

THE LANGUAGES OF KEI, TANIMBAR AND ARU:
A LEXICOSTATISTIC CLASSIFICATION

by

Jock Hughes

Summer Institute of Linguistics

and

Pattimura University, Ambon

This article is a lexicostatistic analysis of data gathered from all the indigenous languages of the Kei, Tanimbar and Aru Islands of south-east Maluku.

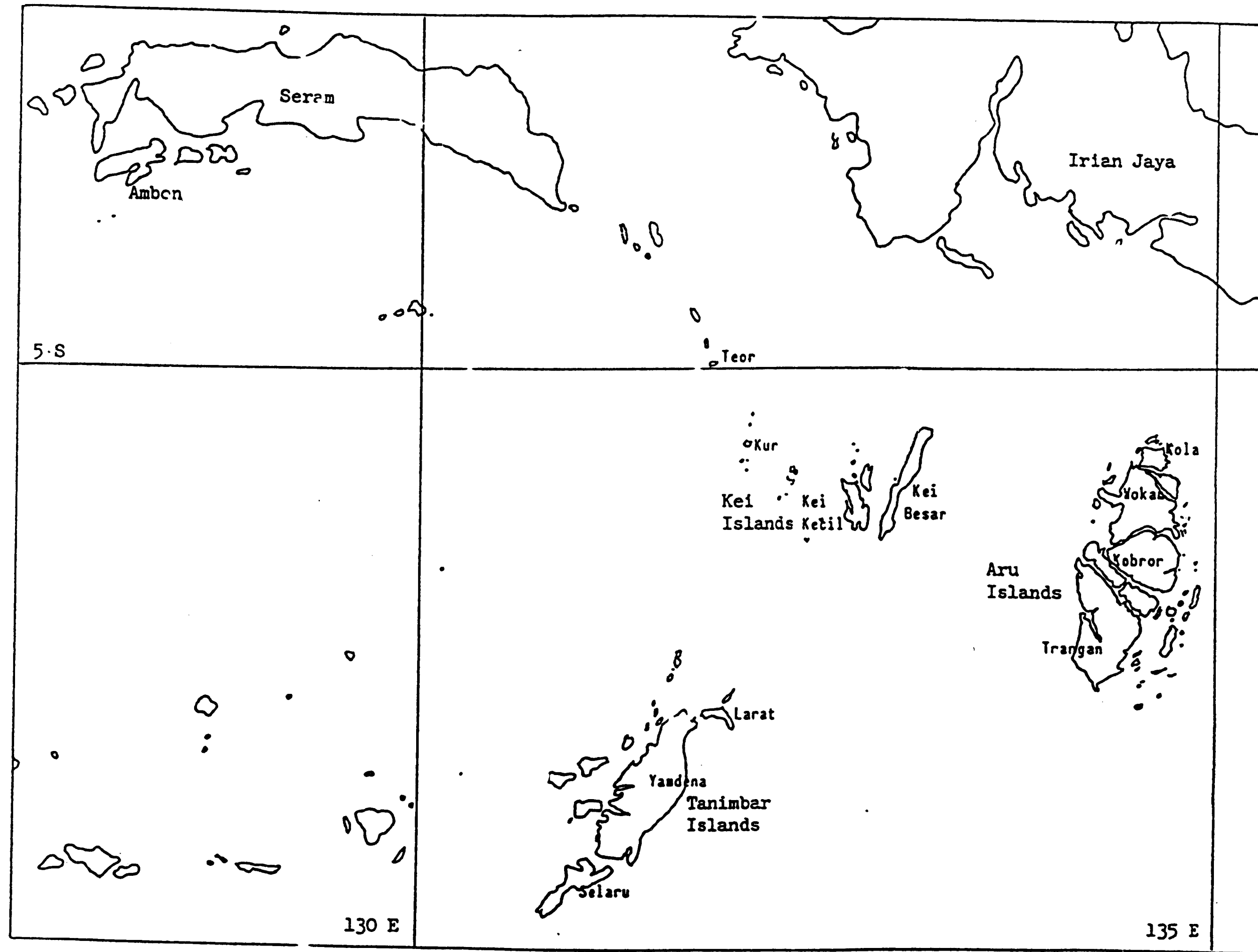
It is explained that the aim of this article is to give a synchronic rather than a diachronic understanding of the relationships between the languages studied. The method of data collection is described as well as that of determining the cognates and classifying the languages.

0. INTRODUCTION

This paper is a report and analysis of the findings of a language survey conducted under the auspices of the Summer Institute of Linguistics and Pattimura University, Ambon. The survey covered the three archipelagos in the south-east of the Indonesian Province of Maluku: the Kei Islands, the Tanimbar Islands and the Aru Islands (See map 1).

To date very little classificatory work has been done on the languages of Kei, Tanimbar, and Aru. James T. Collins, whose main research has been in the Central Moluccas, includes a short section on the South-East Moluccas in Linguistic Research in Maluku: A Report of Recent Field Work, to which reference will be made in this paper. Collins' classification as presented in this work is that used in Wurm and Hattori's Language Atlas of the Pacific Area (1981). Kei, Tanimbar, and Aru were also covered in Salzner's Sprachenatlas des indopazifischen Raumes (1960), but I will not be referring to his classification save to say here that it differs considerably from mine. Other references on the languages of the area are a short section on the languages of Tanimbar as part of Susan McKinnon's PhD thesis, Hierarchy, Alliance and Exchange in the Tanimbar Islands (1983), and various wordlists and language samples by Dutchmen: Brumond (1843) on Aru, Eijbergen (1865) on Aru and Kei, Geurtjens (1912) on the Kei language, and Drabbe (1932a, 1932b and 1932c) on the Fordata, Yamdena and Selaru languages respectively.

The purpose of the survey was to define the linguistic situation as it is today in these islands; that is, to discover what and how many languages are spoken and where the language boundaries lie. The aim was also to investigate possible areas for in-depth descriptive studies of individual languages by linguists working under the Cooperative Programme between Pattimura University and the Summer Institute of Linguistics (UNPATTI-SIL). Two such studies have now begun in the area with a third in prospect.



MAP 1: South-East Maluku

The results of this survey will not be the final word on the languages in the area; it is intended that the above mentioned linguists of the Summer Institute of Linguistics and those who follow them will carry out more detailed language and dialect surveys in the immediate area of their study. These surveys will answer some of the questions as yet unanswered and give an even clearer picture of the linguistic situation in the area.

1. PROCEDURES

1.1 Method

The method used was that of lexicostatistics from a synchronic perspective since the purpose was to measure the lexical similarity between the languages at this point in time rather than to establish the historical relationships of the languages. When lexicostatistics is employed with a diachronic perspective the linguist is concerned to establish the genetic relationship between languages through time. Therefore words borrowed into the language have to be eliminated, since they do not reflect the historical relationships of the languages. However from a synchronic perspective borrowed words do not need to be eliminated as we are concerned to establish the relationship between speech communities as related to intelligibility. This is done by counting the percentage of similar words in a representative sample of the lexicon. It is the phonetic similarity of a word in one language with the same word in another language that allows comprehension between the two speech communities (Sanders 1977:34). Therefore, in deciding whether or not words were lexically similar (cognate for our purposes), I compared the phonetic similarity of the two words. I took note of regular sound correspondences and judged the words to be phonetically similar when the changes were due to these regular correspondences. This procedure is consistent with the method being used, since speakers of the languages easily recognise many regular correspondences. For example I have heard comments like, 'We use a "g," but they use an "h".' Inevitably the use of the cognate set method (described below) means that there will be chains of lexically similar words where the words at each end of the chain will not be very similar phonetically. This is because, rather than comparing wordlists in pairs, each item in a given wordlist is compared with that item in all the other wordlists simultaneously. It is recognised that lexicostatistics is not a highly precise tool for classification and any given percentage of lexical similarity is best thought of as an approximate value. However it is a very useful tool for use in such a diverse linguistic situation covering a wide geographical area, and it enables one to make a valid classification of languages, especially when patterns in the matrices are taken into account.

1.2 Data Collection

1.2.1 The Wordlist

The basic wordlist used was a 210-item wordlist developed for use by UNPATTI-SIL teams for linguistic survey throughout the Moluccas. It incorporates the standard Swadesh 100-word list and much of the Swadesh 200-word list, with a few additions thought to be appropriate to the Moluccas. Of the items on the Swadesh 100-word list, horn, heart and seed were eliminated due to previous experience of confusion and hesitation

upon elicitation, and feather was eliminated due to its polymorphemic form both in Indonesian, the language of elicitation, and in many of the vernaculars; both morphemes were elicited separately. After elicitation seven words were eliminated from the 210-word list, in six cases due to morphemic duplication with other items, and in one case because of consistent confusion on the part of the informants when the word was elicited.² Please refer to appendix A for the resulting list of 203 words used.

1.2.2 The Survey

Fifty-two wordlists were collected, twelve in Kei, seventeen in Tanimbar, and twenty-three in Aru. However one of the wordlists from Aru was considered unreliable and was not used in the analysis, reducing the Aru wordlists to twenty-two, and the total to fifty-one. Most wordlists were taken in the village where the dialect or language is spoken, and all were taken near the area (see appendix B).

Each primary informant was screened before the wordlist was taken to ensure that he and his parents were native speakers of the language, born in the village in question. The majority of wordlists were elicited from a group of people rather than a single individual.

The survey was conducted in two parts by three field linguists, the first part, covering Kei and Tanimbar, by the author and Edgar Travis in January 1985, and the second, covering Aru, by the author and Yushin Tøguchi in March and April 1985. All the wordlists were recorded phonetically.

1.3 Data Analysis

1.3.1 Determining Cognates

For the purposes of a synchronic study, cognates are held to be phonetically similar words with identical or similar meanings,³ and are determined using the inspection method (Gudchinsky 1956, Sanders 1977).

For each item on the basic wordlist, the vernacular words were grouped into cognate sets. Each cognate set is a group of phonetically similar words with the same meaning.⁴ These cognate sets were entered into a computer. A program was then used to compute the percentage of cognate words between each pair of wordlists. The output from this program was a complete matrix of cognate percentages for all the villages.⁵ Another computer program was used to permute the matrix into the desired order.

1.3.2 Classification

There are no hard and fast boundaries for taxonomic levels, however I have decided to use the levels as used by Grimes and Grimes in their classification of the languages of South Sulawesi, adapted from Smith's taxonomic levels for his classification of the languages of Sabah. The range of cognate percentages covered by each of these levels is not absolute and allowance is made for patterns and groupings that emerge within the data. I have, for example, added the level of substock between stock and family since this fits the data for Kei and Tanimbar. The following are the groupings as they are being used:

Percentage of Shared
Lexical Similarity

Classification

0 - 15%	Separate linguistic phyla.
15 - 25%	Separate superstocks of a common phylum.
25 - 45%	Separate stocks of a common superstock.
45 - 55%	Separate substocks of a common stock.
55 - 60%	Separate families of a common substock.
60 - 75%	Separate subfamilies of a common family.
75 - 80%	Separate languages of a common subfamily.
80 - 90%	Separate dialects of a common language.
90 - 100%	Common dialect.

The following are some comments to exemplify the various taxonomic levels as they occur in the area of study:

- Phylum: All speech communities in this study fall within the Austro-nesian Phylum.
- Superstock: All speech communities in this study fall within the Central Malayo-Polynesian Superstock (following Blust 1977).
- Stock: There are four stocks represented in the data, the Kei-Tanimbar Stock, the South Tanimbar Stock, Banda which shares a stock with no other language in the area, and the Aru Family, a stock-level grouping.
- Substock: Within the Kei-Tanimbar Stock I have posited the Kur-Fordata Substock, as its members, while not closely enough related to be of the same family, are much more closely related to each other than they are to the Yamdena Language, which is still a member of the stock. For the same reasons I have posited the South-West Tanimbar Substock within the South Tanimbar Stock.
- Family: The languages of Aru are all members of the Aru Family, and the Kei-Fordata Family consists of two members. All the other languages are family-level isolates, that is, they relate to all other languages at less than 60%.
- Subfamily: Within the Aru Family there are a number of languages that relate to each other at subfamily level. Members of a common subfamily are usually approaching the threshold of intelligibility, and whether or not they are mutually intelligible needs to be established using other methods such as dialect intelligibility testing (Casad: 1974), and sociolinguistic surveys.
- Language: Speech communities falling within this range are typically mutually intelligible, although, as with members of a subfamily, this needs to be checked using other methods.
- Dialect: Speech communities relating to each other above 90% are almost certainly mutually intelligible. Percentages of lexical similarity are, however, only a very tentative method of establishing dialect boundaries; many other factors need to be taken into account such as consistent phonological changes between cognate words.

1.3.2.1 Classification where chaining occurs

In these islands chaining is rare at the language level, however at higher levels of the taxonomy there are clear patterns of chaining, which

make classification difficult at the level in question. I have dealt with this in the following ways:

(a) Chaining at subfamily level

In the Aru Islands chaining occurs at subfamily level, but it is very difficult to posit subfamilies because the ends of the chain relate to each other well below subfamily level; indeed the relationship at subfamily level is only to the immediately neighbouring language in the chain. I consider that in this case trying to posit subfamilies serves no useful purpose, but rather I simply state that all the languages of Aru belong to the Aru Family. I then use the matrices and tree diagrams to display where the subfamily relationships occur, and indeed how the members of the chain do interrelate with one another.

(b) Chaining at stock level

There is chaining at stock level between the Kei-Tanimbar Stock and the South Tanimbar Stock. The Yamdena Language, a family-level isolate, relates to members of both stocks at stock level. I have illustrated this relationship by saying that Yamdena is a member of both stocks. I thus avoid the problem of which stock to put Yamdena in, and at the same time maintain the distinctiveness of the two stocks.

Chaining does not occur at other levels of the taxonomy within these data.

1.3.2.2 Classification on the periphery of the surveyed area

There are two speech community groupings that cannot be fully classified solely on the basis of data from within the area surveyed.

The first of these is the Banda Language. From our data it can be determined that Banda does not share a stock with any of the other speech communities in the Kei, Tanimbar or Aru Islands. James T. Collins (1983c:122) suggests that Banda might be subgrouped with the languages of Geser and Watubela in East Seram. Some data from these languages were obtained and compared with the Banda wordlist. This will be discussed in section 2.1.1.1 below, under Banda.

