PART I

THE LANGUAGES OF SABAH:
A TENTATIVE LEXICOSTATISTICAL CLASSIFICATION

Kenneth D. Smith

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

© 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 wordlist[6] 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.
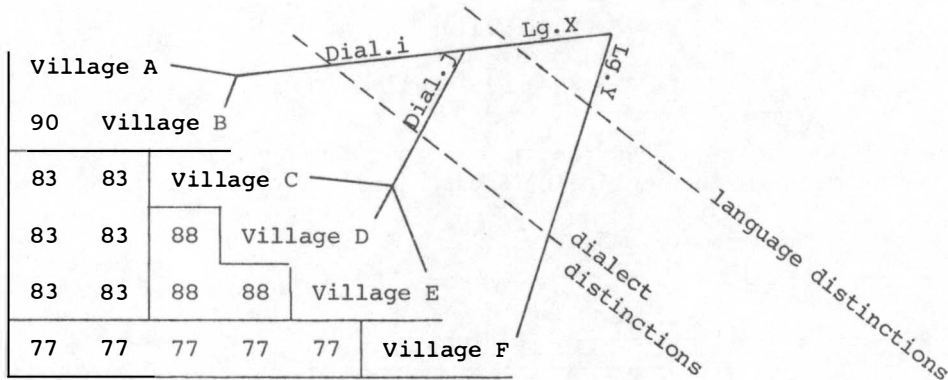

Figure 1: Hypothetical PSC of six villages forming consistent grouping
          of villages into dialects and dialects into languages. (In
          this and the following figures the intersection of a column
          and a row indicates the PSC between the two villages; for
          example, Village 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.
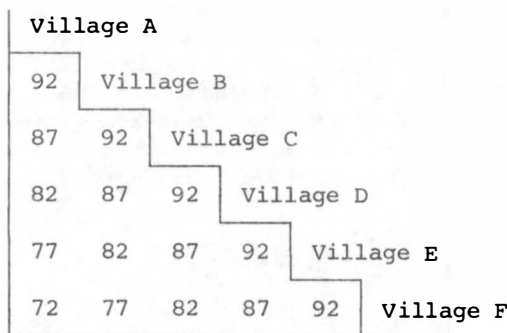
```
Village A
    ┌──────────
 92 │ Village B
    │    ┌──────────
 87 │ 92 │ Village C
    │    │    ┌──────────
 82 │ 87 │ 92 │ Village D
    │    │    │    ┌──────────
 77 │ 82 │ 87 │ 92 │ Village E
    │    │    │    │    ┌──────────
 72 │ 77 │ 82 │ 87 │ 92 │ Village F
```

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

```
                          Dialect i        Dialect j      Dialect k
                                                                        Dialect l
 Village A   Village B   Village C   Village D   Village E   Village F
```
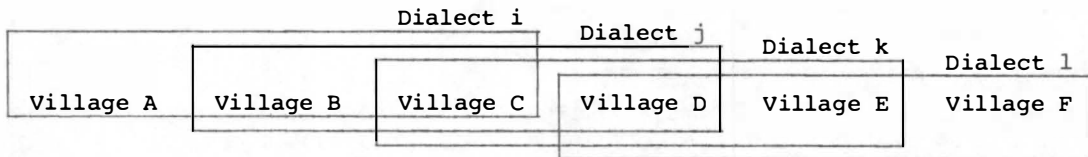
Figure 3: Hypothetical dialect chain derived from Figure 2

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 re-
placement) 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 revis-
iting 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.

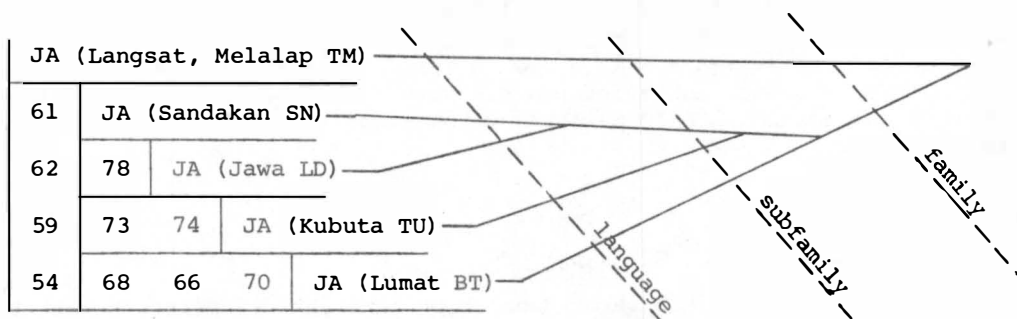| JA (Langsat, Melalap TM)				
61	JA (Sandakan SN)			
62	78	JA (Jawa LD)		
59	73	74	JA (Kubuta TU)	
54	68	66	70	JA (Lumat BT)

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.)

```
LN (Mendolong SG)
26  BO (Kg Bugis LD)
28  36  BI (Kg Brunei BT)
28  36  46  BU (Mapan-Mapan PS)
26  31  37  43  SK (Pulau Lubokan SN)
25  35  36  38  40  IN (Rampayan Laut KB)
25  31  36  38  36  38  ID (Segangan LD)
26  33  41  42  36  36  45  BG (Lok Agong KT)
29  30  35  36  34  34  41  44  DN (Bongkud RU)
```
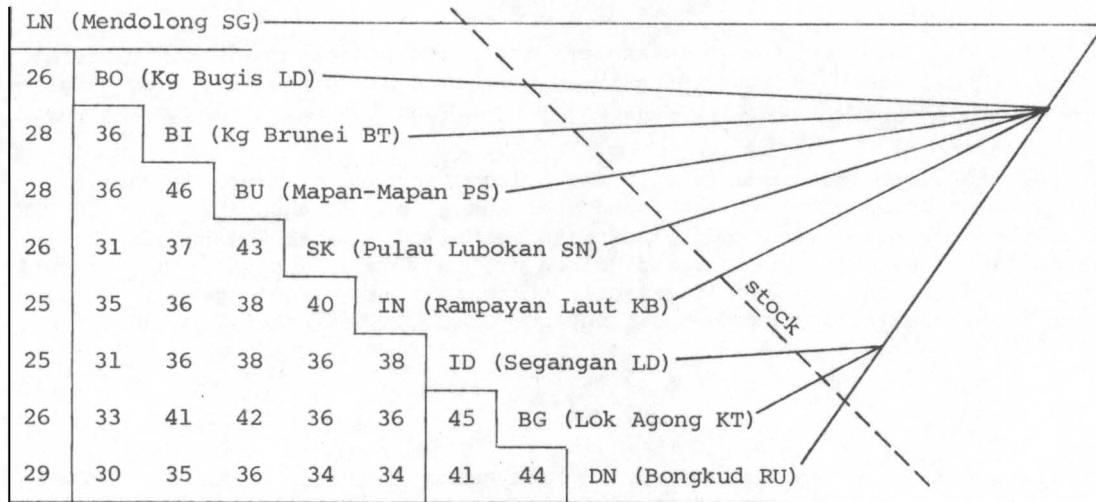
Figure 6: PSC of languages representing the nine linguistic stocks of the North-western Austronesian superstock in Sabah. (LN = Lundayeh language; BO = Bugis language, representing Bugis subfamily; BI = Brunei language, representing the Malayic family; BU = Bajau, West Coast, representing the Bajau family; 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.

```
LN (Kuala Punti KU)
88  LN (Mendolong SG)
88  89  LL (Kawang SG)
88  87  88  LN (Baru Jumpa TM)
```

