murut area of Keningau and Tenom as well as in the south-east coastal Tawau District and, in the case of the widespread Tagal language, also in Pensiangan and Sipitang.

The languages which comprise the Murutic family are Kolod, Gana, Apin-Apin Kuijau, Kalabakan Murut, Sembakung Murut, Serudung Murut, Tagal and the Central Murut languages. The Central Murut languages are Takapan, Paluan, Timugon, Beaufort Murut, Dusun-Murut, Sook Murut, Baukan and Nabay. Figure 25 gives the PSC figures for representative dialects of each of these 15 languages. For each wordlist or village only the ethnonym given by the language assistant is given in the chart; therefore the very common "Murut" response is more specifically defined by either location or the ethnonym for a related dialect.

2.3.9.3.1 Kolod language

The Kolod language (KD), whose linguistic and cultural centre is in Kalimantan, is represented here by only one wordlist from Baru Jumpa TM. Its highest relation with any other wordlist is 73 PSC with the Tagal language as spoken in Tomani TM, but its relation with the other Murutic languages is otherwise in the range of 56-72 PSC.

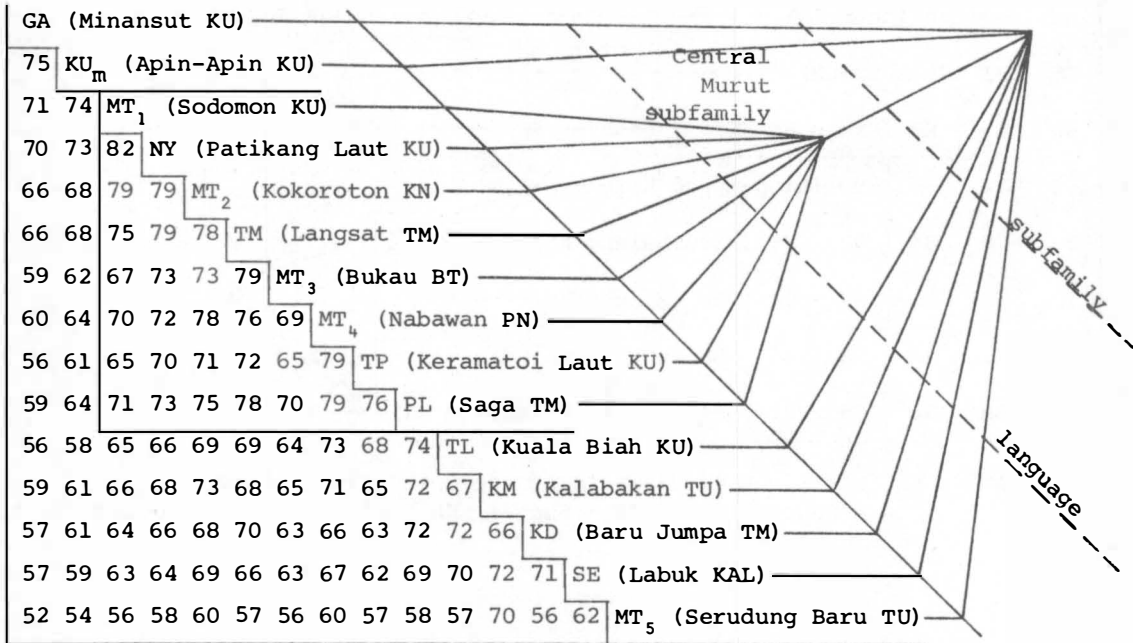


Figure 25: PSC relations of representative dialects of 15 Murutic languages.

(GA = Gana; KU_m = Apin-Apin Kuijau (subscript_m designates this Murutic language in contrast to KU which designates the Dusunic Kuijau language); MT_1 = "Murut" representing Dusun-Murut; NY = Nabay; MT_2 = "Murut" representing Baukan; TM = Timugon; MT_3 = "Murut" representing Beaufort Murut; MT_4 = "Murut" representing Sook Murut; TP = Takapan; PL = Paluan; TL = Tagal; KM = Kalabakan Murut; KD = Kolod; SE = Sembakung Murut; MT_5 = "Murut" representing Serudung Murut.)

2.3.9.3.2 Gana language

The Gana language (GA) is represented here by only one wordlist from Minansut KU. Its highest relation with any other wordlist is 75 PSC with Apin-Apin Kuijau, but its relation with the Murutic languages is otherwise in the range of 56-71 PSC.

2.3.9.3.3 Apin-Apin Kuijau language

Some "Kuijau" dialects are classified as Dusunic (see section 2.3.9.4.7 below) but that spoken in Apin-Apin (KU_m) is lexically closer to Murutic. Its highest relation with any other wordlist is 77 PSC with Dusun-Murut as spoken in Liau-Laut KU, but its relation with the other Murutic languages is otherwise in the range of 54-75 PSC.

2.3.9.3.4 Kalabakan Murut language

The Kalabakan Murut language (KM) is represented here by only one wordlist from Kalabakan TU. Its highest relation with any other wordlist is 77 PSC with the Baukan language as spoken in Tulid KU, but its relation with the Murutic languages is otherwise in the range of 59-73 PSC.

2.3.9.3.5 Sembakung Murut language

The Sembakung Murut language (SE) is represented here by only one wordlist from Kg Labuk of Kalimantan given by speakers in Serudung Baru TU. Its highest relation with any other wordlist is 72 PSC with Kalabakan Murut, also of Tawau, but its relation with the Murutic languages is otherwise in the range of 57-71 PSC.

2.3.9.3.6 Serudung Murut language

The Serudung Murut language (MT₅) is represented here by only one wordlist from Serudung Baru TU. Its highest relation with any other wordlist is 70 PSC with Kalabakan Murut, also of Tawau, but its relation with the Murutic languages is otherwise in the range of 52-62 PSC.

2.3.9.3.7 Tagal language

The Tagal language (TL, MT₆) is here represented by nine wordlists representing villages in Keningau, Pensiangan, Tenom and Sipitang districts. The language appears to have two dialects not geographically distinct but, interestingly, distinguished by the language name given by the speakers; for the three instances of one dialect "Murut" was given, whereas for five of six instances in the other "Tagal" was given. Of these nine Tagal villages the language as spoken in Kuala Biah KU has the highest average relation with the other Tagal dialects; consequently this dialect was chosen as the representative Tagal dialect in Figure 25. Figure 26 gives the PSC relations of the Tagal dialects.

MT ₆ (Sapulut PN)								
89	MT ₆ (Pensiangan PN)							
82	84	MT ₆ (Ansip KU)						
77	80	78	MT ₆ (Kg Lima PN)					
75	76	78	81	TL (Salalir River PN)				
82	82	82	83	84	TL (Kuala Biah KU)			
76	78	80	82	79	85	TL (Masanoi TM)		
75	74	74	77	79	83	79	TL (Maligan SG)	
70	73	73	76	74	77	83	73	TL (Tomani TM)

Figure 26: PSC relations of nine villages of the Tagal language. (MT₆ = Murut, representing Tagal; TL = Tagal.)

2.3.9.3.8 Central Murut subfamily

The Central Murut subfamily consists of eight languages which are predominantly related within the 75-80 PSC range as given in Figure 25.

2.3.9.3.8(a) Takapan language

The Takapan language is represented here by only one wordlist from Keramatoi Laut KU. Its highest relation with any other wordlist is 79 PSC with Sook Murut (i.e. with both villages listed in 2.3.9.3.8(f) below) but its relation with the Central Murut languages is otherwise 65-76 PSC and with the other Murutic languages 56-68 PSC.

2.3.9.3.8(b) Paluan language

The Paluan language (PL) is here represented by only one wordlist from Saga TM. Its highest relation with any other wordlist is 79 PSC with both Sook Murut as spoken in Nabawan PN and Tagal as spoken in Ansip KU, serving thus as a link between Sook Murut and Tagal. Its relation with the Central Murut languages is otherwise 70-78 PSC and with the other Murutic languages 58-74 PSC.

2.3.9.3.8(c) Timugon language

The Timugon language (TM) is here represented by two wordlists from Entabuan TM and Langsat TM which have 85 PSC. The highest relation that either has with any other wordlist is 80 PSC between the latter and Nabay as spoken in Kg Keningau KU, but its relation with the Central Murut languages is otherwise 72-79 PSC and with the other Murutic languages 57-70 PSC.

2.3.9.3.8(d) Beaufort Murut language

The Beaufort Murut language (MT₃) is here represented by two wordlists from Bukau BT and Malalugus BT which have 86 PSC. The next highest relation of both is 79 PSC with the Timugon language as spoken in Langsat TM. Using the speech of Bukau BT as representative, the relation with the other Central Murut languages is 65-73 PSC, notably lower than the Timugon relation, suggesting that the Beaufort community migrated from a Timugon community in Tenom. The relation of Beaufort Murut with the other Murutic languages is 56-65 PSC.

2.3.9.3.8(e) Dusun-Murut language

The Dusun-Murut language (MT₁) is here represented by three wordlists from Ambual KU, Sodomon KU and Liau Laut KU. Only the last village was said to be Dusun-Murut; the others were simply given as "Murut" - but their close relation with Dusun-Murut provides a more convenient ethnonym. The PSC relation of these three villages with each other is given in Figure 27. Of these three villages Sodomon KU has the highest relation with other languages: 82 PSC with the Nabay language as spoken in Patikang Laut KU and Kadalakan KU. Its relation with the Central Murut languages is otherwise 67-79 PSC and with the other Murutic languages 56-74 PSC.

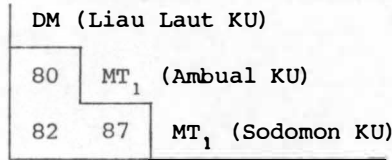


Figure 27: PSC relations of three villages comprising the Dusun-Murut language. (DM = Dusun-Murut; MT₁ = "Murut" representing Dusun-Murut.)

2.3.9.3.8(f) Sook Murut language

The Sook Murut language (MT₄) is here represented by two wordlists from Nabawan PN and Sook KU which have 83 PSC. The highest relation that either has with any other wordlist is 80 PSC between the former and Murut (here called Baukan) as spoken in Inarad KN, but its relation with the Central Murut languages is otherwise 69-79 PSC, and with the other Murutic languages 60-73 PSC.

2.3.9.3.8(g) Baukan language

The Baukan language (MT₂) is here represented by three wordlists from Kokoroton KN, Inarad KN and Tulid KU - only the last was stated as being Baukan (the others were "Murut"). The PSC relation of these three villages is given in Figure 28. Apart from these, the highest relation of these villages elsewhere is Tulid KU with the Nabay language as spoken at Masak KU at 81 PSC. There are relations of 80 PSC with Sook Murut and other Nabay. Their relation with the Central Murut languages is otherwise 71-79 PSC and with the other Murutic languages 60-73 PSC.

The "Baukan" (BN) of Limbawan KU is classified with the Nabay language (see the following paragraph).

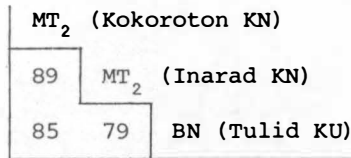


Figure 28: PSC relations of three villages comprising the Baukan language. (MT₂ = Murut, representing Baukan; BN = Baukan.)

2.3.9.3.8(h) Nabay language

The Nabay language (NY) is here represented by seven wordlists. The villages and PSC relations are shown in Figure 29. Nabay of Patikang Laut KU, representative of this group, has the highest relation with Dusun-Murut of 82 PSC, but with the other Central Murut languages at 70-79 PSC and with the other Murutic languages at 58-73 PSC. Though the dialect of Limbawan KU was stated to be Baukan, lexically it is seen here to be equivalent to Nabay.

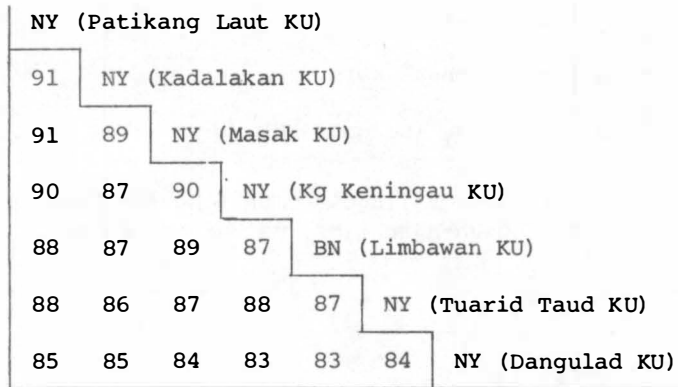


Figure 29: PSC relations of seven villages comprising the Nabay language. (NY = Nabay; BN = Baukan.)

2.3.9.4. Dusunic family

Of the 344 wordlists included in this study 165 wordlists are classified as Dusunic. The Dusunic family is widespread, populous and influential so this high number of wordlists is not out of proportion to the linguistic situation in Sabah.

The Dusunic family of languages consists of ten languages: Papar, Dumpas, Kadazan-Tambanua, Lotud, Bisaya, Tatana, Kuijau, Eastern Kadazan, Rungus, and Kadazan/Dusun. Most of the PSC relations between these various languages are in the 60-75% range - the depth of subfamily relationship, emphasising their distinctness. The Kadazan/Dusun language has 13 notable dialects in this analysis. The PSC relations of these Dusunic languages and of some of the Kadazan/Dusun dialects are given in Figure 30.

2.3.9.4.1 Papar language

The Papar language (PR) is represented by only one wordlist from Tinambak KP. Its highest relation with any other wordlist is 72 PSC with the Tatana language as spoken in Bundu KP, but its relation with the Dusunic languages is otherwise in the range of 51-66 PSC. Its relation with languages of the Murutic family (Section 2.3.9.3) ranges from 53-66 PSC. Speakers of the Papar language are sometimes referred to as Bajau Bukit, but the language relates to West Coast Bajau at a range of only 39-44 PSC.

2.3.9.4.2 Dumpas language

The Dumpas language (DS) is represented here by only one wordlist from Perancangan LS. Its highest relation with any other wordlist is 80 PSC with "Sungai" (Eastern Kadazan) of Buang Sayang KN, but its relation with the Dusunic languages is otherwise in the range of 53-75 PSC. Its relationship with the Paitanic family of languages (Section 2.3.9.2) is 61-79 PSC.

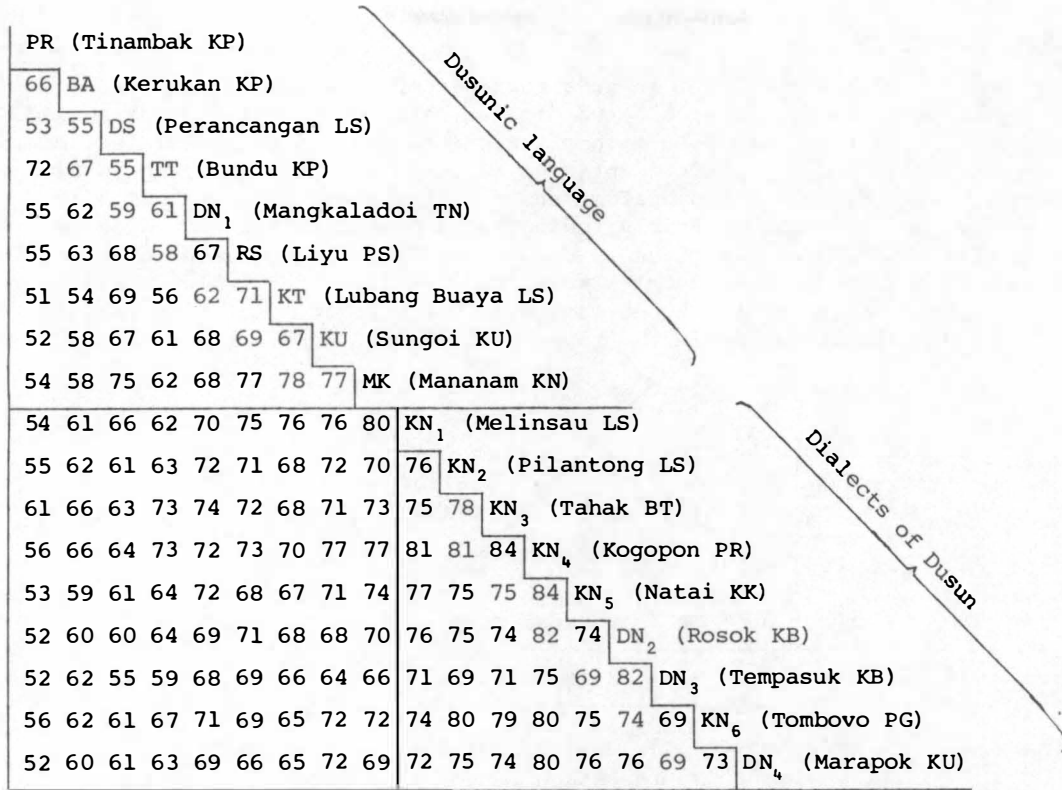


Figure 30: PSC relations of nine Dusunic languages including some dialects of the Dusun language. (PR = Papar; BA = Bisaya; DS = Dumpas; TT = Tatana; DN₁ = "Dusun" representing Lotud; RS = Rungus; KT = Kadazan-Tambanua; KU = Kuijau; MK = "Mangkaak" representing Eastern Kadazan dialect; KN₁ = "Kadazan" representing Sugut Kadazan dialect; KN₂ = "Kadazan" representing Pilantong Kadazan dialect; KN₃ = "Kadazan" representing Beaufort Kadazan dialect; KN₄ = "Kadazan" representing Central Dusun dialect; KN₅ = "Kadazan" representing Kota Kinabalu Kadazan dialect; DN₂ = "Dusun" representing Kota Belud Dusun dialect; DN₃ = "Dusun" representing Tempasuk (KB) Dusun dialect; KN₆ = "Kadazan" representing Tombovo (PG) Kadazan dialect; DN₄ = "Dusun" representing Keningau Dusun dialect.)

2.3.9.4.3 Kadazan-Tambanua language

The Kadazan-Tambanua language (KT) is here represented by only one wordlist from Lubang Buaya LS. Its highest relation with any other wordlist is 78 PSC with Mangkaak (Eastern Kadazan) of Mananam KN, but its relation with the Dusunic languages is otherwise in the range of 51-76 PSC. Its relation to the Tambanua language (Section 2.3.9.2.4) is 64-67 PSC.

2.3.9.4.4 Lotud language

The Lotud language (DN₁, LD) is here represented by six wordlists. The Lotud language has two dialects; a Lotud dialect chain found only in Tuaran and represented here by five wordlists each of which are related to the dialect chain by at least 85 PSC, and a dialect identified as Dusun Kadayan as spoken in Sumagit KK which is related to the Lotud dialect chain in the range of 80-82 PSC - an admittedly low threshold and possibly indicating a distinct language. The PSC relations of these dialects is given in Figure 31. The language as spoken in Mangkaladoi TN has the highest average relation with the other villages and is chosen as representative of the Lotud language in Figure 30. Lotud is related to the other Dusunic languages in the range of 55-74 PSC.

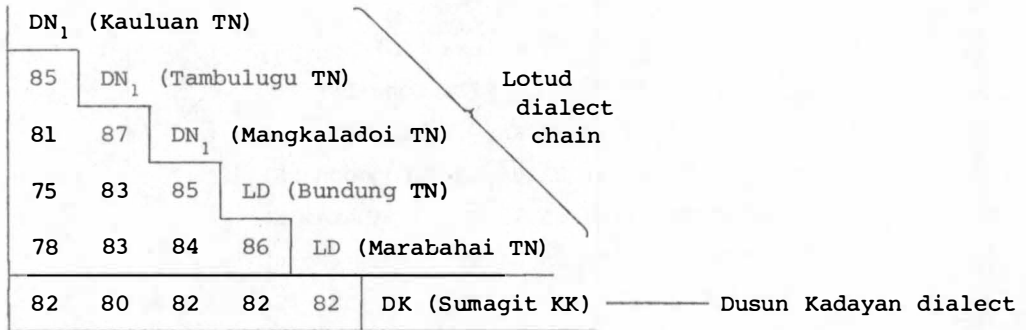


Figure 31: PSC relations of six villages of the Lotud language.
 (DN₁ = "Dusun" representing Lotud; LD = Lotud; DK = Dusun Kadayan.)

2.3.9.4.5 Bisaya language

The Bisaya language (BA) of Beaufort and Kuala Penyu districts is represented here by seven wordlists which indicate a close-knit language without dialect distinctions. The wordlist from Kerukan KP is chosen as representative of the language for Figure 30. Its relation with the other Dusunic languages is in the range 54-67 PSC. The PSC relation of the Bisaya villages is given in Figure 32.

A sample Bisaya wordlist from Limbawang River, Sarawak has 64-68 PSC with the Sabah Bisaya wordlists.

Because of the frequent query, a Philippine Visaya wordlist was obtained and compared with the Sabah Bisaya language. "Binisaya" of Nasuli, Bukidnon Province, Philippines has only 31-33 PSC with the Sabah Bisaya wordlists.

2.3.9.4.6 Tatana language

The Tatana language (TT) spoken by two-thirds of the population of Kuala Penyu District is here represented by two wordlists from Menunggang KP and Bundu KP which have 83 PSC. The former of the two understandably has an elevated 86 PSC relation with the Bisaya language wordlist taken in the same village indicating that the two language communities have borrowed words from each other. All the other PSC figures of this Tatana village with the other Bisaya wordlists

Included among the Murutic languages is Apin-Apin Kuijau (KU_m). It is included in Figure 33 here for comparison and is notably more remote: with the Dusunic Kuijau language Apin-Apin has 64-70 PSC whereas with the Murutic languages it has 54-77 PSC. It is not clear whether Apin-Apin Kuijau, though it is now closer to some Murutic languages than to any Dusunic language, was basically a Dusunic language which has borrowed from its Murutic neighbours, a Murutic language which has borrowed from its Dusunic neighbours, or a hybridisation evolving within a community composed of speakers of both languages. (Though survey procedures called for the exclusion of borrowed words, it was difficult in languages like Kuijau to determine which words were borrowed pending further study.)

2.3.9.4.8 Eastern Kadazan language

The Eastern Kadazan language (KN_7) spoken in Kinabatangan, Labuk-Sugut and Sandakan districts is here represented by 18 wordlists known variously as Kadazan, Mangkaak, Sukang and Sungai. The following nine villages form a single "Mangkaak" dialect because all their PSC relations are within 83-99%. The three villages with ** are used later as representative of this dialect:

"Mangkaak" (MK): **Mananam KN
Langkabung KN
"Sukang": **Entilibon KN
Karamuak KN
"Kadazan": **Telupid LS
Kuala Sapi LS
Panimbanan LS
Kiabau LS
Buis LS

The following four villages form a single Eastern Kadazan Sungai dialect because all their PSC relations are within 84-91%; the last is later used as representative of the dialect:

"Sungai": Batu Putih KN
Balat KN
Kuala Lokan KN
**Buang Sayang KN

There are five other dialects each represented here by only one wordlist. The PSC relation of the seven dialects of Eastern Kadazan are shown in Figure 34. Using the Mangkaak village of Mananam KN as representative of the Eastern Kadazan language because it has the highest average relation with the other Eastern Kadazan dialects, the relation of the Eastern Kadazan language with other Dusunic languages is 54-80 PSC; the highest relation is with the Sugut Kadazan dialect of the Kadazan/Dusun language.

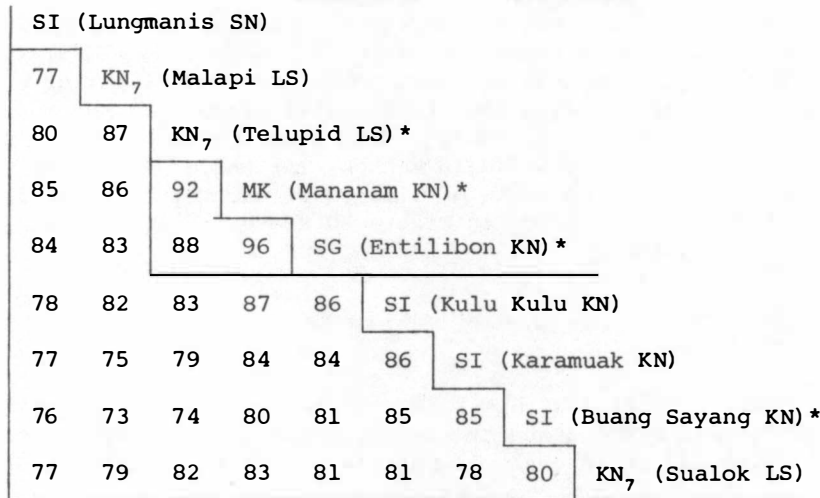


Figure 34: PSC relations of seven dialects of the Eastern Kadazan language. (* indicates village representing more than one location. SI = Sungai; KN₇ = "Kadazan" representing Eastern Kadazan; MK = Mangkaak; SG = Sukang.)

2.3.9.4.9 Rungus language

The Rungus language (RS) spoken in Kudat, Pitas, and Labuk-Sugut districts, is here represented by 12 wordlists. The Rungus language, as indicated by these samples, is a close-knit well-defined language without notable dialects. All the PSC relations are within 86-96%. The village samples are:

- | | | |
|--------------------|------------|-------------|
| Tinangol KT | Dallas PS | Batangun LS |
| Taringai KT | Liyu PS | |
| Kimihang KT | Rokom PS | |
| Lotong KT | Sinukab PS | |
| Lodung KT | | |
| Barambangan KT | | |
| Indarasan Darat KT | | |

Using Liyu PS as representative because it alone has 90 PSC-plus with every other Rungus wordlist, the Rungus language has a relation of 55-77 PSC with the other Dusunic languages.

(Though Bintasan KM village was identified as "Rungus" it is here classified as Kota Belud Dusun dialect of the Dusun language; see 2.3.9.4.10(g) below.)

2.3.9.4.10 Kadazan/Dusun language

Kadazan and Dusun are ethnonyms representing communities of speakers of the Kadazan/Dusun language. The Kadazan/Dusun language is represented here by 110 wordlists. Each wordlist has a relation of at least 85 PSC with another wordlist classified in this language except for the three wordlists (dialects) described in 2.3.9.4.10(b), 2.3.9.4.10(c) and 2.3.9.4.10(d) below. The Dusun language is

spoken in over half of the districts of the State of Sabah: Kota Kinabalu, Papar, Kota Belud, Tuaran, Labuk-Sugut, Ranau, Kinabatangan, Tawau, Keningau, Tambunan, Tenom, Beaufort, Penampang, and Kota Marudu. There are 13 dialects of Dusun: Central Dusun, Keningau Dusun, Pilantong (LS) Kadazan, Tempasuk (KB) Dusun, Tombovo (PG) Kadazan, Kota Kinabalu Kadazan, Kota Belud Dusun, Beaufort Kadazan, Sugut Kadazan, Paginatan (RU) Dusun, Kuala Monsok (TA) Dusun, Pemilaan (TM) Dusun and Bungaraya (KU) Dusun. Representative villages of each of these dialects except for the last four are included in Figure 30 above; their PSC relations are within the range of 69-84 PSC.

2.3.9.4.10(a) Central Dusun dialect

The Central Dusun dialect (KN₄, DN₅) is somewhat arbitrarily established by including 74 villages whose wordlists are strongly interrelated by 85 PSC. There are both core villages and peripheral villages in the Central Dusun dialect.

Fourteen core villages (KN_{4C}, DN_{5C}) are identified as those whose wordlists are related by 85 PSC-plus with 25 or more other Central Dusun wordlists; the ethnonyms given in these villages are Dusun, Ulu Tuaran, Dusun Sinulihan and Kadazan.¹²

"Dusun":	"Ulu Tuaran":
Panaitan KM	Kipouvu PG
Taginambur KB	Lumpoho TN
Sunsuron TA	
Togop Darat RU	"Dusun Sinulihan":
Bundu Tuhan RU	Notoruss PG
Lumat BT	
Batu 60 BT	"Kadazan":
Kionsom KK	Kogopon PR
Kiulu Baru TU	Ulu Lumagar PR

Sixty peripheral villages (KN_{4P}, DN_{5P}) were identified as those whose wordlists are related by 85 PSC-plus with at least one of the above core villages. Some of the villages listed below are followed with an alphabetic code; these are the villages to which the Dusun dialects described below are most closely related indicating something of the interrelations of these peripheral villages to each other and to the Central Dusun dialect. The ethnonyms given in these peripheral villages are Dusun, Kadazan, Tebilung, Pahu, Tindal, Kiundu, Sokid and Kadazan-Tagaro.

"Dusun":		"Kadazan":	
Bongkud RU	i-j-m	Mandalipau PR	
Parancangan RU	i-j-m	Limputong PR	h
Randagong Lama RU	m	Penampang Baru PR	e
Lohan RU		Sabandil PR	h
Pinampadan RU		Lingan PR	
Kilimu RU	j	Himbutong PR	e
Toboh RU		Limbahau	h
Kundasang RU	m	Gana-Manggis PR	h
Kibbas RU		Gadung PR	h
Sagindai RU	i-j	Penampang PG	e
Nalapak RU	i-j	Bunduon PG	e
Kinapulidan RU		Tuavon PG	e

Taiwan KM		Terawi PG	e
Lotong KM		Inobong PG	e
Tangkol KM		Pogunon PG	e
Piasau KB	g	Kipovu PG	
Kelawat KB	g	Moyog PG	
Bangkahak KB	g	Babagon PG	
Kitau TA	i-m	Sugud PG	e
Tangaban TA		Timpanggoh Sugud PG	
Toboh TA		Buayan PG	
Jimpanga BT	i-k-l-m	Lembah Poring BT	e
Berdatuk KK		Takapan BT	h
Pulutan KK		Tandui BT	h
Paniang TN	f	Talungan KK	f
Entilibon KN		T.A.S.S. TU	e
Pahu TA			
"Tebilung":		"Pahu":	
Mengaris KM	g	Longkogunon PG	
"Tindal":		"Kiundu":	
Tenghilan TN		Topokoon TN	
Malangang Baru TN			
"Sokid":		"Kadazan-Tagaro":	
Balabakan TN		Bambangan KK	

Those Dusun language wordlists not included in the core or periphery of the Central Dusun dialect above are assigned to other Dusun dialects as classified in the following paragraphs. Those dialects described in paragraphs 2.3.9.4.10(b) through 2.3.9.4.10(d), being less than 85 PSC with other dialects, are distinct Dusun dialects; those dialects described in paragraphs 2.3.9.4.10(e) through 2.3.9.4.10(m), being related to the peripheral villages by 85 PSC-plus, form a dialect chain relationship.

The PSC relations of nine of these Dusun dialects to each other and to the other Dusunic languages are given above in Figure 30.

2.3.9.4.10(b) Keningau Dusun dialect

The Keningau Dusun dialect (DN₄) is here represented by only one wordlist from Marapok KU. Its highest relation to another wordlist is 84 PSC with Beaufort Dusun dialect as spoken in Lumat BT, but its relation with the Dusun dialects is otherwise in the range 69-80 PSC and with the other Dusunic languages 60-72 PSC.

2.3.9.4.10(c) Pilantong (LS) Kadazan dialect

The Pilantong (LS) Kadazan (KN₂) dialect is here represented by only one wordlist from Pilantong LS. Its highest relation to another wordlist is 84 PSC with the "Kadazan" (Central Dusun dialect) spoken in Ulu Lumagar PR (these LS residents migrated from Penampang District in 1976), but its relation with the Dusun dialects is otherwise in the range 69-81 PSC, and with the other Dusunic languages 61-72 PSC.

2.3.9.4.10(d) Tempasuk (KB) Dusun dialect

The Tempasuk (KB) Dusun (DN₃) dialect is here represented by only one wordlist from Tempasuk KB. Its highest relation to another wordlist is 82 PSC with the Kota Belud Dusun dialect spoken in Piasau, Rosok and Tambulian Laut (all KB), but its relation with the Dusun dialects is otherwise 69-76 PSC and with the other Dusunic languages 55-69 PSC.

2.3.9.4.10(e) Tombovo (PG) Kadazan dialect

This and the next eight Dusun dialects are defined by those wordlists which are not within 85 PSC with a core Central Dusun dialect village described above, but are within 85 PSC of a peripheral Central Dusun dialect. Each of these dialects is represented by from one to 12 wordlists and is related to a specific group of from one to 11 wordlists of the peripheral group (as indicated in the listing of peripheral villages above) giving a distinct character to each dialect.

The Tombovo (PG) Kadazan dialect (KN₆) is here represented by two wordlists from Tombovo PG and Kolopis PG which have 85 PSC. These are related by at least 85 PSC with the 11 peripheral Central Dusun villages noted above with the code (e) - which are "Kadazan" from Papar, Penampang, Beaufort and Tawau. But neither of these wordlists is within that range of a core Central Dusun village. These PSC relations are shown in Figure 35. In this and the following figures the number in the bottom left corner box is less than 85 - indicating the need for a linking dialect to form the dialect chain to the Central Dusun core. Not all numbers in the bottom right box are 85-plus - because not all peripheral villages are within that range of all core villages.

KN ₆ (Tombovo PG)					
85	KN ₆ (Kolopis PG)				
90	84	KN _{4p} (Penampang Baru PR)			
87	86	90	KN _{4p} (Penampang PG)		
90	88	93	92	KN _{4p} (Lembah Poring BT)	
85	82	89	85	91	KN _{4p} (T.A.S.S. TU)
80	82	86	83	87	84 KN _{4c} (Kogopon PR)

Figure 35: PSC relations of two Tombovo (PG) Kadazan dialect villages with representative peripheral and core Central Dusun dialect villages. (KN₆ = "Kadazan" representing Tombovo (PG) Kadazan dialect; KN_{4p} = "Kadazan" representing peripheral Central Dusun dialect; KN_{4c} = "Kadazan" representing core Central Dusun dialect.)

2.3.9.4.10(f) Kota Kinabalu Kadazan dialect

The Kota Kinabalu Kadazan dialect (KN₅) is here represented by one wordlist from Natai KK. This is related by at least 85 PSC with the two peripheral Central Dusun villages noted above with the code (f) - which are "Kadazan" from Kota Kinabalu District and "Dusun" from Tuaran District. These PSC relations are shown in Figure 36.

KN ₅ (Natai KK)			
88	KN _{4P} (Paniang TN)		
91	93	KN _{4P} (Talungan KK)	
84	87	87	KN _{4C} (Kogopon PR)

Figure 36: PSC relations of a Kota Kinabalu Kadazan dialect village with peripheral and core Central Dusun dialect villages.

(KN₅ = "Kadazan" representing Kota Kinabalu Kadazan dialect;
 KN_{4P} = "Kadazan" representing peripheral Central Dusun
 dialect; KN_{4C} = "Kadazan" representing core Central
 Dusun dialect.)

2.3.9.4.10(g) Kota Belud Dusun dialect

The Kota Belud Dusun (DN₂) dialect is here represented by seven wordlists from Kota Belud District, six being "Dusun" and one, "Rungus". These are related by at least 85 PSC with the four peripheral Central Dusun villages noted above with the code (g) - three of which are "Dusun" from Kota Belud District and one "Tebilung" from Kota Marudu District. These PSC relations are shown in Figure 37.

2.3.9.4.10(h) Beaufort Kadazan dialect

The Beaufort Kadazan dialect (KN₃) is here represented by six wordlists from Beaufort District. These are related by at least 85 PSC with the seven peripheral Central Dusun dialect villages noted above with the code (h) - which are "Kadazan" from Papar and Beaufort districts. These PSC relations are shown in Figure 38.

RS (Bintasan KM)												
86	DN ₂ (Dudar KB)											
86	91	DN ₂ (Taburan KB)										
80	86	86	DN ₂ (Mantanau KB)									
76	81	80	85	DN ₂ (Mandap KB)								
81	84	83	86	88	DN ₂ (Rosok KB)							
80	82	81	87	89	90	DN ₂ (Tambulian Laut KB)						
TG (Mengaris KM)												
79	82	81	86	86	90	91	81	DN _{5p} (Piasau KB)				
78	81	79	84	90	87	88	81	87	DN _{5p} (Kelawat KB)			
80	83	82	86	87	93	89	83	89	85	DN _{5p} (Bangkahak KB)		
81	80	79	83	84	83	83	84	85	85	83	DN _{5c} (Taginambur KB)	

Figure 37: PSC relations of seven Kota Belud Dusun dialect villages with peripheral and core Central Dusun dialect villages. (RS = Rungus; DN₂ = "Dusun" representing Kota Belud Dusun dialect; TG = Tebilung; DN_{5p} = "Dusun" representing peripheral Central Dusun dialect (equivalent to KN_{4p}); DN_{5c} = "Dusun" representing core Central Dusun dialect (equivalent to KN_{4c}.)

KN ₃ (Tahak BT)										
86	KN ₃ (Takuli BT)									
88	83	KN ₃ (Tibabar-Sumbiling PR)								
89	85	89	KN ₃ (Pinopok BT)							
92	87	89	91	KN ₃ (Mandangin BT)						
86	89	84	84	86	KN ₃ (Kinamam BT)					
KN _{4p} (Limputong PR)										
92	97	91	92	93	86	87	KN _{4p} (Takapan BT)			
KN _{4c} (Kogopon PR)										
84	84	84	83	84	81	88	85	KN _{4c} (Kogopon PR)		

Figure 38: PSC relations of six Beaufort Kadazan dialect villages with peripheral and core Central Dusun dialect villages. (KN₃ = "Kadazan" representing Beaufort Kadazan dialect; KN_{4p} = "Kadazan" representing peripheral Central Dusun dialect; KN_{4c} = "Kadazan" representing core Central Dusun dialect.)

2.3.9.4.10(i) Sugut Kadazan dialect

The Sugut Kadazan dialect (KN_1 , DN_6) is here represented by 12 wordlists which are divided into two subdialects based upon their relative distance through the dialect chain from the Central Dusun dialect. The wordlists of the Inner Sugut Kadazan subdialect are within 85 PSC of six peripheral Central Dusun villages which are noted with the code (i) - which are "Dusun" in Ranau, Tambunan and Beaufort. The nine Inner Sugut Kadazan subdialect villages are:

- "Kadazan" (KN_{1i}):
 Kaingaran LS
 Melinsau LS
 Basai LS
 Kaingaran Baru LS
 Tinangian LS
 Karaganan LS (the foregoing six villages are interrelated at 90 PSC-plus; Melinsau LS is representative in the accompanying figures)
- "Tilau-Ilau" (TU):
 Basai LS
- "Minokok" (MO):
 Moiwod KN
 Entilibon KN (the foregoing two villages have 96 PSC; Moiwod KN is representative)

The three wordlists of Outer Sugut Kadazan subdialect are not within 85 PSC of a peripheral Central Dusun dialect village but are linked through the Inner Sugut Kadazan subdialect and include:

- "Dusun" (DN_{6o}):
 Parong KM
- "Kimaragang" (KG):
 Dandun PS
- "Garo" (GO):
 Talantang KM

The PSC relations of the two Sugut Kadazan subdialects are shown in Figure 39.

2.3.9.4.10(j) Paginatan (RU) Dusun dialect

The Paginatan (RU) Dusun dialect (DN_7) is here represented by only one wordlist which has 85-86 PSC with the five peripheral Central Dusun dialect villages noted above with the code (j) - which are all "Dusun" from Ranau. The PSC relations are shown in Figure 40.

GO (Talang KM)										
80	KG (Dandun PS)									
85	86	DN _{6O} (Parong KM)								
83	81	87	TU (Basai LS)							
82	76	82	83	MO (Moiwod KN)						
83	81	87	87	85	KN _{1i} (Melinsau LS)					
84	74	80	82	83	85	DN _{5P} (Bongkud RU)				
83	76	81	84	85	85	87	DN _{5P} (Jimpanga BT)			
79	75	80	80	82	81	82	83	KN _{4C} (Kogopon PR)		
83	73	78	77	80	80	87	82	84	DN _{5C} (Togop Darat RU)	
80	73	77	79	81	79	82	86	87	85	DN _{5C} (Batu 60 BT)

Figure 39: PSC relations of three Outer and three representative Inner Sugut Kadazan dialect villages with peripheral and core Central Dusun dialect villages. (GO = Garo; KG = Kimaragang; DN_{6O} = "Dusun" representing Outer Sugut Kadazan dialect; TU = Tilau-Ilau; KN_{1i} = "Kadazan" representing Inner Sugut Kadazan dialect; DN_{5P} = "Dusun" representing peripheral Central Dusun dialect; KN_{4C} = "Kadazan" representing core Central Dusun dialect; DN_{5C} = "Dusun" representing core Central Dusun dialect.)

DN ₇ (Paginatan RU)		
85	DN _{5P} (Bongkud RU)	
83	87	DN _{5C} (Togop Darat RU)

Figure 40: PSC relations of the Paginatan (RU) Dusun dialect village with peripheral and core Central Dusun dialect villages. (DN₇ = "Dusun" representing Paginatan (RU) Dusun dialect; DN_{5P} = "Dusun" representing peripheral Central Dusun dialect; DN_{5C} = "Dusun" representing core Central Dusun dialect.)

2.3.9.4.10(k) Kuala Monsok (TA) Dusun dialect

The Kuala Monsok (TA) Dusun dialect (DN₈) is here represented by only one wordlist which has 85 PSC with the one peripheral Central Dusun dialect village noted above with the code (k) - which is "Dusun" in Beaufort. The PSC relations are shown in Figure 41.

DN ₈ (Kuala Monsok TA)		
85	DN _{5p} (Jimpanga BT)	
83	86	DN _{5c} (Batu 60 BT)

Figure 41: PSC relations of the Kuala Monsok (TA) dialect village with peripheral and core Central Dusun dialect villages.
 (DN₈ = "Dusun" representing Kuala Monsok (TA) Dusun dialect;
 DN_{5p} = "Dusun" representing peripheral Central Dusun dialect; DN_{5c} = "Dusun" representing core Central Dusun dialect.)

2.3.9.4.10(1) Pemilaan (TM) Dusun dialect

The Pemilaan (TM) Dusun dialect (DN₉) is here represented by only one word-list which has 85 PSC with the one peripheral Central Dusun village noted above with the code (1) - which is "Dusun" in Beaufort. The PSC relations are shown in Figure 42.

DN ₉ (Pemilaan TM)		
85	DN _{5p} (Jimpanga BT)	
80	86	DN _{5c} (Batu 60 BT)

Figure 42: PSC relations of the Pemilaan (TM) Dusun dialect village with peripheral and core Central Dusun dialect villages.
 (DN₉ = "Dusun" representing Pemilaan (TM) Dusun dialect;
 DN_{5p} = "Dusun" representing peripheral Central Dusun dialect; DN_{5c} = "Dusun" representing core Central Dusun dialect.)

2.3.9.4.10(m) Bungaraya (KU) Dusun dialect

The Bungaraya (KU) Dusun dialect (DN₁₀) is here represented by two word-lists which have 87 PSC. The two villages are Bungaraya KU and Luanti Baru KU. One or both of them have 85-86 PSC with the six peripheral Central Dusun dialect villages noted above with the code (m) - which are all "Dusun" in Ranau, Tambunan and Beaufort. The PSC relations are shown in Figure 43.

DN ₁₀ (Luanti Baru KU)			
87	DN ₁₀ (Bungaraya KU)		
85	85	DN _{5p} (Randagong Lama RU)	
83	80	87	DN _{5c} (Togop Darat RU)

Figure 43: PSC relations of the two Bungaraya (KU) Dusun dialect villages with peripheral and core Central Dusun dialect villages. (DN₁₀ = "Dusun" representing Bungaraya (KU) Dusun dialect; DN_{5p} = "Dusun" representing peripheral Central Dusun dialect; DN_{5c} = "Dusun" representing core Central Dusun dialect.)

3. SUMMARY

The foregoing classification of languages spoken by residents whose cultural centres are in Sabah or in the South-east Asia insular world distinguishes two phyla of languages, four superstocks, 12 stocks, 16 families, 36 subfamilies and 51 languages (of which 33 are considered, pending further research, to have their cultural centre in Sabah). Language as well as dialect divisions noted above distinguish 83 differing forms of speech in Sabah. The overall language classification is summarised in Figure 44.

Within the language terminology there are various ambiguities, among which are:

- (1) "Jawa" represents five languages;
 "Bugis" represents two languages;
 "Bajau" represents two languages;
 "Murut" represents 15 languages;
 "Dusun" represents perhaps four languages and many dialects;
 "Kadazan" represents perhaps four languages and many dialects.
- (2) "Tambanua" represents the Tambanua language (Paitanic family) as well as the Kadazan-Tambanua language (Dusunic family).
- (3) "Kuijau" represents the Kuijau language (Dusunic) as well as the Apin-Apin Kuijau language (Murutic).
- (4) "Baukan" represents the Baukan language (Murutic) as well as a subdialect of the Nabay language (Murutic).
- (5) "Rungus" represents the Rungus language (Dusunic) as well as a subdialect of the Kota Belud Dusun dialect of the Dusun (Kadazan) language.
- (6) "Sungai", the most misleading term, being a cultural rather than linguistic designation, represents:
 - (a) Ida'an Sungai language (Ida'an subfamily);
 - (b) Abai Sungai language (Paitanic family);
 - (c) Tambanua Sungai dialect of Tambanua language (Paitanic family);
 - (d) Gum-Gum Sungai dialect of Upper Kinabatangan language (Paitanic family);
 - (e) Kuamat Sungai of "Makiang" dialect of Upper Kinabatangan language (Paitanic family);
 - (f) Sungai dialect of Eastern Kadazan language (Dusunic family).

Phylum	Super-stock	Stock	Family	Sub-family	Language	No. of dialects
0-15 PSC	15-25 PSC	25-45 PSC	45-60 PSC	60-75 PSC	75-80 PSC	80-85 PSC
1. Indo-European					Chabacano	
2. Austronesian						
A.					Butung	
B. Javanese stock					Jawa TM	
Javanese family					Jawa BT	
					Jawa TU	
Javanese sub-family					Jawa SN	
					Jawa LD	
C North-western Austronesian superstock						
1.					Lundayeh	
2.					*Banggi	
3.					*Illanun	2
4.					Suluk	
5. Bugis sub-family					Bugis TU	
					Bugis LD/SN	2
6. Ida'an sub-family					*Ida'an/Begahak	
					*Ida'an Sungai	4
7. Malayic family					a. Iban	
					b. Malayic subfamily	
					Cocos Malay/Bahasa Malaysia	
					Brunei/Kedayan	
8. Bajau family					*West Coast Bajau	chain
					East Coast Bajau	

Phylum	Super-stock	Stock	Family	Sub-family	Language	No. of dialects
		9. Bornean stock				
			a.		Tidong	chain
			b. Paitanic family		*Lingkabau	
					*Lobu	
					*Abai Sungai	
					*Tambanua	chain
					*Upper Kinabatangan	4
			c. Murutic family	(1)	Kolod	
				(2)	*Gana	
				(3)	*Apin-Apin Kuijau	
				(4)	*Kalabakan Murut	
				(5)	Sembakung Murut	
				(6)	*Serudung Murut	
				(7)	*Tagal	2
			(8) Central Murut subfamily			
					*Takapan	
					*Paluan	
					*Timugon	
					*Beaufort Murut	
					*Dusun-Murut	
					*Sook Murut	
					*Baukan	
					*Nabay	

Phylum	Super-stock	Stock	Family	Sub-family	Language	No. of dialects
			d. Dusunic family	(1)	*Papar	
				(2)	*Dumpas	
				(3)	*Kadazan-Tambanua	
				(4)	*Lotud	2
				(5)	*Bisaya	
				(6)	*Tatana	
				(7)	*Kuijau	5
				(8)	*Eastern Kadazan	7
				(9)	*Rungus	
				(10)	*Kadazan/Dusun	13
Total distinctions:						
2	4	12	16	36	51 *33	83

Figure 44: Language classification of Sabah. (* identifies languages whose cultural centre is considered to be in Sabah.)

Lexicostatistics is not a sufficient means of differentiating languages and dialects, if "languages" implies lack of mutual intelligibility and "dialect" the presence of mutual intelligibility (barring the case of a dialect chain within a language). The following points need to be clarified by dialect intelligibility testing:

- (1) intelligibility between the Jawa languages in Sabah (as well as enquiring whether more Javanese languages are present in Sabah);
- (2) intelligibility between the two Illanun dialects;
- (3) intelligibility between the Bugis dialects, whether there are one or two Bugis languages in Sabah;
- (4) intelligibility between Ida'an/Begahak and Ida'an Sungai languages, and between the four dialects of the latter;
- (5) intelligibility along the dialect chain of West Coast Bajau;
- (6) intelligibility along the dialect chain of East Coast Bajau and with Philippine Sama languages;
- (7) intelligibility of the three Tidong dialects;
- (8) intelligibility along the language chain of five Paitanic languages;

- (9) intelligibility between the two Tambanua dialects;
- (10) intelligibility along the dialect chain of the Upper Kinabatangan language;
- (11) intelligibility of Murutic Apin-Apin Kuijau with Dusunic Kuijau language;
- (12) intelligibility between the two Tagal dialects;
- (13) intelligibility between the eight Central Murut languages;
- (14) intelligibility between the two Lotud dialects;
- (15) intelligibility between the five Kuijau (Dusunic) dialects;
- (16) intelligibility between the seven Eastern Kadazan dialects;
- (17) intelligibility between the 13 Dusun (Kadazan) dialects.

Inter-language borrowing is evident at several points in Sabah:

- (1) Brunei language of Berhala Darat SN from Bahasa Malaysia;
- (2) East Coast Bajau language from the Suluk language;
- (3) West Coast Bajau language from the Malayic family;
- (4) Apin-Apin Kuijau from the Murutic languages;
- (5) Tatana language of Menunggang KP from Bisaya language of the same village.

Two appendices are included listing the ethnonyms used by the people of Sabah and listing the serial language abbreviations (KN₁, etc.).

NOTES

1. For similar studies of other linguistic areas and for methodological details of the lexicostatistical analysis see Smith (1974), Walton (1977), and Smith (1978).
2. This writer has only analysed the linguistic field data collected and encoded by the other members of the linguistic team. The field linguists who collected the wordlists and made the cognateness decisions are: Julie Blom, Patrick Cohen, Phyllis Dunn, Hope Hurlbut, John and Carolyn Miller, and Inka Pekkanen. These are all members of the Summer Institute of Linguistics. The technical aspects of the type of language survey undertaken by the Institute in Sabah is described in Casad (1974); the survey results from the Kudat Division are described in Blom (1979) and from the upper Kinabatangan River area in Hurlbut and Pekkanen (to appear).
3. The districts of Sabah, the abbreviations used in this paper for the districts, and the number of wordlists collected from each district are:

Beaufort	BT 26	Labuan	LN 2	Sandakan	SN 15
Keningau	KU 28	Labuk-Sugut	LS 33	Semporna	SA 8
Kinabatangan	KN 31	Lahad Datu	LD 20	Sipitang	SG 5
Kota Belud	KB 16	Papar	PR 19	Tambunan	TA 6
Kota Kinabalu	KK 9	Penampang	PG 18	Tawau	TU 14
Kota Marudu	KM 9	Pensiangan	PN 3	Tenom	TM 11
Kuala Penyu	KP 11	Pitas	PS 9	Tuaran	TN 13
Kudat	KT 9	Ranau	RU 16		
				total	331

4. The wordlist used for this study differs from the 372-item wordlist given in Reid (1971) which has been used extensively in the Philippines only by the exclusion of the following five items: abaca (Manila hemp), carabao, penis, vagina, charcoal.
5. For some widely occurring forms like bapa *father* only one root was found among all the wordlists for a given meaning; but for other meanings like *anger*, *fast* and *to fight* as many as 41 different roots were found among the wordlists.
6. The 367-item wordlist used in the field collection of data (see Note 4) was reduced for a variety of reasons by the following 40 meanings to a 327-item wordlist for the lexicostatistical comparison:

ankle	finger	repeat	throw
anus	hold	rub	today
blanket	if	slave	twenty
breakfast	lake	smooth	waterfall
climb	learn	soup	bamboo water container
cut	lonely	squeeze	water jar
delouse	nipah	string	weak
dust	none	summit	weave
earthquake	now	sweet potato	what-you-may-call-it
eel	raincloud	swidden	word

7. Dyen (1965:18) set a maximum of 69.9% for "the highest score between dialects of different languages" and termed it the Provisional Language Limit. The higher value used in this paper is justified by a presumed reduction of errors for the following reasons: (1) the wordlists were obtained by a small group of trained linguists interacting with each other to assure identical understanding of the basic wordlist; (2) each linguist completed an average of about 50 wordlists assuring familiarity with the intended meaning and expected form being elicited; (3) each linguist continually reviewed the isolated forms not identified with cognate sets looking for evidence of cognateness in the growing body of variant forms within each cognate set assuring maximum opportunity to assign isolated forms to established sets as well as to merge previously separate cognate sets; (4) the range of known variance within given cognate sets was monitored using Reid (1971) and Wurm and Wilson (1975); (5) questionable items on each wordlist were deleted from consideration; (6) as noted in Notes 4 and 6 additional items were deleted from any consideration. The effect of each of these procedures was to raise the overall cognate level. For example, the effect of deleting the 40 problem words (Note 6) was an average increase of 1.4 PSC; the effect of merging cognate sets upon evidence noted after the initial comparison were computed was an average increase of 3.8 PSC. This is not to deny a level of error of several percentage points arising from misunderstanding of the Bahasa Malaysia term by the language assistants, confusion of general versus specific terminology and other reasons.
8. Dyen (1965:18) likewise included dialect chains: "if a chain of percentages 70.0 or higher connects a set of lists, they are all assigned to the same language."
9. Here and at other places the classification of Sabah languages as part of the widespread Austronesian linguistic world may differ by a level or two from other classifications of these languages. This classification is based almost solely upon cognate percentages obtained from these languages in

Sabah without inclusion of data from outside of Sabah. The classification is thus more significant for the indigenous languages of Sabah than for the migrant languages recently introduced into Sabah.

10. The six Philippine wordlists were obtained from Reid (1971) or from the office of the Summer Institute of Linguistics in Manila.
11. Prentice (1971:3) adopts Appell's (1968) use of the term Idahan as a language family name and states: "The known languages of the Idahan family are at present divided into three sub-families: Paitanic, Dusunic and Murutic." In this analysis the Ida'an (Idahan) language is related to the Paitanic, Murutic and Dusunic families at a more remote level; so the name Idahan for this group seems inappropriate; Prentice himself has discontinued this use of the term Idahan, according to correspondence from him. Here "Bornean stock" is equivalent to Prentice's "Idahan family".
12. It is interesting to note that should one draw a circle on a map of Sabah choosing as a centre a point near Kg Tobobon TN and establishing a radius of 15 miles, six of the 14 core villages (Sunsuron TA, Bundu Tuhan RU, Kionsom KK, Lumpoho TN, Kipouvu PG, and Notoruss PG) would fall within the circle.

Five of the villages (Lumat BT, Kiulu Baru TU, Panaitan KM, Batu 60 BT, and Taginambur KB) migrated from areas within the circle.

Two villages (Ulu Lumagar PR and Kogoapon PR) have some residents who came from areas within the circle. Only one village (Togop Darat RU) has no known connection with areas within the circle and lies about 21 miles from the edge of the circle.

APPENDIX I: Ethnonyms: an alphabetical listing

Bajau	two languages of Bajau family
Bajau Banaran	dialect of East Coast Bajau dialect chain
Bajau Balangingi	dialect of East Coast Bajau dialect chain
Bajau Darat	dialect of East Coast Bajau dialect chain
Bajau Laut	dialect of East Coast Bajau dialect chain
Bajau Semporna	dialect of East Coast Bajau dialect chain
Banggi	language of North-western Austronesian superstock
Baukan	language of Central Murut subfamily; also dialect of Nabay language
Begahak	language of Ida'an subfamily; same language as Ida'an
Bisaya	language of Dusunic family
Brunei	language of Malayic subfamily; same language as Kedayan
Brunei-Kedayan	subdialect of Brunei/Kedayan language of Malayic subfamily
Bugis	two languages of North-western Austronesian superstock from Sulawesi

Butung	language of Austronesian phylum from Sulawesi
Chabacano	Indo-European creole language from Zamboanga, Philippines
Cocos	language of Malayic subfamily from Cocos Islands; same language as Bahasa Malaysia; also Cocos Malay
Dumpas	language of Dusunic family
Dusun	language of Dusunic family; same language as Kadazan; refers also to other Dusunic languages and dialects
Dusun Kadayan	dialect of Lotud language
Dusun-Murut	language of Central Murutic subfamily
Dusun Segama	dialect of Upper Kinabatangan language
Gana	language of Murutic family
Garo	part of Outer Sugut Kadazan subdialect
Iban	language of Malayic family, from Sarawak
Ida'an	language of Ida'an subfamily; same language as Begahak
Illanun	language of North-western Austronesian superstock
Jama Mapun	dialect of East Coast Bajau dialect chain; also found in Philippines
Jawa	five languages of Javanese stock; from Indonesia
Jawa Halus	High Javanese dialects; from Indonesia
Kadazan	language of Dusunic family; same language as Dusun; refers also to other Dusunic languages and dialects
Kadazan-Tambanua	language of Dusunic family
Kagayan	dialect of East Coast Bajau dialect chain
Kalabakan Murut	language of Murutic family
Kedayan	language of Malayic subfamily; same language as Brunei
Kimaragang	part of Outer Sugut Kadazan subdialect
Kiundu	part of peripheral Central Dusun dialect
Kolobuan	part of "Makiang" dialect of Upper Kinabatangan language
Kolod	language of Murutic family
Kuijau	language of Dusunic family; also language of Murutic family
Laminusa	dialect of East Coast Bajau dialect chain
Lingkabau	language of Paitanic family
Lobu	language of Paitanic family
Lotud	language of Dusunic family
Lundayeh	language of North-western Austronesian superstock
Lun Lod	subdialect of Lundayeh language
Makiang	dialect of Upper Kinabatangan language, cf. Kolobuan, Sinarupa, Sungai

Mangkaak	dialect of Eastern Kadazan language
Minokok	part of Inner Sugut Kadazan subdialect
Murut	family of languages of Bornean stock, refers to various Murutic languages and dialects
Nabay	language of Central Murutic subfamily
Pahu	part of peripheral Central Dusun dialect
Paitan	subdialect of Tambanua language
Paluan	language of Central Murutic subfamily
Papar	language of Dusunic family
Rumanau	part of Sinabu/Rumanau, dialect of Upper Kinabatangan language
Rungus	language of Dusunic family, also subdialect of Kota Belud Dusun dialect
Sama	Philippine language, corresponds to Sabah East Coast Bajau
Sembakung Murut	language of Murutic family
Sibutu	dialect of East Coast Bajau dialect chain
Sikubung	dialect of East Coast Bajau dialect chain
Simunul	dialect of East Coast Bajau dialect chain
Sinabu	part of Sinabu/Rumanau dialect of Upper Kinabatangan language
Sinarupa	part of "Makiang" dialect of Upper Kinabatangan language
Sinulihan	part of core Central Dusun dialect
Sokid	part of peripheral Central Dusun dialect
Sukang	part of "Mangkaak" dialect of Eastern Kadazan language
Suluk	language of North-western Austronesian superstock; same language as Tausug from Philippines
Sungai	various communities on Sabah's East Coast rivers belonging to several languages and dialects
Tagal	language of Murutic family
Tagaro	part of peripheral Central Dusun dialect
Takapan	language of Central Murutic subfamily
Tambanua	language of Paitanic family; cf. Kadazan-Tambanua
Tatana	language of Dusunic family
Tausug	Philippine language of North-western Austronesian superstock, corresponds to Sabah Suluk
Tebilung	part of peripheral Central Dusun dialect
Tidong	language of Bornean stock
Tilau-Ilau	part of Inner Sugut Kadazan subdialect
Timugon	language of Central Murutic subfamily

Tindal	part of peripheral Central Dusun dialect
Tobu	refers to people who speak the Lobu language
Ubian	dialect of East Coast Bajau dialect chain
Ulu Tuaran	part of core Central Dusun dialect

APPENDIX II: Serial language abbreviations

DN ₁	"Dusun" representing	Lotud language
DN ₂	" "	Kota Belud Dusun dialect
DN ₃	" "	Tempasuk (KB) Dusun dialect
DN ₄	" "	Keningau Dusun dialect
DN ₅	" "	Central Dusun dialect (KN ₄)
DN _{5C}	" "	core Central Dusun dialect (KN _{4C})
DN _{5P}	" "	peripheral Central Dusun dialect (KN _{4P})
DN ₆	" "	Sugut Kadazan dialect (KN ₁)
DN _{6O}	" "	Outer Sugut Kadazan subdialect (KN _{1O})
DN ₇	" "	Paginatan (RU) Dusun dialect
DN ₈	" "	Kuala Monsok (TA) Dusun dialect
DN ₉	" "	Pemilaan (TM) Dusun dialect
DN ₁₀	" "	Bungaraya (KU) Dusun dialect
KN ₁	"Kadazan" representing	Sugut Kadazan dialect (DN ₆)
KN _{1i}	" "	Inner Sugut Kadazan subdialect (DN _{6O})
KN ₂	" "	Pilantong Kadazan dialect
KN ₃	" "	Beaufort Kadazan dialect
KN ₄	" "	Central Dusun dialect (DN ₅)
KN _{4C}	" "	core Central Dusun dialect (DN _{5C})
KN _{4P}	" "	peripheral Central Dusun dialect (DN _{5P})
KN ₅	" "	Kota Kinabalu Kadazan dialect
KN ₆	" "	Tombovo (PG) Kadazan dialect
KN ₇	" "	Eastern Kadazan language
KU	"Kuijau" representing	Kuijau language (Dusunic)
KU _m	" "	Apin-Apin Kuijau (Murutic)
MT ₁	"Murut" representing	Dusun-Murut language
MT ₂	" "	Baukan language
MT ₃	" "	Beaufort Murut language
MT ₄	" "	Sook Murut language
MT ₅	" "	Serudung Murut language
MT ₆	" "	Tagal language

PART II

THE LANGUAGES OF SABAH:
INTELLIGIBILITY TESTING

INTRODUCTION

Carolyn P. Miller

0. GENERAL BACKGROUND

The island of Borneo has long been of interest to those studying the movements of people throughout history and the diversity of languages and customs. Because of its strategic position in the large area of the world where Austro-nesian languages are spoken, stretching all the way from Madagascar to Easter Island and encompassing the entire greater South-east Asia region, and because of the evidence of early habitation of the island (Harrisson 1958), Borneo remains "one of the more interesting areas of the world, ethnologically speaking" (LeBar 1972:147).

In the reference work which he edited entitled *Ethnic groups of insular Southeast Asia*, Frank M. LeBar has stated that "considerable reliance has been placed on linguistic classification, since it is felt that demonstrated genetic relationships among languages remain the best indicators of present or past cultural ties among the speakers of those languages" (1972:v). Concerning the island of Borneo he states that a major obstacle to the attempt to synthesise descriptions of the ethnic groups of Borneo has been "the virtual absence of modern linguistic field surveys" (1972:147).

It is hoped that this volume will help toward a better understanding of the relationships among the languages spoken in the northern part of the island, that is in the state of Sabah, East Malaysia. Along with other articles already produced,¹ it presents the results of a linguistic survey of Sabah conducted by the Summer Institute of Linguistics in cooperation with the state government, of the languages spoken throughout the state.

In his article entitled "The languages of Sabah: a tentative lexicostatis-tical classification", Kenneth D. Smith (in this volume) presents the results of the first stage of the survey. Using a computer-assisted comparison of wordlists collected in 325 locations throughout the state, he proposed a tentative classification of Sabah languages. In that article he states that many of these proposed classifications need to be clarified by intelligibility testing.

The other articles present the results of such testing for most of the languages which Smith included in his classification. Excluded from testing are those languages whose linguistic centre is elsewhere than Sabah, such as Javanese, Bugis, Chabacano, Butung, and the various Chinese, Indian, and European languages. Also, the Kadazan-Tambanua language is not reported on here

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 51-57. *Pacific Linguistics*, C-78, 1984.

© Carolyn P. Miller

since that small linguistic community is adapting to the dominant Paitanic language in its area. And, the results of the Banggi language survey and intelligibility testing will be reported in a separate article (Boutin and Boutin, to appear).

In a final article, David Moody discusses the implications of the intelligibility testing on the tentative language classification proposed by Smith. Certain ambiguities and uncertainties remain, but much has been learned which we trust will lead to greater understanding of the linguistic diversity within the state of Sabah.

As far as is possible in a collection of articles by different authors, consistency in format and presentation of material has been maintained throughout the volume to assist the reader in comparing information from article to article.

Each article contains an introduction to give some background about the language group; a section on how the specific language was classified by Smith; test procedures; the results of the intelligibility testing; national language intelligibility testing results, and a conclusion.

Several devices are used throughout the articles which must be explained here to facilitate reading and study of the articles.

First of all, all village names are followed by a two-letter district abbreviation or in the case of villages outside of Sabah, a three-letter state, island or country abbreviation. Those abbreviations are as follows:

BT	Beaufort	PN	Pensiangan	IND	Indonesia
KB	Kota Belud	PR	Papar	KAL	Kalimantan
KK	Kota Kinabalu	PS	Pitas	PHL	Philippines
KM	Kota Marudu	RU	Ranau	SAR	Sarawak
KN	Kinabatangan	SA	Semporna		
KP	Kuala Penyu	SG	Sipitang		
KT	Kudat	SN	Sandakan		
KU	Keningau	TA	Tambunan		
LD	Lahad Datu	TM	Tenom		
LN	Labuan	TN	Tuaran		
LS	Labuk-Sugut	TU	Tawau		
PG	Penampang				

A second device used throughout the articles is that when referring to another article from this compendium an arrow followed by the language name will appear, as (→ Kadazan/Dusun).

1. SURVEY BACKGROUND

The Summer Institute of Linguistics (SIL) first began work in Sabah in June 1978 under a memorandum of understanding with the state government whereby SIL would, in part, undertake a linguistic survey of the indigenous languages of Sabah to determine their location, approximate size, and degree of difference; conduct field research in the lesser-known indigenous languages of Sabah including phonological, grammatical, and semantic studies; and make instrumental recordings of speech, music, and folklore.

An understanding of the languages spoken throughout the state and the relationship of these to each other as well as to languages spoken in other

parts of South-east Asia seemed important to future studies of specific languages of Sabah.

The purpose of the survey was to 1) determine dialect boundaries within defined geographical boundaries comprising the entire state of Sabah, 2) to determine more precisely via lexicostatistics and intelligibility testing the degrees of intelligibility across major and minor dialect boundaries, and 3) to attempt to determine the level of understanding and the extent of the use of the national language in villages across the state.

Before survey personnel visited villages in any district, a letter was sent by the Chief Minister's office to the District Officer introducing the SIL technicians and requesting help. Without the support of government officials and the extensive help they gave, such a survey would have been impossible. Generally District Officers wrote letters of introduction and set up appointments in the villages where there were language or dialect differences. Often someone from the district office who spoke the local dialect was assigned to accompany SIL personnel and assist them in communicating with the people in the villages and in transcribing and translating the material collected.

2. SURVEY METHOD

The basic procedures used in the survey were those presented by Eugene Casad in his handbook entitled *Dialect intelligibility testing* (1974). Stated very simply, this method required two trips to each language and dialect area under consideration. On the first trip (the collection trip), the technicians collected a phonetically transcribed list of 367 words in the local dialect using a standard list which had been prepared in English and Bahasa Malaysia.² They also collected samples of the speech of the people in that place by tape-recording short personal experience 'stories' from a member of each community visited. These accounts were then transcribed and translated into Bahasa Malaysia. Questionnaires and biographical data showing patterns of language use as well as relevant sociological information were also collected.

After material collected on this first trip was evaluated and compared with that collected in other locations, a second trip (the testing trip) was made to selected locations throughout the same area. On this second visit ten persons were asked to listen to tape-recorded 'stories' from other areas where the language was to a greater or lesser degree related to their own. Technicians attempted to get a cross-section of the community to take the test (i.e. young and old, male and female, educated and uneducated). Their understanding of a given story was calculated on the basis of their ability to answer questions concerning the content of the story.³ The questions were interspersed throughout the story to immediately follow the information being questioned. A story collected on the earlier trip from the test subject's own village (called the hometown tape) was played first in order to familiarise him with the testing method. If a subject did not score high on his hometown tape the technicians knew that he did not understand the testing procedures well. The hometown tape was either played again or the test subject was disqualified. The other local language tapes (called reference tapes) were then played, followed by the Bahasa Malaysia story tape (called the national language tape).

3. SELECTION OF TEST POINTS AND REFERENCE TAPES

The selection of test points and reference tapes to be played at each point was important to ensure the effectiveness of the testing procedure. A committee set up guidelines for selecting these.

Generally communities were not chosen as test points if the inhabitants had recently migrated from other areas.⁴ However, testing was done for Lundayeh (Sarawak Murut), even though most of them have been in Sabah for only one generation, to confirm their distinctness from Sabah Murut and to check for possible dialect distinctions between Lundayeh groups in Sabah.

In determining what tapes to play at a given test point, the committee considered three main factors: 1) the degree of difference between the two places as reflected in the percent of shared cognates (PSC) in the wordlists collected at each place, 2) the geographical proximity of the two places, and 3) sociolinguistic factors which might affect understanding.

Generally tapes were not selected for use in a particular location if the PSC between the two places was above 90 or below 70, since results from such testing would be predictable. That is, if PSC relations between wordlists from two villages were 90 or above high intelligibility was expected and if relations were below 70 PSC it was expected that intelligibility would be low.

Occasionally testing was done to determine the level of understanding between groups who seemed to be linguistically very close, but where geographical or sociological factors would seem to have kept them apart. For this reason Ida'an and Begahak speakers (→ Ida'an) were tested with tapes from each other's areas as were also Brunei and Kedayan speakers (→ Malayic).

Because of the difficulty test subjects had in listening to a large number of tapes, the committee decided to limit the total number of tapes to be played at any test point to seven, one of these being a hometown tape and one being the national language tape. This left a possible five tapes which could be used for testing the understanding of a range of related dialects and languages.

It was not always possible to keep all these factors in balance. For some language groups which were closely contained geographically and clearly defined linguistically with few dialect divisions, the choice was relatively simple, and it was easy to stay within the maximum number of tapes allowed for each test point. But in the larger, less well-defined groups such as Kadazan/Dusun the situation became more complex (→ Kadazan/Dusun).

4. DIFFERENCES FROM CASAD

The survey procedures used in Sabah differed in some ways from the method presented by Casad (1974). The method as described by Casad is executed by a trained linguist who is familiar with at least one of the dialects being tested. In Sabah survey teams were dependent on communication in the national language or on the services of an interpreter during the testing process.

The procedure as outlined by Casad suggests using two introductory tapes in the local dialect. The first is to explain the purpose of the survey and the second is to give a sample test story with questions. Though such tapes were prepared for use at the start of the Sabah survey, they were not used thereafter because of their limited effectiveness. Instead, either the SIL technicians explained the purpose and procedure of the survey in the national

language or the guide from the district did so in the local dialect. The hometown tape then became the introductory tape to the testing procedure.

The intelligibility testing results have also been presented differently from Casad's "cost matrix" which is then displayed on a contour map of networks. Casad suggests that establishing a threshold of 80 percent as the basis for proposing language distinctions is inadequate since other linguistic and sociological factors should also be taken into consideration. He recommends that the threshold be expressed as a "series of values corresponding to a range of intelligibility levels" (Casad 1974:46) lying between 75 and 85 percent. In this study, however, the threshold of 80 percent has generally been used in order to correspond with the threshold used by Smith in his lexicostatistical classification.⁵ Authors used either 85 or 90 percent as a threshold for distinguishing different dialects depending on which better fit the language situation with which they were dealing.

5. CHANGES IN STRATEGY

In several respects the survey method used in Sabah evolved as the survey progressed. Initially it was felt that the testing trip should proceed almost immediately after the collection trip. This procedure was followed in the testing of the Kudat Division (Blom 1979) and the upper Kinabatangan River area (Hurlbut and Pekkanen, to appear). After the technicians had collected the data, they returned shortly to the same area to test intelligibility between groups within that geographic area. However, because language boundaries seldom parallel geographic or administrative ones, it became apparent that not all relevant distinctions would be tested where other geographic areas were also involved, and that almost certainly retesting would have to be done later involving the broader language relationships. The decision was then made to complete collection of language data throughout the state before proceeding with further testing. When this was done, testing was mapped out on the basis of statewide comparisons. The articles of this volume which report the results are not done by geographical or administrative areas but by related language groupings since these were the distinctions being tested.

6. HINDRANCES

Throughout the survey, technicians felt the pressure of trying to obtain accurate, quality-controlled material. Several factors contributed to this frustration of trying to achieve the 'ideal' in the reality of field conditions.

Sometimes mechanical malfunctions prevented technicians from getting material of the quality they desired. The quality of tape-recorders and recordings varied and the repair of equipment was a problem. Travel conditions encountered on survey were not conducive to maintaining delicate equipment. There were times when a technician accepted a recording he knew to be of inferior quality simply because it was the best he could do at that time.

Physical conditions for testing were often far from ideal. Most testing was done in a place where people were standing around watching and listening. The subjects sometimes felt embarrassed by the audience and very often were distracted by the noise and confusion.

Some people declined to take the test at all despite reassurance by the technician that there were no 'wrong' answers and that he merely wanted to see how much they could understand. Still the desire to do well on what was obviously a 'test' of some sort often resulted in the selection of only those who were recognised as being clever, self-confident or well-travelled. This undoubtedly had a tendency to elevate scores.

On some occasions it was difficult for the survey technicians and the people in the villages to coordinate time schedules in order to complete the necessary work. Also there were frustrations on the part of technicians and village people when the processes of telling, transcribing and translating stories and taking the tests became more time consuming than was expected. Sometimes those helping with the work would even have to leave before the work was completed.

7. VALIDITY

In the face of all the problems and variables mentioned, the reader might question the validity of the results of such a survey.

Two factors should be kept in mind in assessing the results of the survey. The first is the extremely large corpus which forms the data base from which conclusions are drawn. Smith points out that even after excluding from comparison 40 words of the 367-word list because of problems and ambiguities, the remaining 327 words yield almost 60,000 comparisons. With a corpus of that size, errors can be introduced without appreciably affecting the overall relationships.

Similarly, if at every test point there are ten persons giving ten answers to seven stories, this represents a total of some 700 answers for one test point. Given two test points per language and as many as 40 in the large Kadazan/Dusun language group, the number of comparisons becomes somewhat staggering. The chance of the outcome being greatly affected by the factors mentioned above seems less likely.

The second factor which should be kept in mind is that in the final analysis much of what is presented reflects the subjective evaluation not only of the linguistic technicians, but of the real authorities on the languages of Sabah - the people of Sabah. In travelling from place to place and establishing friendships with individuals from many areas and language communities, members of SIL have had opportunity to observe and to question. Those observations and the answers to the questions do not plot neatly on a chart, but they are an important part of what has been learned.

NOTES

1. Blom (1979); Hurlbut and Pekkanen (to appear); Pekkanen (1981); Smith (in this volume).
2. The list used was adapted from the wordlist given in Reid 1971 to facilitate comparison with lists collected throughout the Philippines.

3. Normally ten questions were used and those questions were translated into the language where the testing was being done. This procedure may, however, have been changed slightly on occasion at the discretion of the technicians.
4. No testing was done in Javanese or Bugis languages despite the statement by Smith (in this volume) that such testing was needed since it was not possible to do this in the context of the areas from which they came.
5. In several instances individual authors have found it necessary to alter these thresholds in order to deal with complicated situations where factors influencing scores may have been different from the norm. Such instances will be explained in the specific reports.

THE LUNDAYEH LANGUAGE

David C. Moody

0. INTRODUCTION

The Lundayeh are relatively recent arrivals in Sabah. They began settlement in the Ulu Padas area of Sipitang District (particularly at Long Pa Sia and Kuala Miau) about 100 years ago. Since then and mostly within the last 60 years they have also settled along the Menalong River. Smaller Lundayeh populations are located in Tenom District at Baru Jumpa, Sapung Estate, Sugiang Estate, and Kuala Tomani, and in Keningau District at Kuala Puntii.

Current Sabah census figures do not indicate the number of Lundayeh living within the state. Crain reckoned that there were 2,165 Lundayeh living in Sabah according to his own personal census in 1968 (Crain 1978:139n2). Based on his figure, and allowing for continued immigration and natural population growth, it is reasonable to estimate the present Lundayeh population in Sabah to be 2,500-3,000.

The Lundayeh of this present study are grouped as part of a larger linguistic and cultural nexus whose heartland has been defined as the Kelabit-Kerayan highland of north central Borneo, particularly the area loosely demarcated by the coordinates 4°15'-3°20'N and 115°20'-116°0'E (LeBar 1972:159). Harrisson suggests this larger group migrated into this area from the south-east, becoming its original settlers sometime in the first centuries of the Christian era. In the 17th century their migrations toward the south and west were halted by the northward advances of Kayan and Kenyah populations. By the early 19th century they had firmly established themselves in the Trusan and Lawas Damit valleys. Migration from interior to coastal regions, including those areas in Sabah where they are now located, took place in the periods following the Second World War, the Indonesian Revolution and the Indonesian-Malaysian Confrontation (Crain 1978:124-126; LeBar 1972:159). The collection and test points for this present study are shown in Figure 1.

Several investigators have been concerned with linguistic aspects of Lundayeh study. Prentice (1970, 1971) and others argue that the distinction between the Murutic languages of Sabah and Lundayeh - often labelled as Murut - should be more carefully maintained. Lees (1959) presents a phonological description of Lundayeh, from phoneme to word levels. Clayre (1972) presents a comparative phonology of Lundayeh and Sa'ban, a language of Sarawak. In an earlier work Clayre (1970) compares several languages of Sabah and Sarawak, including Lundayeh, as to how they mark participant focus. Garman, Griffiths

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 59-65. *Pacific Linguistics*, C-78, 1984.

© David C. Moody

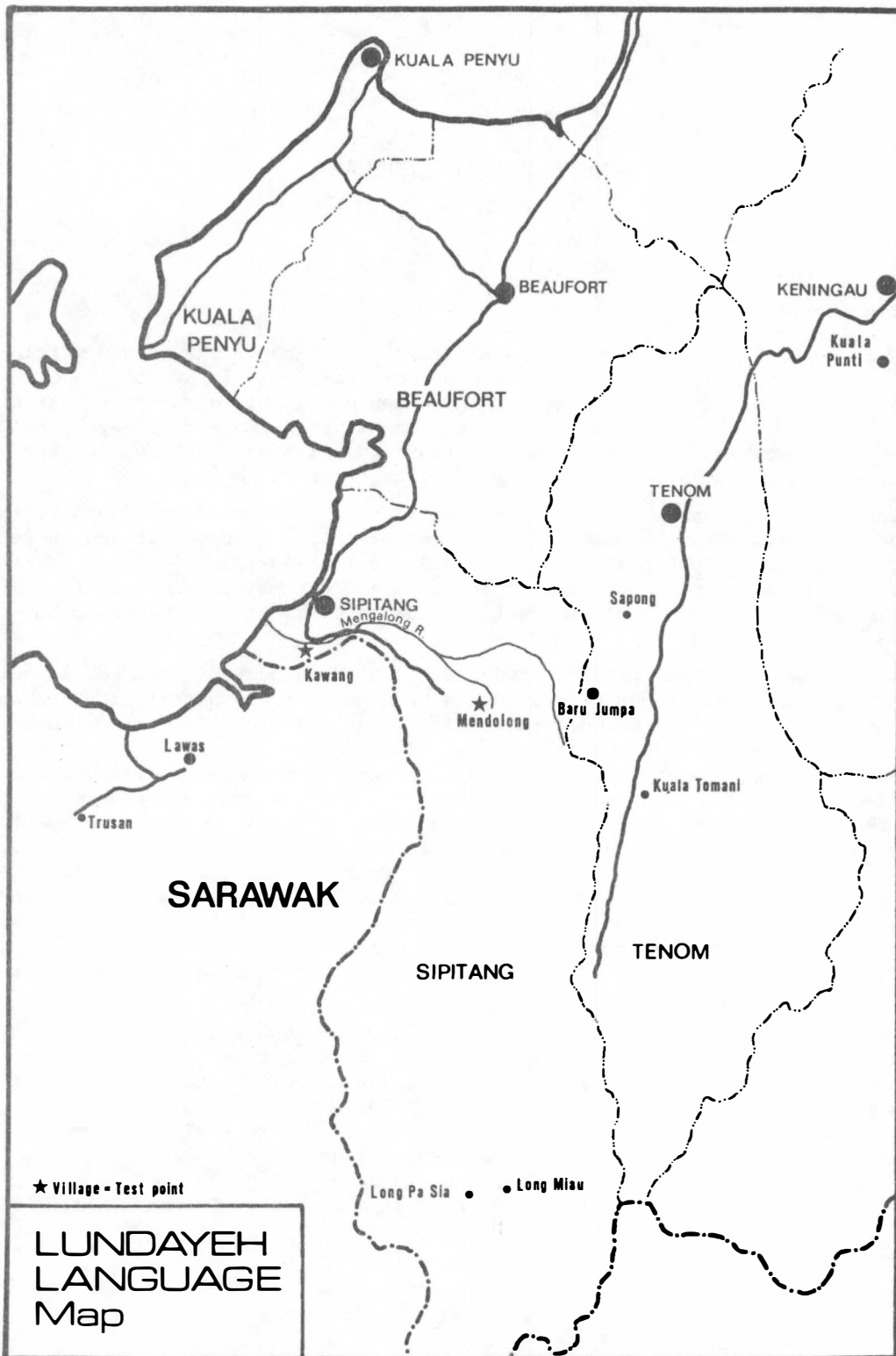


Figure 1: Lundayeh language map

and Wales (1970) present results from their study of language acquisition among Lundayeh children. A phrase book (Padan 1971) and dictionary (Pur 1961) have also been published.

Study of the Lundayeh has been somewhat confusing because of the many terms which have been employed to designate them and the language they speak. In addition to Lundayeh, the list includes Lun Lod, Lun Bawang, Kemaloh, Kelabitic, Kemaloh-Kelabit, Kelabitic Murut, Sarawak Murut, Southern Murut, Murut, Tagal, Potok and Dayak. Most of these terms are applied either in Sarawak or East Kalimantan. The people of Kemaloh East Kalimantan, refer to themselves as Lundayeh and several researchers indicate their dialect is the most widespread Lundayeh dialect (LeBar 1972:159; Crain 1978:139n2). Kelabit refers to a closely related dialect spoken in Sarawak's Fourth Division. This writer is not certain exactly how the terms Potok and Dayak have application to the Lundayeh; they, too, are used in East Kalimantan. (See Deegan 1970:264; Crain 1979:139n7; LeBar 1972:159.)

The terms which have some use as self-designations among Lundayeh in Sabah are Lun Lod, Lun Bawang and Murut. The use of the term 'Murut' whether by Lundayeh themselves or others is particularly frustrating. 'Murut' in a proper linguistic sense refers to a number of languages belonging to a single language family and spoken primarily in Sabah (→ Murutic). The Murutic family is only distantly related to Lundayeh by common inclusion in the North-western Austronesian superstock (→ Section 1). An Assistant District Officer in Sipitang District explained that many Lundayeh had referred to themselves as 'Murut' in the 1980 census-taking so that they would have a place with a larger minority grouping and would not lose their identity altogether. In reality, however, the Lundayeh do not think of themselves as Murut¹ (see also Prentice 1972:154 and Crain 1978: 123-124, 139n3).

The remaining terms Lundayeh, Lun Lod and Lun Bawang are used as genuine self-referents by Lundayeh in Sabah. In the Lundayeh language *lun* means *people*, *dayeh* means *upriver*, *lod* means *downriver*, and *bawang* means *region*, *area* or *locality*. Thus *lun dayeh* means *the upriver people*, *lun lod* means *the downriver people*, and *lun bawang* means *the people of an/this area* or *the local people*.

Both *lun dayeh* and *lun lod* have a non-technical usage in referring to relative placement of people along a river or stream, and certainly their specialised usage has been introduced only as broadening relations within larger social frameworks have necessitated an established identity (Deegan 1979:72n5). The term Lundayeh has preference as a self-referent in that it most aptly describes their historical background as riverine dwellers in interior areas. The term Lun Bawang has limited use outside Sarawak. In some Sipitang communities it bears ecclesio-political connotations, and there as well as in upland, interior areas indicates longtime residents at a place as opposed to Lundayeh (Crain 1978:139n7).

The Sabah Lundayeh Association (Persatuan Lundayeh Sabah)² was established early in 1979 for the purpose of preserving Lundayeh culture and ensuring that information about it is propagated accurately. The Association has officially taken the one-word spelling of Lundayeh as the preferred self-designation of Lundayeh people in Sabah. They feel this term is most descriptive of Lundayeh origins. The one-word spelling is intended to reflect an ethnic identity as opposed to the relational connotations suggested from two-word spellings (cf. Lun Dayeh, Lun Daya, Lun Dayah, Lun Daye, Lun Dayoh; also Lundaya - Dunn 1980).

1. LEXICOSTATISTICAL CLASSIFICATION

According to Smith's tentative lexicostatistical classification (in this volume), Lundayeh represents one of nine linguistic stocks, within the North-western Austronesian superstock, which are found in Sabah. Within this group of nine, Lundayeh has the lowest relations based on percentage of shared cognates (PSC) with any of the other linguistic stocks. Comparing a representative Lundayeh wordlist with representatives of the other eight North-western Austronesian stocks, Smith demonstrated its range of relations with them to be 25-29 PSC. This indicates that Lundayeh is the most distantly related language within the superstock, and by way of corollary, also the most unique linguistically.

In a more direct comparison of 35 Murutic wordlists with four Lundayeh wordlists, the range of relations among them is 28-35 PSC, with an average of 30.8 PSC for the 140 comparisons. If 11 wordlists representative of Murutic languages are compared with a single representative Lundayeh wordlist, the 11 comparisons yield a slightly lower average of 30.3 PSC and a range of 28-33 PSC.³ These comparisons clearly show the distinctiveness of Lundayeh from any form of Sabah Murut.

Wordlists were collected from four Lundayeh villages in Sabah. The villages are Kuala Puntı KU, Baru Jumpa TM, Mendolong SG, and Kawang SG. The Lundayeh in all of these villages except Kawang SG reportedly had come to Sabah within the last 30 years. Kawang SG is more than 50 years old. The person from whom the wordlist was elicited in Kawang SG called his language Lun Lod. Language assistants in the other three villages all spoke of their language as Lundayeh. The PSC relations between the four villages are displayed in Figure 2.



Figure 2: PSC relations between Lundayeh villages

The range of PSC relations among the four wordlists is high, 87-89 PSC, indicating the four represent a single language, and even a single dialect. According to project criteria (→ Introduction) such high internal relations and low external relations as those discussed above would not necessitate intelligibility testing in Lundayeh villages. Nevertheless, it was decided to do intelligibility testing to determine the extent of language learning by Lundayeh speakers since they are living in areas where Murut is the predominant language, and also to test the degree of homogeneity among Lundayeh speakers, especially where different autonyms had been used.

2. TESTING PROCEDURE

Testing was done in two Lundayeh villages, Kawang and Mendolong, both in Sipitang District. Kawang SG represented an earlier wave of immigration and the use of the self-referent Lun Lod. Mendolong SG by contrast represented both a more recent wave of immigration and used the autonym Lundayeh.

The test set included Lundayeh stories from each hometown, the other Lundayeh test point, Baru Jumpa TM, and Lubiduan SAR,⁴ a village located near Trusan in the Lawas district of Sarawak's Fifth Division. The Mendolong SG story was used with only four test subjects in Kawang SG, and likewise, the Kawang SG story was used with only four subjects in Mendolong SG, since the lexical relation between them was high, they are geographically close, and initial test scores were likewise high. The story from Ansip in Keningau District was chosen to represent Murut, as it is also Tagal, which is geographically the nearest Murutic neighbour to Lundayeh in Sabah. Each of the tapes was judged to be of clear quality. The content of each was good. Kawang SG seemed to be an easier story to understand.

Testing went well in both Kawang SG and Mendolong SG with the exception that in Kawang SG the youngest subject was 36 years old and only nine subjects were tested. These factors did not seem to significantly influence test scores.

3. TEST RESULTS

The results of dialect intelligibility testing in Kawang SG and Mendolong SG are shown in Figure 3. It is noteworthy that test subjects in Kawang SG used the self-designations Lun Lod and Lun Bawang as well as Lundayeh. The Lubiduan SAR story-teller also referred to himself as Lun Bawang. All subjects in Mendolong SG used the self-referent Lundayeh. The range for the average score received for each Lundayeh reference tape in each of the two test points was 89-100%. Neither the difference in self-designations nor comparative ages of the settlements proved to be significant. The greatest difference between the average scores received on the same reference tape for the two villages was 4%. The scores confirm that the Lundayeh of Sabah are linguistically homogeneous and speak a single dialect even though different autonyms are in use. Also, the scores suggest that this homogeneity extends across the border into Sarawak. Further investigation should determine the full extent of this linguistic conformity. Of particular interest is the relationship of Sabah Lundayeh to those dialects still spoken within its linguistic heartland in the Kelabit-Kerayan highland of Sarawak and East Kalimantan.

The testing of Lundayeh speakers' understanding of a Murut story indicates that language learning has been minimal. Even though dialect intelligibility testing alone is not an accurate tool for measuring relations between languages outside of the same subfamily, these low scores nevertheless support Smith's conclusion that Lundayeh and the Murutic languages are only distantly related.

REFERENCE TAPE TEST POINT	Lundayeh				Murut
	Kawang SG	Mendolong SG	Baru Jumpa TM	Lubiđuan SAR	Ansip KU
Kawang SG	99 (100)	100* (89)	91 (88)	89	38 (30)
Mendolong SG	95* (89)	98 (100)	95 (87)	91	33 (29)

* average score for four subjects only

Figure 3: Results of intelligibility testing in the Lundayeh villages of Kawang SG and Mendolong SG. (The reference tapes used in the test set are listed along the top. PSC relations are given in parentheses.)

4. NATIONAL LANGUAGE INTELLIGIBILITY

In both Kawang SG and Mendolong SG subjects were tested for their ability to understand a story told in Bahasa Malaysia. Figure 4 summarises the test scores and some sociological information about the test subjects.

	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Kawang SG	82	5/4	45	36-60	3	2.7	0.9
Mendolong SG	80	5/5	34	15-65	8	7.8	6.2

Figure 4: Comprehension of the national language in two Lundayeh villages with sociological data. (AVG. SCORE is given as percentage. RANGE indicates youngest and oldest test subject. EDUCATION figures indicate the number of subjects who had received some formal education (ED.), the average number of years of education those subjects had received (A.E.R.), and the average number of years of education of all subjects (A.E.S.) in the corpus.)

Subjects in both villages understood the national language story well. Though the test corpus at Kawang SG had a higher average age and, per subject, had received significantly less formal education, their average intelligibility score was slightly higher than that for Mendolong SG. This is certainly attributable in part to the proximity of Kawang SG to Malayic-speaking communities with which there is some contact, and its location just off the main road to Sipitang town.

5. CONCLUSION

Dialect intelligibility testing in two Lundayeh villages confirms the conclusions made by Smith in his tentative lexicostatistical analysis of Sabah languages. Lundayeh as it is spoken in various communities in Sabah represents a single dialect. Test results further indicate that a similar degree of homogeneity extends as well to Lundayeh communities in Sarawak. Testing also confirms the distant relationship propounded for Lundayeh and the Murutic languages of Sabah. Application of the term Murut to the Lundayeh people only blurs a clear-cut linguistic distinction.

NOTES

1. The writer wishes to acknowledge the valuable assistance received from Dawar bin Sodom, the Assistant District Officer for Rural Affairs in Sipitang District, and from Charles Ayub Tabad, President of the Sabah Lundayeh Association (Persatuan Lundayeh Sabah).
2. Persatuan Lundayeh Sabah is registered as a 'friendly' society in Sabah with the federal registrar of societies (Pendaftar Pertubuhan Malaysia) in accordance with the Society Act of 1966.
3. The 11 Murut villages are: Minansut KU, representing the Gana language; Bukau BT, representing the Beaufort Murut language; Langsat TM, representing the Timugon language; Pensiangan PN, representing the Tagal language; Kadalakan KU, representing the Nabay language; Sook KU, representing the Paluan language; Baru Jumpa TM, representing the Kolod language; Labuk KAL, representing Sembakung Murut; Kalabakan TU, representing Kalabakan Murut; Serudung TU, representing Serudung Murut; and Kokoroton KN, representing the Baukan language (→ Murutic).
4. The Lubiduan SAR story was tape-recorded by Robin Labo, now a teacher in Miri, Sarawak. He also provided a Lundayeh transcription and English translation for the story.

THE ILLANUN LANGUAGE

John E. Banker

0. INTRODUCTION

The Illanun presently living in Sabah originally came from Mindanao in the Philippines. Forrest (1969) states that before 1667 there was much suffering from volcanic eruption in the Illanun districts in the Philippines so that many fled to Sulu in the southern Philippines and also to Tampassook (Tempasuk) and Tawarran (Tuaran) in Sabah. Forrest also notes that the name Illanun¹ is derived from their homeland around Laka Lanao and the shores of Illana Bay.

Wright (1979-80) states that the Illanun were established in Sabah in "the latter part of the eighteenth century at Tempasuk and Pandasan on the northwest coast, at Marudu Bay, and in the Tungku River on the southern part of the Unsang Peninsula."² These areas which began as "pirate harbours" have evolved into the present-day Illanun settlements in Sabah (Figure 1).

For this present study, data from Illanun settlements in Kota Belud District and Lahad Datu District are considered. Appell (1970) discusses two other Illanun communities in Kudat District, but those villages, Marimbau Laut KT and Indarasan Laut KT, are not under discussion here. In fact, a total of 17 villages of Illanun are known to exist in Sabah at present. The total population of Illanun in Sabah at this time is estimated to be 5,000.³

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classifies the Illanun spoken in Sabah on the basis of four wordlists from Kota Belud and Lahad Datu districts as being comprised of two distinct dialects, one spoken in each district (Figure 2). The range in the percentage of shared cognate (PSC) relations between the two is 77-81. The lower figures for Kulambai KB may be due to the fact that Kulambai KB is a mixed village, with both Illanun and Bajau people living there.

A Maranao wordlist from the Philippines was also compared with the Illanun wordlists in Sabah, and the range was from 65-68 PSC. However Fleischman (1981), doing research on the Magindanao language in Mindanao PHL, compared the Lahad Datu and Kota Belud wordlists with another Maranao wordlist and found 79 PSC between the Illanun of Lahad Datu and Maranao, and 80 PSC between the Illanun of Kota Belud and Maranao. An even closer relationship of 82 PSC was found between Iranun of the Philippines and the Illanun of Lahad Datu. However, the PSC relation of Kota Belud Illanun and Iranun was only 78 PSC.⁴

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 67-74. *Pacific Linguistics*, C-78, 1984.