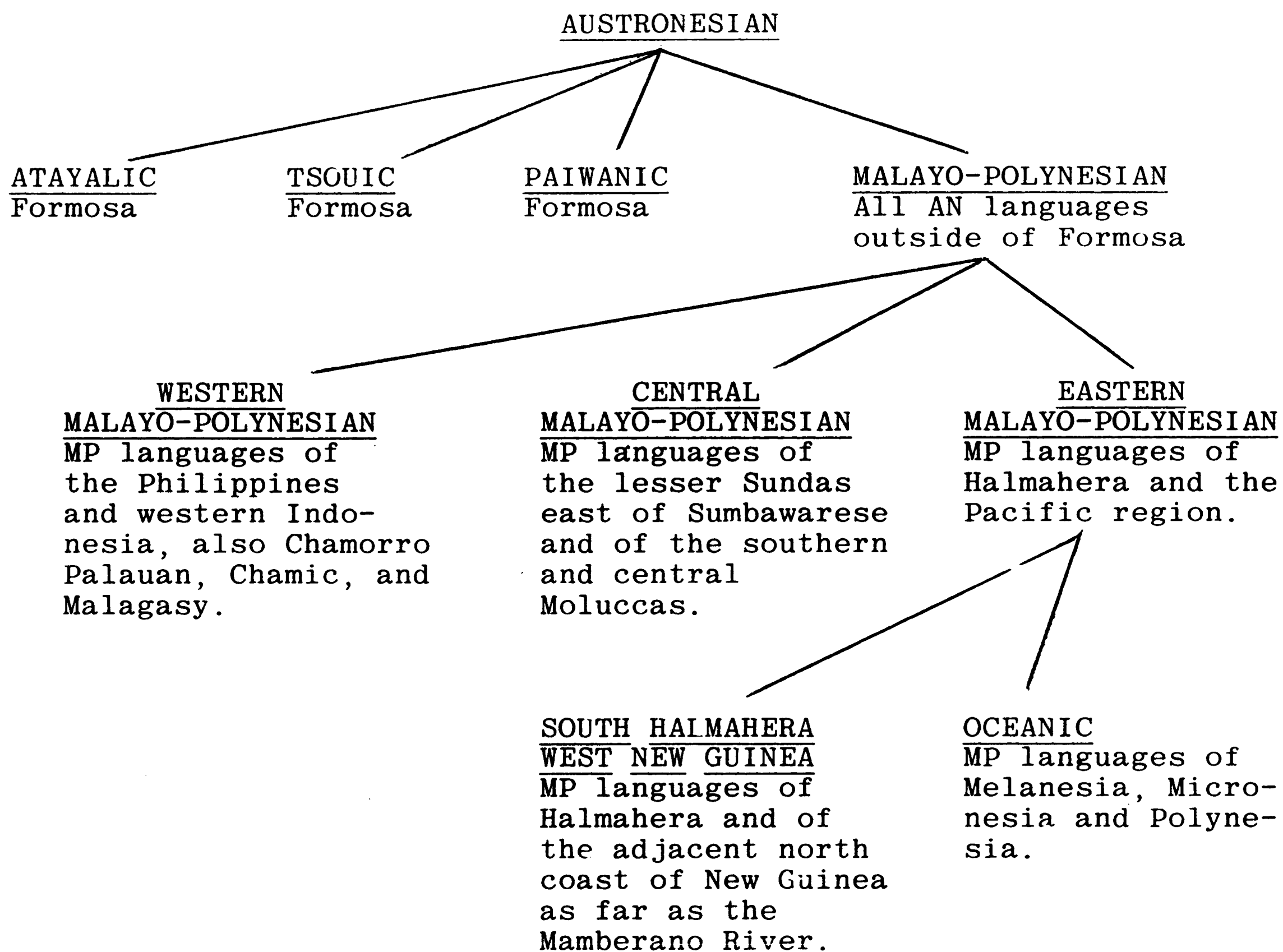
The second area that cannot be classified for stock without recourse to data from outside the area is the entire Aru Family. Again, from our data we can state that the languages of the Aru Family do not share a stock with any of the languages of Kei or Tanimbar. If the Aru languages were to be more closely related to any languages outside the South-East Moluccas one might expect them to be related to the Austronesian languages spoken on the southern part of the Bird's Neck of Irian Jaya. Data were therefore requested from the Cenderawasih University and Summer Institute of Linguistics co-operative programme in Jayapura. Wordlists were sent for three Austronesian languages spoken in the Kaimana area and the coastal areas of the Bomberai Peninsula, Irarutu, Uruangnirin and Koiwai. These data were compared with our data from the South-East Moluccas, and I was thus able to establish the position of the Aru Family (see section 2.2.1).

2. THE LANGUAGES OF KEI, TANIMBAR AND ARU

The data presented in this paper show all the languages of the three archipelagos covered, Kei, Tanimbar and Aru, to be members of the same phylum, that is, the Austronesian Phylum. They also show all the languages to be members of the same superstock. This is shown in matrix 1 by the fact that all the percentages of lexical similarity are 24% or above, although the figures between the Selaru Language of Tanimbar and the languages of Aru are the lowest they could be whilst remaining in the same superstock. This superstock I shall call the Central Malayo-Polynesian.

Superstock, following Blust (1977, 1981) (see figure 1, which is taken from Grimes and Grimes:1984).

FIGURE 1: Blust's (1981) Divisions of Austronesian

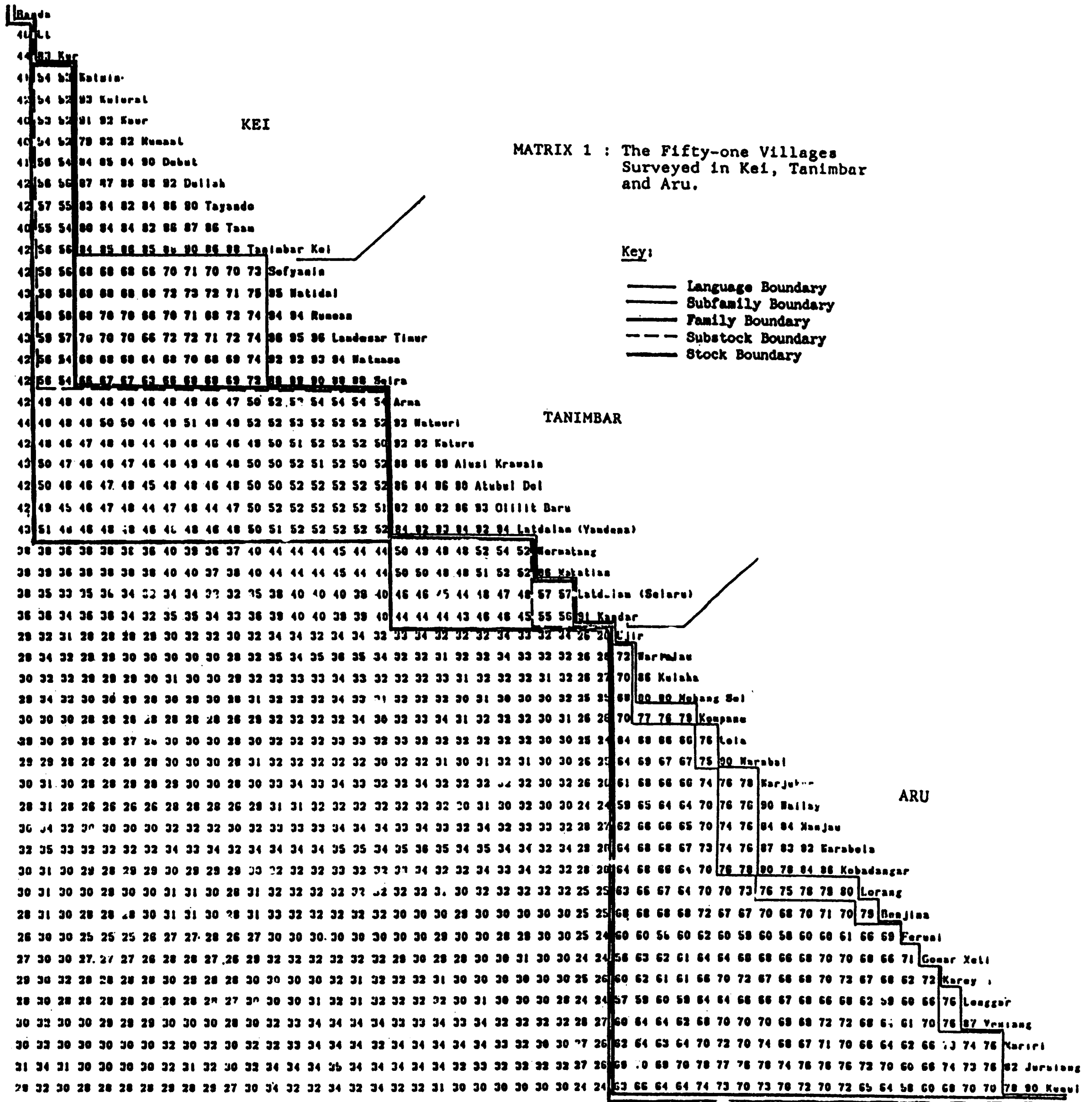


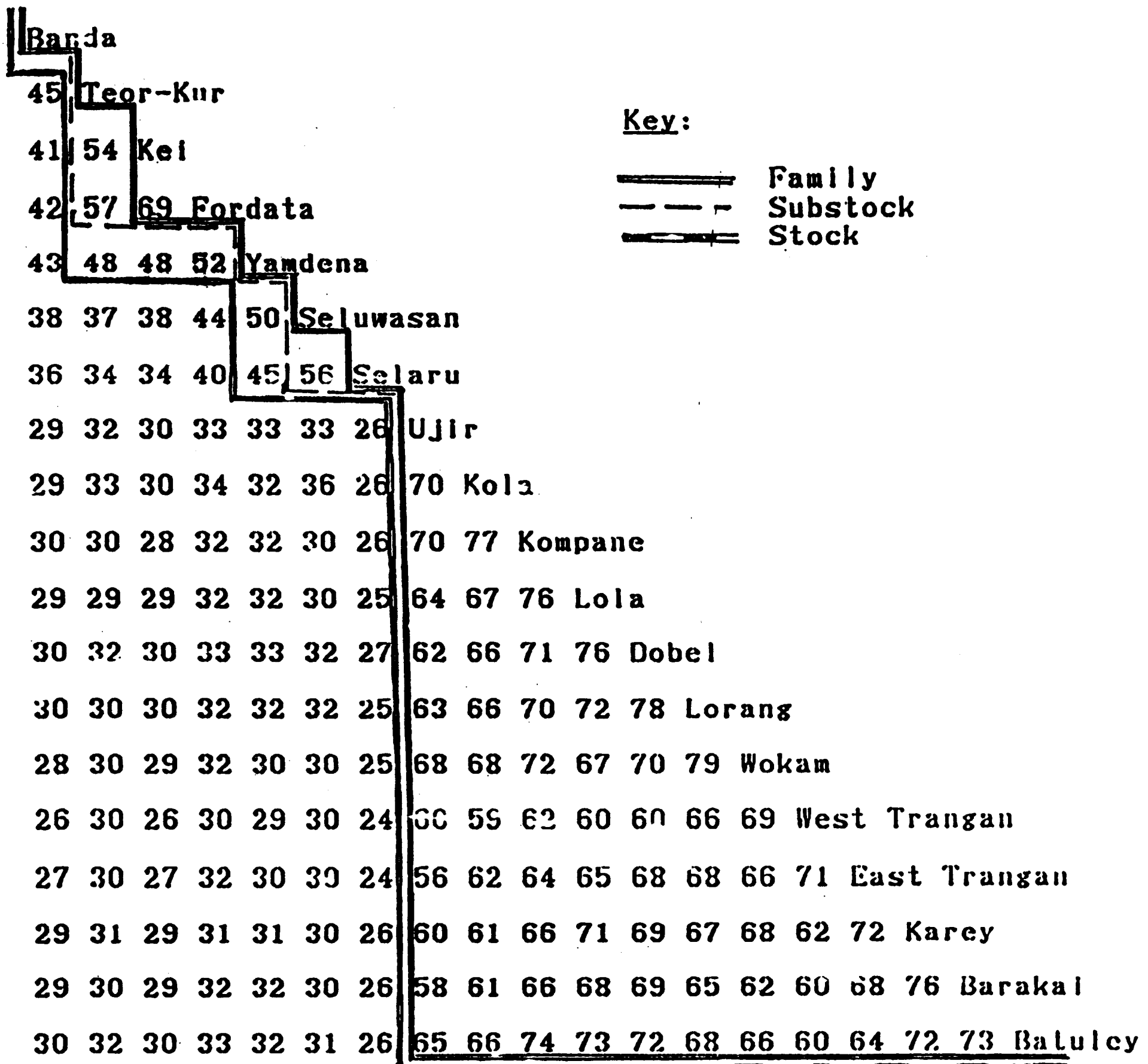
Matrix 1 (p. 78) shows the percentage of lexical similarity between all the villages surveyed and the different taxonomic levels as demarcated by lines that have been drawn on the matrix. To ease the reader in comparing any two languages, however, matrix 1 has been reduced, by grouping the villages into languages and averaging all the figures between the villages surveyed in the two languages. This gives us an average percentage of lexical similarity between the two languages, as represented by the villages surveyed. These percentages are shown in matrix 2.

2.1 The Languages of Kei and Tanimbar

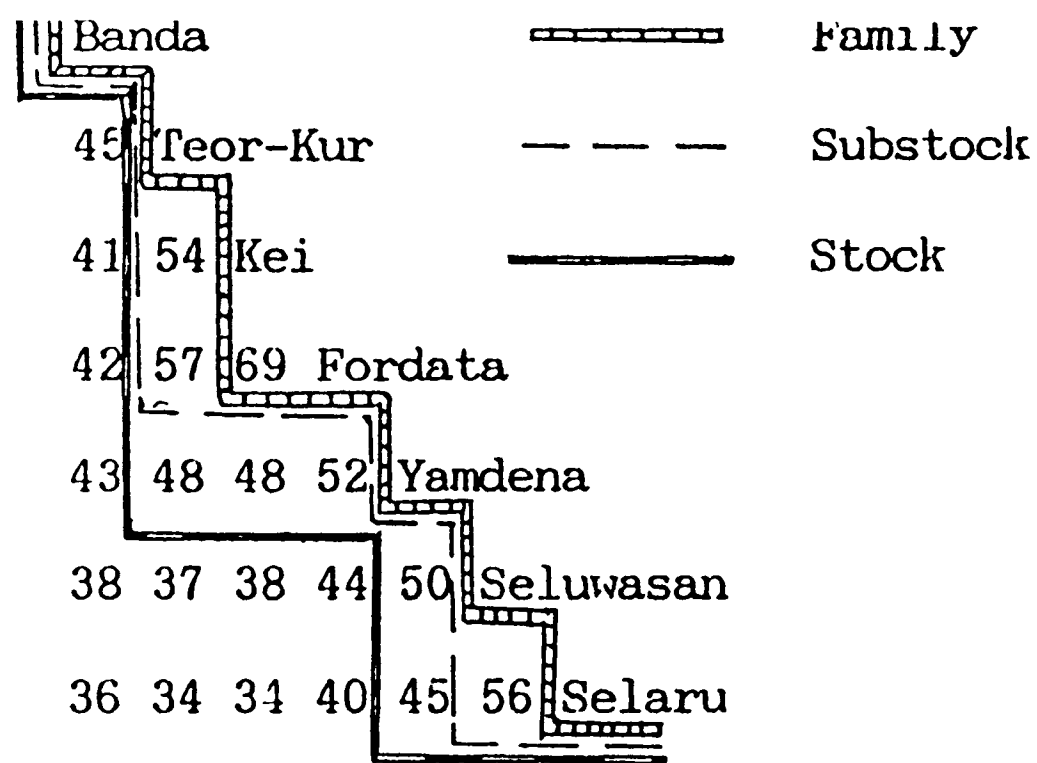
2.1.1 Language Relationships within Kei and Tanimbar

Matrix 3 (below) is for reference throughout this section. It is a portion of the reduced matrix for the whole area (matrix 2) in which the relationships for Kei and Tanimbar are shown language by language.





