

## DIALECT COMPARISON AND INTELLIGIBILITY TESTING IN THE UPPER KINABATANGAN RIVER AREA (SABAH)

*HOPE M. HURLBUT and INKA PEKKANEN*  
*Summer Institute of Linguistics*

### 1. INTRODUCTION

The purpose of this paper is to give a report of conclusions drawn from the linguistic investigation and dialect intelligibility testing carried out in the Upper Kinabatangan River area of the Kinabatangan District in the State of Sabah, Malaysia. The work was part of an extensive language survey initiated by the Summer Institute of Linguistics.<sup>1</sup> Due to difficulties of travel and communication, the survey in the Kinabatangan District was divided into two phases, the first of which is reported here and was carried out by the authors in the Upper Kinabatangan from October 1978 to March 1979.

### 2. DATA COLLECTION AND ASSESSMENT

#### 2.1. THE SURVEY AREA

The survey area in this study includes three main rivers: the Tongod, the Pinangah, and the Milian,<sup>2</sup> and the area along the Pinangah-Telupid road as far north as Entilibon.

Two major processes were carried out in separate stages: first a linguistic survey and secondly dialect intelligibility testing. Each stage required a trip through the entire area.

#### 2.2. THE SURVEY TRIP

Preliminary to the survey trip through the area, decisions were made as to which villages to visit and from which to collect data. These decisions were made with the help of district and local officials who were familiar with the area.

The purpose of this trip was solely that of data gathering. Four types of material were collected from each village visited during this trip: a 367-item wordlist, an ethnographic questionnaire, a linguistic map of the area, and a short tape-recorded story in the local dialect.<sup>3</sup>

<sup>1</sup>The writers express their appreciation to Dr. Kenneth D. Smith for his help in the organization of the material in this study.

<sup>2</sup>Milian is the local name for the Kinabatangan down as far as Koromuak. The village of Koromuak at the junction of Koromuak and Kinabatangan rivers was also visited, and data were collected in two dialects, Sungai and Sukang. According to local sources, the dialects are similar to the Sungai dialect spoken further downstream on the Kinabatangan. It was therefore decided to exclude Sungai and Sukang from this study and to include them in the survey and intelligibility testing of the Lower Kinabatangan area.

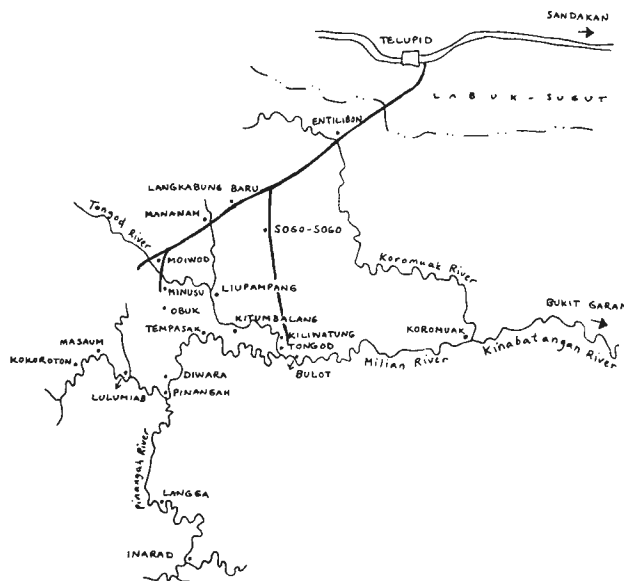
<sup>3</sup>For a description of the wordlists and questionnaires see Blom (1979). The entire procedure of dialect intelligibility testing is described in Casad (1974).

2.3. DATA PROCESSING

The distribution of villages in the Kinabatangan District from which data were collected can be seen on Map 1.

The officials in the district center of Tongod suggested that there were as many as seven different languages in the area: Makiang, Sinabu, Mangkaak, Minokok, Rumanau, Sinarupa, and Murut. Murut was said to be spoken in only one village. However, on visiting the upper Milian a second Murut village was visited where the people stated that their language was different from that spoken in the first Murut village. And lastly, speakers of the Kolobuan language were encountered in a few villages. Data was therefore collected from at least one village of each of the different language groups.

Each wordlist was compared with every other wordlist collected within the survey area to determine the percentage of shared cognates between them.<sup>4</sup> Chart 1 shows the cognate percentages between all the villages visited, and between each of them and the national language, which is Bahasa Malaysia.



Map 1. Villages<sup>5</sup> visited during the survey trip.

- = village<sup>5</sup>
- - - = district boundary
- ==== = main road
- ===== = secondary road

<sup>4</sup>For the purpose of this study two words are considered cognate if they follow the criteria outlined in Gudschinsky (1956). Wordlists were compared across all dialect boundaries. There was some problem at first in computing the cognate percentages because it was discovered that many of the language helpers had misunderstood the words being asked which resulted in lower cognate percentages. Therefore the languages and dialects were initially thought to be more different than they actually were. The wordlists and calculations were subsequently corrected.