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.

IN (Barigas LD)				
93	IN (Nala LD)			
81	80	IN (Rampayan KB)		
77	78	89	IN (Kulambai KB)	
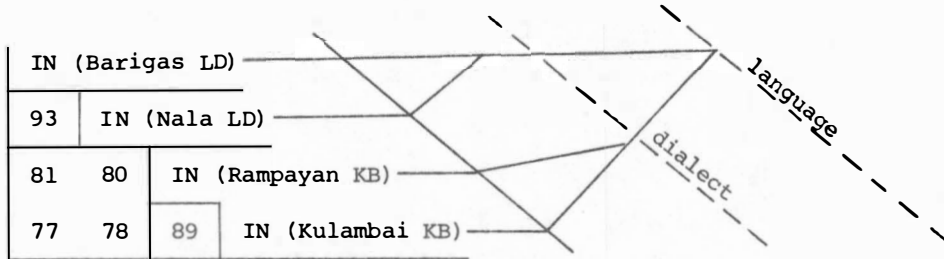
Figure 8: PSC of four samples, two dialects of Illanun. (IN = Illanun.)

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 break-down in communication with such differing vocabularies must be determined from intelligibility testing; see Figure 10.

```
BO (Kg Bugis LD)
┌──────┐
│  81  │ BO (Parpadoam SN)
│      ├──────┐
│  76  │  79  │ BO (Muhibbah Raya TU)
└──────┴──────┘
```
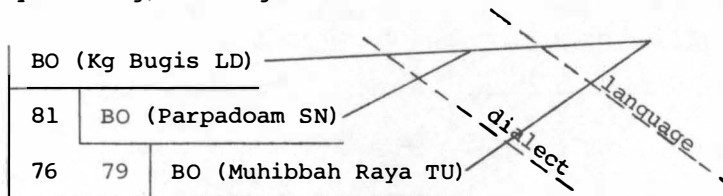
Figure 10: PSC of three samples of the Bugis subfamily. (BO = Bugis.)

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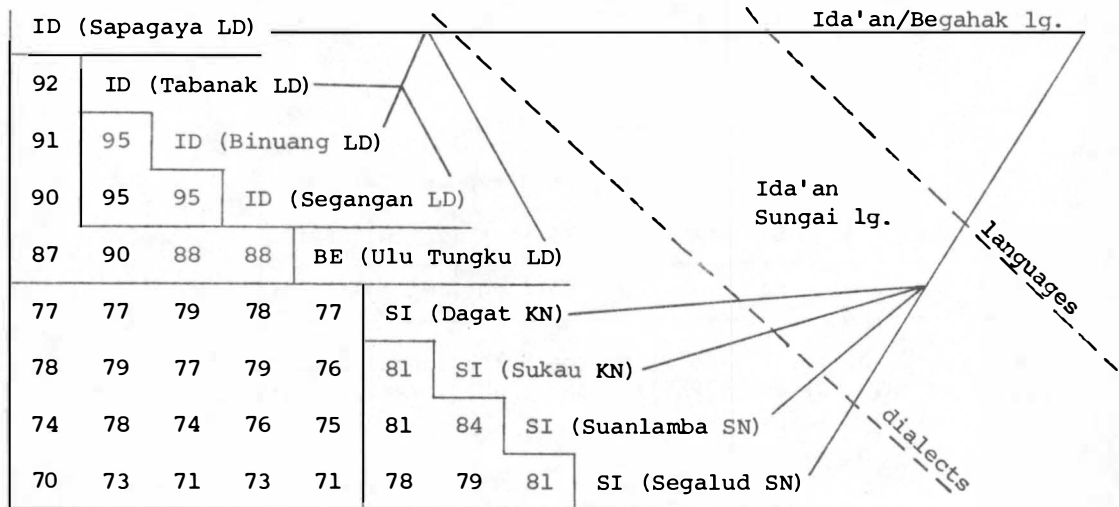
Figure 11: PSC of languages and dialects within the Ida'an subfamily.
(ID = Ida'an; BE = Begahak; SI = Sungai.)

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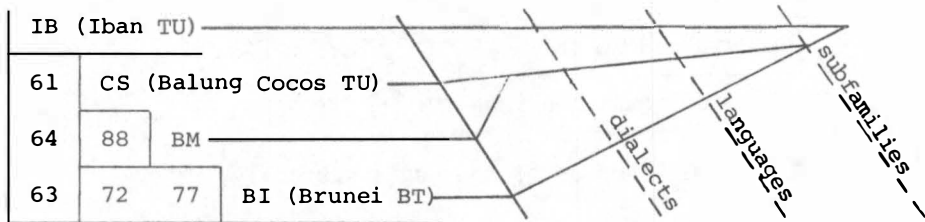
Figure 12: PSC relations of sample Sabah languages within the Malayic family. (IB = Iban; CS = Cocos Malay; BM = Bahasa Malaysia; BI = Brunei.)

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 Sibu, 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.)

```
KY (Lembaga Lumadan BT)
   |
 89 | KY (Layang-Layangan LN)
   |      |
 89 | 91 | BI (Seladan-Tarap PR)
   |      |      |
 83 | 88 | 88 | BI (Kolapis LS)
   |      |      |      |
 79 | 83 | 85 | 86 | BI (Berhala Darat SN)
   |      |      |      |
 74 | 80 | 81 | 83 | 93 | BM
   |      |      |      |      |
 70 | 73 | 74 | 77 | 83 | 88 | CS (Balung Cocos TU)
   |      |      |      |      |      |
 67 | 69 | 71 | 73 | 79 | 82 | 91 | CS (Cocos LD)
```

Figure 13: PSC of representative dialects and subdialects within the Malayic
        subfamily. (KY = Kedayan; BI = Brunei; BM = Bahasa Malaysia;
        CS = Cocos Malay.)