© John E. Banker

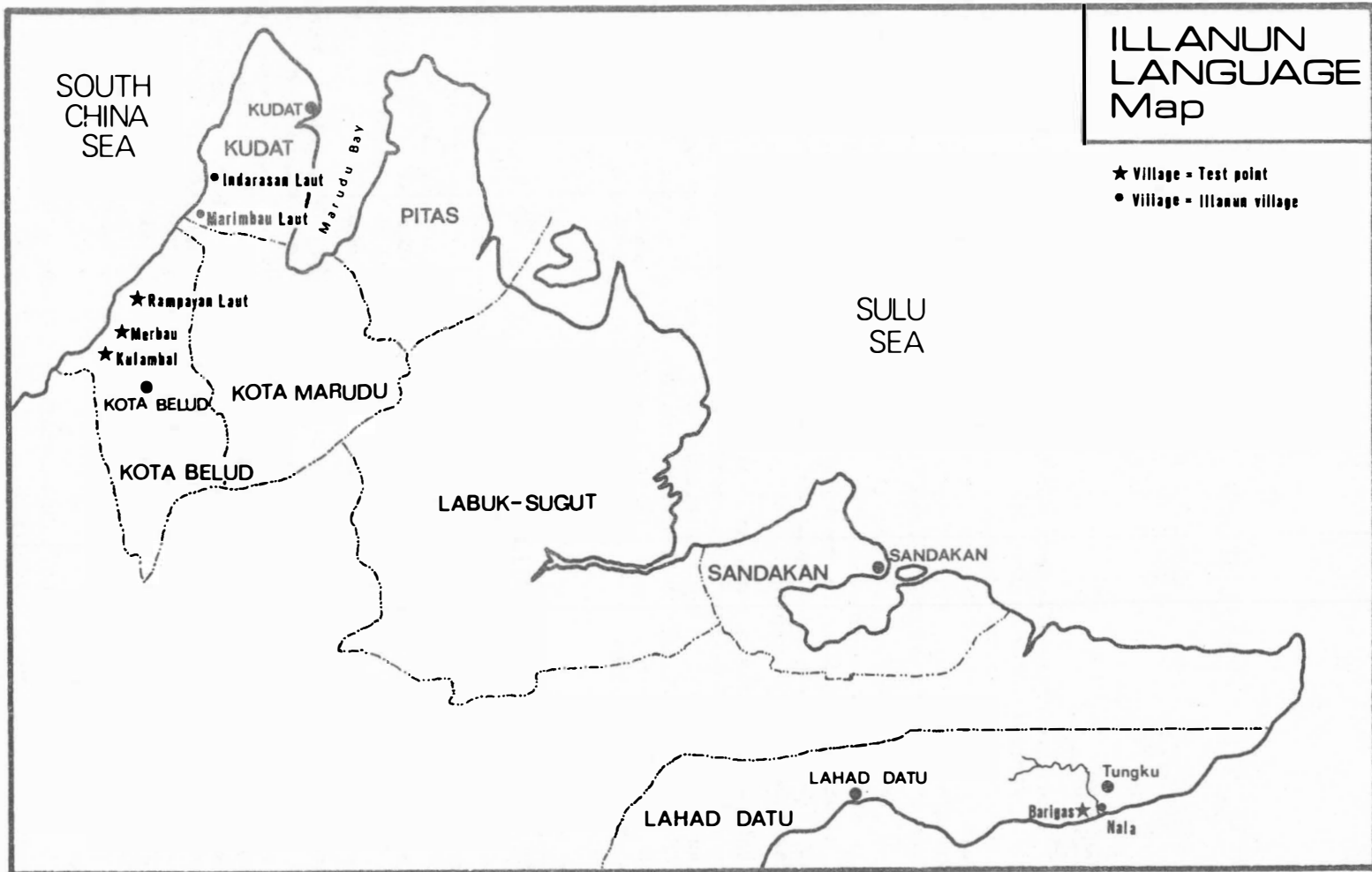


Figure 1: Iillanun language map

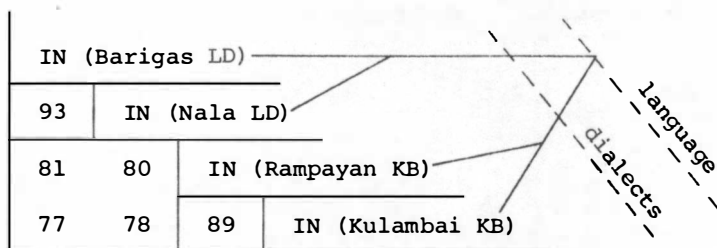


Figure 2: PSC of four samples, two dialects of Illanun (Smith, in this volume.) (IN = Illanun.)

2. TESTING PROCEDURE

At the beginning stages of the intelligibility testing, only Sabah Illanun testing was taken into consideration. It was only considered necessary to test the Illanun of Kota Belud with the Illanun of Lahad Datu since the original PSC figures indicated that they belonged to two distinct dialects. The village of Barigas LD was tested for its understanding of the tape-recorded story from Rampayan Laut KB.

Since, however, the Illanun of Sabah came originally from the Philippines, an objective of this study was also to try to establish the relationship of the Illanun language of Sabah today with the Danao languages - Iranun, Maranao, and Magindanao - to which it relates historically.

At the time intelligibility testing was done in Rampayan Laut KB, an Iranun tape from the Philippines was available, and both this tape and the Barigas LD tape were tested at Rampayan Laut KB.

After this testing was completed, two more tapes of the Maranao and Magindanao languages were acquired from the Philippines. These two tapes were only used to test one Illanun speaker from the Kota Belud area.

All of the tapes used in Illanun testing were considered clear with regard to content and technical quality. It should be noted however that the two Sabah Illanun tapes are shorter than the ones from the Philippines and are therefore easier to understand.

3. TEST RESULTS

Figure 3 displays the results of intelligibility testing of the Illanun language. In the testing done in Barigas LD, the test subjects scored 97% in average understanding of the Rampayan Laut KB tape. In the cross-testing, Rampayan Laut KB scored 93% on the Barigas LD story. Thus both of these tests indicate a high degree of intelligibility between these two Illanun areas of Sabah despite the fact that there seems to be little contact between them. The shortness and simplicity of these two stories however may have artificially raised the scores.

Rampayan Laut KB scored 75% on the Iranun tape from the Philippines, which would indicate that the two represent distinct languages with limited intelligibility.

REFERENCE TAPES TEST POINTS	Rampayan Laut KB (Illanun)	Barigas LD (Illanun)	Maranao PHL (Maranao)	Iranun PHL (Iranun)	Magindanao PHL (Magindanao)
Rampayan Laut KB (Illanun)		93 (81)		75 (78)	
Merbau KB (Illanun)	*100	*100 (81)	*85 (80)	*77 (78)	*58 (71)
Barigas LD (Illanun)	97 (81)				

* scores received for one test subject only

Figure 3: Intelligibility testing results from three Illanun villages in Sabah. (Scores are given as percentages. PSC figures are in parentheses.⁵)

The only other testing which was done in Sabah was that of a full test set of Lahad Datu Illanun and the Philippine languages - Iranun, Maranao and Magindanao - given to only one subject from Merbau KB. The subject stated that he had never had contact with the Lahad Datu Illanun or any person from the Philippine language groups in the test set. His scores on Barigas LD and Philippine Iranun were quite close to the average village score on the ten people tested at Rampayan Laut KB. The one man from Merbau KB scored 100% on the Barigas LD tape and 77% on the Iranun tape (Figure 3).⁶ In addition he also scored 85% on the Maranao tape and only 58% on the Magindanao tape.

Although only one person was tested on the Maranao and Magindanao tapes, it would be safe to assume that his intelligibility of the Danao languages in the Philippines would be close to the average Kota Belud Illanun's intelligibility of those languages. On the basis of that, it appears that Sabah Illanun is a distinct language from Magindanao.

The average score of 75% of 11 Kota Belud Illanun on the Iranun tape from the Philippines indicates that Kota Belud Illanun is a distinct language from Philippine Iranun, even though in cross-testing, the Iranun scored 97% on the Kota Belud Illanun story. The higher score could likely be attributed to the fact that the Kota Belud Illanun story was shorter and simpler than the Iranun story from the Philippines. The PSC relation of 78 between the two would also back up the hypothesis that they are distinct languages.⁷

The 85% intelligibility between Maranao of the Philippines and Kota Belud Illanun would suggest that Kota Belud Illanun and Maranao are dialects of the same language. This remains open to question however, since only one person in Sabah was tested on the one Maranao story. Fleischman's (1981) PSC figure for Kota Belud Illanun and Maranao is 80 PSC, which is just within the same language threshold (→ Introduction).

But if Fleischman's hypothesis is considered accurate, that Maranao and Iranun are distinct languages, and Iranun is closer to Maranao both in PSC and intelligibility scores than Kota Belud Illanun is to Maranao, it would follow that Kota Belud Illanun should be considered a distinct language from both Maranao and Iranun, with a closer relationship to Maranao than to Iranun.

The Kota Belud Illanun recognise a relationship with the Maranao. An Illanun leader in the Kota Belud area stated that the Illanun originally came from "Ranao" (Lanao, Mindanao, a Maranao area). Another man now living in a Kota Belud Illanun village, who had come from the Philippines a few years ago, was interviewed and said that he was a Maranao.

The PSC figures suggest that Lahad Datu Illunan is more closely related to Iranun than Kota Belud Illanun is, and not quite as close to Maranao as Kota Belud Illanun is. Lahad Datu Illanun is 3 PSC closer to Iranun than it is to Maranao (→ Note 4).

It may be that the Lahad Datu Illanun originated from the Iranun but with long years in Sabah have borrowed from Malay, as has Kota Belud Illanun, and so now Lahad Datu Illanun is lexically closer to Kota Belud Illanun than it is to Iranun. As has been mentioned before, at the present time Lahad Datu and Kota Belud Illanun seem to have very little contact with each other. In fact, some of the persons interviewed did not even know of the other Illanun community.

It may be that both of these Sabah Illanun communities originated from a mixture of Iranun and Maranao as is suggested by Fleischman (1981).

Based on the PSC figures and intelligibility testing results, Kota Belud Illanun and Lahad Datu Illanun should be considered the same language. However, further testing should be carried on between the two to determine whether they are distinct dialects or are one dialect only. The fact that both stories were understood so well by the Iranun of the Philippines may indicate that the stories were too easy to provide a good test.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Intelligibility scores on the Bahasa Malaysia story were quite high. Rampayan Laut KB registered an average score of 91% with individual scores ranging from 55-100%, and Barigas LD registered an average score of 84% with individual scores ranging from 35-100% (Figure 4).

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Rampayan Laut KB	91	6/4	34	16-60	6	9	5
Barigas LD	84	8/2	36	18-70	5	7	3

Figure 4: National language intelligibility testing results for two Illanun villages with sociological data. (Test scores are given as percentages. Under EDUCATION, ED. is the number of subjects in each corpus who had received some formal education, A.E.R. is the average number of years of education those subjects had received and A.E.S. is the average number of years of education per subject for the corpus as a whole.)

For those who attended school, the Rampayan Laut KB average score was 99%, and the Barigas LD average score was 96%. For those who did not attend school, the Rampayan Laut KB average score was 79%, the Barigas LD score was 72%.

The slightly higher scores in Rampayan Laut KB probably reflect the fact that Rampayan Laut KB is closer to the economic and educational centre of Kota Belud than Barigas LD is to the economic and educational centre of Lahad Datu.

5. CONCLUSION

The Illanun language of Sabah is a distinct language in the Danao family. Based on the data included in this paper, there appear to be two dialects of Illanun. One is spoken in the Kota Belud area, the other spoken in the Tungku area of Lahad Datu. Kota Belud Illanun is closer to Maranao than it is to Iranun. PSC figures suggest that Lahad Datu Illanun is more closely related to Iranun than it is to Maranao but further intelligibility testing needs to be done to prove this.

Most of the Illanun tested were proficient in Bahasa Malaysia, but they still speak Illanun when communicating with one another.

NOTES

1. Also referred to as Ilanun, Illanoan, Illanoon, Illanos, Iranon Maranao, Iranum, Iranun, Lanoon, Lanun, Magindanao/Iranun, Ylanos, according to Dunn (1980).
2. Sopher 1965:138; Warren 1975:259.
3. This is only a rough estimate based on the following: (a) In the 1960 Census of Sabah, a total of 3,731 Illanun were recorded (Jones 1962). Later census reports have not distinguished the Illanun from larger ethnic groups. (b) The total number of known Illanun villages is 17, including two in Kudat for which Appell (1970) gives information on the population. The average population of the six villages for which figures are known is 296. Seventeen villages this size would be 5,032.
4. The following chart shows the PSC relation between Lahad Datu and Kota Belud as 88. This figure is 7-11 percent higher than the PSC determined in the research done in Sabah using the same wordlists. At the time of writing, the actual procedure by which Fleischman obtained this different PSC is not known. In reworking the comparison of Rampayan Laut KB with Barigas LD on the basis of these two wordlists only, this writer calculated a figure of 83.5 PSC, 15 of the comparisons on the 367-item wordlist being eliminated for various reasons. To end up with 88 PSC many more would have to be eliminated or more cognates recognised.

IN (Kota Belud)

88	IN (Lahad Datu)								
80	79	MA (Maranao PHL)							
80	80	92	IR (Isebanganen PHL)						
78	82	85	92	IR (Iranun PHL)					
71	76	76	82	92	MN (Ilud PHL)				
67	69	74	81	89	97	MN (Laya PHL)			
72	73	75	80	87	94	90	MN (Biwangan PHL)		
73	73	76	84	91	97	93	95	MN (Sibuguey PHL)	

PSC relations between various Danao languages/dialects
(from Fleischman 1981)

IN = Illanun, MA = Maranao, IR = Iranun, MN = Magindanao

5. The PSC figures given for the villages of Merbau KB are based on a comparison between the village of Rampayan Laut KB which is geographically and linguistically closest to Merbau KB and figures given in Note 4, which has been adapted from Fleischman (1981) using the Iranun example from Iranun PHL and the Magindanao example from Ilud PHL. In the case of Barigas LD however, the SIL figure of 81 was used.
6. In testing the Iranun story at Rampayan Laut one of the 11 questions had to be deleted because it did not seem to fit the story. In testing the Iranun story with the one subject later on, another question was used to substitute for the deleted one, and thus, the one subject's percentage is based on one more question than the average percentage scored at Rampayan Laut. If only the ten questions used at Rampayan Laut are considered, then the one subject received a percentage of 75, only one half percent different from the average score of 74.5% of the ten Rampayan Laut subjects.
7. Results of intelligibility testing conducted in the Philippines among the Danao languages (Fleischman 1981). (See Page 74.)

REFERENCE TAPES	Rampayan Laut KB (Ilianan)	Barigas ID (Ilianan)	Maranao PHL (Maranao)	Isebanganen PHL (Iranun)	Iranun PHL (Iranun)	Ilud PHL (Magindanao)	Laya PHL (Magindanao)	Biwangan PHL (Magindanao)
TEST POINTS								
MA Maranao PHL (Maranao)				94	87	52		
IR Iranun PHL (Iranun)	97	96	95			98	98	
MN Ilud PHL (Magindanao)			60		84		96	
MN Laya PHL (Magindanao)			43		92	97		99

Results of intelligibility testing conducted in the Philippines among the Danao languages (Fleischman 1981).

THE SULUK (TAUSUG) LANGUAGE

David C. Moody

0. INTRODUCTION

The origins of the Suluk (Tausug) people can be traced to the north-east coast of Mindanao in the Philippines, the result of an outward expansion from the Bisayan Islands about 1500 years ago. About 1200 A.D. there was a movement of Suluk-speaking people into the Sulu Archipelago via contact with Sama-Bajau traders. By the early 16th century, the Suluk people had established themselves and their language in Sulu, and increased their influence during the Malay-patterned sultanate of the 16th and 17th centuries. Suluk dominance in the archipelago has continued since that time (Pallesen 1977:363-375).¹

The Suluk people of Sabah inhabit communities located in the east coast districts of Labuk-Sugut, Lahad Datu, Sandakan, Semporna and Tawau (Figure 1) as well as in several west coast communities. Though many of these people have been at home in Sabah their whole lives, as have generations of their ancestors, most have immigrated more recently. The 1960 census listed more than 10,000 locally-born "Sulu" people living in Sabah, a figure which remained stable through the 1970 census-taking.² By contrast there are now over 100,000 Suluk people living in Sabah's south-east districts alone. A district official indicated that within the last ten years the Suluk population of Tawau District had grown to about 40,000 while that of Semporna District had increased to 70,000.³

The term Suluk has come into official use in Sabah to designate these immigrants from the Sulu Archipelago. The people speaking the same language in the Sulu Archipelago refer to themselves as "Tausug", which means *men of the current*.⁴ Although Tausug remains the preferred self-referent, they are adapting to the use of the term Suluk. The practice in this paper will be to use the term Suluk generally, qualifying it parenthetically with the term Tausug when the referent's locus of habitation is in the Philippines, i.e. Suluk (Tausug). The term Suluk is sometimes spelled Sooloo, or Sulu. The term Tausug may also be seen as Tau Sug or Taosug (Dunn 1980).

The Suluk language of Sabah has received limited attention in the linguistic literature. A wordlist collected by Anson Cowie was published in 1880.⁵ More recently, Asmah Haji Omar and M.B. Hardaker have individually authored sketches of Suluk phonology and grammar, the latter also including a compilation of phrases. In a separate article, Hardaker presented an introductory Suluk vocabulary.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 75-84. *Pacific Linguistics*, C-78, 1984.

© David C. Moody

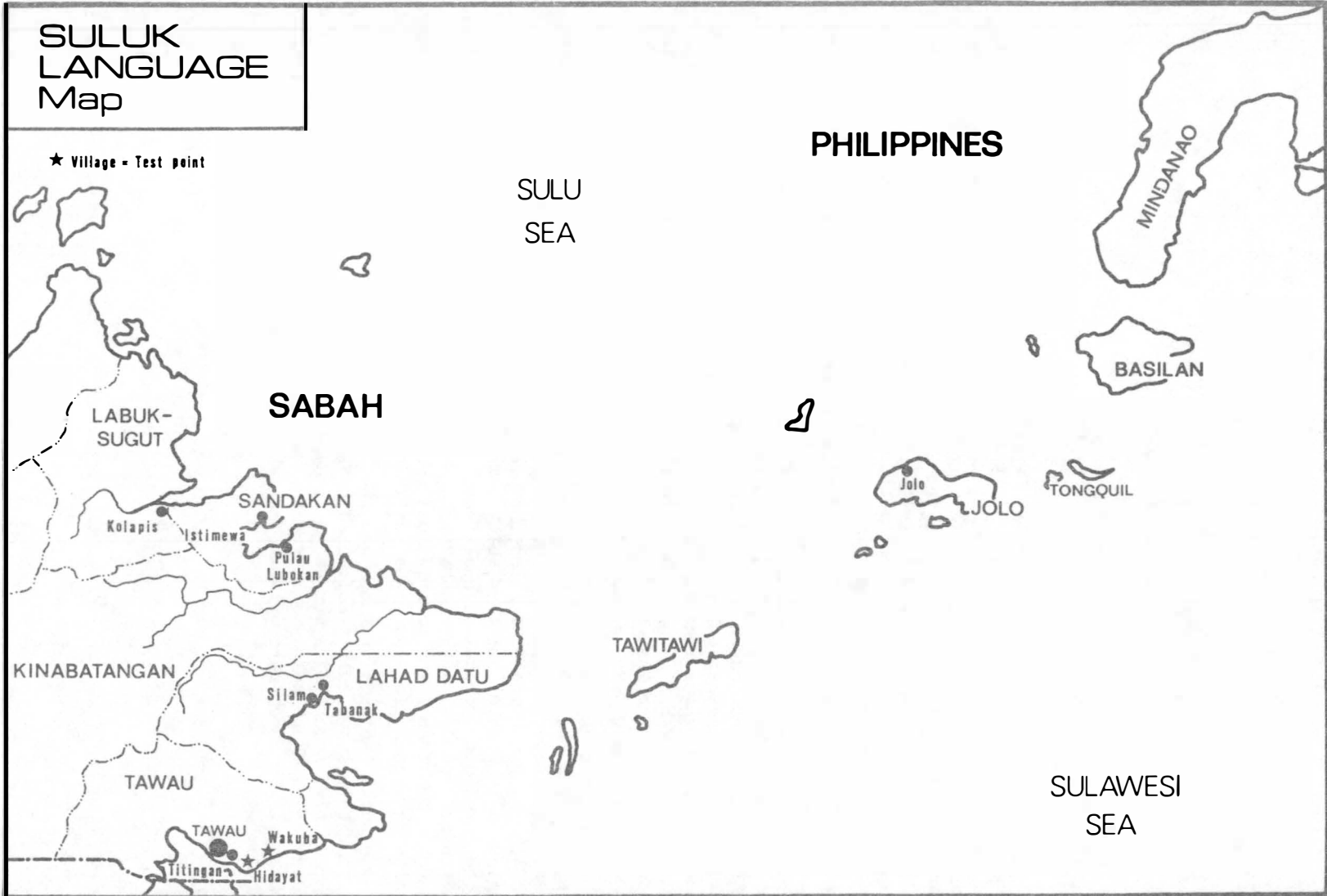


Figure 1: Suluk language map

More research has been done on the Suluk (Tausug) language in the Philippines. Seymour and Lois Ashley have produced articles pertaining to phonology, orthography, verbal cases, and sentence types. Along with Irene Hassan and Nurhadan Halud they have also compiled a Tausug-English dictionary of about 3,500 entries. A four-language phrase book which includes Suluk (Tausug) has also been published for visitors to Zamboanga City, Sulu Province. Literacy related materials including primers, readers and teaching guides by various authors have been published by the Summer Institute of Linguistics of the Philippines. A. Kemp Pallesen's doctoral dissertation is a very satisfying historico-comparative study of Suluk (Tausug) and Sama-Bajau languages in which he establishes evidence for the linguistic convergence between them and draws conclusions pertaining to the history and nature of the contact between the two cultures.

1. LEXICOSTATISTICAL CLASSIFICATION

Suluk (Tausug) is classified as a member of the East Mindanao subgroup of the Central Philippine languages, a descendant of a Southern Bisayan language and most closely related to the Butuanun language from which it separated some 900 years ago (Pallesen 1977:23,333f). Within Sabah Smith (in this volume) has classified Suluk as a language singularly representing one of the nine stock subgroupings of the North-western Austronesian superstock.

The Suluk data collected for lexicostatistical comparison in the present study were gathered from seven communities in Sabah with an eighth wordlist (Jolo) being supplied from the Philippine Branch of the Summer Institute of Linguistics (Figure 1). In Smith's classification, the eight lists represent a single language.

The percentage of shared cognate relations (PSC) for the eight lists are displayed in Figure 2. The range for all eight lists is 77-92 PSC, with each sample relating at 87 PSC with at least one other sample of the group. The four lists from Tabanak LD, Pulau Lubokan SN, Titingan TU, and Silam LD relate to each other in the range 86-92 PSC and thus represent a single core dialect, 85 PSC being Smith's single dialect threshold. The relations between the eight speech samples may be pictured as a dialect chain as in Figure 3. Wordlists 'linked' closest to each other in the 'chain' are most closely related according to cognate percentages. Conversely, the end 'links' are most distantly related lexically. Note that Kolapis LS is obliquely related to the rest of the chain. Its closely relation is 87 PSC with Pulau Lubokan SN of the core 'link', while relating to other Suluk wordlists in the range 79-84 PSC.

For the seven wordlists representative of Sabah Suluk the lexical relations are in the range 82-92 PSC, and mutual intelligibility can be predicted to exist among them. Though the Suluk (Tausug) sample from Jolo PHL is considered as representing the same language, mutual intelligibility would not be predicted between it and the Suluk as spoken in Istimewa SN and Kolapis LS with which it shares relations of below 80 PSC.

The cognate relations between the representatives of Sabah Suluk would make dialect intelligibility testing superfluous according to the survey guidelines. Nevertheless testing was desired in order to confirm the one-language hypothesis for representatives of the language geographically dispersed in Sabah, as well as to ascertain the degree of intelligibility between Sabah Suluk and representatives of other Philippine languages as they are spoken in Sabah.⁶

SK (Istimewa SN)						
88	SK (Tabanak LD)					
89	91	SK (Pulau Lubokan SN)				
87	89	92	SK (Titingan TU)			
84	86	91	88	SK (Silam LD)		
82	84	89	89	86	SK (Semporna SA)	
82	83	87	84	84	82	SK (Kolapis LS)
77	81	84	84	82	87	79 TG (Jolo PHL)

Figure 2: PSC relations between Suluk (SK) and Tausug (TG) wordlists

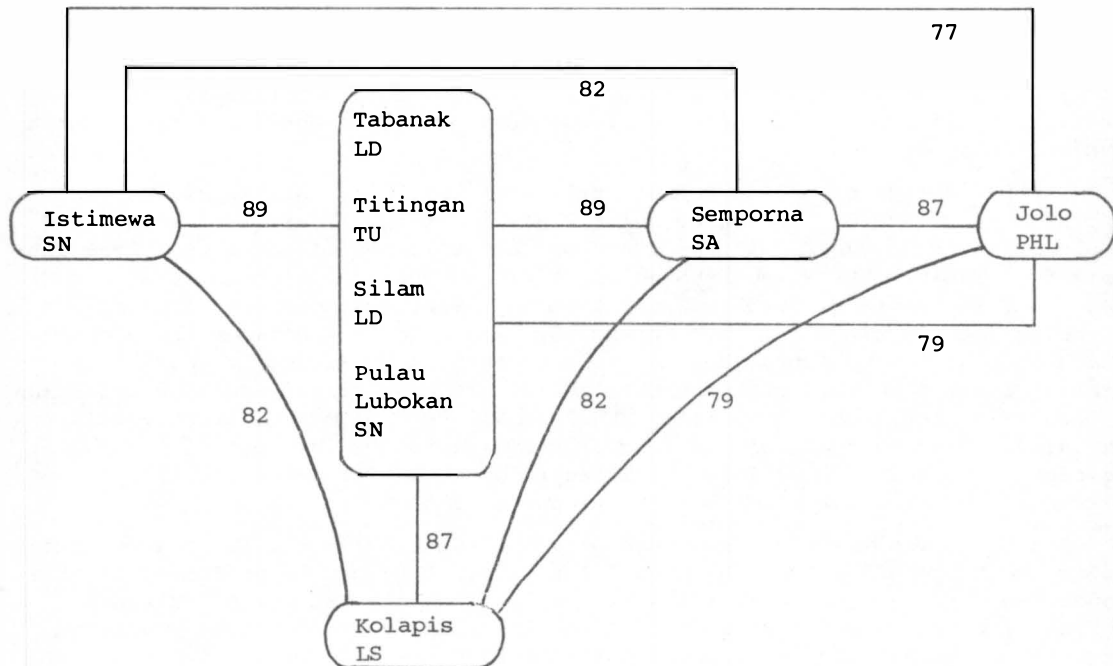


Figure 3: Lexical relations between eight Suluk (Tausug) wordlists presented as a dialect chain

2. TESTING PROCEDURE

Testing was subsequently accomplished in two Suluk villages, both located in Tawau District. The first village was Wakuba TU (Mengkuba in the vernacular) which was established more than 50 years ago. The second was Hidayat TU, a new housing project for the resettlement of Philippine immigrants from the village of Titingan TU in the city of Tawau. Its residents for the most part have lived in Sabah less than ten years, a condition which normally would have prevented its inclusion as a test point. It was included nevertheless because it was desirable to have a second Suluk test point and because it was considered that the comparison between short-term and long-term Suluk residents in Sabah might prove interesting.

As reference tapes had not been previously collected at either test point, the Suluk story recorded at Titingan TU, geographically close to both test points, was used as the hometown tape for both testing situations. The story from Kolapis LS was chosen as a second example of Sabah Suluk, representing also the most geographically distant of the other in-state samples. Two Sama-Bajau stories were included in the test set. One was a Bajau Banaran story which had also been recorded at Titingan TU. The Bajau Banaran language is spoken primarily in the smaller islands to the south of Tawitawi Island's west end. The wordlist representing the same language showed the highest overall cognate relations (53-57 PSC) of any Sabah Sama-Bajau language with Sabah Suluk wordlists. The second was a story recorded in Look PHL, a community in the Tongquil Island group of north-east Sulu Province and representing Sama (or Bajau) Balangingi, a dialect which has had prolonged contact with the Suluk (Tausug) language in the Philippines. Its relationship with the Jolo PHL wordlist was 62 PSC.⁷

3. TEST RESULTS

In both Wakuba TU and Hidayat TU village leaders assembled a complete corpus of ten test subjects. The results of the intelligibility testing are presented in Figure 4. Suluk comprehension of the two Suluk stories is in the range 91-100%, well within the range of single dialect intelligibility. There is significantly less intelligibility of the Sama-Bajau stories, the average scores ranging between 22-62%.

The corpora were not completely homogeneous with respect to the long-term versus short-term residency distinctions. Figure 5 reflects this distinction better, using a period of ten years residency as the criterion for dividing all test subjects into the two groups. For both intelligibility of the Suluk stories and intelligibility of the Sama-Bajau stories the range of average scores is broadened by the regrouping. Short-term residents averaged 89% on the Kolapis LS story, decreasing the lower bound for Suluk intelligibility of Suluk. The new range of 89-100%, while indicating perhaps greater dialect diversity, is still well within the range of single language intelligibility.

REFERENCE TAPE TEST POINT	SULUK		SAMA-BAJAU	
	Titingan TU	Kolapis LS	Titingan TU -Banaran	Look PHL -Balangingi
Hidayat TU	95 (--)	91 (84)	62 (57)	42 (51)
Wakuba TU	100 (--)	93 (84)	52 (57)	22 (51)

Figure 4: Results of intelligibility testing in two Suluk communities. (The PSC relations indicated (in parentheses) are those for villages represented by 'reference tape' and Titingan TU (Suluk).)

REFERENCE TAPE TEST POINT	SULUK		SAMA-BAJAU	
	Titingan TU	Kolapis LS	Titingan TU -Banaran	Look PHL -Balangingi
Suluks resident less than ten years	95	89	65	47
	6-10 10	5½-10 9½	2½-8½ 7½	2-8 5
Suluks resident more than ten years	100	94	48	16
	10 10	6-10 10	1-8 5½	0-4 2

Figure 5: Dialect intelligibility testing results for Suluk speakers grouped according to length of residency in Sabah. (The top score in each box is the average score (cited as a percentage) for all subjects in the group. Below each average score the range of individual scores and the median score (out of 10) are given.)

The new range for Suluk intelligibility of Sama-Bajau is 16-65%. Short-term residents averaged 17% better on the Bajau Banaran story and 31% better on the Sama-Balangingi story. The differences reflect in general the greater contact and bilingualism between the Suluk (Tausug) and Sama-Bajau languages in the Philippines than here in Sabah, though contact in some Sabah communities (e.g. Titingan TU) is not insignificant. The significantly lower scores on the Sama-Balangingi story reflect its more distant geographical locus from Sabah and consequently less frequent contact with Sabah Suluk speakers. Among the nine subjects grouped as long-term residents all but one, a 14 year old girl, lived in Wakuba TU. Although Wakuba TU has a mixed population, its non-Suluk residents are also not Sama-Bajau. Its Suluk residents therefore have much less contact with speakers of Sama-Bajau languages than do the Suluk residents of Hidayat TU who are outnumbered by Sama-Bajau speakers, as was the case in Titingan TU from where they had recently moved. The higher scores for both Hidayat TU and short-term subjects can reasonably be attributed to language learning.

It is yet of interest to determine the relation between Sabah Suluk and Suluk (Tausug) as spoken in its primary centres of the Sulu Archipelago. Further testing between long-term Sabah Suluk residents and Suluk (Tausug) communities of the Philippines is desired in order to establish the current degree of intelligibility between them.

4. NATIONAL LANGUAGE INTELLIGIBILITY

The results of testing Suluk comprehension of the Bahasa Malaysia test tape are displayed in Figure 6. Subjects in Hidayat TU received an average score of only 41% compared with the 69% scored by subjects in Wakuba TU. Again, more insight is gained by regrouping the scores of subjects. If the subjects are grouped according to the length of residency criterion as shown in Figure 7, the difference in the scores is widened. Subjects who have resided in Sabah for less than ten years averaged only 29% comprehension of the national language story. Long-term residents, on the other hand, averaged 86% comprehension.

	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Hidayat TU	41	5/5	40	14-65	8	8	7
Wakuba TU	69	5/5	34	16-60	6	8	5

Figure 6: Comprehension of the national language tape in two Suluk villages with sociological data. (RANGE indicates youngest and oldest test subject. EDUCATION figures indicate the number of subjects who had received some formal education (ED.), the average number of years of education those subjects had received (A.E.R.), and the average number of years of education for all subjects in the corpus (A.E.S.).)

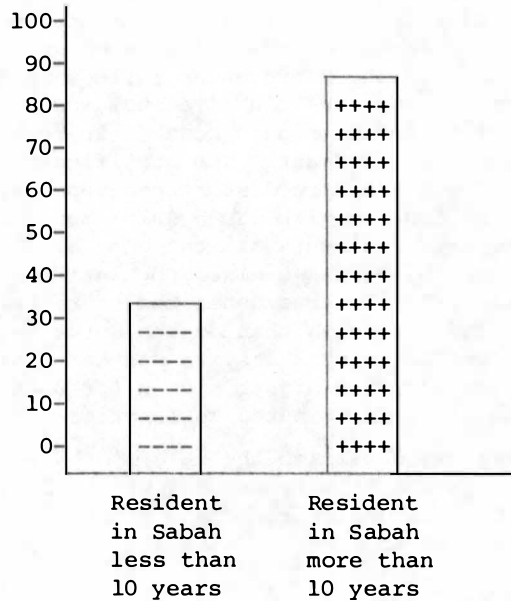


Figure 7: Comprehension of national language in Suluk villages in relation to length of residence in Sabah

A similar comparison can be made if the scores are grouped according to where subjects had received their education. Only the scores of subjects who had received some formal education are considered. This comparison is illustrated in Figure 8. Subjects educated in the Philippines scored 40% as a group, while subjects educated in Sabah scored 100% on the national language story. In both cases the factor of education significantly elevated the scores. Philippine-educated subjects averaged 8.5 years of education per subject and were 83% of the subjects who resided in the Philippines during their school-age years. Sabah-educated subjects averaged 6.3 years of education per subject and were 50% of the subjects who resided in Sabah during school-age years.

5. CONCLUSION

The results of dialect intelligibility testing in the Suluk villages of Hidayat TU and Wakuba TU confirm the lexicostatistical classification proposed by Smith. Two Sabah Suluk speech samples were understood by both long-term and short-term Suluk residents at levels well within the bound of single language intelligibility. Suluk intelligibility of Sama-Bajau speech samples was significantly less, particularly by long-term Suluk residents in Sabah who have been isolated from Sama-Bajau populations.

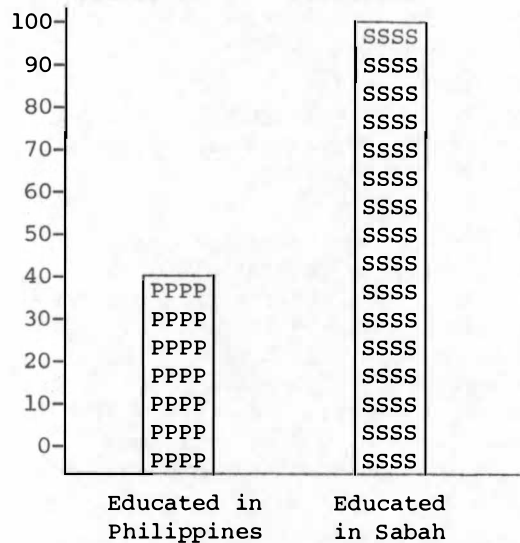


Figure 8: Comprehension of national language in Suluk villages for subjects with formal education in relation to the country in which they had been educated

NOTES

1. Pallesen's reconstruction of Suluk (Tausug) history differs from other popular versions which, he says, indicate the Suluk (Tausug) people pre-date the Sama-Bajau peoples in the Sulu Archipelago, the latter having supposedly come from Johore on the Malay peninsula. Pallesen's conclusions are based on reconstructions of the respective Suluk (Tausug) and Sama-Bajau parent languages and the tracing of subsequent borrowings between the daughter languages. He is primarily concerned with the nature of linguistic convergence between the Suluk (Tausug) language and the Sama-Bajau language group (Pallesen 1977:1).
2. The 1960 census figure is taken from Hardaker (1963:138 editor's note). The 1970 Sabah census listed 10,907 Suluk within the state.
3. This reported growth of the Suluk people in Sabah is consistent with the situation recorded for the Philippines. Pallesen (1977:11) notes that the Suluk (Tausug) population in the Philippines was about 325,000 in 1972, of whom 190,000 lived on Jolo Island. Since that time, however, the population of the island has been considerably decreased due to a forced dispersion of its residents. It is reasonable that many have come to Sabah where Suluk communities already existed, and where there would be relief from the pressures which burdened them in the Philippines. The Suluk (Tausug) language has broader influence and use throughout the Sulu Archipelago as a trade language.

4. For further notes on the etymology of the terms "Suluk" and "Tausug", see Mohring (1967:243) and Kiefer (1968:438).
5. Mr Cowie collected the wordlist for W.H. Treacher. F.A. Swettenham brought this and other wordlists together for his article published in 1880 by the *Journal of Straits Branch Royal Asiatic Society*. Swettenham's lists were reproduced in *The natives of Sarawak and British North Borneo* by H. Ling Roth, published in 1896 in London.
6. The relationship between Sama-Bajau languages and Suluk has been classified by Smith (in this volume) as that of different linguistic stocks, sharing membership in the North-western Austronesian superstock. This classification is confirmed by Pallesen's data (1977:151). Pallesen states that the two stocks have been in contact for about 700 years, during which convergence between them has occurred. This convergence is still on-going. Bilingualism is common between speakers of the two stocks, occurring in both directions, due to Suluk (Tausug) dominance throughout the archipelago and Sama-Bajau majorities in some areas of the archipelago (Pallesen 1977:23,40).
7. Both the wordlist and tape-recorded story from Look, Tongquil PHL, were provided by the Philippine Branch of the Summer Institute of Linguistics.

THE IDA'AN LANGUAGE

John E. Banker

0. INTRODUCTION

The Ida'an may be the earliest inhabitants in the eastern coastal area of Sabah, but very little has been written about their origins. Pallesen (1977: 175-176) notes that in Central Sulu Sama PHL folklore refers to earlier inhabitants of the Sulu Islands as Iraḡan, who are said to be related to the inland people of North Borneo. Sibutuḡ PHL legends also say that the original inhabitants of the island were Iraḡan or Kadazan from North Borneo. The term Iraḡan could refer to present day Ida'an, or it could be a more general term. Rutter writes "The old Bajau name for all the coast pagans was Ida'an." (1929:30). Several authors state that Ida'an is a general name for Dusun people or all indigenous languages of Sabah (Dunn 1980:29).

The oldest document written in a local Sabah language is the 'Idahan' Origin Myth. Harrisson suggests that this myth is closely related to the origin myths of the Kayans who now live over the Kalimantan border (1969-70:229-232). Though their origin is not clear, the Ida'an are known to have lived around the Lahad Datu area since before the 13th century.

Ownership of large birds' nest caves was a significant aspect of the culture of the 'Idahan' and other 'coastal pagans'. In the mid-1700s the 'Idahans' were threatened by the more powerful Suluks. In order to survive the 'Idahans' became Moslem and intermarried with the Suluks. Other 'pagan groups' may have moved inland at the time as a retreat from the Suluks (Harrisson 1969-70:234).

The Ida'an currently number between 5,000 and 6,000¹ and their main centres are around the city of Lahad Datu, the Ulu Tungku area of Lahad Datu District, the lower Kinabatangan River from Bilit to the mouth of the river (except for Abai), and on the Segaliud and Suanlamba Rivers that flow into Sandakan Bay. There are also Ida'an villages farther east on the north side of the Dent Peninsula at Dagat KN. At least one village is also reported to be located in the Sugut River area far to the north (Figure 1).

The people of this language group use three different autonyms. In the area around the town of Lahad Datu they are generally known as Ida'an, but the non-Muslim communities living in the Tungku LD area call their language Begahak and sometimes refer to themselves as Begahak Kadazan. In the lower Kinabatangan area and in Sandakan they call themselves Sungai, as do a number of Muslim communities belonging to several different language groups. Other spellings for Ida'an found in the literature are Eraan, Idaan, Idahan, Idan and Idayan (Dunn 1980).

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 85-90. *Pacific Linguistics*, C-78, 1984.

© John E. Banker

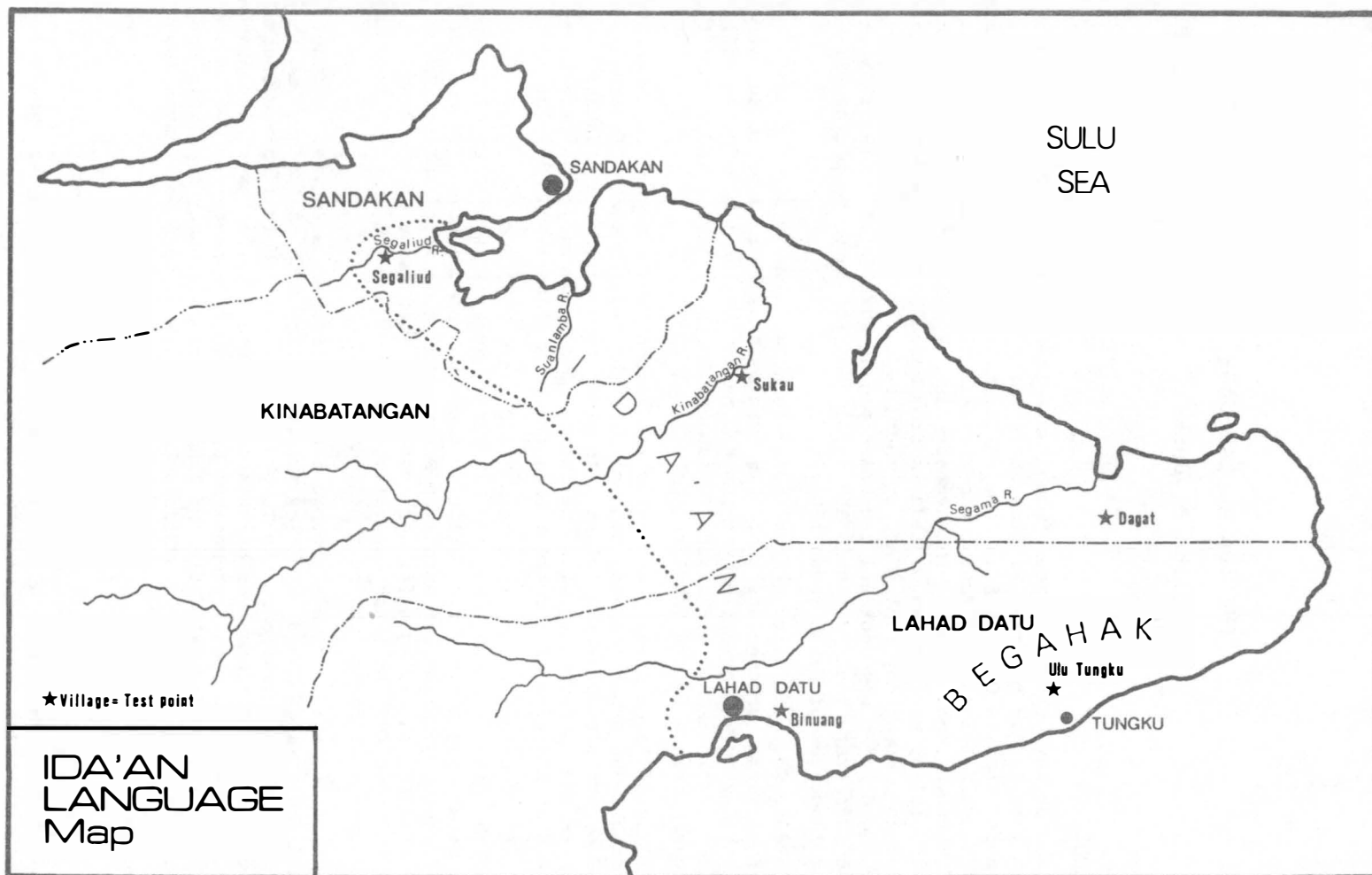


Figure 1: Ida'an language map

Tom and Barbara Harrisson (1969-70) have done extensive research of Ida'an caves and burial sites and have used much of this information as a basis for Sabah prehistory. Roth (1896) includes an Ida'an wordlist compiled by Spenser St. John in 1858 as an Appendix.

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classified Ida'an as a subfamily of the North-western Austronesian superstock. Its closest relation with other Sabah languages is in the range of 41-45 PSC with representatives of the Bornean stock of languages and the Banggi language.

Figure 2 shows the percentage of shared cognates (PSC) between various Ida'an/Begahak and Ida'an Sungai villages. Based on this lexicostatistical data Smith classifies Ida'an/Begahak and Ida'an Sungai as two different languages since no Ida'an Sungai village has more than 80 PSC with the Ida'an/Begahak villages. Smith said, however, that intelligibility testing is needed to clarify this point. The five villages in Lahad Datu District are all closely related to each other (87-95 PSC) and can be considered one homogeneous dialect, while the four Sungai villages show clear dialect distinctions (78-84 PSC).

In this study it will be shown that, based on the results of intelligibility testing, Ida'an, Begahak and Ida'an Sungai can all be classified as one language.

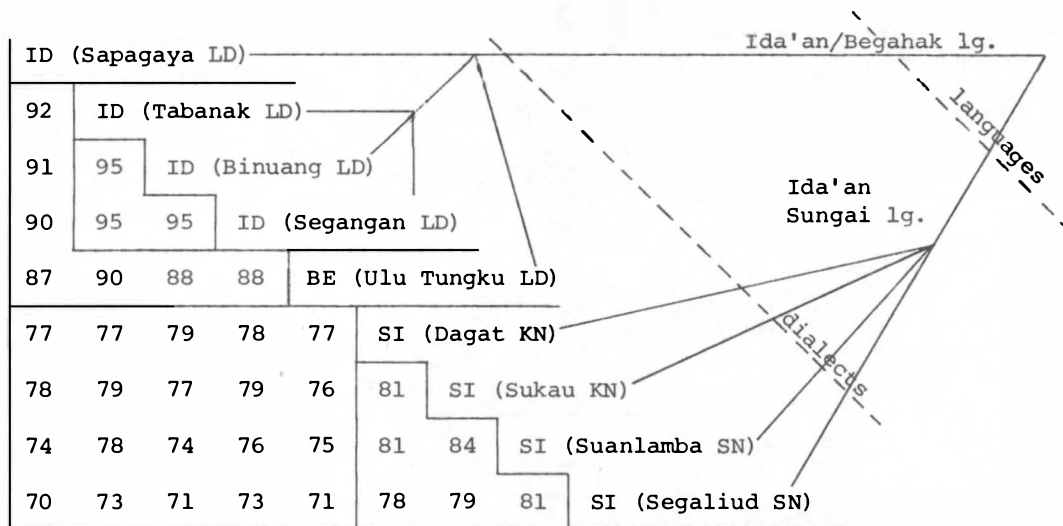


Figure 2: PSC of languages and dialects within the Ida'an subfamily.
(per Smith, in this volume) (ID = Ida'an, BE = Begahak,
SI = Sungai.)

2. TESTING PROCEDURE

Intelligibility testing was conducted in five villages: Binuang LD and Sapagaya LD, representing Ida'an; Ulu Tungku LD representing Begahak; and Sukau KN and Segaliud SN, representing Ida'an Sungai. All of the test sets included a hometown tape and a national language tape as well as reference tapes from Binuang LD, Ulu Tungku LD and Sukau KN.

The reference tapes from Sukau KN and Binuang LD were of good quality. The Sukau hunting story was good and an accurate test of intelligibility. The Binuang LD fishing story had more common terminology than would be desired, but was rated as a fair test. The Ulu Tungku LD tape was of poor quality, so test results from this village were considered unreliable.

3. TEST RESULTS

All five villages showed high intelligibility of Binuang LD and Sukau KN reference tapes (Figure 3). This was expected between the Lahad Datu villages, but the test results also show high intelligibility with Sukau KN even though relations only range 77-79 PSC between the Lahad Datu villages and Sukau KN. Though the results from the Ulu Tungku LD test story are not considered reliable, Ulu Tungku is shown to be clearly within the Ida'an language both by its relation of 87-90 PSC with other Lahad Datu Ida'an villages and its 99% comprehension of the Binuang LD story and 96% comprehension of the Sukau KN story. Intelligibility testing seems to clearly show that Ida'an, Begahak and Ida'an Sungai are all within one language.

REFERENCE TAPES TEST POINTS	Binuang LD	Ulu Tungku LD	Sukau KN
Binuang LD	--	76* (88)	93 (79)
Sapagaya LD	(91)	61* (87)	100 (78)
Ulu Tungku LD	99 (88)	--	96 (76)
Sukau KN	86 (77)	68* (76)	--
Segaliud SN	97 (71)	70* (71)	99 (79)

Figure 3: Results of intelligibility testing in five Ida'an villages. (Scores are given as percentages, PSC relations are in parentheses. * indicates unreliable scores due to a poor test tape.)

That the Ida'an/Begahak language is a homogeneous group with only sub-dialect distinctions is substantiated by intelligibility test results. Since the PSC relations of Ulu Tungku LD with the other four villages are consistently lower than the relations of the villages to each other, and the Ulu Tungku LD test scores were unreliable, that testing needs to be redone. Ulu Tungku LD should also be tested with a Sapagaya LD reference tape as a comparison with the 99% Binuang LD intelligibility score. Also the four Ida'an Sungai dialects should be cross-tested. The one test conducted at Segaliud SN with a Sukau KN reference tape showed 99% intelligibility, but with only one test point no conclusions can be drawn about the homogeneity of the Ida'an Sungai dialects.

Further investigation could be done to determine the validity of the lower PSC relations of Ida'an Sungai dialects with Lahad Datu villages. One example of possible distortion of the wordlists is that the Segaliud SN wordlist was elicited from a teenager studying in Bahasa Malaysia. This could reflect only one person's limited knowledge of Ida'an or it could be indicative of the Ida'an Sungai dialects being heavily influenced by Bahasa Malaysia. Were the latter true, then Sukau KN's low intelligibility of the national language is unexplainable. Either case could cause lower PSC relations with the Lahad Datu villages which are not as heavily influenced by Bahasa Malaysia.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Ida'an understanding of Bahasa Malaysia was tested at each of the five test points. Four villages showed high comprehension (82-92%) of the national language tape, but Sukau KN subjects averaged only 57% (Figure 4). Sukau KN is more remote than the other test points, but showed a somewhat higher PSC relation with Bahasa Malaysia than the Lahad Datu villages.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Binuang LD	89	7/3	40	17-67	6	9	5
Sapagaya LD	92	3/2	32	17-50	2	7	3
Ulu Tungku LD	90	6/4	34	17-65	4	6	2
Sukau KN	57	8/2	43	22-65	4	8	3
Segaliud SN	91	6/1	28	18-51	6	7	6

Figure 4: National language intelligibility at five Ida'an test points with sociological data. (Scores are given as percentages. The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average number of years of education which those subjects received (A.E.R.), and the average number of years of education per subject in the test corpus as a whole (A.E.S.).)

The significant sociological factor in these test results is education using Bahasa Malaysia. Figure 5 displays the average intelligibility scores for the five test points, contrasting subjects with some education with those with no education. Even in Sukau KN where overall scores are low, subjects with some education scored almost twice as high as those with no education. Further investigation into the Sukau KN dialect and comparison with another Ida'an Sungai village's (such as Dagat KN) understanding of Bahasa Malaysia might be useful in clarifying some unanswered questions about national language intelligibility test results.

TEST POINTS	SOME EDUCATION	NO EDUCATION
Binuang LD	100% (6)	73% (4)
Sapagaya LD	95% (2)	90% (3)
Ulu Tungku LD	100% (4)	83% (6)
Sukau KN	75% (4)	38% (6)
Segaliud SN	98% (5)	50% (1)

Figure 5: Ida'an intelligibility of the national language based on education. (The number of test subjects for each case is given in parentheses under the average intelligibility score.)

5. CONCLUSION

In summary, the high intelligibility scores registered by those tested from each of the Ida'an, Begahak and Ida'an Sungai villages tested show that Ida'an, Begahak and Ida'an Sungai are one language. All five Ida'an villages in Lahad Datu form one homogeneous dialect, but further testing is needed to clarify the dialect distinctions of the Kinabatangan and Sandakan Ida'an Sungai villages.

NOTE

1. This population figure is an estimate based on a report from the Lahad Datu District Officer (June 15, 1982) listing all villages with their population and the languages spoken in them.

THE MALAYIC LANGUAGE FAMILY

David C. Moody

0. INTRODUCTION

The Malayic language family is represented in Sabah by two subfamilies. The first is singly represented by the Iban language. The second is called the Malayic subfamily (→ Smith, in this volume).

0.1 Iban

The Iban, numbering approximately 350,000 (Seymour 1977:178) are the most numerous of Borneo's indigenous peoples. However, they reside primarily in Sarawak where they constitute nearly one-third of that state's population. In Sabah, they number under 500. Only one village of Iban speakers was found, located in the Tawau District (Figure 1).

The Iban (also Hivan, Needham 1955:169) are also known as Sea Dayak (also Daya, Dáyá, Dayak, Dáyák, Dayer, Diak, Dyak, Maxwell 1970:93). The term Iban is, according to LeBar (1972:180), a linguistic borrowing from Kayan. The material available on the Iban is primarily ethnological in nature, though as early as 1896 there is record of a published Iban wordlist (Roth 1896).

Dr Asmah Haji Omar has given considerable attention to the Iban language. In addition to her doctoral thesis, she has written several smaller articles including a grammatical sketch and a comparative study of numeral classifiers in Iban and Malay.

Other Iban materials include an English-Iban phrase book (Barry n.d.), a translation of the New Testament, and numerous Iban language materials published by the Sarawak division of Dewan Bahasa dan Pustaka.

0.2 The Malayic subfamily

The Malayic subfamily as represented in Sabah comprises two languages. The first is termed Bahasa Malaysia/Cocos Malay by Smith (in this volume). The second is called Brunei/Kedayan.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 91-100. *Pacific Linguistics*, C-78, 1984.

© David C. Moody

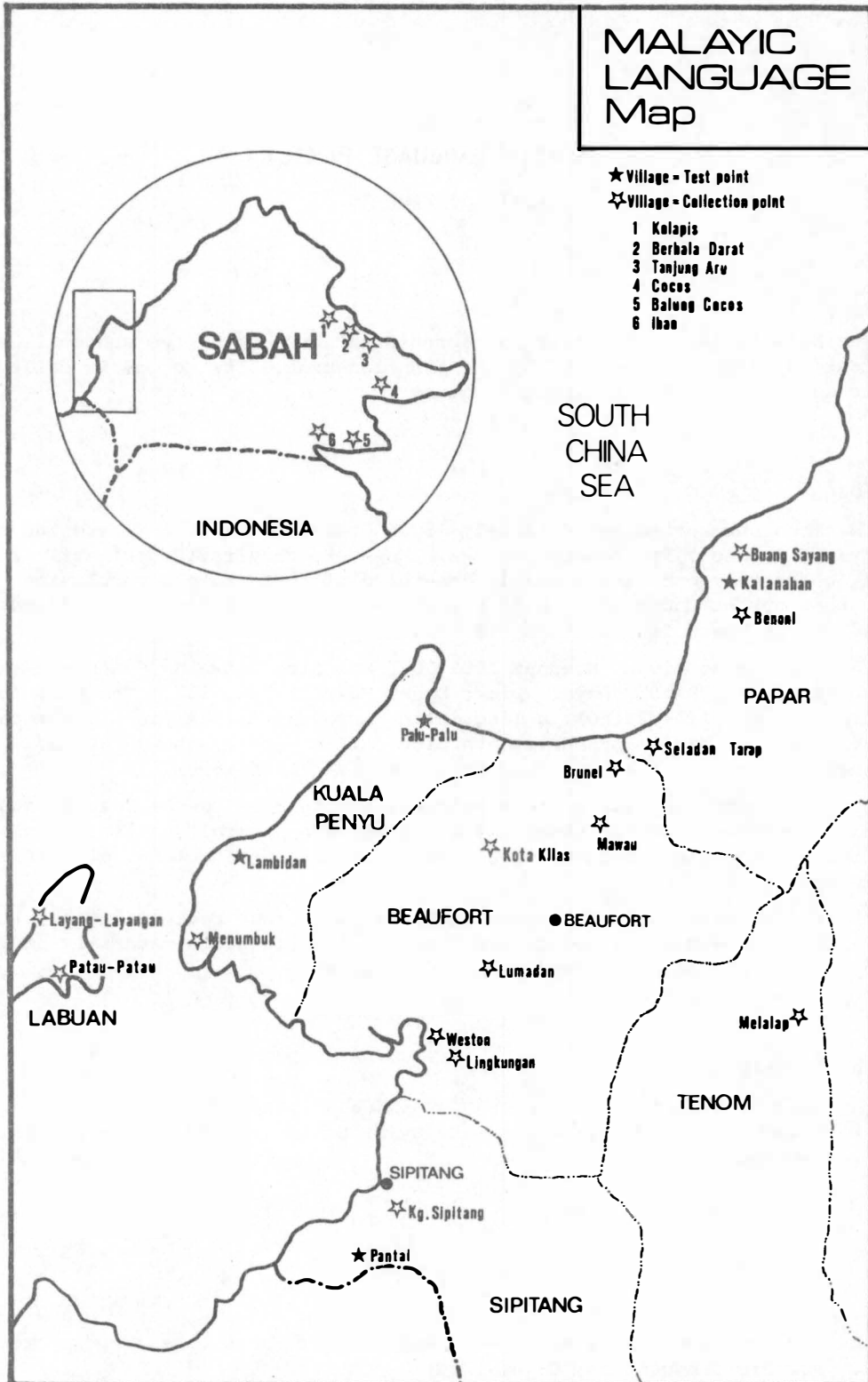


Figure 1: Malayic language map

0.2.1 Bahasa Malaysia/Cocos Malay

The Cocos Islanders are found in the Tawau and Lahad Datu districts and now number over 3,000.¹ Their origins are known to be in the small group of islands known as the Cocos or Keeling Islands, located to the south of Sumatra in the Indian Ocean (Smith, in this volume).

Bahasa Malaysia, the national language, is spoken as a first language by approximately 23,000 people in Sabah.² There is a relatively large and still-growing body of literature concerning Bahasa Malaysia. Considered as one of the world's major languages, Malay is spoken by over 140 million people,³ primarily in Malaysia and Indonesia. Dr Asmah Haji Omar (1975:350-354) has compiled a list of 83 dictionaries alone. Other facets of the language are likewise being studied by an ever-growing list of capable linguists.

0.2.2 Brunei/Kedayan

The Brunei or Kedayan or Brunei-Kedayan people live mainly along Sabah's south-west coast in the districts of Papar, Beaufort, Kuala Penyu, Sipitang, and Labuan (Figure 1). Additional groups live in the Tenom, Sandakan and Labuk-Sugut districts. Current population figures for the 'Brunei' and 'Kedayan' groups are estimated at 35,000 and 11,500 respectively.⁴

The Brunei people are very early residents in Borneo. James Ongkili (1972) writes that the history of Sabah (as well as Sarawak and Brunei) for several hundred years prior to the establishment of the North Borneo Chartered Company, was largely the history of the Brunei Sultanate. During that period the sultanate claimed suzerainty over Sabah, though effective control was probably exercised only in coastal and riverine areas.⁵

Shariffuddin (1969:15ff.) notes that the usual version of the origin of the Kedayans indicates that they were rice farmers in Java. About five centuries ago, Sultan Bulkiah of Brunei visited the island and was impressed with the importance of rice to the Javanese. He thus recruited some of them to return with him to Brunei to teach his own people how to cultivate the crop. Shariffuddin notes the meaning of the term Kedayan, a *retainer*, fits this version of their history. Ongkili (1972:9) renders a different version of the Kedayan migration, which it may be noted is not necessarily contradictory with Shariffuddin's. He indicates that Sultan Bulkiah returned from one of his many voyages with a new bride, a Javanese princess. Her followers intermarried with the Brunei people and they and their descendants became known as Kedayans.

1. LEXICOSTATISTICAL CLASSIFICATION

A total of 24 wordlists were elicited from persons who referred to their mother-tongue as 'Iban', 'Cocos', 'Brunei', 'Kedayan' or 'Brunei-Kedayan'. 'Iban' is represented by one list, 'Cocos' by two, 'Brunei' by 15, 'Kedayan' by five and 'Brunei-Kedayan' by one.⁶

Each of these lists represented a speech form of a particular locality. A Bahasa Malaysia wordlist which reflects standard Sabah usage of the national language was also prepared. The comparison of 344 wordlists from throughout the state demonstrated the above lists were more closely related among themselves than with other wordlists.

1.1 Iban

The 'Iban' list related to all other lists in the range of 59-65 percent of shared cognates (PSC). Smith (in this volume) defined the range which distinguishes between separate subfamilies as 60-75 PSC, and on that basis established the Iban language as the sole representative within Sabah of its subfamily grouping.

1.2 The Malayic subfamily

The counterpart of Iban, Smith called the Malayic subfamily. Within this subfamily there is a further division into two language groupings. The wordlists labelled 'Brunei', 'Kedayan', and 'Brunei-Kedayan' constitute one grouping, while the two 'Cocos' wordlists and the Bahasa Malaysia list constitute the other. The average relation between these two divisions is 73 PSC.

1.2.1 Bahasa Malaysia/Cocos Malay

The two 'Cocos' wordlists demonstrated 91 PSC between themselves and 82 and 88 PSC with Bahasa Malaysia. For Smith, the language threshold was 80 PSC; that is, dialects exhibiting lexical relations below 80 PSC were considered to be dialects of different languages between which it would not be expected to find mutual intelligibility. Conversely, wordlists whose shared lexicon was greater than 80 PSC were designated as dialects of the same language between which mutual intelligibility was expected. Therefore, the two 'Cocos' wordlists were grouped with the Bahasa Malaysia wordlist as a single language, designated by Smith as Bahasa Malaysia/Cocos Malay. Wordlists which showed a sharing of cognates as great as 85 PSC or greater were considered to represent the same dialect. Therefore, the three Bahasa Malaysia/Cocos Malay wordlists may be perceived as a chaining of two dialects, the dialect represented by the two 'Cocos' wordlists representing two subdialects, one of which has a closer relation (88 PSC) to Bahasa Malaysia than the other (82 PSC).

1.2.2 Brunei/Kedayan

The 21 wordlists labelled 'Brunei', 'Kedayan', and 'Brunei-Kedayan' showed a high degree of common vocabulary. Thirteen of the lists, all from the districts of Labuan, Papar, Beaufort and Tenom, are related with each other in the range 87-94 PSC and relate with six or more of the 21 lists at 90 PSC or higher. Of these 13, four wordlists from villages in Labuan, Papar and Beaufort relate with each other in the very high range 93-94 PSC. Each of the three autonyms is represented in the group of 13 wordlists. The remaining lists, from the districts of Kuala Penyu, Sipitang, Labuk-Sugut and Sandakan, show relations of 90 PSC or higher with four or less of the 21 lists. Five of these eight do not relate at 90 PSC or above with any other list.

The lexical relations mentioned above are indicated in Figure 2. The most prominent feature of the chart is the breadth of linguistic homogeneity among the 21 'Brunei', 'Kedayan' and 'Brunei-Kedayan' samples. And, if Berhala Darat SN is excluded, all but 17 of the 190 relations for the remaining 20 wordlists exhibit percentages of shared cognates within the limits indicative of a single dialect. Diversity is thus limited to the finer distinctions manifested by

subdialects and does not coincide directly with autonymical designations. Smith thus classified the 21 wordlists as representing a single language, Brunei/Kedayan, which was without overt dialect distinctions.

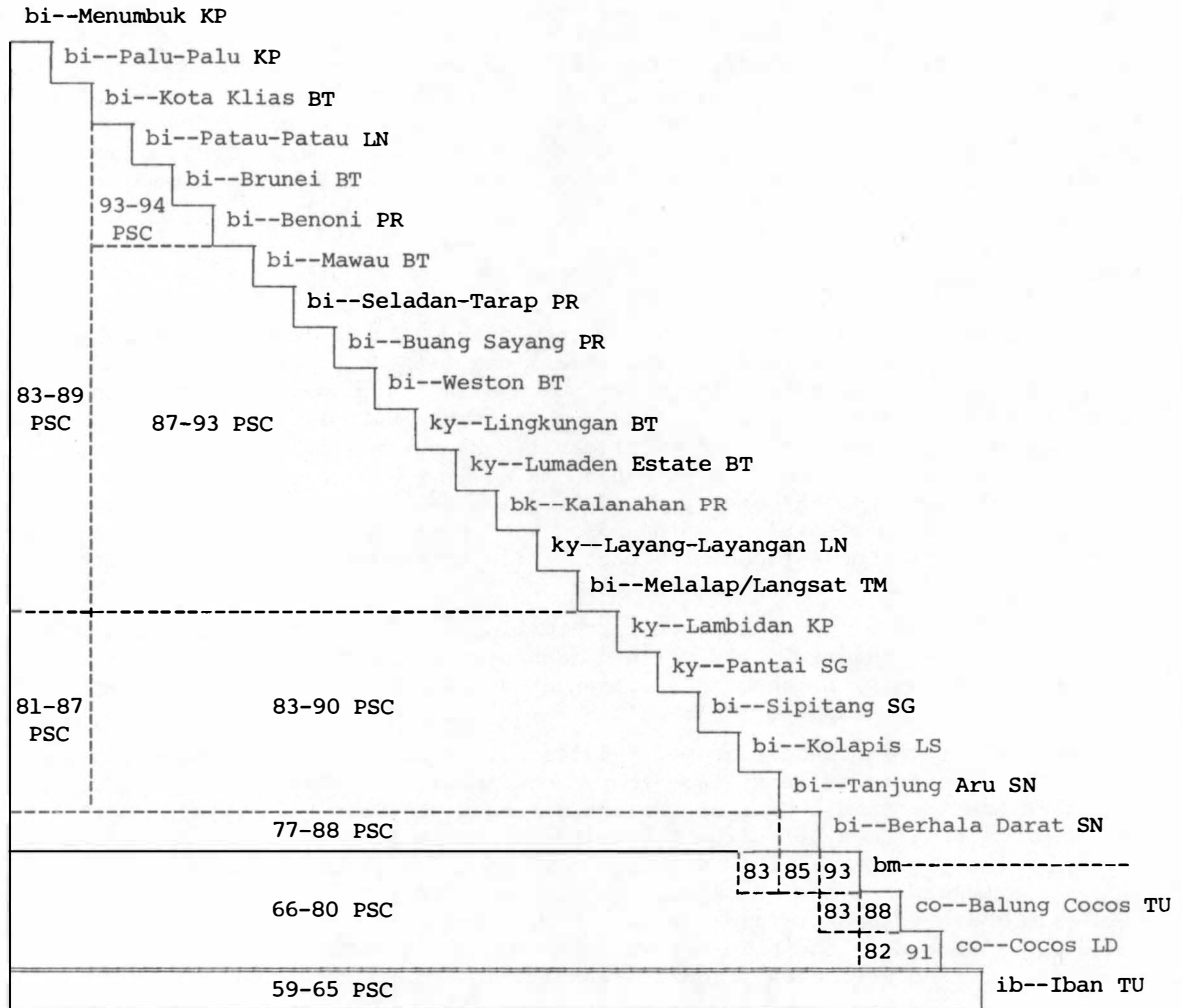


Figure 2: PSC relations within the Malayic family. (Miniscule abbreviations indicate autonyms as follows: bi = 'Brunei', ky = 'Kedayan', bk = 'Brunei-Kedayan', bm = Bahasa Malaysia, co = 'Cocos', and ib = 'Iban'.)

Berhala Darat SN occupies a notable position within the chart. Its highest relation is with Bahasa Malaysia (93 PSC), yet it demonstrates relations with 14 of the other 20 Brunei/Kedayan wordlists in the range 80-88 PSC. This seems a clear case of heavy borrowing of vocabulary items from the national language in the Berhala Darat SN lexicon. This borrowing has effected not only inflated scores of shared cognates with Bahasa Malaysia/Cocos Malay wordlists, but also deflation of its lexical similarity with the other Brunei/Kedayan wordlists. The same dual phenomena have likewise occurred, to a lesser degree, in the wordlists collected from the other east coast villages of Tanjung Aru SN (85 PSC with Bahasa Malaysia) and Kolapis LS (83 PSC with Bahasa Malaysia). These inflation-deflation factors in no way detract from the substance of Smith's classification.

2. TESTING PROCEDURES

The purpose of intelligibility testing in the Malayic language family was primarily to establish whether mutual intelligibility would confirm the homogeneity of the Brunei/Kedayan language. Bahasa Malaysia is treated here only as the national language, and not as a member of the Malayic subfamily. A refinement of the testing procedures, primarily the control of test corpora to exclude subjects whose exposure to Bahasa Malaysia would result in inflated scores, would have been necessary to distinguish comprehension based on linguistic similarity from comprehension based on the status and widespread use of Bahasa Malaysia as the national language. Effective control of test corpora in such a manner would indeed be difficult.

The test points for intelligibility testing included a representative of each of the three autonymically defined groups of Brunei/Kedayan. Palu-Palu KP represented 'Brunei'; Kalanahan PR represented 'Brunei-Kedayan'; and Pantai SG and Lambidan KP represented 'Kedayan'.

The test set for each test point included, in addition to the hometown tape, reference tapes from at least two other Brunei/Kedayan villages, the Iban reference tape and the Bahasa Malaysia reference tape. In Palu-Palu KP ('Brunei') and Lambidan KP ('Kedayan'), the 'Brunei' story from Berhala Darat SN and the 'Kedayan' story from Pantai SG were used, in addition to the hometown reference tape, as samples of Brunei/Kedayan. In Kalanahan PR ('Brunei-Kedayan') the Palu-Palu ('Brunei') story and the Pantai SG ('Kedayan') story were used. In Pantai SG ('Kedayan') the Palu-Palu KP ('Brunei') and Kalanahan PR ('Brunei-Kedayan') stories were used (Figure 3).

During testing in Palu-Palu KP and Lambidan KP it was noted that the Pantai SG reference tape used for some test subjects was not clear. Furthermore, in Lambidan KP the test corpus was much older than the average, the median age being 60. Both of these factors could predictably lower scoring. However, in Palu-Palu KP the average score of those tested with the clear Pantai SG reference tape was 65%, not significantly higher than the overall 56% average for all test subjects in the corpus; the level of comprehension is still below 80%. The overall average on the same reference tape in Kalanahan PR, including those who listened to a poorer quality tape, was 95%. Thus, the quality of the Pantai SG reference tape was not considered a significant factor in the lower average scores on that tape in Palu-Palu KP and Lambidan KP. Likewise, the advanced age of the Lambidan KP corpus did not seem to invalidate their results.

REFERENCE TAPES TEST POINTS	BRUNEI/KEDAYAN					IBAN
	Palu-Palu KP ('Brunei')	Berhala Darat SN ('Brunei')	Kalanahan PR ('Brunei-Kedayan')	Pantai SG ('Kedayan')	Lambidan KP ('Kedayan')	Iban TU
Palu-Palu KP ('Brunei')	100 (--)	93 (78)		56 (85)		36 (62)
Kalanahan PR ('Brunei-Kedayan')	100 (86)		100 (--)	95 (87)		60 (62)
Pantai SG ('Kedayan')	84 (85)		88 (87)	97 (--)		61 (61)
Lambidan KP ('Kedayan')		92 (83)		68 (86)	100 (--)	45 (61)

Figure 3: Results of dialect intelligibility testing (shown as percentages) in Brunei/Kedayan villages. (Autonyms are enclosed in quotation marks. PSC relations are in parentheses; those marked (--) indicate hometowns.)

3. TEST RESULTS

Figure 3 displays the results of the intelligibility testing in Brunei/Kedayan villages. The test results were not completely predictable.

Initially it is clear that the testing of the Iban story in the Brunei/Kedayan villages produced predictable results. Intelligibility scores range from 36% to 61%. The low scores are very much in line with lexicostatic scores in the range 61-62 PSC.

The departure from the expected is observed in the testing between Brunei/Kedayan villages. With the exception of two scores, intelligibility of the Brunei/Kedayan reference tapes at the Brunei/Kedayan test points is in the range 84-100%. Scores in this range generally indicate good intelligibility and linguistic homogeneity, using 80% comprehension as the lower bound of single language intelligibility. However, the 56% and 68% scores of Palu-Palu KP and Lambidan KP on the Pantai SG reference tape are much lower than expected. The technical quality of the tape has already been eliminated as a factor. The lexical relations between Pantai SG and Palu-Palu KP and Lambidan KP are 85 PSC and 86 PSC respectively which would indicate homogeneity at the single dialect level. Furthermore, both Lambidan KP and Pantai SG use the self-referent 'Kedayan'.

Although the high degree of shared vocabulary and common autonym lead us to expect mutual intelligibility, the aberrant scores are explainable.

Both Palu-Palu and Lambidan are located in Kuala Penyu District; Pantai is located in Sipitang District. Even though the lexical relations are high, it is likely that there are areal differences of pronunciation and intonation which affect comprehension. These regional differences apparently hinder comprehension more than the shared lexicon facilitates it. The effect of regional differences may also account for the lower intelligibility of Palu-Palu (36%) and Lambidan (45%) on the Iban story; Kalanahan PR and Pantai SG scored 60% and 61% respectively.

Dialect intelligibility testing has demonstrated that the homogeneity of the Brunei/Kedayan speech varieties is not as complete as uniformly high percentages of shared cognates would indicate. Intelligibility tends toward unidirectionality between Pantai SG and the two samples from Kuala Penyu. Furthermore, these results indicate the skewing is geographically determined rather than being determined by cultural or etiological factors associated with the variety of autonyms. The testing thus indicates the existence of at least two dialects of Brunei/Kedayan between which there is limited intelligibility. Further testing is desirable to determine more completely the range of dialect diversity within the Brunei/Kedayan language.

Finally, lack of intelligibility of Iban by the Brunei/Kedayan test corpora at this point confirms Smith's classification of the two as representing separate linguistic subfamilies, though the proof of the classification is beyond the capability of the testing procedures used in this study. Although the relation between Brunei/Kedayan and Bahasa Malaysia/Cocos Malay was not tested because of the difficulties involved in setting up an unbiased test corpus in Brunei/Kedayan villages, further investigation of this relation is desired. The use of Cocos reference tapes at Brunei/Kedayan test points would give an indication of unbiased intelligibility in one direction, and mother-tongue Malay speakers could be tested on their comprehension of Brunei/Kedayan reference tapes.

4. NATIONAL LANGUAGE INTELLIGIBILITY

The four Brunei/Kedayan test corpora were also tested on their ability to understand the national language. The results of the testing on the Bahasa Malaysia reference tape are displayed in Figure 4. The test results generally reflect a high level of competence in understanding Bahasa Malaysia in these four Brunei/Kedayan villages. The range of average intelligibility scores is 78-98% and the median score for all subjects is 87%.

The raw data can be sorted in various ways to reveal that persons in the categories 'Male', 'Under 35 Years of Age', and 'Formally Educated' generally do better than their counterparts in the opposing categories, though this is not always the case. Figure 5 displays this information. It may also be noted that 58% of the males in the combined corpora had received some formal education (including those who had received only adult education) while the corresponding figure for females is 22%.

TEST POINT	AVG. SCORE	SEX M/F	AGE			EDUCATION		
			AVG.	MED.	RANGE	ED.	A.E.R.	A.E.S.
Palu-Palu KP	78 (74)	4/5	39.1	40	22-60	5	7.8	4.3
Kalanahan PR	98 (77)	5/3	40.8	44	14-67	5	6.6	4.1
Pantai SG	86 (78)	5/5	38.5	26	18-78	3	10.7	3.2
Lambidan KP	89 (79)	5/5	48.7	60	23-80	2	4.5	0.9

Figure 4: Results of national language intelligibility testing with socio-logical data for four Brunei/Kedayan villages. (Scores are given as percentages. PSC relations are shown in parentheses. The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average years of education which those subjects had received (A.E.R.), and the average years of education per subject in the test corpus as a whole (A.E.S.).)

TEST POINT	CORPUS SCORE	SEX		AGE		EDUCATION	
		AVG.M SCORE	AVG.F SCORE	AVG. UNDER 35 SCORE	AVG. OVER 35 SCORE	AVG. ED. SCORE	AVG. NON-ED. SCORE
Palu-Palu KP	78/9	90/4	69/5	83/3	76/6	90/5 3M, 2F	64/4
Kalanahan PR	98/8	98/5	97/3	100/3	96/5	98/5 3M, 2F	97/3
Pantai SG	86/10	80/5	92/5	98/5	74/5	97/3 3M	81/7
Lambidan KP	89/10	95/5	82/5	88/4	89/6	100/2 2M	86/8
COMBINED AVERAGES	87/37	91/19	84/18	93/15	84/22	95/15 11M, 4F	82/22

Figure 5: National language intelligibility results in Brunei/Kedayan villages displayed according to sex, age and education differentials. (The average score for each category (shown as percentage) is followed by the number of subjects for which the average applies. The boxes containing scores of those subjects who had received formal education (AVG. ED. SCORE) also show the number of males and females included in the educated group. NON-ED. indicates subjects without formal education.)

5. CONCLUSION

Dialect intelligibility testing among four Brunei/Kedayan villages essentially confirms Smith's lexicostatistical analysis which classified them as representatives of a single language. There is not, however, complete mutual intelligibility among these representative villages. Dialect variations indicated by the testing seem to be determined by regional rather than ethnic factors. Intelligibility of the Iban language was low, also confirming Smith's classification. Intelligibility of Bahasa Malaysia was generally high.

NOTES

1. The Cocos population of Sabah was listed as 2,731 in 1970. Using that figure as a base and assuming an annual population growth of two percent, the present population would be about 3,400. Two percent population growth is an arbitrary choice, intended only to give a rough estimate of the current population.
2. The 1970 Sabah census enumerated 18,365 Malays. Allowing for two percent annual population growth, the present Malay population would be about 22,800.
3. The figure 140 million is given in *Kamus lengkap*, edited by Drs Awang Sudjai Hairul and Yusoff Khan, Petaling Jaya: Pustaka Zaman Sdn. Bhd. (1977). The number of mother-tongue Malay speakers would be considerably less. Compare Voegelin and Voegelin (1977:179), who indicate 10,000,000 native speakers.
4. The 1970 Sabah census enumerated the two groups separately. The 'Brunei' numbered 28,152 while the 'Kedayan' population was recorded as 9,624. The estimated figures in the text assume two percent annual population growth added to these 1970 census figures.
5. In 1704, the Sultan of Brunei allegedly gave control over the territory of North Borneo to the Sultan of Sulu as reward for the latter's services in the Brunei civil war. The Sulu claim to North Borneo and Brunei's denial of the cession of its territory from that date has resulted in an entangled controversy, which most recently has been fired again in the Philippine claim to Sabah. Details of the controversy can be found in numerous articles, among which are the following: Martin Meadows "The Philippine claim to North Borneo", *Political Science Quarterly* 78/3:321-335, 1962; Leigh R. Wright, "Historical notes on the North Borneo dispute", *Journal of Asian Studies* 25/3:471-484, 1966; Brock K. Short, "Brunei, Sulu and Sabah: an analysis of rival claims", *Brunei Museum Journal* 1:133-146, 1969?; H.G. Tregonning, "The Philippine claim to Sabah", *Journal of the Malaysian Branch of the Royal Asiatic Society* 43(Part 1)/217:161-170, 1970.
6. In this paper, the convention of quotation marks around autonyms is used to distinguish them from language classifications.

THE WEST COAST BAJAU LANGUAGE

Elizabeth F. Banker

0. INTRODUCTION

The Bajau are a culturally and linguistically diverse people living in the Southern Philippines, Eastern Indonesia and Sabah, Malaysia. Knowledge of the origin of the Bajau is obscure, based primarily on oral tradition. As early as 1780 an English captain, Thomas Forrest, on a trip to New Guinea, encountered Bajau fishermen along the north-east coast of Borneo who were said to have come originally from Johore, at the entrance of the Straits of Malacca (Warren 1972).

In this paper, only one of the Bajau language groups will be discussed, namely the group designated by Smith (in this volume) as the West Coast Bajau of Sabah.¹ This group uses the self-referents Sama and Bajau.

It is not known when the West Coast Bajau first arrived in Sabah, but when Ivor Evans first became acquainted with the Bajau in the Kota Belud area in 1910, their settlements seemed to be well established, and he supposed that they had been there for at least 200 years. Those whom he interviewed claimed that only seven generations, including their own, had passed since their ancestors arrived in Borneo (Evans 1952).

Evans (1952) referred to two other Bajau language groups known as "Samah-Samah" or "Samar Lambuh" and "Samar Laiyun" who lived on Sibutu Island and Musa Island. Their languages were mutually intelligible with Kota Belud Bajau. Sather (1965) described "West Coast Sama" as the language spoken from Papar to Kudat, subdivided into regional varieties, each the product of years of isolation.

The West Coast Bajau number approximately 40,000.² They live along the Sabah coast from Kuala Penyu in the south-west as far as Terusan LS. There is also a settlement of West Coast Bajau on Jambongan Island (Figures 1 and 2).

1. LEXICOSTATISTICAL CLASSIFICATION

According to Smith's (in this volume) lexicostatistical classification of West Coast Bajau, there is a central network of seven wordlists linked by relationships of 90 percent of shared cognates (PSC) or higher (Figure 3).

Using Mengkabong TN as a representative sample, the remaining seven non-central wordlists are connected to this central network by relationships of 73-86 PSC (Figure 4).

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 101-112. *Pacific Linguistics*, C-78, 1984.

© Elizabeth F. Banker

101



Figure 1: West Coast Bajau language map

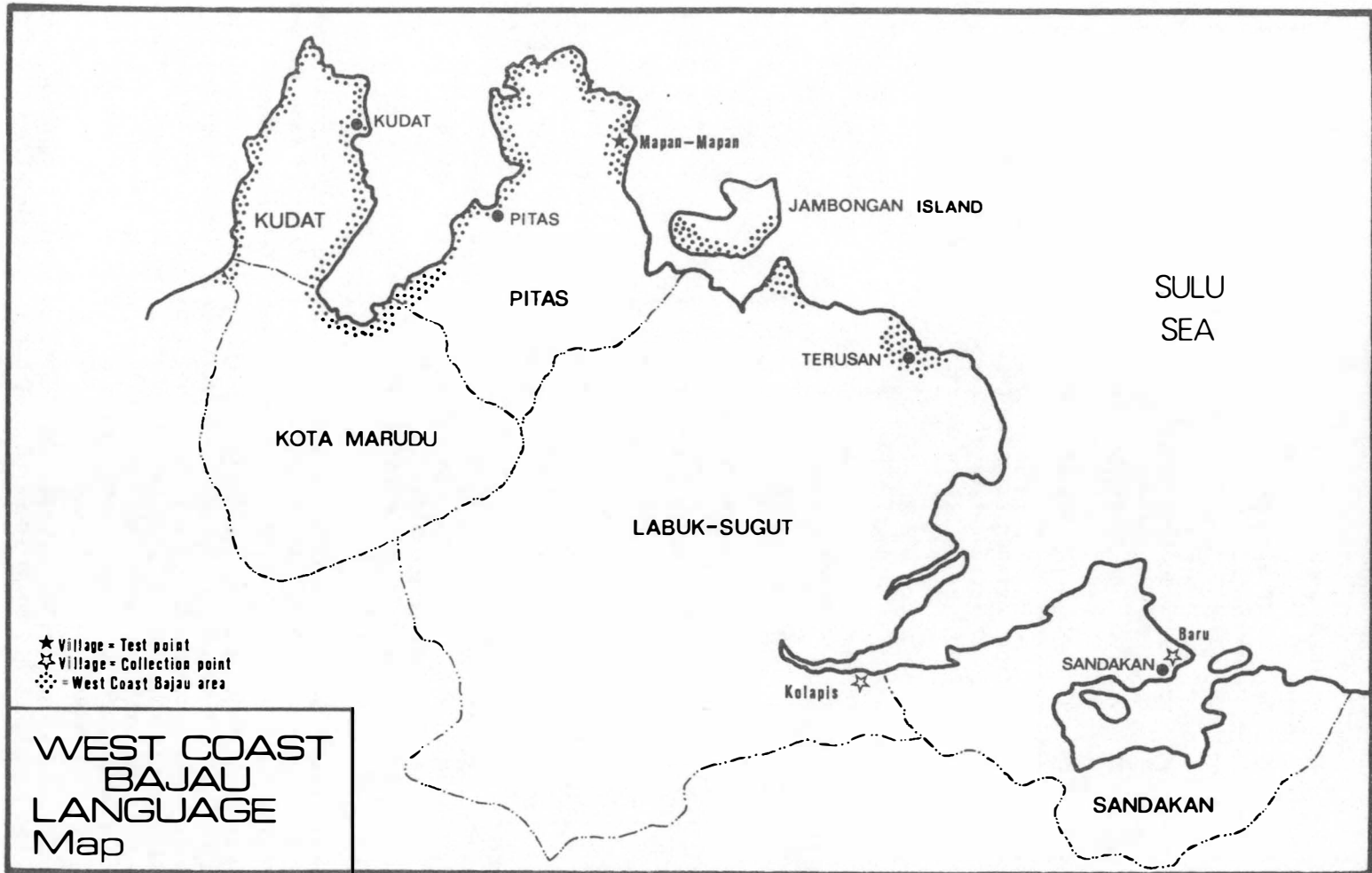


Figure 2: West Coast Bajau language map

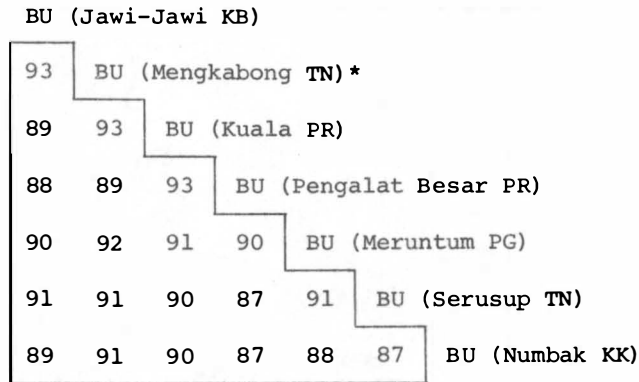


Figure 3: PSC relations between seven central West Coast Bajau villages. (* representative village of this group in Figure 4; BU = Bajau.) (from Smith, in this volume)

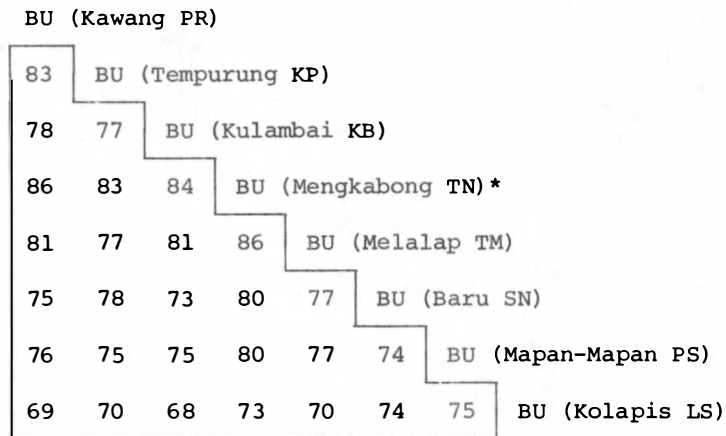


Figure 4: PSC relations between seven non-central West Coast Bajau villages and one central West Coast Bajau village. (* representative village from central West Coast Bajau, Figure 3.) (from Smith, in this volume)

The West Coast Bajau intelligibility testing described in this paper however, was based on a somewhat different lexicostatistical classification made independently from the above classification, though having basic similarities.

The West Coast Bajau language as defined in this paper is composed of the Central West Coast Bajau dialect and two other dialects. Central West Coast Bajau is here represented by 11 villages whose wordlists are interrelated by at least 85 PSC. Eight of these are villages whose wordlists are strongly related by 85 PSC-plus with at least seven other wordlists of the same group. The other three are villages whose wordlists are 85 PSC or more with at least one of the eight more closely related lists (Figure 5).

Two other villages, Baru SN and Mapan-Mapan PS, represent West Coast Bajau dialects whose wordlists are 80-81 PSC with at least three Central West Coast Bajau wordlists.

BU (Kawang PR)										
88	BU (Pengalat Besar PR)									
88	93	BU (Kuala PR)								
86	87	90	BU (Numbak KK)							
85	90	91	88	BU (Maruntum PG)						
86	89	93	91	92	BU (Mengkabong TN)					
84	87	90	87	91	91	BU (Serusup TN)				
86	88	89	89	90	93	91	BU (Jawi-Jawi KB)			
81	82	84	84	83	86	84	85	BU* (Melalap TM)		
83	85	86	81	83	83	81	81	77	BU* (Tempurung KP)	
78	79	83	80	82	84	84	86	81	77	BU* (Kulambai KB)

Figure 5: PSC relations between the 11 central West Coast Bajau villages. (BU = eight more closely related villages; BU* = three more distantly related villages.)

The Baru West Coast Bajau dialect is here represented by only one wordlist from Baru SN. Its relationship with other West Coast Bajau dialects is in the range of 73-81 PSC.

The Mapan-Mapan West Coast Bajau dialect is here represented by only one wordlist from Mapan-Mapan PS. Its relationship with other West Coast Bajau dialects is in the range of 74-80 PSC.

There is one other wordlist from Kolapis LS whose closest relationship to West Coast Bajau is 75 PSC with Mapan-Mapan PS. Its relationship to West Coast Bajau is otherwise in the range of 68-74 PSC. This relationship is too low to be considered a dialect of West Coast Bajau (Figure 6).

Kolapis LS	Mapan-Mapan PS (Mapan-Mapan West Coast Bajau dialect)	Baru SN (Baru West Coast Bajau dialect)		
75				
74	74			
72	80	81	Serusup TN	Central West Coast Bajau dialect
73	80	80	Mengkabong TN	
73	80	79	Jawi-Jawi KB	
74	79	81	Maruntum PG	
73	79	80	Kuala PR	
72	77	80	Pengalat Besar PR	
72	79	79	Numbak KK	
69	76	75	Kawang PR	
70	75	78	Tempurung KP	
70	77	77	Melalap TM	
68	75	73	Kulambai KB	

Figure 6: PSC relations between central West Coast Bajau, two other West Coast Bajau dialects, and Kolapis LS

2. TESTING PROCEDURES

The purpose of the West Coast Bajau intelligibility testing was to answer the following questions. First, do the West Coast Bajau wordlists represent only one language or more than one; and how many dialects of West Coast Bajau are there? Secondly, how well can the West Coast Bajau understand the East Coast Bajau language? And thirdly, what is the West Coast Bajau intelligibility level of Bahasa Malaysia, the national language?

In order to determine how many languages and dialects are represented by the West Coast Bajau wordlists, tape-recorded stories from the three villages of Kawang PR, Kulambai KB and Mapan-Mapan PS were selected as reference tapes for the test set. These particular stories were chosen on the basis of the relationship of the Bajau language as spoken in these villages to that spoken in the other West Coast Bajau villages; on the suitability of the content of the stories and length adequate for forming questions; and on the quality of the recorded tapes.

The story from Kawang PR was chosen to represent the eight more closely related central West Coast Bajau villages, the story from Kulambai KB to represent the three more distantly related villages, and the story from Mapan-Mapan PS to represent that proposed dialect of West Coast Bajau. These stories were also selected because they gave as wide a geographical representation as possible of the West Coast Bajau language area.

The Bajau Semporna story from Semporna SA was chosen to represent the East Coast Bajau language. This particular story was selected because it was representative of the East Coast Bajau language also known as 'Sama Kubang' or 'Bajau Semporna' found to be indigenous to Sabah (→ East Coast Bajau).

The story representing standard Bahasa Malaysia as spoken in Sabah was included in the test set to determine the West Coast Bajau people's understanding of the national language.

The villages where West Coast Bajau testing was done were chosen on the basis of their relationship to the proposed centre of the West Coast Bajau language, their historical background and their geographical location. A test point was chosen in each of the major West Coast Bajau language areas in the Papar, Tuaran, Kota Belud and Pitas districts.

Seven of the eight closely related West Coast Bajau wordlists have an average relationship of 88-91 PSC. Serusup TN was chosen to represent this group and to determine its relationship to more distantly related villages.

The one other of the eight villages, Kawang PR, has a lower average relationship to the others (86 PSC). It was chosen as a test point to determine how strong its relationship to the other seven villages was, as well as its relationship to the three more distantly related central West Coast Bajau villages and the Baru SN and Mapan-Mapan PS dialects.

Kulambai KB, having an average relationship of 83 PSC to the eight wordlists, was chosen as a test point to determine whether its relationship was sufficiently strong to establish its proposed position as part of the central West Coast Bajau dialect.

Mapan-Mapan PS, distantly removed from the other major West Coast Bajau language areas and having an average relationship of only 79 PSC to the eight closely-related wordlists, was chosen to determine whether it represented a distinct dialect.

Tempurung KP was not included in the test set because it represented a small community of people who originally came from Papar district. Melalap TM was not included because it represented an immigrant group of people originating from various other West Coast Bajau language areas. Baru SN was not included because it represented a mixed language and had been in its present location for only 15 years. Kolapis LS was excluded on the basis of its low relationship of 68-75 PSC with the 13 West Coast Bajau wordlists and because it also is a mixed language community.

3. TEST RESULTS

3.1 Mutual intelligibility of West Coast Bajau villages

Figure 7 displays the intelligibility testing results of the testing done at West Coast Bajau villages.

REFERENCE TAPES TEST POINTS	Kawang PR (central West Coast Bajau)	Kulambai KB (central West Coast Bajau)	Mapan-Mapan PS (Mapan-Mapan West Coast Bajau)
Serusup TN	100 (84)	94 (84)	89 (80)
Kawang PR	-	87 (78)	84 (76)
Kulambai KB	100 (78)	-	65 (75)
Mapan-Mapan PS	79 (76)	74 (75)	-

Figure 7: Intelligibility testing results in West Coast Bajau villages tested with West Coast Bajau reference tapes. (Results are given as percentages; PSC figures are in parentheses.)

It is interesting to note that of the villages tested, the village of Serusup TN, which is one of the eight more closely-related villages (→ Section 1) had the highest intelligibility of both the village of Kulambai KB, which is one of the three more distantly-related villages, and the proposed dialect of Mapan-Mapan PS. Serusup TN scored 94% on the Kulambai KB tape and 89% on the Mapan-Mapan PS tape.

The mutual intelligibility of the village of Serusup TN and the village of Kawang PR was 100% thus establishing the strong relationship of Kawang PR to the centre of the West Coast Bajau language.

The intelligibility at Kawang PR of the Kulambai KB story was 87%, just barely placing it within central West Coast Bajau. However the 100% scored by Kulambai KB on the Kawang PR test tape coupled with the 94% scored by Serusup TN on the Kulambai KB test tape firmly establishes that Kulambai KB is a member of central West Coast Bajau.

Although Kawang PR scored 84% intelligibility of the Mapan-Mapan PS test tape, and Serusup TN scored 89% intelligibility of the same tape, Mapan-Mapan PS had only a 79% intelligibility of Kawang PR, confirming that Mapan-Mapan PS is a separate dialect of West Coast Bajau.

Kulambai KB and Mapan-Mapan PS are more distantly related to each other as is shown in Figure 7 above.

3.2 West Coast Bajau intelligibility of East Coast Bajau

The Bajau Semporna story from Semporna SA, representing the East Coast Bajau language for purposes of this test, was used to determine the West Coast Bajau understanding of East Coast Bajau. The East Coast Bajau test tape was used for testing at all four West Coast Bajau test points. Figure 8 shows the results of that testing. Figure 9 shows more clearly how the test results relate to the corresponding PSC relations.

REFERENCE TAPE	Semporna SA (East Coast Bajau)
TEST POINTS	
Serusup TN	35 (65)
Kawang PR	64 (63)
Kulambai KB	64 (61)
Mapan-Mapan PS	82 (63)

Figure 8: West Coast Bajau intelligibility of East Coast Bajau. (Intelligibility scores are given as percentages; PSC relations are in parentheses.)

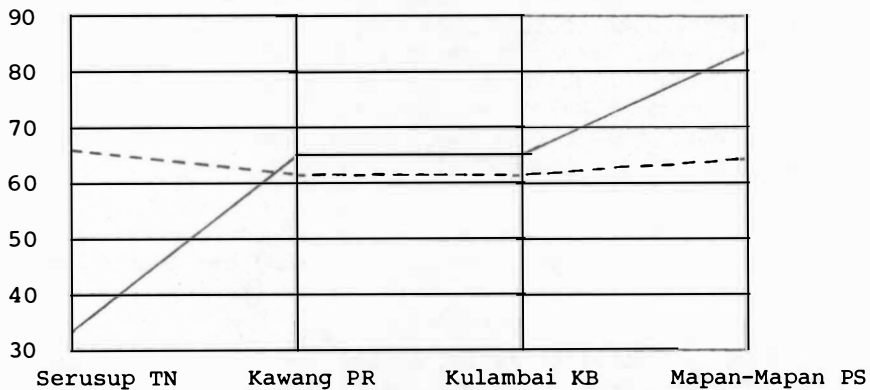


Figure 9: Comparison of intelligibility testing results and PSC relations shown when four West Coast Bajau villages were tested for their understanding of East Coast Bajau. (——— = intelligibility; - - - - - = PSC)

The subjects tested in Kawang PR and Kulambai KB showed a level of intelligibility of Bajau Semporna that was very close to their PSC relationship with that language. The intelligibility shown by Mapan-Mapan PS of Bajau Semporna was 19% higher than its PSC relationship, whereas Serusup TN dipped 30% below its PSC relationship. Mapan-Mapan PS would normally have much more contact with East Coast Bajau speakers because of its location near them, and was therefore expected to have a higher level of understanding of their language. The opposite was true of Serusup TN and was reflected in their low level of understanding of East Coast Bajau.