<sup>5</sup>In that there are no road maps of the Upper Kinabatangan area at present, the roads and villages on the roads which have been marked on Map 5 are located only approximately. All these villages are located by a road except Obuk. At the time of the survey the road to Tongod had not been completed.



3.1. DETERMINING TEST SETS AND TEST POINTS

The cognate percentages, ethnographic questionnaires and maps were analyzed to decide from which villages stories would be selected to make up test sets for the testing stage.

Different test sets were established for the following areas: Pinangah River, Milian River, Tongod River, and Pinangah Road. Maps 2, 3, 4, and 5 show the villages where each set was tested. In determining what each test set should include, several factors had to be kept in mind. The test set should 1) verify distinctness of dialects, 2) determine the extent of language learning between these, and 3) note whether this has been a one-way or two-way process. It was decided that within each of the four geographic divisions, one sample of each dialect found within that area should be tested with all other dialects within that area. The villages from which those sample stories were selected are called reference points.<sup>6</sup>

Once the test sets were determined, villages were selected as test points where the intelligibility testing would be done. For the summary of reference points and test points see Chart 2. When only one or two villages of a language or dialect were found within the survey area, each of these was tested. For Makiang and Sinabu languages representative villages were chosen to obtain as wide a geographical spread as possible.

The Pinangah River test set (Map 2) was tested in the two villages on the Pinangah River.<sup>7</sup>

The Milian River test set (Map 3) contained a sample of each dialect on the Milian River: Makiang, Kolobuan, Rumanau, plus a tape of the Murut dialect on the Pinangah River and of the Sinabu dialect on the Tongod River.

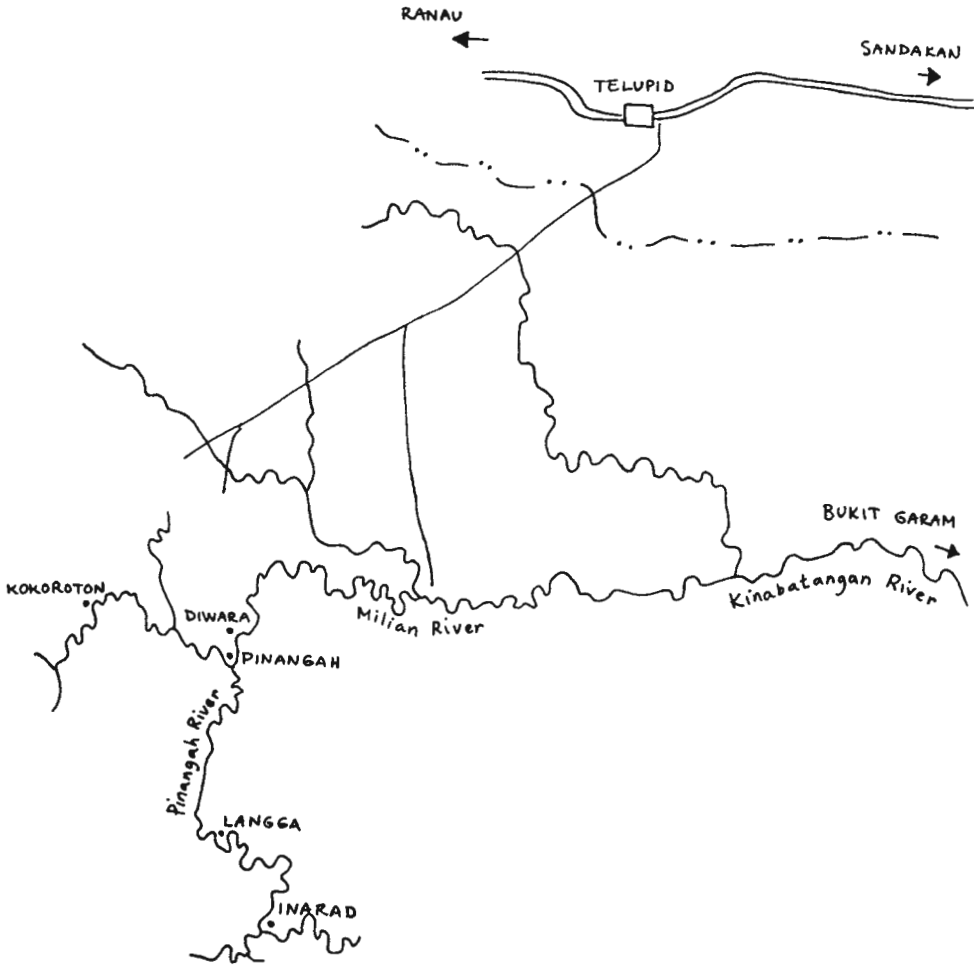
The Tongod River test set (Map 4) contained a sample of each dialect on the Tongod River and its tributaries: Makiang, Sinabu, Rumanau, Minokok, and Mangkaak. In the village of Bulot the Kolobuan dialect was also tested. But by the time the other villages were visited, many of the wordlists had been checked and revised and it was determined that Kolobuan was so closely related to Makiang that it seemed unnecessary to continue using both the Kolobuan and Makiang tapes in testing, and the Makiang tape was selected to represent both.

