

## A SURVEY OF THE ARAPESH LANGUAGE FAMILY OF PAPUA NEW GUINEA

Robert J. Conrad  
 Summer Institute of Linguistics

- 0 INTRODUCTION
- 1 PROCEDURE
- 2 RESULTS
- 3 EVALUATION
- 4 CONCLUSION

## 0 INTRODUCTION

The Arapesh language family of the Maprik and Yangoru Sub-Provinces, East Sepik Province, is in the Kombio Stock of the Torricelli Phylum<sup>1</sup> of Papua New Guinea. It consists of three languages: Bukiyip (Mt. Arapesh), Muhiang (Southern Arapesh<sup>2</sup>), and Bumbita. The approximate populations are 10,300, 10,600, and 2,300 respectively<sup>3</sup>.

The major purpose of this survey was to estimate the coverage of vernacular literature in the Muhiang dialect spoken at Albinama 3. This was accomplished by investigating the major dialect differences in the Muhiang language of the Arapesh family.

The secondary purpose was to ascertain the extreme dialect differences within the Bukiyip (Mt. Arapesh) language and also confirm the place of Bumbita as a separate language.

The main focus of the survey was on the Muhiang language and its dialects, since this is the first language in the family in which national translators have begun working. Data was also collected by helicopter<sup>4</sup> from the coastal dialect of Bukiyip from the villages of Matapau, Malin, Balof, Walum, and Walahiga. I wish to thank Dr. Robert Brown and Mr. Arden Sanders for taking the word lists on this part of the survey in which I was unable to participate due to illness.

## 1 PROCEDURE

The villages of Namango, Balif, Amahop, and Bonahoi were visited during the survey. The remaining lists were taken at the South Seas Evangelical Church's Bible School at Brugam. The Dagua list was obtained from a student at the Lahara Course and the Bubuamo list from a close friend, some years ago, when the author resided there.

For control purposes, two lists (AA and AL) were taken for Albinama 3 and two more (BG and IT) for Ilahita 4. Both pairs of these lists scored 95% probable cognates.

The standard SIL Survey Word List containing 190 words and phrases was modified by deleting those words which are difficult to elicit and/or cause problems in the Arapesh language family.

The words deleted included most of the verbs and also the following other words: heart, liver, bone, horn, feather, wing, tail, old man, old woman, person, frog, bean axe, mountain, seed, tobacco; all colours except white and black; good, bad, long, short, heavy, light, cold, warm, old, new, many, all, this, that, what, who, when, where, round, wet, dry, full, not. This left a list of 127 words. An attempt was made to elicit the full 127 word list except where comparison showed very close similarity with a previous list already taken. In such cases the first 50 words only were elicited. Due to limitations of time in the helicopter, and limitations of the SPEED program for counting cognates, only 115 words were actually compared in 10 villages. A separate comparison was made using 44 words from 26 villages. A final comparison was made using 72 words from 18 villages.

These words were elicited in Pidgin English, which was well known by all the speakers. The words were then compared in the following way: two words were counted as cognate if they had half or more of the phonemes (sounds) the same or regularly corresponding and in the same order. These decisions were then counted using the SPEED program at the Ukarumpa base of the Summer Institute of Linguistics. These counts were then converted into percentages by the use of a calculator.

In addition, many of the speakers were interviewed using a portion of the sociolinguistic questionnaire written by Arden Sanders.

Finally, a simple intelligibility test involving 20 items was used with most of Muhiang speakers to estimate their intelligibility with respect to Albinama 3. A few intelligibility tests were made with Ilahita, and a few with Bubuamo. Albinama 3 and Ilahita were chosen for the intelligibility testing because the two national translators chosen by the South Seas Evangelical Church are from these two villages.

The intelligibility tests were administered in the following way: a tape recording of 20 simple statements or questions in the vernacular to be tested was played, one question at a time, and the speaker's translation into Pidgin was evaluated as correct, half correct, or wrong.

## 2 RESULTS

The following abbreviations were used for village names:

AA	Albinama 3 (Joshua Lukas)
AL	Albinama 3 (Andrew)
BA	Balif
AM	Amahop 2
WA	Walahuta
SU	Supari 2
IL	Ilahop
BG	Balanga (Ilahita 4) list from John Alungum
IT	Ilahita 4
MU	Muwi
NI	Ningalimbi
NO	Nomangu (Namango)
NK	Nagipeim
NB	Ningalemb
AH	Ami (Amih)
IW	Iwam
WO	Womsak 1
WL	Walahiga
MA	Malin
BF	Balof (Balup)
MT	Matapau
DA	Dagua (Urip)
BO	Bonahoi
UR	Urita
TI	Timingir
BU	Bubuamo

The cognate percentages obtained are listed in Figures 1-3. In each case, the percentage of average shared cognates is listed at the right of each village name abbreviation. The results of the intelligibility test are shown in Figures 4 and 5.