Intelligibility testing confirmed that West Coast Bajau and East Coast Bajau are indeed separate languages.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Figure 10 shows the results of national language intelligibility testing carried out in the four West Coast Bajau villages of Serusup TN, Kawang PR, Kulambai KB, and Mapan-Mapan PS. Forty subjects were tested, ten in each village. Their intelligibility of Bahasa Malaysia ranged from 25-100%, the average being 89%. This is an admittedly high level of understanding especially considering that only 23 of the 40 subjects tested had any formal education.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Serusup TN	100	7/3	26	15-45	10	7	7
Kawang PR	100	7/3	36	17-69	6	10	6
Kulambai KB	85	5/5	45	15-63	4	6	2
Mapan-Mapan PS	73	4/6	34	15-57	3	7	2

Figure 10: Comprehension of the national language in four West Coast Bajau villages, with sociological data. (Test scores are given as percentages. Under EDUCATION, ED. = the number of subjects in each corpus who had received some formal education, A.E.R. = the average number of years of education those subjects had received, and A.E.S. = the average number of years of education per subject for the corpus as a whole.)

The education of 21 of the subjects tested ranged from Primary 3 to graduation from Gaya College. The other two educated subjects tested had had several years of Adult Education. The average education of these 23 was eight years or Form 2 level.

This high level of education reflects the fact that most of the West Coast Bajau live near well-developed education centres and because of that, their intelligibility of Bahasa Malaysia is high. For example, all of the subjects tested in Serusup TN who were under 40 years of age (five men and three women) had a Primary 6 or higher level of education, and they all scored 100% on the Bahasa Malaysia test.

The Bahasa Malaysia intelligibility among the 23 subjects with formal education ranged from 70-100%.

As would normally be expected, the average intelligibility of Bahasa Malaysia among those with no formal education was lower than among those with formal education, the relationship being 81% and 96% respectively.

The sex of the subjects tested was found to be an influencing factor in the level of understanding of Bahasa Malaysia among the group with no formal education. The ratio of men to women tested was almost equal - 53% to 47%. The men's understanding of Bahasa Malaysia ranged from 70-100% with an average intelligibility of 93%, whereas the women's understanding ranged from 25-100% with an average intelligibility of 67%. The age of the subjects tested was found to have some influence on the level of understanding of Bahasa Malaysia, particularly among the older women (Figure 11).

Age	Average Intelligibility of Men	Average Intelligibility of Women
24-34	100% (1)	70% (2)
35-49	95% (2)	80% (3)
50-	91.7% (6)	51.7% (3)

Figure 11: Uneducated subjects' understanding of the national language. (Numbers in parentheses indicate the number of subjects in each category.)

It is interesting to note that only eight of the subjects tested said they used some Bahasa Malaysia in their homes; one said he used only Bahasa Malaysia in his home. So although most of the West Coast Bajau who were tested were quite proficient in the use of the national language, most of them use their mother tongue in communicating with each other.

5. CONCLUSION

The West Coast Bajau language of Sabah is a distinct language within the Bajau language family. There appear to be three dialects of West Coast Bajau; one large central dialect spoken in the districts along the west coast of Sabah, one dialect spoken in the Pitas District (Mapan-Mapan PS) and a third spoken in the village of Baru SN. However intelligibility testing in Baru SN would be necessary in order to confirm this classification. Intelligibility testing also confirmed that West Coast Bajau is a distinct language from East Coast Bajau. PSC figures suggest that Kolapis LS is not a dialect of West Coast Bajau, but a separate language. Intelligibility testing would also need to be done to prove this.

And finally, intelligibility testing showed that the West Coast Bajau have a high understanding of Bahasa Malaysia, the national language.

NOTES

1. The term West Coast Bajau is not a strictly geographical designation. Some settlements of 'West Coast' Bajau are located on Sabah's east coast.
2. This is a rough estimate based on the 1970 population and housing census of Malaysia and on information collected by members of the Summer Institute of Linguistics during 1978-81 while doing a language survey of Sabah.

THE EAST COAST BAJAU LANGUAGES

Janice Walton and David C. Moody

0. INTRODUCTION

The origins of the Bajau-speaking peoples are not clearly known. In his master's thesis Warren (1971) offers merely that long ago a land-dwelling people, probably after prolonged contact with the sea, abandoned their former orientation to adopt a boat-dwelling lifestyle as sea nomads. The Brunei chronicles and Sulu tarsilas (genealogical records) share the common tradition that the Bajau emigrated from Johore on the Malay Peninsula, an event one investigator has dated early in the 14th century.¹ This relatively late dating for the arrival of the Bajau coincides also with the belief held commonly throughout the Sulu Archipelago that the Bajau arrived in the Sulu Sea subsequent to the Tausug (or Suluk as they are known in Sabah) people.

More recently this view has been challenged. Tom Harrisson (1973-74) and S.J. O'Connor have suggested the possibility that Chinese trade in the Sarawak River delta between 700 AD and 1350 AD may have been carried in Bajau vessels. Their conclusions that Bajau influence in the area of North Borneo may be much earlier than traditional accounts is based on archaeological evidence.²

Support for an earlier migration also comes from Pallesen (1977) whose historico-comparative analysis of Bajau and Tausug languages provides linguistic evidence that Bajau-speaking peoples had become well established in the southern Zamboanga-Basilan area of the Sulu Archipelago by 800 AD, predating Tausug habitation of the area. He hypothesises further that there was an outward expansion from this area toward the south-west, such that as early as the start of the 12th century subgroups could have been settled in Cagayan Sulu and in both Indonesian and North Borneo, areas in which they remain to this day.

The linguistic literature based on study of the East Coast Bajau dialects of Sabah is quite limited. Besides Schneeberger's (1937) short vocabulary only Abdul Ghani bin Bagul (1950) and Sather (1965, 1968) have published linguistic articles dealing with Sabah Bajau. Only Sather's articles are concerned with East Coast Bajau, particularly Bajau Laut as spoken at Bangau-Bangau SA. His earlier article deals specifically with numbers and adjectives of quantity, while the latter sketches elements of Bajau phonology and grammar. Linguistic descriptions based on Bajau dialects spoken in the Philippines include Pallesen's (1965) phonological description of Central Sama and Allison's (1977) discourse study of a Southern Samal text. Pallesen (1977) includes a reconstruction of the Bajau parent language which he uses to demonstrate socio-historical relationships between the Bajau and Tausug (Suluk) groups.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 113-123. *Pacific Linguistics*, C-78, 1984.

© Janice Walton and David C. Moody

113

The term Bajau (spelled Bajaw in the Philippines) is cited (Evans 1952) as an exonym applied originally by the Brunei Malays. Most speakers (particularly those in the Philippines) refer to themselves as Sama (also Samah, Samar, and Samal) and to their language as 'Bahasa Sama' or 'Bahasa Bajau' in Indonesia and Sabah and 'Sinama' in the Philippines. Other designations include Palaqu, Luwaqan, Sea Gypsies and Orang Laut (*sea people*).³ The term Bajau has gained wide acceptance in Sabah, and is used in this paper as a cover term for all speakers of Bajau or Sama dialects.

The population of the Bajau in Sabah was listed in the 1970 census as 72,563, a figure which included both the East Coast and West Coast Bajau communities. More recent estimates based on the 1980 census suggest there may be that many East Coast Bajau alone currently living in Sabah.

The primary areas included in the survey were the districts of Lahad Datu, Semporna and Sandakan (Figure 1). In Lahad Datu District there is a large Bajau population of varied origins. The dialect names indicate for the most part the location from which the people originated in the Philippines. The following dialects are found in significant numbers: Balangingi, which originated from the island of Balangingi in the Tongquil group of Northern Sulu PHL; Sikubung, which originated from the island of Sikubung near the north-eastern end of Tawi-Tawi PHL; Simunul and Sibutu, who for the most part have recently come from those places in the Philippines; and Bajau Asli, which is also known as Sama Kubang or Bajau of Semporna. In the Kunak LD area there are numerous Simunul and Sibutu people, most of whom came to the area within the last 20 years. The original inhabitants of the area are said to be the Bajau Asli (Sama Kubang) people. Within the past decade West Coast Bajau people have been resettled in several schemes in the Kunak LD area and in Sandakan District.

Semporna District apparently has the largest Bajau population of any of the areas included in the survey, although no population statistics are available at present. The original residents of the Semporna area were evidently Bajau (Sama Kubang) and Bajau is spoken by the vast majority of the people, including many who are not Bajau. In addition to the Semporna Bajau (Sama Kubang) there are several other large communities of Bajau who trace their origins to locations in the Philippines. These include people of Simunul and Sibutu who are largely located in Simunul SA and Kg Air SA; Bajau Laut, who trace their origin to Sitangkay PHL, are located in Bangau-Bangau SA; Ubian people are located in Terusan Baru SA on Bumbum Island and trace their origin to South Ubian Island in the Tawi-Tawi group.

In Sandakan District there are three Bajau groups: West Coast Bajau, which were not included in the survey; Kagayan, or Jama Mapun; and Simunul. The Kagayan population is quite extensive in the Sandakan area and has been there for a relatively long time. Pallesen (1977:171f.) suggests that the migration route of the Bajau group which eventually settled in Cagayan Sulu was via the coast of North Borneo, indicating the Kagayan people may have been living in the area for nearly 800 years. They are now found in Sibuga Besar SN, Simsina SN, Nunuyan Island and Libaran Island, and are reported to be in numerous other places along the northern coast extending as far as Jambongan Island and Kudat. The Simunul people for the most part live in Bokara SA and Kg Air SA and have been there for a long time.

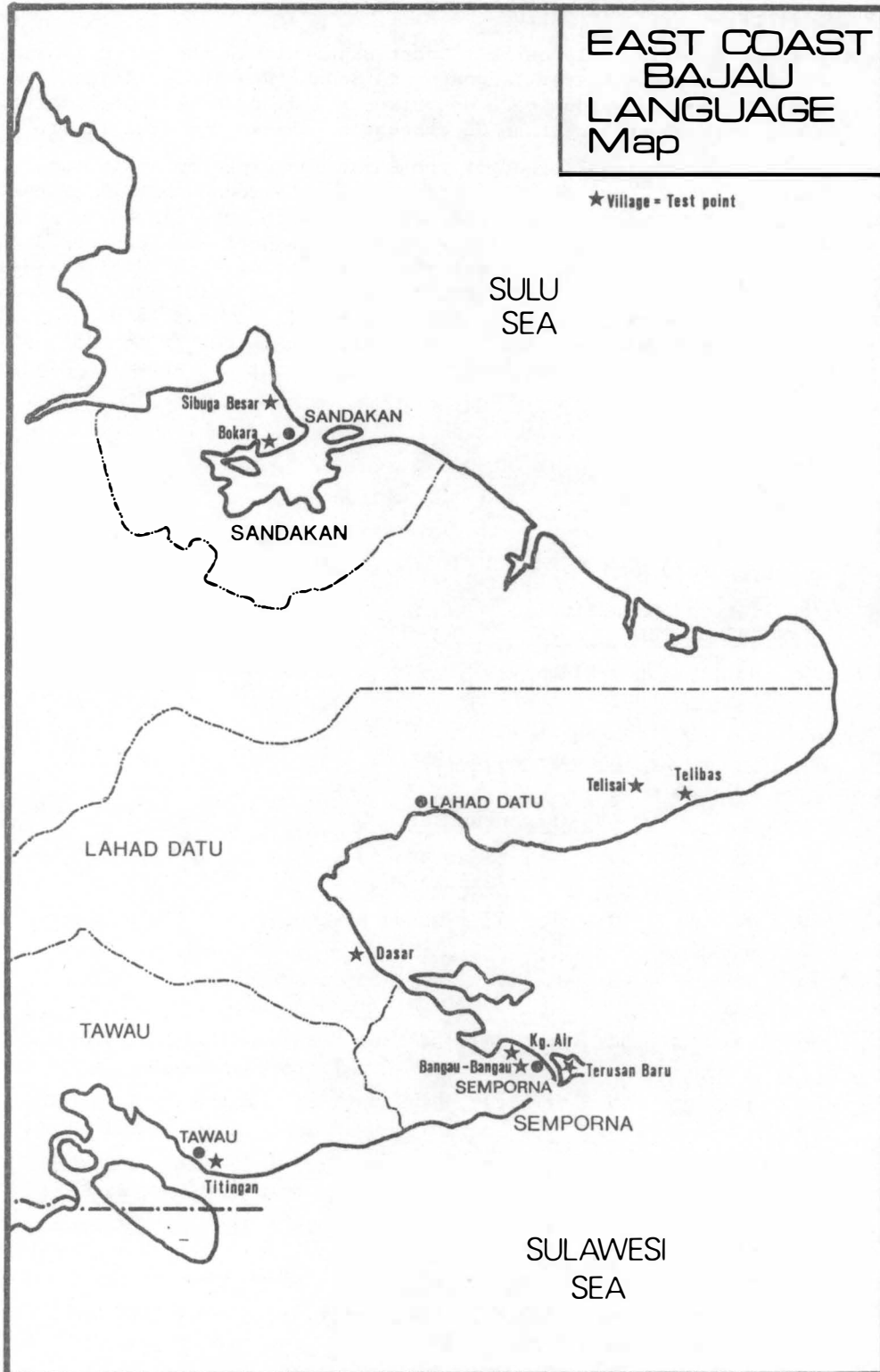


Figure 1: East Coast Bajau language map

The West Coast Bajau list is clearly distinct, relating at 57-69 PSC with the East Coast Bajau samples. Within the East Coast Bajau language group the Mapun (or Kagayan), Simunul and Balangingi lists may be said to represent separate dialects from the group representing Southern Sama, none of them relating at higher than 82 PSC with a Southern Sama list. The wordlist from Siasi PHL is designated Central Sama, based on Philippine information, though it exhibits a high (92 PSC) relationship with the Sitangkay PHL list representing Southern Sama.⁸

2. TESTING PROCEDURE

The purpose of the survey was to test the intelligibility of various Bajau dialects in the selected communities of long-term resident native speakers of Bajau languages on the eastern coast of Sabah (i.e. in Lahad Datu, Semporna and Sandakan districts). The testing was expected to indicate (1) which dialects of Bajau the Sabah communities could best understand, (2) how closely the indigenous Semporna Bajau dialect is related to Southern Sama of the Philippines, and (3) the extent of language change that has occurred since certain communities migrated from the Philippines.

A location for administering intelligibility tests was chosen on the basis of the percentage of cognates shared by the language of that location and other languages of the area and the length of time the community had been in Sabah. Communities whose speakers had not been in Sabah for more than 40 years were not chosen for testing since less than that amount of time was considered insufficient to allow for a significant amount of language change to take place.

The communities selected for intelligibility testing were: Telisai LD, Telibas LD, Bangau-Bangau SA, Terusan Baru SA, Kubang Baru SA, Bokara SN and Sibuga Besar SN.⁹

Selection of test sets was made on the basis of a knowledge of the Bajau dialects of the Philippines, the percentage of cognates shared by the various dialects in the area and with Bajau dialects of the Philippines, and a knowledge of the community to be tested. The test sets were different for each community tested since each had a different point of origin and showed a different cognate relationship to the other dialects.

3. TEST RESULTS

The results of intelligibility testing in East Coast Bajau villages are displayed in Figure 3. Average scores on the hometown reference tapes were in the range 94-98%, well within the expected norm.

3.1 Balangingi

The Balangingi scores from Telisai LD were as follows: 92% on Philippine Balangingi, 72% on Philippine Simunul, 81% on Bajau Laut, and 27% on West Coast Bajau.

REFERENCE TAPES TEST POINTS	Balangingi (Balangingi PHL)	Simunul (Simunul PHL)	Sikubung Semporna (Terusan Tengah SA)	Bajau Semporna (Kubang Baru SA)	Bajau Laut (Bangau-Bangau SA)	Central Sama/Sinama (Siasi PHL)	Jama Mapun (Cagayan Sulu PHL)	West Coast Bajau (Maruntum PG)
Balangingi (Telisai LD)	92	72			81 (72)			27 (63)
Sikubung (Telibas LD)		81	90 (84)		72 (79)	42 (76)		
Bajau Laut (Bangau-Bangau SA)		85	92 (80)	96 (86)		70 (80)		
Bajau Semporna (Kubang Baru SA)		83	92 (84)		90 (86)	46 (81)		15 (64)
Ubian (Terusan Baru SA)		83	92 (91)	95 (84)		46 (81)		
Ubian (Buah Pandai KB)		91		98 (71) *		56 (70) *	87 (60) *	
Simunul (Bokara SN)		82		89 (74)				40 (64)
Kagayan (Sibuğa Besar SN)		51					91 (83)	60 (68)
* PSC scores from Kuala Abai KB are substituted for Buah Pandai KB								

Figure 3: Results of intelligibility testing in East Coast Bajau villages.
(Intelligibility scores are shown as percentages. PSC scores are in parentheses.)

The Balangingi of Telisai LD were tested mainly to find out if the language of the community had changed sufficiently to render Balangingi of northern Sulu PHL unintelligible to them. All the subjects' parents had been born in Malaysia. Some trace their lineage in Malaysia back five and six generations, others more recently. Most of the test subjects stated that their ancestors came from Boan Island in the Philippines. However, the origin of the Boan Balangingi community is known to be from the Tongquil group of islands in northern Sulu. All test subjects said that Balangingi was the language they use in their homes.

The scores of 72% on Philippine Simunul and 81% on Bajau Laut are quite high and are probably due to their living close to Sikubung people who speak a variety of the Southern Samal dialect.

3.2 Sikubung

Sikubung scores at Telibas LD were as follows: 81% on Philippine Simunul, 90% on Sikubung Semporna, 72% on Bajau Laut and 42% on Central Sama. The Sikubung people of Telibas LD were tested to find whether their dialect was more closely related to Southern or Central Sama as the scores indicate. Four of the subjects' parents had been born in the Philippines. Most test subjects trace their origins to Sikubung Island near Tawi-Tawi Island PHL. Many present marriages are arranged with relatives on Sikubung Island PHL. All test subjects said that Sikubung was presently the language they spoke in their homes.

Sikubung is clearly a variety of Southern Sama by its score of 81% with Philippine Simunul, which probably should be adjusted upward 5% because of a poor question which required an obscure term for an answer. The score of 90% with Sikubung Semporna indicates that the language of the two communities has undergone little change during the more than 40 years of isolation from each other.

3.3 Bajau Laut

Bajau Laut was tested in Bangau-Bangau SA and the scores were as follows: 85% on Philippine Simunul, 92% on Sikubung Semporna, 96% on Bajau Semporna, and 70% on Central Sama. Seven of the test subjects' parents were born in Malaysia; three of the test subjects' parents were born in the Philippines. All test subjects trace their lineage to Sitangkay Island PHL. Many present marriages are arranged with relatives there. All test subjects said that Bajau Laut is the language presently used in their homes.

Since these people refer to themselves as Sama Dilaut, as do the people of the Central Sama dialect in Siasi, Sulu PHL, and because of the high percentage of cognates shared by Central Sama and Sitangkay, the ancestral home of the Bangau-Bangau SA Bajau Laut (92 PSC), it would be expected that Bangau-Bangau SA people would score high on Central Sama. This however was not the case. The high scores with Bajau Semporna and Philippine Simunul indicate that the present linguistic relationship is closer to Semporna. This is further verified by the PSC relations. Bajau Laut of Bangau-Bangau SA is clearly part of the Southern Sama dialect area.

3.4 Bajau Semporna

Bajau Semporna (Kubang) people were tested in Kubang Baru SA with the following results: 83% on Philippine Simunul, 92% on Sikubung Semporna, 90% on Bajau Laut, 46% on Central Sama, and 15% on West Coast Bajau. All the test subjects' parents had been born in the Semporna area (with the exception of one who was from Tarakan KAL), their ancestors being from Bumbang Island and other off-shore islands and the coastal area near Semporna. All test subjects said that Bajau Kubang is the language used in their homes.

The test scores clearly indicate that this dialect is a part of the Southern Samal dialect since it has high scores with Philippine Simunul, Sikubung Semporna and Bajau Laut.

3.5 Bajau Ubian

Bajau Ubian was tested in two locations. The Bajau Ubian of Terusan Baru SA were tested to determine their relationship to Samal dialects of the Philippines. The Ubian of Gaya Island KK and Buah Pandai KB were tested to determine the language relationship between these villages whose residents call themselves Ubian and the Bajau language groups living along the east coast of Sabah, as well as the Samal languages of the southern Philippines.

3.5.1 Bajau Ubian of Terusan Baru SA

The Bajau Ubian of Terusan Baru SA showed the following test results: 83% on Philippine Simunul, 92% on Sikubung Semporna, 95% on Bajau Semporna, and 46% on Central Sama. Six of the test subjects had parents who had been born in the Philippines, two had one parent born in Malaysia and one parent born in the Philippines, and two subjects said both parents had been born in Malaysia. Their ancestors came from South Ubian Island PHL between 40 and 100 years ago. All test subjects said that Ubian is the language used in their homes.

The scores indicate that the Ubian dialect is clearly a part of the Southern Samal dialects as evidenced by the high scores with Philippine Simunul, Sikubung Semporna and Bajau Semporna. There has been little linguistic change since these speakers left South Ubian Island PHL.

3.5.2 West Coast Ubian

It was difficult to find Ubian people to test on Gaya Island KK who had been born there or whose parents had been born there. Of the five people tested there was only one who along with his parents had been born on Gaya Island KK. Most others had been born on Banggi Island KT and their parents had been born elsewhere. With the exception of the score on the Central Sama story (14%) the scores are quite high (96-100%), but because of their short term of residence in the area the scores were not considered valid.

The Ubian people of Buah Pandai KB scored as follows: 94% on Ubian of Terusan Baru SA, 98% on Bajau Semporna, 87% on Jama Mapun of the Philippines, 91% on Simunul of the Philippines and 56% on Central Sama. Most of the people tested were born on Mantanani Island KB, the recognised home of the Ubian people in Kota Belud District.¹⁰ There is regular interaction between the people living in Buah Pandai KB and those living on Mantanani Island KB with intermarriage.

The high test scores reveal high intelligibility of Ubian of Terusan Baru SA as well as the Southern Samal dialects represented by Bajau Semporna (98%) and Philippine Simunul (91%). The relationship with Central Sama is clearly more distant (56%). The intelligibility of Jama Mapun of the Philippines was very high, as would be expected from their close relationship with the Cagayan people living in Kuala Abai KB, their market and school town.

3.6 Simunul

The Simunul test subjects of Bokara SN showed the following results: 82% on Simunul of the Philippines, 89% on Bajau Semporna and 40% on West Coast Bajau.

Nine of the test subjects' parents had been born in Malaysia; one had one parent born in Malaysia and one parent born on Simunul Island PHL. Ancestors of four to five generations ago migrated to Sandakan from Simunul Island PHL. Their language is mixed now with Malay, and there is intermarriage with people of other language groups.

The fact that Bokara SN people only scored 82% on a story from their ancestral home in the Philippines is probably due to the fact that one question required an archaic or little known word for an answer and so may have deflated the scores by 5%. The lower score may also indicate language change is affecting intelligibility. Even though the Bokara SN community has been isolated from the rest of the Southern Samal dialects, it has not undergone sufficient change to affect intelligibility greatly. This is evidenced by the 89% score with Bajau Semporna even though the two communities are separated by a great distance and there is little travel to the Semporna area.

3.7 Kagayan (Mapun)

Kagayan (Mapun) people of Sibuga Besar SN were tested with the following results: 51% on Simunul of the Philippines, 91% on Jama Mapun of the Philippines and 60% on West Coast Bajau. For three of the test subjects, both parents had been born in Malaysia; four had one parent who was born in Malaysia and one in Cagayan Sulu PHL. One had a parent born in Indonesia and the birthplace of the other parent was unknown. Two subjects said both parents were born in Cagayan Sulu PHL. Some test subjects said their ancestors had migrated to the Sandakan area three or more generations ago, having come from Cagayan Sulu PHL. All subjects said that Kagayan is the language used in their homes.

The test scores show clearly the distinction between the Southern Samal and Kagayan dialects of Bajau. It is interesting to note that the Kagayan dialect has not changed sufficiently during the years of separation from its ancestral home to affect intelligibility. Also, Kagayan speakers scored higher than the East Coast Bajau groups on the West Coast Bajau story, even though this affinity is reflected only slightly (4-5 percentage points) in the cognate percentages.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Figure 4 displays the results of testing to determine the degree of understanding of a story told in Bahasa Malaysia in the eight East Coast Bajau communities. Taken district by district, the Kota Belud and Sandakan communities demonstrate the highest average test scores. The Sandakan scores correlate with higher educational statistics. The Kota Belud scores point to a high degree of language learning on the part of a minority language community who must rely on the national language for communication with other groups. For all Bajau villages, the average test score was 80%. Over half of the test subjects had received some formal education. The average education for the combined eight corpora was nearly four years.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Telisai LD	85	6/4	30	19-50	5	6	3
Telibas LD	65	8/2	32	18-55	3	5	2
Bangau-Bangau SA	71	7/2	26	14-42	4	5	2
Kubang Baru SA	75	7/3	26*	15-45*	7*	6*	4*
Terusan Baru SA	76	6/4	29	17-50	7	8	5
Buah Pandai KB	96	8/2	23	15-30	6	5	3
Bokara SN	88	5/5	31	15-64	7	9	6
Sibuga Besar SN	85	7/3	27	14-49	6	8	5
* Age and education statistics were not obtained from one subject							

Figure 4: Comprehension of the national language in eight East Coast Bajau villages with sociological data. (Test scores are given as percentages. EDUCATION figures indicate the number of subjects having received some formal education (ED.), the average number of years of education those subjects had received (A.E.R.), and the average number of years of education for all subjects in the corpus (A.E.S.).)

5. CONCLUSION

Intelligibility testing in eight Bajau communities has confirmed Smith's West Coast-East Coast distinction. Moreover, the identification of most of the Sabah samples with the Southern Sama language can be firmly established. Language level distinctions can also be posited for the Balangingi and Kagayan (including Mapun) dialects. The testing further demonstrated that Sabah varieties of East Coast Bajau are distinct from the Central Sama language of the Philippines. Finally, within the Southern Sama varieties of East Coast Bajau, a Simunul dialect is apparent and is showing definite change from the language spoken in the ancestral homeland.

NOTES

1. The investigator (cited by Warren) was Najeeb M. Saleeby, who during his service as a medical doctor in the American administration in Sulu during the years 1901-1935, translated the tarsilas into English.
2. The evidence presented by Harrisson and O'Connor includes the absence of Chinese artifacts other than their ceramics and coins, as well as the abundance of Bajau-type pottery (Harrisson 1973-74:39-40).
3. Harrisson (1976) argues for a close relationship between the Bajau and the Ngaju of Kalimantan. He lists the following forms, appearing in literature as far back as the 17th century and covering territory from the South China Sea to Australia, as pertinent to his discussion: Bajau, Bajou, Baju, Badjau, Badjaw, Badjoo, Baijini, Byajo, Biajo, Beaju, Beadje, Viadje, Ngadju, Ngaju and Lutao.
4. A language survey of the Bajau languages of the southern Philippines conducted in 1970-71, using lexicostatistics and intelligibility testing identified seven Bajau dialects. The dialects were distinguished on the basis of mutual intelligibility scores of 80% or lower and PSC relationships of less than 70. The seven dialects are as follows:
 - Abaknon, spoken on Capul Island, Northern Samar;
 - Yakan, spoken in Basilan, Sakul Island and Bitali, Zamboanga del Sur;
 - Balangingi (Bangingi), spoken in the Tongquil group, along the coast of Zamboanga and neighbouring islands;
 - Central Sama (Sama Dilaut, Sinama), spoken on the islands in the Siasi area and extending to North Ubian Island;
 - Southern Sama, spoken in the Tawi-Tawi group, extending from Ungus Matata to Sibutu and Sitangkay with the communication centre in Simunul Island;
 - Pangutaran Sama, located on the Pangutaran group with resettled communities in Cagayan Sulu and the Brookes Point area of Palawan;
 - Mapun (Kagayan), spoken in Cagayan Sulu and Southern Palawan.
5. The self-designations represented within this central group were: Sikubung (2 wordlists), Ubian (3), Sibutu (2), Sama (4) and Bajau Banaran, Bajau Darat, Bajau Laut and Bajau Semporna (1 list each).
6. The one exception is the list from Siasi PHL, which fits better with the Central Sama dialect. Its relation of 92 PSC with the Sama (Bajau Sitangkay) wordlist seems aberrantly high.
7. The self-designations given for these ten wordlists were: Ubian (3 lists - all on Sabah's west coast), Simunul (2), Kagayan (2) and Jama Mapun, Balangingi and Laminusa (1 list each).
8. The relationship of 92 PSC between the Siasi PHL and Sitangkay PHL wordlists is established on the basis of 168 comparable items, with over 199 non-comparable items occurring on the two lists. The relationships with the other Southern Sama lists (80-85 PSC) are more in line with expectations.
9. Testing was not done in Tawau District since data collected during the initial phase of the survey indicated that those communities had not been in the area for a sufficient length of time to warrant testing.
10. This was indicated by district officers, village leaders, etc.

THE TIDONG LANGUAGE

David C. Moody

0. INTRODUCTION

Tidong speakers claim they have always lived in Sabah. Indeed, though some Sabah Tidong are recent migrants from Indonesia, others can establish a long history of settlement in the state. One village leader asserted the residents of his village had been in Sabah for 100 years following their immigration from Indonesia. Clifford Sather (1972:168) cites Spanish records which indicate Tidong involvement in raids against Spanish settlements and shipping in the early 17th century, though it is not said whether these activities occurred in the territory of present-day Sabah. He also refers to Tidong traditions of a politically autonomous Tidong "state" with a fortified capital on Sebatik Island.

Sather (1972:167) lists the areas of Tidong habitation as the east coast of Borneo from the mouth of the Bolongan River north to the coastal area surrounding Cowie Harbour in the Tawau district, on the Labuk River opposite the town of Klagan, and along the Sembakung and Sibuku Rivers of East Kalimantan. Members of the Summer Institute of Linguistics (SIL) report additional communities in the districts of Labuk-Sugut and Sandakan in Sabah.

The Tidong population of Sabah was listed as 7,755 in the 1970 census. Based on an arbitrary figure for annual population growth of two percent, it is estimated that the population has increased to 9,500-10,000 at present. The Tidong communities visited during the present study are shown in Figure 1.

The Tidong language was first described by M.W.H. Beech in 1908, though a wordlist was published by W. Aernout in 1885. Prentice (1970:378) cited other wordlists by Genderen Stort (1916), and O. Rutter (1929, Tenggara). Prentice himself compared Tidong wordlists along with lists of other Bornean languages and classified them according to their lexical relationships.

Tidong, which means *hill* or *hill people* in the Tarakan dialect, is also referred to in the literature as Tidong, Tedong, Zedong, Tidoeng, Tiroon, Tirones, Tiran and Camucones, the latter term from early Spanish records (Sather 1972:167).

Terms which have been used in the literature to designate dialects of Tidong are shown in Figure 2. Tidong speakers in Indarasabak TU listed four dialects of their language, and noted that speakers of the various dialects had difficulty understanding each other. The dialects they listed are: Tarakan (their own), Sembakung, Bolongan and Sesayap. The first three designations

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 125-137. *Pacific Linguistics*, C-78, 1984.

© David C. Moody

125



Figure 1: Tidong language map

parallel those made by Beech in 1908; the Indarasabak TU listing mentions Sesayap (another spelling is Sesajap) while omitting Nonukan. Sesayap was also given as a dialect designation in Rancangan Blok No.31 TU, as was also the designation Tidong Proper ('Tidong tulin').

Beech (1908) "dialects"	Genderen Stort (1916) "dialects"	Prentice (1970) "languages"
Tarakan	Tarakan	Tarakan - located along Borneo's east coast from Tarakan Island to Cowie Harbour
Sembakung (other spellings: Sembakong, Sembakoeng)	Sembakung - located along the middle reaches of the Sembakung River and its tributaries	Tinggalan (with distribution similar to Sembakung)
Nonukan (located on or around Nonukan Island-?) (other spellings: Noenoekan, Nunukan)	Penchangan - cf. Pensiangan, in the Murutic language family (→ Murutic)	Tanggaraq - Rutter identifies this with Murut (→ Murutic) (other spellings: Tenggaraq, Tenggarah, Tangara', Tengara, Tingara)
Bolongan (located along the Bolongan River/ Island-?) - Prentice says this is neither Tidong nor Murutic	Sedalir - cf. Salalir, in the Murutic language family (→ Murutic) (other spellings: Selalir, Saralir)	
	Tidong proper - located along the upper reaches of the Sembakung River - limited data available	

Figure 2: Tidong dialect/language designations proposed by three authors. (There seems to be agreement for those designations given above the broken line and lack of agreement for those listed below the broken line. (Adapted from material in Sather, Prentice, and Dunn.))

1. LEXICOSTATISTICAL CLASSIFICATION

During the first phase of the survey wordlists were elicited and personal experience stories were tape-recorded from Tidong speakers in three different villages, Lidung LS, Tanjung Aru SN, and Rancangan Blok No.31 TU. The percentage of shared cognate (PSC) relations for these three lists are shown in Figure 3.

Tanjung Aru SN		
76	Lidung LS	
67	74	Rancangan Blok No.31 TU

Figure 3: PSC relations between three Tidong villages

On the basis of these scores Smith wrote, "In terms of a language defined by a dialect chain, these three dialects are marginally considered a single language; mutual intelligibility is doubtful and must be tested." Smith classified the Tidong language as one of four family-level constituents of the Bornean division of the North-western Austronesian superstock (Smith, in this volume).

Prentice (1970:375) placed Tidong within his Murutic subfamily, referring to it as a "language group", which he defined as "a group of languages within a subfamily which show closer relationships with each other than with other languages of the same subfamily." His Tidong language group included three languages: Tarakan, Tinggalan and Tanggaraq (Figure 2). The first two language designations parallel dialect designations made by Beech and Genderen Stort, Tinggalan being tantamount to Sembakung (Sather 1972:167). The Tanggaraq (Tenggara) were described by Rutter (1929:35) as a Murutic group who lived in the area of the upper Kinabatangan and Kuamut Rivers. Data from this present survey confirms Rutter's classification. Tenggara is a term still in use to designate inhabitants of Inarad KN and Kokoroton KN, villages located in the upper Kinabatangan region. Spitzack (→ Murutic) has classified their dialect as part of the Baukan language, within the Murutic family. Prentice further suggested that the people living along the Kalabakan and Serudung Rivers, both of whom Rutter had designated as Tawau Muruts, would probably fit within his Tidong group of languages.

In comparing all 344 wordlists collected initially in SIL's statewide survey, only two lists showed a PSC relation higher than 60 with a Tidong village. Kampung Labuk KAL, located on the Sembakung River in Indonesia (the wordlist was actually elicited from a speaker in Serudung Baru TU), and Kalabakan TU manifested relations of 67 and 66 PSC respectively with the Tidong village of Rancangan Blok No.31 TU. A group of 17 wordlists showed relations in the range of 55-60 PSC with a Tidong wordlist. Included in this group were the lists from Serudung TU¹ and Inarad KN. These relations are displayed in Figure 4.

Figure 4 helps to clarify the relationship between the Tidong and Murutic languages. The three Tidong villages from which wordlists had been obtained are listed above the bold horizontal line. Murut villages are listed to the right of the bold vertical line. Each of the top four Murut villages is representative of a Murutic language which has been suggested as a possible candidate for inclusion within the Tidong classification. "Other Murut" designates seven Murut villages which are representative of remaining Murutic languages.²

TIDONG (Tanjung Aru SN)							
76	TIDONG (Lidung LS)						
67	74	TIDONG (Rancangan Blok No.31 TU)					
55	62	67	MURUT (Labuk, Sembakung R. KAL)				
54	62	66	72	MURUT (Kalabakan TU)			
51	58	59	69	77	MURUT (Inarad KN)		
48	55	58	62	70	63	MURUT (Serudung TU)	
49	54	55	65	66	72	56	OTHER MURUT
47-51	52-57	53-57	57-71	59-69	63-79	52-58	

Figure 4: PSC relations among Tidong and representative Murut villages.

(Murut representatives named by village have been suggested as possible candidates for inclusion within the Tidong classification. "Other Murut" designates seven villages which are representative of remaining Murutic languages, for which the range of PSC relations between those villages and the village designated by the column is given just below the average of those PSC relations.)

Smith (in this volume) establishes the range of 45-60 PSC as that which separates language families. The lexical relations between the three Tidong villages and the 11 representative Murut villages are in the range 47-67 PSC. Only two Murut villages show a closer lexical relationship to Tidong villages than expected for languages representing different language families. These are Labuk KAL representing Sembakung Murut, and Kalabakan TU representing Kalabakan Murut. Labuk KAL relates to the three Tidong villages with an average of 61 PSC and to the ten other Murut representatives with an average of 66 PSC. Likewise, while Kalabakan TU relates to the three Tidong villages with an average of 61 PSC, it relates to the other Murut villages with an average of 68 PSC. On the other hand, Rancangan Blok No.31 TU, which shows the highest lexical relationships with Labuk KAL and Kalabakan TU at 67 PSC and 66 PSC respectively, averages over 71 PSC with the other Tidong wordlists. Thus, while some representatives of the Tidong language and the Murutic language family demonstrate closer inter-family relations than do the remaining representatives of their respective language families (which is natural and expected generally of inter-family relations), Smith's classification places each representative in the family with which it has most in common.³

Both Smith and Prentice observed the marginal nature of the Tidong dialects/languages based on lexical relationships. Smith (in this volume) preferred to

call them dialects of one language even though he could have established each as a separate language. He suggested that the Tidong "dialects" were related to each other like links in a chain doubting that there would be mutual intelligibility between them, particularly between the more distant dialects. Prentice, on the other hand, chose to designate each dialect as a separate language which then together formed a group in which the members were more closely knit to each other than to languages outside the group but within the same subfamily.

It should be noted, however, that the Tidong data presented here may actually represent only one of the languages referred to by Prentice. SIL researchers were told there are both Tidong and Murut living along the Sembakung River, in which case the data may not include Tidong from that region; the person giving the Sembakung list called himself Murut. Also, it is not known to which language Prentice would assign the Tidong groups living in the Labuk-Sugut and Sandakan districts. Conclusions from the present study may therefore neither substantiate nor repudiate his proposed intra-Tidong classification.

The purpose of dialect intelligibility testing in this study initially was to clarify the relationships among the Tidong dialects in Sabah represented by the villages from which data had been collected. Do people from one Tidong village understand the speech of Tidong speakers from another village enough to establish that they speak the same language, or does the degree of mutual intelligibility indicate that they speak different languages? It was also decided to test the ability of Tidong speakers in one village to understand speech samples of two Murutic languages.

2. TESTING PROCEDURE

Each Tidong village from which a wordlist had been obtained was chosen as a test point in order to determine whether mutual intelligibility existed among them. The stories collected in each of the three villages were chosen as reference tapes. For the testing in Tawau, the story from the Murut village of Inarad KN and the story from Serudung TU were selected to represent Murutic languages.

All of the stories were judged good with respect to quality of reproduction. The Tanjung Aru SN story was simple and straightforward in its content, but sprinkled generously with Malay loan words. The Lidung LS and Rancangan Blok No.31 TU stories were good, being less predictable in their content and the Lidung LS story particularly having fewer Malay loan words. Both the Kalabakan TU and Serudung TU stories were short and simple; the Serudung TU story was selected as being less predictable. The Inarad KN story was also judged to be good in terms of its content.

Two test points had to be changed in the process of making arrangements for the actual testing. The researchers who went to Tanjung Aru SN were re-directed to Rancangan Sungai Manila SN. In Tawau, the research team was re-directed to Indarasabak TU as representing a more homogeneous Tidong village than Rancangan Blok No.31 TU.

The test corpus in Rancangan Sungai Manila SN comprised only eight persons, of whom only one was female. This deviation from the ideal could be expected to produce higher scores. The group tested in Lidung LS was younger than average, five of the ten subjects ranging in age from 15-18 years, which could be expected to lower scores. Conversely, the corpus gathered in Indarasabak TU was older, averaging 42 years of age per subject. The age range of the corpus was 17-60 however, and did not seem to cause a deficiency in their performance.

3. TEST RESULTS

The test scores for all of the Tidong intelligibility testing are displayed in Figure 5. Figure 5 includes the PSC data available following the collecting of an additional Tidong wordlist in Indarasabak TU during the testing phase. The data strongly supports Smith's hypothesis of a dialect chain. Lidung LS and Indarasabak TU occupy more central positions within the dialect chain relating to each other with 85 PSC and in the range 74-84 PSC with Tanjung Aru SN and Rancangan Blok No.31 TU. Tanjung Aru SN and Rancangan Blok No.31 TU represent the end links in the chain, being most distantly related with each other at 67 PSC. The whole set of PSC relations is included in Figures 6 and 7.

REFERENCE TAPE		TIDONG			MURUT	
		Tanjung Aru SN	Lidung LS	Rancangan Blok No.31 TU	Serudung TU	Inarad KN
TEST POINT						
TIDONG	Rancangan Sungai Manila SN	99 (--)*	99 (76)*	99 (67)*		
	Lidung LS	93 (76)	100 (--)	82 (74)		
	Indarasabak TU	92 (76)	92 (85)	85 (84)	20 (58)	7 (60)
MURUT	Serudung Baru TU		29 (55)		98 (--)	65 (63)
* PSC relation is between reference tape village and Tanjung Aru SN						

Figure 5: Intelligibility scores (given as percentages) for testing in Tidong villages and the Murut village of Serudung Baru TU. (PSC relations are shown in parentheses.)

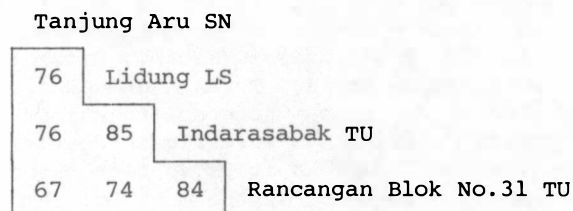


Figure 6: Revised PSC relations between Tidong villages including Indarasabak TU

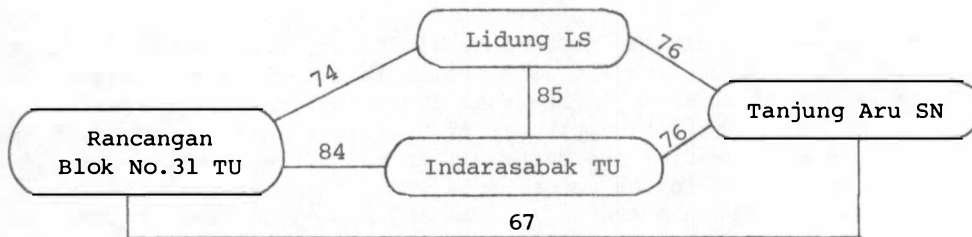


Figure 7: Relations among the Tidong villages shown as a dialect chain

The evidence added by intelligibility testing further shows that the four Tidong villages represent the same language as all of the intra-Tidong scores were above the language threshold of 80% (→ Introduction). The relatively low 82% intelligibility score registered by Lidung LS subjects on the Rancangan Blok No.31 TU reference tape is perhaps explained by the youthful corpus, as the younger generation of speakers generally speak a less pure form of their mother tongue. Though they would be able to understand completely their parents' speech, they would be less likely to pick out slight variations in the speech patterns of more distant speech communities.

The 84 PSC between Rancangan Blok No.31 TU and Indarasabak TU makes the intelligibility score of 85% which Indarasabak TU subjects averaged on the Rancangan Blok No.31 TU story seem low also. This is especially so as the two villages are within several miles of each other. When the individual scores of subjects were studied, it was observed that two low scores were received by female subjects who also had very low scores on the national language story. Thus it is possible that this is a case where the use of the national language to question the subjects resulted in a lower average test score. It is known however that residents of Rancangan Blok No.31 TU come from the Sesayap River in East Kalimantan, and thus may represent a distinct dialect from the Tarakan dialect spoken in Indarasabak TU. The lower score for Lidung LS subjects on the Rancangan Blok No.31 TU story would support this distinction.

The Murut as spoken in Serudung TU and Inarad KN was plainly unintelligible to the Tidong speakers at Indarasabak TU. The intelligibility scores were 20% and 7% respectively. The subject with the highest individual score on the Serudung TU story volunteered the comparison that it sounded "like Dusun". Similarly, the Tidong story from Lidung LS was not understood by subjects in Serudung Baru TU whose average score was only 29%. Serudung Baru TU test subjects scored higher on stories told in other Murutic languages as attested by the 65% for the Inarad KN story. (→ Murutic, for a fuller discussion of the placement of Serudung within that family.)

Though the testing clearly demonstrates the four Tidong villages constitute a single language, there are yet unresolved questions regarding Tidong. The present study deals only with Tidong villages in Sabah. Dialect intelligibility testing indicates two dialects may be spoken, Tarakan and Sesayap. Further investigation will be required to determine the complete number of Tidong dialects/languages. It is also desirable to investigate further the relationship of Sabah Tidong to Kalimantan Tidong as it is spoken on the coast around Tarakan, around Nunukan Island, in the area at the mouth of the Bolongan River, and along the Sesayap, Sibuku and Sembakung Rivers further inland. These two areas of study will undoubtedly entail considerable overlap. It will also be of

benefit to clarify the related issue of whether Tidong and Murut do in fact both dwell along the Sembakung River, and, if such is the case, the linguistic relationship between them.

4. NATIONAL LANGUAGE INTELLIGIBILITY

The Bahasa Malaysia story was included in the test set for each of the three Tidong villages. A summary of the test scores and sociolinguistic information is given in Figure 8. Figure 8 raises several questions, the most apparent being why Rancangan Sungai Manila SN would comprehend the national language so much better (92%) than the other two villages, particularly Lidung LS (62%) whose subjects on the average had received more formal education.

VILLAGE	SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Rancangan Sungai Manila SN	93	7/1	37	22-54	3	5.7	2.1
Lidung LS	62	6/4	26	15-58	4	6.0	2.4
Indarasabak TU	65	5/5	42	17-60	3*	5.0	1.5
* includes one subject whose education comprised three years of adult education							

Figure 8: National language intelligibility scores with sociolinguistic information for three Tidong villages. (RANGE indicates youngest and oldest test subject. EDUCATION figures indicate the number of subjects having received some formal education (ED.), the average number of years of education those subjects had received (A.E.R.), and the average number of years of education for all subjects in the corpus (A.E.S.).)

A breakdown of the data given in Figure 8 is presented in Figures 9-12. Figure 9 shows the effect of sex on the comprehension of the national language. Figure 10 illustrates how the parameter of age affects intelligibility scoring. Figure 11 indicates the influence of education on the ability of Tidong test subjects to understand Bahasa Malaysia. Figures 9-11 each show the effect of the stated parameter on the entire test corpora and also for each village. Figure 12 presents a further breakdown on the related effects of education and sex on the comprehension of the Bahasa Malaysia story.

Comparison of Figures 9-11 reveals that the age differential has less influence on the test scores than do the parameters of sex and education. It appears that sex is indeed a differentiating factor especially among those who have received no formal education. It seems reasonable to suspect that education would level out this difference; the data available as shown in Figure 12 prevents any sound conclusion because of the limited number of educated females.

It is interesting to note that the scores from Rancangan Sungai Manila SN are consistently high for each parameter mentioned. This would suggest that

some other factor such as location near a mixed population centre or in a more highly developed area has an overriding effect on the parameters discussed above.

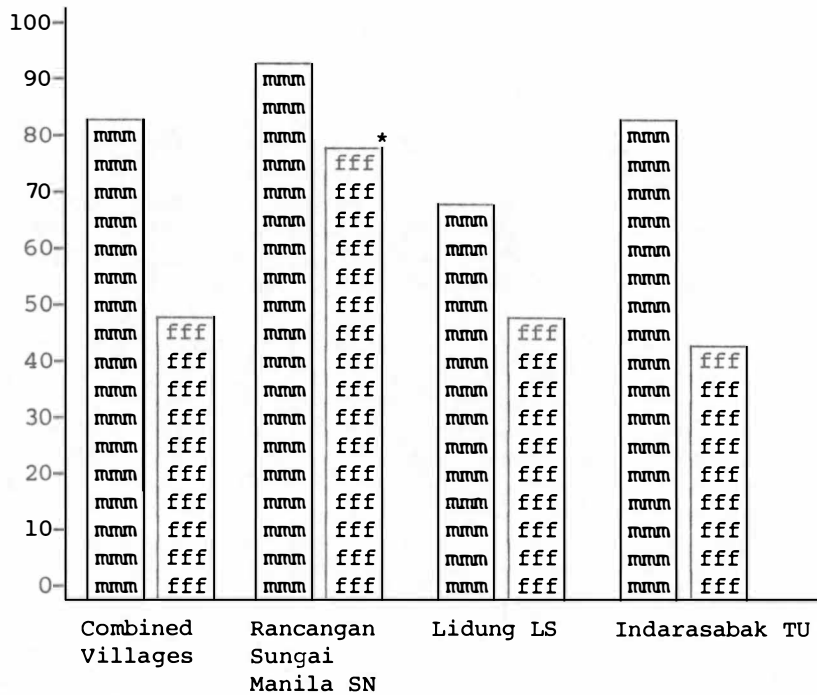


Figure 9: The influence of sex as a parameter in national language intelligibility testing. (The first column of each pair represents the average male score; the second column represents the average female score. * indicates a single score.)

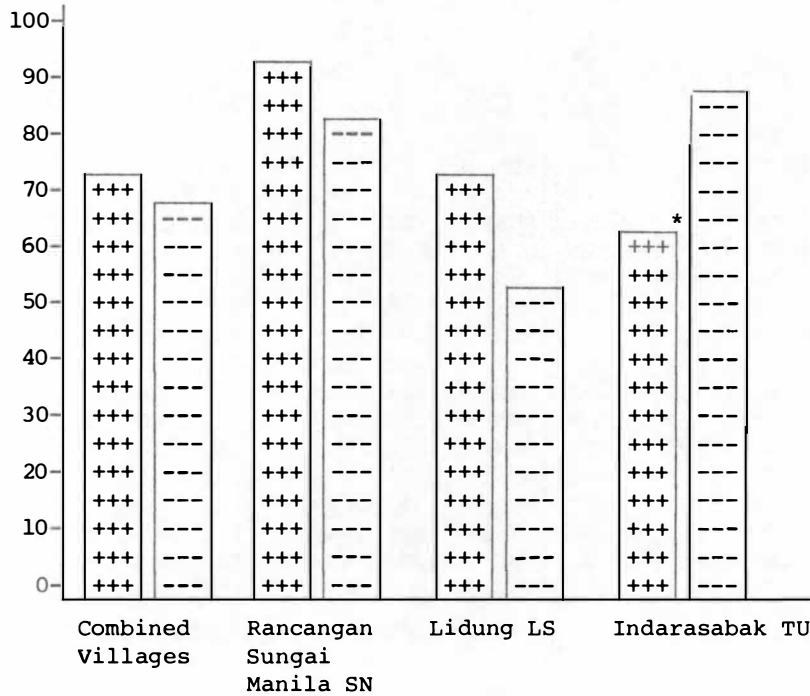


Figure 10: The influence of age as a parameter in national language intelligibility testing. (The first column of each pair represents the average score for subjects over 30 years of age; the second column represents the average score for subjects under 30. * indicates a single score.)

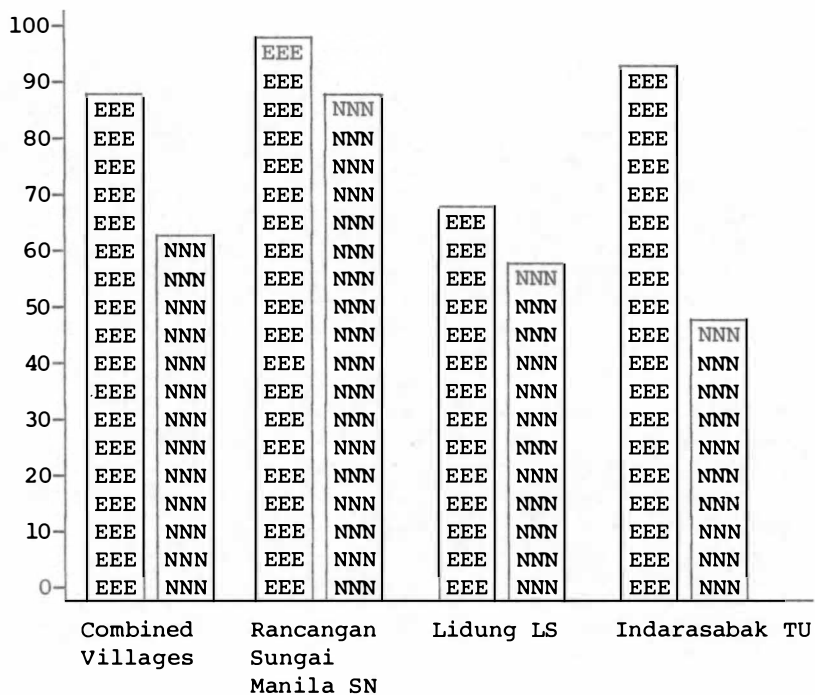


Figure 11: The influence of education as a parameter in national language intelligibility testing. (The first column of each pair represents the average score for subjects who had received some formal education; the second column represents the average score of those subjects who had not received any formal education.)

	Rancangan Sungai Manila SN	Lidung LS	Indarasabak TU
Educated males	100	67	100
Educated females	--	80*	90*
Non-educated males	90	73	75
Non-educated females	80*	40	34

Figure 12: The influence of education according to sex in national language intelligibility testing. (* indicates a single score.)

5. CONCLUSION

The testing within the Tidong villages of Sabah has shown that a single language is spoken among them, supporting Smith's earlier classification. The PSC relations among them indicate that they are related to each other like the links of a chain, and intelligibility testing demonstrates that there is mutual intelligibility between even those villages representing the links at either end of the chain. Smith's decision to include Serudung Murut within the Murutic family has also been confirmed.

NOTES

1. The wordlist from the village of Labuk KAL was actually collected in Serudung Baru TU. The wordlist labelled Serudung TU was also collected in Serudung Baru TU, a resettlement scheme with a Serudung Murut population of about 200 at the time intelligibility testing was done there. The Serudung Murut population is still in flux between the older village (Serudung) which is located on the Serudung River, and the resettlement scheme; the movement is bi-directional.
2. The seven Murut villages are: Minansut KU, representing the Gana language; Bukau BT, representing the Beaufort Murut language; Langsat TM, representing the Timugon language; Pensiangan NN-PN, representing the Tagal language; Kadalakan KU, representing the Nabay language; Sook KU, representing the Paluan language; and Baru Jumpa TM, representing the Kolod language.
3. Pallesen (1977:378) suggests that whenever a language manifests a "transition" position between distantly-related subgroups based on lexicostatistical data, it should be recognised as having special significance for the study of convergence and historical movements between the two subgroups. Sembakung Murut and Kalabakan Murut appear to occupy such a position between the Tidong and Murutic language families. A topic of interest for future research is the nature of the divergence between the Tidong and Murutic peoples and subsequent contact between them.

THE PAITANIC LANGUAGE FAMILY

Julie K. King

0. INTRODUCTION

The people of Sabah who speak languages here classified as members of the Paitanic language family, live throughout the northern and west-central parts of the state. They are generally an inland people who live along rivers in the interior of Sabah, though a few coastal settlements also exist (Figure 1).

The origin of this group is uncertain since most of the persons interviewed maintained that they had originated in their present locations. However, some people in the villages visited during the survey spoke of a migration of a group of people from the upper Kinabatangan River area to the village of Lanas KU across the Wittti Range. From Lanas KU and surrounding villages another group later moved to Tampias RU and other villages in that area because of village rivalry and warring. This migration is attested to by linguistic data collected in those areas. There is greater linguistic similarity between those two villages than between them and other villages of the Paitanic family.

The only other known movements of people speaking Paitanic languages involve some government resettlement within the upper Kinabatangan River area, from very remote areas along the river's tributaries to locations more easily accessible along the Kinabatangan River.

The people speaking languages of the Paitanic family have a complicated system of nomenclature. Ethnonyms for this large Paitanic family of languages are numerous and vary from one village to another based on sociological and/or linguistic factors.

In the upper Kinabatangan River area, Paitanic groups have names which they say were adopted by their group after the introduction of Christianity to the area. Some of these names are 'Sinabu' (also 'Sinobu'), 'Sinarupa' (also 'Sinarupo'), 'Makiang', 'Rumanau' (also 'Romanau', 'Roomarrows' and 'Rumanau Alab'), 'Kolobuan', and 'Sungai' (also 'Sungei', 'Orang Sungai', and 'Orang Sungei'). Several persons who were interviewed from these groups said that their former name was 'Tambanua' (also 'Tambanuo', 'Tambanuva', 'Tambanwas', 'Tambenua', 'Tambunwas', 'Tembenua', 'Tombonuva', 'Tunbumohas', and 'Tunbunwha') but that when they adopted Christianity, the name Tambanua was changed to a new name in order to distinguish themselves. At present the name 'Sungai' (*river*) or 'Orang Sungai' (*people of the river*) is being widely accepted in the upper Kinabatangan River area as a generic autonym.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 139-153. *Pacific Linguistics*, C-78, 1984.

© Julie K. King

139

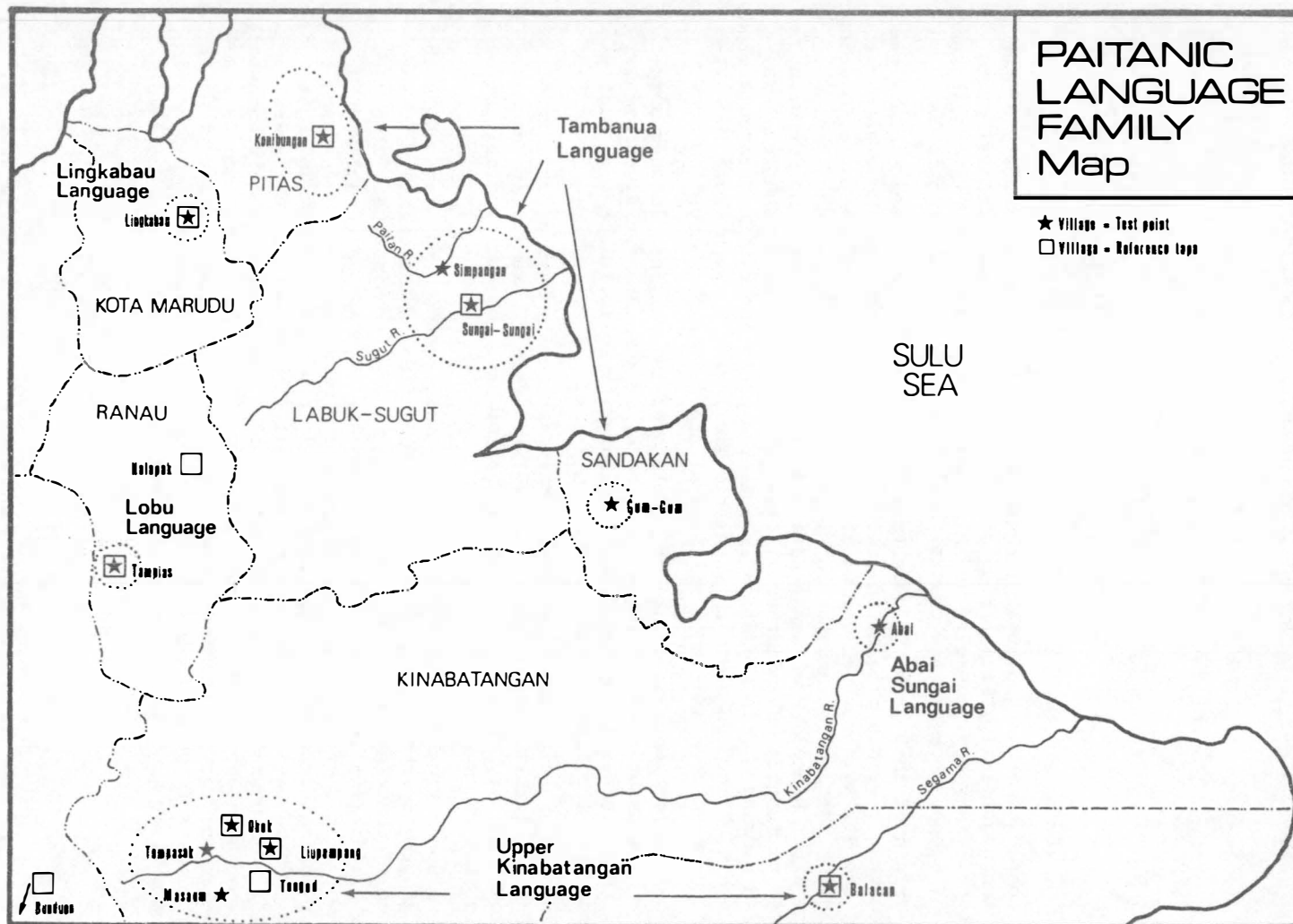


Figure 1: Paitanic language family map showing the general area where the five Paitanic languages are spoken, the villages where intelligibility testing was done and the sources of the reference tapes used in the testing

In the lower Kinabatangan River area as well as the regions along the Sugut and Paitan Rivers and on the Bengkoka Peninsula (Pitas District) where languages of the Paitanic family are spoken, the distinction is quite clear between the 'Tambanua' people who consider themselves religionless and the 'Orang Sungai' who follow the religion of Islam. The 'Orang Sungai' who were interviewed said that they changed their autonym when they converted to Islam.

The people living in Lanas KU and surrounding villages and those speaking the same language in Tampias RU and the surrounding area call their language 'Lobu', which in the Paitanic languages means *people*.

Two other groups speaking languages that are here classified as Paitanic are the 'Lingkabau' (also 'Linkabau') people living in a village of the same name and other villages in the surrounding area in southern Kota Marudu District; and the 'Dusun Segama' (also 'Saga-i', 'Saghai', 'Segai', and 'Segama Dusun') people living along the Segama River in Lahad Datu District.

The people of Parancangan LS speak the 'Dumpas' language (→ Dumpas). This language patterns lexicostatistically on the border between Dusunic and Paitanic. Though it had been originally classified as Dusunic, intelligibility testing results indicated a closer affinity to the Tambanua (Paitanic) language. Since no cross-testing has been done, Dumpas will continue to be considered Dusunic.

In the whole of Sabah there are approximately 20,000 people speaking the languages that are here classified as Paitanic.¹

The only known published material about Paitanic languages are a linguistic sketch by Asmah Haji Omar, in which she describes the 'Paitan' language of Membangan BT,² and a description of dialect intelligibility testing among the languages of the upper Kinabatangan River area by Hurlbut and Pekkanen (to appear) which includes one of the languages here classified as Paitanic.

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classifies the Paitanic language family as consisting of five languages: Upper Kinabatangan (which includes also Dusun Segama), Tambanua, Abai Sungai, Lingkabau, and Lobu. He further states that these languages relate to each other as a language chain. Each of the five languages forms a link in the chain and each link relates to at least one other in the range of 75-80 percent of shared cognates (PSC). The Upper Kinabatangan language is said to form the central link in this language chain (Figure 2). Most of the inter-language relationships are well below 75 PSC as can be seen in Figure 3.

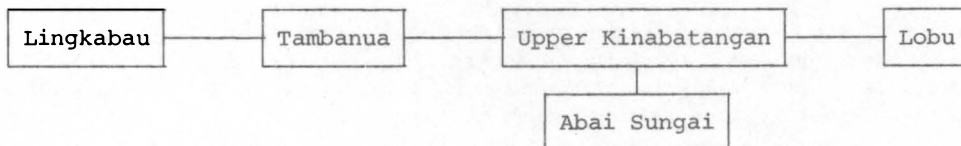


Figure 2: Language chain of the languages in the Paitanic family (from Smith, in this volume)

									LU (Lingkabau KM) ————— Lingkabau Language
70									SI (Pulau Jambongan LS) ————— Tambanua Language
79	80								TA (Konibungan PS)
65	70	77							SI (Abai LS) ————— Abai Sungai Language
64	67	70	73						DA (Bukit Balacan LD)
71	72	76	79	85					KB (Pinangah KN) ————— Upper Kinabatangan Language
71	76	79	73	75	85				SI (Gum-Gum SN)
68	67	71	67	69	80	74			SU (Liupampang KN)
66	60	69	58	60	67	65	72		LO (Tampias RU) ————— Lobu Language

Figure 3: PSC relations between representative samples of the five languages in the Paitanic language family. (from Smith, in this volume) (LU = Lingkabau; SI = Sungai; TA = Tambanua; DA = Dusun Segama; KB = Kolobuan; SU = Sinabu; LO = Lobu.)

Because of these relatively low PSC relations between the languages of the Paitanic family, it was necessary to do testing to determine the level of intelligibility between these related languages. In the case of language chaining such as this noted in the Paitanic language family, there is often found to be a loss of intelligibility between the more distant links.

2. TESTING PROCEDURES

Ideally at least two villages should have been chosen as test points for each of the five Paitanic languages. However, in the cases of Lingkabau, Lobu and Abai Sungai languages only one village was tested for each.³ More than two test points each were chosen for the Upper Kinabatangan and Tambanua languages. This was due to the significantly larger size and geographical spread of the groups speaking these languages and also to the great linguistic and sociological diversity found within these two language groups.⁴

The taped stories used for testing differed somewhat from language to language and also sometimes from village to village within a given language. An attempt was made to test and cross-test each of the five Paitanic languages. This was not done, however, in the case of the Abai Sungai language (→ Section 3.3) nor the Lobu language (→ Section 3.5).

Figure 4 gives a summary of the testing done in Paitanic language villages, showing the reference tapes used in each case. The test results in each of the five Paitanic languages will be discussed more thoroughly in Section 3 showing the specific reference tapes used at each test point.

All of the tapes used for the intelligibility testing were considered good with regard to technical quality. The content of all of the stories was also good, although in the case of the story from Liupampang KN the subject matter and progression of the story may have been too easy to provide a good test. In all cases where that tape was used subjects scored very high.

REFERENCE TAPES TEST POINTS	UPPER KINABATANGAN	TAMBANUA	ABAI SUNGAI	LINGKABAU	LOBU
UPPER KINABATANGAN Liupampang KN Tempasak KN Masaum KN Bukit Balacan LD	10 xx xxx xxx xx	7 xx xx xx x	-	1 x	2 x x
TAMBANUA Konibungan PS Sungai-Sungai LS Simpangan Paitan LS Gum-Gum SN	12 xxx xxx xxx xxx	5 x x xx x	-	1 x	2 x x
ABAI SUNGAI Abai KN	3 xxx	2 xx	1 x	1 x	
LINGKABAU Lingkabau KM	3 xxx	2 xx	-	1 x	1 x
LOBU Tampias RU	2 xx				

Figure 4: Summary of Paitanic language family test points and reference tapes used for testing at each point. (The numerals indicate the number of reference tapes used in the testing within each language group. x's indicate the number of reference tapes used at each test point.)

3. TEST RESULTS

3.1 Upper Kinabatangan language

Intelligibility testing was conducted in four villages which had been classified as representing the Upper Kinabatangan language. The results of this testing are shown in Figure 5.⁵

Three of these test points, Liupampang KN, Tempasak KN, and Masaum KN are located in the upper Kinabatangan River area and refer to themselves and their languages as 'Sinabu', 'Makiang', and 'Rumanau' respectively. The fourth test point village was Bukit Balacan LD, located near the Segama River in Lahad Datu District. The people of this village refer to themselves as 'Dusun Segama'.

REFERENCE TAPES TEST POINTS	UPPER KINABATANGAN			TAMBANUA		LING- KABAU	LOBU
	Liupampang KN 'Sinabu'	Tongod KN 'Makiang'	Bukit Balacan LD 'Dusun Segama'	Konibungan PS 'Tambanua'	Sungai-Sungai LS 'Sungai'	Lingkabau KM 'Lingkabau'	Tampias RU 'Lobu'
Liupampang KN 'Sinabu'		87 (80)	60 (69)	64 (71)	58 (69)		70 (72)
Tempasak KN 'Makiang'	85 (79)	94 (95)	70 (81)	67 (77)	64 (76)		54 (67)
Masaum KN 'Rumanau'	87 (85)	86 (74)	67 (62)	59 (68)	62 (65)	41 (65)	
Bukit Balacan LD 'Dusun Segama'	91 (69)	91 (82)		60 (70)			

Figure 5: Intelligibility testing results from four Upper Kinabatangan language villages tested with the Upper Kinabatangan language and other Paitanic languages. (Scores are given as percentages. PSC figures are in parentheses. An autonym for each village is given under the village name.)

All four of the test points obtained high scores on the stories from Liupampang KN and Tongod KN, both of which had been classified as representing the Upper Kinabatangan language. However, results on the story from the village of Bukit Balacan LD, which was also classified as representing the Upper Kinabatangan language, are quite low.

There are two possible explanations for the low scores recorded on the Bukit Balacan LD tape. One possible reason is that the village of Bukit Balacan LD is geographically quite distant from the other villages which were tested or from which reference-tape stories were taken. But, since the test subjects at Bukit Balacan LD scored very high on the other two Upper Kinabatangan language stories, this reason may not be viable.

A second possible explanation is that Bukit Balacan LD represents a distinct dialect or even a distinct language from the Upper Kinabatangan language samples. This seems possible based on the relatively low PSC relations between Bukit Balacan LD and the other Upper Kinabatangan language samples (62, 69, 81 and 82 PSC). But since the Bukit Balacan LD test subjects scored so high on the other Upper Kinabatangan language samples, and language learning is an unlikely explanation for that, any reclassification of the Bukit Balacan LD sample will have to wait until more data can be gathered and more thorough testing can be done.

The results of testing at Upper Kinabatangan villages with stories from the other Paitanic languages (Tambanua, Lingkabau and Lobu) are significantly lower than their respective PSC figures in all cases, thus confirming that the four are separate languages with only limited intelligibility between them.

When the Tambanua, Lingkabau, Lobu, and Abai Sungai villages were tested for their understanding of the Upper Kinabatangan language (Figure 6) scores were also quite low and definitely below the same-language threshold (→ Introduction) in all but two situations. As can be seen from Figure 6, there were three high scores on the Liupampang KN story. This however can be explained by the ease of that story, and the scores are therefore considered to be somewhat misleading (Section 2). The other situation where the test scores appear to be higher than expected were the results from the Lobu village of Tampias RU (→ Section 3.5).

REFERENCE TAPES		UPPER KINABATANGAN			
		Liupampang KN 'Sinabu'	Tongod KN 'Makiang'	Bukit Balacan LD 'Dusun Segama'	Obuk KN 'Rumanau'
TAMBANUA	Konibungan PS 'Tambanua'	76 (71)	45 (76)	56 (70)	
	Sungai-Sungai LS 'Sungai'	66 (69)	48 (75)	55 (71)	
	Gum-Gum SN 'Sungai'	93 (74)	39 (82)	76 (75)	
ABAI SUNGAI	Abai KN 'Abai Sungai'		66 (77)	62 (73)	
LING- KABAU	Lingkabau KM 'Lingkabau'	89 (68)	61 (70)	64 (64)	
LOBU	Tampias RU 'Lobu'	95 (72)			85 (76)

Figure 6: Intelligibility testing results when Upper Kinabatangan language reference tapes were used in testing at Tambanua, Abai Sungai, Lingkabau and Lobu villages. (Scores are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

3.2 Tambanua language

Three villages where the language classified by Smith as Tambanua is spoken were tested for their understanding of the Tambanua language of different areas than their own and their understanding of other Paitanic languages. In addition,

the village of Gum-Gum SN is included as a Tambanua test point even though it had initially been classified lexicostatistically as representing the Upper Kinabatangan language. During the intelligibility testing it was discovered that the wordlist on which the lexicostatistical classification was based, had been elicited from a person who had originated in the Upper Kinabatangan area. However, the majority of the residents of Gum-Gum SN who speak a Paitanic language (all of those who were tested) originally came from the Sugut River area and speak the Tambanua language. Figure 7 displays the results of that testing. Figure 8 shows the results of cross-testing Tambanua reference tapes in villages where other Paitanic languages are spoken.

From Figure 7 it can be seen that in testing within the Tambanua language, results were very high, so that it appears that there is no dialect distinction between the language as it is spoken in Pitas District and Labuk-Sugut District. The results at Gum-Gum SN, though lower, are still within the single-language boundary.

When Tambanua test subjects were tested for their understanding of the Upper Kinabatangan, Lobu, and Lingkabau languages, test results were generally below the same-language threshold and significantly lower than respective PSC relations. These results, then, would confirm that Tambanua is a distinct language from the Upper Kinabatangan, Lobu, and Lingkabau languages.

REFERENCE TAPES		TAMBANUA		UPPER KINABATANGAN			LOBU	LING- KABAU
		Konibungan PS 'Tambanua'	Sungai-Sungai LS 'Sungai'	Bukit Balacan LD 'Dusun Segama'	Liupampang KN 'Sinabu'	Tongod KN 'Makiang'	Tampias RU 'Lobu'	Lingkabau KM 'Lingkabau'
TEST POINTS								
TAMBANUA	Konibungan PS 'Tambanua'		93 (91)	56 (70)	76 (71)	45 (76)		
	Sungai-Sungai LS 'Sungai'	93 (91)		55 (71)	66 (69)	48 (75)	60 (67)	
	Simpangan Paitan LS 'Sungai'	93 (90)	93 (91)	46 (70)	52 (73)	51 (75)		50 (77)
	Gum-Gum SN 'Sungai'	84 (79)		76 (75)	93 (74)	39 (82)	47 (65)	

Figure 7: Intelligibility testing results at four Tambanua villages. (Results are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

REFERENCE TAPES TEST POINTS		TAMBANUA	
		Konibungan PS 'Tambanua'	Sungai-Sungai IS 'Sungai'
UPPER KINABATANGAN	Liupampang KN 'Sinabu'	64 (71)	58 (69)
	Tempasak KN 'Makiang'	67 (77)	64 (76)
	Masaum KN 'Rumanau'	59 (68)	62 (65)
	Bukit Balacan LD 'Dusun Segama'	60 (70)	
ABAI SUNGAI	Abai KN 'Abai Sungai'	57 (72)	76 (72)
LING- KABAU	Lingkabau KM 'Lingkabau'	69 (79)	64 (77)

Figure 8: Intelligibility testing results for villages where a Tambanua reference tape was used for testing. (Scores are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

3.3 Abai Sungai language

The only village tested where the Abai Sungai language is spoken was the village of Abai KN. Results of that testing are shown in Figure 9. As can be seen from the chart, test results were low in all cases. This is likely due to the fact that this language group is isolated from other Paitanic language communities. It should also be noted that persons from Abai KN stated that members of the community are making an effort to use the national language (Bahasa Malaysia) more than their local dialect. This too could influence the test scores.

Based on the limited test results for Abai Sungai, there is no reason for reclassification of this language. Further intelligibility testing and cross-testing would be necessary in order to make a more complete statement about this language community. And further, a sociolinguistic survey of the language community could help in gaining a better understanding of language use in the Abai Sungai community.

REFERENCE TAPES	UPPER KINABATANGAN		TAMBANUA		LING- KABAU	LOBU
	Tongod KN 'Makiang'	Bukit Balacan LD 'Dusun Segama'	Konibungan PS 'Tambanua'	Sungai-Sungai LS 'Sungai'	Lingkabau KM 'Lingkabau'	Tampias RU 'Lobu'
Abai KN 'Abai Sungai'	66 (77)	62 (73)	57 (72)	76 (72)	44 (65)	44 (58)

Figure 9: Intelligibility testing results from the Abai Sungai village of Abai KN. (Results are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

3.4 Lingkabau language

The village of Lingkabau KM (not to be confused with Lingkabau LS, where the Tambanua language is spoken) was the only village tested where the Lingkabau language is spoken. The results of that testing are shown in Figure 10.

REFERENCE TAPES	UPPER KINABATANGAN			TAMBANUA		LOBU
	Tongod KN 'Makiang'	Liupampang KN 'Sinabu'	Bukit Balacan LD 'Dusun Segama'	Konibungan PS 'Tambanua'	Sungai-Sungai LS 'Sungai'	Tampias RU 'Lobu'
Lingkabau KM 'Lingkabau'	61 (70)	89 (68)	64 (64)	69 (79)	64 (77)	57 (66)

Figure 10: Intelligibility testing results at the Lingkabau village of Lingkabau KM. (Results are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

Intelligibility testing results were low in every case except for the test on the Liupampang KN story (→ Section 2). All the persons tested stated that each of the stories was difficult, even though they recognised some words. They felt that they really did not understand the tapes and that all the languages on the tapes were different from their own.

Intelligibility testing results, therefore, confirm the classification of the Lingkabau language as different from the other Paitanic languages with which it was tested.

3.5 Lobu language

Tampias RU was the only Lobu-speaking village where intelligibility testing was done. The results of that testing are shown in Figure 11.

REFERENCE TAPES TEST POINT	UPPER KINABATANGAN		KADAZAN/DUSUN	
	Liupampang KN 'Sinabu'	Obuk KN 'Rumanau'	Nalapak RU 'Dusun'	Bunduon PG 'Kadazan'
Tampias RU 'Lobu'	95 (72)	85 (76)	95 (59)	54 (55)

Figure 11: Intelligibility testing results at the single Lobu language test point of Tampias RU. (Results are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.)

The test set used at the village of Tampias RU was made up of two stories from villages where the (Paitanic) Upper Kinabatangan language is spoken and two stories from villages where Dusunic languages are spoken. The Dusunic languages were chosen because the village of Tampias RU is one of about three villages where the Lobu language is spoken in an area where Dusun is the dominant language. The testing of Lobu speakers with Dusunic language tapes was done only to demonstrate the amount of language learning which has taken place in the Lobu community.

The only Paitanic language tapes used in this testing were two representing the Upper Kinabatangan language. These were the only two with PSC relations high enough to warrant testing. When this testing was done, the existence of language chaining in the Paitanic language family had not yet been discovered. Further testing of Lobu with the other Paitanic languages would better reveal the position of the Lobu language in the larger Paitanic language family.

It is also interesting to note that although the Lobu test subjects scored very high on the two Upper Kinabatangan language stories, in cross-testing the Upper Kinabatangan test subjects scored only about 70% intelligibility on the Lobu story (Figure 5). The 95% intelligibility scored on the Dusun story from Nalapak RU also reveals that there has been a great deal of language learning on the part of the Lobu language community.

4. NATIONAL LANGUAGE INTELLIGIBILITY

National language intelligibility testing was done in each of the 11 Paitanic test points. The results of that testing, including some sociological data about each test corpus are displayed in Figure 12.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Liupampang KN (Upper Kinabatangan)	49	7/3	37	17-62	2	4	1
Tempasak KN (Upper Kinabatangan)	50	5/5	35	21-55	0	-	-
Masaum KN (Upper Kinabatangan)	42	7/3	33	15-50	6	6	4
Bukit Balacan LD (Upper Kinabatangan)	88	5/6	35	13-65	4	6	2
Konibungan PS (Tambanua)	81	6/4	37	14-64	3	6	2
Sungai-Sungai LS (Tambanua)	80	7/3	37	16-64	4	6	2
Simpangan Paitan LS (Tambanua)	70	8/2	37	25-62	1	6	1
Gum-Gum SN (Tambanua)	96	6/4	32	16-48	5	8	4
Abai KN (Abai Sungai)	62	8/2	38	16-60	2	10	2
Lingkabau KM (Lingkabau)	35	6/4	37	23-55	1	6	1
Tampias RU (Lobu)	68	4/3	39	23-55	4	4	2

Figure 12: Comprehension of the national language in 11 Paitanic villages with sociological data. (Test scores are given as percentages. Under EDUCATION, ED. = the number of subjects who had received formal education, A.E.R. = the average number of education those subjects had received, and A.E.S. = the average number of years of education per subject for the corpus as a whole. Language names, following the classification in this paper, are given in parentheses beneath the village names.)

As would be expected, intelligibility of the national language was generally higher for persons who have had some formal education. This was true except in the case of the test subjects from Konibungan PS, where the persons who had some formal education averaged 77% intelligibility of the national language test story and persons with no formal education averaged 83% intelligibility. On the basis of the data gathered in this study no explanation of this seeming discrepancy can be offered here.

The village of Gum-Gum SN scored the highest on this test (96%). The village of Gum-Gum SN is a mixed language community, so that residents of that village must use the national language in order to communicate with persons from language groups other than their own. In addition, Gum-Gum SN is located in an area with easy access to the city of Sandakan SN, a major economic and educational centre.

The results of the national language test at Abai KN were surprisingly low, since the people in that village said they are making a conscious effort to make Bahasa Malaysia the primary means of communication in the community. It should be noted, however, that of the ten persons tested at the village of Abai KN, only one had been to school, and that factor alone could account for the discrepancy.

5. CONCLUSION

On the basis of the intelligibility testing results presented in this paper, it appears that within the Paitanic language family there are four mutually unintelligible languages: the Upper Kinabatangan language, the Tambanua language, the Abai Sungai language, and the Lingkabau language. In addition, the Lobu language may also be a language separate from the rest, but further study is necessary in order to confirm this.

In the Upper Kinabatangan language testing, although the test subjects at the village of Bukit Balacan LD had a very high level of understanding of other Upper Kinabatangan language tapes, the reverse was not true. Further testing is needed to determine more precisely whether the Upper Kinabatangan language of Bukit Balacan LD (the language there is known as Dusun Segama) should be considered a dialect or a separate language.

In all other cases of testing between the Upper Kinabatangan language, the Tambanua language, and the Lingkabau language, with the exception of the skewing caused by the Liupampang KN story (Section 2), the intelligibility testing results show clearly that these three groups have very low mutual intelligibility and should therefore be considered separate languages.

NOTES

1. This figure can be broken down into separate language groups as follows: Upper Kinabatangan language, 5,000; Tambanua language, 10,000; Abai Sungai language, 500; Lingkabau language, 3,000; and Lobu language, 1,500. These figures are based on information given in the villages and from district officials.

2. The inhabitants of Membangan BT originally came from the Paitan River area, Labuk-Sugut District.
3. Other Lingkabau villages were not tested because of the remoteness of the area and the unavailability of guides. No further testing of the Abai Sungai language was done because the people of Abai KN were unaware of other villages where the same language was spoken. No further Lobu language testing was done because technicians were unaware at the time that there were other areas where the Lobu language was spoken. Since the initial intelligibility testing reported in this paper, it has been determined that there are at least two other villages near Tampias RU and approximately seven villages around Lanas KU where the Lobu language is spoken.
4. Initial intelligibility testing was done in the upper Kinabatangan River area of Sabah (Hurlbut and Pekkanen, to appear) prior to the collection of the full set of Paitanic language family data. That initial testing was restricted to the geographical boundaries of that region in order to test the mutual intelligibility of all the languages in that area.
5. Since the only Paitanic language considered by Hurlbut and Pekkanen (to appear) was the Upper Kinabatangan language, their intelligibility testing results (given in the following chart) are not considered relevant for the classification of Paitanic languages in this paper. In the chart results are given as percentages. PSC relations are in parentheses. An autonym for each village is given under the village name.

REFERENCE TAPES							
TEST POINTS		Diwara KN 'Makiang'	Tongod KN 'Makiang'	Pinangah KN 'Kolobuan'	Masaum KN 'Rumanau'	Obuk KN 'Rumanau'	Liupampang KN 'Sinabu'
UPPER KINABATANGAN LANGUAGE	Tempasak KN 'Makiang'		91 (95)				94 (79)
	Diwara KN 'Makiang'		95 (96)	89 (99)	77 (74)		98 (80)
	Langga KN 'Kolobuan'	59 (97)		86 (98)			
	Pinangah KN 'Kolobuan'		95 (98)		74 (74)		94 (80)
	Masaum KN 'Rumanau'		91 (74)	79 (74)			91 (85)
	Lulumiab KN 'Rumanau'		84 (--)	97 (--)	98 (--)		98 (--)
	Obuk KN 'Rumanau'		77 (74)				
	Liupampang KN 'Sinabu'		98 (80)			96 (88)	
	Kitumbalang KN 'Sinabu'		94 (84)			99 (85)	98 (93)
	Minusu KN 'Sinabu'		85 (83)			87 (88)	
	Bulot KN 'Sinarupa'		97 (77)	66 (99)		52 (78)	80 (80)

(Refer Note 5, previous page)

THE MURUTIC LANGUAGE FAMILY

John A. Spitzack

0. INTRODUCTION

This paper will discuss the Murutic languages of Sabah and Kalimantan as tentatively classified by Smith (in this volume). It will present the results of the dialect intelligibility testing by the Summer Institute of Linguistics (SIL) and compare these to Smith's lexicostatistical analysis. The aim of the paper is to present a reclassification of the Murutic languages of Sabah and Kalimantan using the dual parameters of lexicostatistics and intelligibility testing.

Other factors brought to bear on the final classification are previous arrangements of the Murutic languages and peoples by several earlier researchers. In addition, the results of intelligibility testing of the national language (Bahasa Malaysia) by the Murut of Sabah are included.

0.1 Location

The languages of the Murutic family are spoken over the south-western portion of the state of Sabah, including the administrative districts of Keningau, Tenom, Sipitang and Pensiangan and extending south into Indonesian Borneo. There are also scattered populations in Beaufort, Tawau and Kinabatangan districts (Figure 1).

0.2 Historical origins

Some have suggested that the Murutic-type people arrived in Sabah as a first wave of migration by a landbridge from mainland South-east Asia, followed by a second wave of Dusunic-type people who pushed them further inland. Others, who notice Chinese-like characteristics in the modern Kadazan, suggest that the migration is much more recent, coming more directly from China. It is probably safe to say that however men first arrived in Sabah, the present Murut and Kadazan/Dusun population represent the direct descendants of those first inhabitants and that there were no other aboriginal people (Rutter 1929; Lee 1965; Chatfield 1972).

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 155-223. *Pacific Linguistics*, C-78, 1984.

© John A. Spitzack

155

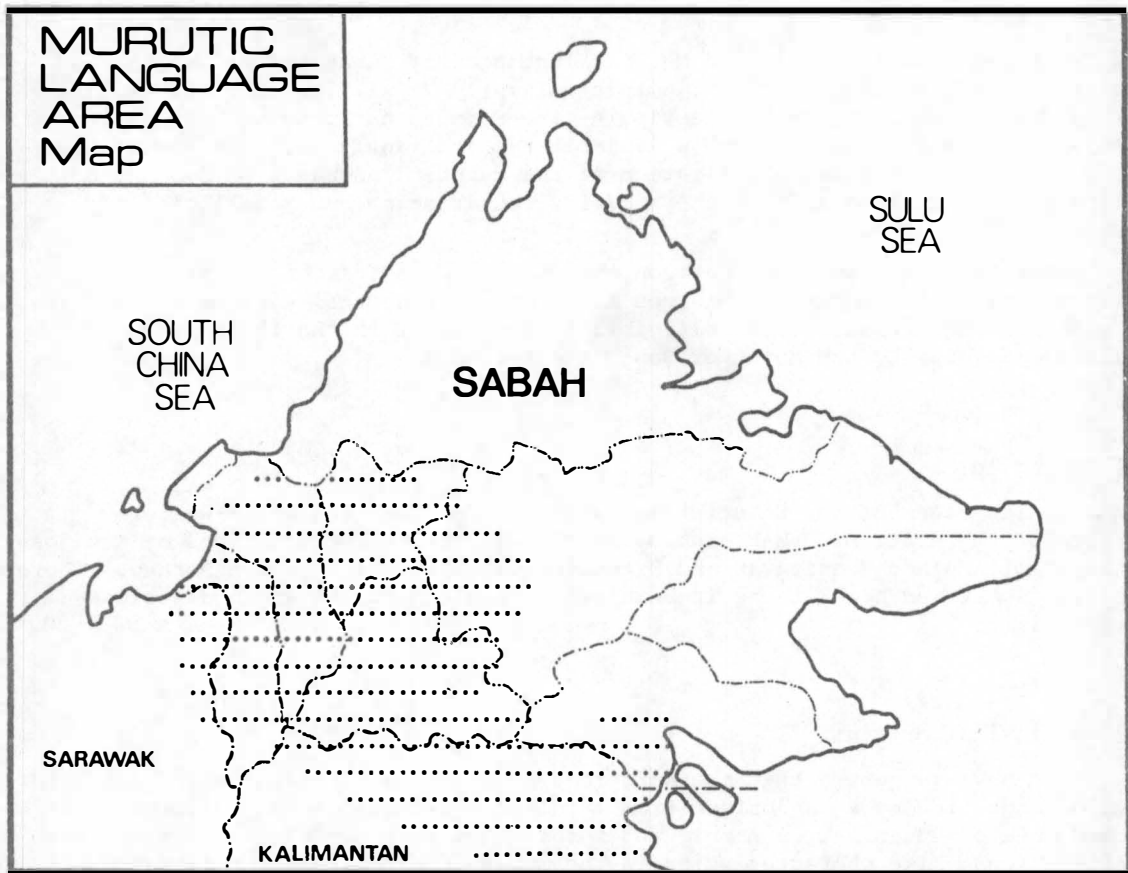


Figure 1: Murutic language area of Sabah, Sarawak and Kalimantan

On the other hand, even to the early researchers it seemed quite apparent that there was a closer relationship between the Murutic and Dusunic peoples of Sabah and some of the nearby groups in the Philippines, Taiwan and mainland South-east Asia (Hose and McDougall 1912; Rutter 1929; Prentice 1970).

The Murut of Sabah also seem to share some cultural affinity with the Lundayeh of Kalimantan and Sarawak (also called 'Murut'), but linguistically they are very distant (→ Lundayeh; LeBar 1972:153). It is probable that the Lundayeh arrived in a totally different migration from that of the Murutic people.

0.3 Population

The total population of Murutic people of Sabah, including the non-indigenous Muruts such as Kolod, but excluding such groups as Lundayeh, Kuijau and Tidong, is estimated to be 36,000 to 38,000 as of 1980.¹ The breakdown according to language designation is as follows:

GANAN	1000-1200
KENINGAU MURUT	4000-5000
BAUKAN	1500-1800
TIMUGON	6000-7000
BEAUFORT MURUT	1200-1700
PALUAN	5500-6500
SUMAMBU/TAGAL	13000-14000
SELUNGA MURUT	600-800
KALABAKAN MURUT	450-600
SERUDUNG MURUT	350-450
Other (SEMBAKUNG, INDONESIAN MURUT, KOLOD)	2800-3800

A topic of considerable discussion by some researchers has been the sharp decline and gradual rise of the Murut population in recent history (Ride 1934; Copeland 1935; Shircore 1937; Clarke 1951; Landgraf 1956; Register 1956; Polunin 1959; Jones 1966, 1967; Crain 1972). Some attribute this to disease, others to culture change and some simply to varying classifications in census taking. It may actually be due to a combination of factors, and has probably affected different ethnic groups at different times, not only the Murut (Appell 1968; Prentice 1972).

0.4 Ethnonyms

The ethnonym 'Murut' (also 'Maroot', 'Marut', 'Meroot', 'Merut', 'Mooroot', 'Moorut', 'Morut', 'Mulut', 'Murat', 'Murút') is applied generically to all languages under discussion in this paper. Literally the term means *hill people*.² Several of the language groups still retain their individual autonyms in common usage, but most apply the more generic term at first when speaking of themselves. When it was possible to discover the autonym it has been used in classifying the language. This presented problems when two villages using different autonyms were found to be linguistically similar. Thus it was often necessary to identify the group by the name of a prominent river or village. In the few cases where a specific autonym was used, it was common for only some in the group to use it (cf. Prentice 1970:370ff).

The misapplication of the term 'Murut' to two very distinct language groups in Sabah and Sarawak has presented great confusion and misunderstanding. Prentice (1971, 1972) has a full discussion of the problems arising from the ethnonyms 'Murut' and 'Tagal' being applied to the Lundayeh of Sarawak (sometimes called Sarawak Murut, Southern Murut or Kelabitic Murut) and to the Murut of Sabah (sometimes called Sabah Murut, Northern Murut or Ida'an Murut).

One suggestion not mentioned by Prentice as to how two such diverse groups came to be referred to by the same name is that the Sarawak Murut were originally called orang murud *people of Murud* after Mt Murud. On the other hand, the Sabah Murut were referred to simply as 'Murut' by outside groups. Eventually the terms became confused and today both are referred to as Murut.³

0.5 Literature

Among the indigenous languages of Sabah, the Murutic languages in general and the Timugon language in particular have received a fair amount of attention, due mainly to the articles and monograph by D.J. Prentice. His work within Murut consists chiefly of a phonology and grammar of Timugon Murut (1971) and numerous other articles relating to various aspects of the language (1965, 1969, 1972, 1981). In his monograph just mentioned he extended his analysis of Timugon by presenting an outline of phonology and morphology of several other Murutic languages. Prentice (1970) also wrote a very important and useful lexicostatistical study which details his classification of the languages of Sabah based on his own wordlists and those gleaned from other literature.

A brief discussion of one aspect of Murut orthography is found in Lees (1966). Clayre (1970) dealt with focus in several East Malaysian languages including the Murutic languages Tagal and Timugon.

Cohen (1981) discussed the relations between 25 wordlists from Keningau District consisting of 80 lexical items each. Of these 25 lists, 14 are considered Murutic in the present paper. His discussion illustrates the difficulty in classifying some of the languages of northern Keningau District as either Murutic or Dusunic.

Several wordlists and vocabularies from Murutic languages have been previously published and are found in Baboneau (1922), Rutter (1929), Keith (1936), Bolang and Harrisson (1949) and the *English-Murut phrase book* (1964). A fuller bibliography, including some non-linguistic material is found in Prentice (1970) and in the bibliography at the end of this book.

1. LEXICOSTATISTICAL CLASSIFICATION

1.1 Smith's classification

The tentative classification of the Murutic languages of Sabah by Smith (in this volume) was based on 36 wordlists collected by linguistic researchers of the Summer Institute of Linguistics (SIL). This first lexicostatistical phase of the survey of Sabah languages by SIL was followed by a second phase of intelligibility testing based partly on the list of points to be clarified found in Smith's summary.

On the basis of relations of 45-60 percent of shared cognates (PSC) with other language families, Smith classified the Murutic family as one of the four

members of the Bornean stock, the others being the Tidong family, the Paitanic family and the Dusunic family. Figure 2 shows the relations between them using representative samples.

TIDONG Family	TIDONG (Lidung LS)			
MURUTIC Family	54	TIMUGON (Entabuan TM)		
PAITANIC Family	46	46	MAKIANG (Diwara KN)	
DUSUNIC Family	45	47	54	DUSUN (Bongkud RU)

Figure 2: PSC relations between representative samples of the four divisions of the Bornean stock (adapted from Smith, in this volume)

The Murutic family was analysed as comprising 15 languages having relations of 52-82 PSC (Figure 3). Seven of these, represented by samples having relations of only 52-75 PSC, were sufficiently different from the others to be separated at the subfamily level. These included Gana, Apin-Apin Kuijau, Tagal, Kalabakan Murut, Kolod, Sembakung Murut and Serudung Murut. Only Tagal was represented by more than one wordlist.

Figure 4 shows nine Tagal wordlists compared with each other and ranging 70-89 PSC. It was found convenient to separate the wordlists into two dialects consisting of Tagal 'A' having relations of 82-89 PSC and Tagal 'B' having relations of 73-85 PSC but having relations between dialects of only 72-82 PSC.⁴

The remaining eight Murutic languages showed greater affinity with one another than to the other seven by having PSC relations of 75-80 with at least one other member in the group. These were classed together under one subfamily called Central Murut. These eight languages, represented by 21 wordlists, included Dusun-Murut, Nabay, Baukan, Timugon, Beaufort Murut, Sook Murut, Takapan⁵ and Paluan.⁶ Figure 5 shows the PSC relations between representative samples of the Central Murut subfamily. No dialect divisions within these languages were noted by Smith.