The Pinangah Road test set (Map 5) contained the same tapes as the Tongod River test set except that it had been determined that Sinabu was so closely related to Makiang that the Makiang tape could be used to represent both. Since the Pinangah Road comes from Telupid in the Labuk-Sugut District, a sample of the Labuk Kadazan dialect was also tested in this area. Several years previously a substantial number of West Coast Dusun, mainly from the Kota Belud and Ranau Districts, had moved into the area; hence, a random sample of their language was also used in testing.

<sup>6</sup>Because of the amount of time required to listen to the stories, it was felt that the number of tapes per set should not exceed 8. During the early stages of testing some decisions as to what should be included were arbitrary as data were not available from the entire area to be surveyed and tested. Thus the test sets were adjusted and modified later on.

<sup>7</sup>Because it was difficult to secure transport to the villages on the Pinangah River, both the initial survey trip and the intelligibility testing were carried out at the same time, though information had not yet been obtained from all the other languages. For this reason only a limited number of stories from neighboring villages were used in testing.

# DIALECT COMPARISON AND INTELLIGIBILITY TESTING



Map 2. Reference and test points for the Pinangah River test set

\* = test

△ = reference point

Villages:

Dialects:

△ Kokoroton

Murut

△ Diwara

Makiang

△ Pinangah

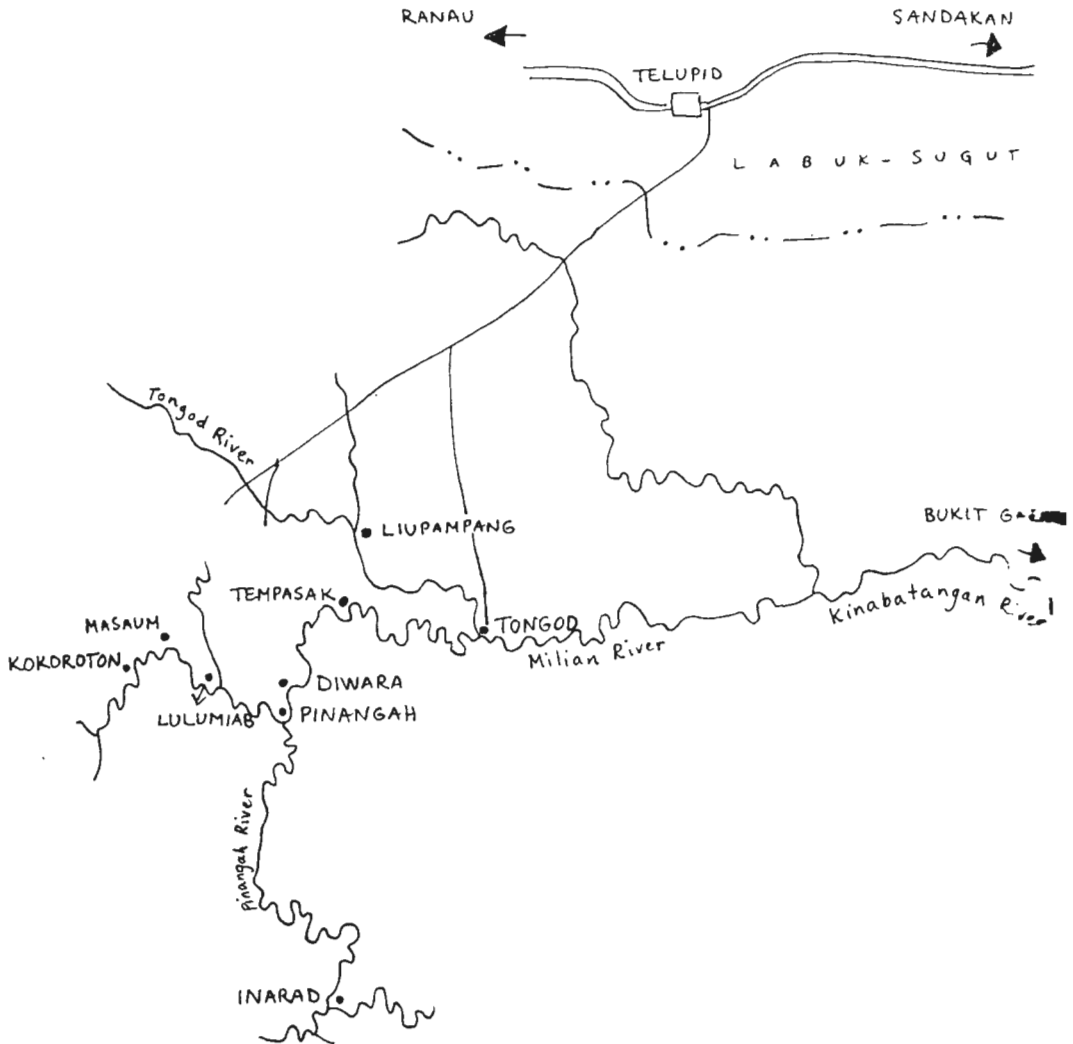
Kolobuan

\* △ Langga

Kolobuan

\* △ Inarad

Murut



Map 3. Reference and test points for the Milian River test set

\* = test point

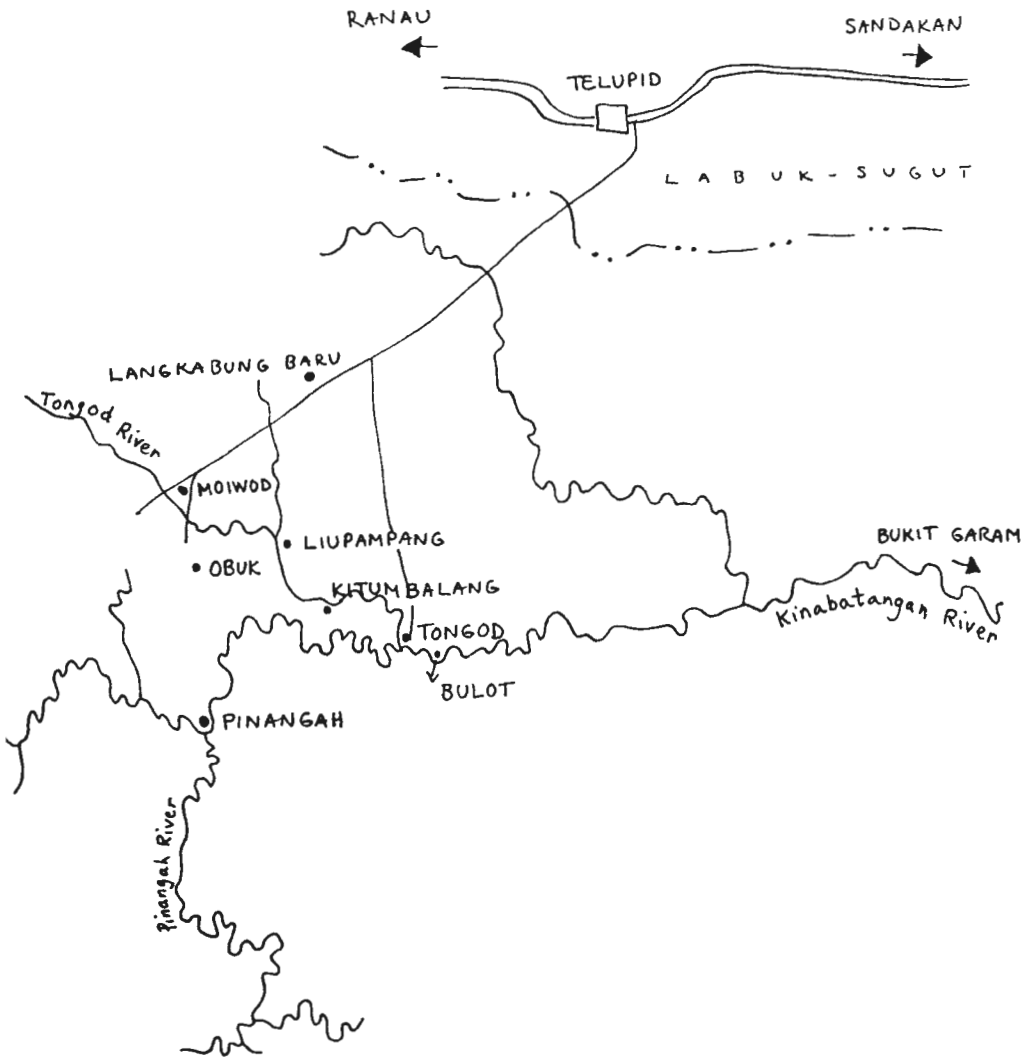
△ = reference point

Villages:

Dialect:

△ Liupampang	Sinabu
* △ Masaum	Rumanau
△ Tongod	Makiang
* △ Kokoroton	Murut
* △ Pinangah	Kolobuan
△ Inarad	Murut
* Tempasak	Makiang
* Diwara	Makiang
* Lulumiab	Rumanau