MATRIX 2 (Reduced) : The Languages of Aru, Kel and Tanimbar



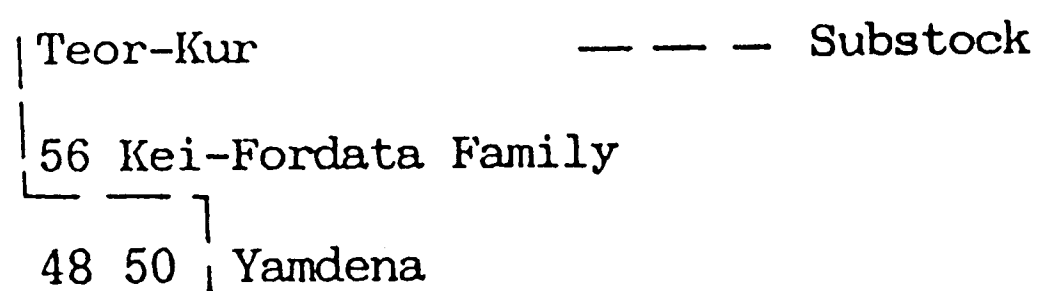
MATRIX 3 (Reduced): The Languages Kei and Tanimbar

2.1.1.1 The Banda Language

The Banda Language is spoken only in two small linguistic enclaves on the island of Kei Besar. However, of more relevance to its relationship with other languages is the fact that the people were fugitives from the Banda Islands some three centuries ago and are thus relative newcomers to Kei Besar. Banda relates to other languages in Kei, Tanimbar or Aru only at superstock level, that is, it does not share a stock with any of them. It relates to members of the Kei-Tanimbar Stock with a percentage of lexical similarity in the low forties, to members of the South Tanimbar Stock in the high thirties, and to the languages of Aru in the high twenties (see matrix 2). Collins (1983c:122) suggests that the Geser-Goram and Watubela languages and Banda may be branches of a single proto-language, Proto-Banda. Data were therefore needed from this area to establish the relationship of Banda to these languages. I was able to obtain a limited amount of data for Geser and Watubela, which I then compared with the Banda data. About 120 words were compared. The results show that Banda is not closely related to either language. It just borders on a stock-level relationship with Geser (46%) and is more distantly related to Watubela (40%). Geser and Watubela appear to relate to each other at family level (64%). On the basis of this information Banda may still be considered a stock-level isolate as its average relationship with the Geser-Watubela Family is 43%, indicating separate stocks. Further comparison of Banda to other East Seram languages might shed more light on Banda's position.

2.1.1.2 The Kei-Tanimbar Stock

The Teor-Kur Language, the Kei-Fordata Family, and the Yamdena Language make up what I am calling the Kei-Tanimbar Stock. All these languages are more than 45% lexically similar to each other, however Teor-Kur and the Kei-Fordata Family relate to each other at a percentage which is approaching the family level and is considerably higher than the percentages at which Yamdena relates to them both. This is shown in matrix 4 (below), in which the total percentages for each family, or family-level isolate have been averaged. I am therefore positing the Kur-Fordata Substock, which consists of the Teor-Kur Language and the Kei-Fordata Family.



MATRIX 4 (Reduced): Family-level Units in the Kei-Tanimbar Stock

2.1.1.2.1 The Kur-Fordata Substock

(a) The Teor-Kur Language

Teor-Kur is spoken in the Kur Islands to the west of Kei and on the island of Teor to the north. Some of the inhabitants of Teor moved to Ut Island near Kei Kecil, from where one of our wordlists was taken; the other was from Kur Besar, the main island in the Kur group. Teor-Kur is a family-level isolate, since it relates to no other language above 60%. The above mentioned data for Geser and Watubela were also compared with the Teor-Kur wordlists, showing Geser to relate to Teor-Kur at 49% and to Watubela at 43%, even though Watubela is geographically closer. With these limited data from East Seram it is unwise to draw too many conclusions, except to note that Teor-Kur has a considerably closer relationship to the Kei-Fordata Family than it has to the Geser-Watubela Family (average 46%).

(b) The Kei-Fordata Family

The Kei-Fordata Family consists of two languages, Kei, spoken throughout the Kei Islands, and Fordata, spoken in the north and west of the Tanimbar Islands. These two languages relate to each other at an average of 69%. The speech community of Tanimbar-Kei is the dialect of Kei that relates most closely to Fordata (average 74%), however it clearly remains a dialect of Kei (84% - 90% with other Kei villages). See matrix 5 (below) for a complete matrix of the villages surveyed in Kei and Fordata.

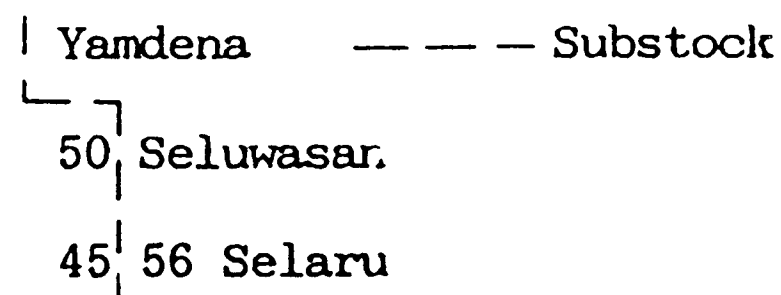
	Language
Watsin	
93 Wulurat	
91 92 Waur	
79 82 82 Rumaat	<u>Kei</u>
84 85 84 90 Debut	
87 87 88 88 92 Dullah	
83 84 82 84 86 90 Tayando	
80 84 84 82 86 87 86 Taam	
84 85 86 85 88 90 86 88 Tanimbar-Kei	
68 68 68 66 70 71 70 70 73	Sofyanin
69 68 68 68 72 73 72 71 75	95 Watidal <u>Fordata</u>
68 70 70 66 70 71 69 72 74	94 94 Rumean
70 70 70 66 72 72 71 72 74	96 95 96 Landesar Timur
68 68 68 64 68 70 68 69 74	92 92 93 94 Watnasa
66 67 67 63 66 68 68 69 72	88 89 90 88 88 Seira

MATRIX 5: Kei and Fordata Languages

Yamdena is a family-level isolate that together with the Kur-Fordata Substock constitutes the Kei-Tanimbar Stock. However Yamdena also relates to the Seluwasan and Selaru Languages at stock level. This two-way relationship is an example of convergence at stock level, which is illustrated in Figure 2, a tree diagram showing the interrelationship of the languages of Kei and Tanimbar. Although Selaru and Seluwasan relate to Yamdena at stock level, they are below stock level in their relationship with other members of the Kei-Tanimbar Stock (see matrix 3). So it will be seen that Yamdena, as well as being a member of the Kei-Tanimbar Stock, also forms a stock with Seluwasan and Selaru. This stock I have called the South Tanimbar Stock. The Kei-Tanimbar Stock and the South Tanimbar Stock are in a chaining relationship with Yamdena as the link in the chain.

2.1.1.3 The South Tanimbar Stock

The South Tanimbar Stock consists, as was shown above, of the Yamdena Language, a family-level isolate, and of the Seluwasan and Selaru languages, both also family-level isolates. As in the case of the Kei-Tanimbar Stock, Yamdena is less closely related to the other members of the stock than they are to each other (see matrix 6).



MATRIX 6 (Reduced): South Tanimbar Stock

2.1.1.3.1 The South-West Tanimbar Substock

From the data above I have posited the South-West Tanimbar Substock which consists of the Selaru and Seluwasan Languages, both of which are family-level isolates. The South-West Tanimbar Substock and the Yamdena Language together constitute the South Tanimbar Stock. The two members of the South-West Tanimbar Substock are both linguistically very isolated from other Austronesian languages. For both of them, their nearest linguistic neighbour is the other, and they relate at only 56%. To Yamdena Seluwasan relates at 50% and Selaru at 45%, the very boundary of the stock level. With all other languages in the area surveyed they share only a superstock. It is noteworthy that Selaru relates at a mere 25.5% (average) with the languages of Aru, the lowest possible percentage while still in the same superstock. This is about 5% lower than any of the other languages of Kei or Tanimbar in their relationship with the Aru languages.

2.1.2 Language Information for Kei and Tanimbar

In this section each of the seven languages found in the Kei and

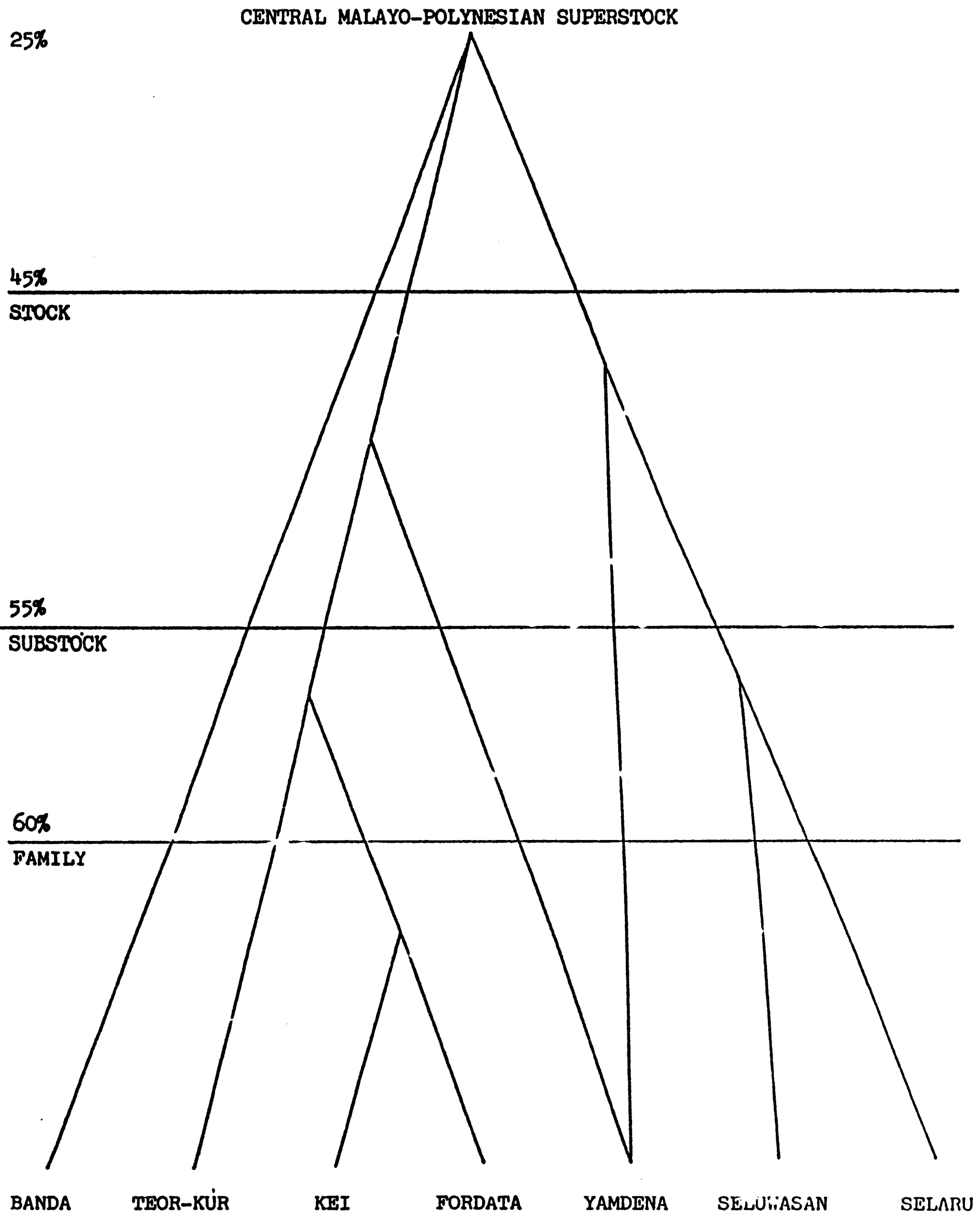


FIGURE 2 : The Languages of Kei and Tanimbar:
Interrelationships.

Tanimbar archipelagos is considered, giving limited information about the people, their location and language usage. There is also dialect information on each language, and some areas that need further study are indicated. The geographical extent of the languages found in the Kei and Tanimbar Archipelagos is shown on maps 2 and 3. These maps also show the villages from which the data were taken.

2.1.2.1 Banda

Banda is spoken on Kei Besar by approximately 2,300 people in the communities of Banda-Eli in the north-east, and Banda-Elat, the administrative capital for the Subdistrict of Kei Besar, in the central western part of the island. The people originally came from the Banda Islands (where the language is no longer spoken), but have lived in Kei Besar for the last three centuries. The people use their own language for all purposes among themselves, but they use the Kei language when communicating with people from the surrounding Kei-speaking villages.