ARAPESH FAMILY - 72 WORD COMPARISON

AA 71  
 94 AL 70  
 85 82 SU 69  
 90 88 85 BG 68  
 93 90 83 92 MU 70  
 85 86 72 81 82 NO 66  
 69 67 75 69 69 61 NK 65  
 76 75 81 74 76 71 72 NB 68  
 82 82 83 78 79 76 72 88 AH 71  
 78 79 79 75 76 72 74 89 89 IW 70  
 62 64 67 57 58 63 71 68 71 74 WL 63  
 57 56 57 51 54 54 71 58 63 61 81 MA 62  
 58 57 58 53 56 57 68 57 63 60 81 81 BF 59  
 63 60 64 60 60 57 68 60 63 61 63 64 64 DA 59  
 51 49 47 49 51 47 44 51 51 50 47 46 44 40 BO 53  
 50 47 44 47 50 44 40 49 50 49 44 43 43 39 90 UR 51  
 47 44 42 44 47 43 39 46 47 46 43 43 43 38 90 97 TI 50  
 68 65 72 69 68 61 74 67 68 71 61 63 60 81 47 43 42 BU 64

Figure 2

ARAPESH FAMILY - 115 WORD COMPARISON

AA 73  
 83 SU 69  
 93 82 BG 72  
 94 80 93 MU 72  
 70 69 69 69 NK 67  
 75 77 72 73 70 NB 66  
 58 54 54 56 68 55 BF 56  
 63 61 61 60 66 58 61 DA 62  
 55 50 53 54 46 52 42 42 BO 49  
 69 66 69 67 73 63 58 86 47 BU 66

Figure 3

PERCENTAGE OF CORRECT ANSWERS TO INTELLIGIBILITY TEST FROM  
ALBINAMA 3

AH	72½	
AM	92½	
BA	100	
BO	57½	(7 answers only half correct)
IL	92½	
IT	95	
IW	62½	
MU	97½	
NB	60	
NI	92½	
SU	100	
TI-UR	42½	
WA	82½	

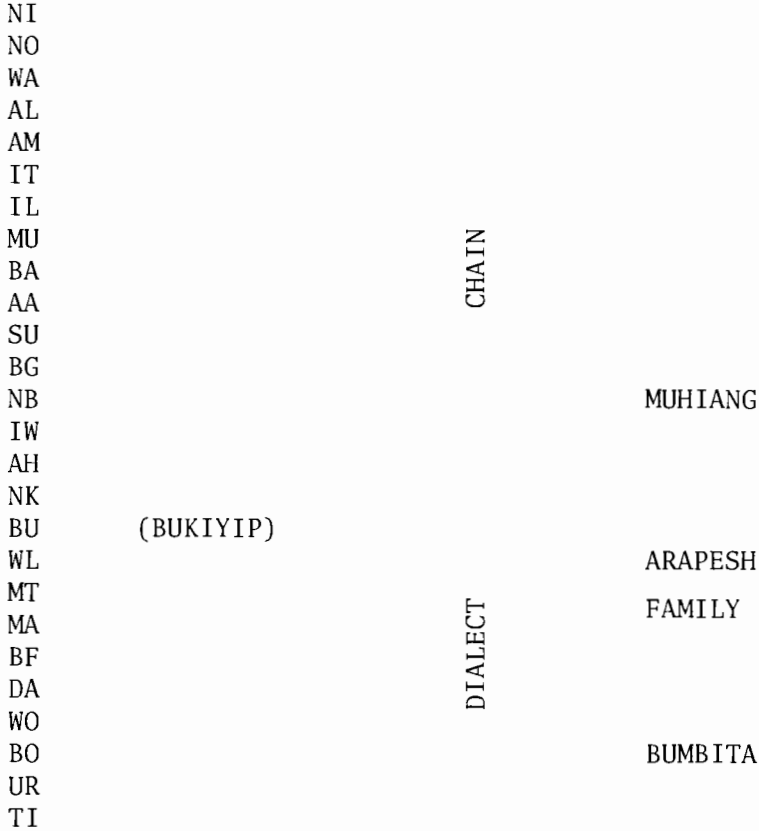
Figure 4

PERCENTAGE OF CORRECT ANSWERS TO INTELLIGIBILITY TEST FROM  
ILAHITA

AH	92½
BA	100
IW	90
TI-UR	80

Figure 5

The data in Figures 1-3 can be summarized in the following chart showing the major relationships:



One approach to a dialect chain was suggested by Gary Simons and is described by Arden Sanders (Sanders 1977:307 ff). In this method the speech communities are arranged according to geographical position on a map and then the graph is drawn beginning with the highest percentage figure. Successive lines are drawn indicating the successive next higher percentages, with the constraint that no line is drawn which will connect two speech communities already connected by another route. As Sanders notes, the result shows the basic pattern of closest relationships between the group of speech communities. These graphs are shown in Figures 6, 7 and 8 respectively, using the 44, 72 and 115 word comparisons.

Following this, the lines from the graph can be used to decide which point is the centre of the pattern. This is done by constructing a matrix whose cells indicate the number of lines on the connected line graph between each pair of speech communities. Then the average scores are calculated and the scores are compared to find the lowest average. This community is taken to be the centre of the language pattern. This was done and the results are shown in Figures 9, 10, and 11 respectively for the three comparisons of 44, 72, and 115 words. Note that AA comes out as the centre in the 44 word comparison only, and is replaced by IW in the 72 word comparison and by NB and SU in the 115 word comparisons. However, in both the latter cases AA is not too far from being the lowest.