# DIALECT COMPARISON AND INTELLIGIBILITY TESTING



Map 4. Reference and test points for the Tongod River test set

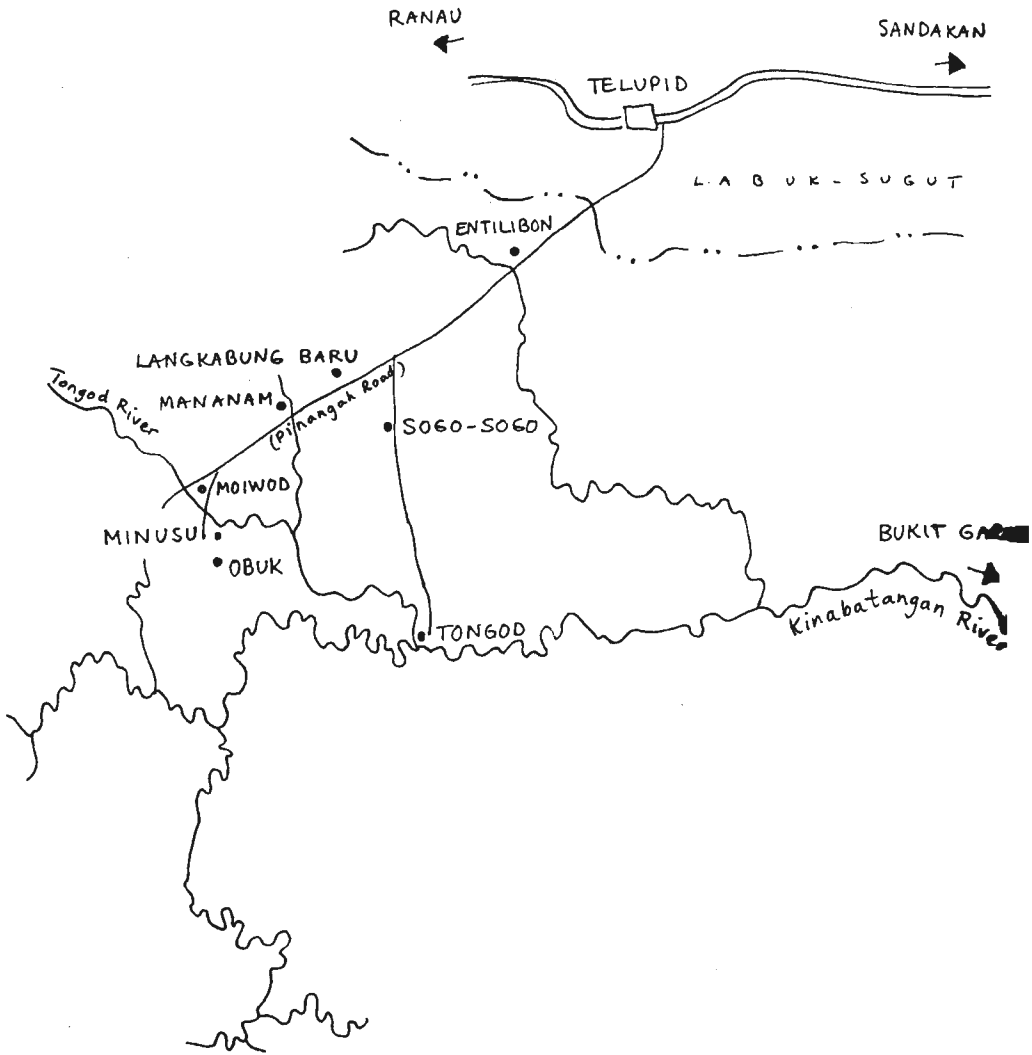
\* = test point

△ = reference point

Villages:

Dialect:

△ Langkabung Baru	Mangkaak
△ Moiwod	Minokok
* △ Liupampang	Sinabu
△ Obuk	Rumanau
△ Tongod	Makiang
△ Pinangah	Kolobuan
* Kitumbalang	Sinabu
* Bulot	Sinarupa



Map 5. Reference and test points for the Pinangah Road test set

\* = test point

△ = reference point

Villages:

Dialect:

- △ Telupid
- \* △ Entilibon
- \* Entilibon
- \* △ Langkabung Baru
- \* △ Moiwod
- \* △ Obuk
- △ Tongod
- \* Sogo-Sogo
- \* Mananam
- \* Minusu

- Labuk
- Dusun
- Minokok
- Mangkaak
- Minokok
- Rumanau
- Makiang
- Mangkaak
- Mangkaak
- Sinabu



DIALECT COMPARISON AND INTELLIGIBILITY TESTING

Reference points:																	
Test points:																	
Village	Dialect	Location		Inarad (Murut)	Kokoroton (Murut)	Pinangah (Kolobuan)	Langga (Kolobuan)	Diwara (Makiang)	Tongod (Makiang)	Liupampang (Sinabu)	Masaum (Rumanau)	Obuk (Rumanau)	Langkabung Baru (Mangkaak)	Telupid (Labuk)	Moiwod (Minokok)	Entilibon (Dusun)	Bahasa Malaysia
Inarad	Murut	P	Rv	x	x	x	x	x									x
Kokoroton	Murut	M	Rv	x	x	x			x	x	x						x
Pinangah	Kolobuan	M	Rv	x	x	x			x	x	x						x
Langga	Kolobuan	P	Rv	x	x	x	x	x									x
Tempasak	Makiang	M	Rv	x	x	x			x	x	x						x
Diwara	Makiang	M	Rv	x	x	x			x	x	x						x
Bulot	Sinarupa	T	Rv			x			x	x	x	x	x		x		x
Liupampang	Sinabu	T	Rv			x			x	x	x	x	x		x		x
Kitumbalang	Sinabu	T	Rv			x			x	x	x	x	x		x		x
Minusu	Sinabu	P	Rd						x		x	x	x	x	x	x	x
Masaum	Rumanau	M	Rv	x	x	x			x	x	x						x
Lulumiab	Rumanau	M	Rv	x	x	x			x	x	x						x
Obuk	Rumanau	P	Rd						x		x	x	x	x	x	x	x
Langkabung Baru	Mangkaak	P	Rd						x		x	x	x	x	x	x	x
Sogo-Sogo	Mangkaak	P	Rd						x		x	x	x	x	x	x	x
Mananam	Mangkaak	P	Rd						x		x	x	x	x	x	x	x
Moiwod	Minokok	P	Rd						x		x	x	x	x	x	x	x
Entilibon	Minokok	P	Rd						x		x	x	x	x	x	x	x
Entilibon	Dusun	P	Rd						x		x	x	x	x	x	x	x