Dialect Information

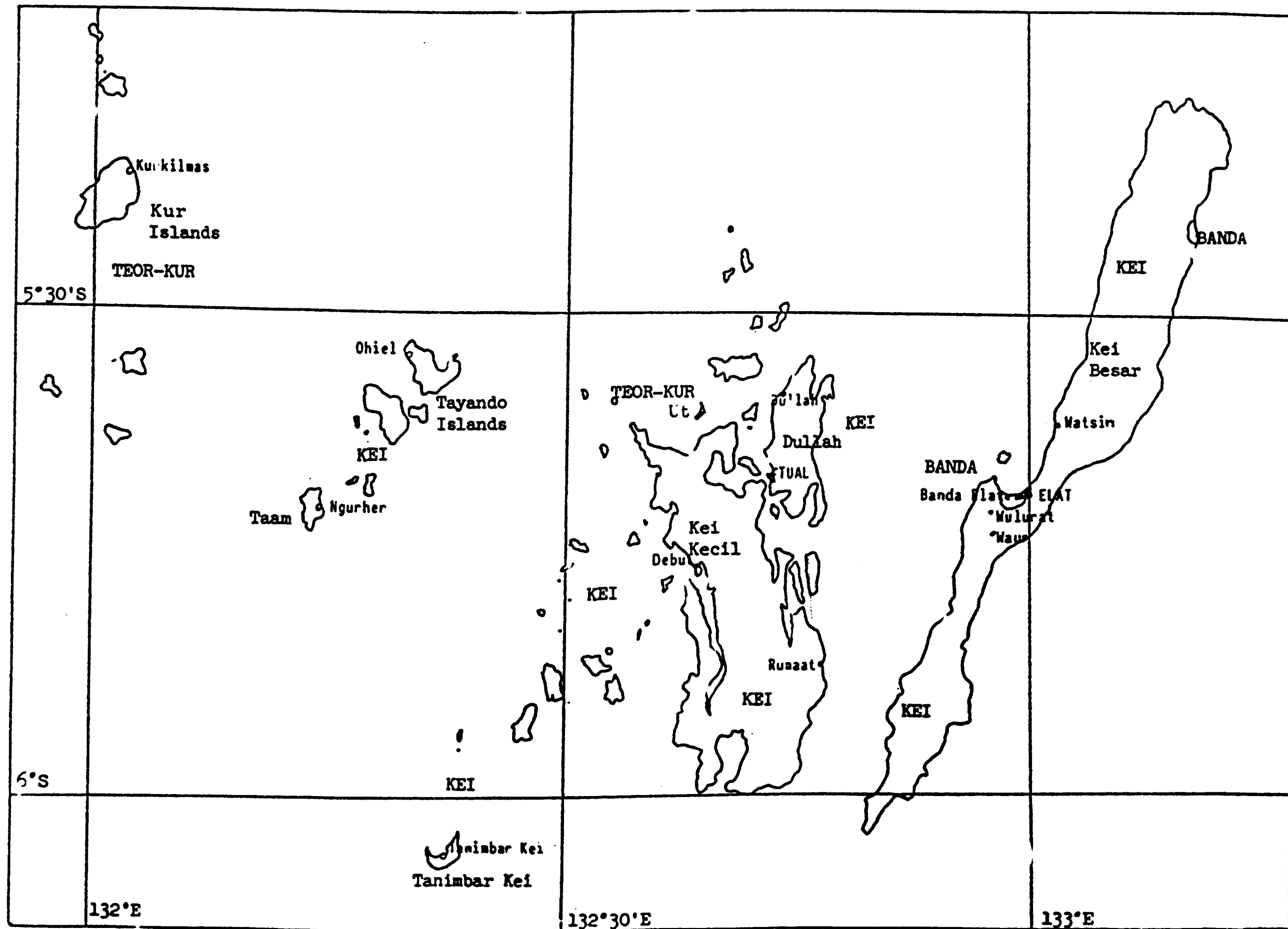
Only one Banda wordlist was taken, and that was from Elat, so it was not possible to compare dialects. However our informants reported differences between Elat and Eli.

2.1.2.2 Teor-Kur

Teor-Kur is spoken on Kur Island and nearby islands, in the far west of the Subdistrict of Kei Kecil, and also on the island of Teor, to the north of Kur, in the Subdistrict of Seram Timur. There are probably 2,000 to 3,000 speakers in several villages in the Kur Islands, and a further thousand on Teor, again in a number of villages. There is also a small community of about a hundred people originally from Teor who now live on the small island of Ut just north of Kei Kecil. It is not known how long they have been there. The people of Ut all speak Kei as well as their own language. Also, in 1979, a small village of Teor folk was established at Sitnohoi in north-west Kei Kecil. Our wordlists were taken from the main island of Kur and from Ut Island, where the dialect is reportedly the same as Teor. In Teor and Ut the language is called Teor by its speakers, whereas in Kur it is called Kur, but there is reportedly complete mutual intelligibility, so I have used the name Teor-Kur for the language. Both wordlists were taken in Tual, and so we were unable to observe it being used in the villages. However, our informants reported vigorous use of the language in both places and that the Kur community in Tual also use the language among themselves.

Dialect Information

The percentage of lexical similarity between Ut and Kur is 83%, which shows them to be separate dialects. Data are needed from Teor Island to check whether Ut and Teor are indeed one dialect. There are some consistent phonetic changes between Kur and Ut, particularly the palatalisation in Ut of word-final alveolar nasals, laterals and voiceless alveolar plosives in Kur. Also, velar nasals in Kur become voiced velar plosives in Ut. Dialect intelligibility testing is needed to establish the precise extent of mutual intelligibility.



MAP 2: The Kei Islands

Key:
 LANGUAGE NAME
 • Village visited
 ◦ Village surveyed but not visited
 Island Name

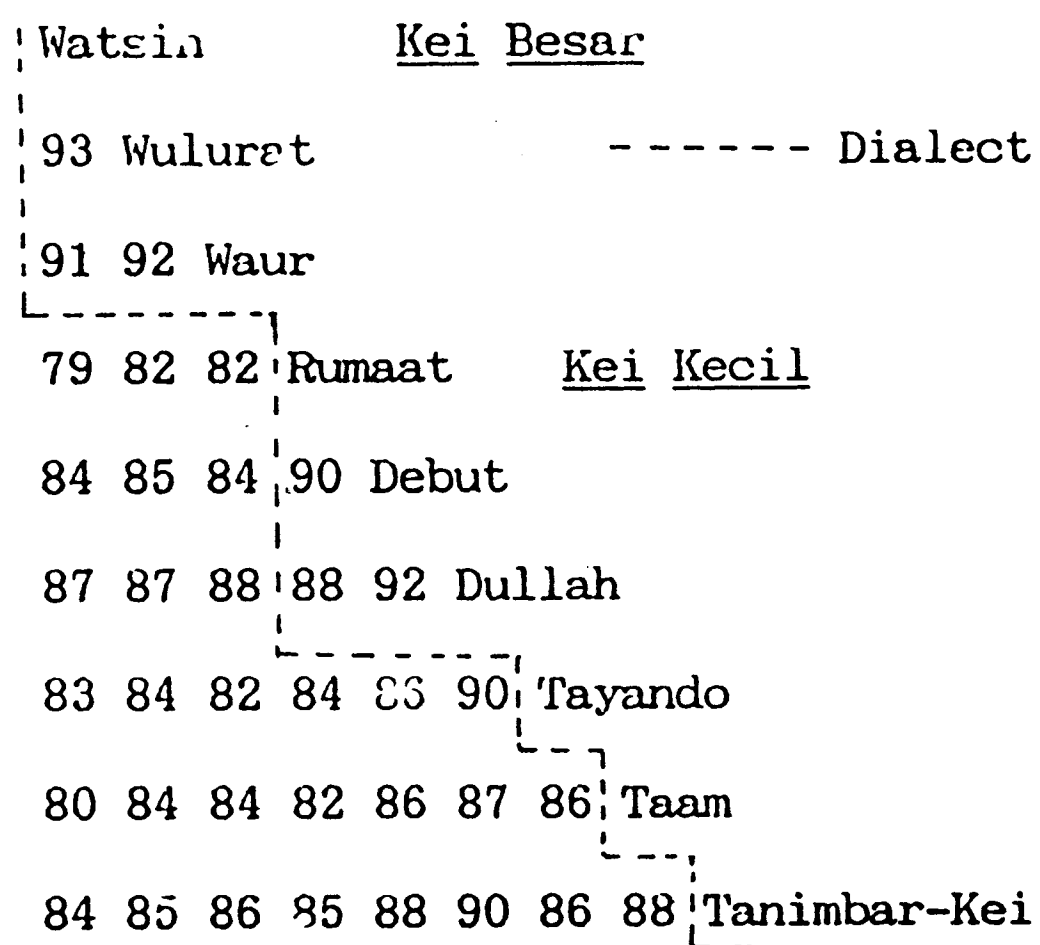
2.1.2.3 Kei

Kei (sometimes spelt Kai), or Evav, as it is known in the vernacular, is spoken as a mother tongue by about 86,000 people in over 200 villages throughout Kei Kecil, Kei Besar and the surrounding islands except, in the two villages of Banda Eli and Banda Elat on Kei Besar, the village of Ut on the island of that name, and in the Kur islands in the extreme west of the Subdistrict of Kei Kecil. In these latter areas it is used as a lingua franca when speaking with people from other parts of Kei.

Kei is used in everyday life, not only in the villages, but often also in the towns of Tual and Elat. We even observed it being used as a language of conversation in government offices. Chinese merchants and other non-natives have also learnt the language.

Dialect Information

There are two large dialect groupings, Kei Kecil and Kei Besar. The northern part of Kei Besar has some prominent phonetic differences from the south, but the three wordlists taken on the island grouped together in one dialect (see matrix 7 below). These three wordlists were all from the central part of the island (see map 2), although one, Watsin, had the phonetic features typical of the northern part. It is likely that dialect differences would be found between the northern and southern extremities of Kei Besar. There are also three small dialect groupings, Tayando, Taam and Tanimbar-Kei, on the islands to the west of Kei Kecil.



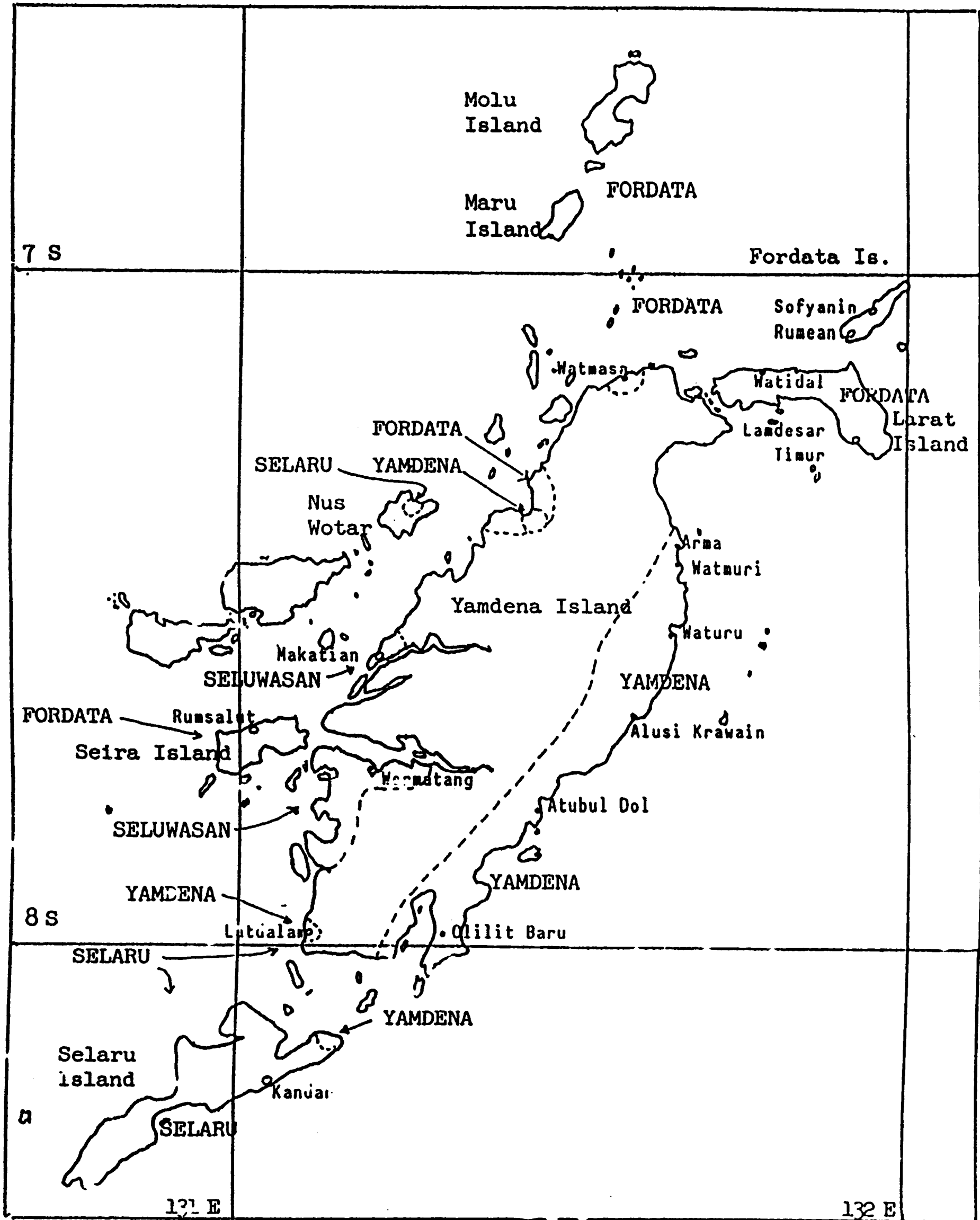
MATRIX 7: The Kei Language

2.1.2.4 Fordata

Fordata (sometimes called Larat) is spoken by about 19,000 people in northern Tanimbar on the islands of Fordata, Larat, the Molu-Maru group, a few villages on the north-west coast of Yamdena, and on Seira, off the west coast of Yamdena. There are thirty Fordata-speaking villages. The language is in constant use in the villages, and we often heard it used in the town of Larat by native speakers (although not by non-natives as in Kei).

Dialect Information

There are two main dialects, the Northern Dialect, spoken in Larat,



MAP 3 : The Tanimbar Islands

Key:

- LANGUAGE NAME
- Village visited
- Village surveyed but not visited
- Island Name

Fordata, the Molu-Maru Group and in the villages on north-west Yamdena, and secondly the Seira Dialect, spoken on the island of that name off the west coast of Yamdena. Susan McKinnon (1983:20) suggests four dialects, dividing the Northern dialect into Larat-Fordata I, Larat-Fordata II, and Molu-Maru. This analysis appears to be based on some consistent phonetic changes between these three groups, which is clearly legitimate. However, when we are concerned with intelligibility these changes may not be very significant. In this survey wordlists were taken from each of McKinnon's areas⁶ and three of her groups relate to each other above 90%, that is, above the boundary for a common dialect, leaving only two dialect groupings according to my analysis. It is recognised however that lexicostatistics is by no means the only means of differentiating dialects, and that McKinnon speaks Fordata, so this present author does not claim to have definitively fixed the dialect boundaries. Having said that, on the basis of this survey I am positing two main dialects, the reasoning for which will be seen from matrix 8 below.

```

| Sofyanin          - - - - - Dialect
|
| 95 Watidal
|
| 94 94 Rumean
|
| 96 95 96 Lamdesar Timur
|
| 92 92 93 94 Watmasa
|-----|
| 88 89 90 88 88 | Seira
|-----|

```

MATRIX 8: The Fordata Language

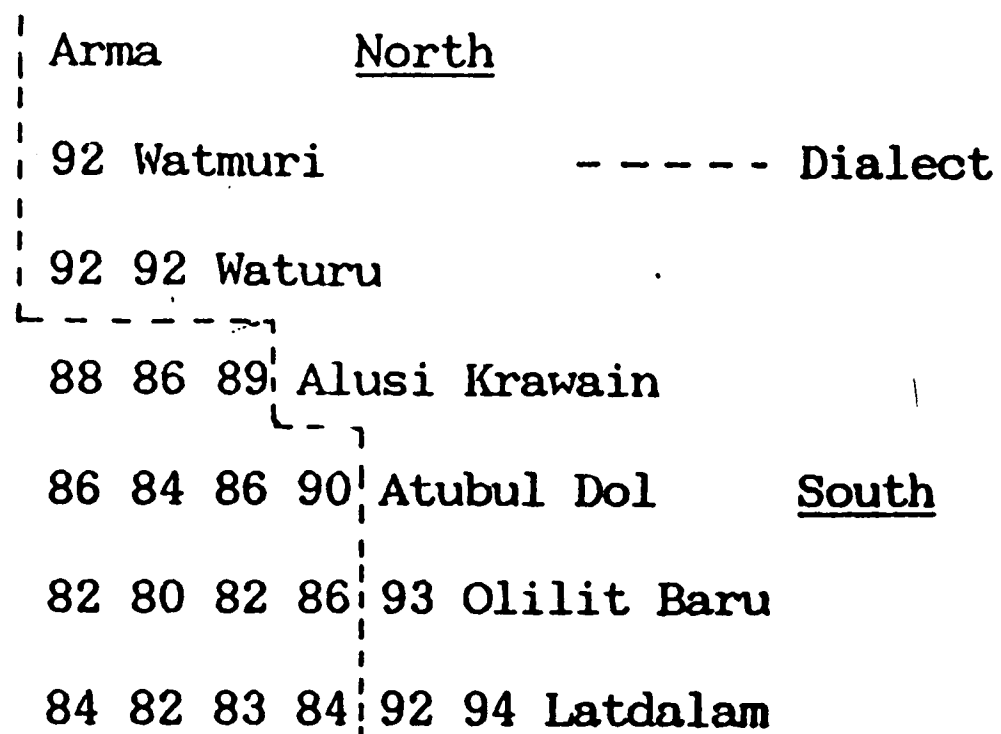
2.1.2.5 Yamdena

Yamdena is spoken all along the east coast of Yamdena Island, the largest island in Tanimbar, in the village of Adaut on Selaru Island, and as one of two languages spoken in the village of Latdalam, in south-west Yamdena. It is spoken by over 30,000 people in 35 villages altogether. The language is in widespread daily use by all age-groups in most of the area, although in some of the villages in the central part of the east coast we noticed some of the younger children were using Malay. When we asked why, the people said that some parents speak to their pre-school children in Ambonese Malay, so that they will be ready for Indonesian when they start school. However, they said they will all learn and use the Yamdena language as they get older. This situation was not found in all the villages we visited. Further south we noticed that young children seemed to be using the Yamdena Language.