2.3.8 Bajau family

    The Bajau family of languages in Sabah comprises two distinct Bajau lan-
guages, each consisting of a chain of Bajau dialects. The dialect chain of both
languages consists of a series of links not less than 75 PSC - an admittedly low
threshold. The most remote parts of the dialect chain have lower PSC figures
with each other, though in some cases one may be closer linguistically to parts
of the other chain than to the remoter parts of its own chain of dialects. The
larger, more complex chain is the East Coast Bajau language which has a closer
tie to the southern Philippine Sama languages. The other chain is the West Coast
Bajau language. No wordlist of one language is within 75 PSC of any wordlist in
the other language; each language is thus an independent dialect chain without
an interconnecting link in this data. (Several links between the East Coast
Bajau and West Coast Bajau languages could be established at the 70-71 PSC
level.)

    The two Bajau languages have been influenced differently because of their
different geographical neighbours. East Coast Bajau has borrowed from Suluk,
the trade language of the Sulu Sea area where the Bajau fish and trade; whereas
the West Coast Bajau have borrowed from the Malayic languages further south on
the west coast of Borneo. Three representative dialects each of Suluk, East
Coast Bajau, West Coast Bajau and the Malayic languages are compared in Figure
14. In this figure the distinctness of the three linguistic stocks is shown in
the 32-47 PSC 'L-shaped' box; the borrowing by the East Coast Bajau and West
Coast Bajau from Suluk and Malayic languages, respectively, is shown in the 45-
57 PSC and 41-53 PSC boxes; and the yet closer relation of East Coast Bajau and
West Coast Bajau is shown in the 55-68 PSC box.

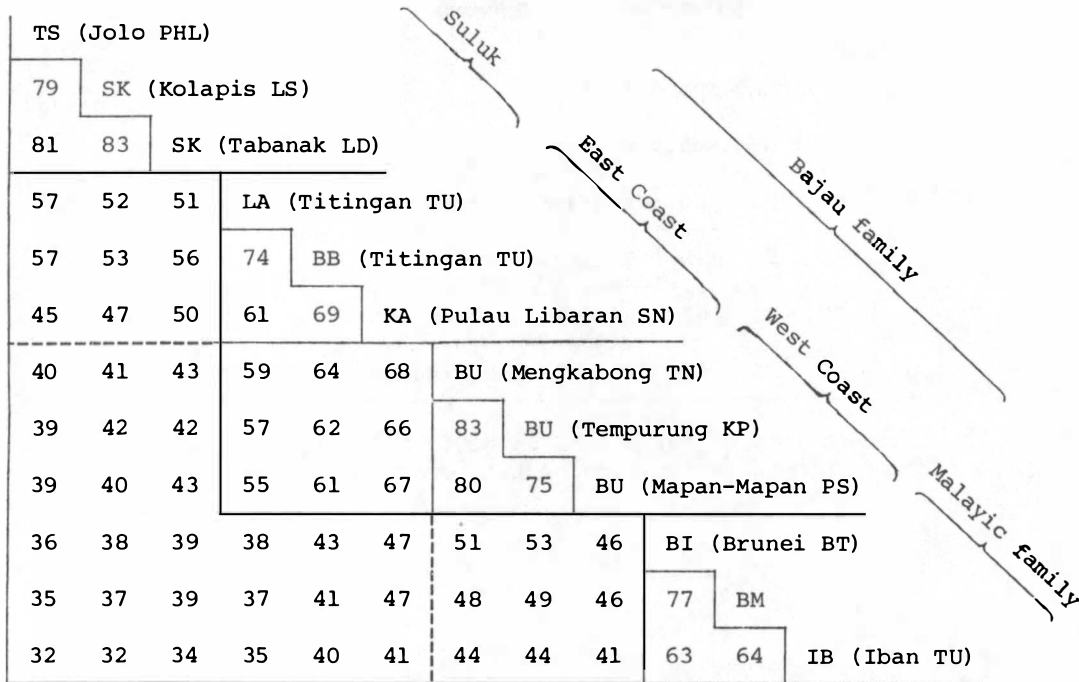| TS (Jolo PHL)										
79	SK (Kolapis LS)									
81	83	SK (Tabanak LD)								
57	52	51	LA (Titingan TU)							
57	53	56	74	BB (Titingan TU)						
45	47	50	61	69	KA (Pulau Libaran SN)					
40	41	43	59	64	68	BU (Mengkabong TN)				
39	42	42	57	62	66	83	BU (Tempurung KP)			
39	40	43	55	61	67	80	75	BU (Mapan-Mapan PS)		
36	38	39	38	43	47	51	53	46	BI (Brunei BT)	
35	37	39	37	41	47	48	49	46	77	BM
32	32	34	35	40	41	44	44	41	63	64	IB (Iban TU)

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.

```
BU (Jawi-Jawi KB)
┌─────
│ 93 │ BU (Mengkabong TN)*
│ 89   93 │ BU (Kuala PR)
│ 88   89   93 │ BU (Pengalat Besar PR)
│ 90   92   91   90 │ BU (Meruntum PG)
│ 91   91   90   87   91 │ BU (Serusup TN)
│ 89   91   90   87   88   87 │ BU (Numbak KK)
```

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

```
BU (Kawang PR)
┌─────
│ 83 │ BU (Tempurung KP)
│ 78   77 │ BU (Kulambai KB)
│ 86   83   84 │ BU (Mengkabong TN)*
│ 81   77   81   86 │ BU (Melalap TM)
│ 75   78   73   80   77 │ BU (Baru SN)
│ 76   75   75   80   77   74 │ BU (Mapan-Mapan PS)
│ 69   70   68   73   70   74   75 │ BU (Kolapis LS)
```