Chart 2. Summary of reference points and test points

- x = test point in which the reference tape was tested
- x = test point which is also a reference point
- P Rv = Pinangah River
- M Rv = Milian River
- T Rv = Tongod River
- P Rd = Pinangah Road

### 3.2. TESTING

Once the reference points and test points were chosen the testing trip through the entire survey area was begun. During the testing trip three things were done in each village:

- (1) the test set was adapted to the village being tested;
- (2) at least ten people were chosen to take the test; and
- (3) the tests were administered to the ten people selected.

Prior to arriving at the test point the technicians prepared ten questions for each reference tape story for use in the testing. In adapting the test set to a particular village the ten questions from each tape in the test set were translated and recorded on tape in the language spoken at that test point.<sup>8</sup> Once the entire set of questions was recorded, the technician dubbed those questions in the local dialect onto the taped stories immediately after the point where the information being questioned is given in the story.<sup>9</sup> Finally the ten people chosen for testing listened to the stories and their answers to the questions were scored.<sup>10</sup>

### 3.3. EVALUATION AND APPLICATION

#### 3.3.1. DATA COMPARISON

For the purpose of establishing dialect boundaries a threshold of cognate percentage and intelligibility test scores must be defined. If the relation of two villages is above the threshold level, those villages are considered to be within the same dialect boundary. If, however, the relation of two villages is below the threshold, they are considered to be of different dialects. For the purpose of this study the threshold level of 80% is used for dialects and the range of 75-80% for languages (see Smith 1980).

The intelligibility test results are displayed in Chart 3. Where there are two numbers, they show the range from lowest to highest score. The cognate percentages are also included in the chart to facilitate comparison. For scores in all the villages of each test set, see the Appendix.

The display in Chart 3 shows distinct groups forming on the hypotenuse of the triangle. The cognate percentages and intelligibility scores within each small triangle are much higher than those falling outside these triangles.

<sup>8</sup>For example, when the village of Tempasak was tested, the questions from Kokoroton, Masaum, Pinangah, Inarad, Tongod, and Liupampang reference tapes and the questions that accompany the national language intelligibility test tape were translated into Makiang as it is spoken in Tempasak by a person from that village. Due to the close proximity of some of the villages sometimes these questions were used in a neighboring village of the same dialect.

<sup>9</sup>The following story will illustrate the insertion of questions:

One day at the time that I was living in the village, I was about eighteen years old and we had some orchards.

1. What did they have?

Our orchards were near the houses.

## DIALECT COMPARISON AND INTELLIGIBILITY TESTING

2. Where were the orchards?

So my father told me to go and see how the orchards were.

3. Who told him to go to the orchard?

I would have gone to the orchard but suddenly there was a (domestic) pig there.

4. What was in the orchard?

The pig, it had given birth.

5. What happened to the pig?

So I went to the orchard, but the pig that had given birth there came after me.

6. What did the pig do?

And there I was screaming and running because that pig that gave birth was chasing me.

7. What did the man do?

And I looked for a way out and there was a tree and I began to climb it and I wasn't bitten by the pig.

8. Where did he go?

So there I was still screaming and my father also arrived there to come to help me by chasing away the pig.

9. Who came to help?

And I wasn't bitten by the pig. And when my father arrived to chase away the pig and when the pig had run away, then I came down from the tree that I had climbed and went home.

10. What did his father do?

(This story was told by Sagun bin Ramburan of Entilbon in the Dusun dialect.)

<sup>10</sup>For further details of the testing procedure see Blom (1979). In most villages on the Upper Kinabatangan it was found that as many as 30% of the people participating either could not learn the procedure or had some physical problem such as a head cold or deafness which seemed to cause unusually low scores. In cases like this the people being tested were allowed to listen to one or more tapes and then thanked for their help. Their scores were not included since they were unable to complete the test. Because that happened frequently, it was found necessary to register the names of approximately 14-18 willing people in the village. This was easier than to call on others as substitutes, if someone could not complete the series successfully, without embarrassing either party.



## DIALECT COMPARISON AND INTELLIGIBILITY TESTING

Lexico-statistically the two Murut 'dialects' are subdialects of the same dialect, but the mutual intelligibility scores are well below the dialect threshold level. As intelligibility testing is considered more conclusive than lexico-statistic comparisons in determining the level of communication between speakers of different 'dialects', the two Murut 'dialects' should not be considered subdialects but rather dialects of different languages.