A SURVEY OF THE ARAPESH LANGUAGE FAMILY

CONNECTED LINE GRAPH - 44 WORD COMPARISON

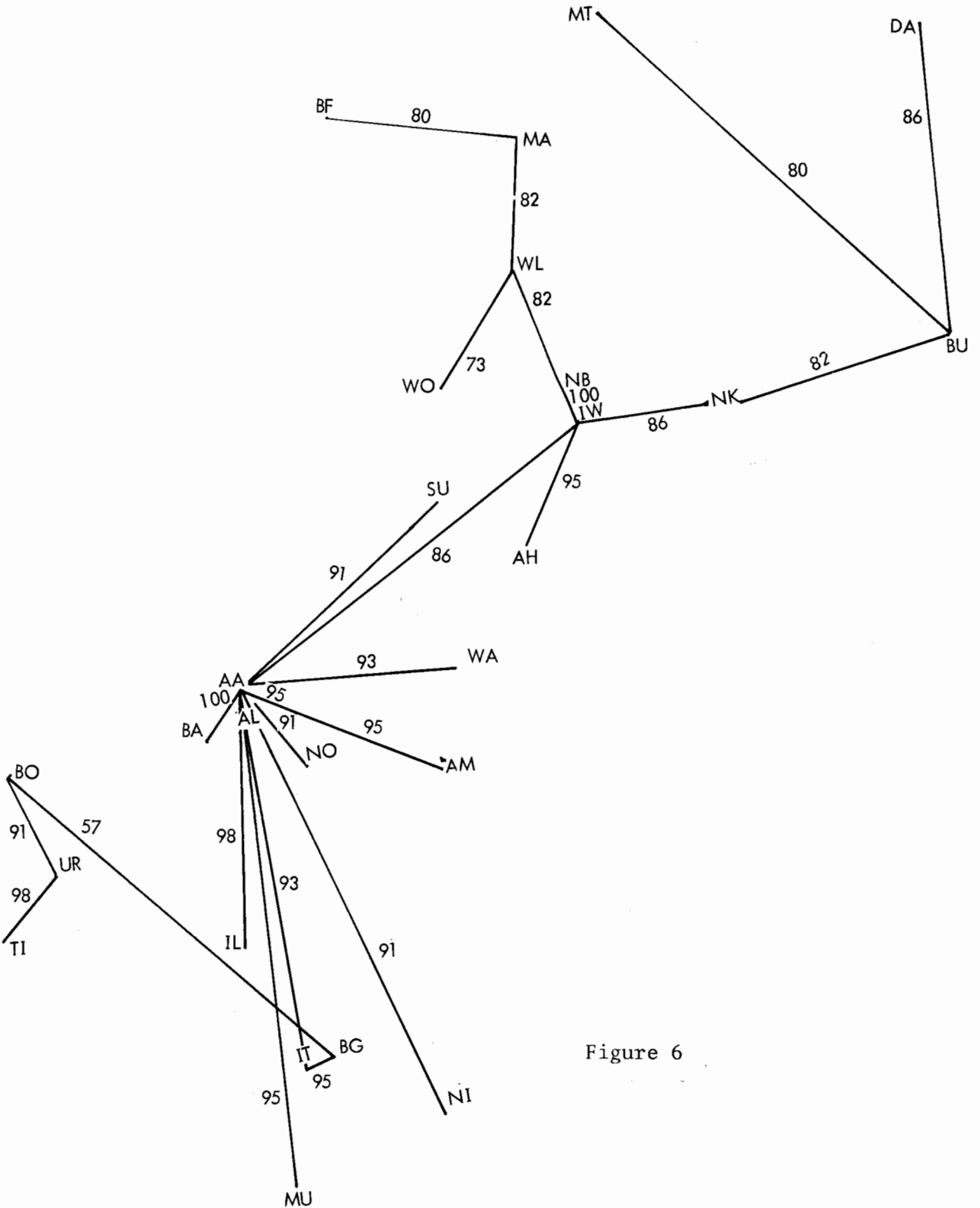


Figure 6



A SURVEY OF THE ARAPESH LANGUAGE FAMILY

CONNECTED LINE GRAPH - 72 WORD COMPARISON

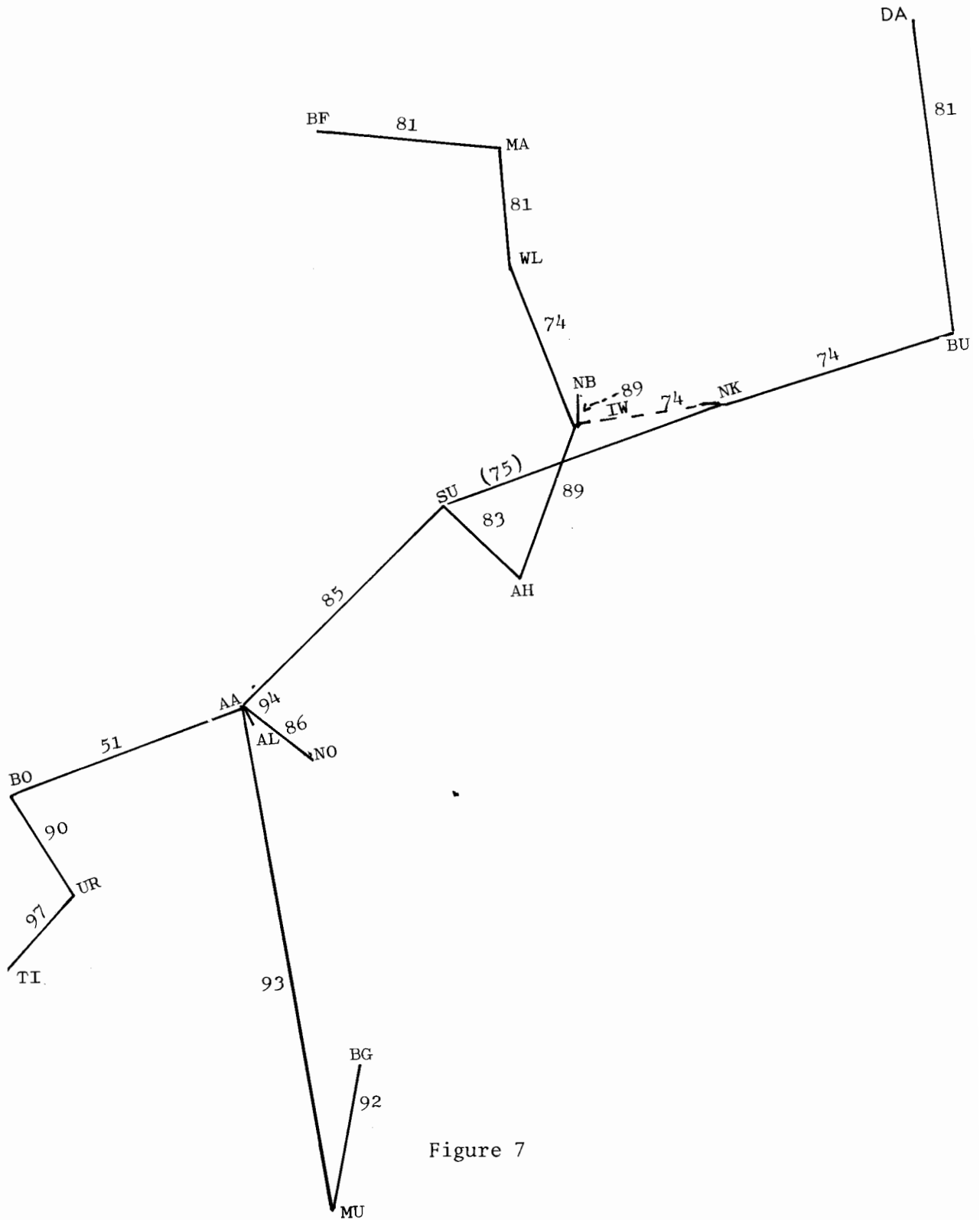


Figure 7

CONNECTED LINE GRAPH - 115 WORD COMPARISON

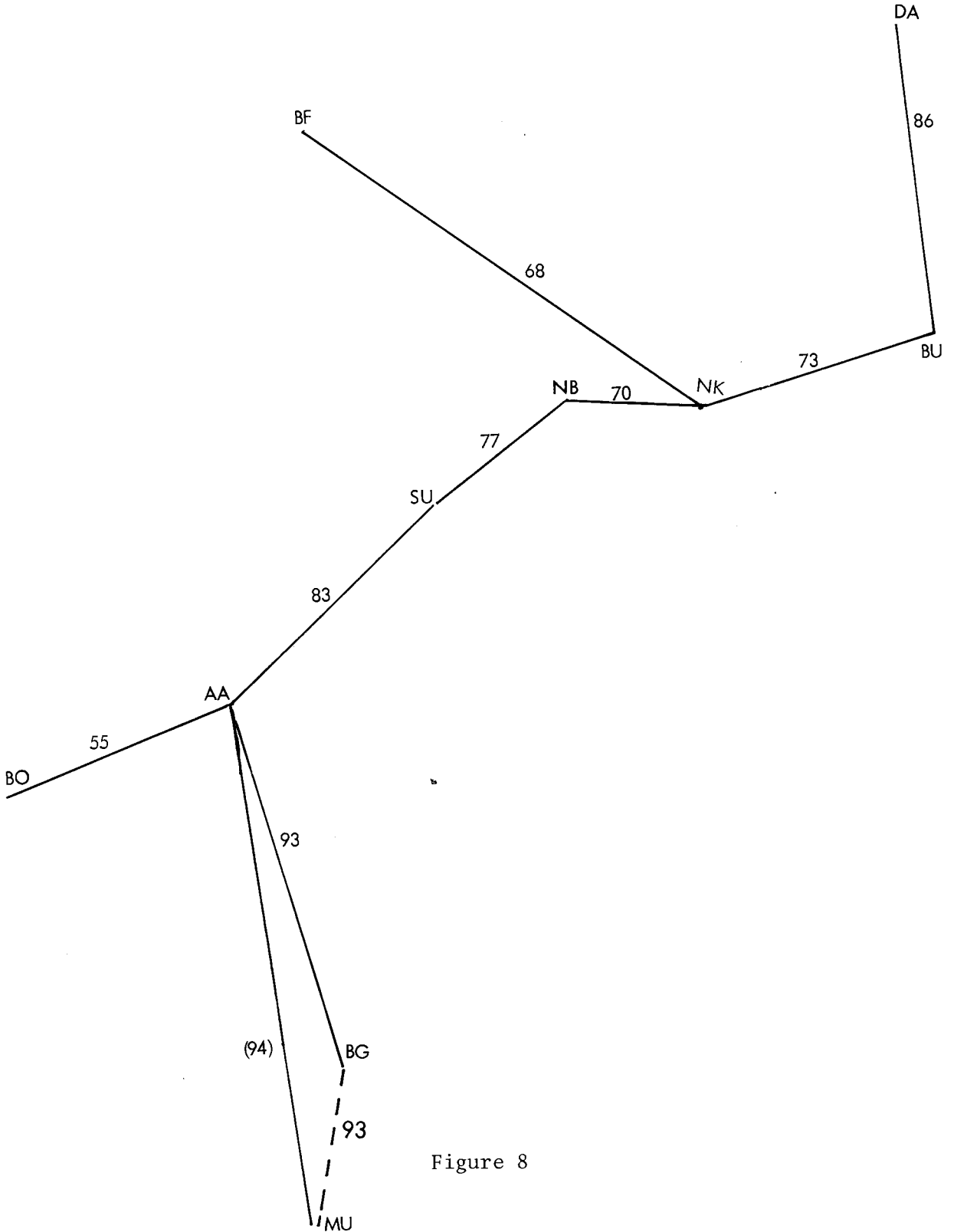


Figure 8





CENTRAL COMMUNITY OF LANGUAGE PATTERN - ARAPESH FAMILY  
 115 WORD COMPARISON

BO	3.3									
2	MU	3.3								
5	5	BF	3.1							
3	2	2	NB	2.2						
2	2	3	1	SU	2.2					
1	1	4	2	1	AA	2.4				
4	4	1	1	2	3	NK	2.4			
5	5	2	2	3	4	1	BU	3.1		
2	2	5	3	2	1	4	5	BG	3.3	
6	6	3	3	4	5	2	1	6	DA	4.0

Figure 11

The cognate percentages from Figure 1 are rearranged in Figure 12 to try and show the rough boundaries of the various dialects of Muhiang and the separate language