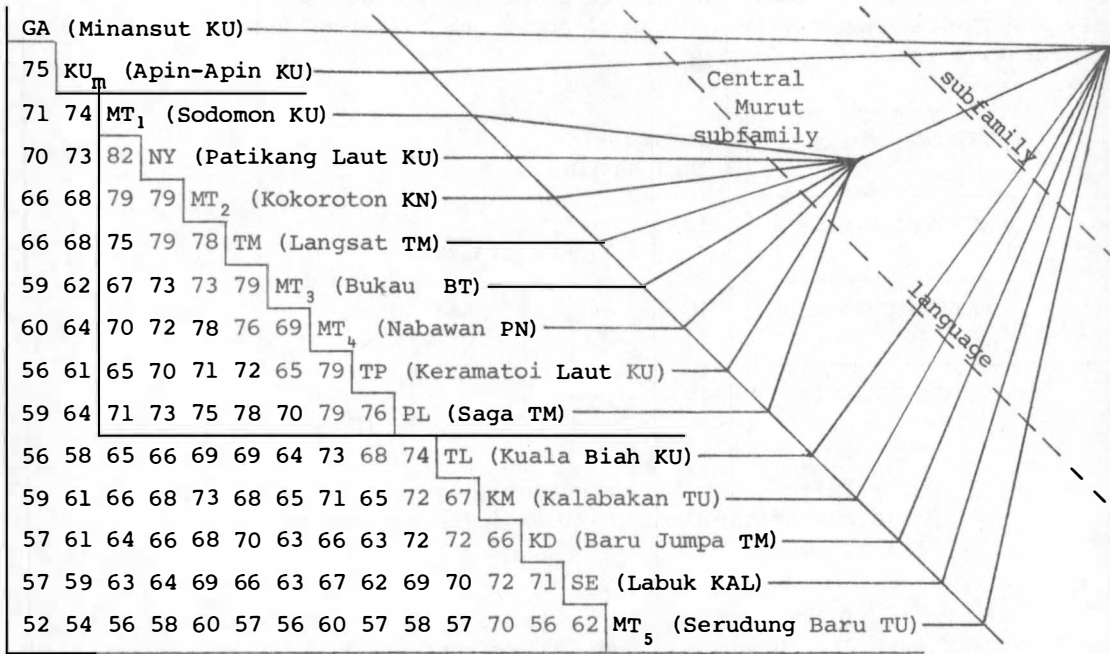


Figure 3: PSC relations of representative dialects of 15 Murutic languages. (from Smith, in this volume) (GA = Gana; KU_m = Apin-Apin Kuijau (subscript m designates this Murutic language in contrast to KU which designates the Dusunic Kuijau language); MT_1 = 'Murut' representing Dusun-Murut; NY = Nabay; MT_2 = 'Murut' representing Baukan; TM = Timugon; MT_3 = 'Murut' representing Beaufort Murut; MT_4 = 'Murut' representing Sook Murut; TP = Takapan; PL = Paluan; TL = Tagal; KM = Kalabakan Murut; KD = Kolod; SE = Sembakung Murut; MT_5 = 'Murut' representing Serudung Murut.)

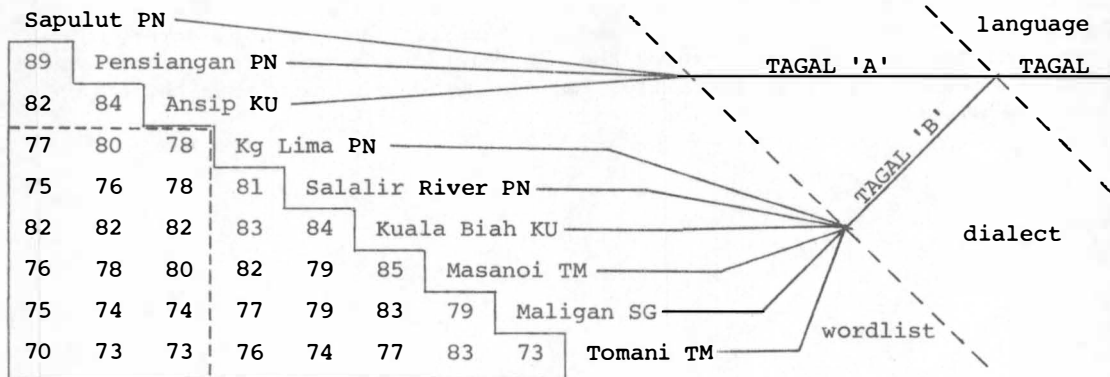


Figure 4: PSC relations of nine samples of the Tagal language representing two dialects (adapted from Smith, in this volume)

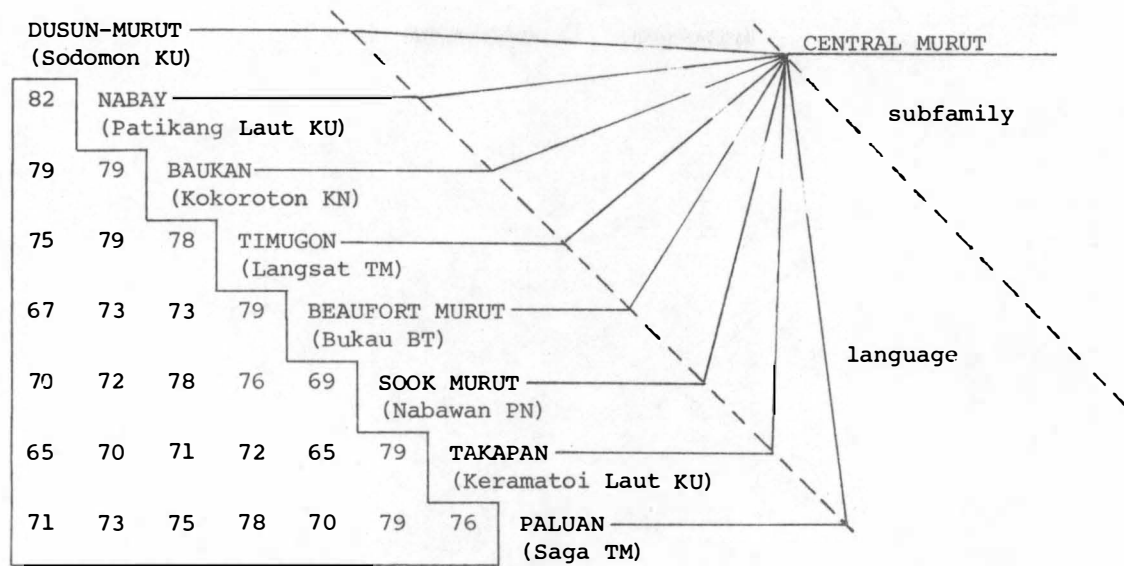


Figure 5: PSC relations of representative samples of eight Murutic languages of the Central Murut subfamily (adapted from Smith, in this volume)

1.2 Revised lexicostatistical classification

During the second phase of the survey it was convenient to check most of the Murutic wordlists and to add several new ones.⁷ The revised lexicostatistical relations presented here take better account of 'synonyms' than the earlier comparisons. Other corrections were also incorporated at this time. Since both Smith's figures and those based on the verified wordlists are cited here and throughout the paper the revised PSC figures are always marked with an asterisk.

The effect these revised figures had on the classification was to raise the PSC relations in all but a few cases. Many of Smith's 'languages' would be described on the 'dialect' level. However, the overall shape of the lexicostatistical classification is only slightly altered.

Figure 6 presents the reclassification which would result if these revised PSC relations were the sole basis for classification.

Figure 7 displays the relationships of 11 villages reclassified as dialect groups of the Central Murut language.

Figure 8 shows the revised PSC relations of five wordlists reclassified as belonging to the Paluan language.

Figure 9 shows the resulting classification based on revised PSC relations of 11 villages which were identified as belonging to the Sumambu/Tagal group.

The initial plan of the intelligibility testing and the discussion of the results in Section 3 centre around Smith's analysis. But the revised lexicostatistical classification is offered here as a further verification of the previous study by Smith and as additional input into the Murutic classification proposed in Section 4, where each dialect/language is discussed in detail. The classification in Section 4 does not necessarily correspond with the classification displayed in Figures 6-9.

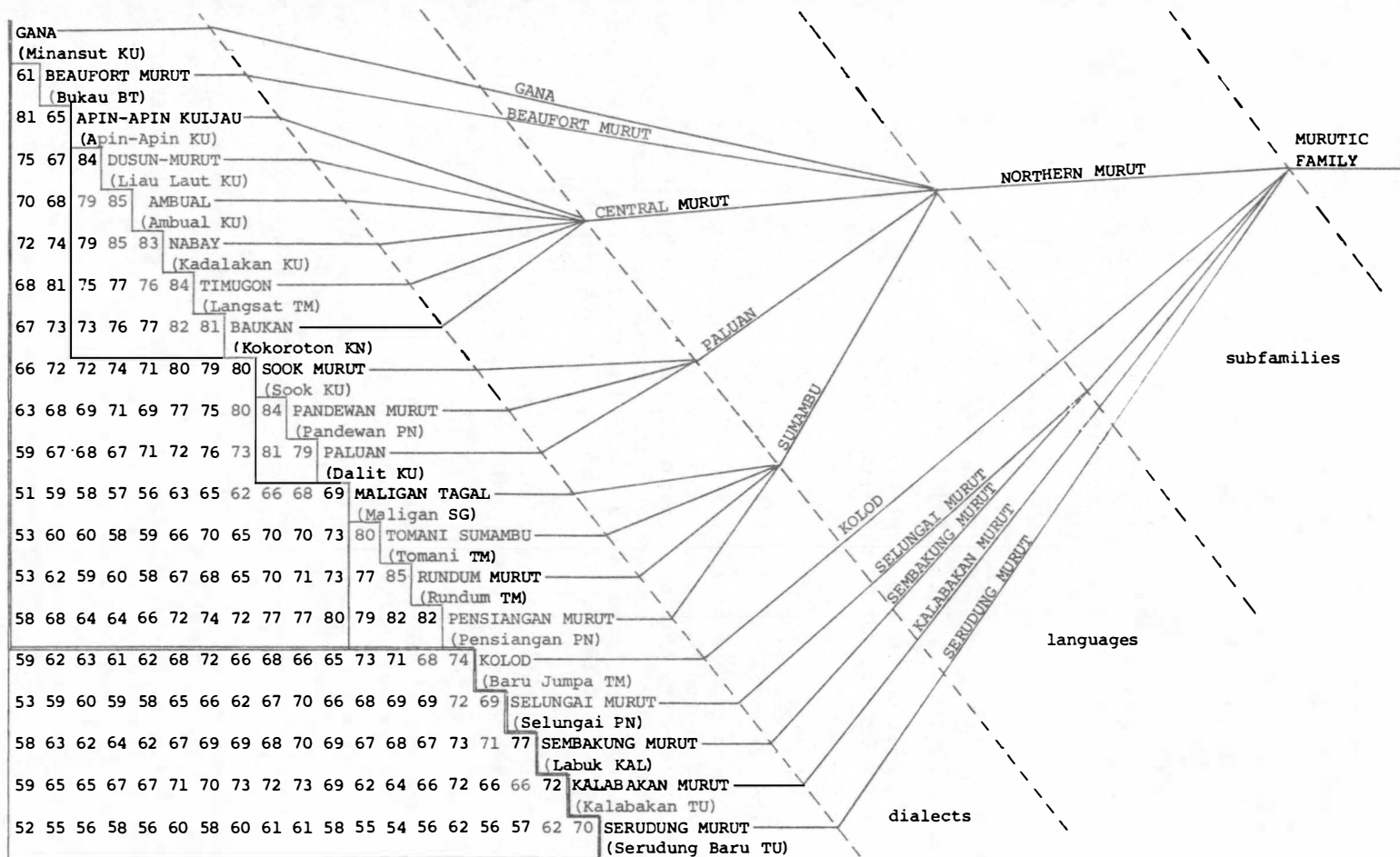


Figure 6: Revised PSC relations between 20 Murutic dialects.

Reclassification based on revised PSC relations.

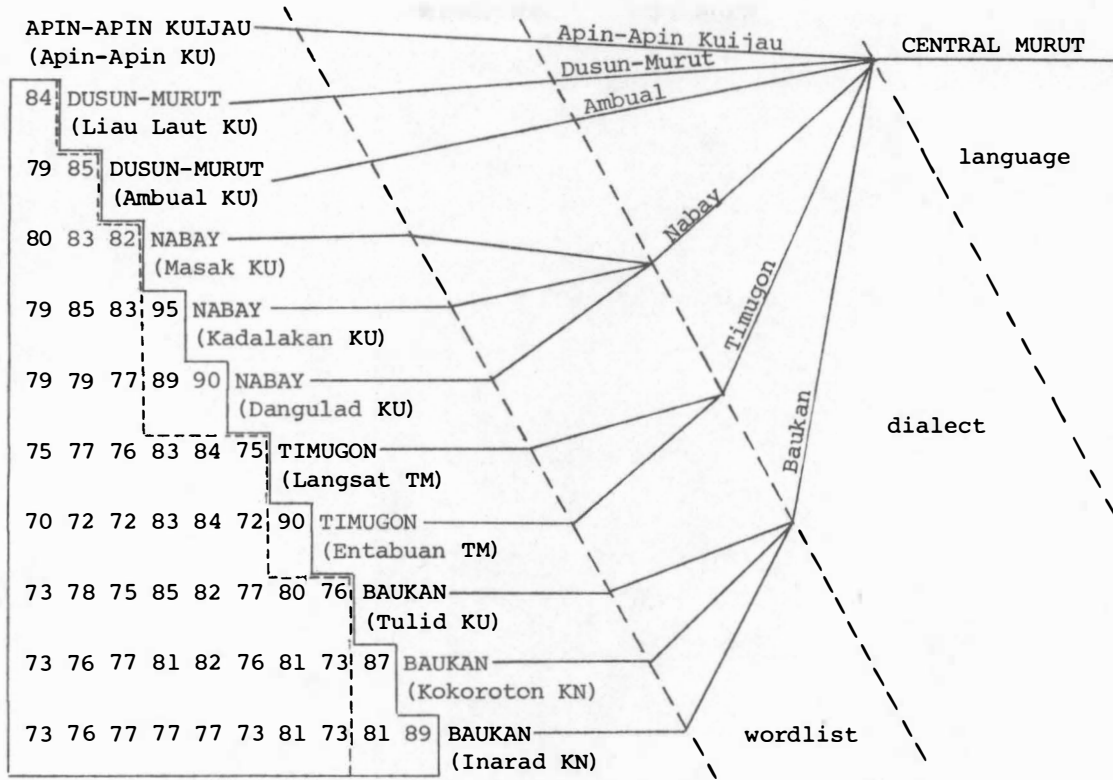


Figure 7: Revised PSC relations of six Central Murut dialects

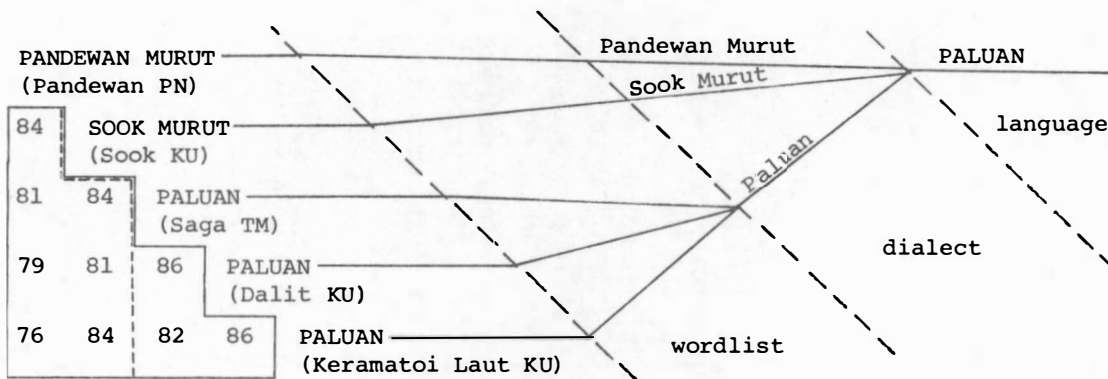


Figure 8: Revised PSC relations and classification of five Paluan wordlists representing three dialects

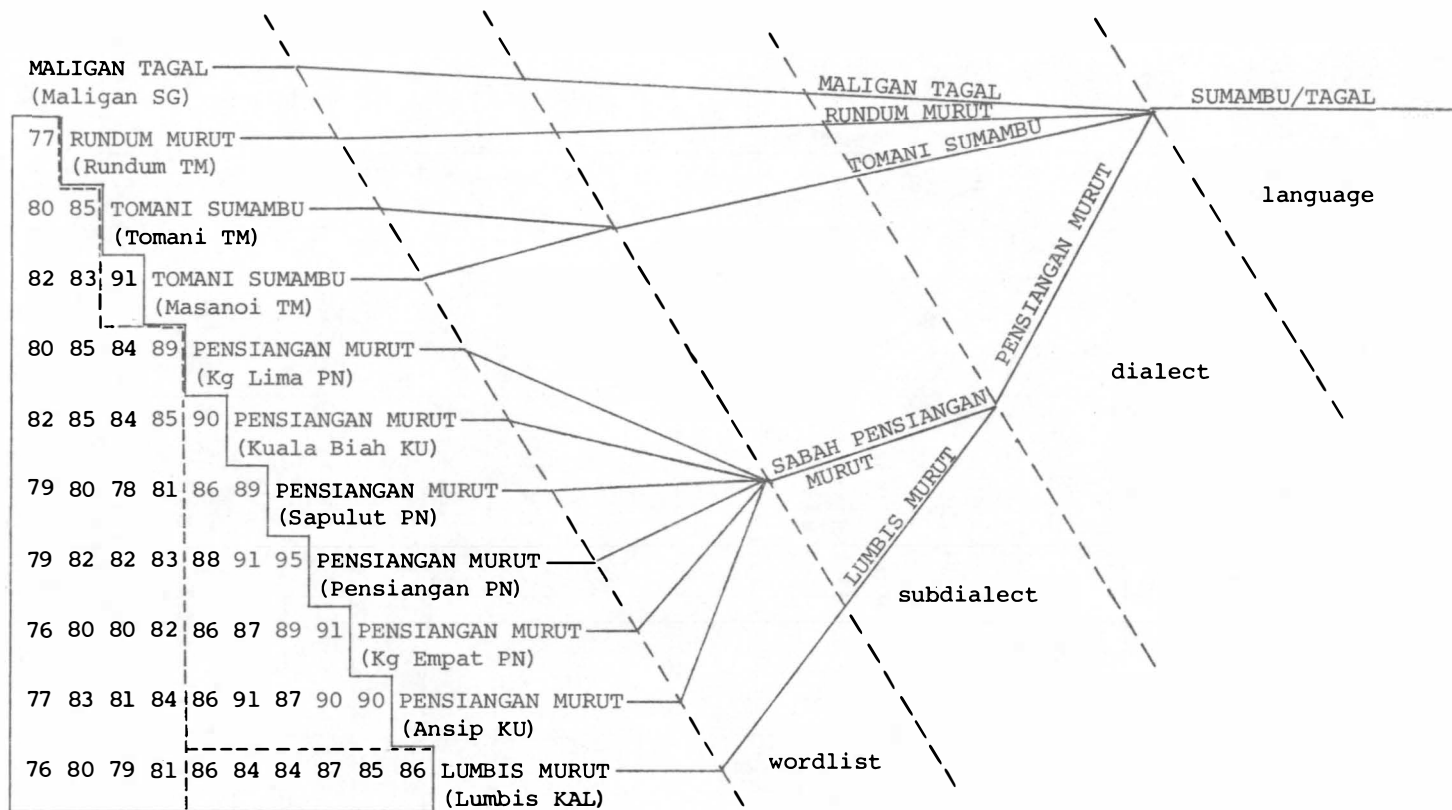


Figure 9: Revised PSC relations and classification of 11 Sumambu/Tagal samples representing four dialects

2. TESTING PROCEDURES

2.1 Purpose of intelligibility testing

The purpose of the intelligibility testing among Murutic languages was to clarify some of the distinctions made by Smith in his lexicostatistical classification. Since intelligibility testing is not precise enough to determine subfamily or subdialect divisions, this testing was limited in purpose to clarifying the language and dialect divisions.

At the end of the first phase of the survey, Apin-Apin Kuijau was specified as needing further clarification to determine its true relation with the Dusunic Kuijau language and the other Murutic languages.

The level of intelligibility between the two Tagal dialects as classified by Smith also needed clarification, since the PSC relation of 70 between the two most remote wordlists is lower than normally expected for mutual intelligibility. The dialect division, or divisions, needed to be more clearly defined. In addition, it was expected that the testing would help to determine which dialect was most widely understood.

Finally, Smith mentioned the need to determine intelligibility among the languages of the Central Murut subfamily. Were they all distinct languages or could some be considered dialects of the same language? Which would be considered central, and which are 'fringe' languages? These are some of the questions the dialect intelligibility testing sought to answer.

Before beginning the testing phase of the Murut survey it was considered advisable to include languages and test points other than those mentioned by Smith. Some of these included testing at Kalabakan TU and Serudung Baru TU both to determine their level of mutual intelligibility and their level of intelligibility of other Murutic languages and the Tidong language. Motivation for the latter purpose came mostly from growing suspicion that these villages could, in fact, be more closely related to the Tidong language than to the Murutic languages (Prentice 1970; Sather 1972).

In addition, some Central Murut reference tapes were used for testing other Murutic subfamilies: Gana, Apin-Apin Kuijau, Kalabakan Murut, Serudung Murut and Tagal. Tagal reference tapes were also used in these as well as in all Central Murut languages. This testing was done in order to determine the mutual intelligibility within the whole Murutic family.

The addition of new data not included in Smith made additional testing necessary since the relative lexical position of the new samples was not known.⁸

2.2 Test points and test sets

The intelligibility testing phase of the survey of the Murutic languages was conducted between the months of December 1980 and June 1981. Six members of the Summer Institute of Linguistics were involved at various times, the writer of this paper being involved in each test. Twenty-nine⁹ tests were done in Murutic language family villages as well as several with the other language families.

It was decided to do intelligibility testing in at least two villages within each language including one from each dialect. This was accomplished in every case except Apin-Apin Kuijau, Kalabakan Murut, Serudung Murut, Takapan

and Beaufort Murut. In all but the last case only one village could be located.¹⁰ Figure 10 gives an overall picture of the Murutic testing by test point showing the languages of each reference tape used in the test set.¹¹

Kolod and Sembakung Murut were excluded from the testing since their cultural and linguistic centres were considered to lie outside the political boundary of Sabah.

Each test set for the Murutic test points consisted of a hometown tape, four or five reference tapes and the national language tape. For those test points for which no hometown tape was available, such as those villages not previously visited, an appropriate substitution was made. It is still considered a reference tape for purposes of intelligibility scores, but may reflect a slightly higher score than expected. Test sets, which included not only Murutic, but also Dusunic reference tapes, were prepared for the Gana and Apin-Apin Kuijau languages, since one of the things under investigation was whether or not these languages had been properly placed in the Murutic rather than the Dusunic family. A Tidong reference tape was included in the Serudung Murut test set to check the possibility of a better understanding of Tidong than Murutic languages.

For most other tests, a standard Murut test set was prepared including a reference tape from each of the following languages: Tagal, Timugon, Baukan and Nabay. It was expected that one of these languages would be more widely understood than the others. In a few cases this was not followed, however. At some of the previously unclassified test points and Tagal test points enough data was already available to project an intelligibility level for the whole group.

2.3 Problems involved

Figure 11 gives an assessment of the reference tapes used in the Murut testing, first as to technical aspects such as tape quality and then as to the content of the story itself. The value of this rating will be seen as it is used to explain some of the more unexpected results in the following section.

REFERENCE TAPES	KADAZAN/ DUSUN	KUIJAU	GANA	APIN-APIN KUIJAU	NABAY	BAUKAN	TIMUGON	BEAUFORT MURUT	SOOK MURUT	TAKAPAN	PALUAN	TAGAL 'A'	TAGAL 'B'	KALABAKAN MURUT	SERUDUNG MURUT	TIDONG
TEST POINTS																
1. KUIJAU Senagang KU			1 x													
2. GANA Minansut KU *Kuangoh KU	4 xx xx	1 x		2 x x	2 x x							1 x				
3. APIN-APIN KUIJAU Apin-Apin KU	1 x			1 x	1 x	1 x	1 x					1 x				
4. DUSUN-MURUT Ambual KU Liau Laut KU					2 x x	2 x x	2 x x					2 x x				
5. NABAY Dangulad KU Masak KU					2 x x	2 x x	2 x x					2 x x				
6. BAUKAN Inarad KN Tulid KU					2 x x	4 xx xx	2 x x					2 x x				
7. TIMUGON Entabuan TM Langsat TM					2 x x	2 x x	1 x	1 x				2 x x	1 x			
8. BEAUFORT MURUT Bukau BT					1 x x	1 x x	1 x x					1 x				
9. SOOK MURUT Sook KU *Pandewan PN					1 x x	2 x x	1 x		1 x	2 x		1 x	1 x			
10. TAKAPAN Keramatoi KU *Dalit KU					1 x x	2 x x	1 x		1 x	1 x	1 x	2 x x				
11. PALUAN Saga TM					1 x	1 x	1 x					1 x	1 x			
12. TAGAL 'A' *Kg Empat PN Pensiangan PN Sapulut PN					1 x		1 x			2 x		4 xx x x	5 xx xx			
13. TAGAL 'B' Kg Lima PN Kuala Biah KU Maligan SG Tomani TM					3 x x x	2 x x x	4 x x x			1 x		4 x x x	3 x x			
14. KALABAKAN MURUT Kalabakan TU					1 x	1 x						1 x			1 x	
15. SERUDUNG MURUT Serudung Baru TU					1 x	1 x						1 x		1 x		1 x
16. PREVIOUSLY UNCL. *Selungai PN										1 x		1 x	2 xx			
17. TIDONG Indarasabak TU						1 x										1 x
18. LUNDAYEH Kawang SG Mendolong SG												2 x x				

* Previously unclassified; new data

Numbers in the boxes indicate the number of times a certain language was tested in a given village. The x in the boxes indicates reference tapes used in each test point.

Note: 'Mini-tests' were carried out at Masanoi TM (for Ansip KU, Kg Lima PN); Kg Lima PN (Lumbis KAL); Rundum TM (Ansip KU, Kg Lima PN, Sapulut PN) and Pensiangan PN (Lumbis KAL), but the corpus is too small and select to be considered on the same level as those listed above. They are mentioned in the discussion where relevant.

Figure 10: Test sets used at each test point having Murutic reference tapes

REFERENCE TAPE	TAPE AND SPEECH QUALITY	STORY CONTENT	RATING
Ansip KU	Good	One hard-to-translate question	8
Apin-Apin KU	Not clear	Two hard questions; reduced to eight questions	8
Bukau BT	Good	Short statements	7
Bundu Tuhan RU	Good	Simple; uses Malay	6
Bunduon PG	Good	Good; substantial content	10
Dangulad KU	Not clear	Fair	7
Entabuan TM	Good, but slow speech	Fair; two weak questions	7
Inarad KN	Good	One redundant question; reduced to nine questions	8
Kadalakan KU	Good	Two redundant questions	7
Kalabakan TU	Good	Simple and short	6
Kg Lima PN	Good, but fast speech	Very good	9
Keramatoi Laut KU	Good	Good	8
Kokoroton KN	Good	Good	10
Langsat TM	Good	Fair; somewhat simple	7
Lidung LS	Good	Uses Malay; reduced to nine questions	8
Minansut KU	Good	Good	8
Pensiangan PN	Good	Very good	10
Rundum TM	Good	Good	9
Sapulut PN	Good	Good; one tricky question	9
Serudung Baru TU	Good	Simple and short	7
Sook KU	Good	Fair; one weak question	7
Tulid KU	Good	Very good; one weak question	9
Tuntumulod KU	Good	Good	9

Figure 11: Rating of the reference tapes used in the Murut testing on a scale of 1 to 10, 1 being very poor and 10 being very good

3. TEST RESULTS

The results of dialect intelligibility testing at 29 Murutic test points are discussed in this section and compared to Smith's classification of Murutic languages. Maps showing the location of the test points and reference points and the approximate language/dialect boundaries as delimited in this paper are found in Section 4. The sociological data for each test point is found in Section 5. The nomenclature of both the linguistic designations and the PSC relations found in this section follow Smith (in this volume).

3.1 Gana language

Intelligibility testing was done in two Gana villages, Minansut KU and Kuangoh KU. Reference tapes from Murutic languages included Gana, Apin-Apin Kuijau, Tagal and Nabay. The Dusunic languages tested were Kuijau and Kadazan/Dusun.

Figure 12 displays the results of intelligibility testing for the two Gana test points. As can be seen from this chart, the test results on Apin-Apin Kuijau, Tagal and Kadazan/Dusun from Bunduon PG are quite consistent with their respective PSC relations. However, the results on Nabay, Kuijau and Kadazan/Dusun from Bundu Tuhan RU are significantly higher than their respective PSC relations.

REFERENCE TAPES TEST POINTS	MURUTIC			DUSUNIC		
	APIN-APIN KUIJAU (Apin-Apin KU)	TAGAL (Ansip KU)	NABAY (Kadalakan KU)	KUIJAU (Tuntumulod KU)	KADAZAN/DUSUN (Bunduon PG)	KADAZAN/DUSUN (Bundu Tuhan RU)
GANNA (Minansut KU)	73 (75)	64 (57)	92 (68)	-- (64)	74 (63)	88 (58)
(Kuangoh KU)	71	--	91	99	67	91

Figure 12: Gana intelligibility of three Murutic and three Dusunic reference tapes. (Scores are given as percentages. PSC relations are given in parentheses.)

In the case of Nabay, the geographical proximity of the Gana and Nabay language areas as well as the redundancy of the Nabay story from Kadalakan KU (Figure 11) likely contribute to the high intelligibility testing results. Nabay intelligibility of Gana was not tested.

The Dusunic language tape representing the Kuijau language, from Tuntumulod KU was also well understood by Gana speakers at Kuangoh KU (99%). Since the Gana and Kuijau language areas are also in close proximity to each other, it is likely that this high level of intelligibility is due to language learning. Kuijau speakers from Senagang KU were also tested for their ability to understand Gana and averaged 82% intelligibility, indicating that this apparent language learning goes in both directions (→ Kuijau).

The very high intelligibility testing results shown by Gana speakers on the Bundu Tuhan RU story (88% and 91%) can only be explained by the low rating of the Bundu Tuhan RU story (Figure 11).

In Figure 13 the Gana data is averaged and displayed according to geography. The results of the Bundu Tuhan RU reference tape are not included. The chart does not finally indicate whether Gana is more Murutic than Dusunic - in fact, it might best be thought of as a borderline language, but it can be said that intelligibility is highest with those languages in its immediate vicinity. Tagal, as spoken in Pensiangan District, is the most remote of the languages tested; Kadazan/Dusun, as spoken in Penampang District, ranks second, followed closely by Apin-Apin Kuijau. The two Keningau languages of Nabay and Kuijau, neither of which has high PSC relations with Gana villages, are most highly intelligible to Gana speakers.

PENSIANGAN DISTRICT		KENINGAU DISTRICT			PENAMPANG DISTRICT	
TAGAL	A-A KUIJAU	NABAY	KUIJAU	CENTRAL DUSUN		
64 (57)	72 (75)	92 (68)	99 (64)	71 (63)		
MURUTIC			DUSUNIC			
← Intelligibility →						

Figure 13: Gana intelligibility of three Murutic reference tapes and two Dusunic reference tapes by geography. (Intelligibility scores are given as percentages. PSC relations are given in parentheses.)

Intelligibility testing showed that Gana has both high and low intelligibility of some Murutic and some Dusunic groups, and that geography more than linguistic affiliation seems to affect its level of intelligibility.

3.2 Apin-Apin Kuijau language

Apin-Apin Kuijau is the name given to that language spoken by about one-fourth of the population of the village of Apin-Apin KU. From the information received, there are no other villages which speak exactly like this linguistic community. Reference tapes from four Murutic languages, including Tagal, Timugon, Baukan and Nabay, and one Dusunic reference tape from the Kuijau language were tested there.

Figure 14 displays the results of intelligibility testing for the single Apin-Apin Kuijau test point. The intelligibility testing results on all of the tapes except for the Baukan language tape from Kokoroton KN, are significantly higher than the respective PSC relations. In the case of the Tagal, Timugon and Nabay tests, since no cross-testing was done, one can only conjecture as to the reasons for the apparent discrepancies. It seems likely that because of the variety of languages spoken in Keningau District and even around the area of the village of Apin-Apin KU, that Apin-Apin Kuijau speakers have had opportunity to learn these Murutic languages. The much lower score on the Kokoroton KN tape is still understandable since that language area is quite remote from the others.

The high intelligibility testing results on the Dusunic Kuijau tape can also be explained by language learning on the part of Apin-Apin Kuijau speakers. This is substantiated by the results of cross-testing Dusunic Kuijau speakers' understanding of Apin-Apin Kuijau. The results of that cross-testing are displayed in Figure 15. The results ranged 66-72% intelligibility compared to a range of 64-70 PSC. These low scores show that intelligibility between the two is not mutual which in turn favors language learning as the cause of the extremely high score attained by Apin-Apin Kuijau speakers of the Dusunic Kuijau tape.

REFERENCE TAPES TEST POINTS	MURUTIC				DUSUNIC
	TAGAL (Ansip KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)	NABAY (Kadalakan KU)	KUIJAU (Tuntumulod KU)
APIN-APIN KUIJAU (Apin-Apin KU)	91 (60)	79 (63)	47 (68)	98 (70)	94 (67)

Figure 14: Apin-Apin Kuijau intelligibility of four Murutic reference tapes and one Dusunic reference tape. (Scores are given as percentages. PSC relations are given in parentheses.)

REFERENCE TAPE TEST POINTS	APIN-APIN KUIJAU (Apin-Apin KU)
KUIJAU (Dusunic) (Tuntumulod KU)	66 (67)
(Sungoi KU)	70 (70)
(Liau Darat KU)	70 (69)
(Senagang KU)	69 (68)
(Linsosod KU)	72 (64)

Figure 15: Gana and Kuijau intelligibility of the Apin-Apin Kuijau reference tape. (Scores are given as percentages. PSC relations are given in parentheses.)

3.3 Dusun-Murut language

The villages of Ambual KU and Liau Laut KU were chosen to represent the Dusun-Murut language for the intelligibility testing. The standard Murut test set used included Tagal, Timugon, Baukan and Nabay reference tapes. Figure 16 displays the results of intelligibility testing for both Dusun-Murut test points.

REFERENCE TAPES TEST POINTS	MURUTIC			
	TAGAL (Ansip KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)	NABAY (Kadalakan KU)
DUSUN-MURUT (Ambual KU)	81 (61)	71 (68)	59 (76)	92 (80)
(Liau Laut KU)	63 (64)	65 (72)	59 (70)	77 (80)

Figure 16: Dusun-Murut intelligibility of four Murutic reference tapes. (Scores are given as percentages. PSC relations are given in parentheses.)

As can be seen from Figure 16, Ambual KU subjects consistently scored a higher average intelligibility than Liau Laut KU subjects on all tapes except for the Baukan tape from Kokoroton KN where both scored the same level of intelligibility. Since no intelligibility testing was done between Ambual KU and Liau Laut KU it is impossible to know whether the different levels of comprehension by these two villages is due to language learning on the part of some of the subjects tested at Ambual KU or to a greater linguistic distinction between the two villages than is indicated by the 80 PSC between them.

From these limited test results it is only possible to conclude that the two Dusun-Murut test points represent a language or languages different from Tagal of Ansip KU, Timugon of Entabuan TM and Baukan of Kokoroton KN. The test results on the Nabay reference tape are marginal. Because of that and the 80 PSC shared between each of the Dusun-Murut samples and Nabay from Kadalakan KU it is possible that Dusun-Murut is a dialect of the Nabay language. Further testing and cross-testing would need to be done in order to confirm this.

3.4 Nabay language

The two Nabay test points Dangulad KU and Masak KU were chosen to demonstrate the Nabay community's intelligibility of the standard Murut test set including one reference tape each from Tagal, Timugon, Baukan and Nabay languages. The only difference between the two test sets was the different Nabay reference tape used.

Masak KU was tested with the reference tape from the other Nabay test point, Dangulad KU. The results (84%) indicate that the two samples represent the same language but the degree of difference may also indicate a dialect distinction between the two.

Figure 17 displays the results of intelligibility testing for the Nabay test points.

REFERENCE TAPES TEST POINTS	MURUTIC				
	NABAY (Dangulad KU)	NABAY (Kadalakan KU)	TAGAL (Ansip KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)
NABAY (Dangulad KU)	+100 (100)	89 (85)	90 (69)	80 (69)	76 (76)
(Masak KU)	84 (84)	+99 (89)	78 (67)	75 (75)	56 (78)

Figure 17: Nabay intelligibility of five Murutic reference tapes.
(Scores are given as percentages. PSC relations are given in parentheses. + indicates hometown tape.)

Figure 18 displays the results of cross-testing in Tagal, Timugon and Baukan villages to determine their understanding of Nabay.

REFERENCE TAPE TEST POINTS		NABAY (Kadalakan KU)
TAGAL	Kuala Biah KU	98 (66)
	Maligan SG	80 (60)
	Tomani TM	77 (61)
	Sapulut PN	74 (67)
TIMUGON	Entabuan TM	78 (75)
	Langsat TM	77 (79)
BAUKAN	Tulid KU	98 (78)
	Inarad KN	74 (75)

Figure 18: Tagal, Timugon and Baukan understanding of Nabay.
(Scores are given as percentages. PSC relations are given in parentheses.)

The results of Nabay-Tagal testing are quite high in both directions when compared to respective PSC relations. The fact of the dominance of these two languages throughout the Murut language area may help to explain the higher intelligibility scores than would have been expected based on the PSC figures alone.

The Nabay-Timugon and Nabay-Baukan test results are quite in line with the respective PSC relations except for the very high 98% scored by Baukan subjects from Tulid KU on the Nabay test tape. This one high score cannot be adequately explained with the present data. Disregarding that one high score then, intelligibility testing confirms the lexicostatistical classification of Nabay as a distinct language from Baukan and Timugon.

3.5 Baukan test points¹²

The two Baukan test points chosen for intelligibility testing included Tulid KU and Inarad KN, representing both extremes of the language group as delimited by Smith (in this volume). In addition to the standard Murut test set, including a Tagal, Timugon, Baukan and Nabay reference tape, a second Baukan reference tape was included in the test set. Therefore, including the hometown tape, three Baukan tapes were tested at each test point.

Figure 19 shows the results of intelligibility testing at the two Baukan test points.

REFERENCE TAPES TEST POINTS	MURUTIC					
	BAUKAN (Inarad KN)	BAUKAN (Kokoroton KN)	BAUKAN (Tulid KU)	TAGAL (Ansip KU)	TIMUGON (Entabuan TM)	NABAY (Kadalakan KU)
BAUKAN (Inarad KN)	+100 (100)	80 (89)	75 (79)	74 (71)	72 (71)	74 (75)
(Tulid KU)	99 (79)	97 (85)	+100 (100)	86 (68)	74 (73)	98 (78)

Figure 19: Baukan intelligibility of six Murutic reference tapes.
(Scores are given as percentages. PSC relations are given in parentheses. + indicates hometown tape.)

The mutual intelligibility testing within Baukan showed that in spite of a high PSC relation in the case of Inarad KN and Kokoroton KN, high intelligibility was not demonstrated and also that Tulid KU consistently showed notably higher intelligibility of the other dialects than Inarad KN did.

Inarad KN showed a marginal 80% intelligibility on the Kokoroton KN tape. Hurlbut and Pekkanen (to appear) recorded only 75% intelligibility of the same

Kokoroton KN reference tape by Inarad KN subjects while Kokoroton KN subjects averaged only 71% on the Inarad KN reference tape. This lack of mutual intelligibility, despite the high PSC, necessitates a dialect distinction between the language spoken in Kokoroton KN and that spoken in Inarad KN.

On the Tulid KU reference tape, Inarad KN registered only 75% intelligibility. However Tulid KU demonstrated near perfect comprehension (99%) on the Inarad KN reference tape. Given a revised PSC relation of *81 the results of the intelligibility testing indicate that a dialect level division should be made separating the two samples. But with intelligibility shown in only one direction a language level distinction is too extreme without further evidence.¹³

Tulid KU subjects also showed 97% intelligibility of the Kokoroton KN reference tape. The Baukan intelligibility scores reverse the interrelation indicated by the PSC scores of samples by linking Kokoroton KN closer to Tulid KU than to Inarad KN. However, since Tulid KU showed equally high intelligibility of the Inarad KN reference tape, a Kokoroton KN test of a Tulid KN story is needed to confirm this. Tentatively, Kokoroton KN and Tulid KU can be grouped together as members of a single dialect.¹⁴

In Figure 19 results of Tagal, Timugon and Nabay testing at Baukan villages can also be seen. Cross-testing results of these languages are shown in Figure 20. All of these scores are either quite consistent with their respective PSC relations or much lower than respective PSC figures, confirming the lexicostatistical classification of Baukan as a separate language from Tagal, Timugon and Nabay. The two aberrant scores at Tulid KU on the Tagal reference tape from Ansip KU (86% intelligibility compared with 68 PSC) and on the Nabay reference tape from Kadalakan KU (98% intelligibility compared with 78 PSC) are likely due to contact with Tagal and Nabay speakers in the Keningau area.

TEST POINTS		REFERENCE TAPE	BAUKAN
TAGAL	Kuala Biah KU		72-K (69)
	Tomani TM		61-T (63)
TIMUGON	Entabuan TM		63-T (73)
	Langsat TM		46-T (75)
NABAY	Dangulad KU		76-K (76)
	Masak KU		56-K (78)

Figure 20: Cross-testing results of Tagal, Timugon and Nabay villages with a Baukan language reference tape. (Scores are given as percentages. PSC relations are given in parentheses. -K indicates the Baukan reference tape from Kokoroton KN was used and -T indicates the Baukan reference tape from Tulid KU was used in the testing.)

3.6 Timugon test points

The two Timugon test points Entabuan TM and Langsung TM were classified by Smith as representing a single dialect. The test sets included reference tapes from Timugon, Tagal, Sook Murut, Beaufort Murut, Baukan and Nabay.

Figure 21 shows the results of testing at the two Timugon test points.

REFERENCE TAPES TEST POINTS	MURUTIC						
	TIMUGON (Langsat TM)	TAGAL 'A' (Sapulut PN)	TAGAL 'B' (Kg Lima PN)	SOOK MURUT (Sook KU)	BEAUFORT MURUT (Bukau BT)	BAUKAN (Tulid KU)	NABAY (Kadalakan KU)
TIMUGON (Entabuan TM)	85 (85)	47 (68)			80 (77)	63 (73)	78 (75)
(Langsat TM)	+97 (100)	58 (71)	43 (66)	71 (74)		46 (75)	77 (79)

Figure 21: Timugon intelligibility of seven Murutic reference tapes. (Scores are given as percentages. PSC relations are given in parentheses. + indicates hometown tape.)

Figure 22 shows the results of cross-testing the same language groups with a Timugon reference tape.

In the Timugon-Tagal testing all of the test results are consistent with or significantly lower than respective PSC relations except in the case of the high understanding of the Timugon story by Tagal test subjects at Kuala Biah KU (88%). This one aberrant score is likely due to language learning so that the distinctiveness between the Timugon and Tagal languages is still maintained.

Timugon-Sook Murut/Takapan/Paluan testing also showed one very high score (90% scored at Sook KU) while all of the other scores are again consistent with PSC relations (→ Section 3.8). The language distinction between Timugon and Sook Murut is still maintained on the basis of the other more consistent scores.

Timugon-Beaufort, Timugon-Nabay (→ Section 3.4) and Timugon-Baukan (→ Section 3.5) test results are quite consistent with respective PSC relations so that the classification of these three languages as being distinct from Timugon is maintained.

REFERENCE TAPE		TIMUGON (Entabuan TM)
TEST POINTS		
TAGAL	Kuala Biah KU	88 (68)
	Tomani TM	76 (65)
	Pensiangan PN	71 (67)
	Kg Lima PN	67 (67)
	Maligan SG	65 (61)
SOOK MURUT	Sook KU	90 (71)
	Pandewan PN	73 (--)
BEAUFORT MURUT	Bukau BT	74 (77)
BAUKAN	Tulid KU	74 (73)
	Inarad KN	72 (71)
NABAY	Dangulad KU	80 (69)
	Masak KU	75 (75)

Figure 22: Cross-testing results in Tagal, Sook Murut, Beaufort Murut, Baukan and Nabay villages with a Timugon reference tape. (Scores are given as percentages. PSC relations are given in parentheses.)

3.7 Beaufort Murut test point

Bukau BT was the single test point chosen for dialect intelligibility testing in the Beaufort Murut language. The standard Murut test set, including reference tapes from Tagal, Timugon, Baukan and Nabay was used.

Figure 23 displays the results of the intelligibility testing conducted at Bukau BT.

REFERENCE TAPES TEST POINT	MURUTIC			
	TAGAL (Ansip KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)	NABAY (Kadalakan KU)
BEAUFORT MURUT (Bukau BT)	82 (64)	74 (77)	51 (73)	87 (72)

Figure 23: Beaufort Murut intelligibility of four Murutic reference tapes. (Scores are given as percentages. PSC relations are given in parentheses.)

While Beaufort Murut has low PSC relations with Tagal and Nabay (64 and 72 respectively), Bukau BT test subjects showed dialect-level intelligibility of those two languages. Since cross-testing was not done in either case, it is not known whether this is due to language learning on the part of Bukau BT test subjects or to a closer relationship between the languages than the PSC relations indicate. Because of this, Beaufort Murut will still be considered a separate language from Nabay and Tagal.

Results of testing Beaufort Murut with Timugon (→ Section 3.6) and with Baukan are consistent with or lower than their respective PSC figures so that language distinctions between the three are maintained.

3.8 Sook Murut, Takapan and Paluan language

Based on the additional data gathered after Smith's classification (→ Section 1, note 6) and the results of intelligibility testing, the discussions of Sook Murut, Takapan and Paluan will be handled under a single heading.

Testing of the Sook Murut language was done in two villages. Pandewan PN had not previously been visited, but Sook KU had been included in Smith's data. Two tests were conducted in the Takapan language group, the first at Dalit KU, and the second at Keramatoi KU (→ Section 2, note 10). The Paluan language was represented by a single test at Saga TM.

The standard Murut test set, including a reference tape from Tagal, Timugon, Baukan and Nabay, was used in most cases with the addition of a Sook Murut, Takapan or Paluan tape.

Figure 24 displays the results of intelligibility testing for Sook Murut, Takapan and Paluan test points. Revised PSC relations (Figure 8) and high intelligibility scores in the few cases tested within Sook Murut, Takapan and Paluan are evidence for reclassifying these languages as dialects of a single language. Further cross-testing is needed to determine if these dialects are mutually intelligible or if the high scores are a result of language learning.

REFERENCE TAPES TEST POINTS	PALUAN (Saga TM)	TAKAPAN (Keramatoi Laut KU)	SOOK MURUT (Sook KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)	BAUKAN (Tulid KU)	NABAY (Kadalakan KU)	TAGAL 'A' (Ansip KU)	TAGAL 'B' (Kg Lima PN)
PALUAN (Saga TM)	+97 (--)			74 (70)		76 (68)	89 (66)	93 (79)	75 (78)
TAKAPAN (Keramatoi KU)		+100 (--)		87 (69)	78 (71)		96 (71)	96 (73)	
(*Dalit KU)	100 (80)	+100 (83)			76			96	
SOOK MURUT (Sook KU)		96 (79)	+100 (--)	90 (74)	85 (75)		98 (75)	96 (72)	
(*Pandewan PN)		80	+90	73		65			57

Figure 24: Takapan, Paluan and Sook Murut intelligibility of various Murutic language tapes. (Scores are given as percentages. PSC relations are in parentheses. * indicates previously unclassified; new data. + indicates tape used as hometown tape in given test set.)

The Timugon reference tape from Entabuan TM was included in the Sook Murut test sets Sook KU and Pandewan PN, the Takapan test set Keramatoi KU and the Paluan test set Saga TM. The PSC relations range 69-71. Intelligibility scores ranged 73-90%. The highest was recorded at Sook KU and the lowest at Pandewan PN. It is interesting to note that both of these had been grouped as Sook Murut. Pandewan PN represents one geographical extreme of the Sook Murut/Takapan/Paluan group and Saga TM represents the other, and both of these have intelligibility scores consistent with their PSC relations.

A Sook Murut reference tape from Sook KU was included in the Timugon test set at Langsung TM (+ Section 3.6). Subjects there averaged 71% which together with a 74 PSC relation distinguishes them as separate languages. Timugon intelligibility of Takapan or Paluan was not tested, but there is no evidence to question the language distinction here.

The Baukan reference tape from Kokoroton KN was included in the Takapan test sets at Keramatoi KU and Dalit KU and in the Sook Murut test set at Sook KU. The Baukan reference tape from Tulid KU was included in the Paluan test set at Saga TM and the Sook Murut test set at Pandewan PN. No significant difference was registered either between reference tapes or language groups. Intelligibility ranged 65-85%. The PSC relations ranged 68-76. As in the case of the Timugon reference tape, the highest score was recorded at Sook KU and the lowest at Pandewan PN, both grouped as Sook Murut villages.

Considering the proximity of Baukan villages to Sook KU, the relatively high intelligibility is understandable. In fact, the 85% intelligibility score together with the 75 PSC relation would allow them to be grouped as representing dialects of one language. However, Sook Murut is better grouped together with Takapan and Paluan, both on the basis of PSC relations and intelligibility scores. Takapan, Paluan and Baukan languages share PSC relations as remote as 68-76 PSC and intelligibility scores of only 76-78%. Taken altogether it is not plausible to include Sook Murut and Baukan in one language. Baukan speakers' intelligibility of Sook Murut/Takapan/Paluan still needs to be tested to more fully understand how far intelligibility extends from the point of Baukan.

The Nabay reference tape from Kadalakan KU was used at the Sook Murut test point Sook KU, the Takapan test point Keramatoi KU and the Paluan test point Saga TM. The intelligibility scores ranging 89-98% are very high compared to the relations of 66-75 PSC. The scores may be adjusted down somewhat due to the redundancy of the tape (Figure 11), but the scores still show appreciable intelligibility. Because of the low PSC relations and the lack of intelligibility results showing Nabay speakers' comprehension of Paluan, no redefinment of the classification between the two will be attempted.

Tagal 'A' and Tagal 'B' reference tapes were used at several Sook Murut, Takapan and Paluan test points. The Tagal 'A' tape from Ansip KU was included in the Sook KU, Keramatoi KU, Dalit KU and Saga TM test points. Scores ranged 93-96% compared to PSC relations of 72-79. The Tagal 'B' tape from Kg Lima PN was used in Pandewan PN and Saga TM. Intelligibility averaged 51% and 75% respectively. The PSC relations, using Sook KU to represent Pandewan PN, were 67 PSC and 78 PSC.

The Takapan reference tape from Keramatoi Laut KU was included in the Tagal 'A' test sets at Kg Empat PN and Sapulut PN (Figure 25). Intelligibility scores were only 49% and 63% respectively. The Sapulut PN sample has 72 PSC. The same tape was included in the Tagal 'B' test set at Maligan SG. Intelligibility registered there was only 59% which is comparable to the 62 PSC.

Important in this testing are the high intelligibility scores received by Sook Murut/Takapan/Paluan on Tagal 'A' stories and the lack of significant intelligibility shown on the Tagal 'B' story as well as the lack of Tagal 'A' and Tagal 'B' intelligibility of Takapan. This indicates the influence of Tagal 'A' but fails to demonstrate linguistic affinity between Tagal dialects and Sook Murut/Takapan/Paluan. No redefinment of the classification of this relationship is offered at this time, but future intelligibility testing of Tagal by Sook Murut and Paluan may show the need for it.

Using the above data, the three groups labelled in Smith as Sook Murut, Takapan and Paluan are considered to represent a single language since high intelligibility among them was shown in every case tested. The PSC relations give a fair indication of the relative intelligibility between these three dialects and any other dialect.

The Timugon and Baukan dialects are ruled out as possible members of the same language as Sook Murut, Takapan and Paluan. Tagal 'B' is also ruled out. However, Tagal 'A' seems to be understood very well, though intelligibility is not mutual in the case of Takapan. For that reason, it should not be considered a member of the same language as Sook Murut, Takapan and Paluan. The Nabay dialect may be candidate as a member of the same language group but Nabay intelligibility of Sook Murut, Takapan and Paluan has not yet been tested and is thus left as belonging to a separate language. Nabay appears to be an important dialect in the area, so language learning by the other dialects is not surprising.

3.9 Tagal test points

Seven test points were chosen for the Tagal intelligibility testing, four in Pensiangan District and one each in Keningau, Tenom and Sipitang districts. The major motivation for testing in the Tagal language was to determine intelligibility between the two dialects of Tagal as well as to investigate comprehension within each dialect. The Tagal 'A' test points included Kg Empat PN, Pensiangan PN and Sapulut PN. Tagal 'B' test points consisted of Kg Lima PN, Kuala Biah KU, Maligan SG and Tomani TM. The testing included both Tagal and non-Tagal Murutic reference tapes.

3.9.1 Intelligibility within the Tagal language

Figure 25 displays the results of the intelligibility testing within the Tagal language. The Tagal 'A' reference tape from Sapulut PN was included in the Tagal 'A' test set at Kg Empat PN. The tape from Pensiangan PN was included in the Kg Empat PN and Sapulut PN test sets. The Ansip KU reference tape was included in the Pensiangan PN test set. The results showed intelligibility of 88-93%. This is slightly higher than the 84-89 PSC relations. None of the subjects' individual scores were below 70%.

REFERENCE TAPES		TAGAL 'A'			TAGAL 'B'	
		Sapulut PN	Pensiangan PN	Ansip KU	Kg Lima PN	*Rundum TM
TEST POINTS						
TAGAL 'A'	*Kg Empat PN	88 (--)	91 (--)		78 (--)	79 (--)
	Sapulut PN		90 (89)		69 (77)	
	Pensiangan PN			93 (84)	80 (80)	80 (--)
TAGAL 'B'	Kg Lima PN		87 (80)			81 (--)
	Kuala Biah KU			99 (82)		
	Maligan SG		89 (74)			93 (--)
	Tomani TM	86 (70)			72 (76)	

Figure 25: Tagal 'A' and Tagal 'B' intelligibility of Tagal reference tapes. (Scores are given as percentages. PSC relations are in parentheses. * indicates previously unclassified; new data.)

The Tagal 'B' reference tape from Kg Lima PN was included in the Tagal 'B' test set at Tomani TM while the Rundum TM reference tape was used at both Kg Lima PN and Maligan SG. The four Tagal 'B' test points registered scores ranging 72-93% intelligibility compared to relations of 74-85 PSC within the dialect.

The Tagal 'B' reference tape from Kg Lima PN was included in each of the Tagal 'A' test sets, while the Rundum TM reference tape was included only in the Kg Empat PN and Pensiangan PN test sets. Intelligibility scores ranged 69-80% for Kg Lima PN and 79-80% for Rundum TM. This compares with relations between these same Tagal samples of 77-80 PSC. Since no significant difference was registered one may conclude that Tagal 'A' only marginally understands Tagal 'B' and a dialect division between the two best reflects the data.

Tagal 'B' intelligibility of Tagal 'A' reflects something different, however. The Tagal 'A' reference tape from Sapulut PN was used at the Tagal 'B' village Tomani TM; the Pensiangan PN tape was used at both Kg Lima PN and Maligan SG; and the Ansip KU tape was used at Kuala Biah KU. Intelligibility of Tagal 'A' by these four test corpora ranged 86-99%. This compares with 70-82 PSC relations for the same samples. The very high score of 99% which was registered at Kuala Biah KU using the Ansip KU tape may be due either to their proximity to one another and the fact that they are both migrant Tagal villages in a Nabay and Paluan area or that Kuala Biah's speech is actually closer to Tagal 'A'. Otherwise, the scores ranged 86-89%, still high enough to conclude that intelligibility of Tagal 'A' by Tagal 'B' villages is good, but as mentioned above, the intelligibility of Tagal 'B' by Tagal 'A' speakers is only marginal.

3.9.2 Intelligibility between Tagal and non-Tagal Murutic languages

Figure 26 displays Tagal 'A' and Tagal 'B' understanding of Takapan, Timugon, Baukan and Nabay.

Figure 27 displays the cross-testing results with these languages.

Tagal test subjects showed intelligibility testing results with the Takapan reference tape which were comparable with or lower than respective PSC relations. However, in cross-testing, Takapan test subjects showed an average intelligibility of 96%, which is much higher than their PSC relation of 73. Based on the dominance of the Tagal language in the Murutic areas of Sabah it is likely that Takapan speakers have learned the Tagal language. Since intelligibility is not mutual between Tagal and Takapan, the two are still considered separate languages.

The results of Tagal-Timugon testing are similar to or lower than respective PSC relations. One aberrant score (88% intelligibility compared with 68 PSC) was registered at Kuala Biah KU. All of the other scores confirm the classification of Tagal as a separate language from Timugon. The one high score at Kuala Biah KU could indicate language learning.

Tagal-Baukan testing in general confirms the classification of these as separate language though some language learning on the part of the Tulid KU test subjects is evidenced by the higher average score (86%) registered there.

The results of Tagal-Nabay testing appeared to be quite high in both directions compared with the PSC relations between them. However, except for the one high score of 98% by Tagal test subjects at Kuala Biah KU on the Nabay reference tape from Kadalakan KU (→ Section 3.4) and in the cross-testing, one high score

of 90% attained by Dangulad KU test subjects, all of the other scores in the Tagal-Nabay testing were either marginal or low enough to maintain the classification of the two as separate languages.

REFERENCE TAPES		TEST POINTS				
		TAKAPAN (Keramatoi Laut KU)	TIMUGON (Entabuan TM)	BAUKAN (Kokoroton KN)	BAUKAN (Tulid KU)	NABAY (Kadalakan KU)
TAGAL 'A'	*Kg Empat PN	49 (--)				
	Sapulut PN	63 (72)				74 (67)
	Pensiangan PN		71 (67)			
TAGAL 'B'	Kg Lima PN		67 (67)			
	Kuala Biah KU		88 (68)	72 (69)		98 (66)
	Maligan SG	59 (62)	65 (61)			80 (60)
	Tomani TM		76 (65)		61 (63)	77 (65)

Figure 26: Tagal intelligibility of five non-Tagal Murutic reference tapes. (Scores are given as percentages. PSC relations are given in parentheses. * indicates previously unclassified; new data.)

TEST POINTS		REFERENCE TAPES	TAGAL 'A'	TAGAL 'B'
TAKAPAN	Keramatoi KU		96-A (73)	
	Dalit KU		96-A (--)	
TIMUGON	Langsat TM		58-S (71)	43-L (66)
	Entabuan TM		47-S (68)	
BAUKAN	Tulid KU		86-A (68)	
	Inarad KN		74-A (71)	
NABAY	Dangulad KU		90-A (69)	
	Masak KU		78-A (67)	

Figure 27: Intelligibility testing results at non-Tagal Murutic test points of Tagal reference tapes. (Test results are given as percentages. PSC relations are given in parentheses. -A indicates the Tagal 'A' reference tape from Ansip KU was used and -S indicates the Tagal 'A' reference tape from Sapulut PN was used in the testing. -L indicates the Tagal 'B' tape from Kg Lima PN was used.)

3.10 Selungai Murut test point

The Selungai PN test point had not been classified by lexicostatistics prior to intelligibility testing since no wordlist had been collected.¹⁵ The test set included one Tagal 'A' reference tape, two Tagal 'B' reference tapes and a Takapan tape.

Figure 28 displays the results of intelligibility testing for the single Selungai Murut test point. Since testing was conducted with the presupposition that this dialect was "like Sapulut Murut", the Sapulut PN reference tape was used as the hometown tape. For this reason, the results of 90% intelligibility must be held suspect and do not necessarily represent Selungai Murut speakers' comprehension of Tagal 'A'.¹⁶

The Tagal 'B' reference tapes from Kg Lima PN and Rundum TM were included in the testing with the results of 68% and 64% intelligibility respectively. This represents a substantially lower degree of intelligibility than on the Tagal 'A' tape, even if the latter is adjusted down due to its use as the hometown tape to teach procedure.

REFERENCE TAPES				
TEST POINT	TAGAL 'A' (Sapulut PN)	TAGAL 'B' (Kg Lima PN)	TAGAL 'B' (Rundum TM)	TAKAPAN (Keramatoi Laut KU)
SELUNGAI MURUT (Selungai PN)	+90	68	64	58

Figure 28: Selungai Murut intelligibility of four Murutic reference tapes. (Scores are given as percentages. + indicates hometown tape.)

The final reference tape was from Keramatoi Laut KU representing Takapan. The subjects showed only 58% intelligibility, but individual scores ranged 0-100%.

The results of the limited intelligibility testing of Selungai Murut give direction for further research. It is quite evident that the Northern Murut languages such as Paluan and other Keningau and Tenom dialects are not closely related nor intelligible to the Selungai Murut community. Likewise, the Tagal 'B' dialect spoken along the Indonesian/Sabah border on the Salalir River is not intelligible to them. The Tagal 'A' dialect as spoken around Sapulut PN was the most highly intelligible, but it is probable, with further research in Indonesia, that other dialects will prove even more closely related to Selungai Murut.

Reference made to Selungai by other Murut in Pensiangan District indicated they spoke "Indonesian Murut". The Selungai subjects themselves mentioned Sembakung KAL as being similar. At the same time they also said "Pensiangan Murut" (presumably Smith's Tagal 'A') was only "a little different". This information is from male subjects who have been up and down the river system quite extensively, and consequently have a much broader comprehension of the various dialects.

3.11 Kalabakan Murut language

Intelligibility testing of Kalabakan Murut was conducted at Ulu Kalabakan TU about two miles upstream from Kalabakan TU. The test set included four Murutic reference tapes besides the hometown and national language tapes.

Figure 29 displays the results of intelligibility testing for the single Kalabakan Murut test point.

These test results show clearly that Kalabakan Murut is a distinct language from Tagal, Baukan and Nabay. It is very unlikely that testing Kalabakan Murut with any other Murutic languages would show higher intelligibility since the PSC relations of the samples that were tested show the closest lexicostatistical relationships of all the Murutic languages with which it was compared.

REFERENCE TAPES TEST POINT	SERUDUNG MURUT (Serudung Baru TU)	TAGAL (Pensiangan PN)	BAUKAN (Inarad KN)	NABAY (Kadalakan KU)
KALABAKAN MURUT (Kalabakan TU)				
Male - 5	84	32	73	49
Female - 4	33	5	18	9
Average - 9	60 (70)	20 (68)	48 (77)	31 (68)

Figure 29: Kalabakan Murut intelligibility of four Murutic reference tapes showing different averages between male and female subjects. (Scores are given as percentages. PSC relations are given in parentheses.)

The highest intelligibility that Kalabakan Murut test subjects showed was with Serudung Murut of Serudung Baru TU. While the average intelligibility was only 60%, male subjects showed 84% intelligibility. In cross-testing, Serudung Baru TU test subjects averaged 91% intelligibility of the Kalabakan Murut reference tape. This high score does not however give an accurate picture of Serudung Murut intelligibility of Kalabakan Murut since the test corpus at Serudung Baru TU was made up of only five subjects, four of whom were men. Based on this and on the distinctiveness which the test subjects at both places perceived between the two languages, Kalabakan Murut will continue to be classified as a separate language from Serudung Murut.

3.12 Serudung Murut language

Intelligibility testing was carried out in the Serudung Murut village of Serudung Baru TU. The test set consisted of four Murutic reference tapes and one Tidong tape. Figure 30 shows the results of the intelligibility testing conducted at Serudung Baru TU.

The Serudung Baru TU subject showed a high percentage of intelligibility of the Kalabakan Murut reference tape. It was concluded, however, that the 91% intelligibility score was partly due to the unbalanced ratio of male to female subjects and its limited size and that it could not be used to show the average level of intelligibility. The nine Kalabakan Murut subjects averaged only 60% intelligibility on the Serudung Baru TU reference tape, confirming the classification based on the lexical relation of 70 PSC which placed them in separate language families (→ Section 3.11).

REFERENCE TAPES TEST POINT	MURUTIC				TIDONG
	KALABAKAN MURUT (Kalabakan TU)	TAGAL (Pensiangan PN)	BAUKAN (Inarad KN)	NABAY (Kadalakan KU)	TIDONG (Lidung LS)
SERUDUNG MURUT (Serudung Baru TU)	91 (70)	14 (57)	65 (63)	39 (57)	29 (55)

Figure 30: Serudung Murut intelligibility of four Murutic reference tapes and one Tidong reference tape. (Scores are given as percentages. PSC relations are in parentheses.)

All of the other test results show clearly that Serudung Murut represents a separate language from Tagal, Baukan, Nabay and Tidong. The only cross-testing of these was at the Tidong village of Indarasabak TU. The 20% intelligibility score registered there confirms the lexical classification of these two languages as belonging to separate language families (→ Tidong).

4. PROPOSED DIALECT CLASSIFICATION

The SIL survey and dialect intelligibility testing have cleared up many questions as to grouping the Murutic people by language. However, other factors were brought to bear on the classification conclusions presented in this section. Previous classifications made by other researchers are seriously considered. Also much information was gathered from the test subjects and village headmen as testing was carried out, and this has been incorporated here.

Figure 31 identifies and displays the general location of the 21 dialects discussed in this section. Eight dialects, including Gana, Beaufort Murut, Timugon, Kolod, Selungai Murut, Sembakung Murut, Kalabakan Murut and Serudung Murut, are sufficiently different from one another and the other dialects to represent separate languages. Apin-Apin Kuijau, Dusun-Murut, Ambual and Nabay are grouped together as four dialects of the single Keningau Murut language. Baukan proper and Tenggara are two dialects of the Baukan language and Paluan proper and Pandewan are two dialects of the Paluan language. Pensiangan Murut, Salalir, Rundum Murut, Tomani Sumambu and Maligan Tagal are grouped together to form the single language Sumambu/Tagal. Figure 32 presents the classification into 12 languages of the 21 dialects distinguished in this section.

The classification of the Murutic dialects as grouped into 16 Northern Murut dialects¹⁷ and five eastern Murutic dialects has not been emphasised here since the latter is a term of convenience and it is not known whether any real linguistic features are present that would distinguish them as a whole from the Northern Murut languages, or whether they have any closer linguistic relationship to each other than to the Northern Murut dialects. The results of

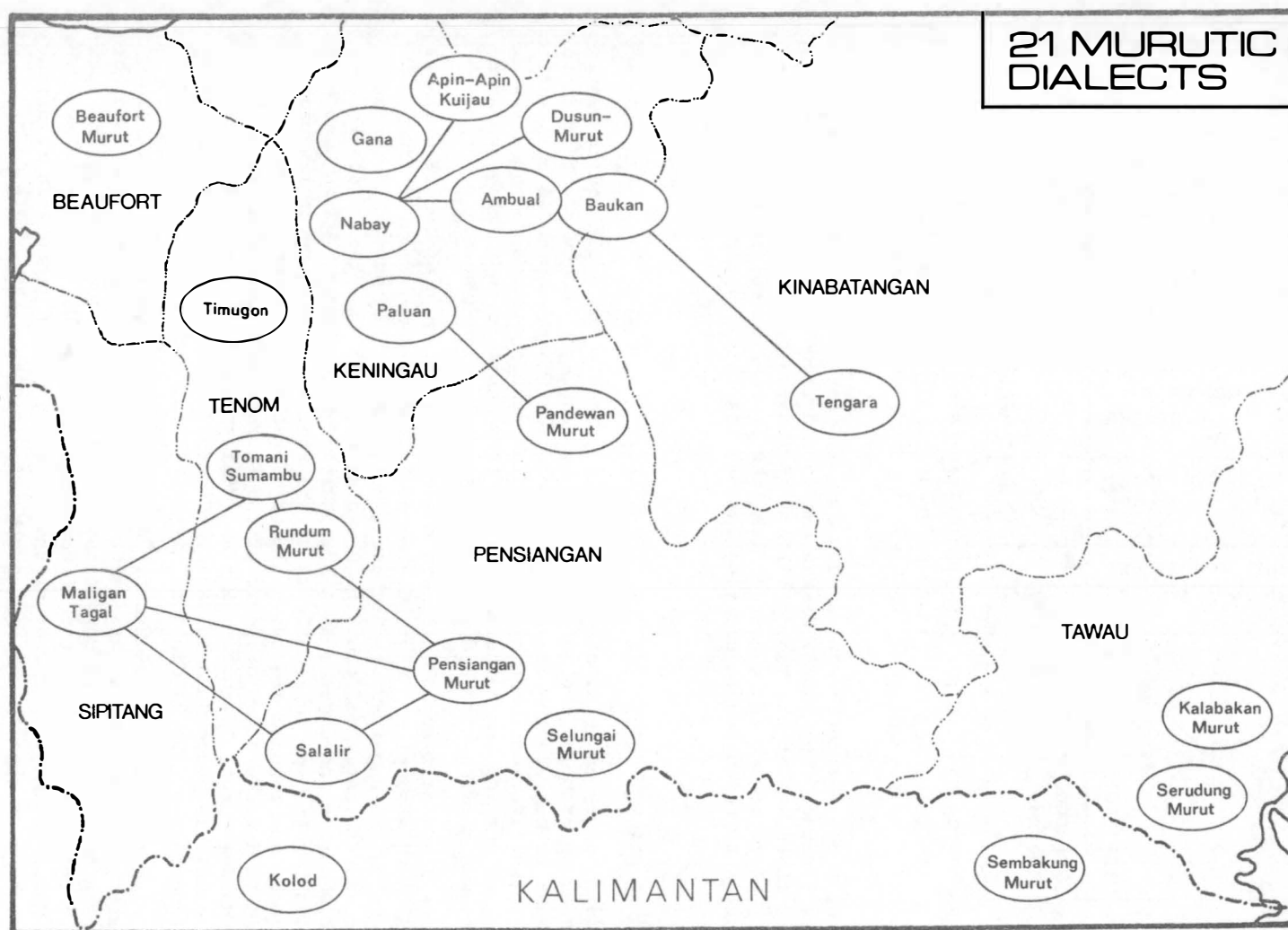


Figure 31: Identification and approximate location of 21 Murutic dialects of Sabah and Kalimantan.
(Dialects belonging to a single language are connected with a solid line.)

intelligibility testing between Northern Murut and eastern Murutic languages are found in Sections 10-12.

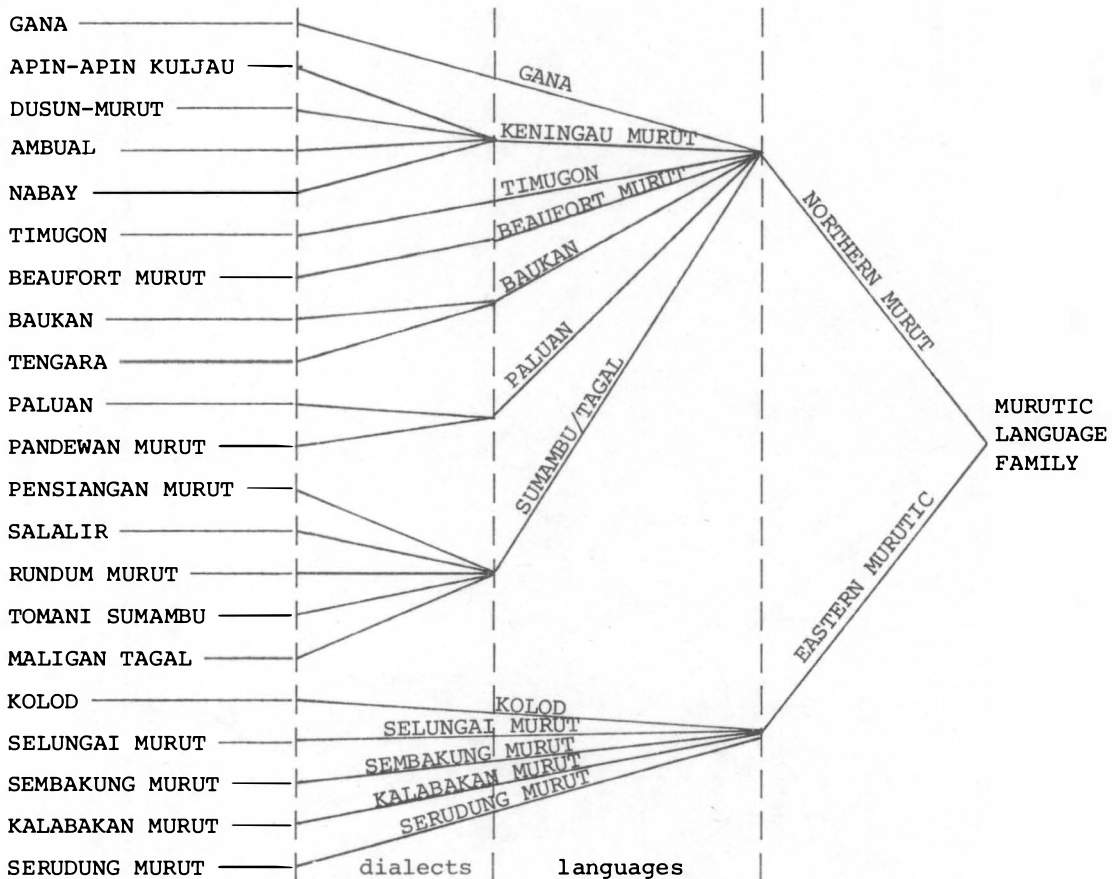


Figure 32: Classification of dialects and languages in the Murutic language family based on the combined factors of lexicostatistical analysis, intelligibility testing, previous research and information from villages tested

4.1 Gana language

This language group traditionally inhabits the valley west of the Pegalan River to the north of Keningau township. Much of the same area is now shared with Kuijau speakers (Figure 33; → Kuijau). A conservative estimate places the present population between 1000 and 1200 first-language speakers,¹⁸ although this is difficult to assess in light of the linguistic complexities around Keningau.

The ethnonym 'Gana' (also Ganaq) means *lowland* or *plain* and refers to plains-dwellers. It has been applied by Gana speakers themselves as well as by surrounding groups. Another ethnonym used by Prentice (1970) is 'Minansut' (→ Kuijau).¹⁹ Generally, it is supposed that others have simply considered them

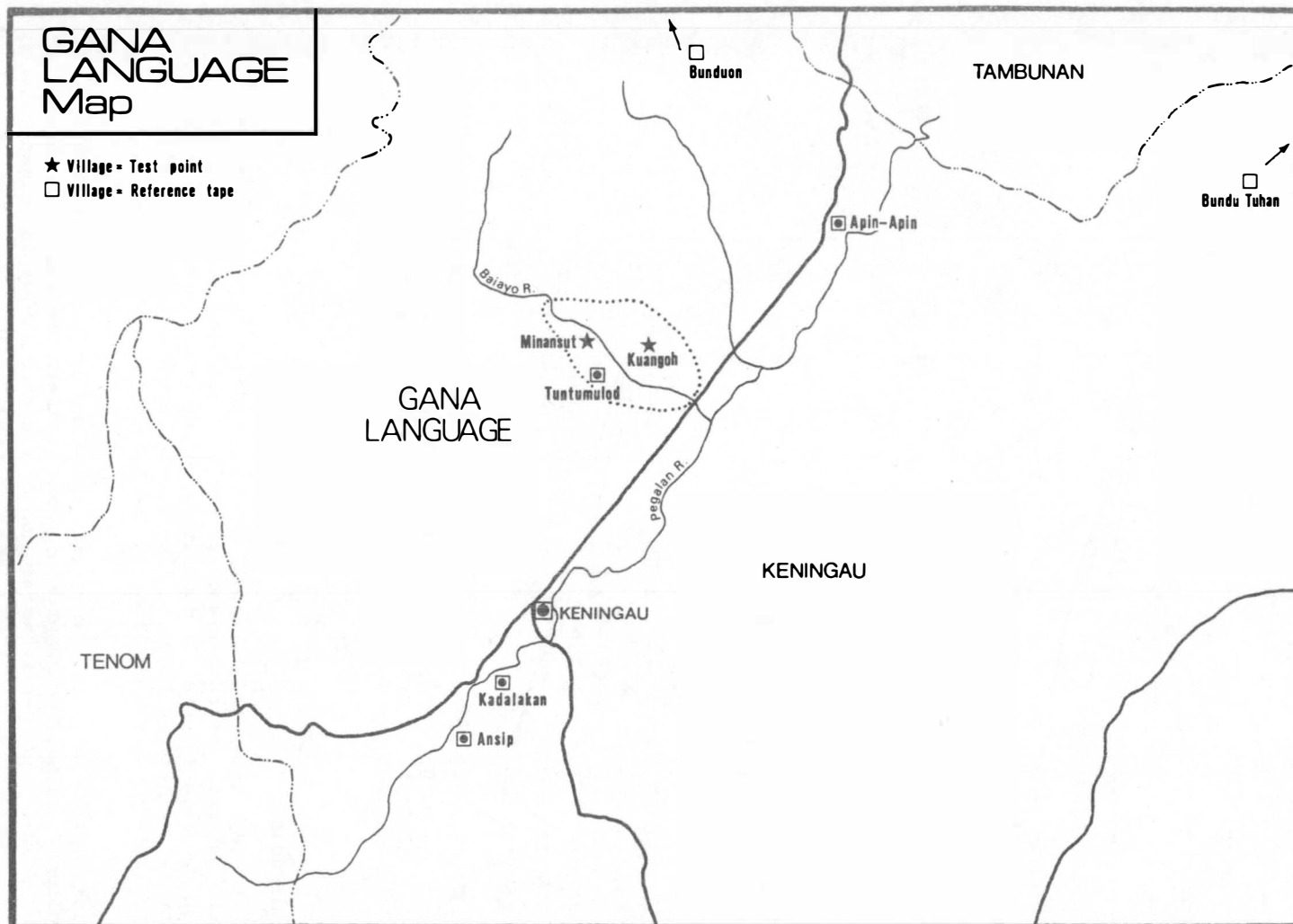


Figure 33: Gana language map showing the two test points, the sources of the six reference tapes and the general location of Gana speakers

part of the group labelled 'Keningau Murut', 'Nabay' and 'Kuijau' (and according to Prentice, 'Kuliow', 'Kuriow' and 'Keningau Dusun').

That Prentice classified 'Minansut' as a Dusun language while Smith classifies 'Gana' as "borderline" Murut emphasises the uncertainty of this classification. The practical mutual intelligibility between Gana and Kuijau (a Dusun language) speakers would argue for classifying them as one group. The distinctiveness of these two is established on a lexicostatistical basis, but widespread bilingualism is evident. The confusing mixture of languages in a limited area of Keningau District and numerous examples of bilingualism disqualify intelligibility testing as a means of clarifying classification distinctions in this case. In-depth fieldwork involving comparison of the grammatical systems of these languages as well as a thorough sociolinguistic survey of this area will be necessary to arrive at definite classification conclusions.

4.2 Keningau Murut language

This designation is used by Baboneau (1922) and Rutter (1929) and is adopted here to collectively describe several dialects in the Keningau township area, including Apin-Apin Kuijau, Dusun-Murut, Ambual and Nabay. Because of the consistent differences in test scores and the use of different ethnonyms, Ambual spoken in Ambual KU is considered here as a separate dialect from Dusun-Murut as spoken in Liau Laut KU (Section 3.3). A high degree of intelligibility of the Nabay dialect was found in the case of two dialects but the Dusun-Murut dialect did not demonstrate this. It is grouped with Keningau Murut, though, because of its close lexical relations (84-85 PSC). There is no evidence for distinguishing any of these four groups above the dialect level.

The whole language group probably numbers around 4000-5000 speakers. Figure 34 shows the general location of the four dialects of the Keningau Murut language.

Baboneau (1922) has an extensive vocabulary gleaned from "various Keningau Muruts" listing nearly 1800 items, sometimes from several dialects.²⁰ Rutter (1929) has a list of 185 items from "Keningau Murut".²¹ Prentice included a phonological and grammatical outline of Nabay with the Timugon grammar (1971: 279-282). Besides these and Cohen (1981) no other linguistic data was noted.

4.2.1 Apin-Apin Kuijau dialect

In light of the revised PSC relations which show Apin-Apin KU lexically more closely related to Dusun-Murut and Nabay samples (Figure 6), this language group is included in the Keningau Murut language. Apin-Apin KU is situated on the river known by the same name, a western tributary of the Pegalan River. It lies at the foothills of the Crocker Range to the west and the Wittti Range further east, about 12 miles north-east of Keningau KU (Figure 34). About one-fourth of the population speaks this dialect, the rest are mostly Dusun or Kuijau speakers. Their present population is estimated to be about 200 and may be restricted to a clan unit.²²

Very little is known about the actual origin of this language group. Although Smith classified it as Murutic lexicostatistically, he hypothesised that its historical roots are Dusunic. If so, prolonged contact of Tambunan Dusuns with Keningau Muruts may have introduced language change or assimilation

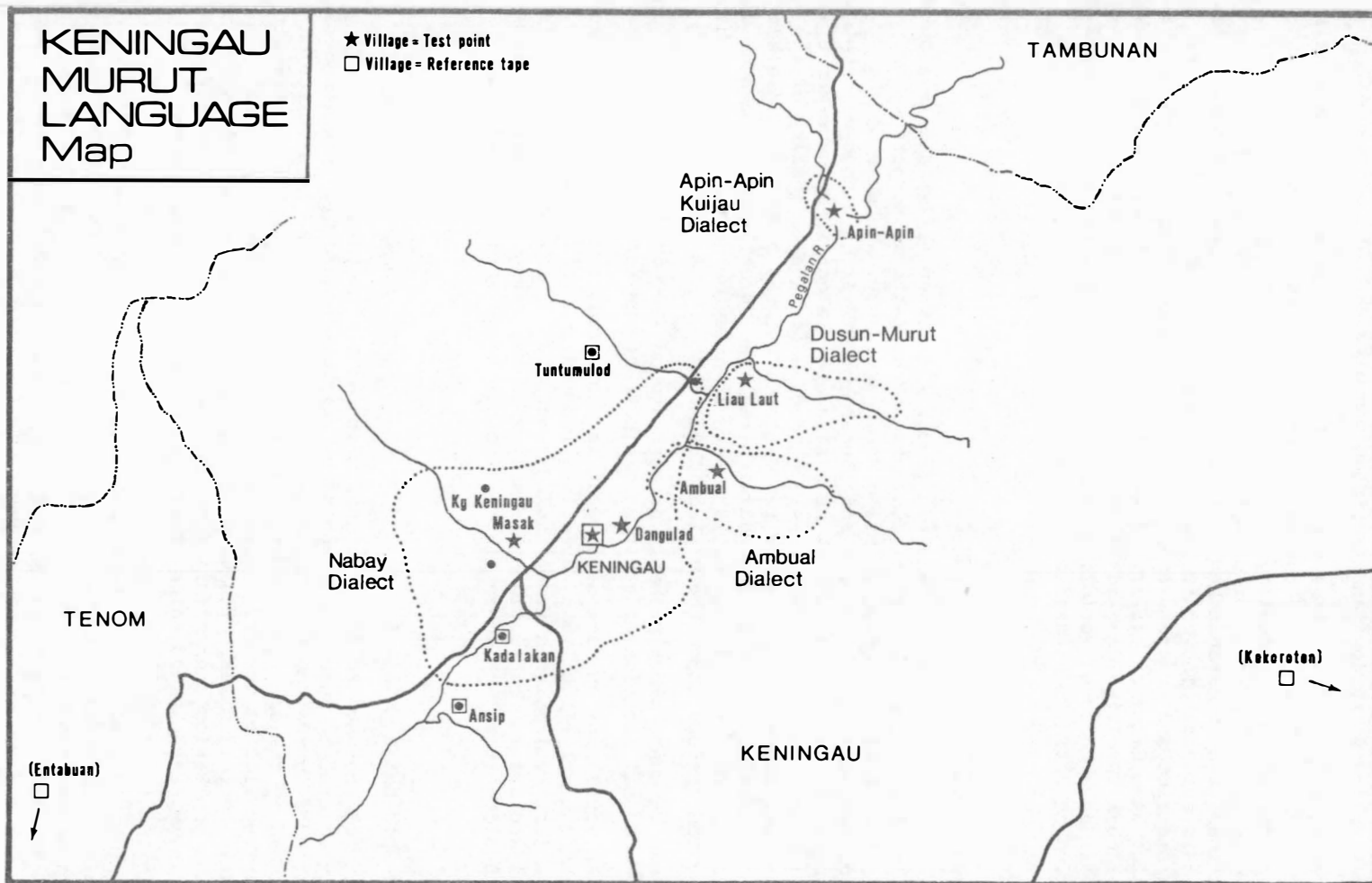


Figure 34: Keningau Murut language map showing test points, the sources of reference tapes used in testing and the general locations of the four Keningau Murut dialects

sufficient that they are now considered more Murutic than Dusunic (Smith, in this volume). On the other hand, its roots may be Murutic with the influence for change coming from the Dusunic groups.²³

In the data collected for this survey the subjects themselves variously called their language 'Dusun', 'Murut', 'Sinimbitan', 'Koroyau', 'Kadazan', 'Kuijau' and 'Kadazan-Murut'. Obviously the speakers perceive their language as an aberrant form of either Dusun to the north or Murut to the south. The term 'Kuijau' is probably used to refer to any 'mixed' or less well defined language such as Gana, Kuijau, Dusun-Murut and the Apin-Apin Kuijau dialect. There does not seem to be a common autonym in use.

Apin-Apin Kuijau speakers' breadth of bilingualism is evidenced in the significant intelligibility results shown not only of Nabay but of Tagal and Kuijau as well.

The relationship of Apin-Apin Kuijau to Dusun on one hand and Murut on the other, together with Gana, Kuijau and Dusun-Murut, form the crux of one of Cohen's (1981) major theses. Other than that there is no clear reference to this language group in the literature already published, except possibly in Ride's (1934) chart as adapted in Appell (1968:269).

4.2.2 Dusun-Murut dialect

The Dusun-Murut dialect group of Keningau Murut inhabits the Pegalan Valley east of the river and north-east of Keningau township between the Mailo and Ambual tributaries. It is closely related to the Ambual dialect spoken just south along the Pegalan River. Liau Laut KU and Sodomon KU represent this group.²⁴

Dusun-Murut probably has its roots in a common language with Nabay and Ambual, but its origin is uncertain. Their extended contact with Tambunan Dusuns who have penetrated the area for the past 50 years may have inhibited their intelligibility of Nabay and may be influencing them away from the main Keningau Murut speech forms.

The population of this language group is estimated to be less than 1000 people. Admittedly, this group is not easily delimited linguistically due to the gradual merging with Ambual and Nabay villages, and the actual population may be considerably less and very local.

Other than Cohen (1981) no other published materials have been found which specifically isolate this group from the other Keningau Murut groups. They themselves used terms such as 'Dusun', 'Dusun-Murut', 'Kuijau' and 'Murut', none serving as a real autonym.

It has been suggested (Section 3.3) that Dusun-Murut could be a dialect of a separate 'Nabay' language, but until further testing is conducted to determine whether this group shows intelligibility of Ambual and Apin-Apin Kuijau or whether those dialects and Nabay can understand Dusun-Murut it must be considered a closely related dialect with the other Keningau Murut dialects.

4.2.3 Ambual dialect

The Ambual dialect of Keningau Murut is spoken along the Ambual tributary of the Pegalan River and south possibly as far as Tuarid Taud KU. It shares a very close relation with Dusun-Murut to the north and Nabay to the south and west of the Pegalan. Nothing definite is known of the origins of the Ambual community but it is probable that they share historical roots with the Nabay and Dusun-Murut community.

Their present population is estimated to be between 1000 and 1200 if Tuarid Taud KU is considered part of this group. But it is possible that that village belongs better with Nabay villages as Smith so classified it and thus the Ambual speakers would number less than 500.

The ethnonym 'Tambual' was applied by Dusun-Murut speakers. Residents of Ambual KU simply called themselves and their language 'Murut' or 'Murut Ambual'.

Lexically this dialect is most closely related to Dusun-Murut, but it is also closely related to Nabay and has demonstrated high intelligibility of that dialect.

As far as is known, this group has never been isolated in the literature except by Cohen (1981). He referred to the language as Ambual and grouped it with Nabay and Baukan villages as 'Northern Murut' in his very local study of Keningau District.

4.2.4 Nabay dialect

The Nabay dialect of Keningau Murut is spoken along the Pegalan River valley and its west bank tributaries just south-west of the Ambual dialect area from Tuarid Taud KU to Banjar KU. Interspersed among the traditional Nabay areas are Kuijau, Tambunan Dusun and, it is said, Bajau villages.

The present population is estimated to be between 2000 and 3000, depending mostly on how the villages are actually divided between Dusun-Murut, Ambual and Nabay.²⁵

The ethnonyms recorded are 'Nabay' (also Nabai and Nebee), 'Dabay' (also Dabai and Dabi) and 'Rabay' (also Rabai). These are also used locally as auto-nyms. Prentice (1970; 1972) also suggests that some are referring to this group when they speak of Gana.

Lexically Nabay is the most central of all the Central Murut dialects which includes the dialects of Keningau Murut as well as Timugon and Baukan and shares a high percentage of lexical items with each. But Nabay speakers' average percentage of intelligibility of the stories from the latter two languages was actually lower than the PSC relations between them. Nabay was found to be a widely understood dialect throughout the Murut-speaking areas, though it does not serve as a trade language. It groups together with Apin-Apin Kuijau, Dusun-Murut and Ambual more on the basis of lexical relations and the statements by the speakers themselves than because of intelligibility testing, since only limited testing was actually conducted.

4.3 Beaufort Murut language

The Beaufort Murut language is spoken in several settlements found in Beaufort District along the Padas River from Saliwangan BT south-east of Beaufort township to Malalugus BT just west of Beaufort BT. Another group of villages is located about ten miles south-west of Beaufort along the Bukau River²⁶ (Figure 35).

Informants could not explain their historical origin but they do recognise a closer linguistic affinity with the Timugon Murut community than with other Murutic peoples. It is possible they moved down the Padas River from Tenom District, but not within the recent past. The areas they now inhabit are surrounded by Bisaya and Kedayan speakers.

The Beaufort Murut language is spoken by an estimated 1,200 to 1,700 people. Those at Bukau BT referred to themselves as 'Sandiwar' while in Malalugus BT only the term 'Murut' was recorded. The name 'Sandiwar' (also Sandewar), means *outside* and was applied by the Murut around Jimpanga BT and Beaufort BT to those who lived outside, or above the Penotal Gorge on the Padas River, a description which better fits the Tenom Timugon. The Timugon in turn called the Jimpanga BT Muruts *ulun dabugus* or 'Dabugus', which means *people toward the sea*. This was also applied to the Bisaya and Brunei in the same area (Woolley 1962:2). Crespigny (1872) referred to one or both of the Beaufort Murut groups as 'Lower Murut' or 'Murut Padass'. They are often incorrectly referred to as 'Bisaya' by the local people, probably due to their proximity to Bisaya villages.

The Sandiwar subgroup of Beaufort Murut found along the Bukau River has been classified by Prentice (1970) as a subdialect of Timugon called 'Bukou'. The Dabugus settlements further north were labelled by Prentice as speaking the subdialect 'Bintaq'. These he grouped together with two subdialects of Tenom Timugon. Prentice further suggested that Crespigny's (1872) 'Murut Padass' wordlist was collected in the Bintaq subdialect of Beaufort Murut.²⁷ He included a brief comparison and outline of the phonology and morphology of the Beaufort Murut groups in his Timugon grammar (1971:276-279).

According to the lexical percentages both the Sandiwar and Dabugus wordlists relate closely enough to one another to be considered members of the same dialect, but may indeed represent subdialects. Mutual intelligibility between them was not checked. Both samples are slightly more remote in relation to the Timugon wordlists. This was verified by intelligibility testing with Sandiwar.

Although both Beaufort Murut groups are lexically close to Tenom Timugon (*81 PSC), they are quite far removed from other Murutic groups (*55-74 PSC), due somewhat to borrowing from the Bisaya around them. Whereas historically Beaufort Murut may have roots in the Timugon language, it is now sufficiently different lexically to be considered a distinct language.

4.4 Timugon language

The Timugon community is concentrated in their traditional area along the Padas River from Melalap TM as far south as Batu TM (Figure 36).²⁸ The most anyone could relate about their origins was that the Timugon have always inhabited this region of the Padas River valley (cf. Woolley 1962:1). The area is traditionally shared with Paluan speakers and now with migrant Dusun groups from Tambunan, Salalir groups from Pensiangan, Kolod and Lumbis people from Kalimantan, and Javanese and Banjar immigrants from other Indonesian islands.