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 lan-
guages 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.

```
SB (Makuau LD)
┌──
│86│ SB (Kg Air SA)*
├──┼──
│72│72│ BB (Titingan TU)*
├──┼──┼──
│74│77│79│ BD (Dasar LD)*
├──┼──┼──┼──
│76│78│81│83│ SN (Terusan Tengah SA)
├──┼──┼──┼──┼──
│79│79│82│83│91│ UN (Terusan Baru SA)*
├──┼──┼──┼──┼──┼──
│82│79│79│80│84│89│ SN (Telibas LD)
├──┼──┼──┼──┼──┼──┼──
│77│76│77│78│84│87│86│ UN (Bakapit LD)
├──┼──┼──┼──┼──┼──┼──┼──
│75│77│77│81│80│82│79│77│ BL (Bangau-Bangau SA)
├──┼──┼──┼──┼──┼──┼──┼──┼──
│74│78│79│83│84│84│80│79│86│ BS (Semporna SA)*
└──┴──┴──┴──┴──┴──┴──┴──┴──
```
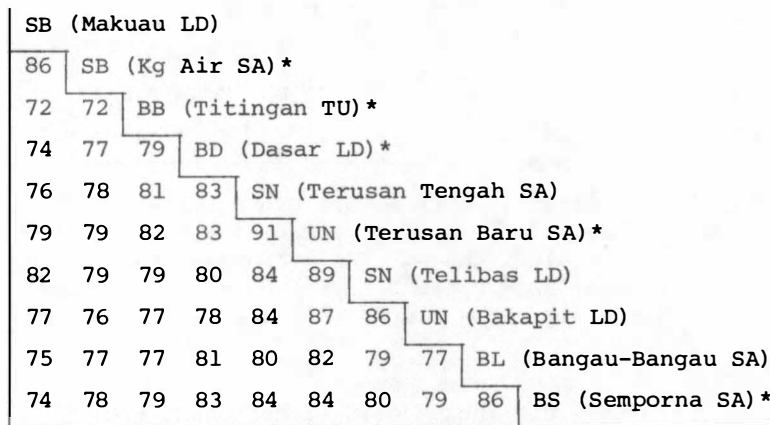
Figure 17: PSC of ten samples of a core group of the East Coast
           Bajau dialect chain. (* identifies villages repeated
           in Figure 18; SB = Sibutu; BB = Bajau Banaran; BD =
           Bajau Darat; SN = Sikubung; UN = Ubian; BL = Bajau Laut;
           BS = Bajau Semporna.)

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]

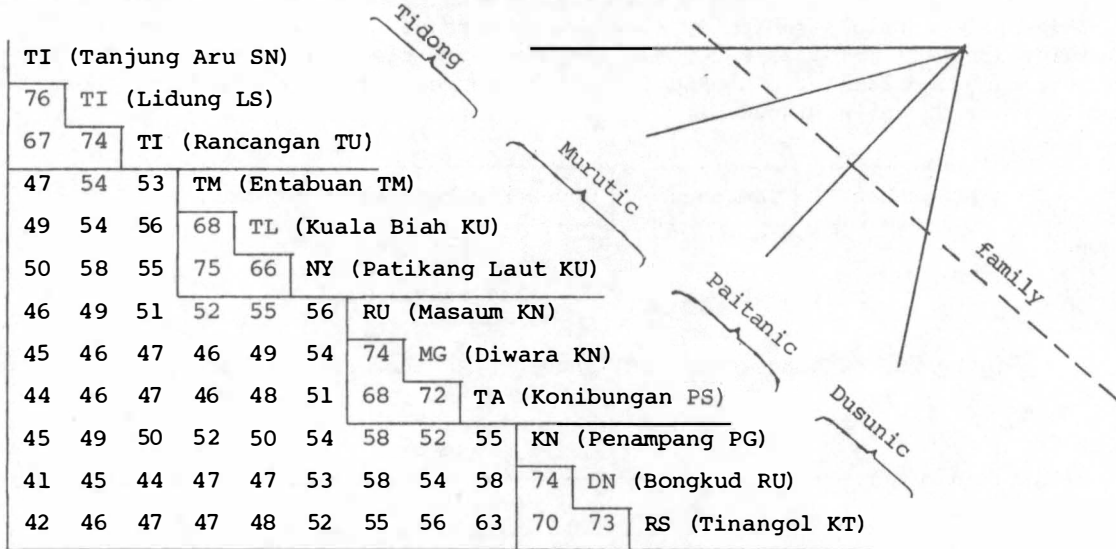| TI (Tanjung Aru SN)										
76	TI (Lidung LS)									
67	74	TI (Rancangan TU)								
47	54	53	TM (Entabuan TM)							
49	54	56	68	TL (Kuala Biah KU)						
50	58	55	75	66	NY (Patikang Laut KU)					
46	49	51	52	55	56	RU (Masaum KN)				
45	46	47	46	49	54	74	MG (Diwara KN)			
44	46	47	46	48	51	68	72	TA (Konibungan PS)		
45	49	50	52	50	54	58	52	55	KN (Penampang PG)	
41	45	44	47	47	53	58	54	58	74	DN (Bongkud RU)
42	46	47	47	48	52	55	56	63	70	73	RS (Tinangol KT)

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 "coastline 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 from 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 rela- tionships 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.
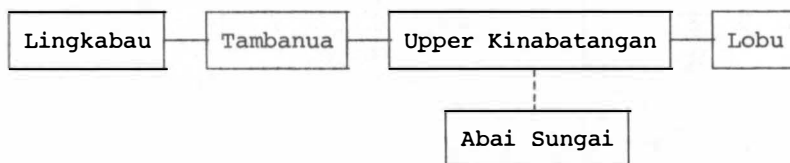
Figure 20: Language chain of languages in the Paitanic family

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: Tampat 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.

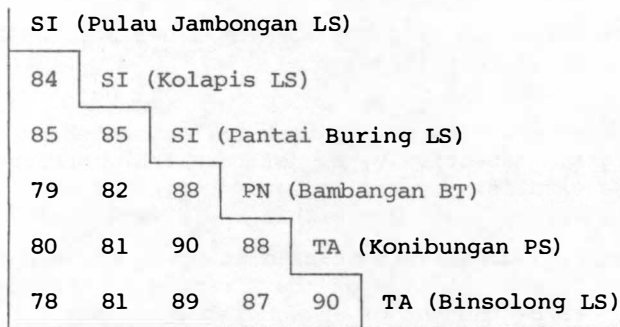|					SI (Pulau Jambongan LS)
84					SI (Kolapis LS)
85	85				SI (Pantai Buring LS)
79	82	88			PN (Bambangan BT)
80	81	90	88		TA (Konibungan PS)
78	81	89	87	90	TA (Binsolong LS)

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).