Bumbita. A quick look at the percentages and the overlap of the lines enclosing a dialect indicate that there is a significant amount of dialect chaining and that some other approach should be used.

Gary Simons (Simons 1977:123 ff) has suggested that the matrix of percentages be rearranged with the average percentage of shared cognates indicated on the diagonal and the highest average positioned in the middle of the matrix. This has been done in Figures 12, 13, and 14 for the three comparisons of 44, 72 and 115 words respectively. Since this is a procedure to show dialect chaining, the three Bumbita villages (BO, UR, and TI) have been left out in making these calculations. Note that in Figure 13 (the 72 word comparison) AH has an average of shared cognates equal to AA, while in the other two comparisons AA comes out as the highest, even though BA equals the AA average in Figure 12. It is also interesting that the average for AA goes down from 85 using 44 words to 76 using 72 words to 75 using 115 words. In general the other villages with the higher averages remain consistently the same (villages) throughout the three comparisons.



## PERCENTAGE OF AVERAGE COGNATES WITH BUMBITA OMITTED - 72 WORDS

BF 62  
 81 WL 67  
 57 63 NO 71  
 53 57 81 BG 73  
 56 58 82 92 MU 74  
 57 64 86 88 90 AL 75  
 58 63 85 90 93 94 AA 76  
 63 71 76 78 79 82 82 AH 76  
 60 74 72 75 76 79 78 89 IW 75  
 58 67 72 85 90 82 85 83 79 SU 74  
 57 68 71 74 76 75 76 88 89 81 NB 72  
 68 71 61 69 69 67 69 72 74 75 72 NK 70  
 60 61 61 69 68 65 68 68 71 72 67 74 BU 68  
 81 81 54 51 54 56 57 63 61 57 58 71 71 MA 66  
 64 63 57 60 56 60 63 63 61 64 65 65 65 64 DA 63