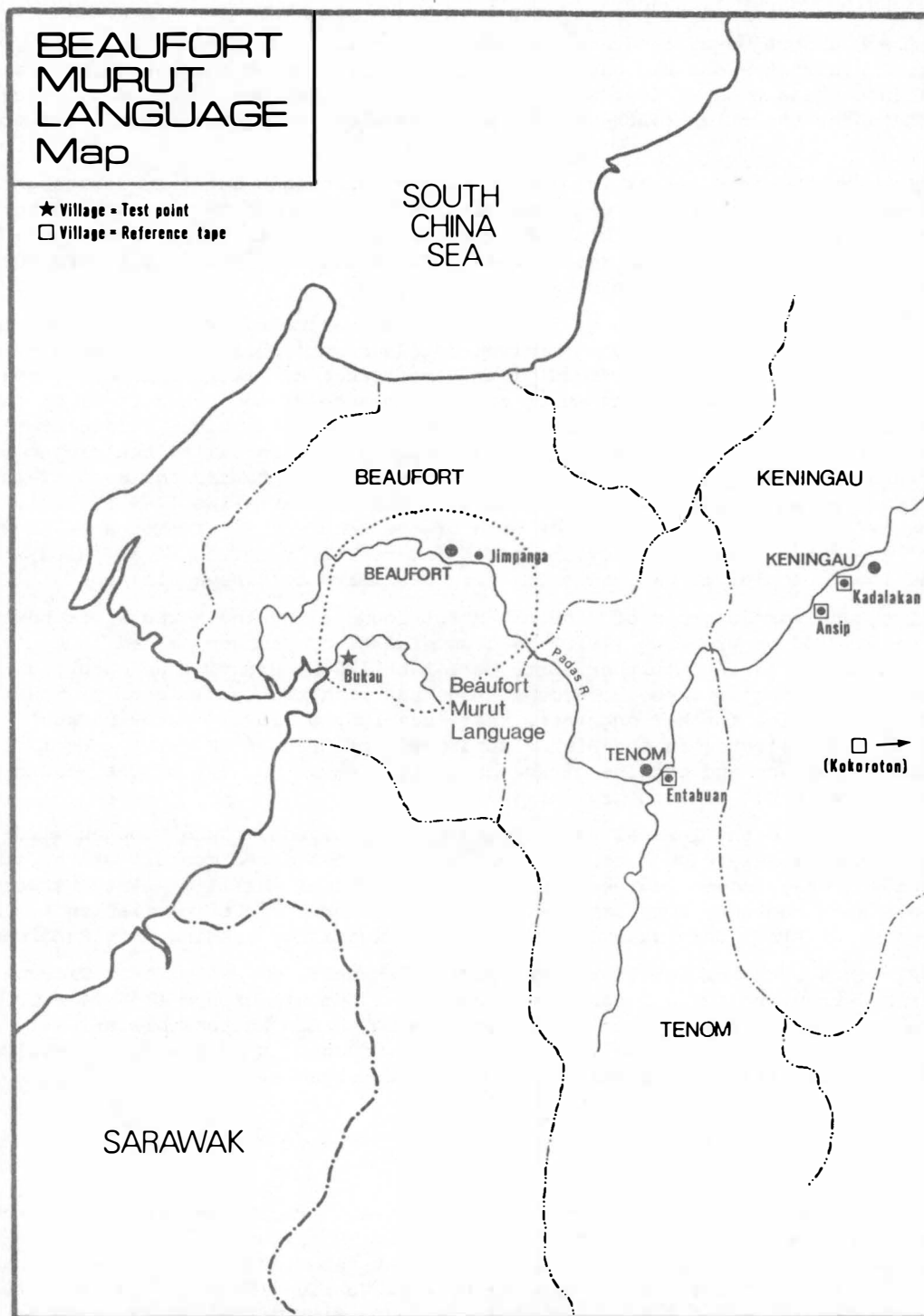


Figure 35: Beaufort Murut language map showing the test point, the source of the reference tapes used for testing and the general area where the language is spoken

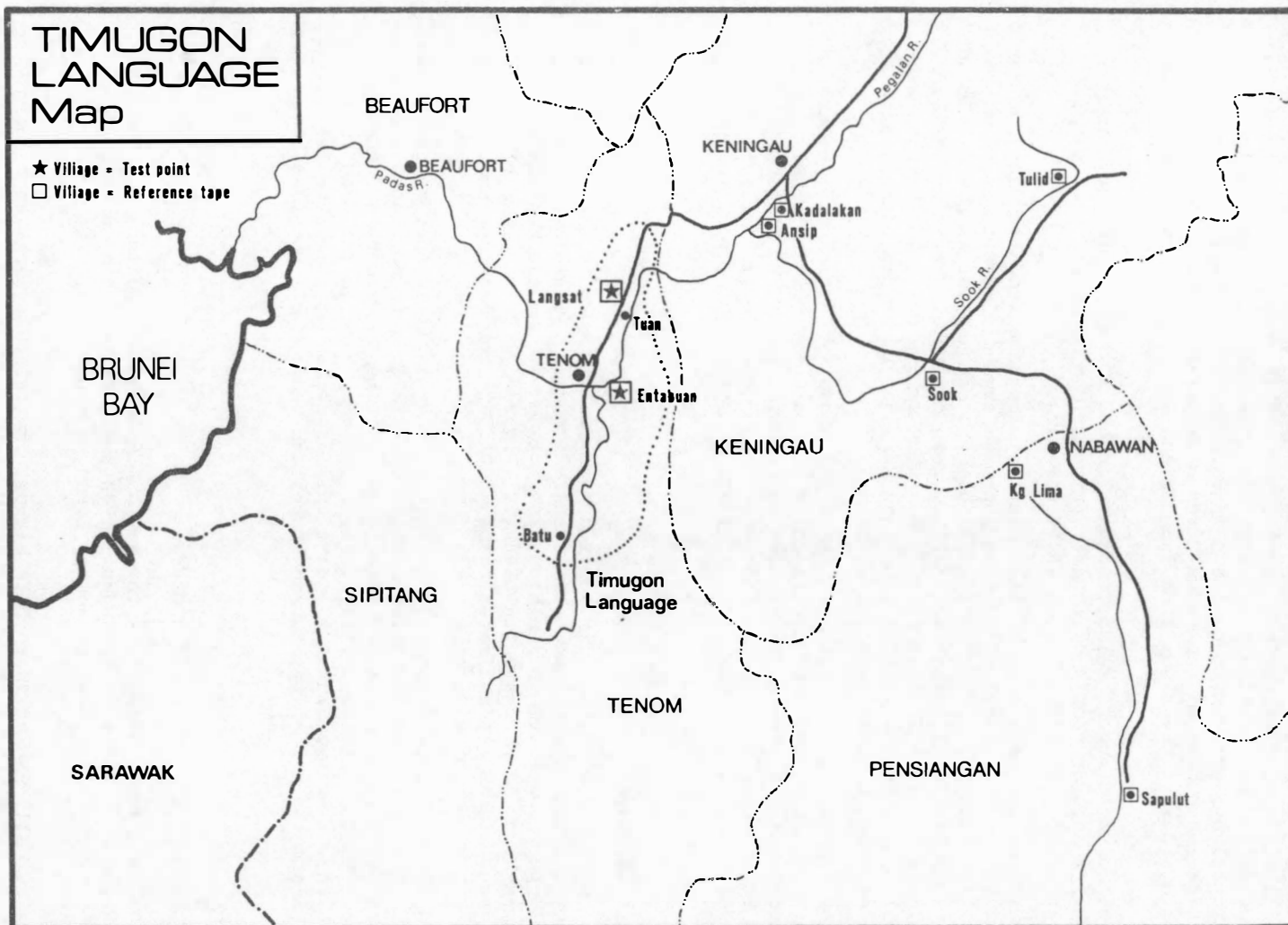


Figure 36: Timugon language map showing test points, sources of the reference tapes used in testing and the general location of the Timugon language area

The present population of this language community is estimated to be 6000 to 7000.

No dialect-level distinctions were noted in this research, but the two sub-dialects delimited by Prentice (1970:380) and also recognised by the local Timugon themselves, are very probable. This includes the 'Kapagalan' group located on the west bank of the Padas River near Melalap TM and the 'Poros' who inhabit both sides of the river from Tuan TM and Makaniton TM as far south as Batu TM.

The ethnonym 'Timugon' (also Temogun, Timogon, Timugan, Timogun, Timigun and Timugun) means *the people of the river* from *timug river* and *ulun people*. To many, this group is simply known as 'Tenom Murut'.

The Timugon language represents a distinct language from the Beaufort Murut language. Though individuals from both language communities show marginal to high intelligibility of the other's language, the communities as a whole do not demonstrate mutual intelligibility. Still, Beaufort Murut shares by far more lexical similarity with Timugon than with any other dialect. The Timugon language is also considered distinct from both Keningau Murut and Baukan, as was demonstrated through intelligibility testing.

All of Prentice's Murut studies mentioned above (→ Section 0.5) centre around this language, and more specifically around the Poros subtribe. Rutter (1929) incorporates a 'Tenom Murut' wordlist consisting of 185 items which according to Prentice is probably from the Kapagalan dialect.²⁹ Clayre (1970) uses some of Prentice's (1965) data to highlight her discussion of focus.

4.5 Baukan language

The generic term 'Baukan' is applied specifically to only one of the two dialects of the language. The other is called Tenggara. These groups are found along the upper Sook River in Keningau District and around the headwaters of the Kinabatangan River in Kinabatangan District (Figure 37). The present population of this language group is estimated to be 1500-1800, the greatest number of whom speak the Baukan proper dialect.

The dialect division must be seen as tentative until further research is conducted. The true relation between these two dialects is not known. Obvious differences exist, however, as is seen in the different lexicons and the speech intonation. Still, intelligibility varied from very high in Tulid KU to minimal and low in the other two Baukan language villages. Due to the acknowledged differences and the ambiguity, Baukan proper and Tenggara are considered separate dialects.

4.5.1 Baukan dialect

The Baukan proper dialect of the Baukan language is spoken along the upper reaches of the Sook River in Keningau.³⁰ Speakers are known by others as well as themselves as 'Baukan' (also Baokan, Bokan, Boken, Bokun, Bukun, Bokon, Ulun-no-Bokon and Ulun-no-Bokan).

Linguistically Baukan shares its highest lexical relation with Tenggara, the other dialect of the Baukan language, and Nabay. However, mutual intelligibility was slightly higher between Baukan and Tenggara than between Baukan and Nabay.

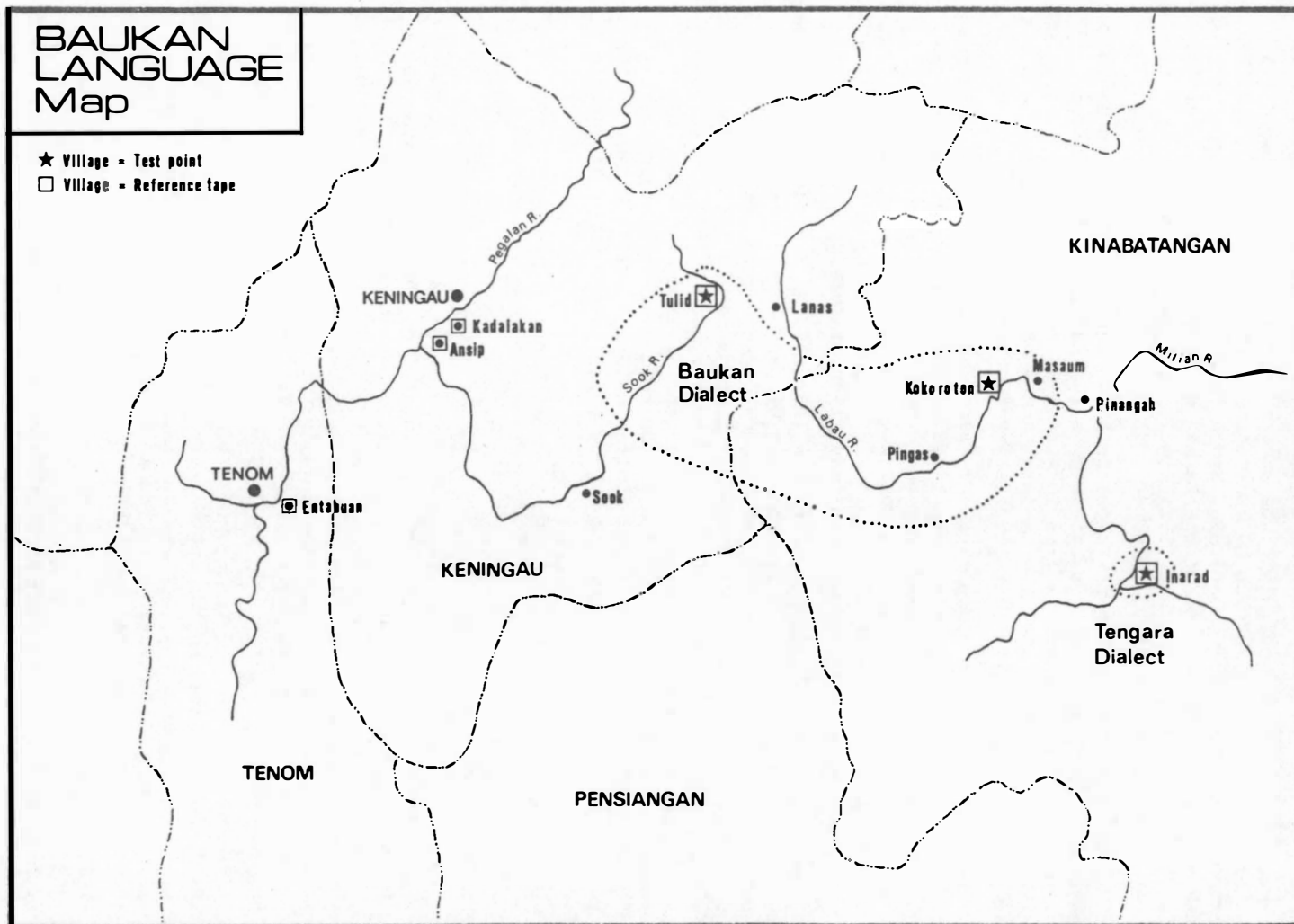


Figure 37: Baukan language map showing test points, sources of reference tapes used in testing and the general area where this language is spoken

Baukan subjects also showed high intelligibility of Tagal. But a relatively low score was recorded on the Timugon story.

Some research has been done among the Baukan groups including a wordlist of natural history collected in Pauh KU and Tiong KU (Keith 1936), along with some anthropological notes on taboo (1936) and a folktale (1936). Prentice (1971:282-285) included an outline of the Baukan dialect phonology and morphology in his Timugon grammar and Cohen (1981) included a single Baukan wordlist in his study of Keningau District dialects.

The Baukan dialect as spoken in Kokoroton KN on the Labau River is no longer found in a concentrated area, but is spoken by a few families in the upper Kinabatangan area, some of whom now live in Masaum KN.³¹ This area was included by Keith (1936) as part of the Baukan language and is included here on the basis of lexical relations and limited intelligibility testing.

The exact origin of this group is not known. However, in Kokoroton KN the SIL researchers were told they originally lived further upriver between Pingas KN and Lanas KU but were asked to move by the government 150(?) years ago. According to one tradition this group itself used to make up a very large population but most died off because of an epidemic a long time ago. It is estimated that now less than 50 speakers of this particular subdialect are left.

This community used no autonym other than 'Murut', nor did the other groups in the area know what to call them if not *orang pingas people from Pingas*. For convenience they have been referred to as 'Kinabatangan Murut', together with the Tenggara Muruts. As noted above, Keith referred to Muruts in this area as 'Baukan' while others have considered them 'Tenggara', though they themselves do not use this as an ethnonym. The local people generally think of the Muruts from Kokoroton KN as more closely related to the Keningau Muruts than to the Tenggara of Inarad KN. Based on this and limited data the Tulid KU test point has been tentatively classified as belonging to this Baukan proper dialect (→ Section 3.5).

4.5.2 Tenggara dialect

The Tenggara dialect is found along the Inarad River, a tributary of the Kinabatangan/Milian River. One other village located on the Kuamut River was mentioned by several people, but it is no longer in existence (Figure 37).³² Others have called the Murut people at Kalabakan TU 'Tenggara'. While several Inarad KN Murut seemed to think their language was very similar to the Tawau District Murut, SIL research showed that this is not a close relationship, nor did the Kalabakan Murut mention any connection. The two groups certainly cannot now be considered members of the same language, let alone the same dialect.

Population figures are not known, but SIL researchers were told that about 150 families lived in Inarad KN and another two or three households from Lenod KN along the Kuamut River now live in Kuamut KN. A recent estimate by the Native Chief was about 350 residents in Inarad KN and about three families in Pinangah KN besides those around Kuamut KN.

Their origins are unknown. Earlier references to this group are sketchy, but Keith (1981) gives an interesting account of one incident in their history which dates back as far as 1890 and in which she connects Tenggara with the Kuamut River and Kalabakan TU Murut. The ethnonym 'Tenggara' (also Tenggara, Tingara, Tenggaraq, Tanggaraq, Tenggara and Tenggara'), is applied to them by the Upper

Kinabatangan groups such as the Kolobuan and the Makiang. They usually refer to themselves as simply 'Murut'.

Aside from the other Baukan dialect, Tenggara shares the greatest number of lexical items with the Paluan language and the Kalabakan Murut. The Tenggara test subjects showed the highest intelligibility of the Kokoroton KN story, while the limited intelligibility of the other Baukan story from Tulid KU as well as stories in Nabay, Timugon and Tagal was about the same. Neither Tidong Kalabakan Murut nor Serudung Murut speakers demonstrated any significant intelligibility of Tenggara.

Though the Tenggara are frequently mentioned in the literature, the only linguistic data for this group is Rutter's (1929) 'Tenggara' wordlist³³ and Hurlbut and Pekkanen's (to appear) study of the Upper Kinabatangan River languages (→ Section 3.5, note 12).

4.6 Paluan language

The Paluan language is spoken from the area around Tenom township east over southern Keningau District and as far south as the Talankai River in Pensiangan District. There has been movement within these boundaries recently but the general area seems to be the Paluan community's traditional habitat (Figure 38).

Intelligibility test results show high mutual intelligibility among the few villages tested, except Pandewan PN, so it was decided to consider Pandewan a separate dialect and to group the four other villages under one Paluan proper dialect (→ Section 3.8). Speakers of the Paluan language recognised a different set of villages for speakers of the dialect used at Pandewan PN. Also the intelligibility test scores were consistently lower at Pandewan PN than the other four villages. Though only two dialect-level divisions are noted in this classification, Paluan proper and Pandewan Murut, differences are also noted lexically between Dalit Murut and Sook Murut and Prentice (1970) lists the Paluan communities just east of the Tenom plain around Saga TM as a separate subdialect. Lexical differences are also recorded between the Dalit KU and Keramatoi Laut KU lists suggesting differences along each river. Thus grouping the four villages under one Paluan proper dialect must be seen as tentative until further research is conducted.

There are varying degrees of mutual intelligibility between the different Paluan areas as well as with their Sumambu/Tagal and Keningau Murut neighbours, but the two dialects discussed here are a fair representation of the divisions on the dialect level of this language.

4.6.1 Paluan dialect

The Paluan proper dialect as isolated here is found along some tributaries of the Padas River around Tenom township east over southern Keningau District along the Dalit and Keramatoi Rivers as well as the lower Sook River valley and south around the headwaters of the Talankai River in Pensiangan District (Figure 38).³⁴ This dialect is estimated to consist of about 4500-5500 speakers at present.

Other ethnonyms and nomenclature applied to this group besides 'Paluan' (also Peluan) are 'Dalit Murut', 'Sook Murut' and 'Takapan'. Prentice (1972) also noted the terms 'Makaheeliga' and 'Makialiga'.

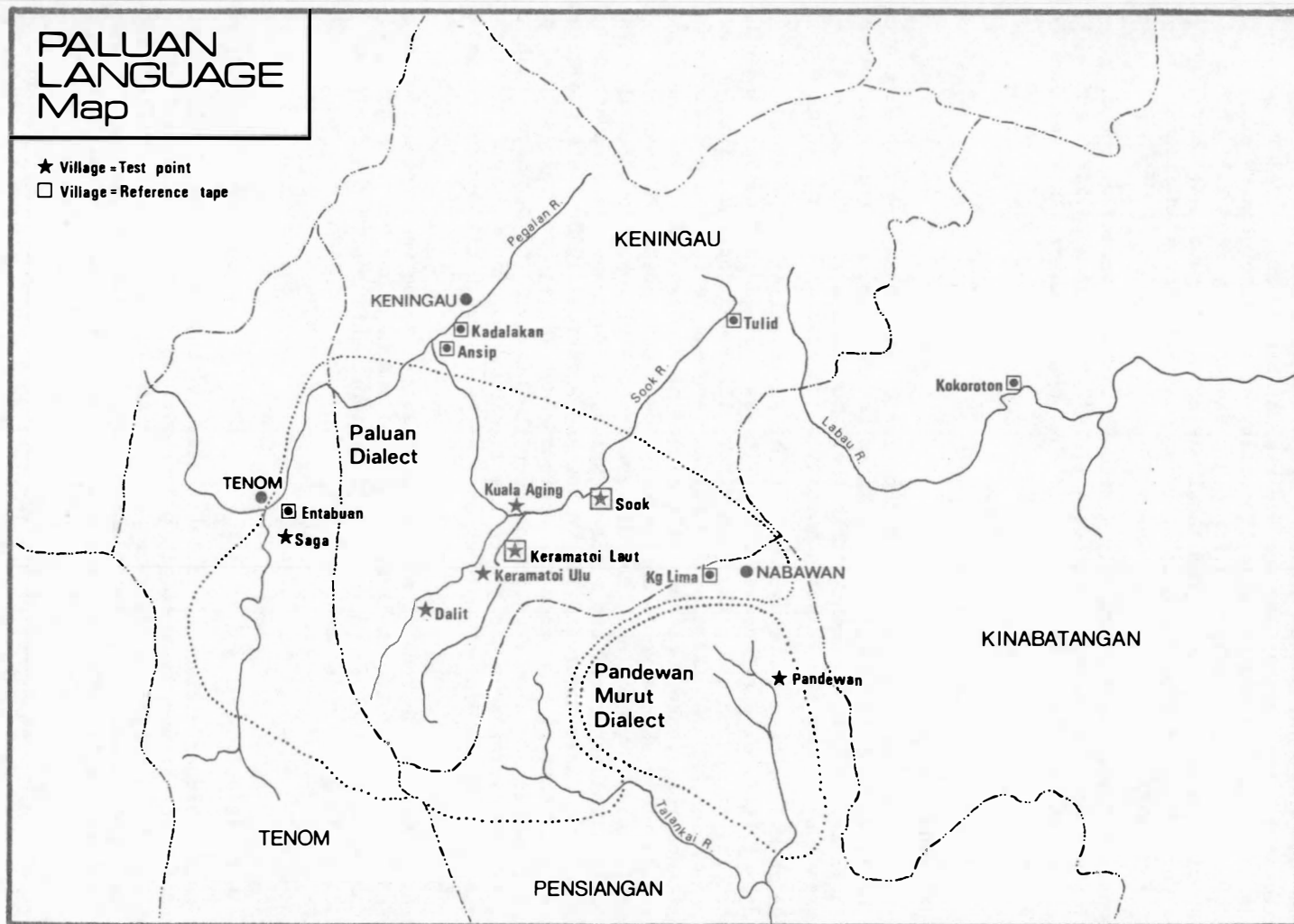


Figure 38: Paluan language map showing test points, sources of the reference tapes used in testing and the general area where this language is spoken

Lexically the Paluan dialect is midway between the dialects of the Sumambu/Tagal language to the south and those of 'Central Murut' to the north. Geographically they act as a buffer between these same two groups. Paluan subjects demonstrated high intelligibility of both Nabay and Pensiangan Murut stories. Their overall intelligibility of Timugon and Baukan stories was lower. Timugon subjects did not register appreciable intelligibility of this dialect.

Rutter (1929) included a 186-item wordlist of the 'Peluan Murut' dialect³⁵ and Woolley (1927; 1928) published three Paluan songs from Dalit KU and he also collected another which is produced in Rutter's (1929) volume. Prentice (1971) identifies the *English-Murut phrase book* (1964) as belonging to this dialect. Lees (1966) made comparison with Paluan and Tagal in her brief orthography study. Cohen (1981) used four wordlists from this dialect in his study which showed Keramatoi Laut KU and Dalit KU more closely related to one another than to either Sook KU or Nabawan KU. This follows their geographical position as well.

4.6.2 Pandewan Murut

The Pandewan Murut dialect is found strictly in Pensiangan District along the Nabawan, Pamentarian and Mesopo Rivers, as far downstream as Matiku PN, about seven miles north of Sapulut PN (Figure 38).³⁶ It is estimated that the Pandewan Murut dialect consists of 1200-1400 speakers. They claim to have always lived in their present location.

In Pandewan PN they referred to their language as Murut Asli *original or authentic Murut* or as 'Pandewan Murut', but never as Paluan or Tagal. This dialect may be partly included in the dialect labelled Sapulut (also Sapulot, Sepulot and Sipulote) by Prentice, Rutter and others.

Aside from Paluan, this dialect shares its closest lexical relations with Baukan, Nabay, Pensiangan Murut and Timugon, in that order. However, no appreciable intelligibility was shown of Baukan, Timugon or Salalir and only marginal intelligibility of Paluan.

4.7 Sumambu/Tagal language

The Sumambu/Tagal language consists of a group of interrelated dialects which vary sufficiently in some cases to inhibit mutual intelligibility between them. It is spoken over the whole south-western region of the Murut area of Sabah and down into Kalimantan (Figure 39). Speakers of this language have begun to migrate north into Timugon, Paluan and Nabay areas during the past 50 years. Other than that, the historical roots of the group as a whole have not been traced.

A fair estimate of the population of the Sumambu/Tagal groups as delimited here is 13,000-14,000, nearly twice the size of any other Murutic group.

The terms 'Tagal' and 'Sumambu' have generally been applied when referring to this group as a whole with only slight variation.³⁷ Rutter (1929:34) used Sumambu to refer to Rundum Murut and Sapulut Murut collectively. Other ethnonyms will be handled under the respective dialects below.

Five dialects are distinguished here, but in a language group such as this each river or vicinity varies somewhat from the next. These dialects include Pensiangan Murut, Salalir, Rundum Murut, Tomani Sumambu and Maligan Tagal.

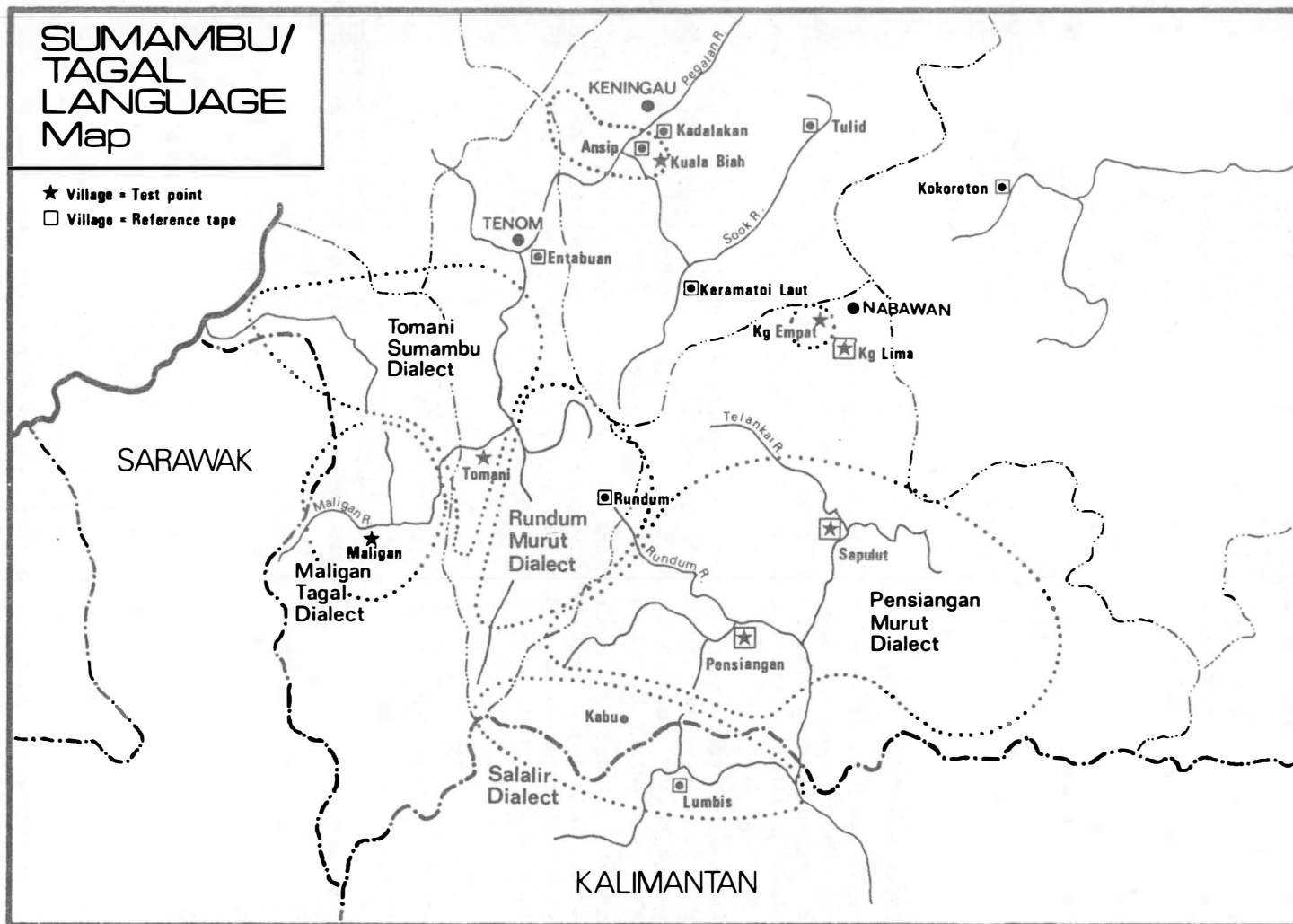


Figure 39: Sumambu/Tagal language map showing test points, source of reference tapes used in this testing and the general locations of the dialects of this language

4.7.1 Pensiangan Murut dialect

The Pensiangan Murut dialect of Sumambu/Tagal is found along the lower reaches of the Sapulut, Talankai and Rundum Rivers, and the Tagul and Sansian Rivers, all in Pensiangan District. There has been a great movement from this area to other areas, most recently to the Nabawan Scheme PN near the Keningau/Pensiangan border³⁸ and to a few villages in the central part of Keningau District.³⁹

This dialect represents the largest number of Sumambu/Tagal speakers. It is roughly estimated to consist of 4000-5000 speakers.

The 'Pensiangan Murut' dialect (also Pěntjángán and Penchangan) is partly referred to as 'Sapulut Murut' (also Sepulot, Sapulot and Sípulote).⁴⁰ It has been called 'Tagal' (also Tagul, Tagol and Taggal),⁴¹ and 'Lagunan Murut'.

Pensiangan Murut is the dialect most widely understood of the Sumambu/Tagal language group. Intelligibility testing in this dialect showed they did not understand Paluan and Baukan, nor Timugon and Nabay if checked in a non-immigrant village. Their intelligibility of the Salalir dialect varied from low to marginal. It was not checked with Tomani Sumambu or Maligan Tagal but showed marginal intelligibility of Rundum Murut. Paluan villages demonstrated intelligibility of Pensiangan Murut, as did some Keningau Murut and Baukan villages, probably due in part to language learning. Rutter (1929:34) felt Pensiangan Murut was closely related to Kalabakan and Serudung Murut, which is not the case.

Genderen Stort (1916) has a wordlist from this dialect⁴² and Woolley collected a Murut chant from the Tagul River area which was published in Rutter 1929. Woolley's (1936) article describes some of the technical hunting terms from this area. Prentice (1971) mentioned a *Murut Catechism* (1960) by the Roman Catholic Mission at Keningau. Clayre's (1970) discussion of focus in the 'Tolokoson dialect' may apply to Pensiangan Murut as well. Prentice's discussion and outline of the phonology and morphology of Highland Murut include this dialect as delimited here (1971:285-300). Cohen (1981) included lists from two migrant Pensiangan Murut villages in his lexical study.

4.7.2 Salalir dialect

The Salalir dialect is spoken along the river of the same name near the Sabah and Kalimantan border (Figure 39). Other than the fact that most Salalir Murut on the Sabah side of the border have migrated from there to other areas in Pensiangan, Tenom and Sipitang districts,⁴³ the origin and population of this group are unknown. Because of that extensive exodus from their homelands a more thorough survey of possible Salalir villages in the other areas is necessary before an accurate assessment can be made of the population.⁴⁴ The people of the Lumbis KAL area along the Lumbis River in Kalimantan are also considered part of this group.⁴⁵

The dialect has been called 'Salalir' (also Sadalir, Sědálir, Selalir and Saralir). One informant now living at Kg Lima PN used the generic term 'Tagal' but generally speakers simply called their language 'Murut' or 'Murut Kabu'. The 'Lumbis' (also Alumbis and Loembis) speakers are usually referred to specifically by that name.

Lexically Salalir shares its closest relation with Pensiangan Murut and Salalir speakers demonstrated relatively high understanding of that dialect as

well as some significant intelligibility of Rundum Murut. They seem to have no trouble understanding Lumbis⁴⁶ but they do not understand Timugon well. Other dialects had more difficulty understanding Salalir. Both Pensiangan Murut and Tomani Sumambu showed low to marginal intelligibility of a Salalir story from Kg Lima PN.

On the basis of PSC relations and Salalir intelligibility of Pensiangan Murut there is no reason to distinguish this dialect from Pensiangan Murut. However, since Pensiangan Murut intelligibility of Salalir was low to marginal and other researchers have recognised a language level distinction (Prentice 1971:302), a dialect division does not seem too extreme since it presupposes some mutual intelligibility.

4.7.3 Rundum Murut

The Rundum Murut dialect as delimited here extends along the Sungih and Kapulu Rivers to the junction with the Telekosang River and south-east along the Rundum River and tributaries. Along the Rundum River it merges with the Pensiangan Murut community (Figure 39). This is somewhat broader than Prentice's (1970) description of the Arundum subdialect but does not include nearly as broad an area as Rutter (1929) has described as Rundum Murut.

Their origin is unknown although one informant ventured the opinion that years ago (100?) they came up from Pensiangan area. There is nothing particularly telling about this view though it is linguistically possible.

This dialect is considered to be one of the smallest two dialects of Sumambu/Tagal but a rough estimate might claim as many as 1,000-1,300 members at present.⁴⁷

No other ethnonym besides 'Rundum Murut' (also Arundum) is known for this group though Clayre (1970) probably includes this when she describes 'Tolokoson'. Others have, of course, used the generic terms 'Murut' or 'Tagal'.

Pensiangan Murut and Salalir speakers showed marginal intelligibility of the Rundum dialect, which confirms the dialect distinction made on the basis of lexical relations. But Maligan Tagal speakers showed high intelligibility of Rundum Murut, though there was a relatively low lexical relation (*75 PSC). More information is needed before determining whether Maligan Tagal, Tomani Sumambu and Rundum Murut should, in fact, be grouped more closely than done here.

Rutter (1929) includes a 'Rundum Murut' wordlist but the geographic boundaries of the group so designated by him are much wider.⁴⁸ This list was used by Bolang and Harrisson (1949) to establish the distinctiveness of Sabah Murut from Sarawak Murut.

4.7.4 Tomani Sumambu dialect

The Tomani Sumambu dialect is spoken along the Padas River between Kemambong TM and Tomani TM including villages along the Tomani River and near the upper Telekosang River.⁴⁹ The sizeable 'Tagal' population shown along the Mengalong River valley from around Sindumin SG eastward may belong to this dialect, or to the Maligan dialect. This area includes at least ten villages and residents of others (Figure 39).

An estimate of the population of this dialect community is difficult since many Salalir Murut have immigrated to the Tenom area where Tomani Sumambu speakers live.⁵⁰

The ethnonym 'Sumambu' (also Sumambuq, Semambu and Semembu) may be the most common one applied to this group by other Murut people, but 'Tagal' is also used, however inappropriate. Prentice (1970) classified this dialect under the linguistic designation 'Tomaniq subdialect' of Sumambu.

Tomani Sumambu test subjects showed fairly high intelligibility of the Pensiangan Murut dialect spoken in Sapulut PN, but not Salalir. Nor did they show any appreciable intelligibility of any non-Sumambu/Tagal dialect tested.

4.7.5 Maligan Tagal Murut

The Maligan Tagal dialect is spoken in the hilly region along the Ibul, Logan and Maligan Rivers in Sipitang District (Figure 39).⁵¹ A rough estimate could assign as many as 1,000-1,200 members to this language community, but it cannot be said with any degree of certainty.

A Salalir informant suggested that this group originated in the Tagul River area, but Maligan informants could not elaborate on this other than to say some had arrived one generation ago. The main immigration has not taken place in recent history since Rutter (1922:66) already has isolated this group.

The 'Maligan Tagal' (also Mauligan) group is sometimes known as 'Tagal' or 'Bol Murut' (also Bole Murut).

Lexical relations with the other Sumambu/Tagal dialects show this dialect to be the most remote of the five. The intelligibility testing demonstrated that they understand both Rundum Murut and Pensiangan Murut well, but as expected, no significant intelligibility of Paluan, Nabay or Timugon dialects was shown.

It is possible that the 'Tagal' translation of the New Testament books and stories, *Kula-Kula' Buku Nu Rahu Nu Tuhan* (1968) and *Kuasa' Nu Tuhan; Tagal* (1975), were done in this dialect (cf. Prentice 1971:293). Other mimeographed materials were prepared using translation helpers from both Maligan SG and Masanoi TM. These include the New Testament books of Philippians and I and II Timothy (*Sambahian Ra Surat Ri Paulus*: 1977), Mark (*Aho Onsoi Sinuratan Ri Markus*: 1977), I John (*Surat Puu-Puun Intor Ri Yahya: Pasukuon Ra Ulun Nu Tuhan*: 1978) and Colossians and Philemon (*Sambahian Ra Surat Ri Paulus*: 1980). The biblical stories of creation (*Kasaukan*: 1977 and *Kawi Nu Atu-Atu Inuma Nu Allah*: 1979) and the Rich Fool (*Baal Nu Ulun Akaya' Am Ookot*: n.d.), as well as a Tagal catechism (*Paniluan Ra Languson Nu Buku Hitu*: 1979) and a prayer directory (*Ibiton Ra Sambayang*: 1979) have been produced. Prentice adds to this list *Buku Manani Rahu Tagal* (n.d.) and *Pangumangan Nu Tuhan Yesus* (?1961). He also suggests that the 'Tagal' wordlist in Ray (1913) and the items mentioned in Clarke (1951) belong to this dialect as well as Lees' (1966) article.

4.8 Kolod language

No intelligibility testing was conducted in the Kolod village of Baru Jumpa TM, but Kolod is clearly a distinct Murut language and is included here as such. The Kolod language group is traditionally located around the headwaters of the

Padas River (Rutter 1929) and in Indonesian Borneo. Prentice (1971) identifies the areas as the headwaters of the Tengoa and the region south of the Lumbis-speaking area. It is not known exactly how far this area extends. Some groups have migrated to Sabah around Tenom, and Rutter (1922) mentioned some in the Bol SG area (Figure 40).

Prentice (1971) estimates the population of the entire group to be 2000 speakers. At least one-fourth of these could be living in Sabah in villages and in various estates in Sipitang, Tenom and Tawau districts.

There is little historical information about this language community but it is certain that the Kolod language relates closer to Northern Murut and eastern Murutic groups than to the Lundayeh of Sarawak,⁵² though it is often assumed that the latter are more closely related to them (Rutter 1929, Voegelin and Voegelin 1965, Grimes 1978, Whelan 1970). Inter-marriage between members of the two communities is evident even in this small sample (→ Section 2, note 11). The ethnonym 'Kolod' (also Kolur, Kolour, Okolor, and Unkolor) is pronounced by Muruts in Tenom okolod.

Prentice (1971:300-303) has included a phonological and morphological outline for this language, but no other linguistic data is known for Kolod.

4.9 Selungai Murut language

Selungai Murut is the language spoken in at least six or seven villages along the Sapulut/Sembakung River between the confluence with the Pensiangan/Tagul River and the Indonesian border (Figure 40).⁵³

The people of Selungai PN trace their recent roots back to Bantul PN, about a mile from the Indonesian border, from which they moved during the Confrontation in 1963. They feel a closer affinity with the groups along the Sembakung River than with those in Sabah, and it is possible that they came up from Indonesia before settling in Bantul PN.

The longhouse at Selungai is considered the largest in Pensiangan District with 161 persons living there. A conservative estimate of the population of this language group in Sabah is 600-800. There may be a significant population of Selungai Murut in Indonesia.

This language is referred to by other Pensiangan Muruts as 'Indonesian Murut'. Other than that, no ethnonym was discovered, though 'Sembakung Murut' drew a favorable reaction from the Selungai Murut themselves.

This language is most closely related to Sembakung Murut, also found along the Sembakung River, but much further downstream. It shares a considerable number of lexical items with Salalir and Pensiangan Murut, but it is significantly more distantly related to other Northern Murut dialects.

4.10 Sembakung Murut

The Sembakung Murut language is believed to be spoken all along the Sembakung River and tributaries in Indonesian Borneo (Figure 40).⁵⁴ No population or historical notes can be offered. Nomenclature other than 'Sembakung Murut' (also Simbakong, Sembakoeng and Sembakong) which has been used to describe this group includes 'Tidong' (also Tidung and Tidoeng) and 'Tinggalan' (also Tinggalum and Tingalun).

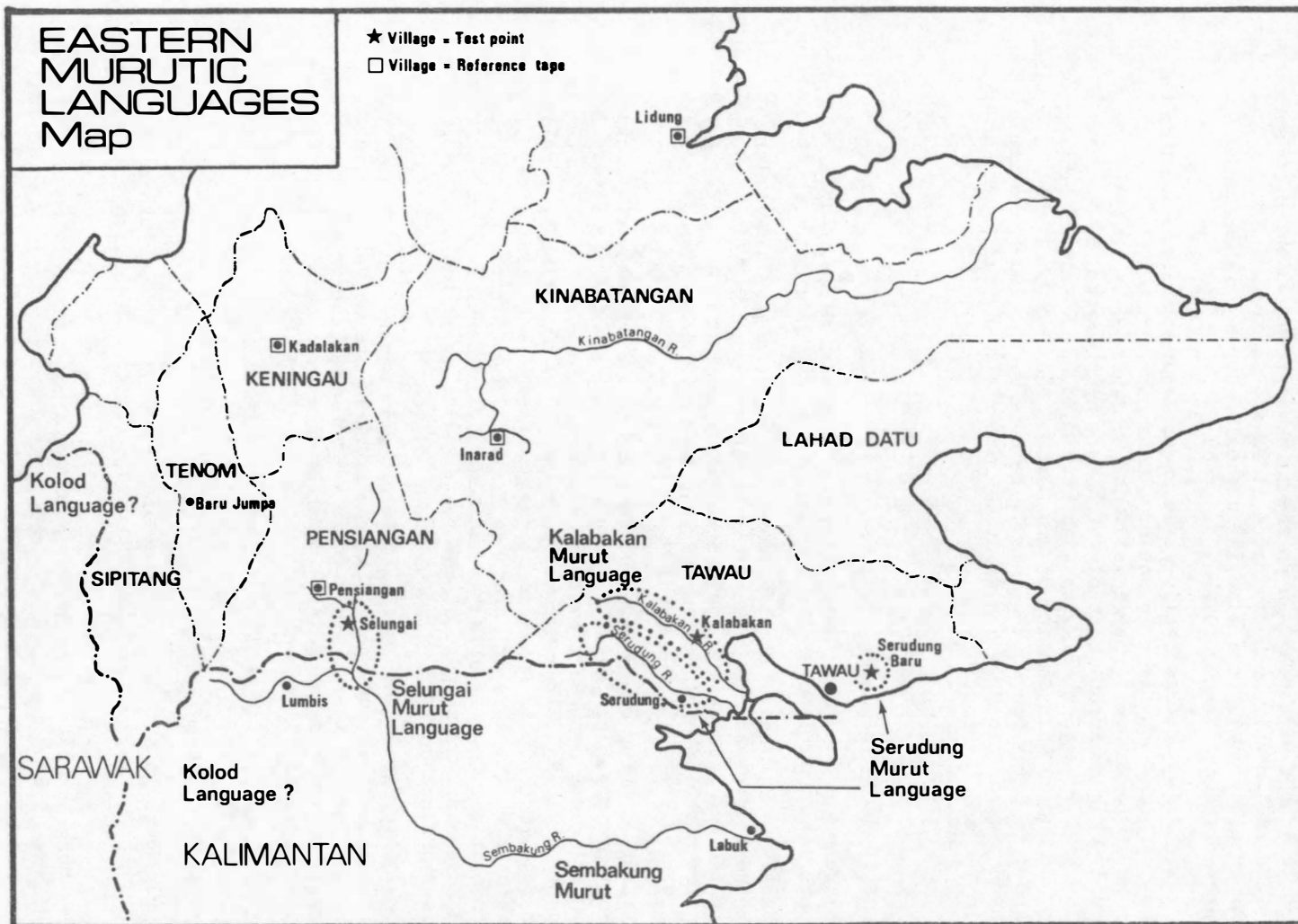


Figure 40: Eastern Murutic languages map showing test points, sources of the reference tapes used for testing and the general locations of the eastern Murutic languages

This dialect has long been considered a member of the Tidong language group, possibly due to religious affiliation, but in fact, it is lexically closer to the Murutic groups of Sabah than to the Tarakan Tidong which is spoken on Tarakan Island and in the Sabah districts of Tawau, Sandakan and Labuk-Sugut. The Bolangan dialect of Tidong, also in Indonesian Borneo, is lexically as distant as Tarakan Tidong (Sather 1972).

Its closest neighbour lexically is Selungai Murut, located much further upstream on the Sembakung River in Sabah. It is more remotely related to Pensiangan Murut (which was grouped as Tidong by Genderen Stort (1916)) and to Kalabakan and Serudung Murut (both of which were hypothesised by Prentice (1970) as belonging to Tidong together with Sembakung Murut). No intelligibility testing was conducted with Sembakung Murut data. The only published linguistic material known to exist concerning this dialect is Beech (1908) who included a few 'Simbakong' words along with his Tarakan material, and the 'Sembakoeng' wordlist in Genderen Stort (1916).

4.11 Kalabakan Murut language

Kalabakan Murut is spoken only along the Kalabakan River in Tawau District. It represents less than 1000 speakers, probably as few as 450-600.

Whelan (1970) and Keith (1981) record some of Kalabakan's turbulent history centred around the government clash with the Murut there in 1891. They were also involved in the conflict during the Confrontation in 1963. But nothing is known of their prior history. Informants all said that as far back as anyone knew they had lived on the Kalabakan River.

The Kalabakan Murut could offer no name for their language other than 'Tawau Murut' or 'Kalabakan Murut'. They adamantly rejected the term Tidong, possibly on religious grounds as much as anything, since Tidong speakers are usually Muslim while Kalabakan Murut do not generally subscribe to a religious faith. In fact, a group of speakers of the Tidong dialect (presumably similar to the Tarakan dialect) live at Kalabakan TU, so the distinctions are very apparent to both groups. Some Ida'an and Sungai informants in Bukit Garam KN considered Kalabakan Muruts part of the Tenggara tribe. It was not possible to verify this at Kalabakan TU.

The lexically closest dialect to this group is indeed the Tenggara dialect of the Baukan language group, but Kalabakan speakers showed very low understanding of the Tenggara story. They registered highest scores on a Serudung Murut story, but even that showed low comprehension. No other linguistic research in this language is known.

4.12 Serudung Murut language

Serudung Murut is spoken along the Serudung River in Tawau District and since 1963 in the village of Serudung Baru TU about 12 miles north-east of Tawau town.

The Serudung Murut language group consists of an estimated 350-450 speakers, about two-thirds of whom now live in Serudung Baru TU. Informants could tell nothing of their place of origin other than "along the Serudung". They mentioned Kalabakan TU as the place where the speech is most similar to their language,

but not the same. They gave no name for themselves other than 'Murut' or 'Serudung Murut'. Others have called them 'Tawau Murut' or 'Tidong', thus grouping them with the Kalabakan Murut.

Lexically, they are the most remote of the Murutic dialects, sharing the highest number of lexical items with Kalabakan Murut, followed much more distantly by Pensiangan Murut and Sembakung Murut. It is conceivable that this language represents a very early, unmixed speech form since until 1963 the whole group was very isolated from the other native groups.

Serudung Murut speakers showed comprehension of a very simple Kalabakan story but this is not mutual. They did not demonstrate intelligibility of either Pensiangan Murut or Nabay, nor of a Tidong story from Lidung LS. No other linguistic research in the dialect is known.

5. NATIONAL LANGUAGE INTELLIGIBILITY

Included in 26 of the 27 Murutic test sets was a national language (NL) reference tape.

Figure 41 gives a summary of the NL testing results by village along with the sociological data for the entire corpus at each test point, whether or not every subject participated in the NL test. A total of 224 subjects from six districts in the state were tested. The village averages for intelligibility scores ranged from 21% in Kg Lima PN to 93% in Masak KU. Both were represented by a fairly balanced corpus. The most obvious difference between the two corpora, as would be expected, is the average educational level attained.⁵⁵ The Masak KU test corpus averaged seven years of education. The Kg Lima PN subjects averaged less than one year of education. The correspondence between education and intelligibility scores, however, is not maintained consistently throughout the village averages. The village averages of sex and age do not show a pattern either. More revealing correspondences are seen in the average of the individual scores as discussed below.

Figure 42 shows the relative level of NL intelligibility among the Murut population by district. Averages range from 47% in Pensiangan District to 81% in Beaufort. However, the Beaufort corpus represents only nine subjects of one Murut village. The high score registered in that village may be due in part to the large Kedayan (→ Malayic) population in the immediate area and the need for a trade language. The low score in Pensiangan District probably reflects the lower degree of development in that district.

The present development pace of Keningau District as well as the "cosmopolitan" flavour of the community as a whole, which creates a need for a trade language, is probably accurately represented in the 74% intelligibility of the NL tape.

The Kinabatangan District average in Figure 42 includes the results from the two test corpora used by Hurlbut and Pekkanen (to appear).⁵⁶ This lower 51% intelligibility probably better represents the Murut community as a whole in that district than the higher 69% shown in Figure 41. In Tenom and Tawau districts the average intelligibility scores of 56% and 50% probably show a true average of the Murut communities at present, but the recent introduction of more schools is fast changing the general use and comprehension of Bahasa Malaysia among the younger generation.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Bukau BT	81	6/3	50	22-71	3	5	2
Ambual KU	62	6/4	51	28-75	3	9	3
Apin-Apin KU	81	5/3	48	31-68	5	5	3
Dalit KU	85	9/1	36	17-58	6	6	4
Dangulad KU	84	6/4	38	24-58	5	9	4
Keramatoi KU	65	8/1	26	17-45	8	6	6
Kuala Biah KU	55	6/4	50	32-78	1	4	0
Kuangoh KU	86	5/5	39	21-62	4	7	3
Liau Laut KU	64	4/6	38	21-58	5	6	3
Masak KU ¹	93	5/5	36	19-70	7	9	7
Minansut KU ²	86	4/4	42	15-62	6	5	4
Sook KU	62	3/5	50	19-68	2	6	2
Tulid KU	68	5/5	32	13-56	7	6	4
Inarad KN	69	7/3	26	14-60	8	9	6
Kg Lima PN	21	5/5	37	20-50	1	3	0
Kg Empat PN	36	5/2	43	17-60	2	6	2
Pandewan PN ¹	60	6/4	32	12-62	6	6	3
Pensiangan PN	70	5/3	40	19-60	4	4	2
Selungai PN ³	50	8/3	31	18-48	4	4	2
Sapulut PN	59	5/5	39	18-59	3	6	2
Maligan SG ⁴	--	6/3	26	13-50	8	7	6
Kalabakan TU ¹	56	5/4	30	14-50	0	0	0
Serudung Baru TU	45	6/4	33	16-46	1	6	1
Entabuan TM	68	3/3	33	18-47	5	7	6
Langsat TM	33	5/5	46	16-63	4	6	2
Saga TM ¹	82	5/1	28	22-46	3	4	2
Tomani TM	59	5/4	30	21-58	6	5	4

Figure 41: Sociological data for the 27 Murut test points showing national language intelligibility scores. (Scores are given as percentages. The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average number of years of education which those subjects had received (A.E.R.), and the average number of years of education per subject in the whole corpus (A.E.S.). Raised numbers after the village indicate irregular testing situations.⁵⁷)

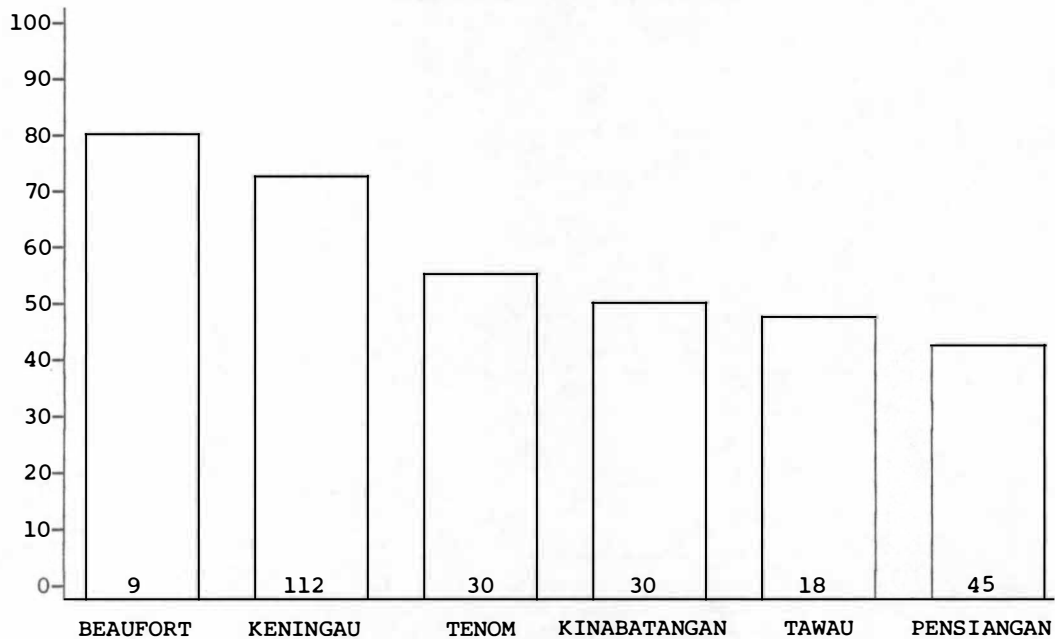


Figure 42: Average intelligibility of the national language tape by district. (The number of subjects considered in each average is noted in the respective column.)

Figure 43 shows the results of intelligibility of the NL tape by the level of education achieved, ranging from "no education" to "10 or more" years of formal training. The intelligibility results were directly influenced by this factor as seen in the steady rise from 47% among the non-educated to 95% among those receiving 10 or more years of education.

National language intelligibility testing demonstrated that the major factor influencing the intelligibility level attained was education in the medium of Bahasa Malaysia. Those with no formal training fared least well while those subjects having had some formal training demonstrated, on the average, intelligibility directly related to the level of education achieved. The factor of sex seemed to be significant only among the non-educated subjects. No other factors can be isolated with any degree of certainty as an influence on intelligibility. The factor of age appears to be of only limited significance by itself.

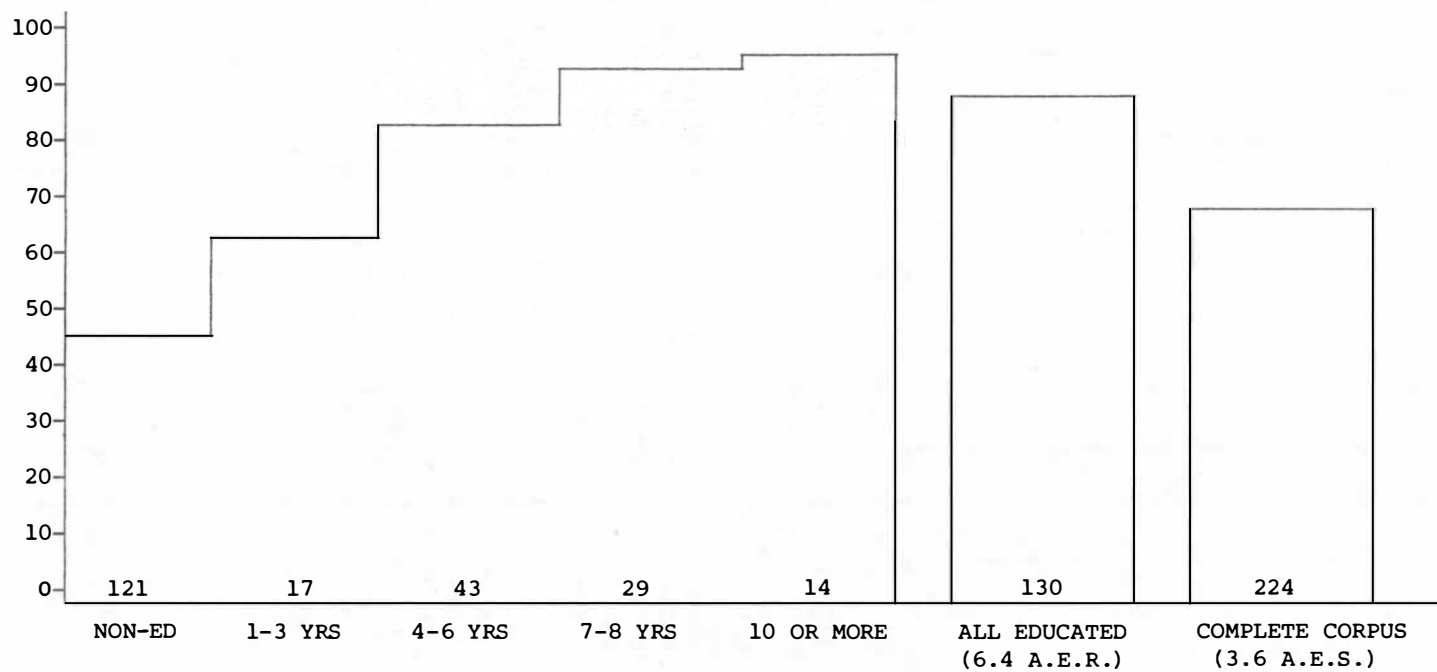


Figure 43: Average intelligibility of the national language tape by the level of education achieved. (The number of subjects considered in each average is noted at the bottom of the respective column.)

6. CONCLUSION

The research of the Summer Institute of Linguistics has gone a long way toward clearing up some questions concerning the classification of the Murutic languages of Sabah and Kalimantan. For instance, it is well established that 'Tawau Murut', while indeed Murutic, in fact, incorporates two Murutic sub-families and is not closely related to Pensiangan Murut, but shares its closest relation with 'Tengara' around the headwaters of the Kinabatangan River.

The delimiting of a distinct Tenggara dialect and its classification with the Baukan language along the upper Sook River is also well established. It relates closer to Northern Murut dialects than to the eastern Murutic dialects or Tidong.

The results of the survey have demonstrated that the relationship between Tidong and Murutic languages is generally on the same level as their relation to the other two Bornean families, Dusunic and Paitanic. It was necessary first, however, to establish the classification of several languages (including Sembakung Murut, Tenggara, Kalabakan Murut and Serudung Murut), formerly thought to be closer to Tidong, as Murutic.

An attempt has also been made to show that the Paluan language, rather than lexically belonging with either the Sumambu/Tagal or Central Murut dialects is situated (both geographically and linguistically) midway between the two language groups. Intelligibility testing seems to indicate the same, though some more testing using Paluan tapes is required.

Another question, not fully answered, is the isolating of four dialects of Keningau Murut. The degree of mutual intelligibility among all these dialects is still unknown. Related to this is the inclusion of Gana with Murutic languages rather than Dusunic. Gana intelligibility of Dusunic Kuijau, as well as Murutic Nabay, is established, making it somewhat of a 'juxtaposition' between Murutic and Dusunic languages.

Even so, much more research needs to be done among the Murutic languages before a more definitive classification can be offered. Specifically, wordlists should be collected from along the Telekosang River in Tenom and the Mengalong River in Sipitang. A wordlist from the Kolod area just west of Sipitang District along the Tengoa River in Sarawak is also needed. If it is possible to collect a wordlist from the Tenggara speakers from the vicinity of the Kuamut River, an interesting lexical link might be found between Kalabakan Murut and Tenggara, and thus other Northern Murut dialects.

In addition to the collection of wordlists, further intelligibility testing is necessary to verify some of the more tentative divisions. For instance, the two Kolod groups, Selungai Murut, Sembakung Murut and Pensiangan Murut should be tested for mutual intelligibility. Since others have considered Lumbis as very different from Salalir it would be prudent to do further testing in this area.

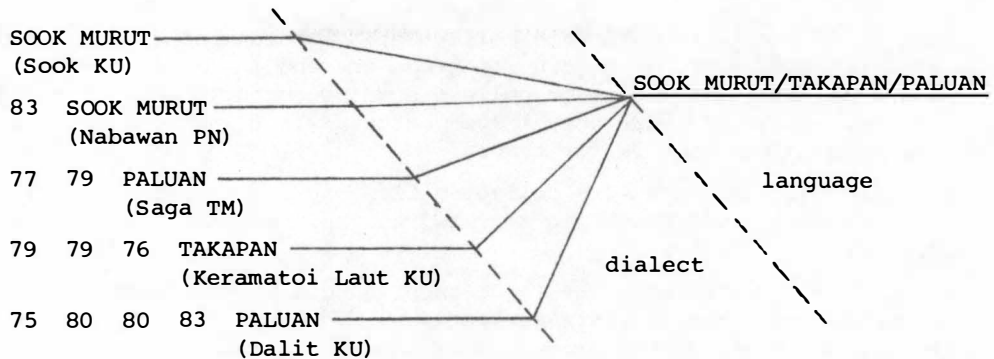
More testing among the Sumambu/Tagal dialects is also necessary. Mutual intelligibility must be tested between Rundum Murut and Tomani Sumambu, Maligan Tagal and Tomani Sumambu, and Maligan Tagal and Salalir. In addition, Salalir intelligibility of Tomani Sumambu and Maligan Tagal intelligibility of Rundum Murut should be tested. If the Telekosang and Mengalong River areas prove lexically distinct, they too should be included in the testing. Other languages which need further clarification by intelligibility testing are the dialects of Keningau Murut and Gana. Since only a Nabay reference tape was used in each of these, and an Apin-Apin Kuijau tape at Gana test points, much more testing could be done to show the present relations among these dialects.

By this time, it seems that the nomenclature used to designate the Murutic language groups is hopelessly confused. But more extensive research among these groups than a general survey allowed may uncover some ethnonyms more descriptive of their ethnic origins and which are agreeable to the local people themselves as well as the professional linguist and anthropologist.

Concerning the intelligibility of the national language, it was shown that whereas comprehension of the NL tape was generally low among the Murut community as a whole, the educational level of the test subjects in the national language was isolated as the single most important factor in learning the language. Those having had some formal education demonstrated nearly twice the comprehension of the NL as those with no formal training. Therefore, as education in the national language continues to expand throughout the Murut community a higher level of intelligibility of the NL is certain to follow.

NOTES

1. The 1970 census listed a total Murut population of 30,947. The 31 December 1978 estimated population was listed at 39,282 (*Statistical handbook* 1978). These census figures tend to mix distinct linguistic and ethnic groups by including the Lundayeh as Murut, the Beaufort Murut as either Murut or Bisaya, and the Gana and Kuijau as either Kuijau, Murut or Kadazan (cf. Ride 1934; Jones 1966; Prentice 1971).
A convergence of several factors was used to arrive at the estimates used in this paper. Some of these factors include: 1) a comparison of the 1960, 1970 and 1980 census figures; 2) a check of available literature (Lee 1965; Jones 1966; Prentice 1971; Sullivan and Regis 1981); 3) estimates from the district officials, village headmen and others of the general populace; and 4) personal visits to more than 40 Murut centres.
2. Rutter (1922:65) derives the term 'Murut' from the Bajau word belud *mountain*. Woolley (1962:1) supposes it was first applied by Brunei or other foreigners. Prentice (1970:370) says only that it was of non-Ida'an origin. Some have thought that it implies *man of woods* (Whelan 1970:44).
3. This folk etymology was related by a young Lundayeh man in Sipitang District and agrees with that noted by Whelan (1970:44).
4. For convenience the two dialects have here been labelled Tagal 'A' and Tagal 'B', though Smith assigned no label to them.
5. This ethnonym could not be verified on subsequent trips to the Murut language area. However, for convenience it will continue to be used until the final analysis in Section 4.
6. A wordlist from Dalit KU was collected after Smith's data was compiled but prior to testing and was manually compared with several of the Central Murut languages. The following chart shows the lexical relations of these comparisons:



On the basis of the above, all five samples could be considered dialects of a single language. The remainder of this paper is written with this in mind. Also, the Dalit KU test point is handled together with Takapan since it shares the highest relation of PSC and is geographically closest to the language group.

7. Of SIL's wordlists, 22 were thoroughly checked and six only partially, but most PSC relations were recomputed. Eight of Smith's lists were not used at all since their high relations with other ones made in unnecessary. In addition, seven new wordlists were added.
8. Wordlists and other data were collected from several Murut villages which had not previously been visited. Intelligibility testing was conducted in five of these: Kuangoh KU (Gana), Kg Empat PN (Tagal), Selungai PN (Selungai Murut), Dalit KU (Takapan) and Pandewan PN (Sook Murut). Data from two remaining villages, Rundum TM and Lumbis KAL, were used in the testing and final classification along with the others.
9. Two of these tests have been included in Hurlbut and Pekkanen (to appear).
10. While there were two Serudung Murut villages, Serudung TU and Serudung Baru TU were considered as representing one village since the latter had migrated from Serudung TU within the past 20 years.

At the Takapan test point it was necessary to visit Keramatoai Laut KU, Ulu Keramatoai KU and Kuala Aging KU, all in the same vicinity, just to find a corpus of nine subjects. For this reason, the test point will be labelled "Keramatoai KU" since all of these villages are located along the Sook/Keramatoai River. The reference tape and wordlist, however, will be labelled "Keramatoai Laut KU" since they were collected in that village.

11. The intelligibility testing was sometimes conducted outside the village for which the wordlist, story or test point was named. Following is a list of such data:

Inarad KN - all the data found in Hurlbut and Pekkanen were collected in Inarad KN, but the second Inarad KN test was conducted at Pinangah KN and Tongod KN using people who had come downstream for trade, medical attention and school.

Labuk KAL - all the data were collected at Serudung Baru TU from a 50 year old man who had spent 14 years in Labuk KAL and from a 40 year old woman who had been born and reared in Labuk KAL.

Lumbis KAL - all the data were collected at Mendalom TM from a Murut man who had been born and reared in Lumbis KAL and spent considerable time there before migrating to Mendalom TM.

Maligan SG - all the data were collected at the town of Sipitang from Murut speakers who had come to town for trade or school.

Rundum TM - all the data were collected at Kemabong TM from a 40 year old man who had moved there only three months earlier and a 42 year old man who came to Kemabong TM for trade.

In addition, some data were collected from migrant villages, i.e., entire village populations which had migrated to their present location relatively recently.

Ansip KU - migrated from the Tagul and Sibangol Rivers in Pensiangan District more than 35 years earlier.

Baru Jumpa TM - migrated from Indonesia. Information and wordlist were obtained from a 40 year old man from Palumomot KAL and a 38 year old man from Kuala Lemumut KAL near Lumbis KAL. Both arrived about 24 years earlier. The former had a Lundayeh wife and the latter had a Lundayeh father and a "Kolod wife from Lumbis".

Kg Empat PN - migrated from Seliu PN on the Rundum/Sibangol River near Pensiangan PN a few years earlier. Kg Empat PN is one of the planned villages of the Nabawan Scheme in Pensiangan District.

Kg Lima PN - migrated from Kabu PN on the Salalir River near the Indonesian border about five years earlier. Kg Lima PN is also a planned village of the Nabawan Scheme.

Kuala Biah KU - migrated from the Tagul River area in Pensiangan District at least 45 years earlier to what used to be a 'Dayak' village.

Saga TM - at least some had migrated from the Dalit KU area about 20 years earlier. The man who gave the wordlist was not in this migration. Both of his parents (he was 52 years old) had originated from the Saga TM area. However, the story and the test subjects had originated from Dalit KU area.

Selungai PN - migrated upstream from Bantul PN, about one mile from the border, during the Confrontation with Indonesia.

12. Results from two Baukan dialect intelligibility tests conducted by SIL personnel using a slightly different approach are found in Hurlbut and Pekkanen (to appear). The tests were at Inarad KN and Kokoroton KN and included Baukan, Kolobuan, Makiang and Rumanau reference tapes. The latter three are representatives of the Upper Kinabatangan language (→ Paitanic). The results of the Murutic testing are included in the discussion in this section.
13. Baukan-Nabay PSC and intelligibility scores support this dialect division (→ Section 3.4). Using either parameter the Baukan village of Tulid KU shows a closer relation to Nabay than does the Baukan village of Inarad KN.
14. This division is further verified in Hurlbut and Pekkanen (to appear). While they record only 44-49% intelligibility of the Makiang and Kolobuan dialects by Kokoroton KN subjects, the Inarad KN subjects scored 82-89% intelligibility on two Kolobuan reference tapes. Not only does this suggest that Inarad KN subjects are assimilating more of the nearby Paitanic language, but it also reinforces the dialect distinction being drawn between Kokoroton KN and Inarad KN. Possibly the two Baukan dialects are diverging even more due to Inarad KN Muruts' frequent contact with Paitanic groups and Tulid KU Muruts' frequent contact with Keningau Murut groups.
15. At the time of testing, it was suspected that Selungai Murut was more closely related to Tagal dialects than the revised lexicostatistics indicate.

16. In addition to the advantage of the hometown tape, another factor that has probably influenced the score is the unbalanced ratio of male subjects to female subjects. The eight male subjects consistently showed higher intelligibility than the three female subjects. In the case of the Sapulut PN tape the males showed 93% intelligibility while the females showed 80%. This fact, together with the advantage of the hometown tape leads one to suspect that a marginal intelligibility score of 80% would more accurately reflect their performance.
17. The Northern Murut subfamily consists of all of the dialects classified by Prentice (1970, 1971) as Murut of his Ida'an family and referred to by authors variously as Northern Murut, Sabah Murut or Ida'an. The former is preferred here because it is most accurately descriptive and lacks the conflict with local ethnonyms that 'Ida'an' has (cf. Smith, in this volume). The use of Northern Murut is not synonymous with Cohen's (1981) use of the term.
18. Some of the Gana villages include Baru KU, Kuangoh KU, Narapot KU, Minansut KU and Tangkungan KU.
19. There is much confusion with ethnonyms applied to the language groups in the Keningau area. Prentice (1970:371) has a very useful chart in which he displays the names used by the various groups to identify one another. He further discusses nomenclature and the use of the ethnonym 'Minansut' for the group here identified as 'Gana'.

The use of the term Gana is retained for two reasons. First of all, Gana is the term commonly heard from the speakers themselves. The term means *lowland* or *plain* and thus accurately describes their present and past situation on the Keningau Plain east of the Pegalan River. Secondly, the source of the term 'Minansut' is not clear. While it may be a legitimate autonym it is not certain to what extent it was ever used. The SIL researchers who have lived and worked in the Keningau area have only heard it in reference to the Minansut KU village which is inhabited by Gana speakers. By Prentice's own admission the use of a particular autonym is often restricted to a single village or clan group.

20. While Prentice (1971) suggests that the items were apparently from a different subdialect than the one represented by his own Nabay list, it is more likely a mixture of several Keningau dialects. Using 294 comparable terms of Baboneau's 1800-item wordlist a revised Kadalakan KU (Nabay) wordlist has a relation of *90 PSC, revised Ambual KU (Ambual) and Liau Laut KU (Dusun-Murut) wordlists both have a relation of *77 PSC, a revised Apin-Apin KU (Apin-Apin Kuijau) wordlist has a relation of *74 PSC and a revised Minansut KU (Gana) wordlist has a relation of *64 PSC.
21. Using 106 comparable terms of Rutter's 185-item 'Keningau Murut' wordlist SIL's revised Nabay lists have relations of *80-85 PSC, with a revised Apin-Apin Kuijau list the relation is *82 PSC, with a revised Dusun-Murut list the relation is *83 PSC, with a revised Ambual list the relation is *85 PSC and with a revised Gana list the relation is *70 PSC.
22. One village leader estimated about 60 individuals.
23. One very old and influential man of the community said his father came from a village named Seraban on the Apin-Apin River and spoke Murut.
24. The position of Sodomon KU between Ambual and Mailo Rivers may contribute to the fact that it is actually more closely related lexically to Ambual KU of the Ambual dialect according to Smith's data. Tuarid Taud KU has also

been mentioned as a member of that dialect. However, lexically and geographically Tuarid Taud KU may better be grouped with Nabay villages. The mixed village of Binaong KU on the west side of the Pegalan River near the road also has some speakers of this dialect.

25. Some of these villages are Masak KU, Patikang Laut KU, Kadalakan KU, Limbawan KU, Dangulad KU, Tuarid Baru KU, and there are some speakers at Kg Keningau KU, Banjar KU and possibly Tuarid Taud KU.
26. Beaufort Muruts along the Padas River inhabit the villages of Saliwangan BT, Menonuk BT, Balibata BT, Papas BT and Malalugus BT. Others live along the Bukau River in Bukau BT, Tamangalong BT and Nalasadan BT.
27. Using 85 comparable terms of Crespigny's 142-item 'Murut Padass' wordlist, a revised Sandiwar wordlist from Bukau BT had a relation of *83 PSC and a revised Timugon wordlist from Langsat TM had a relation of *81 PSC.
28. District and village officials provided the following list of what they thought were Timugon villages: Tuan TM, Dolok TM, Mansalima TM, Naluyan TM, Gunting TM, Baab TM, Kalang TM, Uloi TM, Makakumpai TM, Makalalani TM, Entabuan TM, Amboi TM, Mendalom TM, Tumantalik TM, Mendalom Baru TM, Batu TM, Batu Batu TM, and some speakers in Tenom town TM, Saga TM, Sindugan TM, Mekaniton TM, Kanar TM and Paguokon TM.

Residents of Langsat TM added those villages they considered the same as theirs: Logud TM, Simpangan TM, Mantailang TM, Pagausangon TM and Pulong TM.

29. Using 106 comparable terms of Rutter's 185-item 'Tenom Murut' wordlist, a revised Langsat TM (Kapagalan) wordlist has a relation of *88 PSC and a revised Entabuan TM (Poros) wordlist has a relation of *83 PSC. So in this data also it related more closely to the Kapagalan subdialect.
30. Some of the villages include Tulid KU, Pau KU, Tiong KU and part of Delayan KU. Keith (1936) added Tambalong (KN?) and Kindasan (KN?). Other Murut added Bunang KU, Mapila KU, Kapulu KU and Bersanan KU as Baukan, but these may better fit the Paluan language group. Residents of Limbawan KU called themselves Baukan but lexically the wordlist from there classifies better with Nabay.
31. The inhabitants of Kokoroton KN had dispersed between the time of SIL's first visit in October 1978 and the second visit in March 1981.
32. This village was mentioned by several people from Bukit Garam KN, Pinangah KN and Inarad KN. A district official in Tongod KN stated that they migrated within recent history from Inarad KN itself. However, Pryer (1881), Rutter (1929) and Keith (1981) designated this area as Tenggara and it may be they were there for at least several generations before recently dispersing.
33. Rutter's 'Tenggara' wordlist seems to be about *83 percent cognate with both SIL's Inarad KN and Labuk KAL lists and *82 percent cognate with SIL's Kalabakan TU list. It would be worth comparing the Tenggara language spoken along the upper Kuamut River if a speaker could be found, since Kuamut KN is nearly equidistant from Inarad KN and Kalabakan TU. SIL's samples from these latter two villages have a relation of 77 PSC between them.
34. The various village headmen and district officials gave the following list of villages. These include those "like" Saga TM: Batutingkang TM, Saga Ulu TM, Angalor TM with some speakers in Mansasoh Baru TM, Bambam TM, Sindugan TM and Mansasoh TM. The villages "like" Sook KU included Kuala Binuoh KU,

Babason KU, Tinagalan KU, Binakaan KU and Nabawan KU/PN with some speakers possibly at Bersanan KU, Susulat KU and Tiulan KU. The villages "like" Dalit KU included Kalampon KU, Tantaluan KU, Belinin KU, Ulu Puntih KU, Keramatoi Ulu KU, Keramatoi Laut KU, Punggol KU, Kahabah KU, Linayun KU, Kuala Aging KU, Malaing KU, Ramidi KU and Lumiri KU. It is not certain how far this extended, but some included Maluntok PN, Katalirus PN, Batutai PN, Sapulut PN (not the one on the Sapulut River), Suan PN, Lion PN, Ansapul PN, Damiliew PN and Pun Batu PN.

35. Using 105 comparable terms of Rutter's 186-item 'Peluan' wordlist, a revised Saga TM list has a relation of *89 PSC and a revised Dalit KU list has a relation of *84 PSC.
36. Informants at Pandewan PN included Matiku PN, Kagupil PN, Tataluan PN, Langkuku PN, Ulu Pandewan PN, Pamentarian PN, Takul PN, Mogontok PN, Pimiris PN, Labug PN, Hongkuluson PN and Kagulio PN as having Pandewan Murut speakers.
37. For instance, Smith included both Salalir and Lumbis villages in his classification, Prentice did not. On the other hand, some of Rutter's Sumambu villages would undoubtedly be included in Prentice's Paluan classification.
38. Most of the residents of Seliu PN have moved to Kg Empat PN, part of the Nabawan Scheme PN. Most of the residents of Talingbakas PN, Silom PN, Kinabalau PN, Ulu Saliri PN and others have moved as well.
39. Ansip KU and Kuala Biah KU belong to this migration. More recently, it is reported, the Murut of Senagang Baru KU moved from Pensiangan District to Keningau District. Others have simply moved into other established villages such as Mambulu KU.
40. Harrisson (1967) mentioned two terms offered by his father/son informants, each rejecting the other's term. The son rejected 'Tagal' but suggested Ulun Kirong. The father also rejected 'Tagal' but rejected the latter term as well. He offered classification by river as being significant, recognizing a) Ulun Saliu, b) Ulun Talankai and c) Ulun Sapulut. But, as later discovered, this did not distinguish specific ethnic/cultural groups but only where these groups were living at present. If they moved, the ethnonym would be changed to the name of the new location.
41. Throughout Pensiangan District informants rejected this ethnonym, but were aware that it was applied to them. It seemed to carry no more (or less) meaning than the term 'Murut'. However, in Sipitang, Tenom and Keningau districts they used this to identify themselves. The Tomani Native Chief even used 'Murut Sumambu Tagal', taking no chances.
42. According to Prentice (1970) Genderen Stort's (1916) wordlist belongs to this dialect.
43. The majority of the population of Kabu PN, near the Sabah/Kalimantan border, have moved to Kg Lima PN on the Nabawan Scheme PN. Masanoi TM informants stated that they arrived from the Salalir River area via Maligan SG ten years earlier and they mentioned residents from several other villages in the Tenom area who they thought also came from there, including Kauaran TM, Gumisi TM, Maugos TM, Belumbang TM, Eubai TM, Lampayak TM, Sungai Api TM and Abul TM, but this could not be verified.

Interestingly, the Masanoi TM wordlist patterned more closely with the Tomani TM list than the Salalir sample from Kg Lima PN. It is not certain whether this is due to recent borrowing or their original speech, but for

- that reason, the 'Tagal' or 'Salalir' villages around Tenom TM remain suspect. Other factors may cause them to be grouped better with the Salalir dialect, but further investigation is necessary.
44. The Murut along the Salalir River and in Kg Lima PN probably number about 500, not including any in Kalimantan or those who may have migrated elsewhere. The Tenom migrant villages could number as many as 1000-1200 speakers.
 45. SIL's Lumbis KAL list has been included on the basis of lexical comparisons, geography and some very limited intelligibility testing, but it does show some distinctiveness. Prentice (1970), using other wordlists, separates Lumbis from Salalir on the basis of 72 PSC and because of morphological differences he recognises (1971:299). Prentice (1971:4) estimated the entire Lumbis population in Sabah and Kalimantan as consisting of 3000 (?) members. It is possible that as many as a third of these could be in Sabah.
 46. Three male and two female subjects listened to a simple Lumbis KAL story and averaged 95% intelligibility of the nine questions. They said it was the same as their own dialect, but they also said the same thing about Pensiangan Murut.
 47. This would include such villages as Katubu TM, Kapulu TM, Emagit TM, Rundum TM and Salang TM, all in Tenom District.
 48. Using 106 comparable terms of Rutter's 185-item 'Rundum Murut' wordlist SIL's Rundum TM list has a relation of *84 PSC, the revised Tomani TM list has a relation of *81 PSC and the revised Pensiangan PN list has a relation of *78 PSC.
 49. Prentice (1970) isolated the Telekosang River region as a separate sub-dialect. Since SIL collected no list from that area, researchers had to depend on Tomani TM informants who considered it different from Rundum Murut and similar to their language. Since it is located on a different river, though, it is likely that it represents a unique dialect or sub-dialect.
 50. Researchers were told the Tomani Native Chief held jurisdiction over 876 families. Some of the villages listed as having Tomani Sumambu speakers are Tomani TM, Bangkulan TM, Kalibatang TM, Marais TM, Bekuku TM, Melutut TM, Pangulobon TM, Katambalang TM, Kemabong TM, Ulu Tomani TM and Sugiang Lama TM. The Mengalong River group includes the villages Ulu Muaya SG, Ulu Mengalong SG, Kabawang SG, Kalangsat SG, Solob SG, Marau SG, Mengalong SG, Muaya SG, Kuala Muaya SG, Ulu Malaman SG, Malaman SG, Merapok Ulu SAR and Limpaki SAR. Researchers were told these Murut spoke 'Tagal' or 'Murut' (not Lundayeh) but no additional data was collected. Excluding any of the migrant villages mentioned above (note 41) but including the Mengalong River area, this whole dialect group could number as many as 4000-5000.
 51. This includes such villages as Bol SG, Ulu Bol SG, Ibul SG, Maligan SG and Iburu SG.
 52. The Kolod wordlist relates only 33-35 PSC with the three Lundayeh wordlists but *59-74 PSC with representative lists of the Northern Murut dialects. With eastern Murutic dialects it relates *56-71 PSC. With the Lumbis KAL list it has a relation of only *69 PSC even though they are supposedly very close geographically. However, SIL's Kolod wordlist was not thoroughly revised and it is believed that it may actually share a slightly higher relation with Sumambu/Tagal when the same considerations are afforded it

such as synonyms. If so, it too could be considered a Northern Murut language leaving Selungai Murut, Sembakung Murut, Kalabakan Murut, and Serudung Murut as more truly making up the Eastern Murutic subfamilies.

53. Also included are such villages as Bantul PN, Inpagis PN, Buntulon PN, and Kuala Temalacak PN. One informant listed 17 other villages which he felt were closer to Selungai Murut than to Pensiangan Murut. These are located along the Kuala Sumatalun River and further upstream along the Sapulut River to Agis PN and to Tinandok PN on the Tagul River. The author's impression is that the first information is more accurate, though differences probably appear gradually along the river. Some of the test subjects or their parents were born in the Indonesian villages of Linsayo KAL, Sematalung KAL, Isanal KAL, Kg Sembakung KAL and Simantabul KAL, "40" miles from Selungai.
54. The wordlist and information was collected from informants who originally came from Labuk KAL near the mouth of the Sembakung River. Labuk KAL is about eight days paddling from the Selungai Murut area, yet the two groups remain more closely related to one another (*77 PSC) than to any other group.
55. The number of years of education received includes the current year. No distinction is made between formal education of the national educational program and formal education of an adult educational program or formal religious instruction if it was conducted using Bahasa Malaysia.
56. The NL testing results of Hurlbut and Pekkanen (to appear) are used only on this district chart to provide a larger, more accurate picture. Their results would only minimally affect the much larger corpus using all Murut subjects. Thus it was not considered necessary to include their scores, since it would require considerable time to incorporate into the results presented here.
57. The irregular testing situations were as follows:
 - 1 One person refused to listen to the tape. His results were included in the Murut intelligibility scores and the sociological averages.
 - 2 One man seemed to be hard of hearing and was excluded from all the test results and sociological data. He showed 50% intelligibility.
 - 3 This village included only one subject who took the national language test.
 - 4 This village is included here for the sociological data only. No national language test was conducted.

THE PAPAR LANGUAGE

Phyllis A. Dunn

0. INTRODUCTION

The people of Sabah who speak the language called Papar live in two villages in the north-western section of Kuala Penyu District. The total number of Papar speakers in Sabah is approximately 600-800 (Figure 1).¹

The Papar people are sometimes referred to as the "Bajau Bukit" (lit. *hill Bajau*). This ethnonym serves to cause some confusion since the Papar language relates to the West Coast Bajau language at a range of only 39 to 44 percent of shared cognates (PSC). In addition, the people speaking this language prefer the name Papar when referring to themselves and to their language.

According to a legend, these people originally came from the area near the town of Papar, but the Papar language and the language spoken now in Papar District demonstrate only a very low lexical relationship.

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classifies the Paper language as a separate language within the Dusunic family of languages. Its highest lexical relationship is 72 PSC with the neighbouring Tatana language of Bundu KP. The Papar language relates with the other Dusunic languages in the range of 51-66 PSC, and with the Murutic languages in the range of 53-66 PSC. There are no dialects within the Papar language.

2. TESTING PROCEDURE

The purpose of intelligibility testing at the Papar village of Tinambak KP was to confirm the lexicostatistical classification of Smith that Papar is a distinct language within the Dusunic family of languages. Also since PSC relations between the Papar language and Murutic languages were close to PSC relations between Papar and Dusunic languages, it was considered significant and so desirable to test Papar with a Murut language test tape as well.

Both data collection and intelligibility testing were done only at the Papar village of Tinambak KP since it is the larger of the two Papar villages. Reference tapes from the Tatana language of Bundu KP, the slightly more

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 225-229. *Pacific Linguistics*, C-78, 1984.

© Phyllis A. Dunn

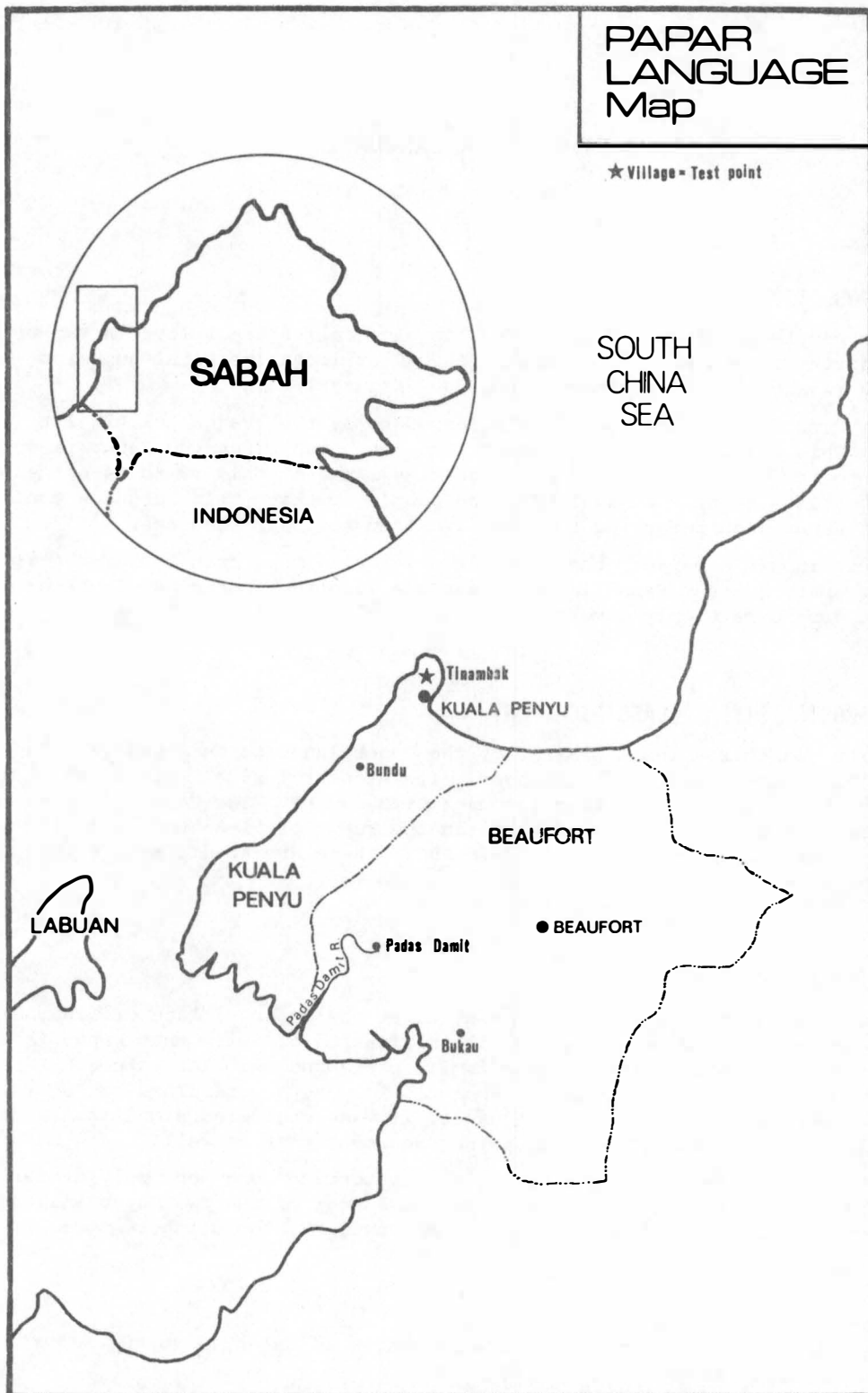


Figure 1: Papar language map

distant Bisaya language of Padas Damit BT, and the Murut language as spoken in Bukau BT were tested. The Murut reference tape from Bukau BT was chosen since it represents the Murut dialect which is closest to the Papar language group both geographically and in terms of PSC relations. In addition, the reference tape from Taginambur KB was used as representing the standard Kadazan/Dusun language and the Bahasa Malaysia tape was used in order to test national language intelligibility.

All of these tapes were considered to be clear with regard to technical quality. The content of the tapes was also considered to be good so as to provide a fair test of intelligibility of the languages under consideration.

Because of the relatively small number of speakers, the Papar language reference tape was not tested in other areas. When Tatana and Bisaya speakers were questioned about the Papar language, however, they maintained that they were not able to understand Papar and so use the national language, Bahasa Malaysia, as a means of communication with the Papar people.

3. TEST RESULTS

Figure 2 gives the results of intelligibility testing at the Papar village of Tinambak KP. The intelligibility testing scores of Papar speakers on the Tatana and Bisaya reference tapes are significantly higher than their respective PSC relations. The intelligibility score on the Tatana story was 96% as compared with a PSC relation of 72. Likewise, average intelligibility of the Bisaya story was 79% compared with 61 PSC. These elevated intelligibility scores appear to be due to language learning on the part of the Papar people. The Papar are a small group, surrounded by predominantly Tatana villages, with Brunei communities located within the area, and Bisaya villages about ten miles away. Most of the Papar people interviewed stated that they were forced to learn another language in order to communicate in the nearby town of Kuala Penyu. Many are fluent in the Tatana language, while others have learned the Brunei language, and the children are studying Bahasa Malaysia in school. Language learning is further encouraged by the frequent intermarriage of the Papar with neighbouring groups. Only two of the subjects interviewed had a Papar father, mother and spouse. The others had at least one member of their immediate family who was not Papar. This has resulted in a nearly total absence of monolingual Papar speakers.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Intelligibility scores on the Bahasa Malaysia reference tape were very high, with only one person in the corpus of eight scoring 80%, and the others scoring 90-100% (Figure 3). The high average score of 93% would support the statements of the Papar people that they have had to learn the national language or the language of a neighbouring group in order to communicate in their daily lives with persons of other language groups. It was also said that many parents encouraged the use of Bahasa Malaysia in the home so that their children will do well in school.

REFERENCE TAPE	DUSUNIC			MURUTIC
	Bundu KP (Tatana)	Padas Damit BT (Bisaya)	Taginambur KB (Kadazan/Dusun)	Bukau BT (Beaufort Murut)
TEST POINT				
Tinambak KP (Papar)	96 (72)	79 (61)	24 (54)	28 (64)

Figure 2: Results of intelligibility testing among speakers of the Papar language in the village of Tinambak KP. (Intelligibility scores are given as percentages; PSC figures are shown in parentheses.)

TEST POINT	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Tinambak KP	93	3/5	45	20-61	4	6	3

Figure 3: Comprehension of the national language in the village of Tinambak KP (Papar language). (AVG. SCORE is given as a percentage. RANGE indicates youngest and oldest test subjects. EDUCATION figures indicate the number of subjects who had received some formal education (ED.), the average number of years of education those subjects had received (A.E.R.), and the average number of years of education for all of the subjects in the test corpus (A.E.S.).)

5. CONCLUSION

While the actual intelligibility scores from Tinambak KP would tend to suggest a closer relationship between the Papar language and the Tatana and Bisaya languages than the PSC relations and the lexicostatistical classification of Smith would indicate, the elevated scores are better explained by language learning than by language affinity. The Papar people themselves stated that through intermarriage and the need to communicate with neighbouring groups, other languages are commonly learned. Those other languages would especially be Tatana, Bisaya, and Brunei which is closely related to the national language, Bahasa Malaysia. The test scores confirm this since the scores in Tatana and Bahasa Malaysia were high and the score in Bisaya was higher than would have

been expected based on the PSC figures alone. The inability of the Tatana and Bisaya speakers to understand the Papar language, and the emphasis by the Papar people that the other languages were learned would seem to rule out the possibility that the languages are closely related linguistically.

NOTE

1. Since the Papar people have not been separated from the other larger ethnic groups in Sabah census figures, this is only a population estimate of the number of Papar speakers, based on interviews conducted by the technicians who did the intelligibility testing at Tinambak KP.

THE DUMPAS LANGUAGE

Julie K. King

0. INTRODUCTION

The Dumpas language is spoken in the village of Perancangan in the Labuk-Sugut District. This village is located on the Terusan Sapi midway between the Sapi River and the Labuk River. During a visit to the village of Perancangan LS in 1980, residents said that Dumpas was not spoken in any other village in Sabah and that the total population of their village was approximately 500. The 1970 Sabah census figures however, state that there were at that time 1,077 Dumpas in Sabah.

When the inhabitants of Perancangan LS were questioned as to the origin of their village, they indicated that they had moved from an area north of their present location more than 50 years ago. Rutter (1929) mentions the Dumpas as a group closely allied with the Tambanua (→ Paitanic). Since the Tambanua presently live along the Sugut and Paitan Rivers and even farther north in Pitas District, it is possible that the Dumpas of Perancangan LS originally migrated from one of those areas (Figure 1).

1. LEXICOSTATISTICAL CLASSIFICATION

In a comparison of 325 wordlists from villages throughout Sabah, the Dumpas wordlist from Perancangan LS had its highest relation of 80 percent of shared cognates (PSC) with the wordlist from the village of Buang Sayang KN (→ Eastern Kadazan). The relations between the Dumpas wordlist and other wordlists of the Dusunic family of languages ranges from 55-75 PSC. But the relations between Dumpas and languages of the Paitanic family (→ Paitanic) range from 61 to 79 PSC. It is therefore unclear from this lexicostatistical classification whether the Dumpas language actually patterns more closely with the Dusunic or the Paitanic languages of Sabah (→ Smith, in this volume).

2. TESTING PROCEDURES

Because of the fact that the Dumpas language seems to pattern lexico-statistically on the border between the Dusunic family of languages and the Paitanic family of languages, it was considered necessary to do intelligibility testing at the Dumpas village, using stories from both Dusunic and Paitanic

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 231-235. *Pacific Linguistics*, C-78, 1984.

© Julie K. King

231

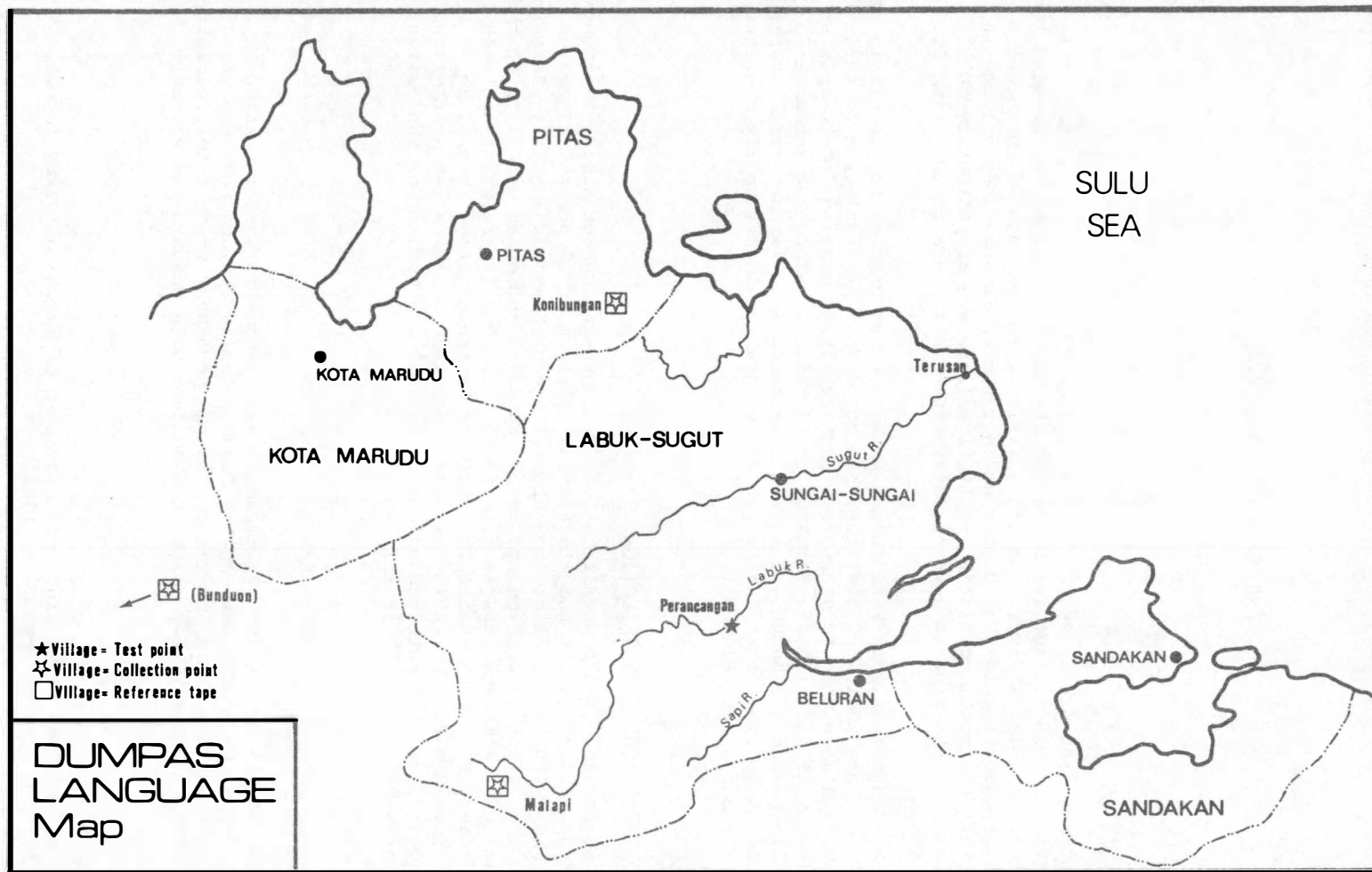


Figure 1: Dumpas language map

villages. It was assumed that intelligibility would be higher with the Dusunic stories than with the Paitanic ones since the village where Dumpas is spoken is located in a predominantly Dusunic-speaking area. Two stories from the Dusunic family of languages were chosen: one was from the village of Malapi LS (→ Eastern Kadazan) and the other was from Bunduon PG (→ Kadazan/Dusun). The latter was chosen since it was representative of the Kadazan/Dusun language which is used on the radio and in the newspaper in Sabah. A third story which was used for testing in this village was the story from Konibungan PS, representing a language from the Paitanic family of languages.