Dialect Information

There is dialect chaining from the north to the south with considerable variation between the extremes. The most northerly and the most southerly villages relate to each other with a lexical similarity percentage in the low eighties reaching right down to 80% between Olilit Baru and Watmuri, that is, approaching the limit of intelligibility. Intelligibility testing might be needed to establish whether or not these extreme dialects are mutually intelligible. It might prove difficult to establish the extent of intelligibility, as the reported information is that speakers of the northern dialect can understand the southern one, but that the

southerners cannot understand the northern dialect. This is probably due to bilingualism by the northerners, since it seems that the southern dialect is perceived to be the prestige dialect. Testers of intelligibility will have to be careful to distinguish between bilingualism and mutual intelligibility of dialects. As will be seen from matrix 9 below, there is a fairly homogeneous southern dialect and a fairly homogeneous northern dialect, with Alusi Krawain in these data representing the bridge between the two.



MATRIX 9: The Yamdena Language

2.1.2.6 Seluwasan

Seluwasan is spoken by about 2,100 people in four villages on the west coast of Yamdena Island. There used to be just three villages, but one of them, Otemer, split into Batu Putih and Marantutul. The wordlists in this survey were taken from the other two villages, Wermatang and Makatian. Makatian is a fairly divergent dialect, and is known all over Tanimbar as being a 'difficult' language. The people are renowned for their skill in hunting. We took our wordlists from groups of people who had just arrived in Saumlaki to bring children in to school, so did not observe the language being used in the villages. However the people we met were using the language freely among themselves, and reported a positive attitude to it.

Dialect Information

There are two dialects. The main dialect is spoken in Wermatang, Batu Putih, and Marantutul, while Makatian forms the other dialect. The latter is fairly divergent from its neighbours. The data show Wermatang and Makatian to be 86% lexically similar, although sometimes there are considerable phonetic differences between cognate words. Most inhabitants of Tanimbar consider Makatian to be a separate language. The Makatian informant said he could understand people from Wermatang, but the Wermatang informant claimed he couldn't understand Makatian. When we told this to the Makatian informant, he indicated that he did not believe the man was telling the truth! Intelligibility testing is needed between Makatian and the other villages.

2.1.2.7 Selaru

Selaru is spoken in six of the seven villages on Selaru Island, the most southerly island of Tanimbar. It is also spoken as one of two languages in the village of Latdalam, in south-west Yamdena, and in the

village of Lingada on Nus-Wotar Island off the west coast of Yamdena. It may also be spoken on some other islands off the north-west of Yamdena. The people of Lingada moved there from Selaru in recent years. The total number of speakers is approximately 8,000. We did not get to any of the villages to observe the language in use, since we took the wordlists from two different groups of people who had just arrived in Saumlaki from their villages, Kandar and Latdalam. They were vigorous in using the language among themselves and also reported that the language was in regular daily use.

Dialect Information

Insufficient data were gathered on this survey to establish what dialects there are in the Selaru Language. Further investigation will be necessary. We were only able to take two wordlists, from Kandar and Latdalam, which were fairly closely related (91%), but the Kandar people reported variations from village to village.

2.2 The Languages of Aru

2.2.1 Relationship with Languages outside Aru

The twelve languages of Aru identified on this survey are all members of a common family. However, although they are fairly closely related to each other, when compared with the languages of Kei and Tanimbar the cognate percentages, that is to say the percentages of lexically similar words, are very low indeed. They range from 24% to 26%, or 26% to 34% in the reduced matrix (see matrices 1 and 2). These figures show that the Aru languages share a superstock with the languages of Kei and Tanimbar, that is they are members of the Central Malayo-Polynesian Superstock. They also show, however, that the Aru languages do not share a stock with any of the languages of Kei and Tanimbar. This low percentage of shared lexical similarity with the languages of Kei and Tanimbar raises this question: Does the Aru Family constitute a stock-level isolate grouping, or does it share a stock with some other families or languages outside of Aru that have not been covered by this survey? The only nearby Austronesian languages outside the survey area are those spoken on the south side of the Bird's Neck of Irian Jaya. So, to answer this question, I requested and obtained data from the Cenderawasih University and Summer Institute of Linguistics Co-operative Programme in Jayapura. The data obtained were from three Austronesian languages spoken in the coastal areas of the Bomberai Peninsula and near Kaimana on the south side of Irian Jaya. The languages are Irarutu, Uruangnirin, and Koiwai. These data were compared with all the data from Aru, Kei and Tanimbar. Irarutu relates to all the languages of the South-East Moluccas ranging from the mid teens to the mid twenties, which suggests that it is a member of a different superstock, in this case the Eastern Malayo-Polynesian Superstock (see figure 1). Uruangnirin and Koiwai, however, appear to be members of the Central Malayo-Polynesian Superstock, as they relate at superstock level to the languages of Aru, Kei and Tanimbar. What, for our purposes, is most noteworthy is that their relationship to the languages of Aru is very low, ranging from just below superstock level (25%) to about 30%, whereas their relationship with the languages of Kei and Tanimbar is considerably higher, ranging from about 30 to 40%. We may also note that, despite the distant relationship (an average of 30% lexical similarity), the languages of Kei and Tanimbar are in fact linguistically the closest neighbours of the Aru languages. The very highest percentage recorded in the data between a language of Aru and one outside Aru is 36%. This confirms the position of the Aru Family as a stock-level isolate grouping within the Central Malayo-Polynesian Superstock.

It should be remembered that in this lexicostatistic comparison borrowed words have not been eliminated, since we are primarily concerned with intelligibility and synchronic relationships, and not historical or diachronic relationships (see §section 1.1). James Collins (1983c:128-9) points out that there are borrowings in the Aru languages from the non-Austronesian languages spoken on the south coast of Irian Jaya, and he gives some examples. However this is likely to be a relatively small percentage of the lexicon, and it is clear that the Aru languages are members of the Austronesian Phylum. Indeed, Collins points out that many of the lexical similarities between the Aru and Irian Jaya languages are due to the Irian Jaya languages borrowing Austronesian forms.

2.2.2 Relationships within the Aru Family

The languages of Aru are all members of the same family, that is, lexicostatistically they relate to each other with a percentage of shared lexical similarity above 60%. Within that family, however, there has been much divergence resulting in twelve separate languages as shown in matrix 10 (see also map 4). This result was arrived at from analysis of the data from our survey of the Aru languages. During this survey data were collected from at least one village in each of the speech communities I have identified as a separate language (see appendix B for a list of villages where data were taken).

These results are somewhat different from the conclusions of Collins (1983c:13), which are those presented in Wurm and Hattori's Linguistic Atlas of the Pacific (Pacific Linguistics C-66, 1981). Collins (page 131) states that there are five languages in Aru, but that 'variation in the larger dialect chains (especially Kola and Wokam-Trangan) is considerable.' He does not define his use of the term language; it seems he does not use the term to mean a speech community of mutually intelligible dialects. Collins distinguishes languages on the basis of phonetic innovation between cognate words, even if the proportion of cognate words between the speech communities in question is relatively low. Therefore, if the phonetic innovation is not great between two speech communities he would consider them to be one language, even if a relatively large percentage of words in each speech community is not cognate with the other. However, I would consider the same two speech communities to be separate languages because the percentage of non-cognate words is high enough to be able to say that they are not mutually intelligible.

The language (or dialect chain) which Collins calls Wokam-Trangan I have designated as five separate languages. It is not simply that I have chosen different labels for the taxonomic levels, there are some fundamental differences in the classification. For example, Collins designates Karey a dialect of Wokam-Trangan whilst he calls Barakai a separate language. My data, however, show that Karey is much more closely related to Barakai than it is to the speech communities designated as Wokam-Trangan by Collins. This is illustrated below in matrix 11 which shows the relationship between languages, as I have designated them, with Collins' designations in parentheses.

Batuley (W-T)
 66 Wokam (W-T)
 60 69 W. Trangan (W-T)
 64 66 71 E. Trangan (W-T)
 72 68 62 72 Karey (W-T)
 73 62 60 68 76 Barakai (Barakai)

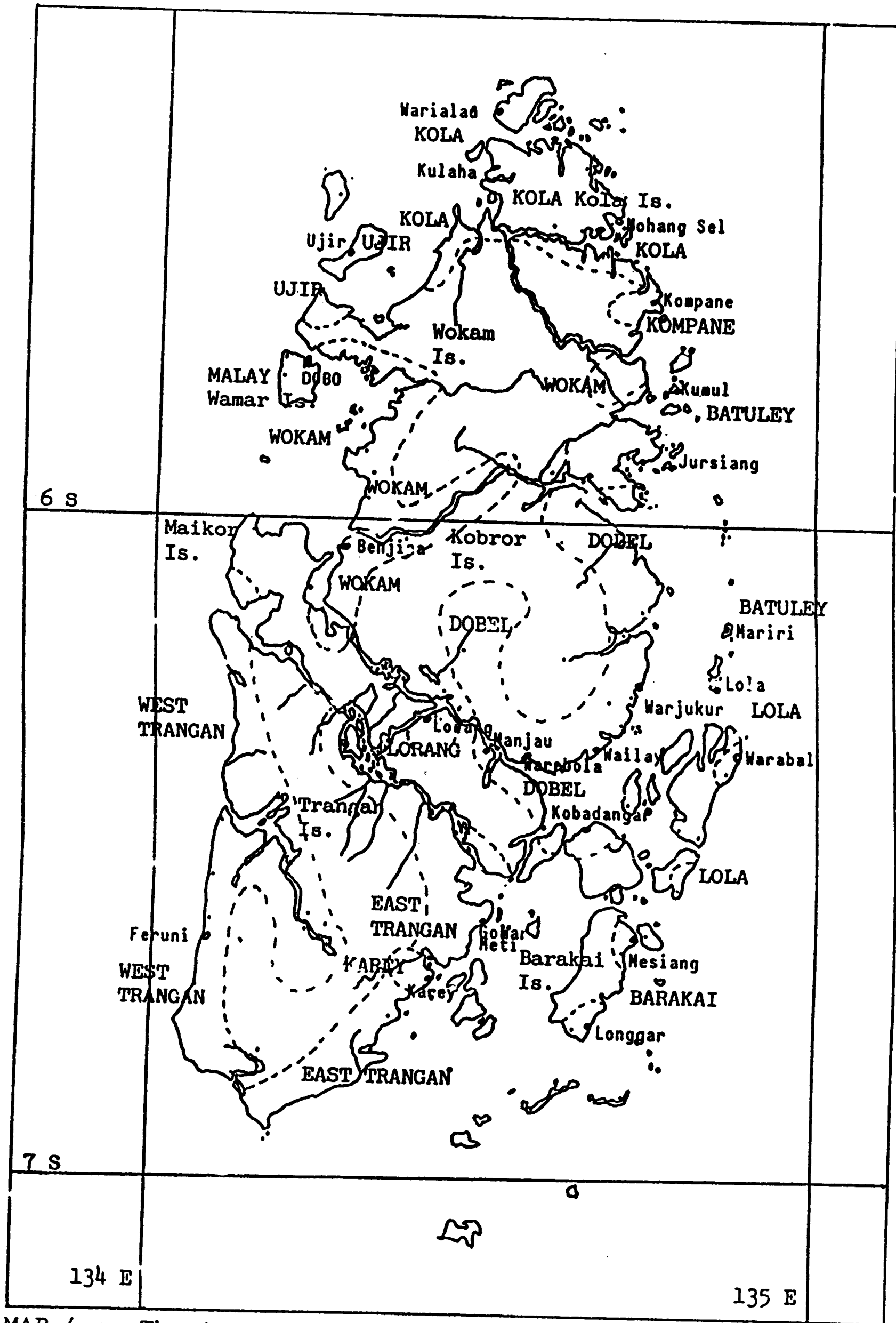
MATRIX 11 (Reduced): Comparison with Collins' Classification

Ujir																					
72	Wari-lau																				
70	86	Kulaha																			
68	90	80	Mohang Sel																		
70	77	76	79	Kompane																	
64	68	66	66	76	Lola																
64	69	67	67	75	90	Warabal															
61	68	66	66	74	76	78	Warjukur														
59	65	64	64	70	76	76	90	Wailay													
62	66	66	65	70	74	76	84	84	Manjau												
64	68	68	67	73	74	76	87	83	92	Warabola											
64	68	66	64	70	76	78	80	78	84	86	Kobadangar										
63	66	67	64	70	70	73	76	75	78	79	80	Lorang									
68	68	68	68	72	67	67	70	68	70	71	70	79	Benjina								
60	60	58	60	62	60	59	60	58	60	60	61	66	69	Feruni							
56	63	62	61	64	64	66	68	66	68	70	70	68	66	71	Gomar Meti						
60	62	61	61	66	70	72	67	66	68	70	72	67	68	62	72	Karey					
57	59	60	59	64	64	66	66	67	68	66	68	62	59	60	66	76	Longgar				
60	64	64	62	68	70	70	70	69	69	72	72	68	64	61	70	76	87	Mosiang			
62	64	63	64	70	72	70	74	68	67	71	70	66	64	62	66	73	74	76	Mariri		
69	70	69	70	78	77	76	78	74	76	76	76	72	70	60	66	74	73	76	82	Jursiang	
63	66	64	64	74	73	70	73	70	72	70	72	65	64	58	60	68	70	70	78	90	Kumul