Figure 13

## PERCENTAGE OF AVERAGE COGNATES WITH BUMBITA OMITTED - 115 WORDS

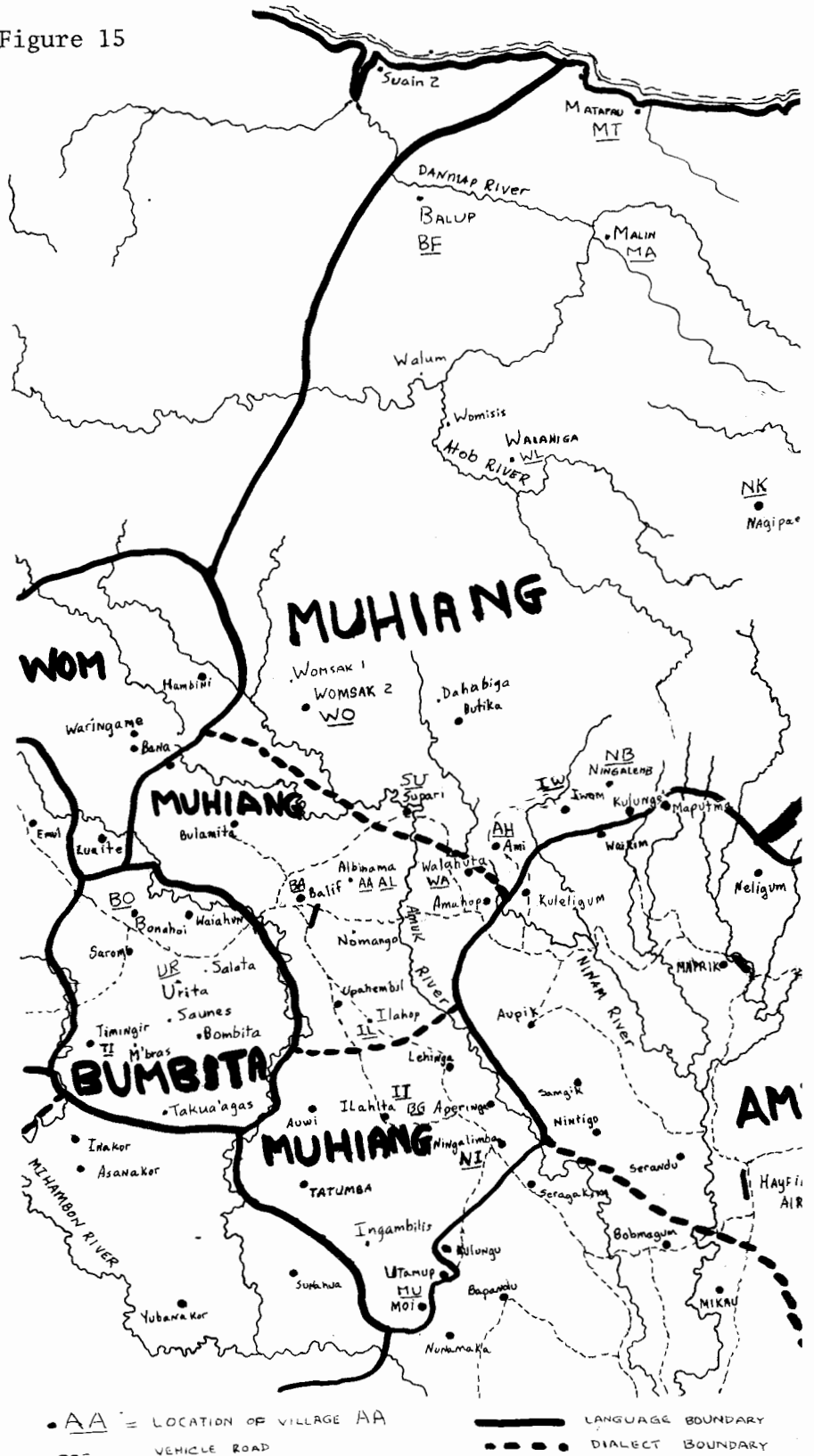
BF 58  
 55 NB 68  
 54 77 SU 71  
 56 73 80 MU 74  
 58 75 83 70 AA 75  
 54 72 82 93 93 BG 74  
 68 70 69 69 70 69 NK 70  
 73 73 66 67 69 69 73 BU 68  
 61 66 61 60 63 61 66 86 DA 64