The second small triangle is comprised of Kolobuan, Makiang, Sinarupa, Sinabu, and Rumanau. There is a distinct break between them and the two Murut dialects. However, the Inarad Murut and the Kolobuan speakers have learned each other's language, for the intelligibility scores are as high as 82-89% and 94-100%, even though the cognate percentage is only 55%. This is likely due to sociolinguistic factors explained by the fact that the Kolobuan village of Pinangah is a meeting place of different ethnic groups as there is a rural clinic, a store, and a school there. The Inarad village headman also lives at Pinangah. The children from the Kolobuan village of Langga go to school at Inarad because the villages are relatively close to one another (3 hours paddling time) but isolated from other villages (2 days paddling time to Pinangah going downstream). The intelligibility scores between Kolobuan, Makiang, Sinarupa, Sinabu, and Rumanau range from 65% to 100%. Except for two, all scores are above the dialect threshold level. The Kolobuan at Langga scored only 68% with Makiang. The low figure may be due to the poor quality of the Makiang tape used there. Later the use of that tape was discontinued. The Sinarupa speakers at Bulot scored 65% with Rumanau. That seems low, considering that Sinarupa scored 100% with Sinabu, which in turn scored 92% with Rumanau. It can be noted that on the average, the Rumanau scored higher with Kolobuan, Makiang, and Sinabu than those did with Rumanau. Even so, the intelligibility scores going both ways between those and Rumanau are higher than the respective cognate percentages. The Rumanau people are culturally quite different from the other groups in the area and consider themselves to be distinct from these. But on the whole, it can be concluded that Kolobuan, Makiang, Sinarupa, Sinabu, and probably also Rumanau are dialects of the same language.

Mangkaak, Labuk, and Minokok form a group which is distinct from the other dialects on the chart. With cognate percentages ranging from 80% to 92% and mutual intelligibility scores from 62% to 97%, they may be considered to be dialects of the same language. Of the two Minokok villages tested, Moiwod scored 62% with Labuk, falling below the threshold level, whereas the Minokok at Entilibon scored 88% with Labuk. It is to be noted that Moiwod is distant from the Labuk-speaking area, whereas Entilibon is geographically close to it. Language learning has apparently taken place between the Rumanau and Mangkaak: Rumanau scored 79-83% with Mangkaak. The Mangkaak people at Mananam scored 91% with Rumanau. Again, the geographical proximity is an influencing factor. It is surprising to note that while Minokok scores ranged from 61% to 91% with Rumanau, it was the more distant village that had a much higher score than the one close by.

Dusun is distinct from all the other dialects displayed on the chart. It does, however, have a rather tenuous relationship with Minokok, in that the cognate percentages range from 78% to 79%. The mutual intelligibility range is 50-98%. The Minokok at Entilibon had the higher score with Dusun. This of course is attributable to the fact that the Dusun live in the same village. The Minokok at Moiwod scored only 50% with Dusun, which may be due to the fact that they have had very little contact with the Dusun. The Dusun scored 79-80% with Minokok, Labuk, and Mangkaak, which is consistent with the cognate percentages. The Dusun were also tested with Rumanau and Makiang tapes, but the scores were considerably lower: 50% and 10%, respectively.

## 3.3.2. NATIONAL LANGUAGE INTELLIGIBILITY

Chart 4 displays the level of bilingualism in Bahasa Malaysia in the villages tested. The Bahasa Malaysia tape was played for each of those taking the test, after they had listened to all the other tapes.

Score	Village	Dialect
76	Entilibon	Dusun
73	Entilibon	Minokok
63	Sogo-Sogo	Mangkaak
60	Pinangah	Kolobuan
52	Mananam	Mangkaak
51	Inarad	Murut
51	Masaum	Rumanau
47	Obuk	Rumanau
46	Diwara	Makiang
44	Lulumiab	Rumanau
42	Tempasak	Makiang
41	Moiwod	Minokok
38	Liupampang	Sinabu
38	Minusu	Sinabu
35	Langkabung Baru	Mangkaak
33	Kokoroton	Murut
32	Bulot	Sinarupa
18	Kitumbalang	Sinabu
14	Langga	Kolobuan

Chart 4. National language intelligibility test results

The national language intelligibility test results range from 14% to 76%, with the average of 45%. From the chart it is apparent that the dialect is not a factor for greater or lesser intelligibility. The following factors seem more likely to have influenced the scores: distance from places where Bahasa Malaysia is used; difficulty in travelling; and scarcity of schools. Until recently the villages in the Upper Kinabatangan have been accessible by river travel only, and that is still true of several of the ones visited during this survey. Many of those who participated in the test had never been outside their dialect area. It is interesting to note that many people were multilingual in languages that differ considerably from their mother tongue, but they could not speak or understand Bahasa Malaysia. Several respondents hesitated taking the test, saying that they did not know Malay. There were others who said they had learned some Malay from their friends and could use it when, for example, buying or selling. The intelligibility scores are generally higher in those villages which have a school or are accessible by road.