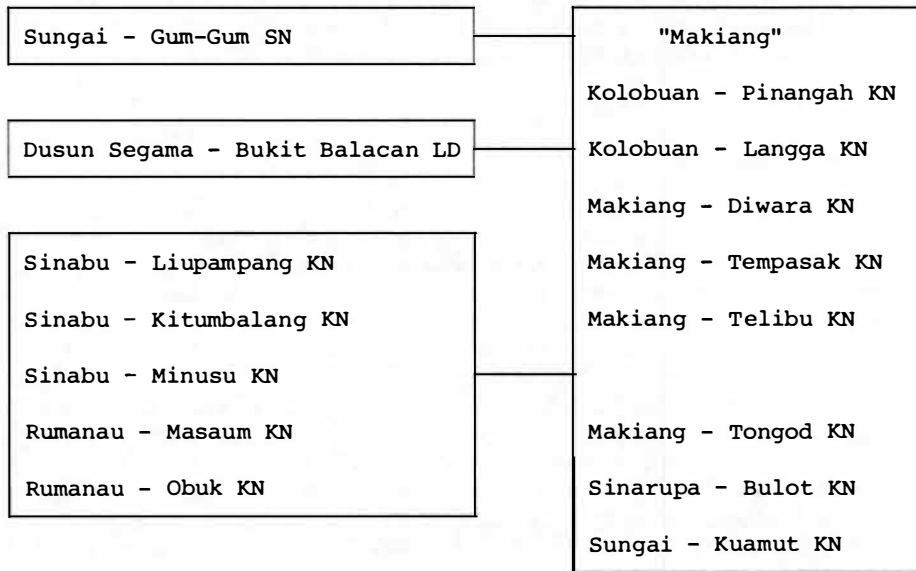|	
Sungai - Gum-Gum SN	"Makiang"
	Kolobuan - Pinangah KN
Dusun Segama - Bukit Balacan LD	Kolobuan - Langga KN
	Makiang - Diwara KN
Sinabu - Liupampang KN	Makiang - Tempasak KN
Sinabu - Kitumbalang KN	Makiang - Telibu KN
Sinabu - Minusu KN	
Rumanau - Masaum KN	Makiang - Tongod KN
Rumanau - Obuk KN	Sinarupa - Bulot KN
	Sungai - Kuamut KN

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.

```
DA (Bukit Balacan LD) ──────────────────────── Dusun Segama

75 │ SI (Gum-Gum SN) ──────────────────────── Gum-Gum Sungai

85   85 │ KB (Pinangah KN) ──────────────────── "Makiang"

69   74   80 │ SU (Liupampang KN)

72   77   84 │ 93 │ SU (Kitumbalang KN)

72   76   83 │ 94   94 │ SU (Minusu KN)

65   71   78   88   85   88 │ RU (Obuk KN)

62   68   74   85   81   84   95 │ RU (Masaum KN)
```
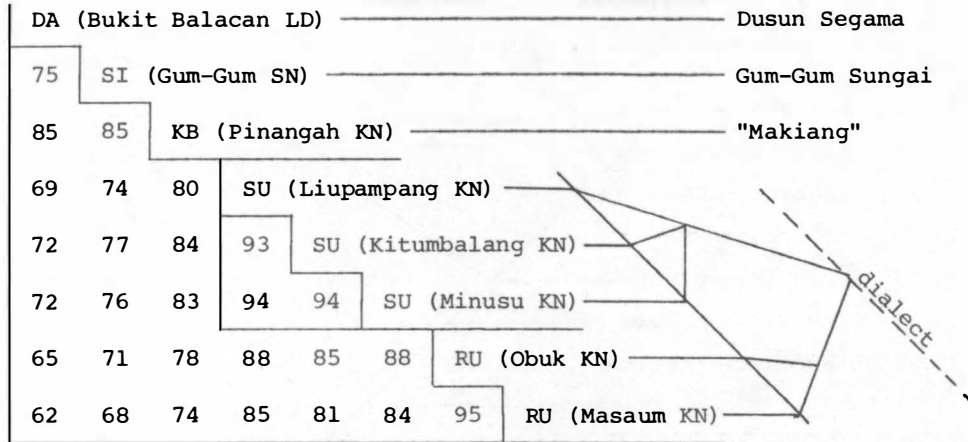
Figure 24: PSC relations of dialects of the Upper Kinabatangan language.
(DA = Dusun Segama; SI = Sungai; KB = Kolobuan; SU = Sinabu;
RU = Rumanau.)

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 southeast 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.
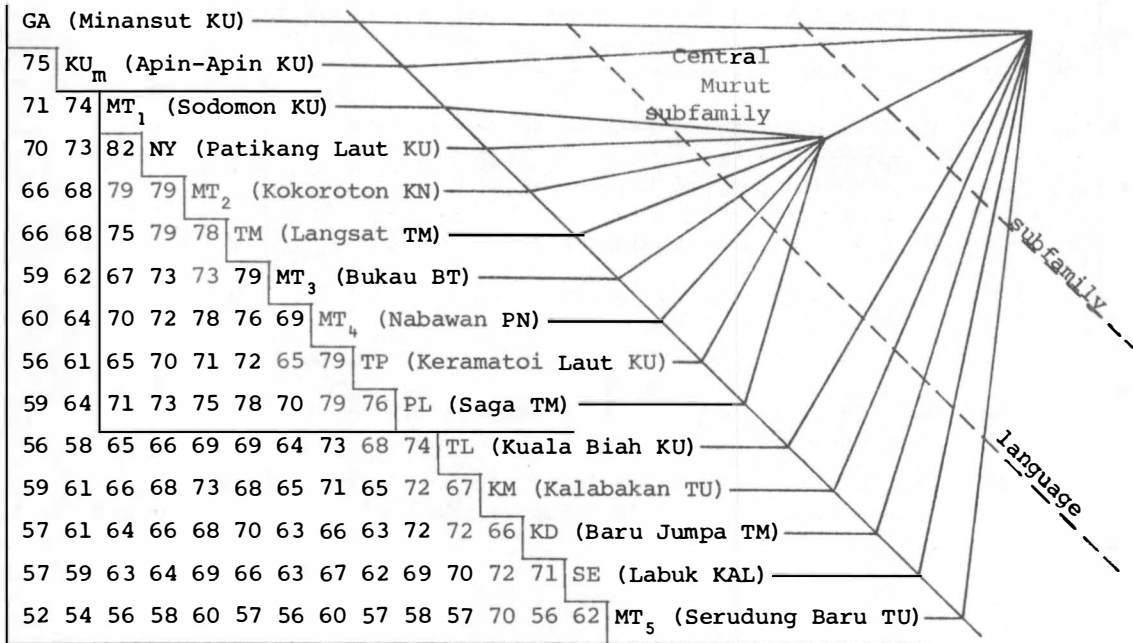
```
GA (Minansut KU) ──────────
75 │ KUₘ (Apin-Apin KU) ──────
71 74 │ MT₁ (Sodomon KU) ──────
70 73 82 │ NY (Patikang Laut KU) ──
66 68 79 79 │ MT₂ (Kokoroton KN) ──
66 68 75 79 78 │ TM (Langsat TM) ──────
59 62 67 73 73 79 │ MT₃ (Bukau BT) ──────
60 64 70 72 78 76 69 │ MT₄ (Nabawan PN) ──────
56 61 65 70 71 72 65 79 │ TP (Keramatoi Laut KU) ──
59 64 71 73 75 78 70 79 76 │ PL (Saga TM) ──────
56 58 65 66 69 69 64 73 68 74 │ TL (Kuala Biah KU) ──────
59 61 66 68 73 68 65 71 65 72 67 │ KM (Kalabakan TU) ──────
57 61 64 66 68 70 63 66 63 72 72 66 │ KD (Baru Jumpa TM) ──────
57 59 63 64 69 66 63 67 62 69 70 72 71 │ SE (Labuk KAL) ──────
52 54 56 58 60 57 56 60 57 58 57 70 56 62 │ MT₅ (Serudung Baru TU) ──
```