Figure 14

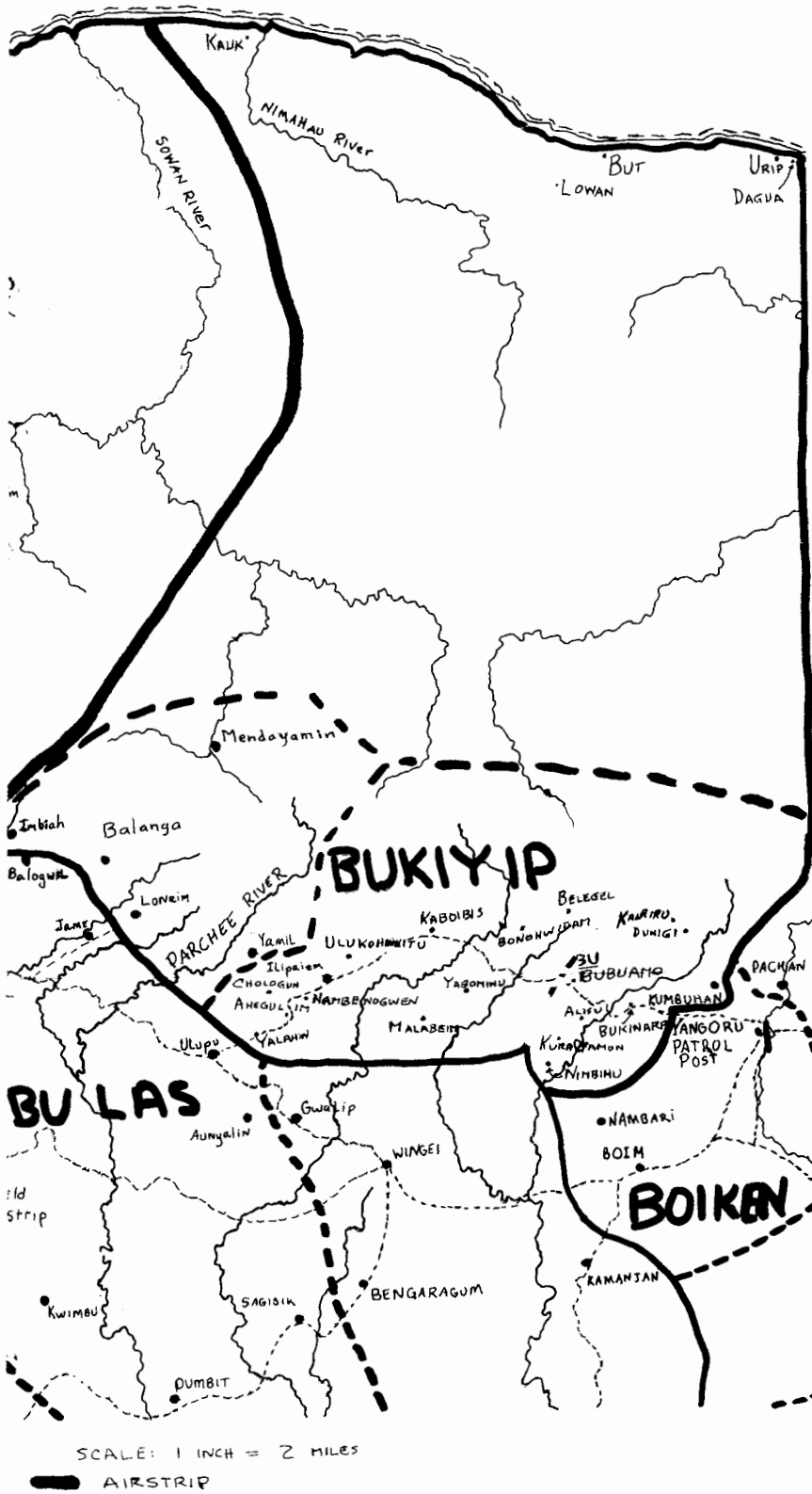
The results of the survey are also shown in the map listed as Figure 15. It is easy to see from the map that Bumbita - villages BO, UR, and TI - is a separate language. This is indicated by the solid line separating it from Muhiang. It is confirmed by the reports of the local people and by the lexicostatistic percentages in Figure 1. The highest percentage of any Bumbita village with any other village is the 57 between BO and IW. By contrast, the percentages between the Bumbita villages range from 89 to 98.

The rest of the map shows that the situation is less clear for Muhiang. The lexicostatistic percentages and the connected line graphs show that the Muhiang language is one long dialect chain going all the way north until it meets the ocean and east until it meets the last Arapesh village of Bubuamo. There is in fact no real place to divide the Muhiang and Bukiyip (Mt. Arapesh) languages. Nagipeim,

Figure 15







with isolation and no percentage higher than 86 with any other language, is chosen somewhat arbitrarily as the boundary. Dotted lines are drawn to indicate the Ilahita dialect and the Iwam-Ami-Ningalemb dialect. The Walihiga-Malin-Balup dialect is also so indicated. This leaves Womsak as a sort of isolated dialect, as confirmed by the fact that its highest lexicostatistic percentage with any other village is 73.

### 3 EVALUATION

The two national translators from AL and IL can converse and communicate in their respective vernaculars with great ease and efficiency. This close relationship is confirmed by the 93% probable cognates shown in Figure 1. Therefore the scores of 95% intelligibility of AL at IL and 100% intelligibility of IL and BA (virtually identical with AL) were to be expected. It was surprising that TI-UR, AH, and IW scored so much higher (80, 92½, and 90) on the IL test (80, 92½, and 90 respectively) than on the AL 3 test (42½, 72½, 62½ respectively). This can probably be explained by the fact that the speakers used for the test of intelligibility with IL had been living at the same Bible school with other IL speakers for a year or two and had learned the changes at IL. Therefore the tentative conclusion is that the scores of these three areas with respect to AL represent the true intelligibility much more accurately than their scores with respect to IL. It would have been better to use speakers who had not had this Bible school contact.

There is some indication that a number of half correct responses correlate with low intelligibility and that half correct responses should not be counted. The best example of this is the comparative response of the speakers from BO and TI-UR. These three villages are very similar linguistically. (This can be seen in Figure 2 where BO scores 90% with the two others which themselves score 97% between each other.) However, the intelligibility with AL was only 42½% for TI-UR while it was 57½% for BO. The higher figure for BO includes 7 responses scored as half correct, while only 3 such responses were scored in this way for TI-UR.

It was unfortunate that we were not able to obtain a list from Womisis, due to lack of time available on the helicopter.