In addition to these three stories, a hometown story was used in testing and a national language test tape story was used. All of the tapes were considered to be good with regard to content and technical quality.

3. TEST RESULTS

Figure 2 displays the results of intelligibility testing at the Dumpas village of Perancangan LS. The results show that the intelligibility testing scores were not predictable from the PSC figures alone.

REFERENCE TAPES TEST POINT	DUSUNIC			PAITANIC
	Dumpas Hometown (Perancangan LS)	Eastern Kadazan (Malapi LS)	Kadazan/Dusun (Bunduon PG)	Tambanua (Konibungan PS)
Dumpas (Perancangan LS)	100 (100)	57 (70)	44 (62)	87 (75)

Figure 2: Dumpas intelligibility of two Dusunic reference tapes and one Paitanic reference tape. (PSC relations are given in parentheses.)

From Figure 2 it can be seen that when ten subjects were tested at the village of Perancangan LS with the Dusunic stories from Malapi LS and Bunduon PG, the intelligibility scores were significantly lower than the corresponding PSC figures. On the other hand, when Dumpas speakers were tested with the Paitanic language family test tape from Konibungan PS, the opposite was true; that is, the intelligibility testing results were higher than the corresponding PSC figure.

Venturing a hypothesis with regard to the origin of the Dumpas people could provide a partial explanation for these seeming discrepancies. It is possible that the Dumpas people are originally from an area on the Sugut River to the north of their present location or from even farther north in the Tambanua area of Pitas District (→ Section 0).

The predominant language spoken in those areas is Tambanua/Sungai which is a Paitanic language. If it is true that the Dumpas people were originally from one of those areas it would also explain the fact that they embrace the religion of Islam, since the Tambanua/Sungai are predominantly Muslims and the other villages surrounding them in their present location either embrace Christianity or consider themselves religionless.

It would then be understandable that when the Dumpas people moved to their present location more than 50 years ago that they began to pick up lexical items from the language of their neighbours. That would explain the high PSC relations of Dumpas with the wordlist of the Eastern Kadazan language in the initial lexicostatistical analysis.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Figure 3 shows the results of testing ten persons in the village of Perancangan LS with the national language intelligibility test story. Four of the ten subjects received a perfect score on this test while one other had only one question marked incorrect. The lowest score received was four incorrect answers out of the ten possible.

TEST POINT	AVG. SCORE	SEX M/F	AGE			EDUCATION		
			AVG.	MED.	RANGE	ED.	A.E.R.	A.E.S.
Perancangan LS	79	6/4	28	33	16-50	7	8.2	5.8

Figure 3: Results of national language intelligibility testing with sociological data for the Dumpas village of Perancangan LS. (Scores are given as percentages. The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average years of education which those subjects had received (A.E.R.), and the average years of education per subject in the test corpus as a whole (A.E.S.).)

Figure 4 displays the same data as Figure 3, but in Figure 4 the data is sorted out with regard to the sex of the subjects, the age of the subjects, and whether or not the subjects had received any formal education. It can be seen that male subjects under the age of 35, who had received some formal education received higher scores on the national language intelligibility test than did their counterparts in the opposing categories.

The rather high overall result on the national language intelligibility test at Perancangan LS is probably also largely due to the fact that this one village is rather isolated in the centre of a language group with whom their intelligibility is quite low. In order for the two communities to communicate over the years, they have had to find a common language, which in this case could very well be the national language.

TEST POINT	CORPUS SCORE	SEX		AGE		EDUCATION	
		AVG. M SCORE	AVG. F SCORE	AVG. UNDER 35 SCORE	AVG. OVER 35 SCORE	AVG. ED. SCORE	AVG. NON-ED. SCORE
Perancangan LS	79/10	93/6	58/4	84/7	66/3	84/7 5M 2F	66/3

Figure 4: National language intelligibility results in the Dumpas village of Perancangan LS displayed according to sex, age and education differentials. (The average score for each category (shown as a percentage) is followed by the number of subjects for which the average applies. The box containing scores of those subjects who had received formal education (AVG.ED. SCORE) also show the the number of males and females included in the educated group. NON-ED. indicates subjects without formal education.)

5. CONCLUSION

On the basis of the intelligibility testing results from the village of Perancangan LS, it is apparent that the Dumpas language has a closer relationship with the Paitanic languages of Sabah than it has with the Dusunic languages. The historical roots of this language then are most likely Paitanic. It should be noted that this classification is contrary to the initial lexicostatistical classification of Smith who placed Dumpas with the Dusunic languages of Sabah. It seems more likely however that Dumpas is historically a Paitanic language and that over the years Dumpas speakers have borrowed lexical items from the language of the neighbouring villages since the intelligibility results as well as the sociological evidence seem to point in this direction.

It is possible that further testing would reveal Dumpas to be a dialect of one of the Paitanic languages or a separate language within the Paitanic family (→ Paitanic). Since, however, no testing was done to determine Paitanic language speakers' intelligibility of Dumpas, this reclassification can only remain speculation. Further testing of Dumpas with other Paitanic languages, and cross-testing of the ability of speakers of other Paitanic languages to understand Dumpas would be necessary in order to warrant such a reclassification.

THE LOTUD LANGUAGE

John E. Banker

0. INTRODUCTION

Little is recorded about the origins of the Lotud language, except that it is related to other 'Dusun' languages. Appell (1968) gives some evidence for cultural ties with Rungus and suggests that Lotud speakers might have migrated to their present location from other Dusunic areas, but this is inconclusive.

Like other languages in Sabah, Lotud is only one of the names given to the language under consideration in this paper. In Dusunic villages around the town of Tuaran TN the language is known as 'Dusun Tuaran'. Other ethnonyms are 'Latud', 'Latod', 'Suang Lotud' and 'Dusun Lotud'. One dialect of this language spoken in the area around the town of Telipok KK is referred to as 'Dusun Kadayan' by its speakers. In this paper the term Lotud is used to refer to all of the villages which appear to belong to this language group.

The number of speakers of Lotud is approximately 5,000,¹ most of whom live in the areas near the towns of Tuaran TN, Tamparuli TN and Telipok KK (Figure 1).

1. LEXICOSTATISTICAL CLASSIFICATION

Smith's classification (in this volume) of Lotud as a distinct language is based on relations with other Dusunic languages which are lower than 75 percent of shared cognates (PSC). Clayre (1966:6), using a different diagnostic word-list, found that Lotud "is quite separate from all other Dusun dialects and must in fact be regarded as a separate but related language."

Within the Lotud language itself Smith postulates two dialects, "a Lotud dialect chain found only in Tuaran and represented ... by five wordlists each of which is related to the dialect chain by at least 85 PSC, and a dialect identified as Dusun Kadayan as spoken in Sumagit KK which is related to the Lotud dialect chain in the range of 80-82 PSC ..." (see p.28 in this volume). Figure 2 shows these PSC relations.

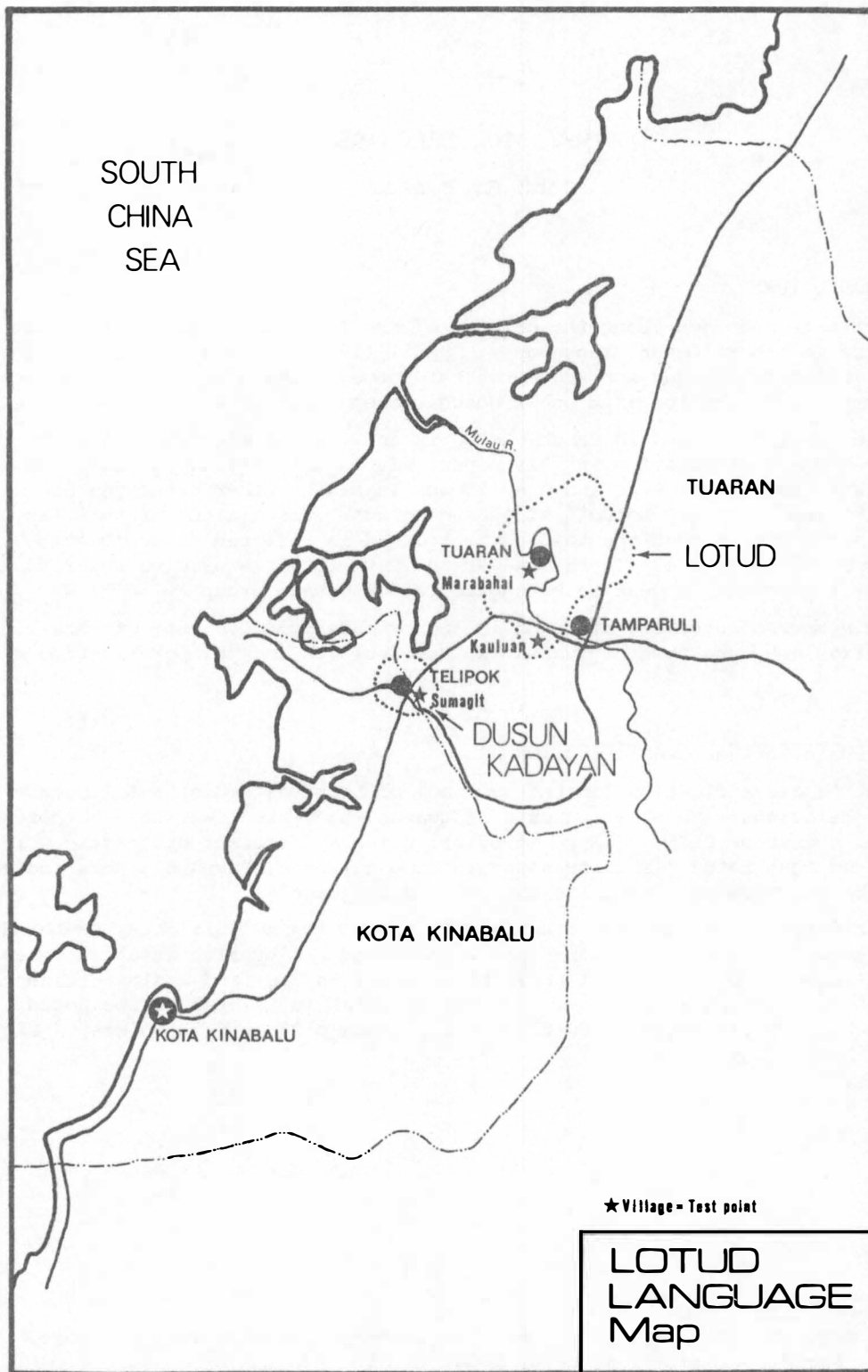


Figure 1: Lotud language map

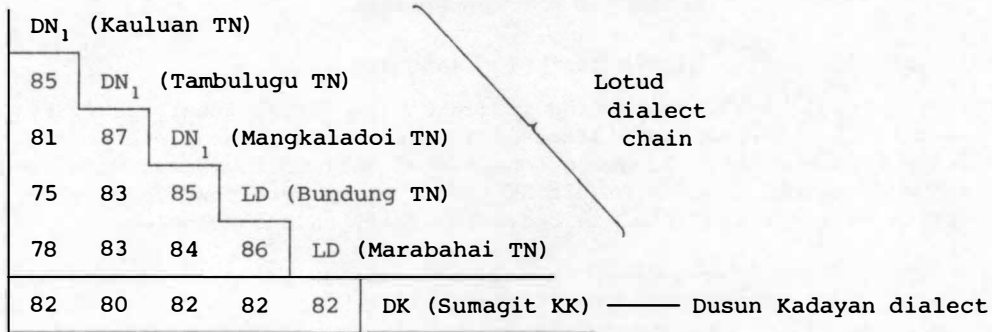


Figure 2: PSC relations of six villages of the Lotud language (per Smith, in this volume). (DN₁ = "Dusun" representing Lotud; LD = Lotud; DK = Dusun Kadayan.)

2. TESTING PROCEDURES

The purpose of dialect intelligibility testing in this study was to clarify the relationships among the Lotud dialects and to better understand the position of Lotud in the Dusunic language family. Testing was also done to determine the status of the Dusun Kadayan dialect within the Lotud language.

Marabahai TN representing Lotud was used as a reference tape for testing the ability of six of the more linguistically central Kadazan/Dusun villages to understand Lotud.

For testing within the Lotud language three test points were chosen: Kauluan TN and Marabahai TN which are on either end of the dialect chain and Sumagit KK representing the Dusun Kadayan dialect. The story tapes from each of these villages were also used as reference tapes (though in the first testing at Sumagit KK, the Kauluan TN reference tape was not used).

Other reference tapes used in the test set were Malangang Baru TN, or Tenghilan TN, representing geographically close Dusun dialects; Bundu Tuhan RU, representing a linguistically central Kadazan/Dusun village and Bunduon PG, representing the Coastal Kadazan dialect of Kadazan/Dusun (the dialect which is used on radio and in the newspapers). The final tape was the national language story.

The Kauluan TN reference story was good. The Marabahai TN story was also good, though possibly a bit easy. The Sumagit KK story was not as good a test since it was a story about the procedures of planting and so the theme and content were familiar. It was also a very short story. All of the tapes were of good technical quality.

The Kadazan/Dusun tapes used in the set were also all of good technical quality. The Bundu Tuhan RU story was a very clear tape, but has been found to be quite easy to understand all over the Kadazan/Dusun area. The story is simple and the words used are common everyday terms. In contrast, the Malangang Baru TN story deals with more specific items that would tend to be different across dialects. The Bunduon PG story is a good test, while the Tenghilan TN story is about the procedures of farming, so would tend to be more easily understood.

3. TEST RESULTS

3.1 Mutual intelligibility within the Lotud language

Results of intelligibility testing indicate high mutual intelligibility between the three Lotud-speaking villages (Figure 3). In general, these figures would indicate that the Lotud language consists of only one homogeneous dialect. Since both the Marabhai TN and Sumagit KK test stories were considered easy, further testing would be essential in order to verify this conclusion.

REFERENCE TAPES TEST POINTS	Marabhai TN	Kauluan TN	Sumagit KK
Marabhai TN	--	89 (78)	98 (82)
Kauluan TN	99 (78)	--	97 (79)
Sumagit KK	97 (82)	97 (79)	--

Figure 3: Mutual intelligibility test results within the Lotud language. (Scores are given as percentages; PSC relations are given in parentheses.)

3.2 Lotud understanding of Kadazan/Dusun

Results of intelligibility testing in Lotud villages using Kadazan/Dusun reference tapes were consistently much higher (80-98%) than the corresponding PSC relations (67-75) (Figure 4). The scores on the Bundu Tuhan RU reference tape are probably higher than they would normally be, due to the very easy story used. The lower 80% score at Sumagit KK using the Malangang Baru TN reference tape cannot be commented on until it is compared with results from further testing of the Malangang Baru TN reference tape.

The major factor in Lotud speakers' understanding of Kadazan/Dusun is the significant amount of contact with the larger Kadazan/Dusun community. Lotud villages are surrounded by Kadazan/Dusun villages, Lotud students go to school with Kadazan/Dusun students, they have contact in the markets, etc. The Lotud speakers generally refer to themselves as Dusun or Kadazan, considering themselves a part of the larger group.

Intelligibility testing shows clearly that Lotud speakers understand Kadazan/Dusun as if it were the same language as their own.

REFERENCE TAPES		KADAZAN/DUSUN			
		Bundu Tuhan RU	Bunduon PG	Malangang Baru TN	Tenghilan TN
TEST POINTS					
LOTUD	Marabahai TN	95 (71)	86 (70)		90 (67)
	Kauluan TN	98 (75)	89 (75)		97 (75)
	Sumagit KK	98 (70)	92 (71)	80 (70)	

Figure 4: Results of intelligibility testing in Lotud villages using reference tapes from Kadazan/Dusun villages. (Scores are given as percentages; PSC relations are in parentheses.)

3.3 Kadazan/Dusun understanding of Lotud

As Figure 5 shows, intelligibility of Lotud by six linguistically central Kadazan/Dusun villages ranges widely from 52% to 89%, while PSC relations only range 69-72 PSC. Two villages show high intelligibility, Kogopon PR (89%) and Pukat TN (89%). Sunsuron TA and Tangaban TA show much lower intelligibility at 56% and 52% respectively. Lack of contact may be a factor in explaining these two lower scores, but cross-testing and more information are needed before any further conclusions can be drawn about Kadazan/Dusun understanding of Lotud.

4. NATIONAL LANGUAGE INTELLIGIBILITY

Intelligibility scores on the Bahasa Malaysia story were generally high (Figure 6). Marabahai TN and Kauluan TN both registered 94%, whereas Sumagit KK showed a somewhat lower score of 82%. The reason for this lower score was that one older woman declined taking the rest of the national language test after missing the first two answers. (If that one score were disregarded, the average would be 91%.)

Higher national language comprehension in the Lotud area is expected since most Lotud villages are near economic centres such as Tuaran TN, Tamparuli TN and Telipok KK, all of which have schools. Education in Bahasa Malaysia is the key factor in intelligibility, but even some with no formal education had high scores.

In the three villages tested an average of one third of the homes now use Bahasa Malaysia for communication at least part of the time, the highest percentage being in Marabahai TN where half of the homes tested do so.

REFERENCE TAPE		LOTUD Marabhai TN
TEST POINTS		
KADAZAN/DUSUN	Kogopon PR	89 (70)
	Pukat TN	89 (71)
	Kionsom KK	81 (70)
	Toboh TA	77 (69)
	Sunsuron TA	56 (72)
	Tangaban TA	52 (70)

Figure 5: Intelligibility testing results in six linguistically central Kadazan/Dusun villages using a Lotud reference tape. (Scores are given as percentages; PSC relations are in parentheses.)

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Marabhai TN	94	5/5	33	19-54	9	8	7
Kauluan TN	94	5/5	35	13-63	7	6	5
Sumagit KK	82	6/4	39	14-70	5	4	4

Figure 6: Lotud intelligibility of the national language tape with sociological data. (Scores are given as percentages; the education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average years of education which those subjects had received (A.E.R.), and the average years of education per subject in the test corpus as a whole (A.E.S.).)

5. CONCLUSION

Lexicostatistical figures clearly show Lotud as a separate language from Kadazan/Dusun, and lower intelligibility scores by linguistically central Kadazan/Dusun villages in their understanding of Lotud confirm this. However, on the local communication level Lotud functions for all practical purposes as a dialect of Kadazan/Dusun with Lotud speakers readily comprehending the larger Kadazan/Dusun language.

Within Lotud itself there seem to be some fundamental differences, i.e. Kauluan TN and/or Sumagit KK may be separate dialects from the rest of Lotud, based on the lexicostatistical comparison. But it is hard to prove by the intelligibility testing that has been done that there is more than one dialect in Lotud. Much contact between speakers of what might be different dialects of Lotud facilitate learning of dialect idiosyncrasies.

Study is needed on a deeper level to understand more fully Lotud's position within the Dusunic languages and the relationship between the various Lotud villages.

NOTE

1. This figure is based on population figures given for each village which speaks Lotud in a report from the District Officer of Tuaran District.

THE BISAYA LANGUAGE

Phyllis A. Dunn

0. INTRODUCTION

The Bisaya (also Bisayah and Besaya) of Borneo are residing in Sabah, Sarawak and Brunei with the majority living in the Beaufort and Kuala Penyu districts of Sabah (Figure 1). Peranio (1972) describes their location as the middle and lower reaches of the rivers in Sarawak and Sabah that flow into Brunei Bay. The major concentrations of Bisaya in Sabah are in the area west of the town of Beaufort along the Padas Damit, Gramma and South Padas Rivers.

There are also many Bisaya villages found along the Klias River intermingled with Kadazan villages. The town of Kota Klias itself is primarily Brunei-speaking.

According to Harrisson (1956), on Labuan Island in Brunei Bay, members of the formerly larger Bisaya population go back many generations. These Bisaya also claim a relationship to Limbang and Klias River Bisaya.

The historical relationships between the Bisaya of north-western Borneo and the groups in the Bisayan Islands PHL are unclear (Harrisson 1962), though the existence of an affinity between the peoples of Borneo and of the Philippines is a long-established fact (Araneta and Bernad 1960).

Headly (1950) and Harrisson (1962) speak of Sabah Bisaya as Muslim wet-rice farmers who inhabit areas along the lower reaches of the Padas and Klias Rivers. Rutter (1922) mentions them around Bundu KP and the town of Kuala Penyu. Bisaya villages today extend along the main road in Kuala Penyu District and throughout the area to the south-west of Menunggang KP.

The 1970 Sabah census lists 14,198 Bisaya in all of Sabah.¹ The majority of Bisaya speakers in Sabah are living in Beaufort District.

Appell (1968) notes the linguistic similarities between Bisaya and Dusun. Prentice (1970) classifies Bisaya as a member of the Idahan language family along with other indigenous languages of Sabah - Murut and Dusun.

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classifies Bisaya as a language within the Dusunic family. Percentage of shared cognate (PSC) figures of Bisaya when compared with other languages within the Dusunic family range 54-67 PSC. A comparison

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 245-250. *Pacific Linguistics*, C-78, 1984.

© Phyllis A. Dunn

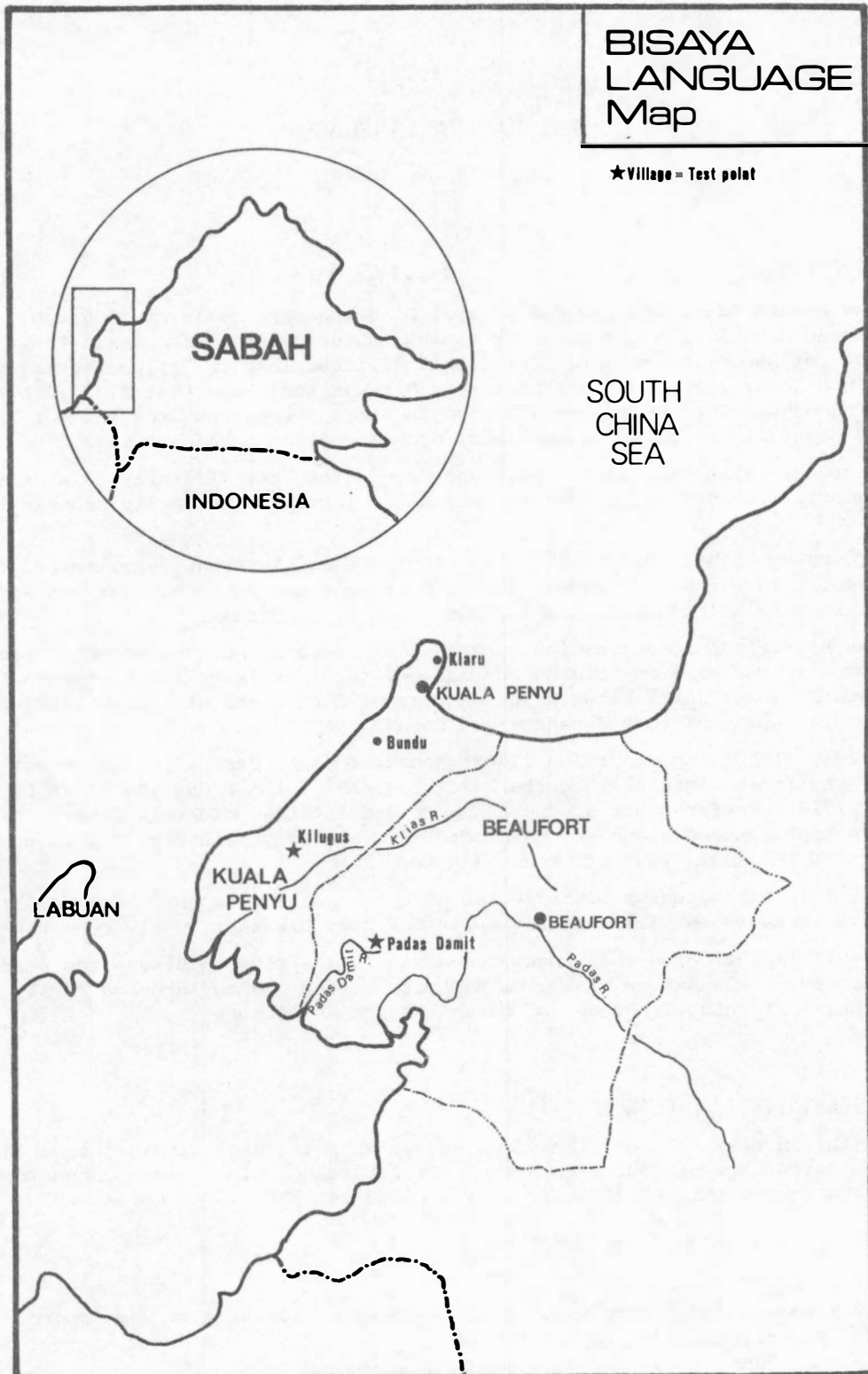


Figure 1: Bisaya language map

of the seven Bisaya wordlists with each other shows high PSC figures, indicating that there are no dialect distinctions between the villages represented by the wordlists compared in Sabah.

Because of frequent queries, sample wordlists of the Bisaya spoken on the Limbawang River in Sarawak and the Visaya language of the Philippines were compared with Sabah Bisaya. The Limbawang River Bisaya was found to have 64-68 PSC with the Bisaya of Sabah and the Binisaya language spoken at Nasuli in Bukidnon Province of the Philippines proved to have only 31-33 PSC with Sabah Bisaya.

2. TESTING PROCEDURES

The purpose of the intelligibility testing was to test the lexicostatistical classification of Bisaya as a distinct language within the Dusunic language family and to determine to what extent Bisaya speakers could understand other languages of the Dusunic family within a wider geographical area.

Two villages were chosen as test points for the Bisaya intelligibility testing. Padas Damit BT was chosen because of its relative isolation from other language groups. It is located 18 miles from the town of Beaufort, near the end of a road which is populated exclusively for the last 12 miles by Bisaya people. Because of the many small rivers and the marshy area to the west of Padas Damit BT, overland contact with the Tatana people in the neighbouring district of Kuala Penyu is minimal or even non-existent. The town of Beaufort is the commercial centre for the Bisaya, and they rarely have reason to visit Kuala Penyu District.

Kilugus KP was chosen as the other test point because of its location in the centre of the Bisaya language area of Kuala Penyu District. This village is situated along the main road between the town of Kuala Penyu and Menumbuk KP, providing the people of Kilugus KP many opportunities for contact with other language groups. Many people from Kilugus KP go to Kuala Penyu regularly where they would naturally have contact with Tatana people.

Tape-recorded stories from four villages were used for the intelligibility testing in the villages of Padas Damit BT and Kilugus KP.

A tape from Taginambur KB was used to represent the Central Dusun dialect. A second tape was from Pinopok BT representing the Coastal Kadazan/Dusun dialect which is geographically the closest Kadazan/Dusun dialect to the Bisaya language area. The Bunduon PG tape was used to represent the Coastal Kadazan/Dusun dialect used in the mass media. And the Bundu KP tape representing the Tatana language was used. Bundu KP, representing the Tatana language has the highest PSC relation with Bisaya and is geographically also very close. The final tape used for testing in the Bisaya villages was the Bahasa Malaysia language tape used to test the intelligibility of the national language.

All of the tapes were considered clear with regard to technical quality and story content. The Taginambur KB story included a few English and Malay words, causing difficulty for a few people. The stories from Pinopok BT and Bundu KP were hunting and fishing stories respectively and so may contain some terminology common to many Sabah languages. The Bunduon PG story was very good since the outcome was not predictable.

3. TEST RESULTS

3.1 Bisaya understanding of Tatana

The Bisaya people were able to understand the Tatana reference tape taken from Bundu KP as well as the Tatana were able to understand the Bisaya tape (→ Tatana). While the PSC figures of Kilugus KP and Padas Damit BT with Bundu KP were only 69 and 62 respectively, intelligibility scores were 86% at Kilugus KP and 85% at Padas Damit BT. It is interesting that the Padas Damit BT score is so high, considering the fact that the people of Padas Damit BT rarely come in contact with Tatana people, while the people of Kilugus KP live within a few miles of Tatana villages, and often meet Tatana speakers in the town of Kuala Penyu.

3.2 Tatana understanding of Bisaya

The Bisaya reference tape from Padas Damit BT was tested in Bundu KP and Kiaru KP, both Tatana villages of Kuala Penyu District. The results were quite surprising since the intelligibility scores were significantly higher than the PSC relations. Tatana wordlists had initially been taken in Bundu KP and Menunggang KP. The PSC between Bundu KP and Padas Damit BT was only 62 while the PSC between Menunggang KP and Padas Damit BT was much higher at 86, indicating lexical borrowing between the Bisaya and Tatana groups living in Menunggang KP. For this reason, Kiaru KP was chosen as the second Tatana test point, since it is geographically one of the most distant from the Bisaya area. The intelligibility scores of Bisaya averaged 82% in Kiaru KP and an even higher 88% in Bundu KP. The higher score at Bundu KP may be explained by the fact that it is located along the main road and so more open to contact with Bisaya people.

3.3 Bisaya understanding of Kadazan/Dusun

Reference tapes of three dialects of the Kadazan/Dusun language were tested in the Bisaya villages. The understanding of these tapes was much lower than their respective PSC figures. The results of this testing are summarised in Figure 2.

The low scores in Kilugus KP may in part be attributed to the fact that the first person tested was very emphatic in his assertion that he could not understand a word of the 'Kadazan' language. After he had listened to the entire tape set, he explained to those who had not yet taken the test that some of the stories were in 'Kadazan'. In the subsequent testing many other test subjects stated that they could not understand 'Kadazan' even before they began to listen to the stories. This problem was only encountered to a mild degree in Padas Damit BT, and although the intelligibility scores are higher, it is still evident that the Bisaya understood very little of the Kadazan/Dusun dialects tested.

REFERENCE TAPES				
		TEST POINTS		
BISAYA	Padas Damit BT	41 (64)	30 (57)	28 (57)
	Kilugus KP	19 (65)	14 (60)	9 (61)
		Pinopok BT (Coastal Kadazan/Dusun)	Bunduon PG (Coastal Kadazan/Dusun)	Taginambur KB (Central Dusun)

Figure 2: Bisaya understanding of Kadazan/Dusun dialects (shown as percentages). (PSC relations are given in parentheses.)

4. NATIONAL LANGUAGE INTELLIGIBILITY

The high scores on the national language intelligibility test of 92% at Padas Damit BT and 97% at Kilugus KP are probably due in part to the fact that a large number of the people tested had attended school (Figure 3). Fifteen of the 20 people tested had completed Primary 6. Many had also attended secondary school. Of the remaining five people, only one elderly person who had not attended school scored lower than 90%, indicating a widespread understanding and use of the national language. Both of the Bisaya villages tested have enjoyed the benefits of government schools for several years.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Padas Damit BT	92	6/3	30	17-65	8	9	8
Kilugus KP	97	8/2	28	17-46	8	7	5

Figure 3: Comprehension of the national language tape in two Bisaya villages.

(AVG. SCORE is given as a percentage; RANGE indicates the youngest and oldest test subjects; EDUCATION figures indicate: the number of subjects who had received some formal education (ED.), the average number of years of education which those subjects had received (A.E.R.), and the average number of years of education for all subjects in the corpus (A.E.S.).)

5. CONCLUSION

The intelligibility testing in Bisaya villages confirmed the fact that the Bisaya language is distinct from the Kadazan/Dusun languages and dialects with which it was tested, and from most other languages within the Dusunic family of languages. It is, however, more highly understood by speakers of the Tatana language within that family than the PSC figures would indicate to be likely. The high degree of intelligibility between Bisaya and Tatana may be due in part to contact between the two groups although the village of Padas Damit BT is very isolated from Tatana villages. Similarity of personal pronouns and grammatical structure are more likely to be important factors influencing the level of intelligibility.

NOTE

1. The 1960 census (Jones 1962) gives a figure of 2,308 Bisaya for Sarawak. Harrison (1964:177) gives an estimate of 7,000 Bisaya in Brunei. No more recent figures for Sabah than those quoted from the 1970 census are available at the time of writing.

THE TATANA LANGUAGE

Phyllis A. Dunn

0. INTRODUCTION

The Tatana-speaking people are living exclusively within the boundaries of the Kuala Penyu District of Sabah. There are approximately 5,500 Tatana, the majority of whom live in the north-eastern half of the district.¹ Tatana villages extend as far south as Menunggang KP, where the Tatana begin to mix with the Bisaya (→ Bisaya). The village of Menunggang KP itself is half Tatana and half Bisaya. Individual Tatana households extend nearly to Menumbuk KP, but these are scattered between Bisaya and Kedayan (→ Malayic). The village of Bundu KP is considered to be in the very heart of the Tatana area. There are also several Tatana villages along the coast of the northern peninsula, and a few are located across the narrow channel of the Kimanis Bay in the interior area extending toward Membakut BT (Figure 1).

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) classifies Tatana as a language within the Dusunic language family. The percentage of shared cognate (PSC) relations between Tatana and other languages within that family range from 55 to 73 PSC. The two Tatana wordlists referred to in Smith's article were taken in Bundu KP and Menunggang KP. Since Menunggang KP is half Tatana and half Bisaya, the relatively low relation of 83 PSC between the two villages reflects the influence of the Bisaya language on the Tatana language as it is spoken in Menunggang KP, rather than indicating different dialects of Tatana. The Tatana themselves recognise no dialect distinctions other than the influence of Bisaya on the Tatana speakers living in the border areas.

2. TESTING PROCEDURES

The purpose of the intelligibility testing was to verify the lexicostatistical classification of Tatana as a distinct language within the Dusunic family, and to see to what extent Tatana speakers could understand other languages of the same language family.

Two test points were selected for this intelligibility testing: Bundu KP, in the heart of the Tatana area, and Kiaru KP, where contact with other languages within the Dusunic family is minimal.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 251-256. *Pacific Linguistics*, C-78, 1984.

© Phyllis A. Dunn

251

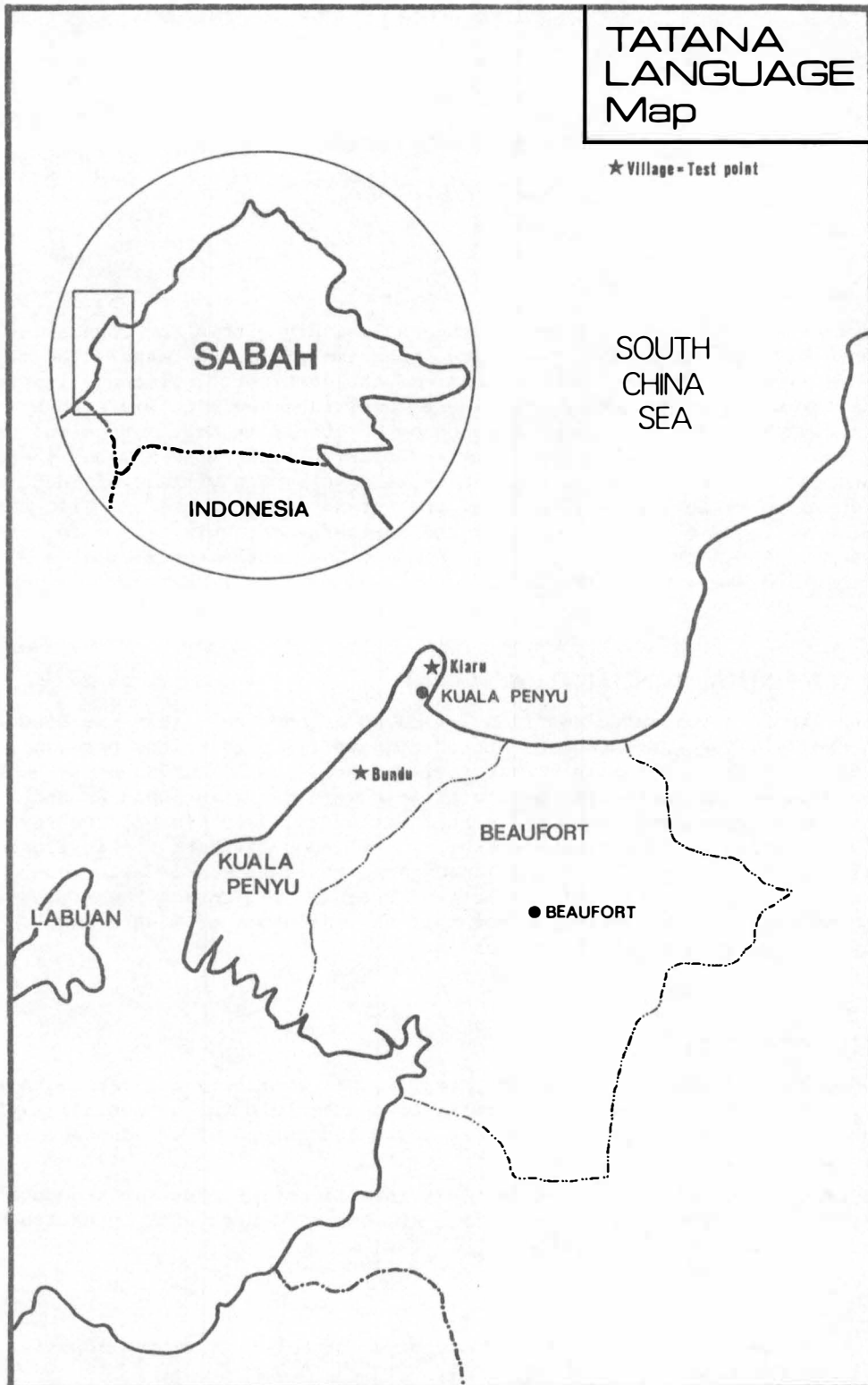


Figure 1: Tatana language map

Reference tapes from four villages were used for testing: 1) Taginambur KB, representing the Central Dusun dialect; 2) Bunduon PG, representing the Coastal Kadazan dialect of Kadazan/Dusun which is used in the mass media; 3) Padas Damit BT, representing the Bisaya language, the geographically closest major language within the Dusunic family; and 4) Takuli BT, the Kadazan/Dusun dialect having the highest PSC relation with the Tatana language. In addition, the Bahasa Malaysia test tape was used for testing in both of the Tatana villages.

All of the tapes used in the testing were considered to be clear with regard to technical quality. The story from Taginambur KB contained a few English and Bahasa Malaysia words, which caused slight problems for a few of the test subjects. Though the story from Takuli BT was about hunting and contained some familiar terminology, the outcome in the testing was not predictable on that basis. The stories from Bunduon PG and Padas Damit BT were also quite unpredictable and so provided a good test for intelligibility of the representative dialects and languages.

3. TEST RESULTS

3.1 Tatana understanding of Kadazan/Dusun

With the exception of the reference tape from Takuli BT, the Tatana test subjects found the Kadazan/Dusun stories difficult to understand. This can be seen from the testing results shown in Figure 2. Intelligibility test scores on the tapes from both Bunduon PG and Taginambur KB were well below the corresponding PSC figures. However with Klias River Kadazan as spoken in Takuli BT, which has the highest percentage of shared cognates with Tatana of any of the other languages or dialects, the level of understanding was well above the PSC.

REFERENCE TAPE		TEST POINT		
		Takuli BT (Klias River Kadazan)	Bunduon PG (Kadazan/Dusun)	Taginambur KB (Central Dusun)
TATANA	Bundu KP	92 (77)	67 (68)	59 (65)
	Kiaru KP	91	55	56

Figure 2: Tatana understanding of Kadazan/Dusun (shown as percentages).
(Corresponding PSC relations are in parentheses.)

3.2 Kadazan/Dusun understanding of Tatana

The Tatana story from Bundu KP was used as a reference tape in two closely related Kadazan/Dusun villages in the district of Tambunan (Figure 3). The villages of Toboh TA and Tangaban TA scored 42% and 41% respectively (→ Kadazan/Dusun).

TEST POINTS		REFERENCE TAPE	Bundu KP (Tatana)
KADAZAN/DUSUN	Toboh TA	42	(63)
	Tangaban TA	41	(66)
	Pukak TN	42	
	Kionsom KK	36	(64)
	Sunsuron TA	37	(67)
	KogoPON PR	58	(73)
BISAYA	Kilugus KP	86	(69)
	Padas Damit BT	85	(62)

Figure 3: Results of intelligibility testing (shown as percentages) using a Tatana story in other Dusunic villages.

Four less closely related Kadazan/Dusun villages were also tested with the reference tape from Bundu KP (Figure 3). Those four villages were Pukak TN, Kionsom KK, Sunsuron TA, and KogoPON PR. The scores from Pukak TN, Kionsom KK, and Sunsuron TA ranged from 36% to 42% (→ Kadazan/Dusun). In KogoPON PR, a higher 58% may reflect the relative geographical proximity of KogoPON PR to the Tatana language area as compared to the others which are geographically more remote from the Tatana language area. All of these villages, however, found the Tatana language very difficult to understand.

3.3 Tatana understanding of Bisaya

The intelligibility scores of the Tatana villages on the reference tape from Padas Damit BT were similar to those of the Bisaya villages when they were tested with a Tatana story (Figure 4). The people of Kiaru KP were able to

understand 82% of the story, while the score at Bundu KP was somewhat higher at 88%. The elevated score at Bundu KP may be due to the fact that Bundu KP is along the main road between the Bisaya area of the district and the town of Kuala Penyu, providing many opportunities for contact between the two groups.

REFERENCE TAPE		Padas Damit BT (Bisaya)
TATANA	Bundu KP	88 (62)
	Kiaru KP	82

Figure 4: Tatana understanding of Bisaya (shown as percentages).
(Corresponding PSC relations are in parentheses.)

3.4 Bisaya understanding of Tatana

The Tatana reference tape from Bundu KP was used for intelligibility testing in the Bisaya villages of Kilugus KP and Padas Damit BT. While the lexical relations of these villages were 69 and 62 PSC respectively with Bundu KP, the intelligibility scores were significantly higher at 86% and 85% (Figure 3). It is interesting to note that the Padas Damit BT score is only slightly lower than that of Kilugus KP, even though the people of Padas Damit BT rarely come in contact with Tatana people, while the people of Kilugus KP live within a few miles of Tatana villages and often meet Tatana speakers in the town of Kuala Penyu.

4. NATIONAL LANGUAGE INTELLIGIBILITY

The scores at the villages of Bundu KP and Kiaru KP on the national language intelligibility test were both 87% (Figure 5). Because of the presence of government schools in both of these villages, many of the test subjects had attended school. Only five of the total of 20 people tested scored below 90%. It is also interesting to note that all of those scoring under 90% were over the age of 35 and had had little or no opportunity for formal education. As can also be seen in Figure 5 however, others in the same age group and with no formal education scored 100% on the national language intelligibility test. This of course indicates a widespread knowledge and usage of the national language by Tatana speakers.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Bundu KP	87	6/4	39	23-66	9	6	6
Kiaru KP	87	6/4	38	17-60	7	8	6

Figure 5: Comprehension of the national language tape in two Tatana villages. (AVG. SCORE is given as a percentage; RANGE indicates the youngest and oldest test subjects; EDUCATION figures indicate: the number of subjects who had received some formal education (ED.), the average number of years of education which those subjects had received (A.E.R.), and the average number of years of education for all subjects in the corpus (A.E.S.).)

5. CONCLUSION

The intelligibility testing that was done at the villages of Bundu KP and Kiaru KP confirms the lexicostatistical classification of Smith that Tatana is a language distinct from the Kadazan/Dusun language and from the other languages within the Dusunic family with which it was tested. The test results of Tatana with the Kadazan/Dusun language within the Dusunic family showed an average of 59% with a range in scores of 55-67%. The one exception to this was the village of Takuli BT which averaged 92% (→ Section 3.1).

Tatana may, however, be more closely related to the Bisaya language than previously thought. The high degree of intelligibility between Bisaya and Tatana may be due in part to contact between the two groups. This however would not be the case for the village of Padas Damit BT which is quite isolated. It is more likely that similarity of personal pronouns and grammatical structures are more significant factors to explain the high level of intelligibility between the two.

The people themselves have their own explanations for their ability to understand each other's languages. A Tatana legend tells of a time when the Tatana and the Bisaya were very close friends. In addition to intermarrying at that time, both groups learned to speak the language of the other group. An older Bisaya man also explained that in ancient times the Bisaya and the Tatana were one people and so the Bisaya are still able to understand the Tatana language.

NOTE

1. In the Sabah census figures, the Tatana are grouped with the larger 'Kadazan' group. This population figure is therefore only an estimate based on interviews in Tatana villages and the author's knowledge of the area.

THE KUIJAU LANGUAGE

John A. Spitzack

0. INTRODUCTION

0.1 Location and origins

The language under discussion here is classified by Smith (in this volume) as Kuijau and is located in Keningau District on the plains between the Wittie Range to the east and the Crocker Range to the west. The Kuijau villages form a crescent around the south-west, north-west and north-east regions of Keningau KU, all within a 12-mile radius (Figure 1).

At least two of the individuals interviewed indicated that the Kuijau people originally inhabited the highlands nearby, presumably the Crocker Range. As late as the 1920s Rutter locates them "... on the western hills, above the Keningau Plains" (1922:66). Fifty years earlier, Crespigny (1872:176) reported a "Koijoe" group "on the Klias River, near Padass ...", which also may be a reference to the Kuijau.

They have since moved to the lowlands around Keningau KU, and inhabit some of the same areas as the Gana people and at times the same villages. The Keningau Murut (→ Murutic) live to the south and east of the Kuijau crescent. To the north-east are the Tambunan Dusun, many of whom have migrated to the Keningau region. To the north-west, over the Crocker Range, are found the Kadazan of the coastal regions of Beaufort, Papar and Penampang districts (→ Kadazan/Dusun.)¹

Because of the Kuijau position between the Dusunic groups and the Murutic groups it is generally held that they represent a "hybridization" of the two peoples (Rutter 1929; Chatfield 1972). However, Woolley (1953:1) cited one of his chief informants as denying such an admixture. Rather, he said, the Kuijau adhere to the same tradition as several other Dusun groups (Woolley 1953; Williams 1961; → Eastern Kadazan), that the legendary cradle of all Dusuns, including the Kuijau, was beside a deep pool on the upper Labuk River near what is now Tampias RU. The place was called Nunukaragang after the enormous nunuk tree which sheltered it.

Williams (1961:73) suggests that the "Koh/hub" group, which is said to have moved to the south from the area north of Tambunan TA about a century ago, may in fact, be the present Kuijau group or a progenitor of it. He cites a Kuijau tradition of migration claiming that they came from the north and first settled in Bingkor KU. He traces the etymology of the original autonym and the present Kuijau (Kwijau) ethnonym as follows: Koh/hub → Kur/re → Kwijau. But he gives no linguistic evidence to support this.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 257-270. *Pacific Linguistics*, C-78, 1984.

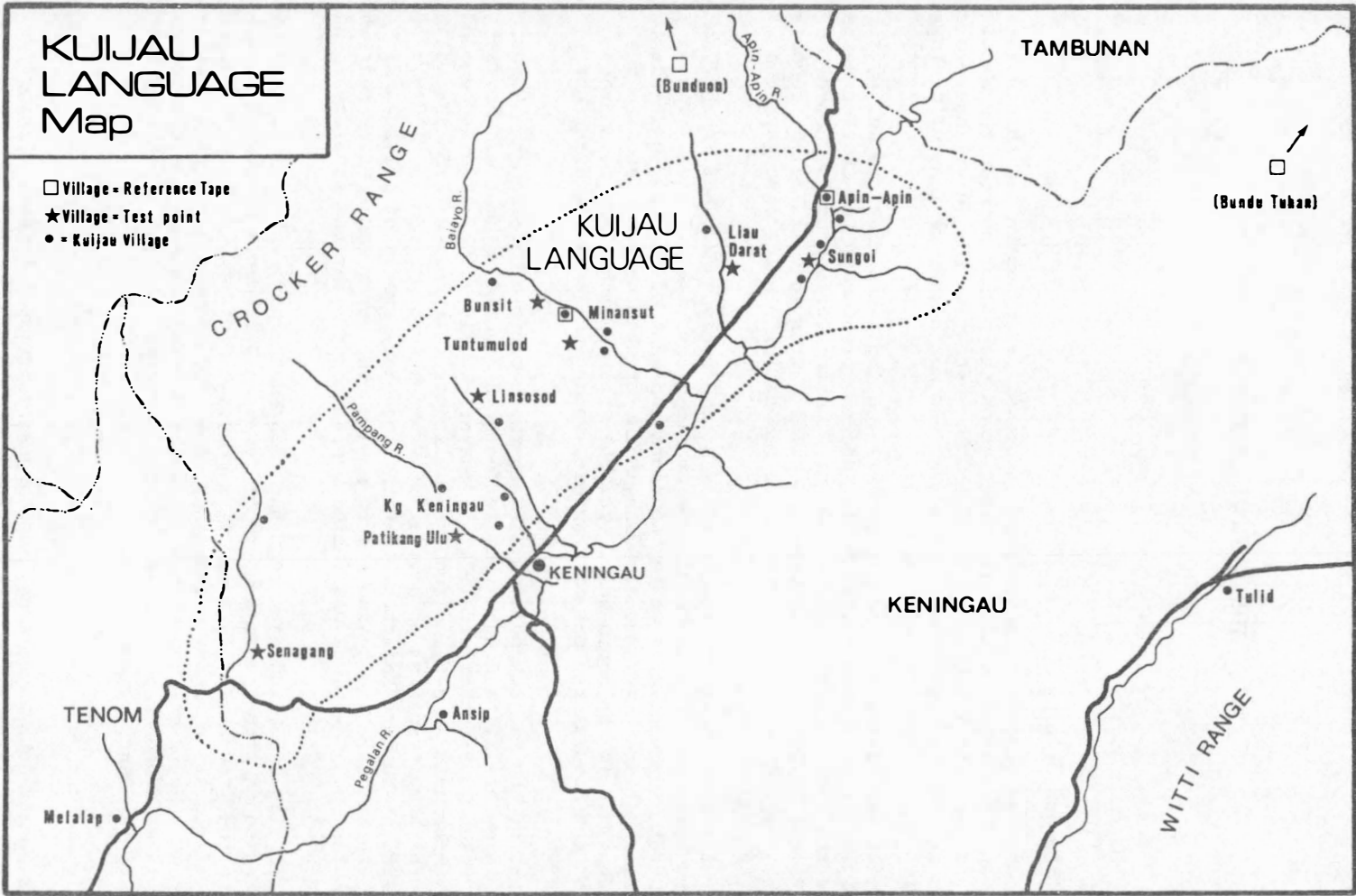


Figure 1: Kuijau language map

0.2 Ethnonyms

According to Rutter (1929:36) the ethnonym "Kuijau" (also Kwijau, Kijau, Kuiyow, Kuliow, Kuriyo and possibly Koijoe) is derived from the Dusun word riau which means *dried up*, because they resided in areas which were completely deforested. The word, as he traces it, is related to the Tambunan Dusun word Keriau, while the Papar Dusun who have the 'z' sound predominant in their dialect, have 'corrupted' it to Kuizau. The Dusun form Kuizau was then anglicised to Kuijau, which is sometimes reduced to Kujau.

The autonym sometimes used by the Kuijau is Menindal. They have also been called Tindal (as have other groups) which comes from the same root. Prentice (1970:371) records the term as Menindaq. The Kuijau are sometimes called Tidong by the lowland Muruts around them, which means *hill*, thus *hill people*. Many people simply refer to them as Hill Dusun, a very broad term used to describe any number of groups.

0.3 Population

The population of the Kuijau language group is estimated to be between 5000 and 6000.² This is difficult to assess though, because of the intermixture of Kuijau speakers with other groups in the same villages and throughout the region. The following list gives the Kuijau population of those villages which were tested and for which estimates have been gathered from the local headmen.

Tuntumulod KU	340
Sungoi KU	125
Liau Darat KU	390
Senagang KU	500
(including Keretan KU)	
Linsosod KU	140
Bunsit KU	300
Patikang Ulu KU	360
Total	2155

The following villages are said to be made up mostly of Kuijau speakers:

Agudon KU	Kepayan KU
Bandukan KU	Liau Ulu KU
Bingkor KU	Liawan KU
Bulusilau KU	Ranggun KU

Other villages which are said to be mixed include:³

Binaong KU	Pengas KU
Apin-Apin KU	Sandapak KU
Kg Keningau KU	

0.4 Literature

References to the Kuijau people go back at least a century, but as far as is known, no serious linguistic description has ever been attempted. As early as 1922 Rutter classified them as Murutic, but also as holding a unique position as "... a kind of buffer state between the west coast Dusuns and the Muruts of

the interior" (p.68). In his 1929 work he placed them midway between the two groups and called them "... a most interesting link between the Muruts and the Dusuns ..." (p.36). He also included a 186-item wordlist from the Kuijau language.

Much of the phonological and grammatical analysis of Dusun dialects/languages which has been completed would be relevant to a study of the Kuijau language (e.g. Grossens 1924; Staal 1926; Majusim 1949; Antonissen 1958; Lees 1964, n.d.; Clayre 1966, 1967, 1970 and Asmah 1978, 1979-80).

Some early lexicostatistical studies included wordlists from the Kuijau language, such as Wellings' (1960) initial Dusun dialect survey, Healey and Healey's (1961) methodological study based on the same data, and Foran's (1967) comparison of some Dusunic lists.

Prentice's (1970) comparison of 73 wordlists, mostly from Sabah languages, included Rutter's Kuijau wordlist mentioned above and his own wordlist of the "Menindaq" dialect (identified as Kuijau in this study). Prentice found Rutter's Kuijau wordlist to be sufficiently different to warrant a language-level distinction. He included Rutter's Kuijau with his Minansut language which has been identified as Gana in this study (→ Murutic). However different Prentice's classification is from that of this study, it was the first real attempt to classify a broad spectrum of the languages of Sabah by lexicostatistics.⁴

More recently Cohen (1981) produced a preliminary study of the linguistic unity and diversity in Keningau District in which he displayed a list of 80 words gathered from 25 villages, including the Kuijau villages discussed here.

1. LEXICOSTATISTICAL CLASSIFICATION

1.1 Smith's classification

Smith (in this volume) classified Kuijau as one of the languages of the Dusunic family on the basis of a relation with representative wordlists from the other Dusunic languages of 52-77 percent of shared cognates (PSC). However, he classified the Kuijau from Apin-Apin KU (hereafter called Apin-Apin Kuijau) as belonging to the Murutic family rather than the Dusunic. This was established on the basis of a relation of only 64-70 PSC with representative wordlists of the Dusunic languages and a relation of 54-77 PSC with representative wordlists of the Murutic languages.⁵

In view of the difficulty of connecting Apin-Apin Kuijau with the other Kuijau language either historically or linguistically, it was decided to handle it in a separate treatment (→ Murutic) and only mention it here as a point of comparison. Figure 2 shows the PSC relations among the five Kuijau villages, Tuntumulod KU, Sungoi KU, Liau Darat KU, Senagang KU and Linsosod KU, compared with their relations to Apin-Apin KU.

Besides separating Apin-Apin Kuijau from the Kuijau of the other villages, Smith also classified each of these five Kuijau samples as representing distinct dialects of one language. This was established on the basis of relations of 77-85 PSC between the wordlists. Since none of them shared cognate relationships of more than 85, which is the threshold for merging dialects, each of the villages were regarded as representing a separate dialect.

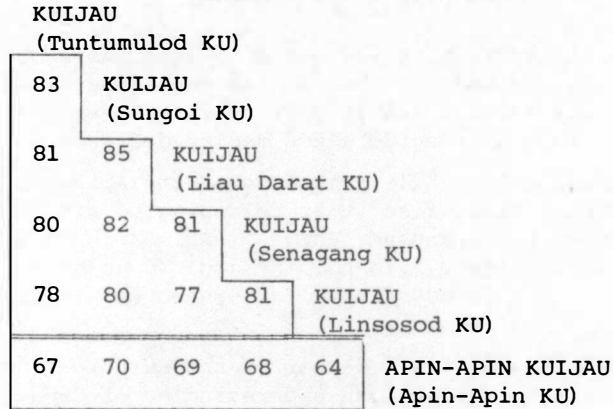


Figure 2: PSC relations of five Dusunic Kuijau samples and the Murutic Apin-Apin Kuijau sample

1.2 Revised classification

During the second phase of the survey the wordlists (except for Liau Darat KU) were checked and synonyms and obvious loan words noted. A manual comparison of these rechecked wordlists has somewhat revised the PSC relations of the Kuijau dialects and Apin-Apin Kuijau. These revised relations are marked with an asterisk throughout this paper.

The revised figures tend to raise the PSC relations of the five Kuijau dialects between 2 and 7 percentage points making it doubtful that they should be classified as distinct dialects. The PSC relations among the dialects now stand at *80-89 percent. Each of the samples is connected to the others by at least two relations above the 80-85 PSC dialect threshold.

The revised PSC relations of Kuijau with Apin-Apin Kuijau tend to be lower, making Smith's classification which placed them in separate language families even more convincing. Figure 3 shows the revised PSC relations.

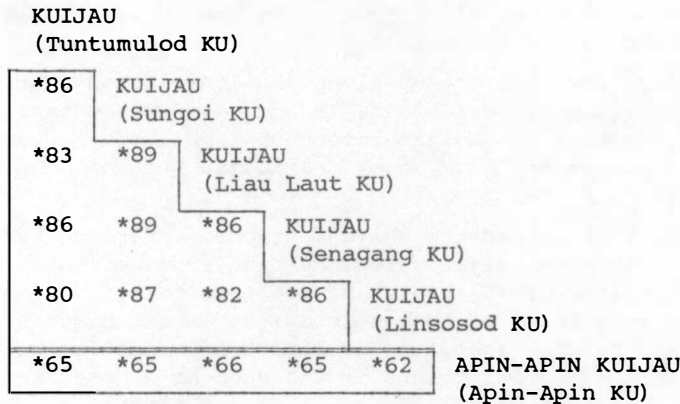


Figure 3: Revised PSC relations (marked with an asterisk) comparing five Kuijau samples and the Apin-Apin Kuijau sample

2. TESTING PROCEDURES

Since the intelligibility testing was set up on the basis of the PSC relations of the initial wordlists and not the revised relations, the first objective in the intelligibility testing was to clarify whether each of the five Kuijau villages represented a distinct dialect and to define its limits.

Several factors made intelligibility testing of the Apin-Apin Kuijau dialect advisable, even though Smith classified it as belonging to a different language family from Kuijau: it uses the ethnonym 'Kuijau'; Apin-Apin KU is located in close proximity to Kuijau villages; Apin-Apin Kuijau had no high PSC relations with other languages and there is no strong link connecting it with the Murutic language family.

Even though the Gana language is a member of the Murutic language family, it was considered desirable to test Kuijau understanding of Gana. The two groups are in close proximity and were reported to have a high level of mutual intelligibility.

It was also important to discover how well the Kuijau community could understand the Kadazan/Dusun language. This testing was done mainly to find out how far the influence of the Kadazan/Dusun language extended. In the case of the Kuala Monsok dialect there was a real possibility that a close linguistic affinity did exist. Therefore it proved desirable to do intelligibility testing in this dialect as well as the Central Dusun dialect.

In setting up the set of reference tapes to be used in the intelligibility testing of the Kuijau language, it was necessary to make choices between what should be tested and what was actually possible. Ideally, every Kuijau hometown tape should have been used as a reference tape at every other Kuijau test point. But since it had already been decided to limit the size of the test sets to no more than five stories besides the hometown and national language tape, it was necessary to limit the test sets to the Kuijau hometown tape and two of the best remaining Kuijau stories as reference tapes.

Since the Tuntumulod KU and Senagang KU tapes were considered the two best Kuijau tapes, they were chosen as the reference tapes representing the variety in Kuijau dialects. When, however, the Tuntumulod KU reference tape was used as the hometown tape, the Linsosod KU story was substituted as the second Kuijau reference tape. By using the same Kuijau stories there was some 'control' by which to evaluate the results and also it was possible to use the best stories and leave the poorer stories as hometown tapes only.

Each of the Kuijau tests was set up along the lines noted above with one exception. For the Senagang KU test a Gana language reference tape from Minansut KU was substituted for one of the Kuijau reference tapes. This was done on the basis of what several people had said about the Kuijau speakers and the Gana speakers.

Besides the hometown tape and the above mentioned reference tapes, each test set included an Apin-Apin Kuijau reference tape from Apin-Apin KU. It is unfortunate that the quality of the story and the tape itself was not better in light of the possible importance of Apin-Apin Kuijau in establishing a link between the Murutic and Dusunic language families. After using the tape at three of the above mentioned Kuijau test points it was used as a hometown tape in Apin-Apin KU itself. There it was discovered that the content on which the last two questions were based was so short that the questions were difficult if not inappropriate. Therefore, in order to safeguard the results of the test and not

call into question its validity, those questions were excluded when analysing results. In most cases this raised the average of the test scores.

The Kadazan/Dusun story from Bunduon PG was also used as a reference tape in each of the Kuijau test sets. Both the content of this story and the technical quality of the tape were very good.

The Kadazan/Dusun reference tape from the village of Bundu Tuhan RU was of good technical quality. The story, however, was very simple with frequent use of Malay terms and terms which are used broadly throughout Sabah languages. This was likely a significant factor in the generally high intelligibility scores.

3. TEST RESULTS

3.1 Mutual intelligibility within Kuijau

Figure 4 shows the results of the intelligibility testing conducted in five Kuijau villages, each representing separate dialects according to Smith's classification. All five test points understood their hometown tape well; three of them registered 100% intelligibility.

REFERENCE TAPES		KUIJAU				ANA	APIN-APIN KUIJAU	KADAZAN/DUSUN	
		Hometown	Tuntumulod KU	Senagang KU	Linsosod KU	Minansut KU	Apin-Apin KU	Bunduon PG (+Coastal Kadazan)	Bundu Tuhan RU (+Central Dusun)
TEST POINTS									
KUIJAU	Tuntumulod KU	100 (--)		96 (*86)	100 (*80)	(*66)	66 (*65)	65 (*75)	62 (*73)
	Sungoi KU	94 (--)	87 (*86)	93 (*89)			70 (*65)	70 (*77)	92 (*75)
	Liau Darat KU	100 (--)	96 (*83)	99 (*86)			70 (*66)	75 (74)	82 (74)
	Senagang KU	100 (--)	100 (*86)			82 (*68)	69 (*65)	70 (*77)	89 (*75)
	Linsosod KU	98 (--)	100 (*80)	99 (*86)			72 (*62)	72 (*71)	96 (*70)

Figure 4: Kuijau intelligibility of Kuijau, Gana, Apin-Apin Kuijau and Central Dusun reference tapes. (PSC relations are given in parentheses and revised PSC relations are marked with an asterisk; + indicates re-classified designations.)

Intelligibility between the Kuijau villages is high, ranging 87-100% compared to PSC relations of *80-89. Individual scores for all 50 subjects averaged 97% with a range of 80-100%. The 100% mutual intelligibility between Tuntumulod KU and Linsosod KU was unexpected since the PSC relation is only *80. The high intelligibility scores at all five test points is a clear demonstration of the homogeneity of the Kuijau language in spite of the lexical differences. The Kuijau villages discussed may be considered members of one homogeneous group.

3.2 Mutual intelligibility of Kuijau and Gana

Kuijau intelligibility of the Gana language was tested in Senagang KU using a reference tape from Minansut KU (Figure 4). The ten subjects averaged 82% intelligibility with individual scores ranging from 60% to 100%. This compares to the significantly lower *68 PSC.

Gana intelligibility of Kuijau was also tested in Kuangoh KU using the Tuntumulod KU reference tape (→ Murutic). The Gana wordlist collected there has a relation of *87 PSC with the other Gana village of Minansut KU and a relation of *66 PSC with the Tuntumulod KU wordlist. The Gana subjects averaged 99% intelligibility of the Kuijau reference tape. While language learning rather than linguistic affinity would at first appear to be the most logical explanation for the high intelligibility, this is not easy to verify in light of the relative isolation of this particular Gana village from Kuijau villages and the presence of only two Kuijau speakers in a village of over 350. It could be that high mutual intelligibility arises from certain important grammatical features of Gana which follow the Dusunic rather than Murutic pattern.

3.3 Mutual intelligibility of Kuijau and Apin-Apin Kuijau

An Apin-Apin Kuijau reference tape from Apin-Apin KU was included in each of the Kuijau test sets (Figure 4). The village averages ranged 66-72% intelligibility compared to *62-66 PSC. The highest average intelligibility (72%) was recorded at Linsosod KU, which had the lowest PSC of the five Kuijau villages. The lowest intelligibility (66%) was recorded at Tuntumulod KU compared to *65 PSC. The 50 Kuijau test subjects averaged 69% intelligibility.

These scores are very much in line with the revised PSC figures and probably reflect an accurate appraisal of Kuijau intelligibility of the Apin-Apin Kuijau language.

The normal reaction of the subjects was one of bewilderment when they first heard the story. Some of the subjects recognised the Apin-Apin Kuijau reference tape as Murutic while others said it was "like" Dusun.

To complete the mutual intelligibility testing of Kuijau and Apin-Apin Kuijau a reference tape from Tuntumulod KU was used at Apin-Apin KU (→ Murutic). The five male and three female subjects averaged 94% intelligibility with individual scores ranging 80-100%. This is much higher than the *65 PSC and probably indicates considerable language learning by a very local language community of 200 (Apin-Apin Kuijau) of the more dominant Kuijau language spoken in and around the same village.

3.4 Mutual intelligibility of Kuijau and Kadazan/Dusun

Kuijau intelligibility of the Kadazan/Dusun language was tested using several Central Dusun dialect (per Smith, in this volume) reference tapes. Intelligibility testing has since helped to clarify the divisions within the Dusunic family (→ Kadazan/Dusun). According to reclassification based on intelligibility testing, the Kuijau test points discussed thus far were actually tested with a Coastal Kadazan reference tape from Bunduon PG and a Central Dusun reference tape from Bundu Tuhan RU. In addition, two other Kuijau villages, Bunsit KU and Patikang Ulu KU, were tested for intelligibility of a different Central Dusun reference tape from Taginambur KB.

3.4.1 Mutual intelligibility of Kuijau and Coastal Kadazan

Intelligibility of the five Kuijau test points of the Coastal Kadazan reference tape from Bunduon PG showed an average of 70%. Village averages ranged from 65-75% compared to relations of *71-77 PSC (Figure 4). These low scores demonstrate that Kuijau and Coastal Kadazan are separate languages.

A test made at Kogopon PR using a Tuntumulod KU reference tape is the only sample of Coastal Kadazan understanding of Kuijau. The intelligibility score here was 75% compared to a *78 PSC relation. This intelligibility score and the 70% average score of Kuijau understanding of Coastal Kadazan demonstrates both a low mutual intelligibility and supports Smith's separate language classification.

3.4.2 Mutual intelligibility of Kuijau and Central Dusun

As seen in Figure 4, the intelligibility results for the Bundu Tuhan RU reference tape representing Central Dusun were substantially higher than the results for the Coastal Kadazan reference tape, averaging 84% for the five test points and ranging 62-96%. Individual scores ranged 0-100% with one subject showing zero percent understanding and 27 showing perfect intelligibility. By comparison the PSC relations ranged *70-75.

Some explanation is needed for the disparity between the very high scores of the test points Sungoi KU, Liau Darat KU, Senagang KU and Linsosod KU, which averaged 90% intelligibility and the very low score of Tuntumulod KU, which showed only 62% intelligibility.

In order to investigate the validity of the high scores two other Kuijau villages, Bunsit KU and Patikang Ulu KU, were chosen, and a cross-section of people in each was tested for their understanding and comprehension of the Bundu Tuhan RU story. In addition, a reference tape from Taginambur KB, another Central Dusun village, was also tested. Figure 5 shows the results of this testing.

The representative Kuijau village Sungoi KU has a relation of *75 PSC with Bundu Tuhan RU and *74 PSC with the sample from Taginambur KB. Smith has classified both of these as "core" Central Dusun. The results of testing with the Bundu Tuhan RU tape at Bunsit KU and Patikang Ulu KU was 100% intelligibility at both villages, even higher than the other Kuijau villages previously tested. Section 2. discusses the problem of using the very simple story from Bundu Tuhan RU as a reference tape. Language learning may be a factor in Kuijau

comprehension of Central Dusun, but it is more likely that the very high intelligibility is due to the simple test story.

REFERENCE TAPES		Central Dusun	
		Bundu Tuhan RU	Taginambur KB
TEST POINTS			
KUIJAU	Bunsit KU	100 (*74)	77 (*75)
	Patikang Ulu KU	100 (*74)	75 (*75)

Figure 5: Two Kuijau villages' intelligibility of two Central Dusun reference tapes. (Revised PSC relations are marked with an asterisk and given in parentheses.)

By contrast, the Bunsit KU subjects averaged only 77% intelligibility of the Taginambur KB tape and the subjects in Patikang Ulu KU averaged only 75% intelligibility of the same tape. The intelligibility score of Bunsit KU and Patikang Ulu KU of the Taginambur KB reference tape was very similar to the PSC score. This fact along with the fact that the Taginambur KB story was observably more difficult than the Bundu Tuhan RU story leads to the conclusion that the Taginambur test results more accurately reflect Kuijau intelligibility of Central Dusun.

That Tuntumulod KU scored a very low 62% average intelligibility of the Bundu Tuhan RU story, cannot be explained from the limited data and warrants further testing and investigation.

To investigate intelligibility in the opposite direction, Central Dusun intelligibility of Kuijau was tested in the villages of Pahu TA, Paginatan RU,⁶ Tangaban TA, Sunsuron TA, Toboh TA, Parancangan RU, Kundasang RU, Kionsom KK and Pukak TN (→ Kadazan/Dusun).

The Liau Darat KU reference tape represented the Kuijau language in the Pahu TA and Paginatan RU test sets, with the results of 86% and 84% respectively compared to PSC relations of 76 and 77. The content of the Liau Darat KU story was considered simple and at both test points male subjects heavily outnumbered female subjects. These results, therefore, are not considered an accurate reflection of the Central Dusun community's intelligibility of Kuijau.

The same number of male and female subjects, (29), made a very balanced corpus from the remaining seven Central Dusun test points. These were all tested for intelligibility of Kuijau using the very good reference tape from Tuntumulod KU. The average intelligibility score for all 58 subjects was 65%.

Given the average PSC of *76, which represents a spread of *74-78 PSC, the 65% intelligibility can be said to be a fair representation of the intelligibility of Kuijau by the Central Dusun community as a whole.

In summary, village averages of Kuijau comprehension of Central Dusun ranges 62-96%. The average of individual scores for all Kuijau subjects tested is 82%. The village averages of Central Dusun comprehension of Kuijau ranges 57-86%, while the average of individual scores for all Central Dusun subjects is 72%. Clearly these figures demonstrate that these are separate languages, though language learning may be evident in a number of situations.

3.4.3 Intelligibility of Kuijau by Kuala Monsok

The wordlist collected at Kuala Monsok TA shows the highest relation with Kuijau of all Dusunic lists, ranging *80-86 PSC (→ Kadazan/Dusun). It is also one of the geographically closest villages to the Kuijau-speaking areas. For these reasons, a Kuijau reference tape from Liau Darat KU was included in the Kuala Monsok TA test set. However, the 62% intelligibility registered there by the corpus of four male and three female subjects demonstrates very low intelligibility in spite of the very simple story and high PSC relation. No testing was conducted in the other direction.

4. NATIONAL LANGUAGE INTELLIGIBILITY

The final test tape in each of the Kuijau test sets was the national language tape in Bahasa Malaysia. This story tape is considered very good in both content and technical quality. Figure 6 summarises the sociological data for the corpus at each test point along with the average percentage of intelligibility of the national language tape for all ten subjects in each village.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Tuntumulod KU	77	6/4	34	17-58	6	8	5
Sungoi KU	88	4/6	28	16-50	9	9	8
Liau Darat KU	80	3/7	31	17-68	6	6	4
Senagang KU	84	5/5	35	16-65	7	6	5
Linsosod KU	85	7/3	32	15-67	4	8	3

Figure 6: Kuijau intelligibility of the national language tape with sociological data. (Scores are given as percentages. The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average years of education which those subjects had received (A.E.R.), and the average years of education per subject in the test corpus as a whole (A.E.S.).)

Intelligibility ranged from 77% in Tuntumulod KU (which also scored lowest on all other non-Kuijau tests) to 88% in Sungoi KU. The latter corpus also had the highest level of education.

In every test case, the group of subjects which had received some formal education averaged higher intelligibility than the group which had received no formal training using the medium of Bahasa Malaysia. The amount of education did not prove significant. Those having received some education averaged 94% intelligibility against only 62% intelligibility shown by those with no formal training. Figure 7 displays these relationships.

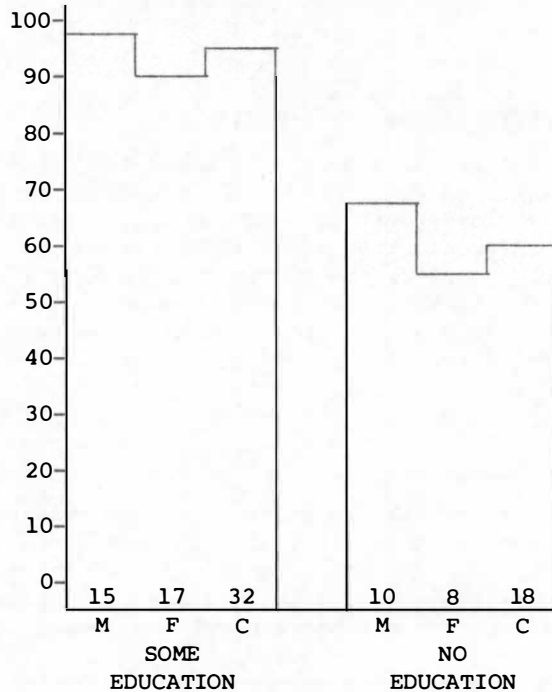


Figure 7: Kuijau intelligibility of the national language tape displayed by level of education. (The three levels in each block represent male (M), female (F) and combined (C) corpus averages. The number within each column represents the size of the corpus used to calculate the intelligibility average of each.)

In Figure 8 the average intelligibility of the national language tape is displayed by the parameter of age. The four age groups which were distinguished (15-30, 31-45, 46-55 and 56-up) show a consistent descending level of intelligibility within both the male and female corpora. This could be due as much to education, though, since only four male and two female subjects of the educated corpus of 15 male and 17 female subjects are found outside the first age group, and they are all included in the second age bracket. Age may show a minimum of influence but it is probably safe to say that the continued spread and use of the national language in school, government and trade will gradually work its way through all age levels.

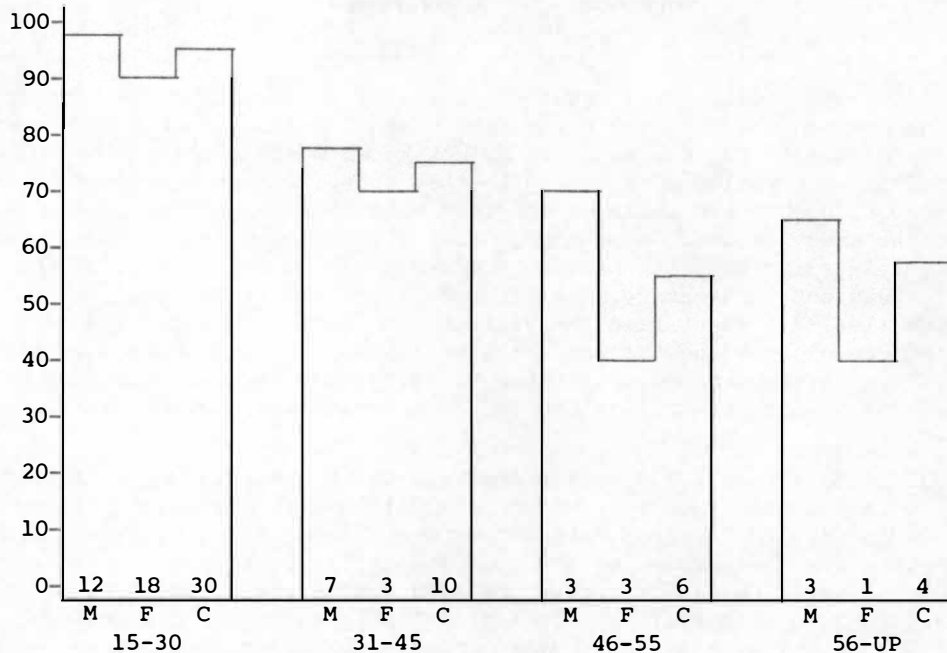


Figure 8: Kuijau intelligibility of the national language tape displayed according to age parameters. (The three levels in each block represent male (M), female (F) and combined (C) corpus averages. The number within each column represents the size of the corpus used to calculate that intelligibility average.)

Since all villages tested are located in the same vicinity and variables such as neighbouring language groups and proximity to the district centre are much the same, the only factor that can positively be shown to be of significance as far as skill in using the national language is concerned is education through the medium of Bahasa Malaysia.

5. CONCLUSION

The results of intelligibility testing as well as the revised lexical relations within the Kuijau language show that all samples discussed above belong to one homogeneous group. The differences which exist in the lexicons of each village, while still apparent in the revised wordlists, are not sufficient to impede mutual intelligibility among these Kuijau villages.

Further intelligibility testing is not needed to determine the status of Kuijau in relation to the other Dusunic languages since the intelligibility testing tends to confirm Smith. However, it would be helpful to understand more clearly the true relationship between Kuijau and the Murutic language of Gana since intelligibility scores indicate good comprehension in both directions. That study will no doubt have to include an investigation of the ethnic origins and relations of each.

NOTES

1. Rutter (1929:38) followed by Prentice (1970:382), identified the Tagaas, who live between the Tambunan Dusun and Coastal Kadazan, as a group very closely related to the Kuijau. It is not known whether any of the collection points were villages belonging to this group. None identified themselves as such, but the Kuala Monsok TA sample (→ Kadazan/Dusun), though a bit to the east, is located in the general vicinity and is the sample most closely related to the SIL Kuijau wordlists (76-84 PSC). This close relation was not born out by intelligibility testing (→ Section 3.4.3). Williams (1961:73) identified the villages of Tobou (Toboh TA) and Kintutul as being occupied by the "Tah/gas" (Tagaas) group. Although no intelligibility testing was conducted at those villages, the lexical relations of only 70-76 PSC with the Kuijau samples do not indicate a close relationship either.
2. The 1931 census lists 4,539 Kuijau speakers while the 1970 census listed only 120 individuals (Sullivan and Regis 1981). This discrepancy is not due to a drastic fall in population but to a change in the Kuijaus' self-designation for the purpose of the census. It is probable that most Kuijaus called themselves "Kadazan" during the 1970 census which listed 15,271 Kadazan in Keningau District. It is also likely that many Kuijau were classed as Muruts as a similar Kuijau population figure decrease was seen between the 1921 and 1931 censuses (Ride 1934; Jones 1966).
3. In addition to these villages, Woolley (1953) mentioned the fact that there were several settlements in the Tambunan plain area and the village of Minansut KU (cf. Prentice 1970).
4. Rutter's (1929) Kuijau wordlist had 102 comparable terms (according to the Malay) with SIL's 367-item wordlists. Of these 102 terms, PSC relations ranged from *81-86 percent, which is within the *80-89 PSC range SIL's samples showed, indicating that his list belongs to the same group. His Kuijau list also had a range of *79-80 with SIL's Gana lists. This is a much closer relation than SIL's Kuijau lists had with the Gana lists (*62-68 PSC), but having used only a subset of the words may have skewed the PSC figures. * denotes, as in the body of the paper, that the revised wordlists were used.
5. Having stated this, though, Smith then surmised that the actual historical roots of Apin-Apin Kuijau are Dusunic. This will have to be supported by more extensive investigation into Apin-Apin Kuijau speakers' migration into the area, a more thorough historical comparison of shared vocabulary and an examination of the grammars of each. When the Apin-Apin KU test subjects were asked what language they spoke, they gave names as diverse as Dusun, Kuijau, Sinimbitan and Murut.
6. This was lexically classified by Smith as belonging to the Paginatan Dusun dialect of the Kadazan/Dusun language but included with the Central Dusun dialect after intelligibility testing showed no significant difference.

THE EASTERN KADAZAN LANGUAGE

John A. Spitzack

0. INTRODUCTION

The language designated Eastern Kadazan is spoken in the Labuk-Sugut, Sandakan and Kinabatangan districts. Villages are located along the Labuk/Liwagu River and tributaries from the east coast of Sabah as far west as Tampias RU and along the Kinabatangan/Milian River system from Bilit KN to its headwaters, in the region of the Tongod River (Figure 1).

The actual historical origins of this language group and their migration are unknown. The Tompulung group from the area of Telupid LS share with several other Dusunic groups the legend that their beginnings are under a large nunuk ragang, or *red fig tree*, which was located near what is now Tampias RU on the upper Labuk River (Woolley 1953; Williams 1961; → Kuijau).

At some point in their history those Eastern Kadazan groups closest to the coastal areas came in contact with Islamic groups, possibly the Bajau or Ida'an who had not long before converted to Islam, and embraced their religion. At that time they began calling themselves orang sungai, or *river people*. Those groups further inland have retained their own particular autonym or use the more generic term "Kadazan" unless they embrace the Islamic religion, at which time they, too, change their name to Orang Sungai.

The problem with nomenclature is complicated somewhat more by the fact that those Paitanic language groups on the Paitan and Sugut Rivers in the north-eastern part of the state and those on the lower Kinabatangan River also call themselves Orang Sungai after converting to Islam. In addition, one group of Orang Sungai living in Kinabatangan and Sandakan districts is linguistically more closely related to the Ida'an language than to any other (→ Paitanic; → Ida'an).

The present population for speakers of the Eastern Kadazan language can only be roughly estimated since no current official population figures are available. The very broad ethnic classifications used for the 1970 census were not used for the 1980 census. With information taken from several sources the population for Eastern Kadazan speakers, including all dialects, is estimated to be 14,000 to 16,000.¹

Some of the Eastern Kadazan dialects are known by various ethnonyms and clan names. The Malapi Kadazan use the autonym "Sogilitan". The autonym "Tompulung" is used by people living on the upper Labuk River near Telupid LS.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 271-282. *Pacific Linguistics*, C-78, 1984.

© John A. Spitzack

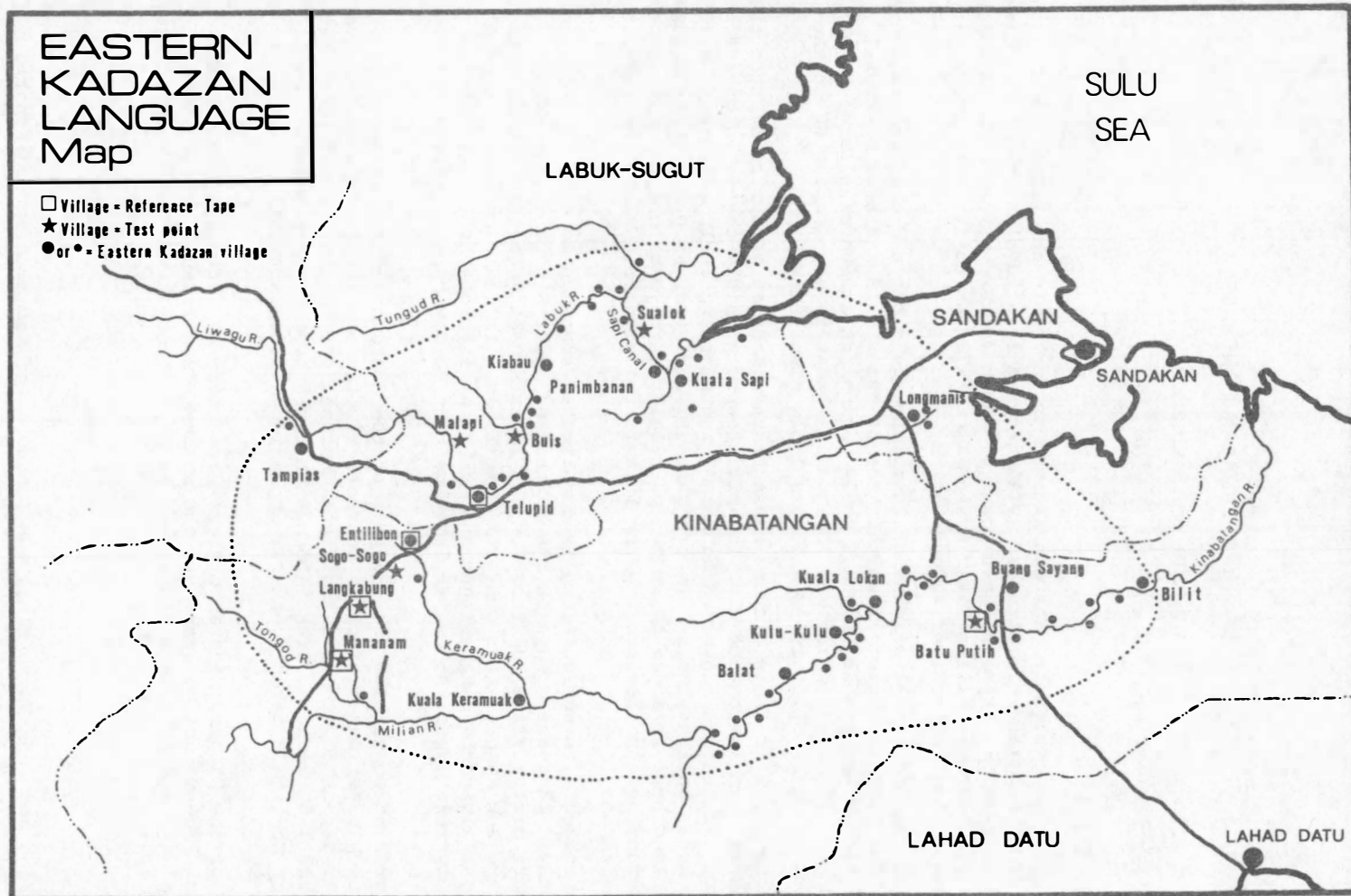


Figure 1: Eastern Kadazan language map

The Terusan Sapi Kadazan living along the Sapi Canal between Sapi LS and the lower Labuk River have called their language "Tindakon" but it is not known how broadly this is applied. Sometimes the term "Labuk Kadazan", or simply "Kadazan", is used for the above groups.

The "Sukang" and "Mangkaak" are found along the Keramuak River in Kinabatangan District. The "Mangkaak" are also found along the Tongod and Malagatan Rivers. Most of the other groups, located on the Segaliud River in Sandakan District and along the Kinabatangan/Milian River from Bilit KN to Kuala Keramuak KN generally refer to themselves as "Orang Sungai". Those in Balat KN and Bukit Permata KN sometimes call themselves "Tindakon" (or Tindakon Sungai) as do those on the Sapi Canal.

The subdialect of Eastern Kadazan spoken along the Labuk River has been under study for the past couple of years by Hope M. Hurlbut of the Summer Institute of Linguistics. To date, a paper on some features of narrative discourse (Hurlbut 1979) and a study of the morphophonemics (Hurlbut 1981) have been published. In preparation is an in-depth study of the verbal affixation of "Labuk Kadazan". In addition, a manuscript copy of a 5000-item English-Malay-Kadazan and Kadazan-Malay-English dictionary and a trilingual thesaurus have been compiled, and over 80 pages of Kadazan story text has been prepared and translated into Malay and English.

Wellings (1960) included data from this dialect in his comparison and survey of 13 Dusunic dialects. Healey and Healey (1961) used the data from Wellings for a methodological study of dialect comparison and suggested some possible classifications. Clayre (1966) also included "Labuk Kadazan" and "Mangkaak" data in her dialect comparison of some Dusunic dialects.

A booklet of five Kadazan folk tales entitled *Cerita dongeng Sabah* (Matius 1979) has been published in diglot form (Kadazan/Malay with an English translation in the back) by the Sabah Museum. Also, a "Labuk Kadazan" translation of the New Testament books of Mark (1976) and Acts (1977) has been published.

1. LEXICOSTATISTICAL CLASSIFICATION

Smith (in this volume) grouped 18 wordlists using the various ethnonyms "Mangkaak" (also Mangkak, Mangkahak and Mangkok), "Sukang", "Kadazan" and "Sungai" (also Sungei) together under the single designation Eastern Kadazan language. These lists have relations of 75-99 PSC. Using his representative Dusunic villages, those most closely related to Eastern Kadazan (71-81 PSC) are members of the Kadazan/Dusun language. In addition, the Rungus, Kadazan-Tambanua and Kuiu samples related at 77-78 PSC.

Nine Eastern Kadazan villages were grouped together to make up the Labuk/Mangkaak/Sukang dialect ranging 83-99 PSC. Mananam KN was chosen to represent this group since it shares the highest relation with all the other samples.

Four other wordlists having relations of 84-91 PSC were grouped together as a single Lamag Sungai dialect. Buang Sayang KN represented this dialect.

The wordlists from Longmanis SN, Malapi LS, Kulu-Kulu KN, Kuala Keramuak KN and Sualok LS each represented a distinct dialect of the Eastern Kadazan language. Some of these dialects are closely related to each other, for example Kulu-Kulu KN and Kuala Keramuak KN with 86 PSC; while others have a close relationship with one of the larger dialects, for example, Sualok LS, representing Terusan

Sapi Kadazan, has a range 79-89 PSC with the Labuk/Mangkaak/Sukang dialect. These relations illustrate the interrelatedness of the dialects within this rather broad language group and the tentativeness of the divisions. Figure 2 shows the PSC relations of the Eastern Kadazan samples.

2. TESTING PROCEDURES

The chief reason for dialect intelligibility testing within Eastern Kadazan was to clarify the relations between the seven dialects posited by Smith. This necessitated testing intelligibility across dialect boundaries. In addition to this, some intelligibility testing within the larger dialects was desirable to ensure that each dialect is represented by only mutually intelligible samples.

Though the major purpose for intelligibility testing within Eastern Kadazan was not to further define its relation to the other Dusunic languages, some testing was deemed advisable with one of the more closely related groups. The Kadazan/Dusun language, as represented by the Central Dusun dialect, was used for this purpose.

Intelligibility testing within the Eastern Kadazan language was conducted in six villages using seven different corpora of ten subjects each. The test points were Batu Putih KN representing the Lamag Sungai dialect, Malapi LS representing the Malapi Kadazan dialect, and Buis LS, Mananam KN, Langkabung KN and Sogo-Sogo KN representing the Labuk/Mangkaak/Sukang dialect.²

In all of the test sets a hometown tape was used to teach procedure and to gauge the subjects' understanding of the question/answer method. All test sets also included a national language tape to check intelligibility of Bahasa Malaysia. The hometown tapes were all fair to good in content and quality and the national language tape was considered good.

Besides the hometown tape and the national language tape, at least one other Eastern Kadazan reference tape was used from either Labuk/Mangkaak/Sukang villages of Mananam KN, Lankabung KN or Telupid LS or the Lamag Sungai village of Batu Putih KN. The Mananam KN, Telupid LS and the Langkabung KN reference tapes were all considered good. The Batu Putih KN tape was simple but satisfactory.

One or two Central Dusun dialect reference tapes were also used from either Bunduon PG, Taginambur KB, Pahu TA or Entilibon KN. The Bunduon PG tape has since been reclassified as representing the language designated Coastal Kadazan (→ Kadazan/Dusun). That tape was considered very good. The Entilibon KN tape was good while the other two had difficulties but were still rated as good.

Except in the case of Batu Putih KN and Mananam KN 'B' test points, all of the corpora are regarded as a good cross-section of each community. These two test points used a higher ratio of male to female subjects and the Mananam KN corpus was unusually young.

3. TEST RESULTS

3.1 Results of intelligibility testing

Figure 3 displays the results of intelligibility testing conducted in the Eastern Kadazan language group. Mutual intelligibility testing was done

	SEGALIUD SUNGAI (Longmanis SN)																		
"Kadazan"	77	MALAPI KADAZAN (Malapi LS)																	
"Kadazan"	80	87	LABUK/MANGKAAK/SUKANG (Telupid LS)																
"Kadazan"	78	85	91	(Kuala Sapi LS)															
"Kadazan"	75	84	87	90	(Kiabau LS)														
"Kadazan"	77	83	86	89	89	(Panimbanan LS)													
"Kadazan"	79	83	85	88	90	87	(Buis LS)												
"Mangkaak"	85	83	91	89	85	87	86	(Langkabung KN)											
"Mangkaak"	85	86	92	91	87	90	88	99	(Mananam KN)										
"Sukang"	84	83	88	87	86	88	86	95	96	(Entilibon KN)									
"Sukang"	83	80	83	83	84	84	83	88	89	90	(Kuala Keramuak KN)								
"Sungai"	78	82	83	83	83	81	82	84	87	86	85	KULU-KULU SUNGAI (Kulu-Kulu KN)							
"Sungai"	77	75	79	78	79	80	79	81	84	84	87	86	KERAMUAK SUNGAI (Kuala Keramuak KN)						
"Sungai"	80	73	75	75	77	78	77	76	79	79	82	83	84	LAMAG SUNGAI (Batu Putih KN)					
"Sungai"	76	73	74	74	77	78	77	77	80	81	81	85	85	86	(Buang Sayang KN)				
"Sungai"	74	73	75	76	76	79	77	77	80	80	81	84	84	86	91	(Kuala Lokan KN)			
"Tindakan"	73	75	78	78	79	84	78	80	83	82	83	86	85	84	88	90	(Balat KN)		
"Tindakan"	77	79	82	85	87	89	88	81	83	81	79	81	78	80	80	81	81	TERUSAN SAPI KADAZAN (Sualok LS)	

Figure 2: PSC relations of seven dialects of the Eastern Kadazan language. (The dialect designation is printed in capital letters and the village self-designation is given in quotation marks. (Adapted from Smith, in this volume.))

between the Lamag Sungai dialect and the Labuk/Mangkaak/Sukang dialect. Testing done at the Batu Putih KN test point, representing Lamag Sungai, showed only 69% intelligibility of a Labuk/Mangkaak/Sukang reference tape from Mananam KN. That compares to a PSC relation of 79. This is far below the 80% intelligibility threshold expected of mutually intelligible dialects.

TEST POINTS		REFERENCE TAPES	EASTERN KADAZAN				KADAZAN/DUSUN			
			LABUK/MANGKAAK/-SUKANG			LAMAG SUNGAI	COASTAL KADAZAN	CENTRAL DUSUN		
			"Mangkaak" (Mananam KN)	"Mangkaak" (Langkabung KN)	"Kadazan" (Telupid LS)	"Sungai" (Batu Putih KN)	"Kadazan" (Bunduon PG)	"Dusun" (Taginambur KB)	"Dusun" (Pahu TA)	"Dusun" (Entilibon KN)
EASTERN KADAZAN	LAMAG SUNGAI	"Sungai" (Batu Putih KN)	69 (79)			--	44 (61)		49 (60)	
	MALAPI KADAZAN	"Kadazan" (Malapi LS)	82 (86)			72 (73)	37 (63)	35 (70)		
	LABUK/MANGKAAK/SUKANG	*"Kadazan" (Buis LS)	80 (88)			87 (77)	41 (70)	34 (70)		
		"Mangkaak" (Mananam KN)	--	+100 (99)	+98 (92)	74 (79)	46 (71)		52 (72)	+58 (73)
		"Mangkaak" (Langkabung KN)		--	81 (80)					41 (72)
		"Mangkaak" (Sogo-Sogo KN)		100	85					42

Figure 3: Eastern Kadazan intelligibility of four Eastern Kadazan reference tapes and four Kadazan/Dusun reference tapes. (+ Marks data from Mananam 'A'; Data from Mananam 'B' is unmarked; PSC relations are given in parentheses; autonyms are given in quotation marks; *"Kadazan" is also referred to as "Labuk Kadazan" in the paper.)