Figure 25: PSC relations of representative dialects of 15 Murutic languages.
(GA = Gana; KUₘ = Apin-Apin Kuijau (subscriptₘ designates this Murutic language in contrast to KU which designates the Dusunic Kuijau language); MT₁ = "Murut" representing Dusun-Murut; NY = Nabay; MT₂ = "Murut" representing Baukan; TM = Timugon; MT₃ = "Murut" representing Beaufort Murut; MT₄ = "Murut" representing Sook Murut; TP = Takapan; PL = Paluan; TL = Tagal; KM = Kalabakan Murut; KD = Kolod; SE = Sembakung Murut; MT₅ = "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 (KUₘ) 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 (MT_5) 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, MT_6) 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.
(MT_6 = 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 (MT_3) 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 (MT_1) 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.

```
      DM (Liau Laut KU)
   ┌─────┐
   │ 80  │ MT₁ (Ambual KU)
   ├─────┼─────┐
   │ 82  │ 87  │ MT₁ (Sodomon KU)
   └─────┴─────┘
```

Figure 27: PSC relations of three villages comprising the Dusun-Murut
language. (DM = Dusun-Murut; MT_1 = "Murut" representing
Dusun-Murut.)

2.3.9.3.8(f) Sook Murut language

The Sook Murut language (MT_4) 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 (MT_2) 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 else-
where 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).

```
      MT₂ (Kokoroton KN)
   ┌─────┐
   │ 89  │ MT₂ (Inarad KN)
   ├─────┼─────┐
   │ 85  │ 79  │ BN (Tulid KU)
   └─────┴─────┘
```

Figure 28: PSC relations of three villages comprising the Baukan
language. (MT_2 = Murut, representing Baukan; BN =
Baukan.)

2.3.9.3.8(h) Nabay language

The Nabay language (NY) is here represented by seven wordlists. The vil-
lages 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
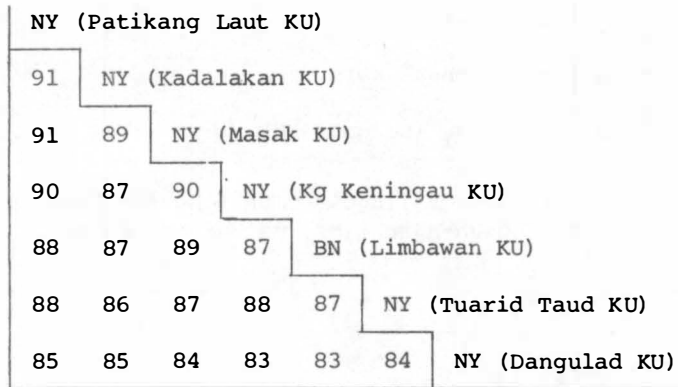be Baukan, lexically it is seen here to be equivalent to Nabay.

```
NY (Patikang Laut KU)

91 │  NY (Kadalakan KU)

91   89 │  NY (Masak KU)

90   87   90 │  NY (Kg Keningau KU)

88   87   89   87 │  BN (Limbawan KU)

88   86   87   88   87 │  NY (Tuarid Taud KU)

85   85   84   83   83   84 │  NY (Dangulad KU)
```