With the exception of these problems listed above, the scores on the cognate percentages generally correlated quite well with the scores on the intelligibility test. This can be seen by the following comparison:

Ranking of Intelligibility with AA (Albinama 3)	Ranking by Percentage of Probable Cognates with AA
BA	BA
SU	IL
MU	MU
IT	AM
IL	IT
AM	WA
NI	SU
WA	NI
AH	IW
IW	NB
NB	AH
BO	BO
TI-UR	UR
	TI

#### 4 CONCLUSION

The general conclusion from the lexicostatistic testing is that Muhiang consists of one very long dialect chain running north from Muwi through Ilahita, Balif, Walahuta, Supari, and Womsak all the way to Matapau on the coast. The chain also continues on eastward from Womisis through Walahiga and on to Nagipeim, Imbia, Lonem, Yamil, Kaboibis, and finally Bubuamo. Along the coast the chain continues eastward from Matapau to Kauk, Bae, But, Dagua, and Woginara. The chain consists of such gradual changes that even at the extreme eastern end BU (Bubuamo) is still 69% cognate with AA (Albinama 3). It is actually difficult to say where Muhiang ends and Bukiyip begins, although probably Nagipeim (NK) is as good a place as any to make the border. The intelligibility test data confirm that intelligibility with Albinama 3 gradually decreases going north-east until it is down to 60% at Ningalemb. An earlier version of the test, testing intelligibility of Namango (NO) at Bubuamo, indicated 35% correct.

The general conclusion from the intelligibility test is that the people from WA, AH, IW, and NB will have some difficulty in using the Albinama 3 materials, while the Bumbita speakers (at BO, TI, and UR) will find these materials virtually impossible. Note the higher range of scores at WA, AH, IW, and NB: from 82½% to 60%. On the other hand, TI and UR scored 42½% and the BO score of 57½% is probably much too

high, as was indicated earlier.

The cognate percentages arranged to show language dialects and families, using the 44 word lists, are shown in Figure 12. The subgroupings are indicated by lines. The Bumbita language is clearly shown at the bottom, and also the Bukiyip (Mt. Arapesh) with its dialect variations. WO has an unusual relationship with cognate percentages of 68 with both IL and AH. The various dialects of Muhiang are shown at the top half of the chart.

The sociolinguistic data in general confirms the results of the intelligibility test. WA and SU seem to have contact with villages reasonably near Albinama 3. However, NB and AH refer to Muhiang as a "slightly different" language, while Iwam #3 calls it "another" language. The lack of contact between Albinama 3 and AH, NB, and IW confirms that the people in those areas will have some difficulty with the Albinama 3 materials.

The estimate of the number of additional translations needed is somewhat problematic. It is clear that a separate translation will be required for Bumbita. The lexicostatistic percentages with the range of 90 to 97 in Figure 2 seem to indicate that one translation is sufficient. The national translator should probably come from Urita, as it is the most linguistically and geographically central.

For the rest of the family, the situation is less clear because of extreme dialect chaining. If another translation is needed for the area north-east of Balif, a national translator from Supari or Iwam should be chosen. Before this is attempted, an honest effort should be made for the people to use the Albinama 3 materials.

The situation for the coastal area is just as problematic. A separate translation will probably be needed. The national translator should be from either Dagua, Malin, or Matapau. Each location has its advantages. Possibly a list from But, or Kauk, or Lowan will indicate that one of those villages will be the best compromise. Probably the national translator could work from the Bukiyip materials, but this is not certain.

The connected line graphs in Figures 6, 7, and 8 and the matrices showing the central community of the language pattern in Figures 9, 10, and 11 confirm AA (Albinama 3) as a quite good choice for a national translator in terms of centrality. However, these graphs and matrices also indicate that SU (Supari), IW (Iwam), AH (Ami), and NB (Ningalemb) are possible choices. One of these others should probably be used as a source of another national translator if it turns out that the translation from AA is not intelligible in the area to the north.

## NOTES

- <sup>1</sup> Laycock, D. C., 1973. Sepik Languages - Checklist and Preliminary Classification. *Pacific Linguistics B* 25.
- <sup>2</sup> The name Southern Arapesh is taken from Glasgow, D. and Loving, R., 1964, *Languages of the Maprik Sub-District*. The people in the Balif and Ilahita areas all say the name of their language is Muhiang.
- <sup>3</sup> These figures are from Laycock, D. C., 1973, as listed in Note <sup>1</sup> The 10,300 figure for Bukiyip includes the villages south into the hills from Matapau, where the range of lexicostatistic percentages is from 68 to 80 inclusive (see Figure 1).
- <sup>4</sup> Special thanks are due to Mr. John Champ, who piloted the helicopter. Without his skill and assistance this part of the survey would have been very difficult or impossible.

I wish to thank Mr. Maarten te Hennepe for his help in this part of the analysis.

## BIBLIOGRAPHY

- Glasgow, D. and Loving, R., 1964. *Languages of the Maprik Sub-District*. Port Moresby: Department of Information and Extension Services.
- Laycock, D. C., 1973. Sepik Languages - Checklist and Preliminary Classification. *Pacific Linguistics B* 25.
- Loving, R., 1977. "Guidelines for Writing Up Language Surveys". *Workpapers in Papua New Guinea Languages* 21:317-24.
- Sanders, A., 1977. "Some Synchronic Analysis Procedures for Language Survey Data". *Workpapers in Papua New Guinea Languages* 21:295-316.
- Simons, G., 1977. "Recognizing Patterns of Divergence and Convergence in a Matrix of Lexicostatistic Relations". *Workpapers in Papua New Guinea Languages* 21:107-34.