At two Labuk/Mangkaak/Sukang test points, Buis LS and Mananam KN, the Lamag Sungai reference tape from Batu Putih KN was used. They showed 87% and 74% intelligibility respectively compared to relations of 77 and 79 PSC. Given the marginal PSC relations this range in intelligibility is not unexpected. But the high intelligibility score at Buis LS of Batu Putih KN was a little surprising since there is no obvious contact between the two groups. In spite of the higher Buis LS intelligibility score, Labuk/Mangkaak/Sukang and Lamag Sungai are not mutually intelligible.

No cross-testing was conducted for the Malapi Kadazan dialect but the Malapi LS test point, representing that dialect, was tested for intelligibility of both Lamag Sungai, using the Batu Putih KN reference tape, and Labuk/Mangkaak/Sukang, using the Mananam KN reference tape. The Malapi LS subjects showed 72% intelligibility of the Lamag Sungai tape which is very similar to the 73 PSC relation. There is no reason to expect high intelligibility of a Malapi Kadazan reference tape at a Lamag Sungai test point.

On the other hand, the Malapi LS subjects showed 82% intelligibility of the Labuk/Mangkaak/Sukang reference tape from Mananam KN compared to a higher 86 PSC score. It is not possible to predict the Labuk/Mangkaak/Sukang intelligibility of a Malapi Kadazan reference tape which could conceivably be higher or lower. However, it is expected that the "Labuk Kadazan" subdialect of the Labuk/-Mangkaak/Sukang dialect would show slightly higher intelligibility of Malapi Kadazan, since they are both located along the same river. In fact, it has been said that the Malapi LS people have already adopted the "Labuk Kadazan" dialect since having moved closer to the East-West Highway.

The "Labuk Kadazan" and "Mangkaak" test points showed marginal to high intelligibility of the reference tapes from within the Labuk/Mangkaak/Sukang dialect. The intelligibility ranged from 80-100% compared to equally high PSC relations of 80-99. No "Sukang" tape or test point was used.

The high PSC relation of 88 between Buis LS and Mananam KN is not reflected in intelligibility. Buis LS subjects registered only a marginal 80% intelligibility of that "Mangkaak" story, possibly indicating some difficulty in comprehension between the Labuk River villages and those in the headwaters region of the Kinabatangan.

In the "Mangkaak" village of Langkabung KN, subjects showed marginal intelligibility of the "Labuk Kadazan" reference tape from Telupid LS, registering only 81% intelligibility, which is quite in line with the marginal 80 PSC. Two other "Mangkaak" villages, Mananam KN and Sogo-Sogo KN, scored 98% and 85% intelligibility respectively on the same "Labuk Kadazan" story. The high Mananam KN score was expected since it has a relation of 92 PSC with the Telupid LS dialect. The data possibly suggests a dialect division between "Mangkaak" and "Labuk Kadazan", but mutual intelligibility can be expected in any case.

All of the Kadazan/Dusun testing showed negligible intelligibility, ranging 34-58%. This is not surprising since the test point samples only had a range of 61-73 PSC. The higher scores recorded on the Pahu TA reference tape (49-52%) as compared to the Taginambur KB reference tape (34-35%) are probably due to the simpler story rather than to greater linguistic similarity. The higher (41-58%) scores for the Entilibon KN reference tape is easily explained by its geographical proximity to the test points. It is significant enough that all test points scored substantially less than 75% intelligibility.

3.2 Further intelligibility testing

Intelligibility testing has not yet been completed for the Eastern Kadazan language. As Figure 4 indicates, an exhaustive investigation would require mutual intelligibility testing of as many as 21 relations. However, ten of these relations (marked with a hexagon) are sufficiently established on the basis of PSC alone to make dialect divisions, though mutual intelligibility would still not be certain. While all the relations are within the testable range of 75-85 PSC, the 11 relations marked with a question mark should be investigated first. Thus far only one of these has been fully tested, the Labuk/Mangkaak/Sukang dialect with the Lamag Sungai dialect. However, Malapi Kadazan intelligibility of Lamag Sungai and Labuk/Mangkaak/Sukang dialects has been tested. To complete this testing these dialects must be tested for comprehension of a Malapi Kadazan reference tape.

SEGALIUD SUNGAI					
⬡	MALAPI KADAZAN				
?	?	LABUK/MANGKAAK/SUKANG			
⬡	?	?	KULU-KULU SUNGAI		
⬡	⬡	?	?	KERAMUAK SUNGAI	
⬡	⬡	⬡	?	?	LAMAG SUNGAI
⬡	⬡	?	?	⬡	?
					TERUSAN SAPI KADAZAN

Figure 4: Required dialect intelligibility testing among seven Eastern Kadazan dialects. (The question mark indicates that the dialect status is doubtful. The hexagon indicates that the dialect status of the relation is established based on PSC relations.)

Within the Labuk/Mangkaak/Sukang grouping it was found that a dialect division between "Labuk Kadazan" and "Mangkaak" may be reasonable. Likewise the divisions between "Labuk Kadazan" and "Sukang", and between "Mangkaak" and "Sukang" should be tested for mutual intelligibility before any one of those subdivisions is used to represent the whole group.

Terusan Sapi Kadazan has very high PSC relations with Labuk/Mangkaak/Sukang and may belong to that group. Segaliud Sungai shows high relations with some members of the Labuk/Mangkaak/Sukang group and should be investigated on that account. Other testing, too, can be done within Smith's tentative classification of the Eastern Kadazan language group to determine the language and dialect boundaries.

4. NATIONAL LANGUAGE INTELLIGIBILITY

A national language tape by a native Sabahan was used at each test point to check that community's average intelligibility of Bahasa Malaysia. Figure 5 shows test point scores ranging 35-83% intelligibility. The test point where the highest intelligibility was recorded was at Batu Putih KN. The highest ratio of male to female subjects was also there, as was the most educated corpus of the seven.

TEST POINTS	AVG. SCORE	SEX M/F	AGE		EDUCATION		
			AVG.	RANGE	ED.	A.E.R.	A.E.S.
Batu Putih KN	83	9/1	28	16-60	8	9	7
Sogo-Sogo KN	63	6/4	34	17-50	4	4	1
Mananam KN 'A'	52	5/5	30	15-45	3	2	1
Malapi LS	49	6/4	35	21-56	2	3	1
Buis LS	41	5/5	38	21-55	1	9	1
Mananam KN 'B'	35	7/3	22	16-26	6	6	3
Langkabung KN	35	5/5	33	17-45	4	3	1

Figure 5: National language intelligibility at seven Eastern Kadazan test points with sociological data. (Scores are given as percentages; The education data for each village are listed as the number of persons in the test corpus who had received formal education (ED.), the average years of education which those subjects had received (A.E.R.), and the average years of education per subject in the test corpus as a whole (A.E.S.).)

The average score of all 70 subjects was 53% with scores ranging from 0-100% being registered. Nine subjects scored zero while the same number showed perfect intelligibility. All zero scores came from subjects not having had any formal education, and all perfect scores came from those having had at least four years of formal training using the medium of Bahasa Malaysia.

Figure 6 shows the intelligibility of the national language tape by education and sex. The significance of education is seen in the chart. All those having received any formal education from one to 11 years are considered under the group "educated". Those with no formal training at all are grouped together as "non-educated".

Whereas the group having some education averaged 81% intelligibility with no significant difference recorded between the 24 male and four female subjects, the group not having received any education averaged only 35% intelligibility, with the 19 male subjects averaging 54% and the 23 female subjects averaging 19%.

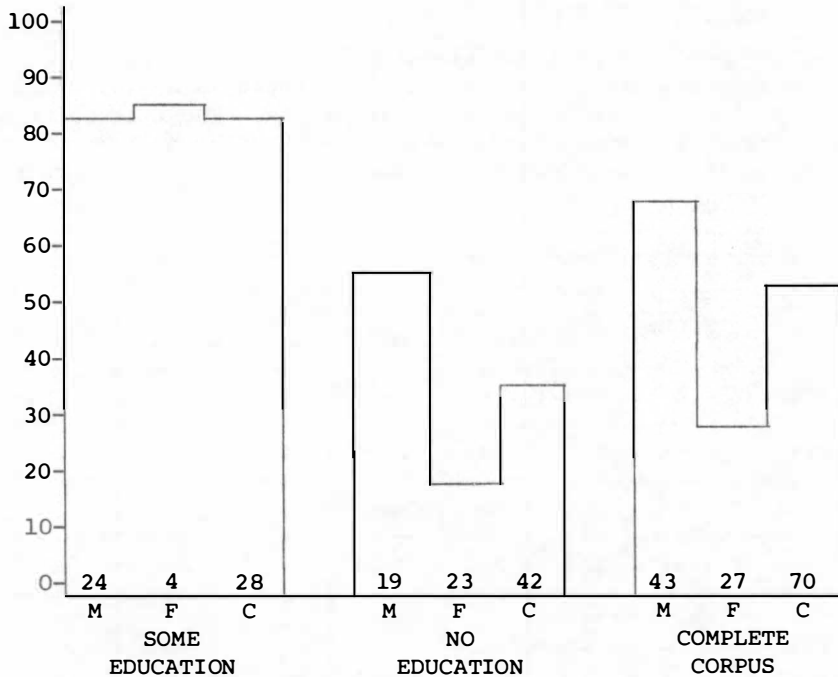


Figure 6: Eastern Kadazan speakers' intelligibility of the national language tape by education and sex. (The three levels in each block represent male (M), female (F) and combined (C) male and female corpus averages. The number within each column represents the size of the corpus used to calculate that intelligibility average.)

Since the corpus is not large enough or balanced enough to make a thorough study of the effect of age on the intelligibility scores, the only possible breakdown is between the "30 years and below" age group and the "above 30 years" age group. Figure 7 shows the averages of each group by sex and education as well as the combined averages.

Among the non-educated subjects age made very little difference in the female corpus. The "30 age under" group averaged 23% intelligibility while the "above 30" group averaged 21%. However, among the non-educated male corpus there is a substantial difference - 75% for the younger group and 45% intelligibility for the older group.

The most obvious conclusion from the test results of the national language tape is that those having received some education using Bahasa Malaysia as the medium of instruction have much higher comprehension of the national language. Secondly, sex makes no difference among the educated corpus but among the non-educated, males demonstrated intelligibility nearly three times as high as female subjects. Finally, the age factor, (outside of education) may indicate higher comprehension of Bahasa Malaysia among males under 30 years of age. No other factors influencing intelligibility have been isolated.

SEX	EDUCATION	AGE	
		30 YEARS AND UNDER	ABOVE 30
MALE	Ed.	84 (18)	72 (6)
	Non-Ed.	75 (6)	45 (13)
	Combined	81 (24)	54 (19)
FEMALE	Ed.	83 (4)	--
	Non-Ed.	23 (12)	21 (11)
	Combined	38 (16)	--
COMBINED		64 (40)	42 (30)

Figure 7: Eastern Kadazan speakers' intelligibility of the national language tape by age, sex and education. (The number within parentheses below each intelligibility average represents the size of the corpus used to calculate that average.)

5. CONCLUSION

The results of the dialect intelligibility testing in the Eastern Kadazan language confirm a division between the Lamag Sungai group and the Labuk/-Mangkaak/Sukang group, but might in fact, suggest a language level distinction rather than simply a dialect difference. A wide range of intelligibility occurred for Labuk/Mangkaak/Sukang comprehension of Lamag Sungai while Lamag Sungai showed only very limited intelligibility of Labuk/Mangkaak/Sukang.

Malapi Kadazan and Labuk/Mangkaak/Sukang testing is incomplete but tends to confirm Smith's dialect division between Malapi Kadazan and the "Mangkaak" member of the Labuk/Mangkaak/Sukang dialect. The results of testing the Malapi Kadazan comprehension of Lamag Sungai indicate either a dialect or language level division, thus confirming Smith's separation of these two dialects.

Within the Labuk/Mangkaak/Sukang dialect group the need for some rearrangement is quite certain. From the testing it is evident that a dialect distinction could be drawn between some members of the "Mangkaak" and the "Labuk Kadazan". These members could represent fringe areas of each and thus might still best be classified as one dialect.

Eastern Kadazan is clearly separate from the Kadazan/Dusun language, and there is no reason to expect it to show a closer relation with any other Dusunic language.

NOTES

1. The population figure of 14,000 to 16,000 is based on the following information (broken down by dialect group) :

SEGALIUD SUNGAI: 800-1000

The population figure given by the headmen at Longmanis (600) and the villages of Panganan and Segaliud in Sandakan District. This is very uncertain.

MALAPI KADAZAN: 100-200

The 1980 census lists 100 people for Malapi LS, but researchers were told that there are others who moved to various villages such as Ensuan LS when old Malapi LS moved closer to the East-West Highway several years ago.

"LABUK KADAZAN": 5000-5500

The 1980 census figures for the "Kadazans" of the following Labuk-Sugut District villages: Ulu Sapi, Ulu Sungai Sapi, Kuala Sapi, Tagas-Tagas Basai, Kamansi, Rumidi, Bilai, Berayong Ensuan, Ulu Ensuan, Buis, Gambaron, Bantu, Wonod, Lumau, Pandan-Pandan, Telupid Batu 4, Telupid Batu 6 and Kiabau. Other villages include a small minority of Labuk Kadazans and thus are not listed with these villages.

"MANGKAAK": 1100-1300

This is based on the 1970 "Mangkaak" population figure (962) plus 2% growth per year. Subjects and their parents came from the villages of Mananam (300), Tolunglokos, Langkabung (400), Sogo-Sogo, Kiliwatong, Kibungkawa and along the Malagatan Kecil and Malagatan Besar Rivers, all in the Kinabatangan District.

"SUKANG": 350-500

The Sukang populations as given by headmen for Entilibon KN (100) and Kuala Keramuak KN (250) as well as scattered individuals in such places as Bukit Garam KN, Kulu-Kulu KN, Telupid LS, and Sapi LS.

LAMAG, KULU-KULU, KERAMUAK SUNGAI: 6000-6500

The population figures were given by the headmen at Kuala Lokan (300), Buang Sayang (400), Kuala Keramuak (150), Kulu-Kulu (302), Batu Putih (300) as well as 21 villages located between Bilit KN and Kuala Keramuak KN along the Kinabatangan River and tributaries. The 1970 "Sungai" census stood at 17,535 which includes the Orang Sungai from all language groups.

TERUSAN SAPI KADAZAN: 700-900

The 1980 census figures were given for the "Kadazan" of Sualok, Panimbanan and Lidung and possibly others, all in Labuk-Sugut District.

2. The results of the Mananam KN 'A', Langkabung KN and Sogo-Sogo KN tests were included in Hurlbut and Pekkanen (to appear) who used a slightly different, but still comparable, approach and are adapted here from their work. Those tests included for the first time in data papers are the Batu Putih KN, Malapi LS, Buis LS and Mananam KN 'B' tests.

THE RUNGUS LANGUAGE

Julie K. King

0. INTRODUCTION

The people of Sabah who refer to themselves and to their language as Rungus or Memogun, reside in the Kudat, Pitas, Kota Marudu and northern Labuk-Sugut districts. The largest numbers of Rungus however live on the Kudat and Bengkoka (formerly Melobong) Peninsulas in the northernmost part of Sabah (Figure 1).

There is considerable variation in the literature with regard to the nomenclature applied to this group of people. G.N. Appell, who has done considerable anthropological and linguistic research among the Rungus, says that this group of people, whom he refers to as the Momogun, "... are divided into sixteen or more different, localized and named sub-groups differentiated from each other by dialectical differences, differences in adat (customary ways), and by the conviction of the members of each of these sub-groups that they belong to a separate ethnic entity." (Appell 1963). Appell further classifies Rungus as a language of the Dusunic family of languages. He says that Dusun is an exonym used to refer to the whole language family and that Rungus is an autonym (Appell 1968).

Prentice (1970) classifies Rungus as a member of the Dusunic subfamily within the Idahan family of languages. He further breaks down the Dusunic subfamily into the Coastal Dusun language and the Marudu dialect. Rungus is considered another name for this dialect and is also the name of one of three subdialects distinguished in Prentice's classification. The other two subdialects he mentions are Nulu, located on the northern tip of the Kudat Peninsula and Gonsomon, to the south of the Kudat Harbour.

Clayre (1966) distinguishes between the Rungus dialect of Dusun spoken on the Kudat Peninsula and the Bengkoka dialect of Dusun of the Bengkoka Peninsula. In her comparison of shared cognates between the two she found them to have 90 percent of lexical similarity.

For the purpose of this present study, the name used in referring to this people and their language will be the autonym used by the village people when technicians collected the linguistic data presented here.¹ This name, Rungus, also reflects most clearly the high level of linguistic unity found to exist among these people.

These problems of nomenclature also effect the demographic analysis of the group under study. The 1970 government census figures gave a total of approximately 11,000 Rungus people in Sabah at that time. At an annual growth rate

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 283-296. *Pacific Linguistics*, C-78, 1984.

© Julie K. King

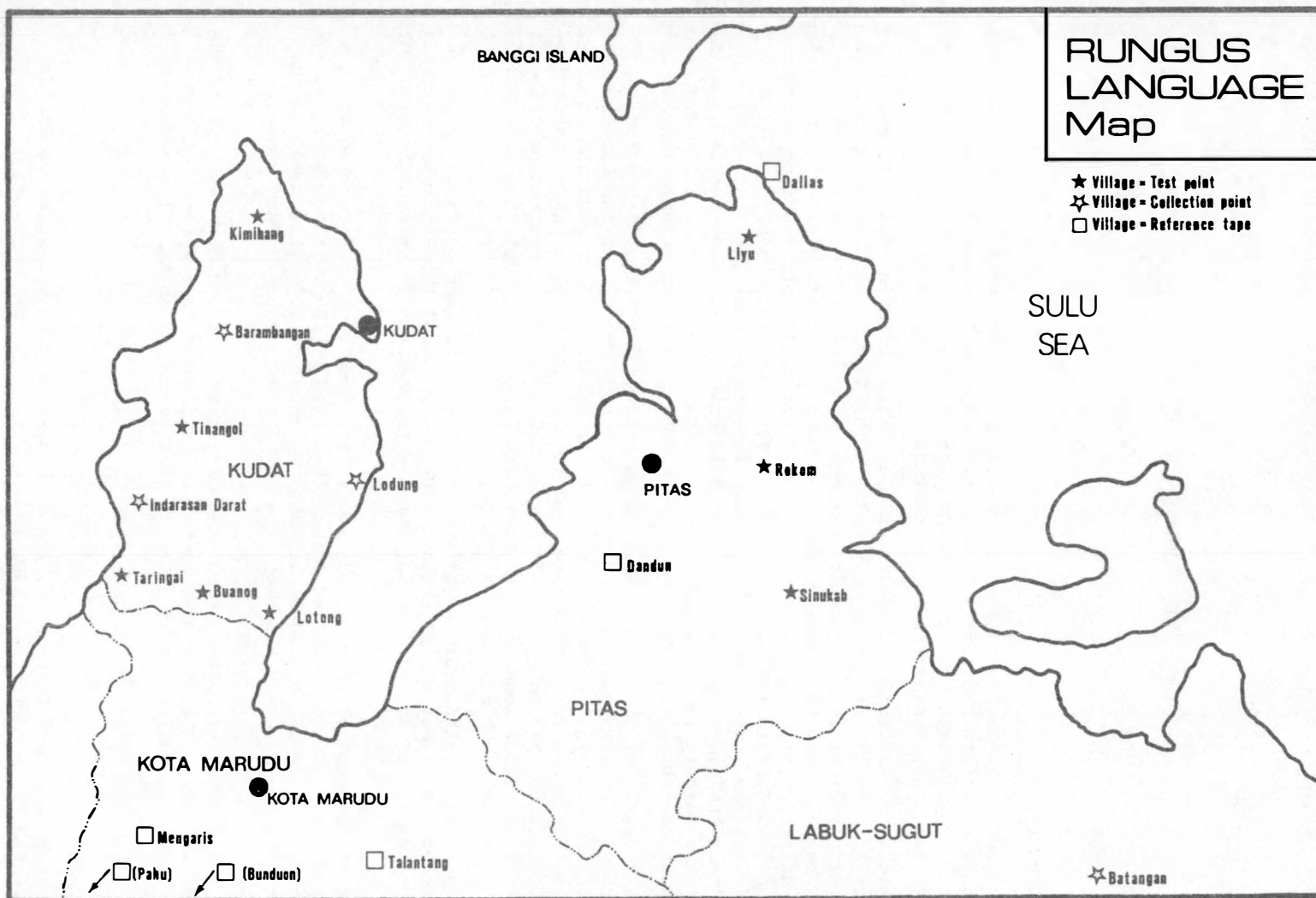


Figure 1: Rungus language map

of about 2%, the total for 1982 should be around 14,000.² The figures from the 1980 census are displayed by district rather than by ethnic group and so offer little assistance in current calculations.³

Because of the problems of nomenclature mentioned above, it is very likely that some speakers of Rungus as it is defined here, gave a variant autonym to census officials and were then listed even in the 1970 census with that other group rather than with Rungus.⁴

It seems that for the purpose of this study, which is concerned with lexical classification of languages and dialect intelligibility testing, one further source is very reliable in indicating who speaks the Rungus language. Church workers who work among this group have translated some written materials into the Rungus language. On the basis of how widely that material is used in the Rungus areas in the Kudat, Pitas and Kota Marudu districts as well as significant research done by the church workers and the self-designation used by these people, they estimate that there are approximately 40,000 speakers of Rungus in Sabah.⁵

Aside from the sources already mentioned, further linguistic work on the Rungus language is primarily in the form of unpublished manuscripts and mimeographed booklets which have been published locally and have had limited distribution.

In addition to Clayre's dialect comparison including Rungus, which was published in 1966, she also did an earlier comparison of the Ranau, Rungus and Ulu Tuaran dialects of Sabah in 1964 which is unpublished. A further work on Rungus by Clayre is a mimeographed booklet of Rungus root words.

An English-Rungus dictionary was published and mimeographed by the Protestant Church of Sabah in 1966. And there is an Outline of Rungus Grammar by T. Forschner dated 1978, also in mimeographed form.

Various portions of the Christian Scriptures and songs have also been translated into Rungus and mimeographed for distribution in the Rungus areas of Sabah.

Aside from the dialect comparisons by Clayre, the description of the language situation in Kudat District and surrounding areas by Appell and the language classification done by Prentice, to date, in the knowledge of this author, there has been no clear and comprehensive published statement as to the degree of intelligibility within this group or between Rungus and other related groups in Sabah (Blom 1979, Appell 1968).

1. LEXICOSTATISTICAL CLASSIFICATION

When 12 wordlists representing the Rungus language from the districts of Kudat, Pitas and Labuk-Sugut were compared, the lexical relations between them ranged from 86-96 percent of shared cognates (PSC). Figure 2 displays this and also shows that Rungus is a close-knit and well-defined unit linguistically, with only minimal lexical variation. Further, Figure 2 shows that the geographically most remote of the villages used in that comparison, Batangun LS, had the consistently lower relations (86-90) PSC. When Batangun LS is excluded from the comparison the range is 88-96 PSC.

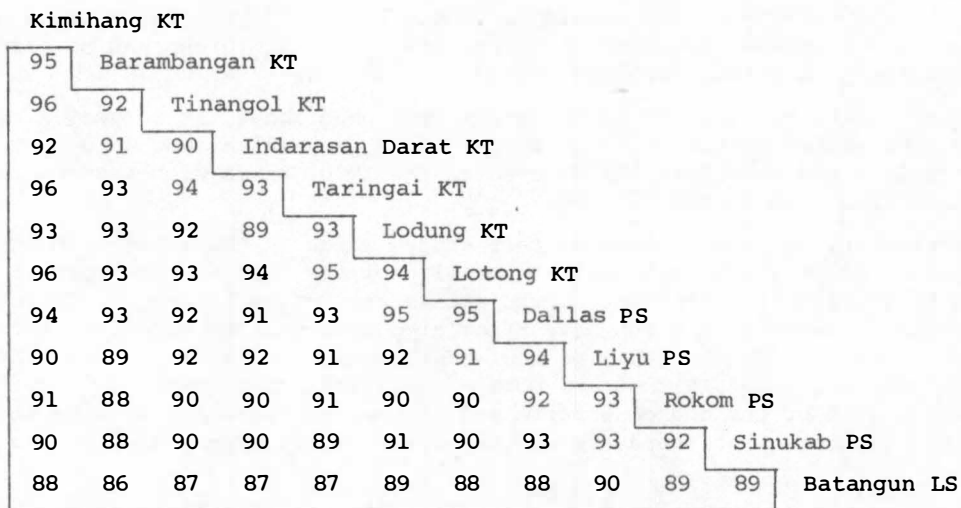
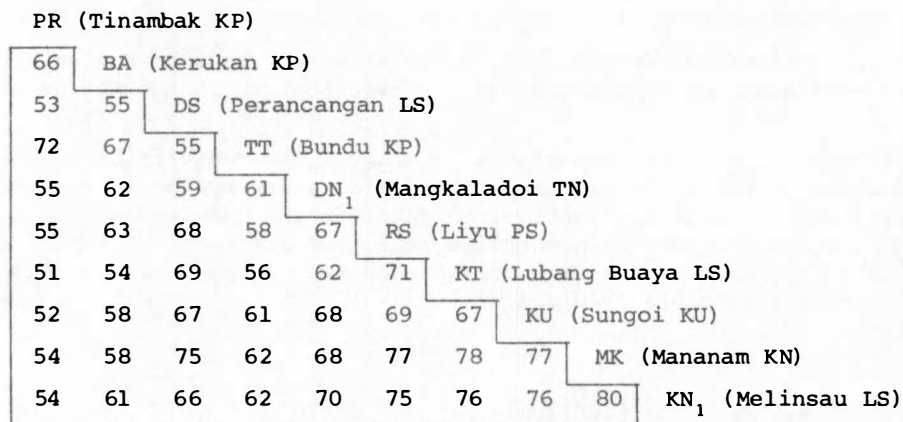


Figure 2: PSC relations of 12 Rungus villages

According to Smith's lexicostatistical comparison (in this volume), Rungus is a language of the Dusunic family of languages. He notes further however that the relation of the Rungus language with other languages of the Dusunic family is 55-77 PSC. Figure 3 shows a comparison of the nine languages and one dialect of the Dusunic family using the wordlist from the village of Liyu PS as a representative of the Rungus language.

Figure 3: Comparison of PSC scores between examples of nine Dusunic languages and one dialect of Dusunic (KN₁ Melinsau LS).

(PR = Papar; BA = Bisaya; DS = Dumpas; TT = Tatana; DN₁ = "Dusun" representing Lotud; RS = Rungus; KT = Kadazan-Tambanua; KU = Kuijau; MK = "Mangkaak" representing Eastern Kadazan; KN₁ = "Kadazan" representing Sugut Kadazan dialect.) (Adapted from Smith, in this volume.)

Clayre (1966) has done similar lexical comparisons of Dusunic languages of Sabah though her investigation with regard to lexical comparisons was less extensive than Smith's. An additional difference between her work and Smith's is that she used the Swadesh Diagnostic wordlist rather than the larger wordlist adapted from the Philippines which was used by Smith.⁶

On the basis of Clayre's lexical analysis and her threshold for same and different language/dialect, Rungus appears to be the same dialect as her Bengkoka, marginal with her Ranau, Bundu, Tambunan, Penampang and Labuk languages and a different language from her Mangkok and Lotud.⁷

It should be noted then that in general the two comparisons are quite consistent except that in several cases where Clayre's figures are near the borderline of what is considered same or different language (80 and 83 PSC), Smith's are significantly lower. Thus, in Smith's lexicostatistical analysis, Rungus was clearly a different language from all other languages within the Dusunic family of languages.

2. TESTING PROCEDURES

Prior to the processing of all of the lexicostatistical data displayed in Figure 3, some preliminary intelligibility testing was done in the Rungus language area to determine the level of intelligibility between Rungus speakers and speakers of other languages within the geographical boundaries of what was then known as the Kudat Division - the Kudat, Pitas and Kota Marudu districts.⁸ So the initial testing in the Rungus villages was only with languages which were represented in those three districts (Blom 1979). Aside from testing Rungus subjects with Rungus reference tapes, in the initial testing the subjects were also tested with a language that Smith classified as a peripheral representative of the Central Dusun dialect from Mengaris KM, and two which have been classified as part of the Outer Sugut Kadazan subdialect - Dandun PS and Talantang KM (Smith in this volume).⁹ Thus the purpose of the initial intelligibility testing in the Rungus area was to determine the level of understanding between speakers of all of the languages in those three districts.

Seven Rungus villages were chosen for the initial testing. The four test villages from Kudat District were Tinangol KT, Kimihang KT, Lotong KT, and Taringai KT. The three Rungus test villages in Pitas District were Liyu PS, Rokom PS, and Sinukab PS. These seven villages were chosen purely on a geographical basis since the minimal lexical variation revealed by the PSC comparisons was not considered to be significant.

At each of these seven villages the test set included a hometown tape, a reference tape from two other Rungus villages, the two reference tapes from the Outer Sugut Kadazan subdialect from Dandun PS and Talantang KM, the peripheral Central Dusun dialect reference tape from Mengaris KM and the Bahasa Malaysia reference tape (Figure 4).¹⁰

When the first phase of the entire survey of Sabah was complete and all of the comparisons included in Figure 3 had been analysed, it was considered necessary to test Rungus speakers for their understanding of other examples of Dusunic languages from outside of the Kudat Division, which were considered to be representative of larger and more well-defined languages of the Dusunic family.

Two reference tapes were chosen from the villages of Pahu TA and Bunduon PG. The languages from both of these villages were classified as peripheral villages of the Central Dusun dialect (Smith, in this volume). But the relation between the two is only 74 PSC.

The purpose of this second testing in the Rungus area then was to test the understanding of Rungus speakers with two major Dusunic languages spoken outside of the Kudat Division (→ Kadazan/Dusun).

Because of the very high PSC relations between all of the Rungus villages which had been visited earlier in the survey, two were chosen rather arbitrarily for this further testing. Those villages were Kimihang KT and Buanog KT.

At these two villages only three tapes were used for testing. The first was the reference tape from Tinangol KT representing the Rungus language. The second was the reference tape from Pahu TA representing a dialect of Central Dusun. And the third was from Bunduon PG and also represented a Central Dusun dialect. It should also be noted that the dialect from Bunduon PG is also the dialect which appears in newspapers in Kota Kinabalu and is used for radio broadcasts.

All of the taped stories used in the initial testing and in the second test in the Rungus language area were considered to be good with regard to content and technical quality.

3. TEST RESULTS

3.1 Dialect intelligibility testing results in Rungus villages

Figure 4 displays the results of the intelligibility testing in Rungus villages.¹¹ The results of the first and second testings are given in that figure. It will be clearer however to discuss these results in three groups: 1) Rungus understanding of Rungus, 2) Rungus understanding of Outer Sugut Kadazan, and, 3) Rungus understanding of Central Dusun.

3.1.1 Rungus understanding of Rungus

The results of using Rungus language reference tapes for testing in Rungus-speaking villages are almost entirely predictable from the PSC relations between them. In only four cases, villages scored below 90% intelligibility on either of the two Rungus language reference tapes. That was at Rokom PS where a score of 89% was achieved, at Taringai KT where scores of 84% and 86% were achieved and at Lotong KT where a score of 88% was achieved. In all of the 11 other cases where the Tinangol KT reference tape or the Dallas PS reference tape was used, the subjects scored over 90% intelligibility and so belong to the same language and dialect.

Though the scores from Rokom PS, Taringai KT and Lotong KT are somewhat lower, they are still considered to be the same language and even the same dialect as the other Rungus villages. This is based on a couple of non-linguistic factors which can help to explain the lower scores.

REFERENCE TAPES	RUNGUS		OUTER SUGUT KADAZAN		CENTRAL DUSUN		
	Tinangol KT ("Rungus")	Dallas PS ("Rungus")	Talatang KM ("Garo")	Dandun PS ("Kimaragang")	Mengaris KM ("Tebilung")	Pahu TA ("Dusun")	Bunduon PG ("Kadazan")
TEST POINTS							
Liyu PS ("Rungus")	91 (92)	97 (94)	60 (78)	61 (79)	60 (76)		
Rokom PS ("Rungus")	89 (90)	94 (92)	63 (77)	63 (78)	70 (75)		
Sinukab PS ("Rungus")	96 (90)	95 (93)	62 (79)	65 (79)	67 (76)		
Tinangol KT ("Rungus")	100 100	91 (93)	69 (76)	70 (79)	67 (77)		
Lotong KT ("Rungus")	98 (93)	88 (95)	66 (77)	64 (79)	65 (76)		
Taringai KT ("Rungus")	84 (94)	86 (93)	65 (77)	67 (78)	62 (75)		
Kimihang KT ("Rungus")	96 *100 (96)	95 (94)	72 (78)	72 (79)	71 (76)	*39 (71)	*45 (72)
Buanog KT ("Rungus")	*99 (98)					*44 (71)	*43 (73)

Figure 4: Results of first and second intelligibility testing (shown as percentages) in Rungus villages. (PSC relations are in parentheses; autonyms are in quotation marks; * indicates a score from the second Rungus testing.)

These villages are considerably more remote than the others which were tested. At the time the testing was done, Rokom PS was still inaccessible by road. Taringai KT at that time, had had a road for only a short time. And Lotong KT, though it was accessible by road, was quite distant from the main road to Kudat town. In all three villages then, communication and transportation between villages was more limited than for the others tested, so that the people did not frequently mix with people from distant Rungus villages.

A second factor tending to lower the scores from these three villages is the lower level of education of the test subjects at Taringai KT, Lotong KT and Rokom PS as compared with the other Rungus test villages (Figure 5). This

factor could influence how well a subject understood the procedures for taking the test and what was expected of him in taking the test.

TEST POINTS	EDUCATION		
	ED.	A.E.R.	A.E.S.
Rokom PS	0	-	-
Taringai KT	0	-	-
Lotong KT	3	4.3	1.3
Liyu PS	4	6	2.4
Sinukab PS	3	4.6	1.4
Tinangol KT	8	6	5
Kimihang KT	7 *4	6.4 *6.2	4.5 *2.5
Buanog KT	*5	*8	*4

Figure 5: Education data for test subjects from eight Rungus villages. (ED. indicates the number of persons in the test corpus who had received some formal education; A.E.R. indicates the average years of education those subjects had received; and A.E.S. indicates the average years of education per subject in the test corpus as a whole; * indicates data from the second test.)

3.1.2 Rungus understanding of Outer Sugut Kadazan

As can be seen in Figure 4, seven Rungus villages were tested for their understanding of Outer Sugut Kadazan. The two Outer Sugut Kadazan reference tapes were from the villages of Talantang KM and Dandun PS. The people from Talantang KM use the autonym "Garo" to refer to themselves and their language. In Dandun PS the autonym "Kimaragang" is used.

In most cases the intelligibility of Outer Sugut Kadazan by Rungus speakers is significantly lower than corresponding PSC relations. In all cases, the intelligibility scores are well below the 80% threshold which defines same or different languages.

So the intelligibility testing confirms the lexicostatistical classification of Smith, who calls Rungus and Outer Sugut Kadazan two separate languages. That the PSC relations are significantly higher than intelligibility testing results should not be disturbing since both figures for each village are well below the point at which two languages are considered to be different.

It is very likely, as the test subjects listened to the tapes from Talantang KM and Dandun PS, that they immediately perceived that the tape was in a

different language from their own. This perceived difference in itself could make the test more difficult and so account for the lower scores.

3.1.3 Rungus understanding of Central Dusun

The results of testing Rungus speakers' understanding of Central Dusun are a bit more difficult to evaluate.

Three Central Dusun reference tapes were used at Rungus villages on two separate occasions. In the initial testing, the reference tape from the village of Mengaris KM was used for testing at seven Rungus villages. The range in the relations between Mengaris KM and the seven Rungus villages is 75-77 PSC. The range of intelligibility scores is 60-71%. Though the intelligibility scores are again consistently lower than the PSC relations, this is not disturbing for the same reasons as those mentioned in Section 3.1.2.

The surprising results, however, were those received when the two other Central Dusun reference tapes from Pahu TA and Bunduon PG were used for testing at two Rungus villages - Kimihang KT and Buanog KT.

The results at those two villages, also shown in Figure 4, gave intelligibility scores much lower than corresponding PSC relations and also much lower than the intelligibility testing results with the Central Dusun reference tape from Mengaris KM.

There are two possible explanations for these results. The first is based on the premise that when similarity between two languages becomes less (e.g. lexical similarity), at a point, intelligibility begins to sharply decline. This would help to explain that although the PSC relations of Pahu TA and Bunduon PG with Rungus villages are a matter of only 4-6 percentage points lower than the PSC relations between Mengaris KM and Rungus villages, the intelligibility is drastically reduced.

Another explanation for the great difference in intelligibility scores with the Mengaris KM reference tape and the reference tapes from Pahu TA and Bunduon PG is the location of the three villages in relation to the Rungus language area. The village of Mengaris KM is right on the borderline between the Rungus language area and a different Dusunic language area. The villages of Pahu TA and Bunduon PG are much farther from any Rungus-speaking areas. This explains why the language from Mengaris KM is more familiar to Rungus speakers than are the others. A higher score would then be expected on the Mengaris KM reference tape.

3.2 Kadazan/Dusun understanding of Rungus

It is very interesting to note that when cross-testing was done and Rungus language reference tapes were used for testing in Kadazan/Dusun villages, a similar phenomenon occurred as when Kadazan/Dusun reference tapes were played in Rungus villages (→ Kadazan/Dusun). Figure 6 displays those test results. The significant factor is that when these Kadazan/Dusun villages have a PSC relation with Rungus that is in the low 70s, the intelligibility testing results are much lower, as was the case in the Rungus village testing (→ Section 3.1.3). And again, when the PSC's are in the upper 70s as with Outer Sugut Kadazan and Rungus, the intelligibility scores are about ten percentage points lower (→ Section 3.1.2).

TEST POINTS		REFERENCE TAPE	RUNGUS
OUTER SUGUT KADAZAN	Talantang KM ("Garó")	68 (78)	
	Dandun PS ("Kimaragang")	70.5 (79)	
KADAZAN/DUSUN	Tangaban TA ("Dusun")	40 (73)	
	Toboh TA ("Dusun")	19 (70)	
	Sunsuron TA ("Dusun")	39 (71)	
	Kogopon PR ("Kadazan")	48 (74)	
	Kionsom KK ("Dusun")	50 (70)	
	Lumpoho/Pukak TN ("Ulu Tuaran")	45 (72)	

Figure 6: Results of intelligibility testing using a Rungus language reference tape at eight Kadazan/Dusun villages. (Scores are given as percentages; PSC relations are in parentheses.)

4. NATIONAL LANGUAGE INTELLIGIBILITY

National language intelligibility testing was done in Rungus villages only in the initial Rungus testing (Blom 1979). Persons from seven Rungus villages in the Kudat and Pitas districts were then tested for their ability to understand the national language. The results of the testing on the Bahasa Malaysia reference tape are displayed in Figure 7.¹²

The range of average intelligibility scores was 42-81%. It appears from Figure 7 that age and education might be affecting the scores.

The raw data has been reorganised in Figure 8 to show: the average score for males as compared to that of females; the average score for subjects under 35 years of age as compared to that of subjects over 35; and to show the average score for subjects with formal education as compared to those with no formal education.

From Figure 8 it can clearly be seen that educated subjects scored significantly higher than non-educated subjects, and also that male subjects consistently scored higher than female subjects.

TEST POINTS	AVG. SCORE	SEX M/F	AGE			EDUCATION		
			AVG.	MED.	RANGE	ED.	A.E.R.	A.E.S.
Tinangol KT	81	6/4	21	23	14-32	8	6	5
Lotong KT	44	5/5	31	32	13-50	3	4.3	1.3
Kimihang KT	69	6/4	25	24	15-32	7	6.4	4.5
Taringai KT	48	5/5	33	33	17-48	0	-	-
Liyu PS	42	6/4	34	36	18-53	4	6	2.4
Sinukab PS	44	7/3	40	44	30-58	3	4.6	1.4
Rokom PS	43	7/3	37	36	22-50	0	-	-

Figure 7: Results of national language intelligibility testing with socio-logical data for seven Rungus villages. (Scores are given as percentages; RANGE indicates the age of the youngest and oldest test subjects; the education data for each village are listed as the number of test subjects who had received formal education (ED.), the average number of years of education which those subjects had received (A.E.R.), and the average number of years of education per subject (A.E.S.).)

TEST POINTS	CORPUS SCORE	SEX		AGE		EDUCATION	
		AVG. M SCORE	AVG. F SCORE	AVG. UNDER 35 SCORE	AVG. OVER 35 SCORE	AVG. ED. SCORE	AVG. NON ED. SCORE
Tinangol KT	81	81/6	77/4	81/10	0/0	86/8 5M 3F	55/2
Lotong KT	44	46/5	40/5	50/7	26/3	76/3 2M 1F	28/7
Kimihang KT	69	73/6	60/4	68/9	70/1	70/7 4M 3F	60/3
Taringai KT	48	54/5	42/5	50/4	46/6	0/0	48/10
Liyu PS	42	48/6	32/4	58/5	26/5	67/4 3M 1F	25/6
Sinukab PS	44	51/7	26/3	43/3	44/7	50/3 3M	41/7
Rokom PS	43	41/7	40/3	44/5	38/5	0/0	43/10

Figure 8: National language intelligibility results in seven Rungus villages displayed according to sex, age and education. (The average score for each category (shown as percentage) is followed by the number of subjects for which the average applies. The boxes containing scores of those who had received formal education (AVG.ED.SCORE) also show the number of males and females in the educated group. NON-ED. indicates subjects without formal education.)

5. CONCLUSION

Dialect intelligibility testing in Rungus villages confirms the lexico-statistical analysis of Smith who classifies Rungus as a "... close-knit, well-defined language without notable dialects." (in this volume, p.31).

In the testing and cross-testing of Rungus with representatives of Kadazan/Dusun languages it was also determined that mutual intelligibility with Outer Sugut Kadazan is scored in the 60s and 70s and mutual intelligibility with Central Dusun languages is even lower. This helps to provide further verification that Rungus is a separate language from those with which it was tested.

NOTES

1. The use of the name Rungus in this paper, as applied to the people and the language they speak, is different from the use of the same term by G.N. Appell. Here Rungus is used to apply to the language and people who themselves used the term to refer to themselves when the data for this paper was first collected. Appell (1963) uses the term Rungus to apply to only one of possibly more than 16 subgroups of "Momogun". He states further that these subgroups are differentiated on the basis of differences in dialects, customary ways, and by their own belief that they are different from the other subgroups.
2. The 1970 Sabah census listed approximately 11,000 Rungus-speaking people living in Sabah. The 2% annual growth rate figure is merely an arbitrary figure used to estimate a more recent population figure, as such figures are not otherwise available.
3. The 1980 census figures for Sabah are published in the Sarawak Gazette and give population figures by district only. The figure given for Kudat District is 41,872 and for Pitas District is 17,164. The figure of 28,249 was given for the district of Kota Marudu. It must be remembered however that especially in the case of Kota Marudu, but also in the cases of the other two districts, there are other language groups represented by those population figures.
4. It was noted that some Rungus individuals would use the term "Dusun" to apply to themselves and their language. This appeared to be an effort to associate themselves with a larger and more well-known group in Sabah. Though this happened only very rarely in the experience of the technicians who collected the data for this paper, it should be kept in mind that this may have been a more frequent occurrence for the government census-takers and so may skew the population figures with regard to ethnic or linguistic differences.
5. This information was obtained from Rev. Poong Shong Khon and was later confirmed by Mr Johnny Sokuroh, both of the Protestant Church of Sabah.
6. The wordlist which was used for this study in Sabah was adapted from the wordlist used by the Summer Institute of Linguistics in the Philippines. This was done in order to facilitate comparison between the languages of both areas. There are 367 items on this list, 340 of which were used for the calculation of percentage of shared cognate (PSC) figures between languages.

It is not possible to do a completely accurate comparison between Clayre and Smith since they used different wordlists, their criterion for determining "cognateness" may have been different, and the exact locations from which Clayre's lists were taken are not known.

Figure 9 (Clayre 1966) shows the PSC relations between some "dialects of Dusun" including Rungus. The abbreviations, language designations and general locations where they are spoken are as follows:

- Rn = Ranau from around the Ranau area
- B = Bundu Tuhan area down the Tuaran valley and along the western and north-western slopes of Mt Kinabalu
- T = Tambunan - Tambunan area
- P = Penampang dialect around Kota Kinabalu area

THE KADAZAN/DUSUN LANGUAGE

John and Elizabeth Banker

0. INTRODUCTION

The group of people whose language is under discussion here are known by the general names Kadazan and Dusun. Various other ethnonyms also exist to distinguish subgroups or clan groupings within the larger group.¹

The name Dusun (also Dusan, Dusum, Dusur) literally means *orchard* and is said to give the connotation of *rural* (LeBar 1972). The name is said to be an exonym which was applied to the local people of Sabah by Malay and English voyagers (Appell and Harrison 1968).

The wide application of the term Dusun makes it difficult to distinguish the groups under discussion here from others in Sabah, especially in early writings. The term has been applied to many of the aboriginal peoples of Sabah and parts of Kalimantan and West Malaysia who have little in common linguistically and culturally.²

The term Kadazan (also Kadasan and Kadayan) appears to have originally referred only to the people of Penampang and Papar. The name is gaining popularity, however, so that many in other areas are using it as a general autonym (LeBar 1972, Whelan 1970).

This paper is concerned with only one of ten languages classified by Smith (in this volume) as Dusunic. Smith called the language under discussion here Kadazan/Dusun, and that language designation will be followed throughout this paper.

The origin of this group is uncertain, though it is generally agreed that Kadazan or Dusun peoples were among the first inhabitants of Sabah (cf. Whelan 1970; LeBar 1972; Chatfield 1972; Staal 1923). Glyn-Jones (1953:12-13) maintains that they were originally an inland people, although they are now also inhabiting coastal areas.

People speaking the Kadazan/Dusun language as it is delimited here live in parts of 13 of the 23 administrative districts of Sabah (Figure 1). There are heavy concentrations of Kadazan/Dusun in the districts of Ranau, Tambunan, Penampang, Papar, Tuaran, Kota Belud, and those parts of Kota Kinabalu outside the city. Speakers of the Kadazan/Dusun language also live in Beaufort, Kinabatangan, Labuk-Sugut, and Keningau districts, with some migrant villages in the districts of Tenom and Tawau.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 297-324. *Pacific Linguistics*, C-78, 1984.

© John and Elizabeth Banker

297

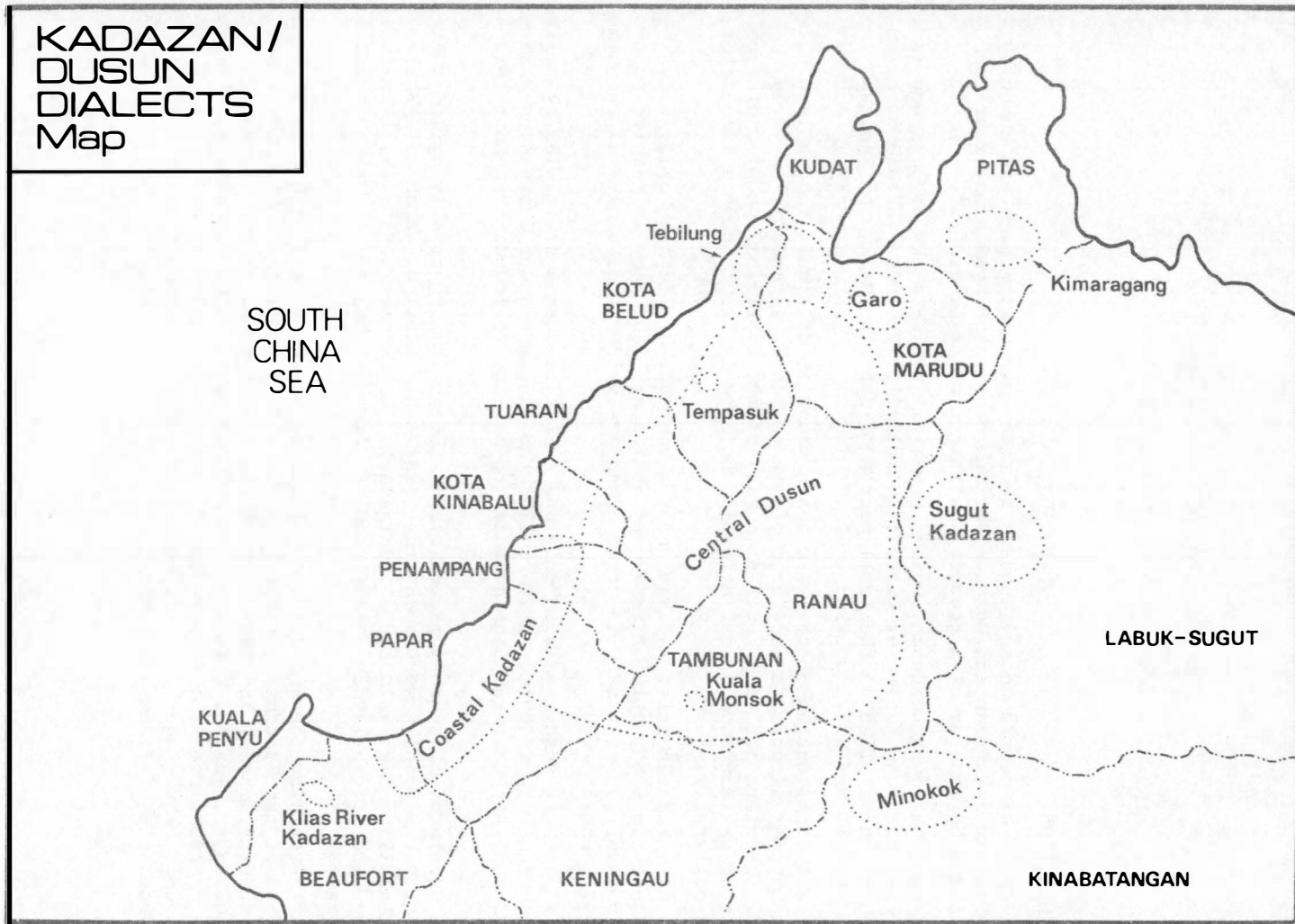


Figure 1: Kadazan/Dusun dialects map

The 1970 Sabah census listed 183,454 Kadazans in Sabah. However, some groups considered in this paper may not be included in that figure, and other groups included in the Kadazan census figures may differ linguistically from the Kadazan/Dusun under discussion here. The Kadazan/Dusun language is the mother tongue of more people than any other indigenous language of Sabah.

A considerable amount of material has been written about Kadazan/Dusun, though much is in the form of unpublished manuscripts and much deals with only a part of the language group.

Appell (1968) discusses the languages of Sabah but pays particular attention to the Rungus language (→ Rungus). Asmah Haji Omar has made a detailed description of verbal affixation in Kadazan (1978) and also a description of the Kadazan language of Tambunan (1980). Clayre has written a number of articles dealing with part of the Kadazan/Dusun language and has included some other Dusunic languages in her dialect comparisons (1964; 1965). In addition to general grammatical descriptions (1965; 1970), she has written on Dusunic clause types (1967) and focus (1970).

Some of the earliest writings about this language were wordlists such as the Dusun wordlist from Kimanis PR, published by Luering (1897), and Swettenham's (1880) wordlist from Tempasuk KB.

Majusim (1949) has discussed morphology and syntax of Dusun, and Pekkanen (1981) has done a study of sound changes in the Kadazan language from three districts of Sabah.³

1. LEXICOSTATISTICAL CLASSIFICATION

In his lexicostatistical classification, Smith (in this volume) delimited the Kadazan/Dusun language as one of ten languages in the Dusunic language family. Figure 2 shows the percent of shared cognates (PSC) between representative samples of the ten Dusunic languages with the figures from Kogopon PR representing the Kadazan/Dusun language.



Figure 2: PSC relations between representative samples of ten Dusunic languages (adapted from Smith, in this volume) (PR = Papar; BA = Bisaya; DS = Dumpas; TT = Tatana; DN = "Dusun" representing Lotud; RS = Rungus; KT = Kadazan-Tambanua; KU = Kuijau; MK = "Mangkaak" representing Eastern Kadazan; KD = Kadazan/Dusun.)

On the basis of 110 wordlists, Smith postulates 13 dialects of Kadazan/-Dusun: Central Dusun, Keningau Dusun, Pilantong (LS) Kadazan, Tempasuk (KB) Dusun, Tombovo (PG) Kadazan, Kota Kinabalu Kadazan, Kota Belud Dusun, Beaufort Kadazan, Sugut Kadazan, Paginatan (RU) Dusun, Kuala Monsok (TA) Dusun, Pemilaan (TM) Dusun, and Bungaraya (KU) Dusun. Figure 3 displays PSC figures between representative samples of the 13 dialects.

Sugut Kadazan (Melinsau LS)

76																		Pilantong Kadazan (Pilantong LS)
75	78																	Beaufort Kadazan (Tahak BT)
81	81	84																Central Dusun (Kogopon PR)
77	75	75	84															Kota Kinabalu Kadazan (Natai KK)
76	75	74	82	74														Kota Belud Dusun (Rosok KB)
71	69	71	75	69	82													Tempasuk Dusun (Tempasuk KB)
74	80	79	80	75	74	69												Tombovo Kadazan (Tombovo PG)
72	75	74	80	76	76	69	73											Keningau Dusun (Marapok KU)
81	73	71	78	77	74	69	72	72										Paginatan Dusun (Paginatan RU)
81	73	74	83	81	73	67	72	75	79									Kuala Monsok Dusun (Kuala Monsok TA)
79	72	72	80	80	75	70	70	75	75	79								Pemilaan Dusun (Pemilaan TM)
79	75	75	80	77	76	70	72	78	77	81	82							Bungaraya Dusun (Bungaraya KU)

Figure 3: PSC relations of 13 Kadazan/Dusun dialects (adapted from Smith, in this volume)

Smith proposes a Central Dusun dialect with a core of 14 villages (wordlists) and a periphery of 60 villages. He then postulates 12 smaller dialects. Three of these are rather large - Beaufort Kadazan, Kota Belud Dusun, and Sugut Kadazan.

2. TESTING PROCEDURES

Since the Kadazan/Dusun language comprises a large geographical area with many dialect distinctions, it was impossible to test every dialect with every other dialect. And at the onset it was unclear what distinctions would be the most important to test. In addition, since the test points and test sets were set up before Smith's 1980 classification, they are based on a preliminary 1979 classification in which Smith used a somewhat different method in determining dialects. This presents some problems in correlating the 1980 lexicostatistical classification with the intelligibility testing. Major patterns of dialect grouping had begun to emerge and it was on the basis of these patterns that the test points and test sets were determined. This means that some of the smaller dialects proposed in the 1980 classification were not tested. These include the Kota Kinabalu Kadazan dialect spoken at Natai KK and four migrant-village dialects: Keningau Dusun of Marapok KU, Pilantong (LS) Kadazan, Pemilaan (TM) Dusun, and Bungaraya (KU) Dusun. Residents of these migrant villages emigrated from their original language areas within the last 30 years. Apparently,

residents of these villages have taken on some new vocabulary from speakers of other dialects and languages with which they now have contact. No intelligibility testing was done there because it was felt that speakers of these "migrant dialects" would still understand the speech of their original areas.

3. TEST RESULTS

Intelligibility testing results indicated the need to alter Smith's (in this volume) classification of the dialects of the Kadazan/Dusun language. The discussion below follows the revised classification system.

3.1 The Kadazan/Dusun language

Intelligibility testing disclosed the existence of at least four dialects of the Kadazan/Dusun language: Central Dusun, Coastal Kadazan, Sugut Kadazan, and Kuala Monsok Dusun (Figure 4).

3.1.1 Central Dusun

Eight Central Dusun reference tapes were tested in other Central Dusun villages (Figure 5). Results are displayed in Figure 6.

A number of villages classified by Smith (in this volume) as members of separate dialects were found to belong to the Central dialect. Paginatan RU demonstrated 95% and 87% understanding of Moyog PG and Piasau KB reference tapes, despite cognate relations of only 76 and 74 PSC. The Paginatan RU reference tape was understood at a level of 93-100% at eight other Central Dusun villages with which its PSC relations range only 69-77. Based on these high intelligibility scores it is now considered to be part of the Central Dusun dialect.

Tempasuk KB, also considered by Smith to represent a distinct dialect of Kadazan/Dusun, demonstrated a 96-100% range of intelligibility of three Central Dusun reference tapes, compared to 69-76 PSC with those same villages. It was thus reclassified as part of the Central Dusun dialect.

Three villages formerly classified as Kota Belud Dusun have been reclassified as Central Dusun. These villages, Tambulian Laut KB, Dudar KB, and Mandap KB demonstrated 93-100% intelligibility of three Central Dusun reference tapes, in contrast to PSC relations of only 73-84.

Figure 7 demonstrates the distinctiveness of Central Dusun from two other dialects of the Kadazan/Dusun language: Coastal Kadazan and Sugut Kadazan. A Kuala Monsok Dusun reference tape was not included in any test set. Results are mixed, but the average score of all Central Dusun test points on Coastal Kadazan reference tapes is 82%. This indicates that Central Dusun should be considered a separate dialect from Coastal Kadazan.

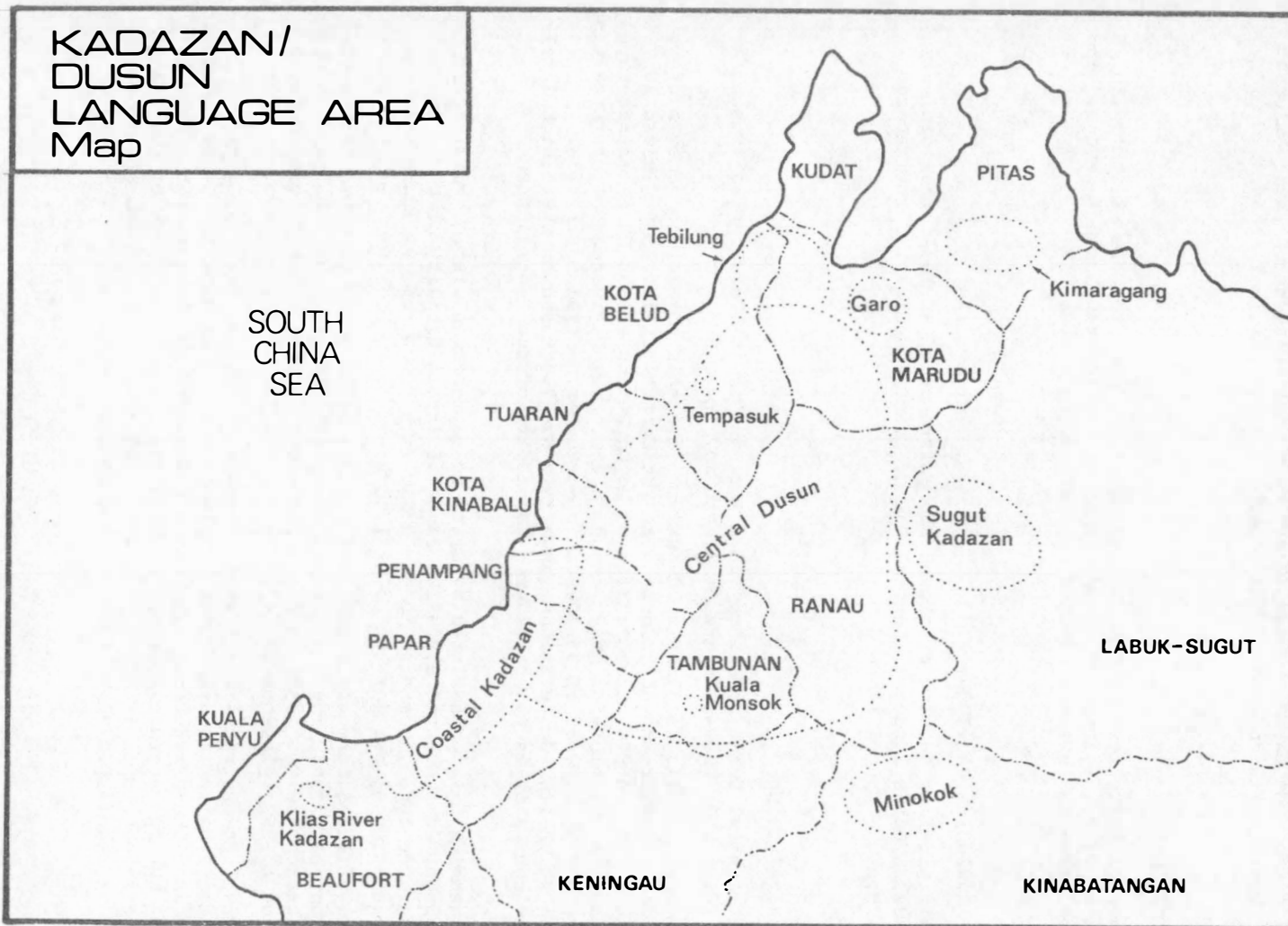


Figure 4: Kadazan/Dusun dialects and closely-related Dusunic languages

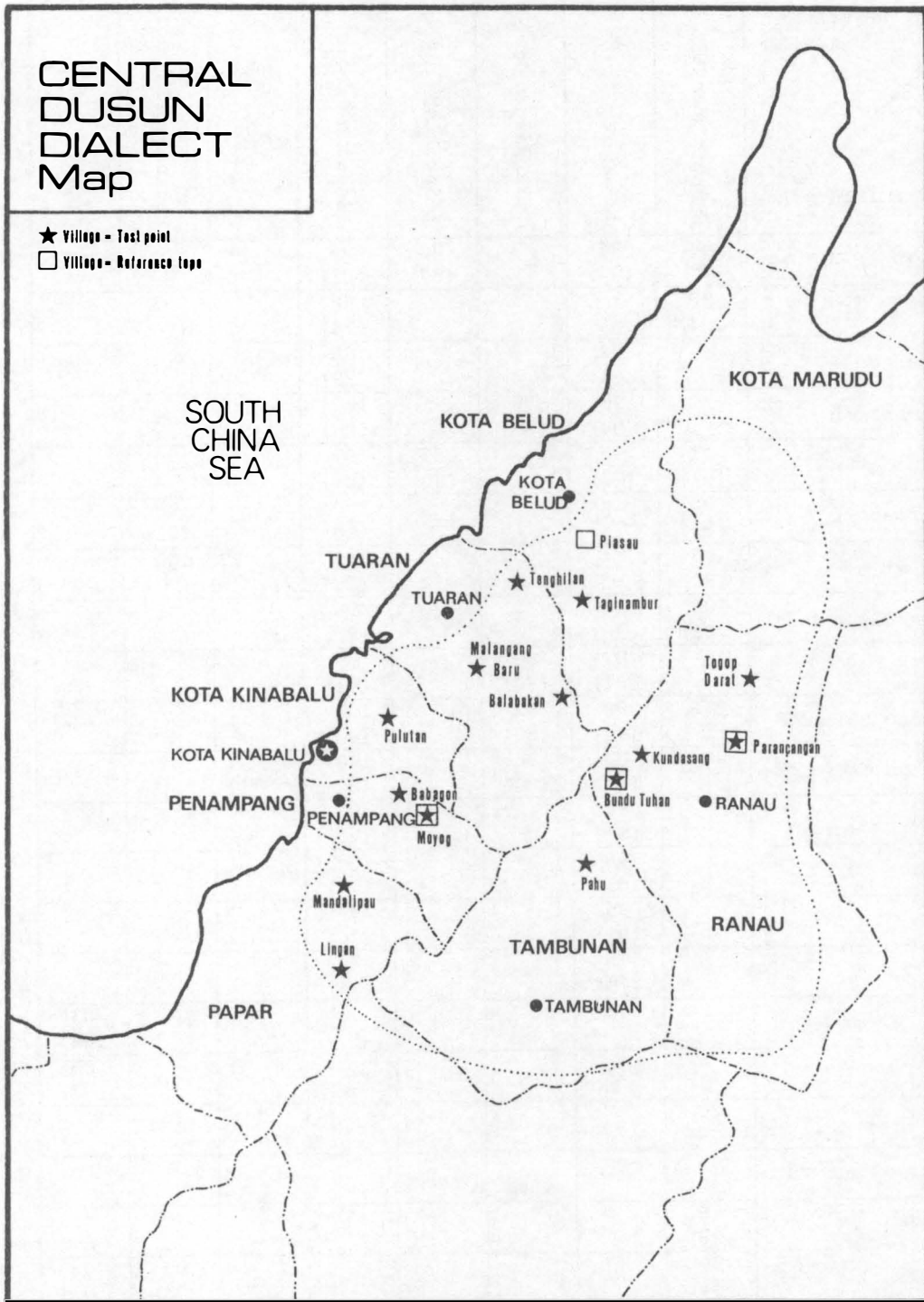


Figure 5: Central Dusun dialect

REFERENCE TAPES TEST POINTS	Bundu Tuhan RU	Moyog PG	Parancangan RU	Nalapak RU	Piasau KB	Taginambur KB	Paginatan RU	Panaitan KM
Bundu Tuhan RU	--	94 (84)		93 (87)	94 (82)			
Moyog PG	98 (84)				98 (79)		95 (76)	
Parancangan RU		92 (81)			88 (79)			
Taginambur KB		95 (84)	96 (83)		100 (85)			
Togop Darat RU	97 (91)	98 (83)		86 (88)	97 (81)			
Mandalipau PR	98 (82)				100 (80)		98 (75)	
Lingan PR	98 (83)				98 (79)		100 (77)	
Babagon/Kibunut PG	100 (82)				99 (80)		96 (76)	
Tenghilan TN		97 (81)	89 (76)		99 (81)			
Melangang Baru TN		92 (91)	82 (80)		100 (80)			
Balabakan TN		88 (84)	86 (80)		100 (80)			
Pulutan KK		97 (82)			95 (75)			
Kundasang RU		98 (81)			100 (77)			
Pahu TA		95 (81)			94 (76)			
Paginatan RU		95 (76)			87 (74)		--	
Tangkol KM								83 (85)
Taiwan KM								85 (87)
Tempasuk KB	100 (75)					96 (76)	96 (69)	
Tambulian Laut KB	99 (82)						96 (73)	
Dudar KB	98 (78)						98 (76)	
Mandap KB	99 (84)						93 (74)	

Figure 6: Intelligibility testing results at 21 Central Dusun villages. (Results are given as percentages; PSC figures are in parentheses.)

REFERENCE TAPES TEST POINTS	COASTAL KADAZAN			SUGUT KADAZAN		
	Bunduon PG	Sabandil PR	Pinopok PR	Basai LS (1)	Basai LS (2)	Moiwod KN
Bundu Tuhan RU	66 (75)	72 (76)				
Moyog PG	100 (80)	87 (77)				
Parancangan RU	56 (75)		67 (74)			
Taginambur KB	88 (76)	80 (78)		90 (79)	50 (79)	
Togop Darat RU	61 (77)	71 (76)				
Mandalipau PR	89 (80)	99 (83)				
Lingan PR	91 (81)	99 (85)				
Babagon/Kibunut PG	96 (81)	86 (78)				
Tenghilan TN	87 (75)	77 (76)				
Melangang Baru TN	93 (77)	74 (78)				
Balabakan TN	85 (76)	75 (76)				
Pulutan KK		86 (79)	89 (76)	100 (78)		
Kundasang RU	73 (72)		100 (73)			
Pahu TA	75 (74)		70 (74)			
Paginatan RU	76 (73)		89 (70)			
*(DN) Entilibon KN						79 (78)
Tempasuk KB	78 (68)					
Tambulian Laut KB	83 (73)		93 (74)			
Dudar KB	78 (76)		81 (74)			
Mandap KB	77 (73)		77 (75)			

Figure 7: Intelligibility testing results at 20 Central Dusun villages. (Results are given as percentages; PSC figures are in parentheses; * represents 'Dusun' speakers from a mixed-language village.)

The high scores on the first Basai LS tape can be explained by the simplicity of the story. The second Basai LS reference tape and the Moiwod KN tape may be much better indicators of Central Dusun's understanding of Sugut Kadazan. Sugut Kadazan should perhaps be considered not merely a separate dialect but a separate language from Central Dusun. But until further testing can clarify the issue, a dialect-level distinction will be maintained between Central Dusun and Sugut Kadazan.

Figure 8 displays Central Dusun understanding of other languages formerly classified as Kadazan/Dusun (Smith, in this volume). All intelligibility test scores are below the 80% same-language threshold and so the languages are still considered distinct from Kadazan/Dusun.

REFERENCE TAPES TEST POINTS	KIMARAGANG	GARO	TEBILUNG
	Dandun PS	Talantang KM	Mengaris KM
Panaitan KM	47 (75)	63 (81)	72 (86)
Tangkol KM	59 (72)	57 (79)	59 (81)
Taiwan KM	69 (74)	73 (80)	65 (85)
Lotong KM	63 (75)	70 (79)	
Tempasuk KB	52 (71)		
Tambulian Laut KB	69 (70)		
Dudar KB	73 (80)		
Mandap KB	51 (69)		

Figure 8: Intelligibility testing results at eight Central Dusun villages. (Results are given as percentages; PSC figures are in parentheses.)

3.1.2 Coastal Kadazan

Intelligibility testing indicated that several villages which Smith (in this volume) had classified as representing separate dialects should be re-grouped as a Coastal Kadazan dialect. Bunduon PG, Sabandil PR, and Terawi PG had been classified as Central Dusun; Pinopok BT, as Beaufort Kadazan; and Tombovo/Potuki PG, as a third distinct dialect (Figure 9). A large group of Kadazan villages in Penampang, Papar and Beaufort districts have distinct phonemic differences from Central Dusun, along with some differences in pronouns (Pekkanen 1981).

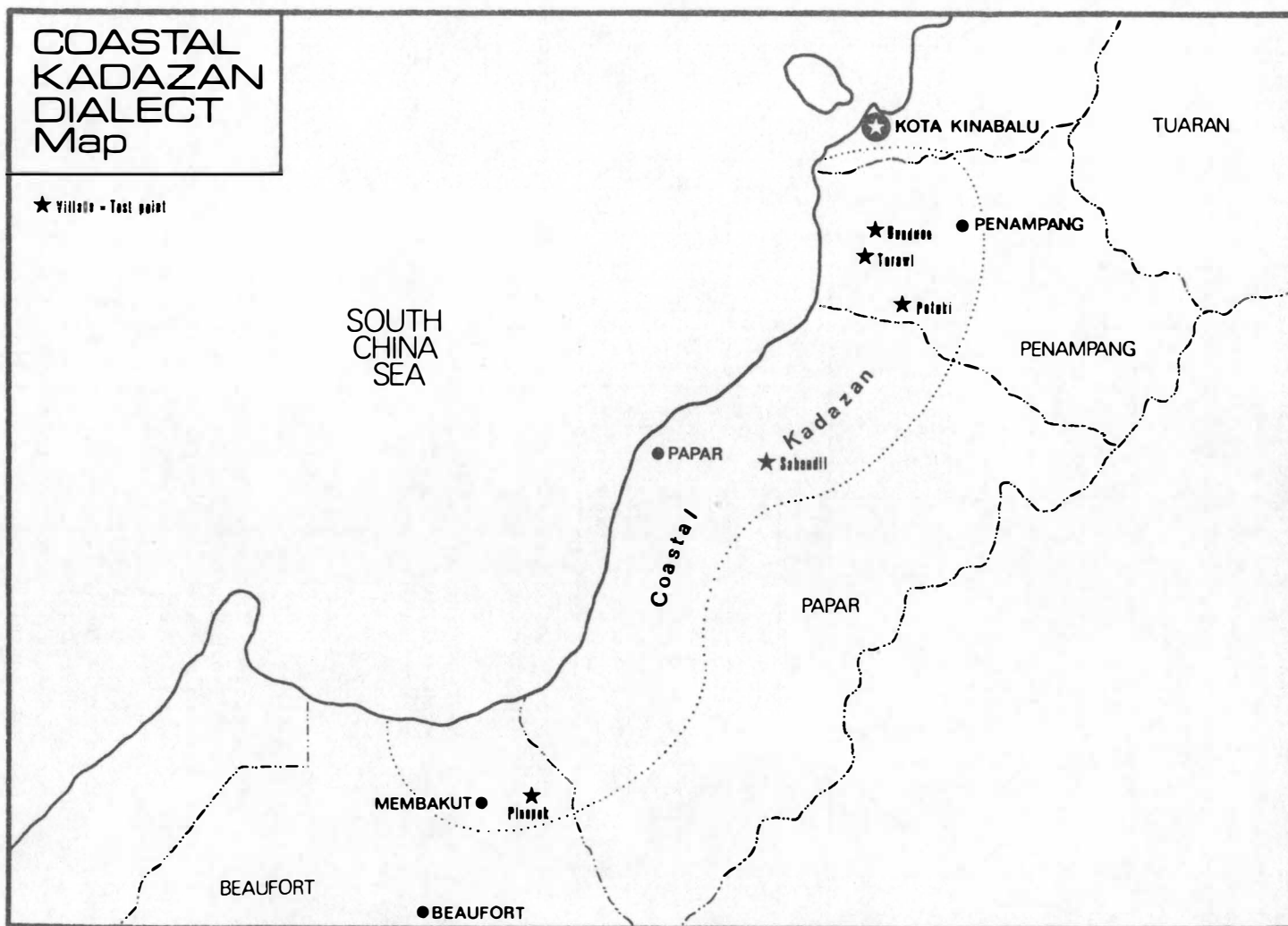


Figure 9: Coastal Kadazan dialect of the Kadazan/Dusun language

Figure 10 displays the results of intelligibility testing at five Coastal Kadazan villages. Coastal Kadazan intelligibility of other Coastal Kadazan reference tapes was at or above the 85% same-dialect threshold in every case. Coastal Kadazan understanding of Central Dusun was also quite high, considering the lower PSC relations. However, Central Dusun understanding of Coastal Kadazan was not as high (cf. Figure 7), indicating that intelligibility between the two dialects is not mutual.

REFERENCE TAPES TEST POINTS	KADAZAN/DUSUN							SUGUT	KLIAS RIVER KADAZAN
	COASTAL			CENTRAL					
	Bunduon PG	Sabandil PR	Pinopok BT	Bundu Tuhan RU	Moyog PG	Piasau KB	Taginambur KB	Basai LS (1)	Takuli BT
Bunduon PG	--	96 (82)	100 (78)	100 (75)		98 (74)		100 (74)	
Sabandil PR		--	99 (84)		93 (77)	91 (75)		95 (74)	
Terawi PG	97 (93)	87 (83)		82 (76)	93 (80)	83 (74)			
Tombovo/Potuki PG	100 (89)	94 (81)		99 (75)	98 (78)	93 (73)			
Pinopok BT	85 (78)	85 (84)	--				85 (77)		80 (85)

Figure 10: Intelligibility testing results at five Coastal Kadazan villages. (Results are given as percentages; PSC figures are in parentheses.)

Coastal Kadazan understanding of the simple Basai LS tape, representing the Sugut Kadazan dialect, was very high. Pinopok BT intelligibility of the geographically close Klias River Kadazan language of Takuli BT was 80%, lower than their PSC relation of 85.

3.1.3 Sugut Kadazan dialect

The Sugut Kadazan dialect of the Kadazan/Dusun language corresponds to Smith's (in this volume) Inner Sugut subdialect. Figure 11 displays the results of intelligibility testing at three Sugut Kadazan villages: Kaingaran LS, Moiwod KN and Entilibon KN. The latter two villages employ the ethnonym 'Minokok' and are geographically distant from Kaingaran LS (Figure 12). Entilibon KN is a mixed-language village, with speakers of both Central Dusun and 'Minokok' Sugut Kadazan living there. The high score obtained in that village is the only example of high intelligibility of Central Dusun found among the Sugut Kadazan villages tested. At all other test points, Sugut Kadazan understanding of Central Dusun and Coastal Kadazan was below the 80% same-language threshold. Central Dusun and Coastal Kadazan test scores on one Sugut

Kadazan story from Basai LS (cf. Figures 7 and 9), however, were very high, ranging 90-100%. Further testing must be done in order to determine Sugut Kadazan's true relationship to Central Dusun and Coastal Kadazan. Until such time, Sugut Kadazan will continue to be classified as a dialect of the Kadazan/Dusun language, following Smith (in this volume).

REFERENCE TAPES	KADAZAN/DUSUN					KIMARAGANG	GARU	
	SUGUT	CENTRAL			COASTAL			
	Basai LS (1)	Taginambur KB	DN*Entilibon KN	Pahu TA	Bunduon PG	Dandun PS	Parong KM	Talantang KM
TEST POINTS								
Kaingaran LS		72 (79)			58 (74)	69 (81)	85 (85)	80 (83)
Moiwod KN	96 (81)		50 (78)	73 (79)	47 (75)			
Entilibon KN			98 (79)					

Figure 11: Intelligibility testing results at three Sugut Kadazan villages.
(Results are given as percentages; PSC figures are in parentheses.)

Further testing should also be done among the Sugut Kadazan villages to determine the degree of intelligibility between Sugut Kadazan as spoken in Kaingaran LS and as spoken in the 'Minokok' villages of Moiwod KN and Entilibon KN.

Test subjects from Kaingaran LS scored 80% and 85% on Garu reference tapes from Talantang KM and Parong KM, classified as "Outer Sugut" by Smith (in this volume). Test subjects at Talantang KM, however, scored only 74% on a reference tape from Moiwod KN (cf. Figure 21). Test results are inconclusive thus far in determining the relationship between Sugut Kadazan and Garu, since there appears to be only a dialect-level difference between Talantang KM and Kaingaran LS, but a language-level difference between Talantang KM and Moiwod KN.

The 69% intelligibility scored on the Kimaragang reference tape from Dandun PS (Figure 11) does indicate clearly that the Sugut Kadazan of Kaingaran LS is a different language from Kimaragang.

3.1.4 Kuala Monsok Dusun dialect

Intelligibility testing at Kuala Monsok TA confirms Smith's (in this volume) lexicostatistical classification of Kuala Monsok Dusun as a distinct dialect of the Kadazan/Dusun language. Kuala Monsok TA test subjects scored 80% and 77% on reference tapes from the Central Dusun villages of Moyog PG and Piasau KB (Figure 13). On Coastal Kadazan tapes from Bunduon PG and Pinopok BT, however,

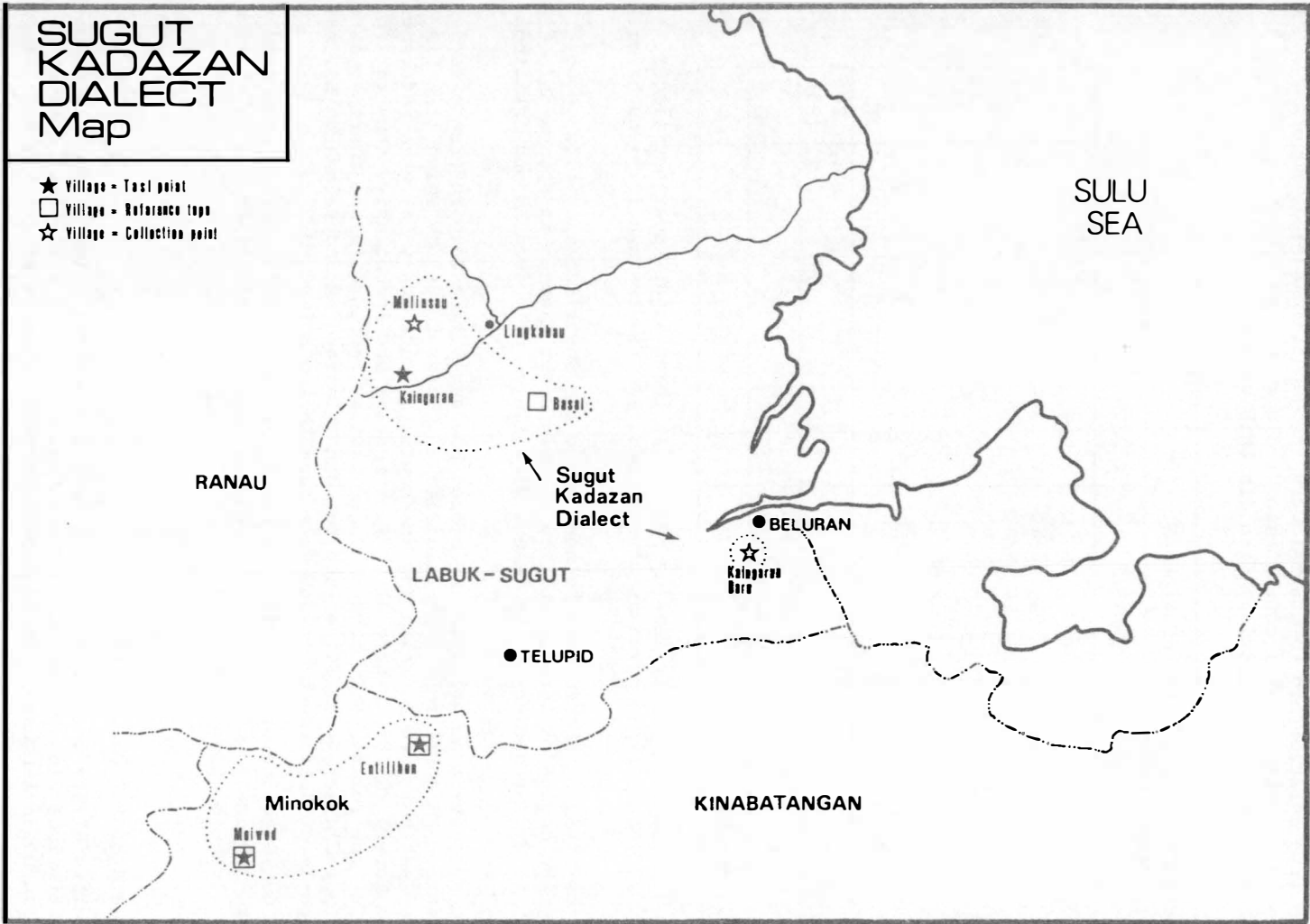


Figure 12: Sugut Kadazan dialect of the Kadazan/Dusun language

Kuala Monsok Dusun speakers scored only 61% and 66%. No cross-testing of a Kuala Monsok reference tape was done.

REFERENCE TAPES TEST POINT	KADAZAN/DUSUN				KUIJAU
	CENTRAL		COASTAL		
	Moyog PG	Piasau KB	Bunduon PG	Pinopok BT	Liau Darat KU
Kuala Monsok TA	80 (79)	77 (72)	61 (75)	66 (73)	62 (82)

Figure 13: Intelligibility testing results at Kuala Monsok TA. (Results are given as percentages; PSC figures are in parentheses.)

Kuala Monsok Dusun has a relationship of 82 PSC with the Kuijau language of Liau Darat KU - higher than its cognate relationship to any of the four Kadazan/Dusun reference tapes with which it was tested. The low 62% intelligibility of the Liau Darat KU reference tape, however, indicates that, despite their higher PSC relation, Kuijau and Kuala Monsok Dusun are different languages.

3.1.5 Kadazan/Dusun language in relation to other Dusunic languages

Reference tapes from five other Dusunic languages were played in six Kadazan/Dusun villages representing the Central Dusun dialect (Figure 14). Results are displayed in Figure 15.

Kadazan/Dusun speakers scored only 19-50% intelligibility of a Rungus language tape from Lodung KT. Similar low scores were repeated when Rungus speakers were tested for their understanding of taped stories from the Central and Coastal dialects of Kadazan/Dusun (Figure 16). Intelligibility testing thus confirms that Rungus and Kadazan/Dusun are separate languages.

The relationship of Kadazan/Dusun and Lotud is not as clear-cut. The Central Dusun villages of KogoPON PR, Kionsom KK and Lumpoho/Pukak TN scored 81-89% intelligibility of the Lotud reference tape from Marabhai TN (Figure 15). These scores are above the same-language threshold, 80%. Toboh TA, Tangaban TA and Sunsuron TA also representing Central Dusun, scored 52-77%, below the same-language threshold. Lotud intelligibility of Central Dusun and Coastal Kadazan reference tapes was 80% or higher in each of the three Lotud villages tested (Figure 16). However, since the average level of intelligibility of Lotud demonstrated by the six Kadazan/Dusun villages tested was only 73%, Lotud is still considered a separate language.

The status of Kuijau, Eastern Kadazan and Tatana as separate languages from Kadazan/Dusun is clearly affirmed by intelligibility testing results (Figures 15 and 16).



Figure 14: Central Dusun dialect area showing villages where other Dusunic languages were tested

Central Dusun intelligibility of a Kuijau story from Tuntumulod KU ranged only 57-75%. Although four of the five Kuijau villages tested scored 82-96% on the Central Dusun reference tape from Bundu Tuhan RU, Tuntumulod KU test subjects scored only 62%. On the Coastal Kadazan tape from Bunduon PG, all Kuijau villages scored significantly lower than 80%, indicating that Kuijau and Coastal Kadazan are different languages.

Central Dusun understanding of Eastern Kadazan ranged 51-71%. In cross-testing, Eastern Kadazan understanding of both Central Dusun and Coastal Kadazan ranged 34-56%. Such low scores confirm that Kadazan/Dusun and Eastern Kadazan are separate languages

Intelligibility test results were similarly low between the Kadazan/Dusun and Tatana villages tested. Central Dusun scores on a Tatana story from Bundu KP ranged 36-58%, and Tatana scores on the two dialects of Kadazan/Dusun tested ranged 55-67%. Kadazan/Dusun and Tatana are clearly demonstrated to be separate languages.

REFERENCE TAPES		TEST POINTS		RUNGUS Lodung KT	LOTUD Marabahai TN	KUIJAU Tuntumulod KU	EASTERN KADAZAN Buis LS	TATANA Bundu KP
CENTRAL DUSUN	Toboh TA	19 (70)	77 (69)	62 (74)	52 (71)	42 (64)		
	Tangaban TA	40 (73)	52 (70)	71 (76)	51 (73)	41 (66)		
	Sunsuron TA	39 (71)	56 (72)	67 (75)	51 (71)	37 (67)		
	Kogopon PR	48 (74)	89 (70)	75 (77)	62 (75)	58 (73)		
	Kionsom KK	50 (70)	81 (70)	64 (73)	71 (70)	36 (64)		
	Lumpoho/Pukak TN	45 (72)	89 (71)	57 (73)	61 (70)	40 (65)		

Figure 15: Intelligibility testing results at six Central Dusun villages with reference tapes from other Dusunic languages. (Results are given as percentages; PSC figures are in parentheses.)

REFERENCE TAPES		CENTRAL DUSUN					COASTAL KADAZAN
		Pahu TA	Bundu Tuhan RU	Melangang Baru TN	Tenghilan TN	Taginambur KB	Bunduon PG
TEST POINTS							
RUNGUS	Kimihang KT	39 (71)					45 (72)
	Buanog KT	44 (--)					43 (--)
LOTUD	Marabahai TN		95 (71)		90 (67)		86 (70)
	Kauluan TN		98 (75)		97 (75)		89 (75)
	Sumagit KK		98 (70)	80 (70)			92 (71)
KUIJAU	Tuntumulod KU		62 (73)				65 (74)
	Sungoi KU		92 (71)				70 (73)
	Liau Darat KU		82 (74)				68 (74)
	Senagang KU		89 (72)				70 (72)
	Linsosod KU		96 (69)				72 (68)
EASTERN KADAZAN	Mananam KN	56 (72)					46 (71)
	Batu Putih KN	49 (60)					44 (61)
	Malapi LS					35 (70)	37 (68)
	Buis LS				34 (70)		41 (70)
TATANA	Bundu KP					59 (65)	67 (68)
	Kiaru KP					56 (--)	55 (--)

Figure 16: Intelligibility testing results at selected Rungus, Lotud, Kuijau, Eastern Kadazan and Tatana villages with Central Dusun and Coastal Kadazan reference tapes. (Results are given as percentages; PSC figures are in parentheses.)

3.2 Klias River Kadazan language

Intelligibility testing at Takuli BT reveals surprisingly different results from those indicated by lexicostatistical findings. Takuli BT is classified as part of the Beaufort Kadazan dialect of Kadazan/Dusun by Smith (in this volume). However, intelligibility testing results seem to indicate that Smith's Beaufort Kadazan dialect should be considered part of the Coastal Kadazan dialect. But when Takuli BT was tested for its understanding of three Coastal Kadazan and one Central Dusun reference tapes (Figure 17), intelligibility scores were too low to consider Takuli BT part of the Kadazan/Dusun language. Cross-testing a Takuli BT tape at the Coastal Kadazan village of Pinopok BT (Figure 10) yields an intelligibility score of 80%, lower than the corresponding PSC figure of 85. Though this score in isolation would indicate that Takuli BT and Pinopok BT are different dialects of the same language, the low intelligibility of other Coastal Kadazan tapes at Takuli BT warrants classifying the Klias River Kadazan spoken at Takuli BT and neighbouring Kinamam BT (Figure 18) as a separate language from Kadazan/Dusun.

REFERENCE TAPES TEST POINT	KADAZAN/DUSUN			
	COASTAL			CENTRAL
	Bunduon PG	Inobong PG	Sabandil PR	Taginambur KB
Takuli BT	53 (78)	46 (79)	71 (83)	64 (77)

Figure 17: Intelligibility testing results at Takuli BT. (Results are given as percentages; PSC figures are in parentheses.)

The Takuli BT tape was also tested in two Tatana villages, Bundu KP and Kiaru KP. Although each of these villages has only 77 PSC with Takuli BT, Bundu KP scored 92% and Kiaru KP 91% on the Takuli BT story. Further testing needs to be done in Takuli BT and/or Kinamam BT to see how well these villages understand Tatana, and to establish the relationship between them and Tatana (→ Tatana).

3.3 Kimaragang language

Smith (in this volume) included Dandun PS as part of his "Outer Sugut sub-dialect" of the Kadazan/Dusun language, along with the villages of Parong KM and Talantang KM (Figure 19). Intelligibility testing indicates that the people of Dandun PS speak Kimaragang, a different language from the Garo spoken in Parong KM and Talantang KM. Test subjects in Dandun PS scored only 76% intelligibility of a reference tape from Talantang KM (Figure 20). Cross-testing of a Dandun PS reference tape at Talantang KM yielded an even lower score, 71% (Figure 21). Clearly, Dandun PS and Talantang KM represent separate languages. A Tebilung tape from Mengaris KM and a Rungus tape from Tinangol KT were also tested at

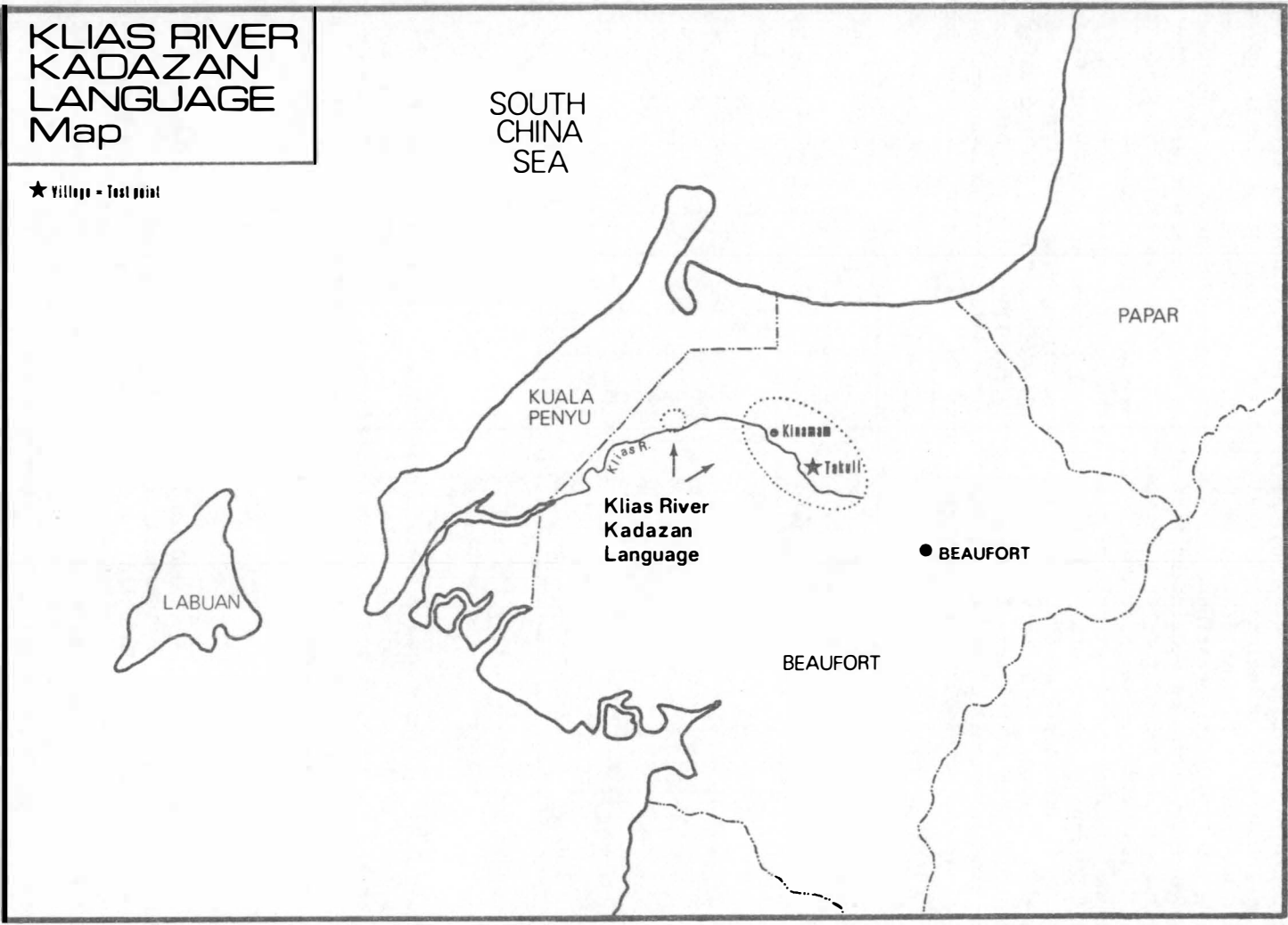


Figure 18: Klias River Kadazan language map

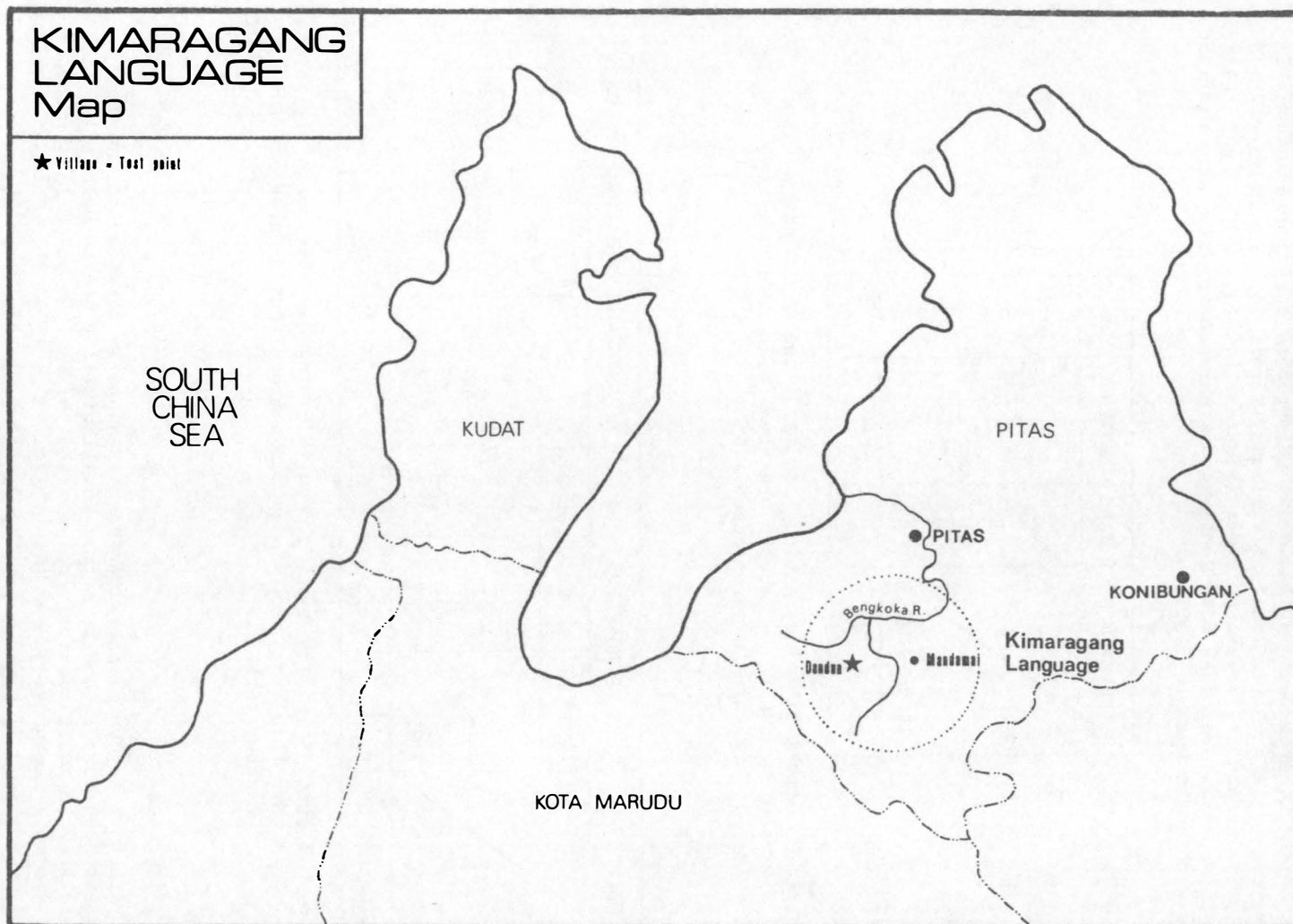


Figure 19: Kimaragang language area

Dandun PS. Scores on both tapes were below 80%, indicating that they too, are different languages from Kimaragang. The Kimaragang reference tape was tested at eight Central Dusun villages (cf. Figure 8), one Sugut Kadazan village (cf. Figure 11) and two Tebilung villages (Figure 24), in addition the Garo village of Talantang KM, mentioned above (Figure 21). In every case, intelligibility scores were lower than 80%, confirming Kimaragang's place as a distinct language within the Dusunic family.