#### 4. CONCLUSION

After the initial data gathering it seemed that the linguistic picture of the Upper Kinabatangan area was rather complex, because several different names had been given as language or dialect names. When all the wordlists had been compared with each other and the intelligibility testing completed, it was found that the number of significant distinctions was reduced. For example, Kolobuan, Makiang, Sinarupa, Sinabu, and probably Rumanau are considered to be within the same language boundary. Lexico-statistically Kolobuan, Makiang, and Sinarupa are subdialects of a dialect which Smith calls 'Makiang' (see Smith 1980). Sinabu and Rumanau form another dialect within the language. The two Murut 'dialects' pose a problem in interpreting both the cognate percentage and intelligibility test results. Lexico-statistically they are closely related dialects of the same language, but intelligibility scores would support their division into two distinct languages. Mangkaak, Labuk, and Minokok are within the same language boundary with each other. Mangkaak and Labuk together comprise a dialect. Dusun is distinct from all the others. Here the division differs somewhat from the one presented by Smith.<sup>11</sup> Generally speaking, the intelligibility scores are consistent with the cognate percentages, except where language learning has taken place, which raises the intelligibility scores. In a few cases it was difficult to determine what factors exactly have influenced the intelligibility scores. As a summary it is noted that four (or perhaps five) languages are spoken in the Upper Kinabatangan. Each of them, with the exception of Dusun, consists of two or more dialects.

#### REFERENCES

- BLOM, JULIE K. 1979. Dialect comparison and bilingualism: the Kudat Division. Paper presented at a seminar on the Application of Linguistic Studies to Sabah Cultures in Kota Kinabalu, February 17, 1979, to be published by the Sabah State Museum.
- CASAD, EUGENE. 1974. Dialect intelligibility testing. Summer Institute of Linguistics, University of Oklahoma, Norman, Oklahoma.
- GUDSCHINSKY, SARAH K. 1956. ABC's of lexicostatistics (glottochronology). *Word* 12.2.175-210.
- SMITH, KENNETH D. 1980. The languages of Sabah: a tentative lexico-statistical classification. Manuscript.

#### APPENDIX

The following charts show the mutual intelligibility testing results in each test set.

<sup>11</sup>Smith presents Mangkaak and Labuk as dialects of the Labuk Kadazan language; Minokok is a subdialect within the Sugut Kadazan dialect of the Dusun (Kadazan) language, and Dusun as it is spoken at Entilibon is within the Central Kadazan-Dusun dialect of the same language.

REFERENCE TAPES  TEST POINTS	MURUT (Inarad)	(Kokoroton)	KOLOBUAN (Langga)	(Pinangah)	MAKIANG (Diwara)
MURUT (Inarad)	100	75	89	82	
KOLOBUAN (Langga)	94		100	99	68

Mutual intelligibility testing results:  
Pinangah River (Map 2)

REFERENCE TAPES  TEST POINTS	MURUT (Inarad)	(Kokoroton)	KOLOBUAN (Pinangah)	MAKIANG (Tongod)	SINABU (Liupampang)	RUMANAU (Masaum)
MURUT (Kokoroton)	71	100	49	44	75	60
KOLOBUAN (Pinangah)	100	65	100	100	100	88
MAKIANG (Diwara)	67	36	98	100	100	85
(Tempasak)	51	15	92	94	97	87
RUMANAU (Masaum)	90	75	89	100	100	100
(Lulumiab)	80	63	100	100	100	100

Mutual intelligibility testing results:  
Milian River (Map 3)



**DIALECT COMPARISON AND INTELLIGIBILITY TESTING**

REFERENCE TAPES  TEST POINTS	KOLOBUAN (Pinangah)	MAKIANG (Tongod)	SINABU (Liupampang)	RUMANAU (Obuk)	MANGKAAK (Langkabung Baru)	MINOKOK (Moiwod)
SINARUPA (Bulot)	82	96	100	65	56	62
SINABU (Kitumbalang)		95	100	100	79	61
(Liupampang)		99	100	98	77	72

Mutual intelligibility testing results:  
Tongod River (Map 4)

REFERENCE TAPES  TEST POINTS	MAKIANG (Tongod)	RUMANAU (Obuk)	MANGKAAK (Langkabung Baru)	LABUK (Telupid)	MINOKOK (Moiwod)	DUSUN (Entilibon)
DUSUN (Entilibon)	10	50	80	79	79	100
MINOKOK (Entilibon)	43	91	92	88	99	98
(Moiwod)	21	61	83	62	100	50
MANGKAAK (Langkabung Baru)	45	60	100	81	78	41
(Sogo-Sogo)	58	82	100	85	96	42
(Mananam)	60	91	100	97	93	58
RUMANAU (Obuk)	83	100	79	65	83	48
SINABU (Minusu)	90	92	77	49	66	25

Mutual intelligibility testing results:  
Pinangah Road (Map 5)