Key:

- Dialect Boundary
- ===== Language Boundary
- ===== Subfamily Boundary
- ===== Family Boundary

MATRIX 10 : The Aru Family



MAP 4 : The Aru Islands

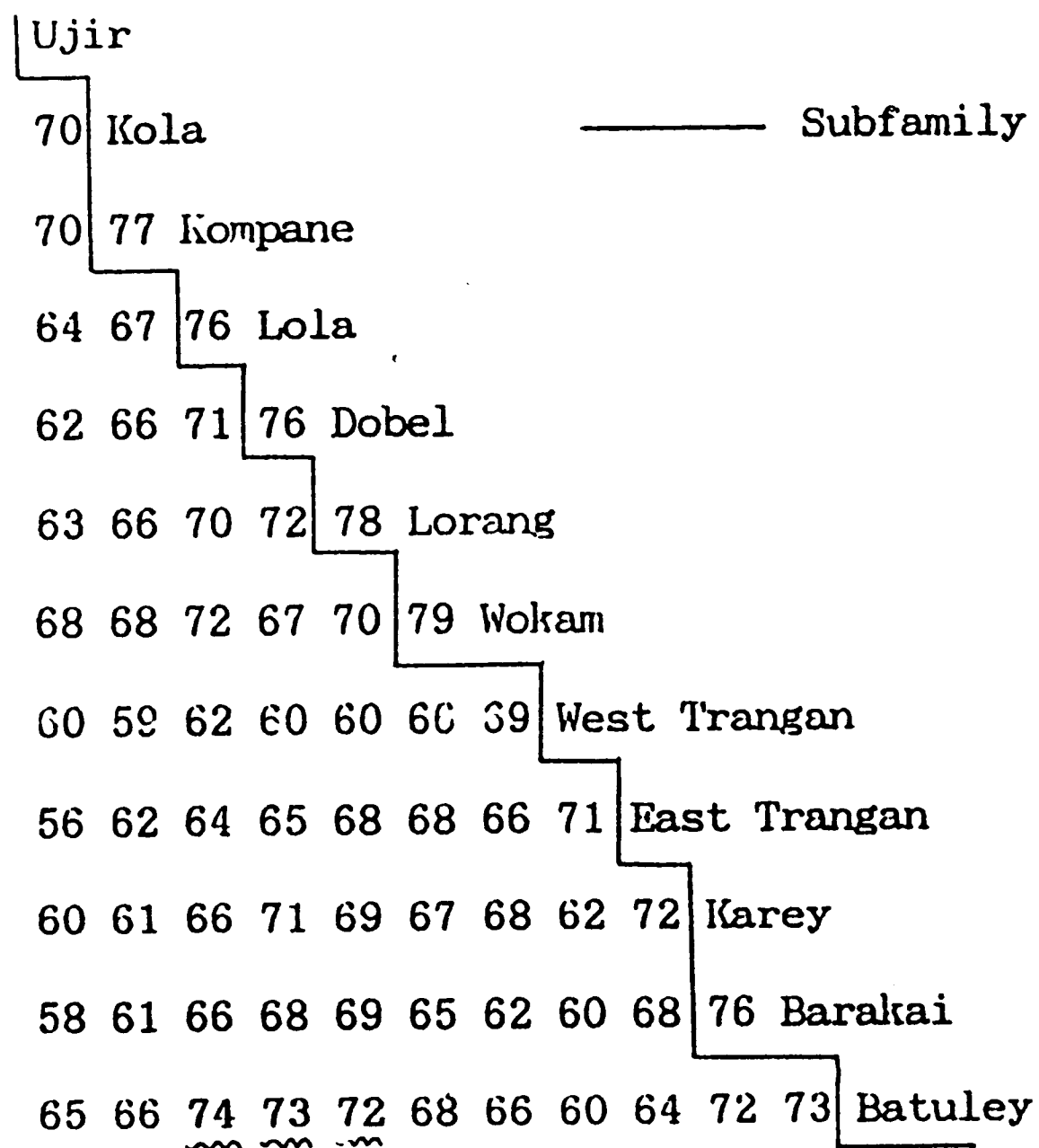
Key:

- LANGUAGE NAME
- Village visited
- Village surveyed but not visited
- Island Name

According to the results of the lexicostatistic analysis, all of these speech communities are different languages, but in matrix 11 the two most closely related villages are Karey and Barakai, which share a subfamily. However, Collins calls Barakai a separate language and the others dialects of a common language. In fact, the lexicostatistic results fit in better with the reported information from the speakers of the languages. For example in Karey, which is situated on Trangan Island alongside villages that speak East Trangan, the people said that their language was very different from the East Trangan language, and that the people in the surrounding villages could not understand them, however they reported limited understanding of the more distant Barakai language. Similarly, Lola, which Collins designates a dialect of Kola, relates to the Kola villages in the north at an average of 67.5% (admittedly it is more closely related to Kompane), whereas it relates to Dobel (which Collins designates a completely different branch of Proto-Aru) at 76%, a subfamily-level relationship.

2.2.2.1 Subgrouping the Aru Languages

The full matrix for Aru showing all the villages surveyed (matrix 10) has been reduced below in matrix 12 by averaging the figures between all the villages for the two languages being compared. The resultant matrix more clearly shows the complexity of the relationships between the languages of the Aru Family.



MATRIX 12 (Reduced): The Aru Family

These relationships can be shown diagrammatically with a tree diagram (see figure 3), in which the complexity of these relationships can be seen. The tree diagram shows that a linguistic chain, with varying degrees of closeness, can be traced through all the languages of the archipelago. (Please use the tree diagram in conjunction with map 4). The linguistic chain can be traced round the islands as follows: it begins at Ujir and goes clockwise round the northern part of the archipelago (Ujir, Kola, Kompane). It then bypasses Batuley, continues down the east coast and cuts

across to the west coast via Lorang (Lola, Dobel, Lorang, Wokam). From Wokam it goes anticlockwise round the southern part of the islands (West Trangan, East Trangan, Karey), and then from Karey it heads northwards via the eastern islands through Barakai to Batuley. This is not a language chain (nor yet a subfamily chain) since there is no overlap at language level (see discussion below), however each language does relate to its neighbour in the chain at around 70% or above. At one point there is chaining at subfamily level, which will now be discussed more fully.

In the reduced language matrix of Aru (matrix 12) and in the tree diagram (figure 3) it will be seen that there is chaining at subfamily level, extending from Kola through Kompane, Lola, Dobel, and Lorang to Wokam, each language relating at subfamily level (75-80%) to the neighbouring language, but at a much lower level to the others in the chain. Each of these languages, however, does relate to all the others at above 70%, with the exception of Kola which relates down in the mid-to-high 60s with all but Kompane. Kola, in fact, groups more readily with Ujir (see below) than it does with these central Aru languages. From this the language of Kompane can be seen to be a bridge between Ujir and Kola to the north, on the one hand, and the central Aru grouping, on the other. This is clearly shown on the tree diagram.

At this point it would be good to discuss the position of the Batuley language. Although it is geographically between the languages of Kompane to the north, and Lola and Dobel to the south (see map 4), it does not fit into the subfamily chain mentioned above. However, considering the geographical distance between them, it relates relatively closely to the Barakai-Karey subfamily in the south-east, and that seems to be where it best fits into the overall relationships between the Aru languages. Batuley actually has a closer relationship to Barakai than the latter has to the nearby East Trangan language (represented in these data by the village of Gomar Meti). It will be noted though, that while Batuley does not relate as closely to Kompane, Lola, and Dobel (its geographical neighbours) as they do to each other, it does relate more closely to them (low 70s) than to other Aru languages, except for the aforementioned Barakai-Karey subfamily. This relatively close relationship with its geographical neighbours is probably due to sporadic convergence,⁸ brought about by the geographical proximity of the languages. The sporadic convergence is illustrated in the tree diagram (figure 3) by the broken line joining Batuley and the Kompane-Lola-Dobel grouping, and in the reduced matrix (matrix 12) by the high figures in bold print. I have put Batuley adjacent to Barakai in the matrix and tree diagram (representing linguistic relationship resulting from divergence) since this seems to be a satisfactory explanation when the relatively close relationship with its neighbours is explained by sporadic convergence due to geographical proximity. If, however, I had put Batuley in the matrix and tree diagram next to its neighbours, a procedure I tried at first, there would be no explanation for its relatively close relationship with the geographically distant Barakai. The present analysis, therefore, seems to be the best.

The languages in the southern part of Aru, West Trangan, East Trangan, Karey, and Barakai, are less closely related to each other than are the languages of the north central grouping described above. The exception is the close relationship of Karey and Barakai which relate to each other at 76% lexical similarity, and thus form a subfamily. East Trangan and West Trangan only relate to each other at 71% (although data were only collected from one village for each), while Karey displays convergence with East Trangan due to geographical proximity, but still relates only at 72%. Karey is in fact a solitary village within a string of East Trangan speaking villages. The relationship of the Barakai language to East and West Trangan is down in the 60s.

Finally we return to the north, where, as I have mentioned, Ujir groups with Kola and Kompane, but only at the fairly low figure of 70%. This is the closest Ujir relates to any other language, making it the most divergent of the Aru languages. Between Ujir and Wokam there is 68% shared lexical similarity, an unexpectedly high figure, probably due to the geographical proximity of the two languages. Ujir has more distant relationships with all the other members of the Aru Family.

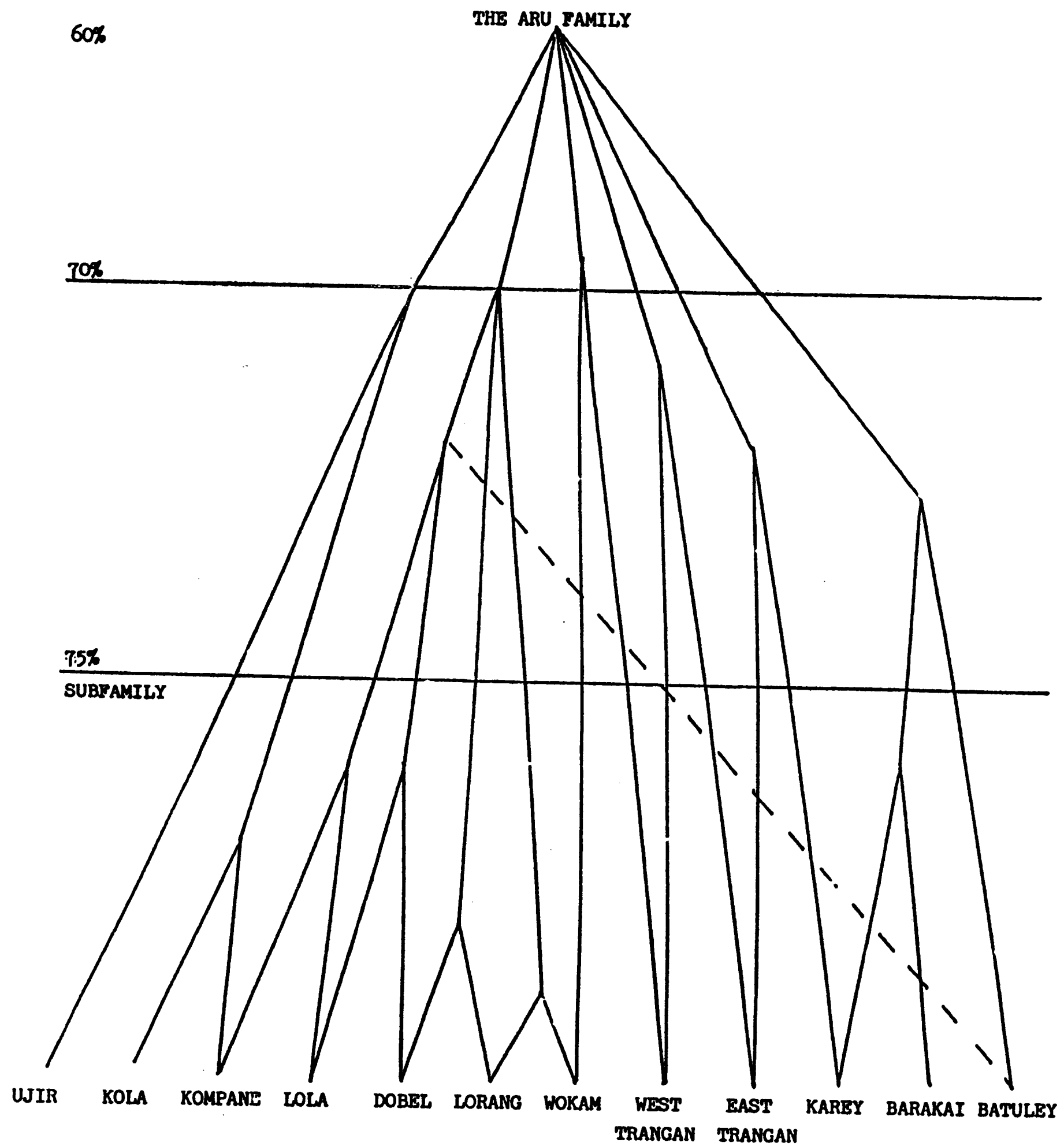


FIGURE 3 : The Languages of Aru:
Interrelationships.

Key:
 ——— Significant along its length
 - - - Significant only at its ends

2.2.2.2 Establishing Language Boundaries

Language boundaries were established, as discussed in the introduction, with the intention of demarcating mutually intelligible speech groups. It is usually found that speech communities with a percentage of lexical similarity of over 80% have mutual intelligibility, and so this figure has been set as the language boundary. Figures around the 80% mark indicate that intelligibility testing is needed. It will be seen from the full matrix of Aru (matrix 10) that the language boundaries are fairly distinct, except in two places. The first of these is the village of Lorang, which relates to the Koba Dialect of Dobel at 80% and to the village of Benjina (the only representative of the Wokam language in the data) at 79%, both at the very boundary of expected intelligibility. Lorang is designated as a separate language because, (1) it relates lower than 80% to the other Dobel villages surveyed (i.e., its average percentage of lexical similarity with the Dobel language is below 80%), (2) geographically Benjina is the second closest Wokam-speaking village to Lorang (and therefore the other Wokam villages are likely to be less closely related to Lorang than Benjina is) and (3) the people of Lorang say that their language is unique to their village, although they do admit that it has similarities with the speech of Kobadangar. Although Dobel and Wokam are not too closely related to each other (70%), Lorang is the bridge between them. The second place where the language boundary is not too distinct is at the village of Kompane. Kompane relates to the village of Mohang Sel, one of the closest Kola speaking villages, at 79%, however it has a lower relationship with the more distant Kola villages. It should be noted, however, that Mohang Sel itself only relates at 80% with the two other Kola villages surveyed. Further investigation, including intelligibility testing, is needed to check whether or not there is more than one language within what I am at present calling Kola. A third place of possible doubt as to the language boundaries is whether Mariri is indeed a dialect of Batuley or a separate language (this will be discussed below under Batuley).