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.

```
PR (Tinambak KP)
66 BA (Kerukan KP)
53 55 DS (Perancangan LS)
72 67 55 TT (Bundu KP)
55 62 59 61 DN₁ (Mangkaladoi TN)
55 63 68 58 67 RS (Liyu PS)
51 54 69 56 62 71 KT (Lubang Buaya LS)
52 58 67 61 68 69 67 KU (Sungoi KU)
54 58 75 62 68 77 78 77 MK (Mananam KN)
54 61 66 62 70 75 76 76 80 KN₁ (Melinsau LS)
55 62 61 63 72 71 68 72 70 | 76 KN₂ (Pilantong LS)
61 66 63 73 74 72 68 71 73 | 75 78 KN₃ (Tahak BT)
56 66 64 73 72 73 70 77 77 | 81 81 84 KN₄ (Kogopon PR)
53 59 61 64 72 68 67 71 74 | 77 75 75 84 KN₅ (Natai KK)
52 60 60 64 69 71 68 68 70 | 76 75 74 82 74 DN₂ (Rosok KB)
52 62 55 59 68 69 66 64 66 | 71 69 71 75 69 82 DN₃ (Tempasuk KB)
56 62 61 67 71 69 65 72 72 | 74 80 79 80 75 74 69 KN₆ (Tombovo PG)
52 60 61 63 69 66 65 72 69 | 72 75 74 80 76 76 69 73 DN₄ (Marapok KU)
```
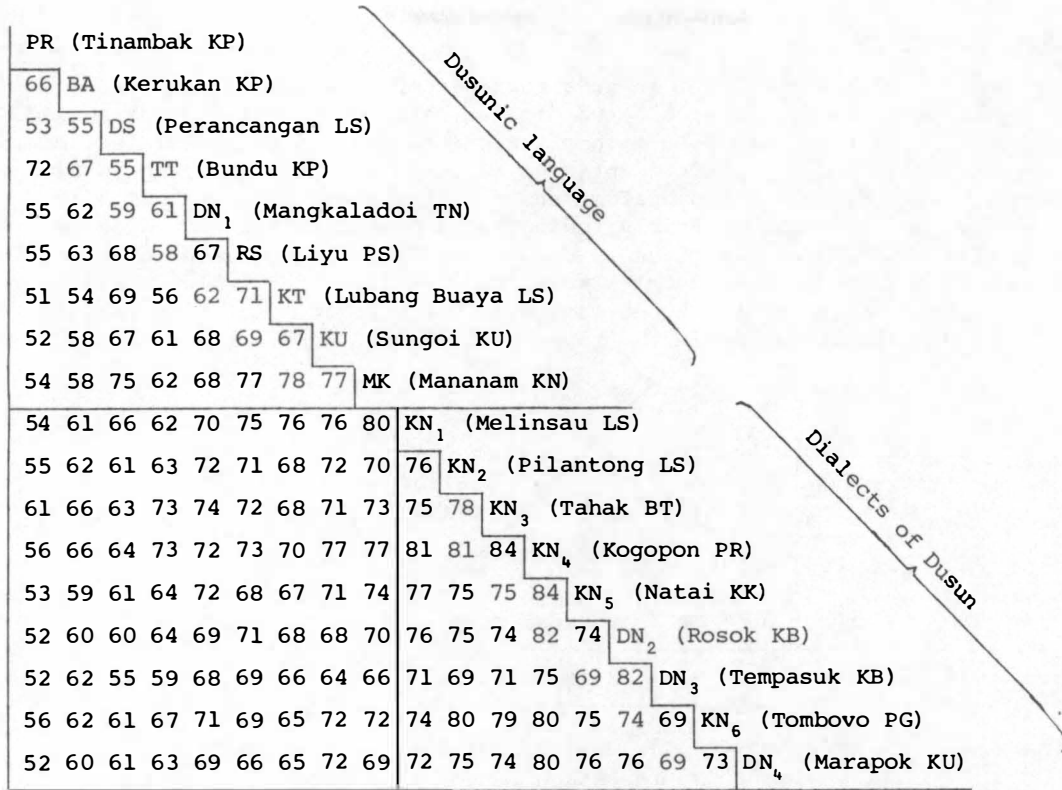
Figure 30: PSC relations of nine Dusunic languages including some dialects of the Dusun language. (PR = Papar; BA = Bisaya; DS = Dumpas; TT = Tatana; DN_1 = "Dusun" representing Lotud; RS = Rungus; KT = Kadazan-Tambanua; KU = Kuijau; MK = "Mangkaak" representing Eastern Kadazan dialect; KN_1 = "Kadazan" representing Sugut Kadazan dialect; KN_2 = "Kadazan" representing Pilantong Kadazan dialect; KN_3 = "Kadazan" representing Beaufort Kadazan dialect; KN_4 = "Kadazan" representing Central Dusun dialect; KN_5 = "Kadazan" representing Kota Kinabalu Kadazan dialect; DN_2 = "Dusun" representing Kota Belud Dusun dialect; DN_3 = "Dusun" representing Tempasuk (KB) Dusun dialect; KN_6 = "Kadazan" representing Tombovo (PG) Kadazan dialect; DN_4 = "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 (DN$_1$, 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.
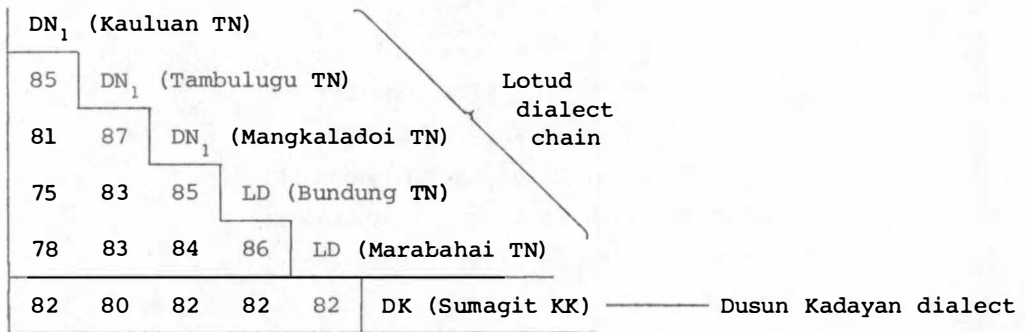

DN$_1$ (Kauluan TN)					
85	DN$_1$ (Tambulugu TN)				
81	87	DN$_1$ (Mangkaladoi TN)			
75	83	85	LD (Bundung TN)		
78	83	84	86	LD (Marabahai TN)	
82	80	82	82	82	DK (Sumagit KK)

Lotud dialect chain

Dusun Kadayan dialect

Figure 31: PSC relations of six villages of the Lotud language.
(DN$_1$ = "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

are lower, and those of the Bundu KP wordlist are even lower yet; thus the 86 PSC figure is not to be considered for classification of languages - Tatana is not closer to Bisaya linguistically than to other Dusunic languages. The PSC relation of the two Tatana wordlists with the Bisaya language samples and with two samples of Dusun are given in Figure 32. Using Bundu KP as representative, the relation of Tatana to the other Dusunic languages is 55-73 PSC.

```
BA (Kerukan KP)
┌──┐
│95│ BA (Kilugus KP)
├──┼──┐
│89│93│ BA (Kukut BT)                    Bisaya
├──┼──┼──┐                                language
│90│92│89│ BA (Limbawang BT)
├──┼──┼──┼──┐
│89│90│89│87│ BA (Batandok BT)
├──┼──┼──┼──┼──┐
│89│90│88│87│89│ BA (Padas Damit BT)
├──┼──┼──┼──┼──┼──┐
│87│89│87│86│84│83│ BA (Menunggang KP)
├──┼──┼──┼──┼──┼──┼──┐
│74│78│78│75│74│72│86│ TT (Menunggang KP) ⎫  Tatana
├──┼──┼──┼──┼──┼──┼──┼──┐                   ⎬  language
│67│69│67│66│65│62│72│83│ TT (Bundu KP)     ⎭
├──┼──┼──┼──┼──┼──┼──┼──┼──┐
│68│69│68│66│66│64│70│73│77│ KN₃ (Takuli BT)  ⎫  Beaufort
├──┼──┼──┼──┼──┼──┼──┼──┼──┼──┐                ⎬  Kadazan
│68│69│71│67│67│66│72│76│80│89│ KN₃ (Kinamam BT) ⎭  dialect
└──┴──┴──┴──┴──┴──┴──┴──┴──┴──┘
```

Figure 32: PSC relations of seven Bisaya language villages, two Tatana language villages and two Beaufort Kadazan dialect villages. (BA = Bisaya; TT = Tatana; KN$_3$ = "Kadazan" representing Beaufort Kadazan dialect.)