REFERENCE TAPES TEST POINT	GARO	TEBILUNG	RUNGUS
	Talantang KM	Mengaris KM	Tinangol KT
Dandun PS	76 (80)	67 (78)	71 (79)

Figure 20: Intelligibility testing results at the Kimaragang village of Dandun PS. (Results are given as percentages; PSC figures are in parentheses.)

REFERENCE TAPES TEST POINT	KADAZAN/DUSUN			KIMARAGANG	TEBILUNG	RUNGUS
	CENTRAL	COASTAL	SUGUT	Dandun PS	Mengaris KM	Tinangol KT
	Pahu TA	Bunduon PG	Moiwod KN			
Talantang KM	61 (81)	64 (76)	74 (82)	71 (80)	70 (84)	68 (78)

Figure 21: Intelligibility testing results at the Garo village of Talantang KM. (Results are given as percentages; PSC figures are in parentheses.)

3.4 Garo language

The distinct nature of the Kimaragang language in relation to the Garo language has been established by mutual intelligibility scores below 80% (Section 3.3). Earlier, lexicostatistical analysis (Smith, in this volume) had grouped these as a single language.

Six reference tapes representing Central Dusun, Coastal Kadazan, Sugut Kadazan, Kimaragang, Tebilung and Rungus were tested at the Garo village

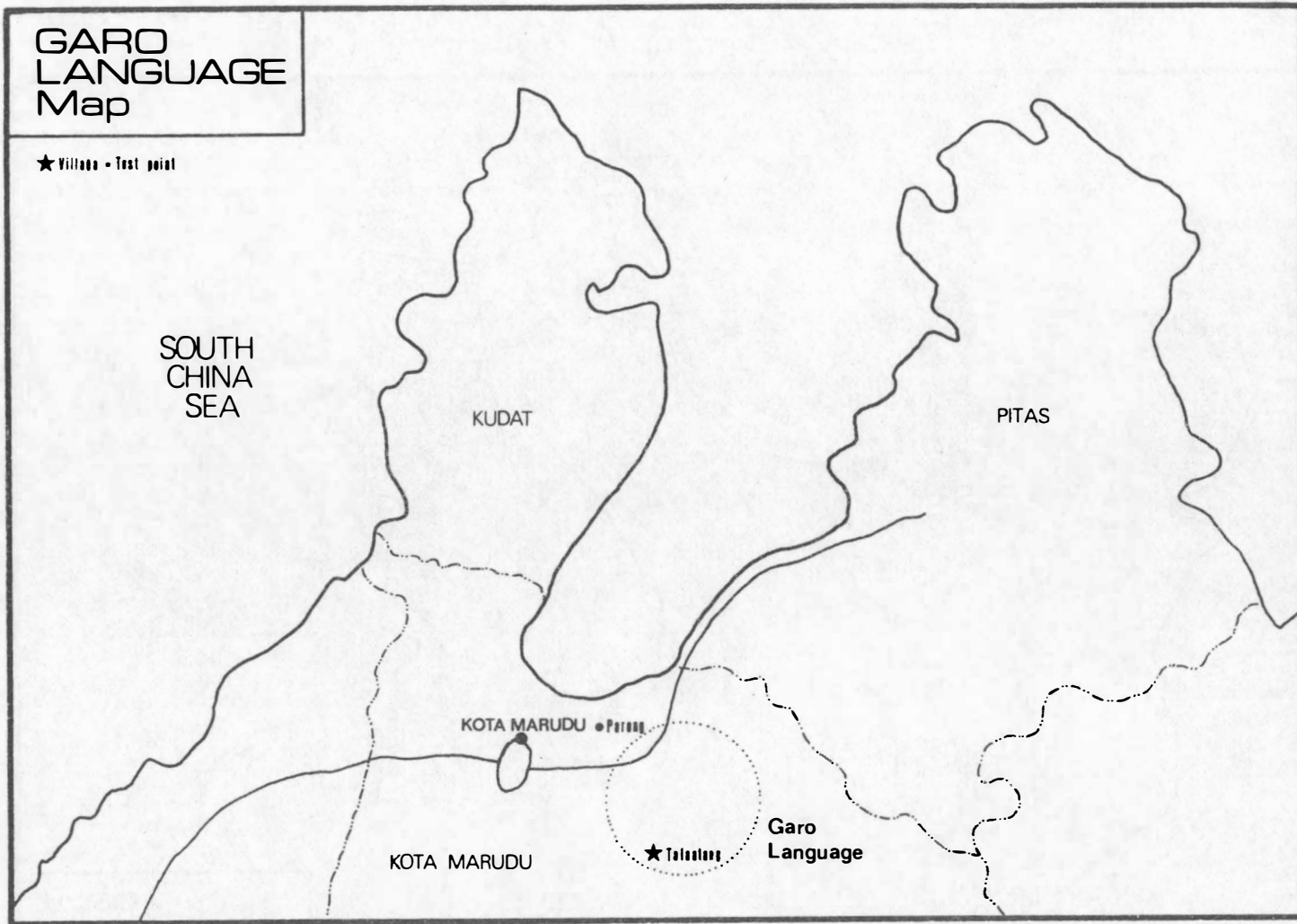


Figure 22: Garo language area

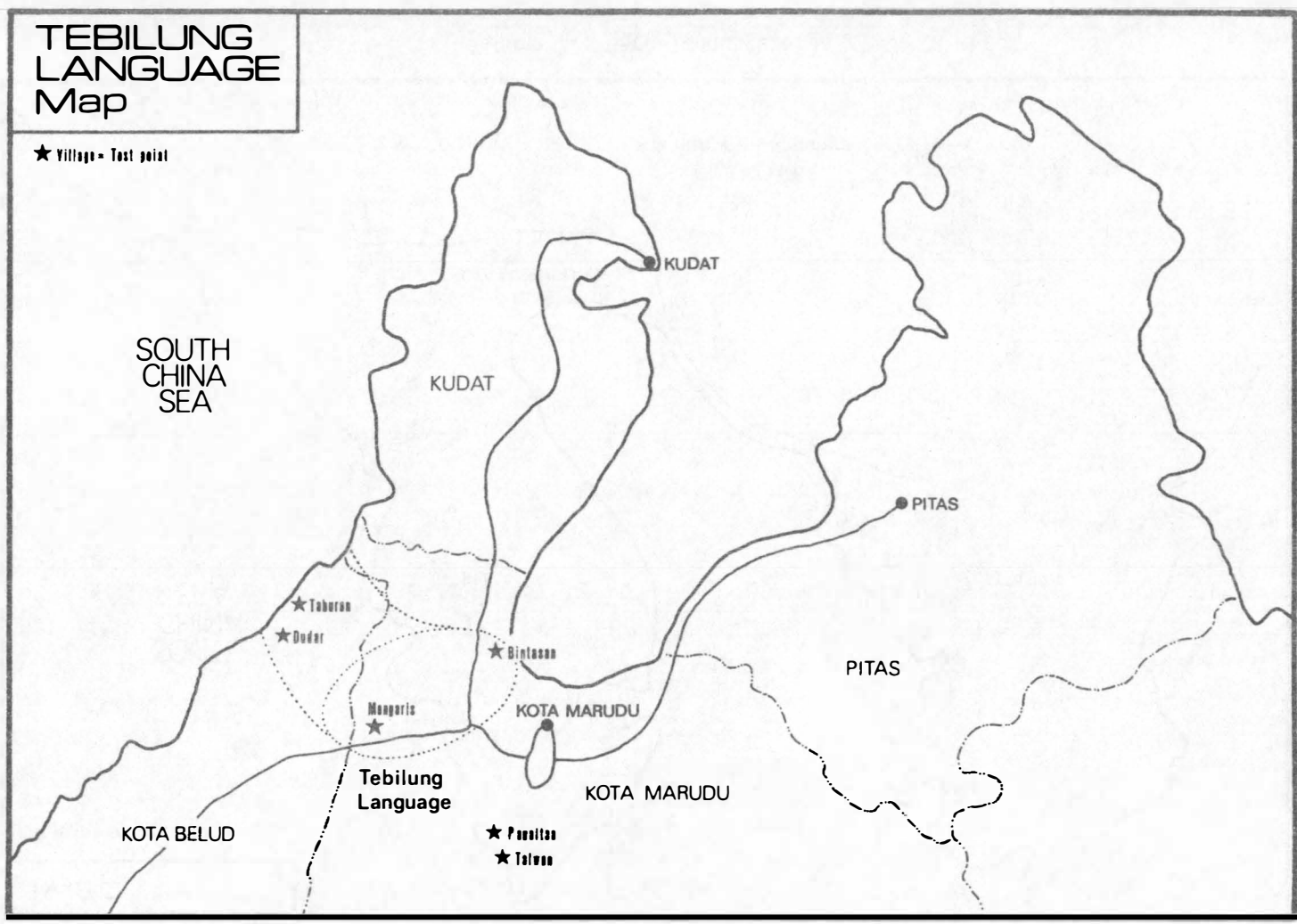


Figure 23: Tebilung language area

of Talantang KM (Figures 21 and 22). In every case, intelligibility scores were below the 80% level established as the same-language threshold. When a Garo reference tape was cross-tested at four Central Dusun villages (cf. Figure 8), one Kimaragang village (cf. Figure 20) and two Tebilung villages (Figure 24), results were similarly low. However, when Sugut Kadazan speakers (Figure 11) were tested on their ability to understand Garo reference tapes from Parong KM and Talantang KM, scores were 85% and 80%. Further testing is needed to clarify the relationship between Garo and the Sugut Kadazan dialect of Kadazan/Dusun.

The position of Parong KM as a member of the Garo language needs verification. It is possible that Parong KM relates more closely to Dandun PS (Kimaragang) or forms a link in a dialect/language chain between Sugut Kadazan, Garo and Kimaragang.

3.5 Tebilung language

Reference tapes representing Central Dusun, Kimaragang and Garo were tested at the Tebilung villages of Bintasan KM and Mengaris KM (Figure 23), previously classified as Kota Belud Dusun (Smith, in this volume). In addition, a Mengaris KM tape was also tested in Bintasan KM. Results of the Central Dusun, Kimaragang and Garo tapes were lower than 80% in every case, indicating that they are indeed separate languages from Tebilung (Figure 24). Cross-testing a Tebilung reference tape in three Central Dusun villages (cf. Figure 8), one Kimaragang village (cf. Figure 20) and one Garo village (cf. Figure 21) yielded similar findings and confirms the assertion that Tebilung is a separate language.

The Tebilung village of Bintasan KM scored only 82% intelligibility of the tape from Mengaris KM, compared to a PSC relation of 92. Further testing is needed to determine the number of dialects within the Tebilung language.

REFERENCE TAPES	TEBILUNG	CENTRAL DUSUN	KIMARAGANG	GARO
TEST POINTS	Mengaris KM	Panaitan KM	Dandun PS	Talantang KM
Bintasan KM	82 (92)	65 (83)	68 (75)	67 (83)
Mengaris KM	--	70 (86)	63 (78)	70 (84)

Figure 24: Intelligibility testing results at two Tebilung villages.
(Results are given as percentages; PSC figures are in parentheses.)

4. NATIONAL LANGUAGE INTELLIGIBILITY

National language intelligibility testing was carried out in 43 Kadazan/-Dusun villages in 12 of Sabah's districts. Results of that testing are displayed in Figure 25. This testing included villages located in developing areas as well as areas already well-developed.

DISTRICT	SEX M/F	AGE RANGE	EDUCATION		AVERAGE INTELLIGIBILITY PERCENTAGE AND AVERAGE AGE		
			ED.	A.E.R.	Educated	Uneducated	All
Beaufort	10/9	16-63	13	5	94 (32)	58 (43)	83 (36)
Kinabatangan	12/8	18-53	10	5	72 (24)	58 (35)	65 (30)
Kota Belud	27/21	13-70	41	8	99 (25)	68 (47)	94 (29)
Kota Kinabalu	14/6	16-78	12	8	96 (31)	63 (55)	83 (41)
Kota Marudu	47/21	13-70	27	7	75 (25)	40 (39)	54 (33)
Labuk-Sugut	5/5	17-67	4	5	95 (24)	53 (42)	70 (35)
Papar	22/15	15-67	30	6	95 (28)	57 (59)	88 (34)
Penampang	28/13	17-70	26	7	98 (34)	71 (49)	88 (39)
Pitas	5/5	13-52	4	5	80 (27)	58 (43)	67 (36)
Ranau	27/12	14-82	25	7	97 (31)	49 (40)	80 (34)
Tambunan	30/17	15-71	25	6	83 (31)	38 (46)	62 (38)
Tuaran	18/14	19-63	25	8	98 (30)	84 (47)	95 (34)

Figure 25: Kadazan/Dusun understanding of the national language tape by district. (Under Education ED. = the number of persons who had received some formal education; A.E.R. = the average number of years of education those subjects had received; numbers in parentheses represent the average age of each specified group.)

Thirty-six of the villages were tested with a story that represented standard Bahasa Malaysia as spoken in Sabah. The other seven villages were tested earlier with a story given by a native speaker of Bahasa Malaysia as spoken in West Malaysia. Therefore, results from both stories are considered valid. Three hundred and ninety-one persons were tested. Their comprehension of the Bahasa Malaysia story ranged 0-100%. The average intelligibility was 77%.

Two factors were found to have a significant effect on test results. One was the educational background of the test subjects and the other was the amount of development in the area.⁴ Figure 26 displays the results of the national language intelligibility testing divided according to the educational level of the test subjects and whether the test subjects were from developing areas or areas considered to be well-developed.

LEVEL OF EDUCATION	DEVELOPING AREAS		WELL-DEVELOPED AREAS	
	AVG. SCORE	NO. TESTED	AVG. SCORE	NO. TESTED
Adult Education (3 mths.-3 years)	--	--	88	7
Primary 1	58	2	70	1
Primary 2	68	3	88	7
Primary 3	63	4	91	10
Primary 4	85	3	100	7
Primary 5	80	6	97	7
Primary 6	74	28	96	54
Form 1 and above	89	20	100	83

Figure 26: Comparison of national language intelligibility test results of educated test subjects in developing areas and in well-developed areas

Two hundred and forty-two test subjects had received some formal education. The educational level ranged from three months of adult education to graduation from Mara Institute of Technology.

The Bahasa Malaysia scores of the group with some formal education ranged 30-100%, with the 30% being scored by a Primary 6 student from an isolated village in Kota Marudu District. Within this group, the intelligibility of Bahasa Malaysia of those living in developing areas was significantly lower than the test results of those with a comparable amount of education living in well-developed areas.

5. CONCLUSION

In comparing the intelligibility testing results with the lexicostatistic classification of Smith (in this volume) the following dialects of the Kadazan/Dusun language are indicated:

1. Central Dusun
2. Coastal Kadazan
3. Sugut Kadazan
4. Kuala Monsok (TA) Dusun

The following languages separate from Kadazan/Dusun (and from any other language) are postulated:

1. Klias River Kadazan
2. Kimaragang
3. Garo
4. Tebilung

There are valid reasons lexicostatistically and from native reaction for considering Tempasuk (KB) Dusun a separate dialect of Kadazan/Dusun, but so far this has not been demonstrated by intelligibility testing.

The position of Minokok (as represented by Moiwod KN) has not yet been verified as to whether it is a part of the Sugut Kadazan dialect or distinct from it. Further testing is also necessary to determine whether Sugut Kadazan should be considered a dialect of Kadazan/Dusun or a distinct language.

The intelligibility testing between Kadazan/Dusun and the Rungus, Lotud, Kuijau, Eastern Kadazan and Tatana languages verifies the findings of lexico-statistics, viz. that Kadazan/Dusun is a distinct language from any of these.

NOTES

1. Where it is appropriate and helpful, the various ethnonyms applied to this language group will be presented in the discussion of intelligibility testing (Section 3). Dunn (1980) has an extensive listing of ethnonyms in her article. LeBar (1972), Appell (1968), Clayre (1966), Prentice (1970), and others have also done classifications of the languages and peoples under discussion here, giving the various ethnonyms applied to this group. Since these classifications tend to vary, even when it appears that the same group or locality is under discussion, the present writers will confine themselves to using Smith (in this volume) as their starting point and will attempt to define designations which are used here.
2. G.N. Appell (1968) has given considerable attention to the matters relating to the use of the term Dusun, tracing the term to nineteenth century sources.
3. This list is not meant to be exhaustive, but rather mentions the works known by the authors to be relevant to the language under consideration here.
4. The 28 villages visited in Beaufort, Kota Belud, Kota Kinabalu, Labuk-Sugut, Papar, Penampang, Ranau and Tuaran were considered well-developed areas. The 15 villages visited in Kinabatangan, Kota Marudu, Pitas, and Tambunan were considered developing areas. Kaingaran LS was included in the first category because it is situated in a well-developed community and the Bahasa Malaysia scores were comparable or better than those of the other villages located in well-developed areas.

CONCLUSION

David C. Moody

1. INTRODUCTION

The preceding papers discuss the results of extensive data gathering and dialect intelligibility testing by the Summer Institute of Linguistics (SIL) to determine the present-day dialect and language relationships in Sabah. This summary article briefly reviews the SIL research and provides an overall perspective to these relationships.

Any scientific study rests upon assumptions held by the researchers whether or not they are explicitly stated. In this study it has been assumed that language is a code of meaning used for communication between at least two parties. One elemental basis of language as a code is that each unit of meaning has a particular phonetic shape or form, or set of forms. The similarity between two lists of these form-meaning composites reflects to some degree the relationship between the languages represented by the lists. It has been assumed that language use assumes intelligibility. The degree of intelligibility between two languages reflects to some degree the relationship between the two languages. A further assumption has been that for any two languages, the measure of their lexical similarity should roughly correlate with the degree of intelligibility between them. Lastly, it has been assumed that in the comparison of any two languages, there may be specific factors which skew the expected correlation between lexical similarity and degree of intelligibility (Pike and Pike 1977).

The starting point for these papers has been the data collected by SIL from 325 villages throughout Sabah and representing 344 speech samples. Smith (in this volume) based his computer-aided lexicostatistical study on this initial collection of data. Additional data has been collected and processed in the analysis of some groupings. As lexicostatistical studies go, Smith's has several significant advantages. The data was collected within a relatively short span of time. It was collected by a small group of technicians working in a common effort. The same basic wordlist was used for all collection points, and the wordlist was relatively long.

Using these 344 wordlists, Smith tentatively established a classification of Sabah's language in which he enumerated 51 languages and 83 dialects distributed into 16 language families and 36 subfamily groupings. Smith anticipated that with additional data, particularly data from other sources (e.g. intelligibility testing, comparative studies, etc.), his arrangement of the linguistic groupings would be improved by altering some of the affiliations he had proposed. Smith suggested that the designations of "language" and "dialect" would be most susceptible to change.

Julie K. King and John Wayne King, eds *Languages of Sabah: a survey report*, 325-337. *Pacific Linguistics*, C-78, 1984.

© David C. Moody

325

Smith assumed that there would be very little ability to communicate between speakers whose dialects represented separate subfamilies. Furthermore, he suggested that the definition of "language" implies the absence of mutual intelligibility with other languages. In the same way, the definition of "dialect" implies the presence of mutual intelligibility with other dialects representing the same language, though exception is made for chaining phenomena.

The relative merit of lexicostatistics as a tool for classifying languages has been much debated.¹ The approach taken by Smith and the authors of the preceding language-specific papers is that lexicostatistics can be used to indicate dialect and language relationships, but that any resultant classification must be considered tentative. The tentativeness of the lexicostatistic classification should not be a deterrent to presenting these findings. In his own defense, Dyen wrote, "The reason that the lexicostatistical classification is inconclusive would be the same that affects any classification; not all the knowable facts are known at the time of classification unless one chooses to wait hopelessly until all knowable facts are known. Based on fewer facts than that a classification remains open to correction as additional facts become available" (Dyen 1966:35). Although such a classification is tentative, it provides a data base and a set of hypotheses which may be used to further examine the relationships between dialect and language groups.

The language-specific papers supplement Smith's lexicostatistical analysis with findings from dialect intelligibility testing. In this testing Smith's analysis served as a starting point. His classification of particular dialects and languages served both as a set of hypotheses to be confirmed or refuted by the evidence provided from intelligibility testing, as well as the basis for determining the overall testing framework, that is, the test sets and test points.

The dialect intelligibility phase of the survey comprised a total of 965 tests conducted in 157 villages. A standard Sabah Malay reference tape was used in all but four villages; thus 812 non-Malay tests were conducted in the 157 villages where testing was done. Among these tests there are 62 reciprocal pairs, or 124 tests, where a reference tape from a test point A was tested in a test point B, and the reference tape from test point B was tested in the test point A.

Like lexicostatistics, dialect intelligibility testing has been variously evaluated as a means of determining the degree of linguistic similarity, or dissimilarity, between languages. It has been shown that intelligibility is a measure of social relations as well as linguistic similarity (Wolff 1964; Simons 1979). The component of linguistic similarity, furthermore, is complex, comprising lexical similarity as well as phonological, grammatical and semantic factors. Where social factors are not present, this component of linguistic similarity is directly related to intelligibility. Thus, in such cases, whereas lexical similarity is one component of linguistic similarity, intelligibility scores represent linguistic similarity directly, being composite measures of the several factors which determine linguistic similarity (Simons 1979:15, 57f., 67-87).

The present study uses intelligibility scores as a check on the language classification obtained by lexicostatistical analysis, the language classification itself being a set of hypotheses concerning the similarity, or dissimilarity, between dialects. The fact that the intelligibility scores do not incorporate any factor for social contact should be kept in mind. Nevertheless, in the majority of cases, the intelligibility scores have confirmed the lexicostatistical hypotheses. There is still a need to complete statistical analyses of the data to determine a more precise correlation between the lexicostatistical data and the intelligibility scores.

2. THE PROPOSED CLASSIFICATION OF THE LANGUAGES OF SABAH

The proposed classification of the languages of Sabah is presented in Figure 1. The revised classification (SIL 1982) is based on both the lexicostatistical evidence as well as the intelligibility data. For the sake of comparison this revised classification is presented alongside Smith's classification (in this volume).² The languages of this proposed classification are located on Figure 2.³

Among the dialects from which wordlists were collected, only those languages which were considered to be either indigenous to Sabah or whose tenure in Sabah has been long term were selected for inclusion in the dialect intelligibility testing phase of the project. On this basis, no testing was conducted in Chabacano, Butung, the five Javanese languages or the two Buginese languages. These languages would likewise have been omitted from testing on the basis of their lexical dissimilarity with the remaining speech samples in the whole study. Generally, testing was not conducted if the cognate percentages between two groups indicated they were in different subfamilies, that is, in the range below 75 percentage of shared cognates (PSC), nor if their cognate relations indicate that they were part of the same subdialect, that is 90 PSC or above. All of the dialect intelligibility testing was conducted in languages of Smith's North-western Austronesian superstock.

Testing within the Lundayeh language, including one sample from a village in Sarawak, confirmed a high degree of homogeneity. External testing of Lundayeh was conducted with a Murutic sample and confirmed the distant relationship which had been established by Smith and others.

Testing within the Illanun language indicated that high mutual intelligibility exists between the two dialects proposed by Smith. External testing with samples from the Philippines indicated a close relationship with some members of the Danao language family.

Testing within the Suluk (Tausug) language confirmed the homogeneity of the samples representing Sabah. External testing with two varieties of East Coast Bajau supported the distinction between these two Philippine-originated groups.

Testing within Smith's Ida'an subfamily did not confirm Smith's division of the samples into two separate languages. On the basis of intelligibility between the samples one Ida'an language is now posited. Dialect-level relationships remain unclear.

Testing within the Malayic family was conducted only in Brunei/Kedayan villages. Testing within the Brunei/Kedayan language indicated a division at the dialect level. External testing supported the division which placed the Iban language in a separate subfamily.

Testing within the Bajau family was conducted in both West Coast and East Coast villages. Testing within the West Coast Bajau language indicated at least two (possibly three) dialects should be distinguished. Testing among the East Coast Bajau villages supported the identification of the Sabah samples with three of the Sama languages distinguished in an earlier Philippine study. The three languages are Balangingi, Southern Sama and Kagayan. Most of the Sabah Bajau samples relate most closely to Southern Sama. Two dialects are proposed for Southern Sama.

The remaining testing was conducted among representatives of Smith's Bornean stock. Smith distinguished four language families within this stock-level grouping: Tidong (represented by a single language), Paitanic, Murutic and Dusunic.

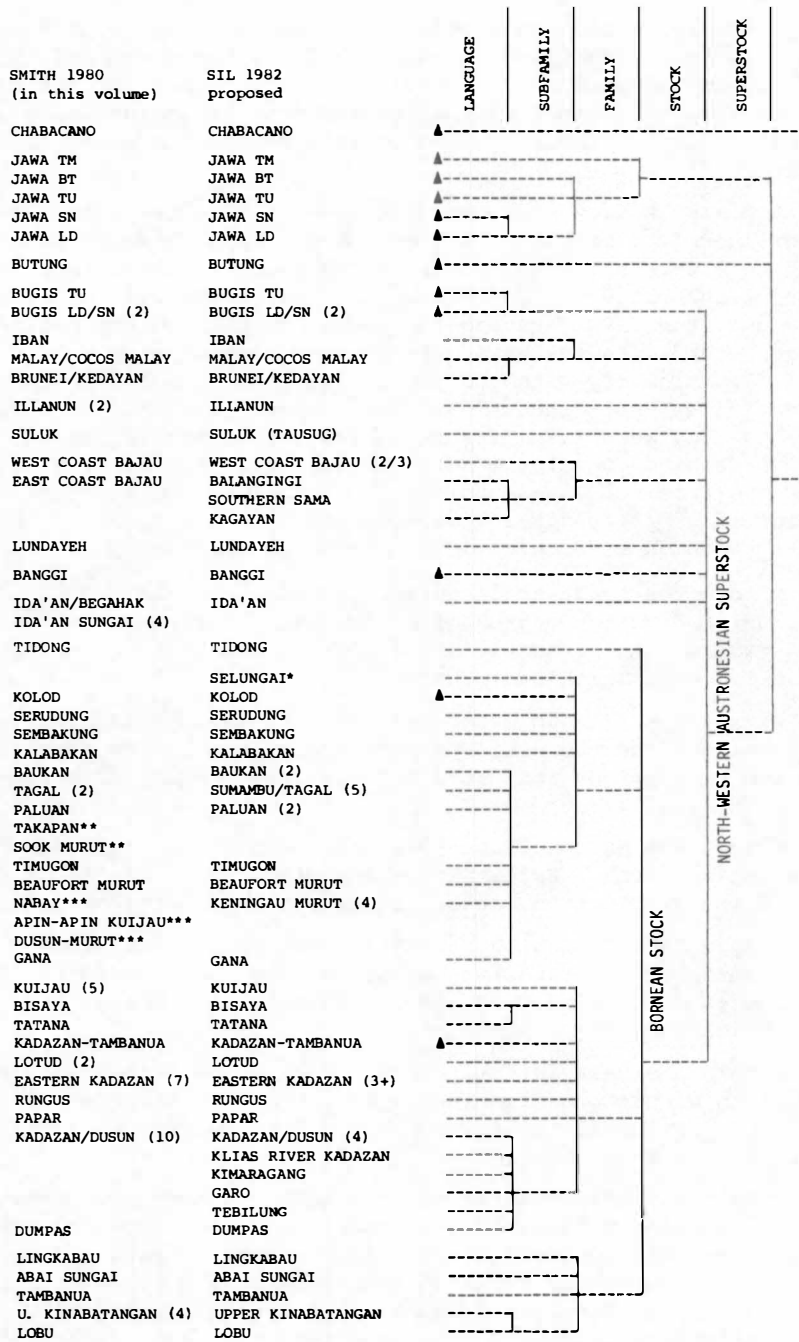


Figure 1: Classifications of the languages of Sabah: 1980 and 1982
 (* classified on the basis of data not available to Smith;
 ** Takapan and Sook Murut have been subsumed as dialects of Paluan in the 1982 classification; *** Nabay, Apin-Apin Kujau and Dusun-Murut have been subsumed as dialects of Keningau Murut in the 1982 classification; (n) numbers in parentheses indicate the number of proposed dialects;
 ▲ indicates a language for which no dialect intelligibility testing was completed.)

Testing within the Tidong language along with additional lexicostatistic data confirmed Smith's single-language classification. A possible dialect-level division was also proposed. External testing with representatives of the Murutic family supported Smith's separation of Tidong from the Murutic languages.

Smith enumerated five Paitanic languages, each relating to the others lexicostatistically at the subfamily level. The five languages were Lingkabau, Lobu, Abai Sungai, Tambanua and Upper Kinabatangan. Lobu remains classified as a separate Paitanic language though testing indicated high Lobu intelligibility of Upper Kinabatangan. Dusun Segama remains classified as a dialect of Upper Kinabatangan, though testing indicated a language distinction may be appropriate.

Smith enumerated 15 Murutic languages which he distributed among eight subfamily divisions. Seven of these subfamily divisions were represented by a single language each. These languages were Kolod, Gana, Apin-Apin Kuijau, Kalabakan Murut, Sembakung Murut, Serudung Murut and Tagal. The eighth subfamily grouping, Central Murut, comprised the following eight languages: Takapan, Paluan, Timugon, Beaufort Murut, Dusun-Murut, Sook Murut, Baukan and Nabay.

Additional lexicostatistical data and analysis and dialect intelligibility testing has produced significant realignment of the Murutic family classification. This revised classification distinguishes 12 languages, which group into the Northern Murut subfamily and an eastern Murutic division.

The Northern Murut subfamily comprises seven languages. The distinctiveness of Gana, Tagal, Paluan, Baukan, Timugon and Beaufort Murut as separate languages is confirmed. Gana has remained distinct on lexicostatistical grounds, although it demonstrated mutual intelligibility with Kuijau, a Dusunic language. Tagal has been renamed Sumambu/Tagal and encompasses five dialects: Pensiangan Murut, the most widely understood of the Sumambu/Tagal dialects, Salalir, Rundum, Tomani Sumambu and Maligan Tagal with the latter three showing greater affinity with each other than with Pensiangan or Salalir. Paluan comprises two dialects: (1) Paluan proper, which includes Smith's Paluan, Takapan and Sook and (2) Pandewan, for which Smith had no data. For Baukan, two dialects have been suggested: Baukan proper and Tenggara. The seventh Northern Murut language is designated Keningau Murut and includes Smith's Nabay, Apin-Apin Kuijau and Dusun-Murut groups as well as Ambual; Nabay is the most central of the four dialects.

The term "eastern Murutic" was coined to designate the remaining five languages of the Murutic family without intending to suggest that they should be identified as a formal unit. No testing was done in the Kolod language. Selungai Murut was distinguished on the basis of data not available to Smith. Testing confirmed the distinctiveness of Sembakung Murut, Kalabakan Murut and Serudung Murut.

Smith enumerated ten Dusunic languages which related to each other at the subfamily level. These ten languages were Papar, Dumpas, Kadazan-Tambanua, Lotud (including two dialects), Bisaya, Tatana, Kuijau (including five dialects), Eastern Kadazan (including seven dialects), Rungus and Kadazan/Dusun (including 13 dialects). Researchers attempting to do intelligibility testing in the only kampung speaking the Kadazan-Tambanua language found that subsequent to the earlier collection of language material the kampung had dispersed. Speakers of the language were scattered throughout the area and no longer functioning as a language community.

Testing within the Dusunic family confirmed the distinctiveness of Papar, Lotud, Bisaya, Tatana, Kuijau, Eastern Kadazan, Rungus and Dumpas. Within this testing several anomalies were noted. Papar speakers demonstrated high intelligibility of Tatana. Lotud seems to function on the local level as a dialect of Kadazan/Dusun as indicated by high intelligibility scores. Furthermore, the

mutual intelligibility between Bisaya and Tatana indicated that a closer relationship may be appropriate.

In addition, Smith's proposed dialect divisions for the Kuijau language were not confirmed by the dialect intelligibility testing. At least three dialects are confirmed for Eastern Kadazan, though others remain unclear. Internal and external testing of Dumpas indicated it may be more appropriately classified as a Paitanic rather than as a Dusunic language.

Smith's Kadazan/Dusun language designation has been significantly realigned. Testing within Kadazan/Dusun indicated four additional languages should be enumerated. These languages are Klias River Kadazan, Kimaragang, Garo and Tebilung. Four dialects are now posited for the Kadazan/Dusun languages; these include Central Dusun, Coastal Kadazan, Sugut Kadazan and Kuala Monsok (TA) Dusun.

Within the Bornean stock several languages seem to function transitionally between family groupings. Between Tidong and Murutic, the Kalabakan and Sembakung Murut languages share this position. Between the Murutic and Dusunic families, Murutic Gana and Dusunic Kuijau appear to be in the margin. Between Paitanic and Dusunic, Dumpas is marginal. Further study of these transitional areas between language families may provide valuable insights into the historical development of the Bornean languages.

In summary, the number of languages representing cultures of insular Southeast Asia has increased by two over Smith's calculation. The changes are identified as an increase from one to three East Coast Bajau languages, a decrease from two to one Ida'an language, a decrease from 15 to 12 Murutic languages and an increase from ten to 14 Dusunic languages.

3. NOTES ON BILINGUALISM

The survey data reflects numerous instances of bilingualism, where bilingualism is defined as the learning of a second language through social contact (a person may be "bilingual" in several languages). Bilingualism is reflected in the data by intelligibility scores which are significantly higher than corresponding PSC values. Sometimes bilingualism is unidirectional, or non-reciprocal, that is, speakers of language A learn language B, but speakers of language B do not learn language A. Bilingualism may also be bidirectional, or reciprocal, that is, speakers of language A learn language B and vice versa. This latter type may also be referred to as mutual intelligibility, though mutual intelligibility is not restricted to bilingualism; mutual intelligibility is expected to occur between very closely related speech samples. Thus bilingualism, whether reciprocal or non-reciprocal, does not contrast with intelligibility, but is rather a particular type of intelligibility (Simons 1979:3).

The survey data includes 62 reciprocal pairs of test and reference points. Among these pairs there are 20 cases where both A and B speakers registered intelligibility scores of 80% or higher on each other's language. These would be examples of mutual intelligibility. In only one case where the PSC relationship between sample A and sample B was below 75 PSC did speakers of both languages demonstrate mutual intelligibility above 80%. The Bisaya villagers of Padas Damit BT registered 85% intelligibility of the Tatana language of Bundu KP, and Bundu KP registered 88% intelligibility of Padas Damit BT; the cognate relation between the respective wordlists is 62 PSC. This appears a strong case for reciprocal bilingualism.

More frequently bilingualism is non-reciprocal. In the data on reciprocal pairs there are ten clear cases of non-reciprocal bilingualism where speakers of sample A registered an average intelligibility score of greater than 80% on sample B, and speakers of B registered intelligibility of less than 80% on A, and the spread between the two scores was greater than ten percentage points. (The range for the difference between the two scores of these reciprocal pairs is 14-35%. The average difference for the ten pairs is 27%.)

Bilingualism is largely a product of social contact between speakers of different languages, though the degree of linguistic similarity may facilitate the language learning process. In addition to cases of bilingualism which have already been cited, dialect intelligibility testing in the Keningau District demonstrated a high rate of bilingualism. Out of 70 comparable cases, there were 26 intelligibility tests where the difference between an intelligibility score above 80% and a cognate relation below 80 PSC was greater than ten percentage points. Within Keningau District, the town of Keningau serves as a market, educational and administrative centre for speakers of both Dusunic and Murutic languages, thus setting the stage for a high degree of social contact between language groups. The nature of the factors affecting bilingualism within Keningau District and elsewhere in Sabah is an area where further study is desired.

Inasmuch as the intelligibility data has not yet been analysed in a way which might account for the effect of social factors, the handling of the data has been subjective in some cases. The value of the intelligibility data, however, cannot be dismissed. In many cases Smith's conclusions were confirmed by the results of intelligibility testing where there is no reason to expect a skewing of the results due to sociolinguistic factors. On the other hand, in some cases the results of intelligibility testing indicate that a reclassification is appropriate. An example of the latter is the Dumpas situation. On lexicostatistical grounds, Smith had classified Dumpas as a Dusunic language; its highest percentage of shared cognates was with a dialect of Eastern Kadazan, a Dusunic language. Dumpas is now spoken in an area where Eastern Kadazan is the dominant language. Intelligibility testing, however, indicated that Dumpas is more closely related to the Paitanic family. This latter indication is likewise supported by sociolinguistic evidence, as well as the lexicostatistical evidence that Dumpas shows lower lexical relationships with Dusunic languages than it does with Paitanic languages.

Testing within the Malayic family provided another example to demonstrate how intelligibility data can supply evidence for reclassification. Lexicostatistically, the Brunei/Kedayan samples tested appeared to be dialectally the same. Mutual intelligibility was expected. The testing results, however, show that the intelligibility is non-reciprocal. A dialect distinction, whether the basis is some yet undiscovered linguistic dissimilarity or sociolinguistic factors, should be recognised.

4. NOTES ON THE TESTING OF NATIONAL LANGUAGE COMPREHENSION

Testing for comprehension of the national language was conducted in all but four villages where intelligibility testing was conducted throughout the state. There were 146 such tests where there was also a cognate relation figured between the wordlist for the dialect being tested and the Sabah Malay wordlist. These scores divided almost evenly into two groups. Of the 146 correlations,

70 of the Malay test scores were between 75% and 100%, and 76 scores were between 12% and 70%. The average score for the higher percentage group was 86.5%, while the average score for the lower percentage group was 52%. The mean for all 146 scores was 68.5%.

Generally, throughout the testing several factors were noted as contributing to higher scores on the Malay reference tape. The most crucial factor seems to be education. In only a few locations did this correlation between education and higher Malay scores not hold. Age as a factor seems directly related to the education factor. The older test subjects generally (though with some notable exceptions) had not received much, if any, formal education. Education also seemed to level the difference in performance between males and females. Among the formally uneducated, males generally outscored females, which seemed to be a result of their greater mobility. Finally, the location of the test point with respect to population centres with their market, educational and administrative functions correlated well with the scores. It is evident that as the government of Sabah continues to develop its resources and its people through improvement of its educational institutions the percentage of people who competently and confidently can communicate in the national language will likewise continue to increase.

5. COMPARISON WITH DYEN

The proposed classification as it is presented here is similar to that produced by Dyen (1965), though there are some differences (Figure 3). All of the languages except Chabacano fit within Dyen's Hesperonesian Linkage. The Bugis and Malayic languages are most incongruous with Dyen's classification. He identifies Bugis (Buginese) with his Celebes Hesion. Butung, a language of southern Celebes which Dyen does not discuss, is placed within a Celebes grouping by Voegelin and Voegelin (1977), and presumably would also fit within Dyen's Celebes Hesion.

Dyen places Malayic coordinate with Javanese in his West Indonesian Cluster. Dyen groups his Celebes Hesion, West Indonesian Cluster and Northwest Hesion as coordinate members of Hesperonesian. In the present study the Bugis subfamily and the Malayic family are subordinate within Smith's North-western Austronesian superstock, which otherwise roughly parallels Dyen's Northwest Hesion. A possible implication of this is that either a great degree of lexical inflation is present in the data under consideration (that is, northern Borneo Bugis and Malay have become more like the remaining languages in the Northwest Hesion or vice versa), or there is actually a lesser degree of language variation for these languages than indicated by Dyen.

Most of Sabah's languages fit within Dyen's Northwest Hesion. Three languages of the present study, each representing a distinct linguistic stock (Smith's terminology), are not found in Dyen. Lundayeh, the most disparate language within Smith's North-western Austronesian superstock (which roughly parallels Dyen's Northwest Hesion) is, according to Clayre (1970), closely related to Kelabit which Dyen did not assign to one of his Hesperonesian divisions, though he notes that the closest relation between a member of his Kelabitic Cluster and a non-member is with Murut (30.2%, a figure which parallels Smith's figure for the relationship between Lundayeh and the Bornean stock languages). The Banggi language may represent Dyen's Palawan subdivision of his Sulic Hesion, as it was found to relate more closely with the Molbog language

spoken on islands in the southern Palawan area than with Sabah languages. This, however, would indicate a closer relationship with Suluk (Tausug) of Dyen's Bisayan Cluster than with Ida'an and the Bornean stock languages, contrary to Smith's findings. The third language representing a separate linguistic stock is the Ida'an language. Ida'an has no known counterpart in Dyen's work.

One other area of comparison between Dyen's study and the SIL study involves the languages Smith has grouped together within his Bornean stock. Dyen grouped Tidong with Murut, whereas the SIL study finds reason to keep them distinct. Although Dyen separated Murutic and Dusunic languages as coordinate members of his Northwest Hesion, they are grouped together within the Bornean stock in the present classification. The Paitanic languages in the present study are not represented in Dyen's work.

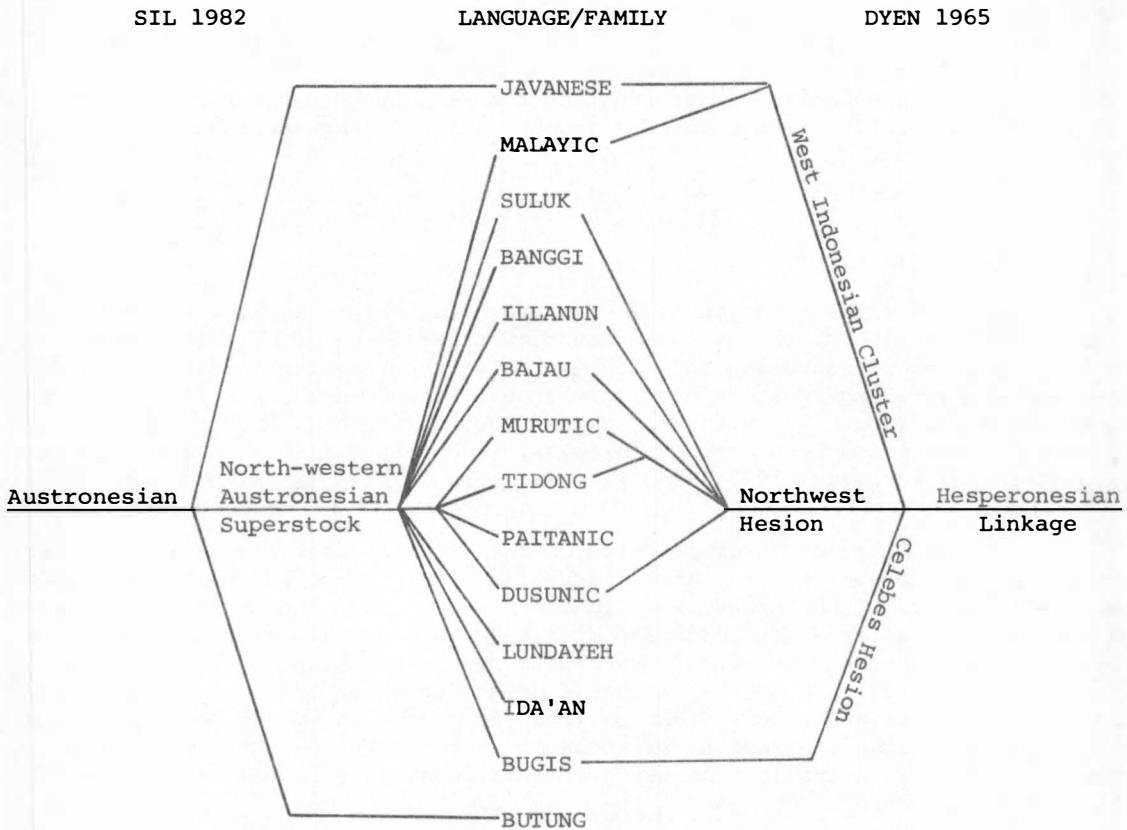


Figure 3: Comparison of SIL's classification of Sabah languages with Dyen's 1965 classification for the same language groups. (The language designations follow SIL's classification.)

6. COMPARISON WITH PRENTICE

The majority of the languages of this study are not known to be represented by communities located outside of northern Borneo. Thus it is most appropriate to compare these results with that of D.J. Prentice (1970) whose classification of Sabah languages has heretofore been the most thorough and detailed. Figure 4 presents this comparison in summary.

Prentice's classification comprises four taxonomic levels: subfamily, language-group, language and dialect. Prentice includes three higher order categories, the Murutic, Dusunic and Paitanic subfamilies.⁴ Using different thresholds (Smith, in this volume), the SIL classification includes stock and family groupings as well, but does not employ Prentice's "language-group" category, thus making "family" and "subfamily" in the current work roughly equivalent to Prentice's "subfamily" and "language-group", respectively.

Perhaps the most apparent of the differences between the two classifications is in the enumeration of languages. Prentice lists 19 languages whose cultural centres could be said to lie in Sabah. The corresponding SIL count is 34 languages. Most of the difference can be attributed to the more extensive collection of uniform-quality data in the SIL survey.⁵ Some differences also result from the application of different thresholds.

Within the two classification several differences can be noted. In his Murutic subfamily Prentice joins the Tidong and Murut "language-groups" (as did Dyen). In the present study, though Tidong's relation with Murut is regarded as closer than with Dusunic or Paitanic, the two are separated at the higher level. Two of Prentice's three Tidong constituents, however, are considered members of the Murutic family in this study. Prentice includes four Bisayan languages,⁶ including Tatana, while SIL distinguished only one Bisaya language which, as Prentice indicated, shows a closer relationship with Tatana than with other languages of the Dusunic family. Prentice includes the Banggi, Paitan and Buludupi languages within his Paitanic grouping. SIL has distinguished Banggi as a separate stock, which as noted above (as also by Prentice) may relate more closely to languages of the southern Philippines than to other Sabah languages. SIL's research has revealed five Paitanic languages rather than one. Buludupi, for which Prentice found nineteenth century wordlists and which is attested to in literature during the time of the Chartered Company rule, has not been positively identified by SIL investigations.⁷

Finally, this survey has attempted to define linguistic entities within the geographical confines of Sabah, not to link them in any formal way with languages found geographically outside of the state. The larger task of clarifying the higher order relationships which exist across Sabah's borders remains to be done. Even within the scope of the present work there are numerous relationships which remain tentative. These have been indicated by the various writers. It is hoped that the presentation of these papers will contribute significantly to the knowledge of Sabah's peoples and languages and will provide incentive to further linguistic research within insular South-east Asia.

NOTES

1. For an example of such debating as it also applies to a particular lexicostatistical classification, see the review of Dyen (1965) by Grace (1966) and the comments by Dyen (1966) and Hymes (1966) in the same volume.
2. Since Smith's lexicostatistical analysis (in this volume) was completed in 1980, the classification arrived at from that analysis is referred to as Smith 1980. Similarly, since the revised classification arising out of the intelligibility testing was completed in 1982, it is here referred to as SIL 1982.
3. The map serves as a general guide to the locations of the various language groups. Only those languages which were considered either indigenous to Sabah or of long-term residence in Sabah, and which have been discussed in this volume, have been included in the map.
4. Appell (1968) listed Bisaya as a separate subfamily along with Dusun and Murut in his more tentative earlier classification.
5. Prentice's study was based on the comparison of 73 wordlists, 64 of which he considered as "Ida'an" (=Bornean stock) languages; 54 of the lists were gathered from published sources dating as far back as 1872.
6. Prentice cites locations in Sarawak for two of his Bisaya groups, Southern Bisaya and Tutong Bisaya.
7. Prentice cites two wordlists, both of inadequate length, which exhibit 50 PSC between them. One of the lists demonstrates high lexical relations (83-87 PSC) with several of SIL's Ida'an wordlists. It is, of course, unlikely that the second list would likewise show such high relations with the same Ida'an lists. It is more likely that "Buludupi" is an early exonym applied to more than a single linguistically distinct speech group (John A. Spitzack, personal communication).

BIBLIOGRAPHY

AERNOUT, W.

- 1885 Een woordenlijste der Tidoengsche taal (A Tidong wordlist). *De Indische Gids* 7:536-550.

ABDUL GHANI bin Bagul

- 1950 Notes on the Bajau language. *Sarawak Museum Journal* 5/2:196-200.

ALLISON, Cyril

- 1966 *Threatened paradise: North Borneo and its people*. London: Robert Hale.

ALLISON, Joe

- 1977 Discourse analysis of a southern Samal text. In Casilda E. Luzares and Austin Hale, eds *Studies in Philippine linguistics*, 1/2:143-169. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.

ALMAN, John H.

- 1960 Bajau saddle making. *Sarawak Museum Journal* 9:619-637.

ANON

- 1900 Babel. *British North Borneo Herald* 18/17:283-285.
- 1964 *English-Murut phrase book*. Sabah: Government Printer.
- n.d. Dusun analysis - Ranau dialect. MS.
- n.d. Dusun - comparative sentences. MS.

ANTONISSEN, A.

- 1958 *Kadazan-English and English-Kadazan dictionary*. Canberra: Government Printing Office.

APPELL, G.N.

- 1963 Myths and legends about the Rungus of the Kudat district. *Sabah Society Journal* 1/4:9-15.

- 1964 The longhouse apartment of the Rungus Dusun. *Sarawak Museum Journal* 11:570-573.
- 1966 The ethnography of Dusun-speaking peoples of Sabah. *Sarawak Museum Journal* 14:376-388.
- 1967 Ethnography of northern Borneo: critical review of some recent publications. *Oceania* 37:178-185.
- 1968 The Dusun languages of North Borneo: the Rungus Dusun and related problems. *Oceanic Linguistics* 7:1-15.
- 1968 Social and medical anthropology of Sabah: retrospect and prospect. *Sabah Society Journal* 3/4:247-286.
- 1970 Ethnographic notes on the Iranon Maranao (Illanun) of Sabah. *Sabah Society Journal* 5/2:77-82.
- 1972 Rungus Dusun. In LeBar, ed. 1972:150-153.

APPELL, G.N., ed.

- 1969 *Studies of the Tausug (Suluk) and Samal-speaking populations of Sabah and the southern Philippine Islands.* *Borneo Research Bulletin* 22/1.

APPELL, G.N. and Robert HARRISON

- 1968 Ethnographic profiles of the Dusun-speaking peoples of Sabah, Malaysia. *Journal of the Malaysian Branch of the Royal Asiatic Society* 41/2:131-147.

ARANETA, F. and M. BERNAD

- 1960 "Bisayans" of Borneo and the "Tagalogs" and "Visayans" of the Philippines. *Sarawak Museum Journal* 9/15-16:542-564.

ASHLEY, Seymour

- 1971 A case classification of Tausug verbs. *Sulu Studies* 2:70-85. Jolo, Sulu: Notre Dame of Jolo College.
- 1973 Notes on Tausug orthography. *Sulu Studies* 2:86-94. Jolo, Sulu: Notre Dame of Jolo College.

ASHLEY, Seymour and Lois ASHLEY

- 1963 The phonemes of Tausug. *Papers on Philippine Languages* 1:7-17. Manila: Institute for Language Teaching and Summer Institute of Linguistics.
- 1971 Outline of sentence types of Tausug. *Philippine Journal of Linguistics* 2/1:44-91. Manila.

ASMAH HAJI OMAR

- 1975 *Essays on Malaysian linguistics.* Kuala Lumpur: Dewan Bahasa dan Pustaka.
- 1977 The Iban language. *Sarawak Museum Journal* 25/46:81-100.
- 1978 The Suluk language of Samporna. *Brunei Museum Journal* 4/2:11-33.

- 1978 Verbal affixes in standard Kadazan. *Sarawak Museum Journal* 26 (New)/47:161-194.
- 1980 The Kadazan language of Tambunan. *Sabah Society Journal* 6/4:178-206.
- 1981 Paitan. In Andrew Gonzalez and David Thomas, eds *Linguistics across continents: studies in honor of Richard S. Pittman*. Manila: Summer Institute of Linguistics and Linguistic Society of the Philippines. LSP Monograph Series Number 11.
- AWANG SUDJAI HAIRUL and Yusoff KHAN, eds
- 1977 *Kamus lengkap*. Petaling Jaya: Pustaka Zaman.
- BABONEAU, N.B.
- 1922 A Murut vocabulary. *Journal of the Straits Branch of the Royal Asiatic Society* 86:343-375. (Includes an introduction by G.C. Woolley.)
- BARRY, Leo J.
- n.d. *English-Iban phrase book*.
- BEECH, M.W.H.
- 1908 *The Tidong dialects of Borneo*. London: Oxford University Press.
- BLOM, Julie K.
- 1979 Dialect comparison and bilingualism: the Kudat division. *Sabah Museum Annals* 1:23-46. Kota Kinabalu: Sabah Museum.
- BOLANG, A. and Tom HARRISSON
- 1949 Murut and related vocabularies. *Sarawak Museum Journal* 5/1:116-124.
- BOUTIN, Michael and Alanna BOUTIN
- 1982 Report on the linguistic situation on Banggi and Balambangan. (to appear).
- BURROUGH, P.A.
- 1970 Stick signs in the Sook plain. *Sabah Society Journal* 5/2:83-97.
- CASAD, Eugene H.
- 1974 *Dialect intelligibility testing*. Summer Institute of Linguistics Publications in Linguistics and Related Fields, Number 38. Norman, Oklahoma: Summer Institute of Linguistics and the University of Oklahoma.
- CATHOLIC MISSION
- 1960 *Murut catechism*. (2nd, revised edn). Keningau R.C. Mission.

CENSE, A.A. and E.M. UHLENBECK

- 1958 *Critical survey of studies on the languages of Borneo*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde, Bibliographical Series 2. The Hague: Martinus Nijhoff.

CHATFIELD, Godfrey A.

- 1972 *Sabah: a general geography*. (Fourth revised edn). London and Singapore: Eastern Universities Press Snd. Bhd.

CLARKE, M.C.

- 1951 Some impressions of the Muruts of North Borneo. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 44/4:453-464.

CLAYRE, Beatrice M.

- 1964 A comparison of three dialects of Dusun: Ranau, Rungus and Ulu Tuaran. MS.
- 1965 A study of Dusun dialects. MS.
- 1965 Towards a description of the grammar of the Dusun language. MS.
- 1966 A comparison of some dialects of Dusun. *Sabah Society Journal* 3/1: 3-12.
- 1967 The two favorite clause types in use in Dusun. *Sabah Society Journal* 3/3:113-127.
- 1970 Focus: a preliminary survey of some languages of eastern Malaysia. *Sarawak Museum Journal* 18(New)/36-37:193-219.
- 1970 Some notes on the Grammar of Dusun. *Sabah Society Journal* 5/2:133-140.
- 1972 A preliminary comparative study of the Lun Bawang (Murut) and Sa'ban languages of Sarawak. *Sarawak Museum Journal* 20:40-41; 45-71.
- n.d. Rungus notes. MS.

CLAYRE, Beatrice M., Jeanne DAVIS and T. FORSCHNER

- n.d. Rungus Dusun notes. MS.

COHEN, Patrick D.

- 1981 Linguistic unity and diversity in Keningau district (Sabah). Paper presented at the Dialek Symposium: Penyelidikan dan Pendidikan, Universiti Kebangsaan Malaysia, 2-3 December 1981.

COMBER, J.

- 1971 Murut and Lun Dayah bird names. *Sabah Society Journal* 5/3:172-176.

COPELAND, A.J.

- 1935 Muruts of North Borneo: malaria and racial extinction. *Lancet* 228: 1233-1239.

CRAIN, Jay B.

- 1972 Murut depopulation and the Sipitang Lun Dayeh. *Journal of the Malaysian Branch of the Royal Asiatic Society* 45/2:110-119.
- 1978 The Lun Dayeh. In Victor T. King, ed. *Essays on Borneo societies*, 7:123-143. Hull monographs on South-East Asia. Oxford: Oxford University Press.

CRESPIGNY, C. de

- 1872 On Northern Borneo. *Proceedings of the Royal Geographic Society* 16: 171-183.

DEEGAN, James L.

- 1970 Some Lun Bawang spirit chants. *Sarawak Museum Journal* 18:36-37; 264-280.
- 1978 A brief description of the Lun Bawang kinship terminology. *Sarawak Museum Journal* 27:48:53-77.

DUNN, Phyllis A.

- 1980 *Sabah people's who's who: a glossary of the terms used for the people and languages of Sabah, East Malaysia*. Sabah Museum Annals No.2. Kota Kinabalu: Sabah Museum.

DYEN, Isidore

- 1965 *A lexicostatistical classification of the Austronesian languages*. *International Journal of American Linguistics*, Memoir 19 (Vol.31, No.1).
- 1966 Comment. *Oceanic Linguistics* 5/1:32-49.

EVANS, Ivor, H.N.

- 1952 Notes on the Bajau and other coastal tribes of North Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 25/1:48-55.

EVANS, Terry

- n.d. Tuaran Dusun dictionary. MS.

FERNANDO, C.

- 1978 The way of life of the Kadazans (Dusun) in the interior of Sabah (Borneo). MS.

FLEISCHMAN, Eric

- 1981 The Danao languages: Magindanaon, Iranun, Maranao, and Illanun. MS.
- 1982 Are there Iranuns in Sabah? Submitted to the *Dansalan Quarterly*. (In press.)

FORAN, J.

- 1967 Comparison of words from dialects known as "Dusun". MS.
- 1967 Comparison of 100 words from dialects known as Kedazan, Rungus and Dusun. MS.

FORREST, Thomas

- 1780 *A voyage to New Guinea and the Moluccas from Balambangan, including an account of Magindano, Sooloo, and other islands.* 2nd edn, 372-374. London.
- 1969 *A voyage to New Guinea and the Moluccas 1774-1776.* Kuala Lumpur: Oxford University Press.

FORSCHNER, T.

- 1978 English-Rungus dictionary. MS.

GARMAN, M.A., P.D. GRIFFITHS and R.J. WALES

- 1970 Prepositions and noun particles in children's speech. *Sarawak Museum Journal* 18:353-376.

GENDEREN STORT, P. van

- 1916 Nederlandsch-Tidoengsch-Tinggalan-Dajaksche woordenlijst. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen* 61/5:i-iv, 1-100. Batavia.

GLYN-JONES, Monica

- 1953 *The Dusun of the Penampang Plains in North Borneo.* London: Colonial Social Science Research Council.

GOLLAN, J.

- n.d. Dusun notes. MS.

GOSENS, A.L.

- 1924 A grammar and vocabulary of the Dusun language. *Journal of the Malayan Branch of the Royal Asiatic Society* 2/2:87-220.

GRACE, George W.

- 1966 Austronesian lexicostatistical classification: a review article. *Oceanic Linguistics* 5/1:13-31.

GRIMES, Barbara F.

- 1974 *Ethnologue.* Huntington Beach: Wycliffe Bible Translators.

HARDAKER, M.B.

- 1963 An introduction to Suluk. *Sarawak Museum Journal* 11/21-22:138-173.
- 1964 An introductory vocabulary of Suluk. *Sarawak Museum Journal* 11/23-24:578-596.

HARRISSON, Tom

- 1956 "Bisaya": Borneo-Philippine impacts of Islam. *Sarawak Museum Journal* 7:43-47.
- 1958 The caves of Niah: a history of prehistory. *Sarawak Museum Journal* 8/12:549-595.

- 1959 The Kelabits and Muruts. In Tom Harrisson, ed. *The peoples of Sarawak*. Kuching: Government Printing Office.
- 1962 "Bisaya" in North Borneo and elsewhere. *Sabah Society Journal* 1/2: 6-13.
- 1964 The peoples of north and west Borneo. In Wang Gungwu, ed. *Malaysia: a Survey*, 163-178. London and New York: Pall Mall Press and Praeger.
- 1967 Ethnological notes on the Muruts of the Sapulut River, Sabah. *Journal of the Malaysian Branch of the Royal Asiatic Society* 40/1: 11-129.
- 1973-74 The Bajaus: their origins and wide importance. *Sabah Society Journal* 6/1:38-41.
- 1976 Ngaju-Bajau: significant early term usages in Southeast Asia. *Journal of the Malaysian Branch of the Royal Asiatic Society* 49/1: 1-14.

HARRISSON, Tom and Barbara HARRISSON

- 1970 The prehistory of Sabah. *Sabah Society Journal* 4:229-237.

HASSAN, Irene, Nurhadan HALUK, Seymour ASHLEY and Lois ASHLEY

- 1975 *Tausug-English dictionary; kabtangan Iban Maana*. *Sulu Studies* 4. Coordinated investigation of Sulu culture and Summer Institute of Linguistics. Manila: Summer Institute of Linguistics.

HEADLEY, Derek

- 1950 Some Bisaya Folklore. *Sarawak Museum Journal* 5:187-192.
- 1951 Some Illanun and Bajau marriage customs in the Kota Belud District, North Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 24/3:159-162.

HEALEY, A. and P.M. HEALEY

- 1961 Dusun dialect comparison. MS.

HOLLY, S.

- 1955 The origin of the Idahan people. *Sarawak Museum Journal* 6/5:257-262.

HOSE, Charles and William McDOUGALL

- 1912 *The pagan tribes of Borneo*. 2 vols. London: Macmillan.

HURLBUT, Hope M.

- 1979 Some features of narrative discourse in Kadazan. In Nguyen Dang Liem, ed. *South-East Asian linguistic studies*, vol.4, *Pacific Linguistics*, C-49:255-282.
- 1981 Morphophonemics of Labuk Kadazan. In Andrew Gonzalez and David Thomas, eds *Linguistics across continents: studies in honor of Richard S. Pittman*. LSP Monograph Series No.2:46-53.
- n.d. Trilingual thesaurus: Kadazan-English-Malay. MS.

- n.d. Verbal affixation in Labuk Kadazan. MS.
- HURLBUT, Hope M. and Inka PEKKANEN
forth- Dialect comparison and intelligibility testing in the upper
coming Kinabatangan River area. To appear.
- HYMES, Dell
1966 Comment. *Oceanic Linguistics* 5/1:50-57.
- JAMES, Harold and Denis SHEIL-SMALL
1979 *The undeclared war: the story of the Indonesian confrontation 1962-1966*. Kuala Lumpur: University of Malaya Co-operative Bookshop.
- JOHNSON, Osa
1967 *Last adventure: the Martin Johnsons in Borneo*. London: Jarrolds.
- JONES, L.W.
1962 *North Borneo: report on the census of population taken on 10 August 1960*. Kuching: Government Printing Office.
1966 *The population of Borneo: a study of the peoples of Sarawak, Sabah and Brunei*. London: The Athlone Press.
1967 The decline and recovery of the Murut tribe of Sabah. Reprinted from *Population Studies* 21/2:133-157.
- KEITH, Agnes Newton
1981 *Land below the wind*. Petaling Jaya: Eastview Productions Sdn. Bhd. Reprinted from the 1939 edition.
- KEITH, H.G.
1936 A few Ulun-no-Bokan (Murut) taboos. *Journal of the Malayan Branch of the Royal Asiatic Society* 14/3:327-329.
1936 Some Ulun-no-Bokan (Murut) charms. *Journal of the Malayan Branch of the Royal Asiatic Society* 14/3:330.
1936 Some Ulun-no-Bokan (Murut) words from North Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 14/3:314-322.
1936 Ulun-no-Bokan (Murut) folklore. *Journal of the Malayan Branch of the Royal Asiatic Society* 14/3:323-326.
- KHOO KAY KIM
1981 *Sabah: history and society*. Kuala Lumpur: Malaysian Historical Society.
- KIEFER, Thomas M.
1968 Concerning the word "Suluk". *Sarawak Museum Journal* 16/32-33:438-439.

KILGOUR, Heather J. (compiler)

- 1978 *Bibliography of the Summer Institute of Linguistics: Philippines, 1953-78.* Manila: Summer Institute of Linguistics.

KROEGER, Paul R.

- 1981 A proto-Dusunic word list. MS.

LANDGRAF, John L.

- 1955 Anthropological research in British Borneo. *Transactions of the New York Academy of Science, Series 2/18:60-68.*
- 1956 *Interim report to the Government of the Colony of British North Borneo on field work done under the supervision of the Department of Medical Services, 1954-1955.* Jesselton: Government Printing Office.

LeBAR, Frank M.

- 1972 Dusun. In LeBar, ed. 1972.
- 1972 Iban. In LeBar, ed. 1972.
- 1972 Kelabitic Murut. In LeBar, ed. 1972:158-163.

LeBAR, Frank M., ed.

- 1972 *Ethnic groups of insular Southeast Asia, vol.1: Indonesia, Andaman Islands, and Madagascar.* New Haven: Human Relations Area Files Press.

LEE YONG LENG

- 1965 *North Borneo (Sabah): a study in settlement geography.* Singapore: Eastern Universities Press.

LEES, Shirley P.

- 1959 Lun Daye phonemics. *Sarawak Museum Journal* 9/13-14:56-62.
- 1964 Apparent and real differences in Dusun linguistics. *Sarawak Museum Journal* 11/23-24:574-577.
- 1966 Murut orthography. *Sabah Society Journal* 3/2:90-97.

LEES, W.C.

- n.d. The Dusun verb. MS.

LUERING, H.L.E.

- 1897 A vocabulary of the Dusun language of Kimanis. *Journal of the Straits Branch of the Royal Asiatic Society* 30:1-29.

MAJUSIM BIN MAJINAL, Edmund

- 1949 Some notes on the Dusun language. *Sarawak Museum Journal* 5:125-129.

MALAYSIA

- 1976 *1970 Population and housing census of Malaysia, vol.1: basic population tables; part 12: Sabah. [Banci Penduduk dan Perumahan Malaysia, vol.1/12.]* Kuala Lumpur: Jabatan Perangkaan Malaysia.

MALAYSIA: DEPARTMENT OF STATISTICS

- 1978 *Statistical Handbook: Sabah*. Kota Kinabalu: Department of Statistics.

MATIUS MATULANG (compiler)

- 1979 *Cerita dongeng Sabah: cerita dongeng Kadazan*. Bahasa Malaysia trans. by Hope Hurlbut and James Muyou. English trans. by Rebbin Noyo. Sabah Museum Annals No.1. Sabah Museum, Kota Kinabalu.

MAXWELL, Allen R.

- 1970 On the various spellings of the word "Kedayan". *Brunei Museum Journal* 2:87-103.

MILLER, Carolyn P.

- 1981-82 A brief look at the languages and dialects of Sabah. *Sabah Society Journal* 7/1:53-64.

MOHRING, Hans

- 1967 Concerning the word "Suluk". *Sarawak Museum Journal* 15/30-31:243-244.

MURDOCK, George B.

- 1964 Genetic classification of the Austronesian languages: a key to oceanic culture history. *Ethnology* 3:117-126.

NEEDHAM, Rodney

- 1955 A note on ethnic classification in Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 27:167-171.

NEW TESTAMENT

- 1968 *Kula-Kula' Buku Nu Rahu Nu Tuhan*. Lawas: Bible Societies in Malaysia, Singapore and Brunei and Sidang Injil Borneo. (Includes Luke, John, Acts, Romans, Ephesians, Philippines, James, and I & II Peter.)
- 1976 *Abar Taandai Tinulis Di Marakus*. (The good news written by Mark.)
- 1977 *Sorita Diri Susuon Di Yisus*. (Acts of the Apostles.)

ONGKILI, James P.

- 1972 Pre-western Brunei, Sarawak and Sabah. *Sarawak Museum Journal* 20/40-41:1-20.

PADAN, James Noah

- 1971 *Lun Dayeh (Murut) phrase book*. Kuching: Borneo Literature Bureau.

PALLESEN, A. Kemp

- 1965 Samal phonemics. MS.
- 1977 Culture contact and language convergence. Ann Arbor: University Microfilms International. (Ph.D. dissertation.)

PEKKANEN, Inka

- 1981 Some consonantal sound changes in the Kadazan dialects of the Penampang, Papar, and Beaufort districts of Sabah, Malaysia. *Borneo Research Bulletin* 13:33-47.

PERANIO, Roger D.

- 1972 Bisaya. In LeBar, ed. 1972.

PIKE, Kenneth L. and Evelyn G. PIKE

- 1977 *Grammatical analysis*. Summer Institute of Linguistics Publications in Linguistics, Number 53. Dallas, Texas: Summer Institute of Linguistics and the University of Texas at Arlington.

POLUNIN, I.

- 1959 The Muruts of North Borneo and their declining population. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 53/4: 312-321.
- 1959 A note on visual non-literary methods of communication among the Muruts of North Borneo. *Man* 59:97-99.

PRENTICE, D.J.

- 1965 Form and function in the verbs of Sabah Murut: a preliminary analysis. *Oceanic Linguistics* 4/1-2:127-156.
- 1969 A word list for use in Borneo. *PL*, A-20:1-8.
- 1969 Verbal inflection in Sabah Murut. *PL*, A-20:9-22.
- 1969 Phonemes of Sabah Murut. *PL*, A-20:23-41.
- 1970 The linguistic situation in northern Borneo. In S.A. Wurm and D.C. Laycock, eds *Pacific Linguistic studies in honour of Arthur Capell*, *PL*, C-13:369-408.
- 1971 *The Murut languages of Sabah*. *PL*, C-18.
- 1972 Idahan Murut. In LeBar, ed. 1972:154-158.
- 1972 Notes on place names and personal names in the song-language of the Timugon Murut. *Sabah Society Journal* 5/4:371-376.
- 1982 Some ludic aspects of Timugon Murut. In Amran Halim, Lois Carrington and S.A. Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics*, *PL*, C-76:319-324.

PRYER, W.B.

- 1881 Diary of a trip up the Kina-Batangan river. Repr. in *Sabah Society Journal*, 1970, 5/2:117-126.
- 1883 Notes on north-eastern Borneo and the Sulu Islands. *Royal Geographic Society Proceedings* 5:90-96.

PUR, Samuel Labo

- 1961 *Kamus Lun Dayeh*. Kuching: Borneo Literature Bureau.

RAY, Sidney H.

- 1913 The languages of Borneo. *Sarawak Museum Journal* 1/4:1-196.

REGESTER, P.T.

- 1956 The study of primitive communities and its relation to social medicine with particular reference to the Muruts and Dusuns of North Borneo. *Journal of the Royal Institute of Public Health and Hygiene* 44/4:453-464.

REID, Lawrence A.

- 1971 *Philippine minor languages: word lists and phonologies. Oceanic Linguistics Special Publication No.8.* Honolulu: University of Hawaii Press.

RIDE, Lindsay T.

- 1934 The problem of depopulation with special reference to British North Borneo. *Population* 1934:36-48.

ROTH, Henry Ling

- 1896 *The natives of Sarawak and British North Borneo.* (2 vols). London: Truslove and Hansen. (Repr. 1968.)

RUTTER, Owen

- 1922 *British North Borneo: an account of its history, resources and native tribes.* London: Constable.
1929 *The pagans of North Borneo.* London: Hutchinson.

SATHER, Clifford A.

- 1965 Bajau numbers of adjectives of quantity. *Sabah Society Journal* 2/4: 194-197.
1968 Some notes concerning Bajau Laut phonology and grammar. *Sabah Society Journal* 3/4:205-224.
1972 Tidong. In LeBar, ed. 1972:154-168.
1978 The Bajau Laut. In Victor T. King, ed. *Essays on Borneo societies*, 172-192. Hull Monograph on South-East Asia No.7. Oxford.

SCHNEEBERGER, W.F.

- 1937 A short vocabulary of the Banggi and Bajau languages. *Journal of the Malayan Branch of the Royal Asiatic Society* 15/3:145-164.

SEYMOUR, J. Madison

- 1977 Urbanization, schooling and psychocultural adaptation: Iban students of Sarawak, Malaysia. *Sarawak Museum Journal* 25/46:177-200.

SHARIFUDDIN, R.M.

- 1969 The Kedayans. *Brunei Museum Journal* 1:1-15.

SHIRCORE, J.O.

- 1937 *Report on native health in North Borneo.* Sandakan: Government Printing Office.

SIMONS, Gary F.

- 1979 *Language variation and limits to communication.* Technical Report No.3. Ithaca: Cornell University.

SMITH, Kenneth D.

- 1974 A computer analysis of Vietnam language relationships. *Workpapers of the Summer Institute of Linguistics, University of North Dakota Session* 18:99-113.
- 1978 An automatic typewriter assist to comparative linguists, with application to 30 Mon-Khmer languages. Paper presented to the Second International Congress on Austronesian Linguistics, Mysore, India, December 19-21, 1978. Mimeo.

SOPHER, David E.

- 1965 *The sea nomads: a study based on the literature of the maritime boat people of South-East Asia.* Memoirs of the National Museum, No.5. Singapore: Lim Bian Han.

STAAL, J.

- 1923-24 The Dusuns of North Borneo. *Anthropos* 18-19:958-977.
- 1926 The Dusun language. *Anthropos* 21:938-951.

SULLIVAN, Anwar and Patricia REGIS

- 1981 Demography. In Anwar Sullivan and Cecilia Leong, eds *Commemorative history of Sabah 1881-1981*, 545-584. The Sabah State Government Centenary Publications Committee.

SWETTENHAM, F.A.

- 1880 Comparative vocabulary of the dialects of the wild tribes inhabiting the Malayan Peninsula, Borneo. *Journal of the Straits Branch of the Royal Asiatic Society* 5:125-156.

TREGONNING, K.G.

- 1960 *North Borneo.* London: Her Majesty's Stationery Office.

VOEGELIN, C.F. and F.M. VOEGELIN

- 1965 *Languages of the world: Indo-Pacific fascicle four.* *Anthropological Linguistics* 7/2.
- 1977 *Classification and index of the world's languages.* New York: Elsevier.

VREELAND, Nena and others

- 1977 *Area handbook for Malaysia.* 3rd edn. Washington: US Government Printing Office.

WALTON, Charles

- 1977 A Philippine language tree. Paper presented to the Austronesian Symposium, University of Hawaii, August 18-20, 1977. (Published in *Anthropological Linguistics*, October 1979.)

WARREN, James F.

- 1971 *The North Borneo Chartered Company's administration of the Bajau, 1878-1909: the pacification of a maritime nomadic people*. Papers in International Studies, Southeast Asia Series No.22. Athens, Ohio: Ohio University Center for International Studies.
- 1972 The North Borneo Chartered Company's administration of the Bajau, 1878-1909. *Sabah Society Journal* 5:288-351.
- 1975 Trade, raid, slave: the socio-economic patterns of the Sulu zone: 1770-1898. Ph.D. dissertation, Australian National University, Canberra.

WELLINGS, D.R.

- 1960 Dusun dialect survey: North Borneo. MS.

WHELAN, F.G.

- 1970 *Sabah*. Singapore: Macmillan.

WILLIAMS, Thomas Rhys

- 1960 A survey of native peoples of North Borneo. *Sociologus* 10:170-174.
- 1961 Folklore texts: a Tambunan Dusun origin myth. *Journal of American Folklore* 74:68-73.
- 1965 *The Dusun: a North Borneo society*. New York: Holt, Rinehart and Winston.

WOLFF, Hans

- 1964 Intelligibility and inter-ethnic attitudes. In Dell Hymes, ed. *Language in culture and society*. New York: Harper and Row. [Repr. with slight revisions from *Anthropological Linguistics* 1/3:34-41.]

WOOLLEY, G.C.

- 1927 Two Murut pantuns from the Dalit district, Keningau, British North Borneo. *Journal of the Malayan Branch of the Royal Asiatic Society* 5/2:366-369.
- 1928 Peluan drinking song. *British North Borneo Herald*, 1 May. (Repr. in Rutter 1929.)
- 1936 Some Murut hunting customs. *Journal of the Malayan Branch of the Royal Asiatic Society* 14/3:307-313.
- 1937 Murut customs; human sacrifice and slavery amongst the Nabai tribe, Keningau. *British North Borneo Herald*, 2 November:236-237.
- 1947 A Murut fairy tale. *Journal of the Malayan Branch of the Royal Asiatic Society* 20:145-152.

- 1953 *Dusun adat: some customs of the Dusuns of Tambunan and Ranau, West Coast Residency, North Borneo.* Native Affairs Bulletin, No.5. Jesselton: Government Printing Department (reprinted).
- 1953 *Kwijau adat: customs regulating inheritance amongst the Kwijau tribe of the Interior.* Native Affairs Bulletin, No.6. Jesselton: Government Printing Department (reprinted).
- 1953 *Murut adat: customs regulating inheritance amongst the Nabai tribe of Keningau and the Timogun tribe of Tenom.* Native Affairs Bulletin, No.2. Jesselton: Government Printing Department (reprinted).
- 1962 *The Timoguns: a Murut tribe of the Interior, North Borneo.* Native Affairs Bulletin, No.1. Jesselton: Government Printing Department (reprinted).

WRIGHT, Leigh

- 1979-80 *The Lanun pirate states of Borneo: their relevance to Southeast Asian history.* *Sabah Society Journal* 6/4:207-217.

WURM, S.A. and T. Basil WILSON

- 1975 *English finderlist of reconstructions in Austronesian languages (post-Brandstetter).* *PL*, C-33.

INDEX

A

Abai Sungai 18,19,40,42,140-145,147,
148,150-152,328,329,330,336

Ambual 162,163,187-189,191-194,219,
330,336; see also Dusun-Murut

Apin-Apin Kuijau 21,22,29,30,40,42,
44,49,159,160,162,163,165-171,187-
189,191-194,215,219,260-264,270,
328,330,336

Austronesian lgs 1,5,41,45,47,51,334

B

Bahasa Malaysia (national lg) 2,5,
10,13,41,44,45,47,53,64,71,72,81,
89-96,98,100,106,107,110,111,121,
133,147,150,151,155,211-214,223,
227,228,234,235,241,242,249,253,
255,256,267-269,274,279-281,294,
296,322-324,332; see also Malay

Bahasa Sama → Bajau, Sama(1)

Bajau F, lg 1,6,7,12-14,40,41,46,101,
106,111,113,114,116,117,121-123,216,
327,334; see also East Coast Bajau
and West Coast Bajau

Bajau Asli → Bajau Semporna

Bajau Balangingi 15,16,46,79,80,114,
117,118,122,123,327-329

Bajau Banaran 13-16,46,79-81,123

Bajau Darat 14-16,46,123

Bajau Kubang → Bajau Semporna

Bajau Laut 14-16,46,113,117-119,123

Bajau Semporna 15,16,46,107,109,110,
114,117-121,123

Bajau Sitangkay → Sama

Bajau Ubian → Ubian

Balangingi → Bajau Balangingi

Bangi 1,6-8,41,46,52,87,328,329,333-
336

Baukan 21-23,25,26,40,42,46,49,65,
128,157,159-163,166,167,170-180,182-
189,194,198-203,205,210,218,220,328-
330,336

Beaufort Dusun 33

Beaufort Kadazan 27,29,32,35,36,49,
300,315

Beaufort Murut 22,24,42,49,65,137,
157,159-162,165-167,176-178,187-189,
195,196,198,216,228,328-330,336

Begahak 9,10,46,47,54,85,87,88,90;
see also Ida'an/Begahak

Bengkoka 283,287,296

Bintaq 195

Bisaya 26-29,43,44,46,195,227-229,
245-251,253-256,286,299,328-331,
335-337

Bolongan 125,127,210

Bornean S 1,6,7,17,42,46,48,87,128,
159,215,327,328,331,334,336,337

Brunei 7,10-13,44,46,47,54,93-97,100,
228

Brunei/Kedayan 10,11,41,46,91,93-100,
327-329,332

Buginese → Bugis

Bugis SF, lg 1,6,7,9,40,41,43,46,51,
57,327,328,333,334

Bukou 195

Buludupi 335,337

Bundu (Tuhan) 287,295

Bungaraya Dusun 32,39,40,49,300

Butuanun 77

Butung 1,5,41,47,51,327,328,333,334

C

Central Dusun 27,32-40,47-49,170,
247,249,253,254,262,263,265-267,
270,274-277,287-289,291,292,294,
298,300-306,308,309,311-315,318,
321,323,331,336

Central Murut SF, lgs 21,22,24,25,
42,44,46-48,159,161,163,165,194,
203,215,216,330

Central Philippine lgs 77

Central Sama 113,117-120,122,123

Central West Coast Bajau 104-106,108

- Chabacano 5,41,47,51,327,328,333, 356
 Chavacano → Chabacano
 Coastal Dusun 283,336
 Coastal Kadazan 247,249,253,263,265,
 274,298,301,302,305-309,311,313-315,
 318,323,331,336
 Cocos Malay 10-12,41,47,91,93-96,98,
 328,329; see also Malay, Bahasa
 Malaysia
- D
 Dabugus 195
 Dalit Murut 201
 Danao F, lgs 69,72,73,327
 Dayak 61,91
 Dumpas 26,27,43,47,141,231-235,286,
 299,328-332,336
 Dusun 5,7,17,27-29,31-40,44,47,49,
 132,149,159,191-193,195,237,239,240,
 245,259,260,264,270,276,283,286,292,
 295-297,299,324,337
 Dusunic F, lgs 6,17,18,26-31,33,34,
 40,43,44,46-49,141,149,158,159,166,
 168-171,191,192,215,225,228,231,
 233,235,237,239,245,247,250-254,
 256,257,260-262,264,265,267,269-
 271,273,274,281,283,286,287,289,
 299,311-313,318,327,329,330-332,
 334-336
 Dusun Kadayan 28,47,237,239
 Dusun-Murut 21,22,24,25,42,47,49,
 159-163,167,171,172,187-189,191,
 193,194,219,328,330,336
 Dusun Segama 18,20,21,47,141-144,
 146-148,151,330
 Dusun Sinulihan 32
 Dyen's classification 333,334
- E
 East Coast Bajau 12-16,41,43,44,46-
 49,106,107,109-111,113-118,121,122,
 327,328,331
 Eastern Kadazan 26,30,31,40,43,44,
 48,49,231,233,234,257,271-281,286,
 299,311,313,314,324,328-332,336
 Eastern Murutic lgs: see p.330 then
 elsewhere
 East Mindanao lgs 77
- G
 Gana 21,22,42,47,65,137,157,159,160,
 162,165-171,187-194,215-218,260,
 262-264,269,270,328-331,336
 Garo 37,38,47,290,292,298,302,306,
 309,315,318,319,321,323,328,329,
 331,336
 Gonsomon 283
 Gum-Gum Sungai 20,21,40; see also
 Sungai
- H
 High Javanese 47
- I
 Iban 10,13,41,47,91,93-95,97,98,100,
 327-329
 Idahan F 46,85,245,283
 Ida'an F, SF, lg 6,7,9,10,40,41,46,
 47,54,85-90,210,216,271,327-329,331,
 334,336,337
 Ida'an/Begahak SF, lg 1,9,10,41,43,
 87,89,328,329; see also Ida'an,
 Begahak
 Ida'an Murut 219
 Ida'an Sungai 40,41,43,87-90,328
 Illanun 1,6-8,41,43,47,67-74,327-329,
 334
 Indo-European lgs 41,47
 Indonesian Murut 157,185,208; see
 also Kolod, Selungai Murut, Sembakung
 Murut
 Inner Sugut Kadazan 37,38,48,49
 Iranun 67,69-74; see also Illanun
- J
 Jama Mapun 15,16,47,114,120,121,123
 Javanese S, F, SF, lg(s) 1,5,6,41,43,
 47,51,57,195,327,333,334; see also
 Jawa
 Jawa 6,40,41,43,47,328; see also
 Javanese
 Jawa Halus 47; see High Javanese
- K
 Kadazan 17,27,29-38,40,44,47,49,149,
 193,240,248,273,275,276,282,286,292,
 297,299,306,336
 Kadazan/Dusun 26,30,31,43,52,54,56,
 149,169,227,228,233,239-243,247-250,
 253,256,257,262,263,265-267,270,273-
 276,281,288,291,292,294,297-302,
 306-311,313,315,318,322-324,328-331,
 336
 Kadazan-Murut 193
 Kadazan-Tagaro 32,33
 Kadazan-Tambanua 26,27,40,43,47,48,
 51,273,286,299,328,330,331,336
 Kagayan 13,15,16,47,114,117,121-123,
 327-329
 Kalabakan (Murut) 21-23,42,47,65,129,
 137,157,159,160,162,165,167,185-189,
 201,205,209-211,215,223,328-331,336
 Kapagalan 198,220
 Kedayan 10-12,46,47,54,93-97,100,195;
 see also Brunei/Kedayan
 Kelabit(ic) (Murut) 61,333
 Kemaloh 61
 Kemaloh-Kelabit 61

- Keningau Dusun 27,32,33,49,191,300
 Keningau Murut 157,185,187,189,191-194,196,198,201,205,215,218,219,328-330,336; see also Nabay
 Keramuak Sungai 275,278,282
 Kimaragang 37,38,47,290,292,298,302,306,309,315,317,318,321,323,328,329,331,336
 Kiundu 32,33,47
 Klias River Kadazan 253,298,302,308,315,316,323,328,329,331,336
 Kolobuan 18,20,21,47,139,153,201,218
 Kolod 21,22,42,47,65,137,157,159,160,162,166,187-189,195,207-209,215,222,328-330,336
 Koroyau 193
 Kota Belud Dusun 27,31,32,34-36,40,48,49,300,321
 Kota Belud Illanun → Illanun
 Kota Kinabalu Kadazan 27,32,35,49,300
 Kuala Monsok Dusun 32,38,39,49,262,267,298,300-302,309,311,323,331,336
 Kuamut Sungai 40
 Kuijau 22,26,27,29,30,40,43,44,47,49,160,165,167,169-171,189,191-193,215,216,257-270,273,286,299,311,313,314,324,328-331,336
 Kulu-Kulu Sungai 275,278,282
 L
 Labuk 287,296
 Labuk Kadazan 273-278,281,282; see also Eastern Kadazan
 Labuk/Mangkaak/Sukang 273-278,281,336; see also Eastern Kadazan, Mangkaak, Sukang
 Lahad Datu Illanun → Illanun
 Lamag Sungai 272,274-278,281,282,336
 Laminusa 13,15,16,47,123
 Lingkabau 18,42,47,140-152,328-330,336
 Lobu 18,19,42,47,49,140-149,151,152,328-330,336
 Lotud 26-28,43,44,47,49,237-243,286,296,299,313,314,324,328-330,336
 Lumbis Murut 164,195,205,206,208,215,221,222,336
 Lun Bawang 61,63
 Lundayeh 1,6,7,41,47,54,59,60-65,167,208,216,222,327-329,333,334
 Lun Lod 7,47,61,62
 M
 Magindanao 67,69-74
 Makiang 17,20,21,40,47,48,139,143,144,146-148,153,159,201,218
 Malapi Kadazan 274,275,277,278,281,282,336
 Malay 71,91,93,98,100,121,130,168,247,263,273,326,328,332,333; see also Cocos Malay, Bahasa Malaysia
 Malayic S, F, SF, lgs 1,6,7,10-13,41,44,46,47,54,64,91,92,94-96,327,332-334
 Maligan Tagal 162,164,187-189,203-207,215,330,336
 Mangkaak 30,31,48,272-278,281,282,286,287,296,299; see also Eastern Kadazan
 Mapan-Mapan West Coast Bajau 105,108; see also West Coast Bajau
 Mapun → Kagayan
 Maranao 8,67,69-74
 Marudu 283,336
 Memogun, Momogun 283,295; see also Rungus
 Menindal, Menindaq → Kuijau
 Minansut 189,191,219,260,336
 Minokok 37,48,298,302,308,309,324; see also Sugut Kadazan
 Molbok 8,333
 Murut 22-25,40,48,49,54,59,61-65,129,131-133,157-160,166,168,171,172,174,177,187,189,191-194,198,200,201,203,205-207,211,216-219,221-223,225,227,245,259,270,333-335,337
 Murutic F, SF, lgs 6,17,21-26,29,30,40,42,44,46-49,59,61-63,65,128-130,132,137,155,156,158-162,165-174,176,178,179,181-189,192,203,208-211,215,216,218,222,225,228,257,259-262,264,269,327,329-332,334-336
 N
 Nabay 17,21,22,24-26,40,42,46,48,65,137,159-163,166,167,169-180,182-189,191-194,198,201,203,205,207,211,215,218-220,328,330,336; see also Keningau Murut
 National language; see Bahasa Malaysia, Malay
 Nonukan 127
 Northern Murut SF 162,185,187,189,194,208,215,219,222,223,330
 North-western Austronesian SS 1,5-7,41,46-48,61,62,77,84,87,128,327,328,333,334
 Nulu 283
 O
 Outer Sugut Kadazan 37,38,47,49,287-292,294,309,315

- P
- Paginatan Dusun 32, 37, 38, 49, 270, 300
- Pahu 32, 33, 48
- Paitan 19, 48, 141, 335, 336
- Paitanic F, lgs 6, 17, 18, 26, 40, 42, 43, 46-48, 139-147, 149-152, 159, 215, 218, 231-235, 271, 296, 327, 329-332, 334-336
- Paluan 21, 22, 24, 42, 48, 65, 137, 157, 159-163, 167, 176, 178-180, 182, 185, 187-189, 195, 201-203, 205, 207, 215, 217, 220, 221, 328-330, 336
- Pandewan Murut 162, 163, 188, 189, 201-203, 221, 330, 336
- Papar 26, 27, 43, 48, 225-229, 259, 286, 299, 328-330, 336
- Pemilaan Dusun 32, 39, 49, 300
- Penampang 287, 295; see also Coastal Kadazan
- Penchangan 127
- Pensiangan Murut 162, 164, 185, 187-189, 203-208, 210, 211, 215, 222, 223, 330, 336
- Philippine lgs 12, 43, 46, 48, 69, 70, 77, 116-120, 327, 335
- Pilantong Kadazan 27, 32, 33, 49, 300
- Poros 220
- Potok 61
- Prentice classification 336
- R
- Ranau 285, 287, 295, 336
- Rumanau 17, 19-21, 48, 139, 143, 144, 147, 149, 153, 218
- Rundum Murut 162, 164, 187-189, 203-206, 215, 222, 330, 336
- Rungus 17, 26, 27, 31, 35, 36, 40, 43, 48, 237, 273, 283-296, 299, 311, 313-315, 318, 324, 328-330
- S
- Sabah Malay → Malay
- Salalir 125, 127, 187-189, 195, 203-208, 215, 221, 222, 330, 336
- Sama(1) lgs 12-16, 43, 48, 101, 114, 120, 123, 327
- Sama-Bajau 77, 79, 81-84; see also Bajau, Sama(1)
- Sama Balangingi 79, 81; see also Bajau Balangingi
- Sandiwar 195, 220
- Sapulut Murut 184, 203, 205
- Sa'ban 59
- Sarawak Malay → Malay
- Sedalir → Salalir
- Segaliud Sungai 275, 278, 282
- Selungai Murut 157, 162, 184, 185, 187-189, 208-210, 215, 217, 218, 223, 328-330, 336
- Sembakung 125, 127, 128, 130; see also Sembakung Murut
- Sembakung Murut 21-23, 42, 48, 65, 129, 130, 137, 157, 159, 160, 162, 166, 187-189, 208-211, 215, 223, 328-331
- Sarudung Murut 21-23, 42, 49, 65, 132, 137, 157, 159, 160, 162, 165-167, 186-189, 201, 205, 209-211, 215, 217, 223, 328-330, 336
- Sesayap 125, 127, 132
- Sibutu 15, 16, 48, 114, 123
- Sikubung 15, 48, 114, 119, 123
- Sikubung Semporna 118-120
- Simunul 15, 16, 48, 114, 117-121, 123
- Sinabu 18, 20, 21, 48, 139, 143, 144, 146-149, 153
- Sinabu/Rumanau 20, 48
- Sinarupa 19, 20, 47, 48, 139, 153
- Sinimbitan 193, 270
- Sinulihan 48; see also Dusun Sinulihan
- Sokid 32, 33, 48
- Sook Murut 21, 22, 24, 25, 42, 49, 159-163, 167, 176-180, 201, 217, 328, 330
- Southern Bisayan lgs 77
- Southern Sama(1) 113, 116-118, 120-123, 327-329
- Sugut Kadazan 27, 30, 32, 37, 49, 286, 298, 300-302, 305, 306, 308-310, 318, 321, 323, 324, 331, 336
- Sukang 30, 31, 48, 272-278, 282
- Suluk 1, 6-9, 12, 13, 41, 44, 48, 75-84, 113, 327-329, 334
- Sumambu/Tagal 157, 161, 162, 164, 187, 189, 201-207, 215, 221, 222, 328-330, 336; see also Tagal
- Sungai 9, 10, 19, 20, 26, 30, 31, 40, 47, 48, 139, 146-148, 210, 273, 275, 276
- T
- Tagal 17, 21-23, 42, 44, 48, 49, 61, 63, 65, 137, 158-160, 165-167, 169-187, 193, 200, 201, 203, 205-207, 216-218, 221, 222, 328, 330; see also Sumambu/Tagal
- Tagaro 48; see also Kadazan-Tagaro
- Takapan 21, 22, 24, 42, 48, 159-161, 165, 167, 176, 178-180, 182-185, 201, 217, 328, 330
- Tambanua 17, 19, 27, 40, 42, 44, 48, 139, 141-148, 150, 151, 231, 233, 234, 296, 328-330, 336
- Tambanua Sungai 19, 40; see also Tambanua
- Tambunan 287, 295, 299, 336
- Tangara → Tenggara
- Tanggaraq → Tenggara
- Tarakan 125, 127, 128, 132, 210, 336

- Tatana 26-29,43,44,48,225,227-229,
248,250-256,286,299,311,313-315,
324,328-331,335,336,359
- Tausug 8,9,13,48,75,77-79,81,83,84,
113,327,328,334; see also Suluk
- Tawau Murut 211,215
- Tebilung 32,33,35,36,48,298,302,306,
315,318,320,321,323,328,329,331,
336
- Tempasuk (Dusun) 27,32,34,49,298,
300,302,324,336
- Tengara 127,187-189,198-201,210,215,
220,330,336
- Terusan Sapi Kadazan 273-275,278,
282
- Tidong 17,42,43,48,125-133,137,159,
165-167,186,187,201,208,210,211,
215,259,327-331,334-336
- Tilau-Ilau 37,48
- Timugon (Murut) 17,21,22,24,42,48,
65,137,157-163,166,167,170-180,
182-184,187-189,191,194,195,197,198,
200,201,203,205-207,220,328-330,336
- Tindakan 272,275
- Tindal 32,33,49,259
- Tinggalan → Sembakung
- Tomani Sumambu 162,164,187-189,203-
207,215,330,336
- Tombovo Kadazan 27,32,34,49,300
- U
- Ubian 14-16,49,120,123
- Ulu Tuaran 32,49,285,292
- Upper Kinabatangan 18-21,40,42,44,
47,48,140-153,201,218,328-330,336
- V
- Visaya 28,247
- W
- West Coast Bajau 12-14,26,41,43,44,
101,103-112,114,116,117,119-122,
225,327-329
- West Coast Ubian 120; see also
Ubian