2.2.3 Language Information for the Aru Archipelago

In this section each of the twelve languages spoken in the Aru Islands is separately considered. Information is given about the people, where they live, and how much they use the language. I also give all the information I can on dialects of each language, indicating some areas where further research is needed.

2.2.3.1 Ujir

Ujir is only spoken in two villages in the north-west of the Aru Islands, Ujir village on Ujir Island and Samang on the end of the large peninsula in the west of Wokam Island. In Ujir we heard the language being used in everyday life, and the village leader (Orang Kaya) told us it was still strong. He also told us that some time ago he instructed the parents of the village to use Ujir, not Malay, when speaking to their children, and apparently they do. It is interesting that he had to do that, though. We were not able to go to Samang but we were told that the language is falling into disuse there because of a large influx of non-Aruese people. The population of Ujir is 500 and that of Samang is 341, but as some of the inhabitants of Samang are not speakers of the language, I can only estimate the number of speakers of Ujir at between 700 and 800.

Dialect Information

It is not known if there is a dialect difference between Ujir and Samang.

2.2.3.2 Kola

Kola is spoken by nearly 6,000 people in twenty-two villages in the very north of Aru, on Kola and adjacent islands. The language appears to be in regular everyday use and to be spoken by everybody. In Warialau village we were told that the non-natives who live there want to learn the local language.

Dialect Information

In the three wordlists we took there was evidence of considerable dialect variation as shown in matrix 13:

```
Warialau    - - - - -Dialect
└───┘
 86└───┘ Kulaha
    └───┘
 80 80└───┘ Mohang Sel
```

MATRIX 13: The Kola Language

The matrix shows each of the three villages surveyed to speak a different dialect of the language. Warialau and Kulaha which are both in the north and west of the language area are more closely related but are still different dialects. However Mohang Sel, in the south-east of Kola island is seen to be only just on the boundary of intelligibility with the other two, relating to them at 80%. As there are a further 19 Kola villages which we did not survey, much more investigation is needed before we can have a clear picture of the dialect situation in the Kola Language. It will be noted however that Mohang Sel is geographically distant from both Warialau and Kulaha; in fact, they are at opposite extremities of the language area, so it is unlikely that there is more than one language there.

2.2.3.3 Kompane

Kompane is spoken in north-east Aru by 254 people in Kompane village on the east coast of Kongan Island, just south of Kola Island and north of Wckam Island. The language is in vigorous daily use. As it is spoken in only one village there are no dialect variations. See above for a discussion of Kompane's close relationship with neighbouring languages.

2.2.3.4 Lola

Lola is spoken by about 470 people in the three villages of Lola, Warabal, and Jambuair, which are on three separate islands to the east of Kobror and Baun Islands. In Lola we were told that this was where the language was strongest. We did observe the language in use but we also noticed a group of teenagers speaking with each other in Malay. They also said their language was not used as much as the local language is on the

big island (Kobror). It was reported to us that in Jambuair some of the young people no longer use the language.

Dialect Information

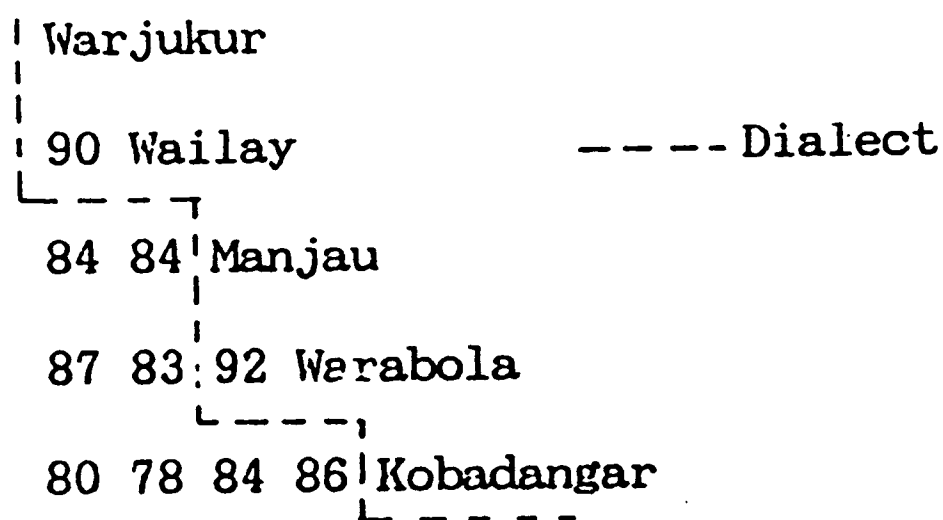
Lola and Warabal show 90% of shared lexical similarity, indicating that any dialect variation between them would be slight. No data were obtained from Jambuair.

2.2.3.5 Dobel

Dobel (sometimes known as Kobror) is spoken by a total of nearly 5,000 people along the whole east coast of Kobror Island, in three interior villages of that island, in one village in the very south-east of Wokam Island, in a number of villages at the eastern end of the Barakai Strait both on Kobror and Koba Islands, and in the three Koba villages on Baun and Fukarel Islands. The total number of villages that speak Dobel is twenty. The language is in constant use and we did not notice the use of Malay except in communicating with us. Many pre-school children do not know Indonesian.

Dialect Information

There is some dialect variation throughout the language area. The data show three dialects, but these data do not cover the whole Dobel area. The three Koba villages are considered by the people to speak a separate language. It will be noted that Kobadangar relates to the nearby Manjau and Warabola villages well within the language boundary, but the percentages for Wailay and Warjukur are on the boundary of intelligibility. As many of the unsurveyed Dobel villages are considerably further away from the Koba area than Wailay and Warjukur, it is possible that their relationship with Koba would be even lower. If this were the case we would have a case of dialect chaining with the extremities of the chain not mutually intelligible. More data are needed to establish dialect patterns. So far no data have been gathered from the north and interior of the language area. The author expects to do further research within the Dobel language. Matrix 14 below shows the relationship of the villages already surveyed in the Dobel area.



MATRIX 14: The Dobel Language

2.2.3.6 Lorang

Lorang is spoken by just over 200 people in the village of Lorang, right in the centre of Aru, half-way down the Barakai Strait, on Koba Island. This is the smallest language community in Aru, indeed in the whole Aru-Kei-Tanimbar area. Lorang people are locally known for being

able to speak a number of the local languages. As the Lorang language is only spoken in one village there is no dialect variation (however see above for close relationships with neighbouring languages).

2.2.3.7 Wokam

Wokam (also known as the Manombai Language) is spoken by about 5,700 speakers along the west coast of Wokam Island from Wokam village (very close to the town of Dobo) southwards, in the villages along both sides of the Manombai Strait as far as Wakua (that is on Wokam and Kobror Islands), and in Benjina (Kobror Island) and Gardakau (Maikor Island) at the western end of the Barakai Strait. It is also reportedly spoken in the small village of Kobamar on the east coast of Wokam Island. It was spoken in the villages of Wangel and Durjela on Wamar Island, but due to a large influx of people from outside Aru, and also possibly due to their proximity to the town of Dobo with which they share an island, the language is no longer spoken there except by a very few old people. We took a wordlist in Wangel but rejected it as unreliable, because our informant, the only person we could find who claimed to speak the language, was an elderly man living about half a kilometre outside the village who was no longer using the language and could not remember many words. As in Durjela and Wangel there were reports of many non-natives living in Wokam village, but I have no information as to what extent the language is still spoken there. Excluding Durjela and Wangel, the Wokam Language is spoken in 21 villages. In all of them, with the possible exception of Wokam village, language usage is reportedly still strong. In Benjina we heard the language in everyday use.

Dialect Information

The only reliable wordlist we obtained from the Wokam Language was from the village of Benjina, so we were not able to compare dialects. Dialect survey will need to be conducted to establish the dialects of this language.

2.2.3.8 West Trangan

West Trangan, the largest language in the Aru Islands, is spoken all down the west coast of the large island of Trangan in the southern part of Aru. There are over 7,000 speakers of the language living in 20 villages. We were unable to visit a West Trangan village, but managed to take a wordlist in the town of Dobo. The language is reportedly strong and is the language of day-to-day communication. It is sometimes reported to be in use as a lingua franca in Aru (see 2.3.4 below).

Dialect Information

This needs further research; there are reports of dialect variation but we have no data for comparison as yet. In fact extensive dialect survey is needed throughout Trangan Island.

2.2.3.9 East Trangan

East Trangan is spoken by nearly 3,300 people in thirteen villages right along the east coast of Trangan Island and in the villages in the Maikor Strait. The language was in constant use in the two villages we visited.

Dialect Information

We visited two East Trangan speaking villages but were only able to take a wordlist in one of them, Gomar Meti, and so although we heard reports of considerable variation, we were unable to confirm this. Dialect survey is therefore still needed in this language.

2.2.3.10 Karey

Karey is spoken by 500 people in the village of Karey which is situated on the east coast of Trangan Island. The language is in strong daily use. As there is only one village there is no dialect variation. It is most closely related to the Barakai Language (see above).

2.2.3.11 Barakai

Barakai is spoken by over 2,400 people on Barakai Island in the south-east of the Aru Islands. Four villages are on Barakai Island and one, Gomo-Gomo, is on an island of its own across a narrow strait from Mesiang in the north-east of Barakai Island. In Longgar we were told that all sections of population use the language, including school children.

Dialect Information

There are two dialects, Mesiang (population: 821) speaks one and the other four villages, Gomo-Gomo, Bemun, Longgar and Aparas speak the other dialect (combined population. 1,613). There are reportedly considerable differences between the two dialects. My analysis shows Mesiang to be 87% cognate with Longgar.

2.2.3.12 Batuley

Batuley is spoken by about 2,400 people in eight villages on small islands off the east coast of Wokam Island, and also in the village of Mariri on the island of that name off the east coast of Kobror Island. We heard the language being used in all situations. In Mariri the children reportedly learn Malay first, but all go on to learn and use the vernacular.

Dialect Information

Mariri is a divergent dialect of Batuley, on the boundaries of intelligibility. The local people consider it a separate language. The mean percentage of lexical similarity between Mariri on the one hand and Jursiang and Kumul on the other is 80%. Intelligibility testing will be needed to establish whether or not the speech of Mariri is mutually comprehensible with that of the seven villages to the north. The other seven villages are probably fairly homogenous, since Kumul and Jursiang, while geographically quite far apart, are 90% lexically similar. The matrix below shows the Batuley villages surveyed.

Mariri	-----Dialect
82	Jursiang
78	90 Kumul

MATRIX 15: The Batuley Language

2.3. Linguae Francae

2.3.1 Malay dialects

The Malay dialect used in Kei and Tanimbar appears to be similar to Ambonese Malay as spoken in Ambon, whilst Aruese Malay, although similar to Ambonese Malay, has a number of different features (for example, beta 'I' is still used, but instead of seng 'no, not' tarada or tara is used). Throughout the area Malay dialects are used between people from different language groups when they want to communicate, unless they have learnt one another's languages (a not uncommon occurrence, especially in Aru) or unless they use one of the languages below.

2.3.2 Kei

Kei is spoken throughout the Kei Islands as a native language as already stated, but it is also used as a lingua franca by speakers of the other languages spoken in the Kei Islands (Banda and Teor-Kur). It is also learnt and used by many of the non-native inhabitants of Kei (Chinese, Buton etc.).

2.3.3 Fordata

Fordata was formerly the lingua franca of the Tanimbar Islands, as well as being the language of ritual (adat) throughout the islands. It is now little known by speakers of the other three languages, although some older people still know it and use it for ritual.

2.3.4 West Trangan

West Trangan is known as the lingua franca of the Aru Islands. In the southern part of Aru it seems that it is still understood by many people who are not native West Trangan speakers. However in central and northern Aru it may only be understood by a few. Aru people can often speak other languages that are near to their own area.

3. SUMMARY

This tentative classification of the languages of the Aru, Kei and Tanimbar Archipelagos is the result of surveys carried out under the Memorandum of Cooperation between Pattimura University, Ambon, and the Summer Institute of Linguistics.

All the languages of these islands belong to the Austronesian Phylum, and, with the exception of the local Malay dialects and Indonesian (which belong to the Western Malayo-Polynesian Superstock), all the languages belong to the Central Malayo-Polynesian Superstock.

The languages of these islands divide into four stock-level groupings. I have posited the following stocks: the Kei-Tanimbar Stock and the South Tanimbar Stock. The Aru Family also represents a stock-level division and so does the Banda language. For a summary of the classification of the nineteen languages of Kei, Tanimbar and Aru see figure 4.

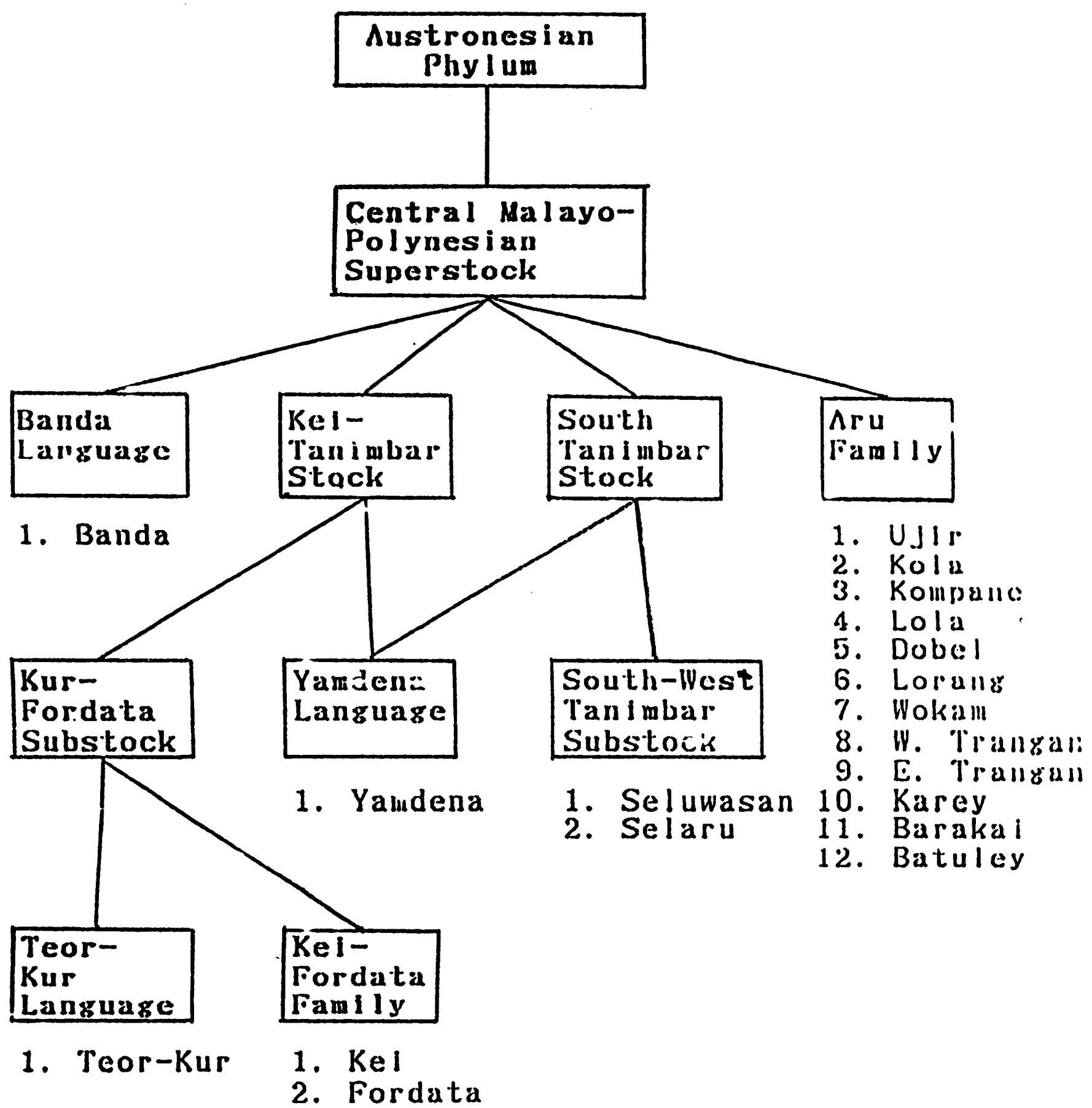


FIGURE 4:
Summary Classification: Aru, Kei and Tanimbar

NOTES

¹This area is known locally as Tenggara Dekat (The near South-East) in contrast to Tenggara Jauh (The far South-East), the islands in the south-west of the province from Babar to Wetar. Both these areas make up the Kabupaten Maluku Tenggara (The South-East Moluccas District).