2.3.9.4.7 Kuijau language

The Kuijau language (KU) spoken in Keningau District is here represented by five wordlists. Each wordlist represents in this analysis a distinct Kuijau dialect since their relations are 77-85 PSC. The PSC relations of these dialects with each other are shown in Figure 33. Using Sungoi KU as representative, the Kuijau language has a relation of 52-79 PSC with the other Dusunic languages.

```
KU (Tuntumulod KU)
┌──┐
│83│ KU (Sungoi KU)
├──┼──┐
│81│85│ KU (Liau Darat KU)
├──┼──┼──┐
│80│82│81│ KU (Senagang KU)
├──┼──┼──┼──┐
│78│80│77│81│ KU (Linsosod KU)
├──┼──┼──┼──┼──┐
│67│70│69│68│64│ KUₘ (Apin-Apin KU)
└──┴──┴──┴──┴──┘
```

Figure 33: PSC relation of five Dusunic Kuijau dialects compared with the Murutic Apin-Apin Kuijau language. (KU = Kuijau (Dusunic); KU$_m$ = Apin-Apin Kuijau (Murutic).)

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.
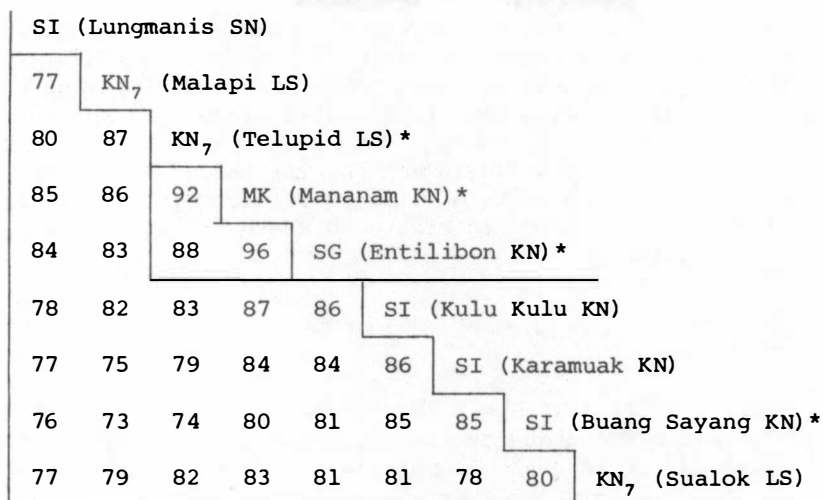
```
SI (Lungmanis SN)

77   KN₇ (Malapi LS)

80   87   KN₇ (Telupid LS)*

85   86   92   MK (Mananam KN)*

84   83   88   96   SG (Entilibon KN)*

78   82   83   87   86   SI (Kulu Kulu KN)

77   75   79   84   84   86   SI (Karamuak KN)

76   73   74   80   81   85   85   SI (Buang Sayang KN)*

77   79   82   83   81   81   78   80   KN₇ (Sualok LS)
```

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 classi-
fied 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 (KN_4, DN_5) 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	Kipouvu 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   Longkogungon 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 (DN_4) 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 (KN_2) 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 (DN$_3$) dialect is here represented by only one word-list 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 (KN$_6$) 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.

```
KN₆ (Tombovo PG)
┌────┐
│ 85 │ KN₆ (Kolopis PG)
├────┼────┐
│ 90 │ 84 │ KN₄ₚ (Penampang Baru PR)
│ 87 │ 86 ├────┐
│    │    │ 90 │ KN₄ₚ (Penampang PG)
│ 90 │ 88 │ 93 │ 92 │ KN₄ₚ (Lembah Poring BT)
│ 85 │ 82 │ 89 │ 85 │ 91 │ KN₄ₚ (T.A.S.S. TU)
├────┼────┼────┼────┼────┤
│ 80 │ 82 │ 86 │ 83 │ 87 │ 84 │ KN₄c (Kogopon PR)
```

Figure 35: PSC relations of two Tombovo (PG) Kadazan dialect villages with representative peripheral and core Central Dusun dialect villages. (KN$_6$ = "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 (KN_5) 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.
(KN_5 = "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 (DN_2) 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 (KN_3) 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 | DN₂ (Dudar KB)
86  91 | DN₂ (Taburan KB)
80  86  86 | DN₂ (Mantanau KB)
76  81  80  85 | DN₂ (Mandap KB)
81  84  83  86  88 | DN₂ (Rosok KB)
80  82  81  87  89  90 | DN₂ (Tambulian Laut KB)
92  86  84  81  79  85  81 | TG (Mengaris KM)
79  82  81  86  86  90  91 | 81 | DN₅ₚ (Piasau KB)
78  81  79  84  90  87  88 | 81  87 | DN₅ₚ (Kelawat KB)
80  83  82  86  87  93  89 | 83  89  85 | DN₅ₚ (Bangkahak KB)
81  80  79  83  84  83  83 | 84  85  85  83 | DN₅ᴄ (Taginambur KB)
```

Figure 37: PSC relations of seven Kota Belud Dusun dialect villages with peripheral and core Central Dusun dialect villages. (RS = Rungus; DN_2 = "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}.)

```
KN₃ (Tahak BT)
86 | KN₃ (Takuli BT)
88  83 | KN₃ (Tibabar-Sumbiling PR)
89  85  89 | KN₃ (Pinopok BT)
92  87  89  91 | KN₃ (Mandangin BT)
86  89  84  84  86 | KN₃ (Kinamam BT)
85  84  86  83  84  83 | KN₄ₚ (Limputong PR)
92  97  91  92  93  86 | 87 | KN₄ₚ (Takapan BT)
84  84  84  83  84  81 | 88  85 | KN₄ᴄ (Kogopon PR)
```

Figure 38: PSC relations of six Beaufort Kadazan dialect villages with peripheral and core Central Dusun dialect villages.
(KN_3 = "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: are:

> "Kadazan" (KN_{1i}):
> Kaingaran LS
> Melinsau LS
> Basai LS
> Kaingaran Baru LS
> Tinangian LS
> Karagasan 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 (Talantang 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" rep-
resenting core Central Dusun dialect; DN$_{5C}$ = "Dusun" rep-
resenting 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. (DN$_7$ = "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 (DN$_8$) 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.
          (DN_8 = "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(l) Pemilaan (TM) Dusun dialect

The Pemilaan (TM) Dusun dialect (DN_9) is here represented by only one word-
list which has 85 PSC with the one peripheral Central Dusun village noted
above with the code (l) - 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.
          (DN_9 = "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 (DN_{10}) 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. (DN_{10} = "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	
				Java-nese 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	chain

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 (KN_1, 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 under-taken 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 Kogopon 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