²The words eliminated after elicitation were:

<u>English</u>	<u>Indonesian</u>	<u>Reason for elimination</u>
bark	kulit kayu	Polymorphemic: <u>skin</u> <u>wood</u> . Both tree and skin are on the list.
twenty	dua puluh	Polymorphemic: <u>ten</u> <u>two</u> . Both morphemes are on the list separately.
here	di sini	Same morpheme as <u>this</u> , <u>ini</u> .
there	di situ	Same morpheme as <u>that</u> , <u>itu</u> .
grandmother	nene	Same morpheme as <u>grandfather</u> , <u>tete</u> .
bathe (trans.)	mandikan	Same morpheme as <u>bathe</u> (intrans.) <u>mandi</u> .
lie down	berbaring	Consistent confusion on elicitation and often the same morpheme as <u>sleep</u> , <u>tidur</u> .

³From a diachronic perspective these would be probable cognates as loan words have not been eliminated. (Grimes and Grimes: Languages of South Sulawesi).

⁴Using the cognate set method we come closer to the cognate as defined from a diachronic perspective. Within a cognate set there may be words that are phonetically fairly different, the phonetic changes from dialect to dialect are not very great but the dialects or languages at the ends of the chain may be far enough apart that they would not be recognised as cognate if these two wordlists were compared in isolation. As this method involves comparing all the wordlists at once it makes it easier to see regular sound correspondences. These help to determine cognates.

⁵For 51 wordlists the total number of percentages computed is 1,275 which for a 200-word list represents 255,000 comparisons of words.

⁶The following were the villages where wordlists were taken, arranged according to McKinnon's dialect divisions:

Larat-Fordata I: Rumean (Rumya'an), Lamdesar Timur
 Larat-Fordata II: Sofyanin, Watidal
 Molu-Maru: Watmasa
 Seira: Rumsalut

⁷There are some marked phonetic differences between cognate words in the data between the two main dialect areas of Yamdena. These phonetic differences put the village of Alusi Krawain clearly in the northern dialect. Here are some examples of phonetic changes from the data:

<u>Indonesian</u>	<u>English</u>	<u>Indonesian</u>	<u>English</u>
11. bunga	flower	73. pendek	short
12. buah	fruit	74. dekat	near
13. kusu-kusu	a tall grass	75. jauh	far
14. sago	sago	76. penuh	full
15. beras	uncooked rice	77. baru	new
16. saya	I	78. baik	good
17. engkau	you (singular)	79. bulat	round
18. kami	we (exclusive)	80. kering	dry
19. kita	we (inclusive)	81. menjemur	to dry in the sun
20. matahari	sun	82. tidak	no
21. bulan	moon	83. ini	this
22. bintang	star	84. itu	that
23. air	water	85. di dalam	inside
24. hujan	rain	86. di atas	above, on
25. batu	stone	87. diluar	outside
26. pasir	sand	88. di bawah	below
27. pulau	island	89. di depan	in front of
28. tanah	earth, land	90. di belakang	behind
29. awan	cloud	91. pinggir	side
30. angin	wind	92. kulit	skin
31. laut	sea	93. daging	meat, flesh
32. hutan	forest	94. darah	blood
33. gunung	mountain	95. tulang	bone
34. api	fire	96. lemak, gemuk	fat
35. asap	smoke	97. rambut	hair
36. abu	ashes	98. bulu	body hair
37. panas	hot	99. kepala	head
38. bakar	bake	100. telinga	ear
39. dingin	cold	101. mata	eye
40. malam	night	102. hidung	nose
41. ikan	fish	103. mulut	mouth
42. burung	bird	104. bibir	lip
43. telur	egg	105. gigi	tooth
44. anjing	dog	106. lidah	tongue
45. marsegu	fruit bat	107. kuku	finger-/toe-nail
46. kutu	louse	108. kaki	leg/foot
47. nyamuk	mosquito	109. lutut	knee
48. tikus	mouse	110. tangan	arm/hand
49. ular	snake	111. perut	stomach
50. ekor	tail	112. leher	neck
51. hitam	black	113. susu	milk
52. putih	white	114. hati	liver
53. merah	red	115. muka	face
54. kuning	yellow	116. tahi	excrement
55. hijau	green	117. air kencing	urine
56. satu	one	118. tuli	deaf
57. dua	two	119. buta	blind
58. tiga	three	120. orang	person
59. empat	four	121. laki-laki	male
60. lima	five	122. perempuan	female
61. enam	six	123. suami	husband
62. tujuh	seven	124. istri	wife
63. delapan	eight	125. bapak, ayah	father
64. sembilan	nine	126. ibu	mother
65. sepuluh	ten	127. nama	name
66. seratus	a hundred	128. tete	grandfather
67. seribu	a thousand	129. kakak	elder sibling
68. semua	all	130. adik	younger sibling
69. banyak	many	131. cucu	grandchild
70. besar	big	132. om, paman	uncle
71. kecil	small	133. tante	aunt
72. panjang	long	134. kawan	friend

<u>Indonesian</u>	<u>English</u>	<u>Indonesian</u>	<u>English</u>
135. tamu	guest	170. mengantuk	sleepy
136. tali	rope	171. anak	child
137. jalan	road	172. langit	sky
138. perahu	sailing canoe	173. terbang	fly
139. layar	sail	174. datang	come
140. parang	machete	175. berdiri	stand
141. pisau	knife	176. batuk	cough
142. lesung	mortar	177. meludah	spit
143. alu	pestle	178. menangis	cry
144. garam	salt	179. bermimpi	dream
145. gula	sugar	180. mengandung	pregnant
146. apa	what	181. tua	old
147. siapa	who	182. sungai	river
148. di mana	where	183. membunuh	kill
149. kenapa	why	184. duduk	sit
150. bagaimana	how	185. mandi	bathe
151. berapa	how many	186. jatuh	fall
152. haus	thirsty	187. jatuhkan	drop
153. menggigit	bite	188. memberi	give
154. minum	drink	189. bicara	speak
155. lapar	hungry	190. lintah	leech
156. makan	eat	191. tebal	thick
157. melihat	see	192. tipis	thin
158. mendengar	hear	193. tajam	sharp
159. tidur	sleep	194. busuk	rotten
160. tahu	know	195. pikul	carry (on shoulder)
161. bangun	wake up	196. lari	run
162. membangunkan	awake (someone)	197. pukul	hit
163. mati	die	198. luka	wound
164. berenang	swim	199. rumah	house
165. berjalan	walk	200. kampung	village
166. bernyanyi	sing	201. otak	brain
167. muntah	vomit	202. berkeringat	sweat
168. tertawa	laugh	203. karang	coral
169. gatal	itch		

APPENDIX B

Wordlists were taken and used from the following villages. Unless stated otherwise the lists were taken in the village concerned. The informants were male unless marked (f).

I. Kei Islands

<u>Village</u>	<u>Where taken</u>	<u>Primary Informant</u>	<u>Age</u> <u>Date</u>	<u>Relia- bility</u>	<u>Linguist</u>
Banda Elat		Humut Suat	45 1.1.85	VG/G	J. Hughes
Pulau Ut	Tual	Abdul Hamit	20 3.10.85	VG	E. Travis
Kur (Kilmas)	Tual	Hasim Sarkol	35 3.1.85	G	E. Travis
Watsin		F. Heatubun	19 31.12.84	VG	J. Hughes
Wulurat		D. Serkol	33 1.1.85	VG	J. Hughes
Waur		F. Sibeubun	53 1.1.85	F	E. Travis

<u>Village</u>	<u>Where taken</u>	<u>Primary Informant</u>	<u>Age</u>	<u>Date</u>	<u>Reliability</u>	<u>Linguist</u>
Rumaat		R. Wellikin	18	3.1.85	G	E. Travis
Debut		Rudy Jamlean	17	2.1.85	G	E. Travis
Dullah		Gasim Renuan	38	4.1.85	G	J. Hughes
Tayando (Ohiel)	Tual	K. Rahantan	45	3.1.85	G	J. Hughes
Taam	Tual	Abdul Haris	29	3.1.85	G	J. Hughes
Tanimbar Kei	Tual	R. Tabalubun	27	3.1.85	G	J. Hughes

II. Tanimbar Islands

Sofyanin	Larat	Petrus Waturu	32	12.1.85	VG	E. Travis
Watidal		A. Kulale'en	53	10.1.85	VG	J. Hughes
Rumean	Larat	Willem Wekan	18	12.1.85	G/F	J. Hughes
Landesar Timur	Larat	Petrun Fun	19	12.1.85	G	E. Travis
Watmasa		Saul Masela	54	11.1.85	VG	J. Hughes
Seira (Rumsalut)	Saumlaki	M. Refinaly	59	7.1.85	G/F	E. Travis
Arma		M. Batmomolin	53	9.1.85	VG	E. Travis
Watumuri		A. Lartutul	55	15.1.85	VG	E. Travis
Waturu		G. Sarwuna	52	15.1.85	VG	J. Hughes
Alusi Krawain		R. Sarluri	50	9.1.85	VG	J. Hughes
Atubul Dol		D. Bulurdity	44	8.1.85	VG	E. Travis
Olilit Baru		M. Fase (f)	30	18.1.85	VG/G	J. Hughes
Latdalam (Yamdena speaking)	Saumlaki	P. Ngilamelek	48	17.1.85	VG	J. Hughes
Wermatang	Saumlaki	D. Samangun	30	6.1.85	VG	J. Hughes
Makatian	Saumlaki	i. Oinanditi	23	6.1.85	G	E. Travis
Latdalam (Selaru Speaking)	Saumlaki	D. Rankoratat	55	17.1.85	G	E. Travis
Kandar	Saumlaki	H. Refwalu	45	7.1.85	VG	J. Hughes

III. Aru Islands

Ujir		N. Hatala	50	4.4.85	VG	J. Hughes
Warialau		D. Sintimir	39	3.4.85	G	J. Hughes
Kulaha		A. Wamit	38	3.4.85	VG	Y. Taguchi
Mohang Sel		S. Matally	38	2.4.85	VG	J. Hughes
Kompane		H. Galgal	35	1.4.85	VG	Y. Taguchi
Kumul		(a few people)		1.4.85	G	J. Hughes
Jursiang		W. Djeeroban	48	31.3.85	VG	Y. Taguchi
Mariri		A. Jamarfui	21	31.3.85	VG	J. Hughes
Lola		H.R. Henkesa	34	30.3.85	VG	Y. Taguchi
Warabal	Lola	D. Mangar	29	30.3.85	VG	H. Hughes
Warjukur		F. Balsala	70	2.10.85	VG/G	J. Hughes
Wailay		Yuda Uyer	26	23.3.85	VG	Y. Taguchi
Manjau		Yahaya Lehui	29	22.3.85	VG	Y. Taguchi
Warabola		Demi Mangar	39	22.3.85	VG	J. Hughes
Kobadangar		H. Dabamon	40	24.3.85	VG	J. Hughes
Lorang		E. Djermor	50	22.3.85	VG	J. Hughes
Mesiang		Sawal Welai	42	25.3.85	VG	T. Taguchi
Longgar		D. Kobawan	40	25.3.85	VG	J. Hughes
Karey		J. Kailansian	36	27.3.85	VG	J. Hughes
Gomar Meti		Mustari Golap	34	26.3.85	VG	Y. Taguchi
Feruni	Dobo	A. Garbanasi	18	4.4.85	F	J. Hughes
Benjina		Zakeus Fukar	38	21.3.85	VG	Y. Taguchi

BIBLIOGRAPHY

- BARRAUD, Cecile
 1979 Tanebar-Evav, une société de maisons tournée vers le large. Cambridge University Press and Éditions de la Maison des Sciences de l'Homme.
- BLUST, Robert A.
 1977 "The Proto-Austronesian Pronouns and Austronesian Subgrouping: A Preliminary Report." Working Papers in Linguistics 9.2:1-15.
- 1978 "Eastern Malayo-Polynesian: a Subgrouping Argument." In: Proceedings of the Second International Conference on Austronesian Linguistics, Fascicle 1. Pacific Linguistics C-61:181-234. Canberra: Australian National University.
- 1981 "The Soyobo Reflexes of Proto-Austronesian *S." In: Historical Linguistics in Indonesian Part I, ed. by R.A. Blust. NUSA, Linguistic Studies in Indonesian and Languages in Indonesia 10:21-30. Jakarta: Badan Penyelenggara Seri NUSA.
- BRUMOND, J. F. G.
 1843 "Proeve over de Aroe-Taal." Tijdschrift voor Nederlandsch Indie 6(2):321-40.
- CASAD, Eugene H.
 1974 Dialect Intelligibility Testing. Norman, Oklahoma: Summer Institute of Linguistics.
- COLLINS, James T.
 1983c "Linguistic Research in Maluku: a Report of Recent Field Work." Oceanic Linguistics, vol. XXI, Nos. 1,2: 120-23, 125-33.
- DRABBE, P.
 1932a Woordenboek der Fordaatsche Taal. Verhandelingen van het Koninklijk Bataviaasch Genootschap 71:1-118.
- 1932b Woordenboek der Jamdeensche Taal. Verhandelingen van het Koninklijk Bataviaasch Genootschap 71:1-122.
- 1932c Beknopte Spraakkunst en Korte Woordenlijst der Slaroesche Taal. Verhandelingen van het Koninklijk Bataviaasch Genootschap 71:1-32.
- DYEN, Isidore
 1962 "Some New Proto-Malayo-Polynesian Initial Phonemes." Journal of the American Oriental Society 82:214-15.
- 1965b A Lexicostatistical Classification of the Austronesian Languages. International Journal of American Linguistics, Memoir 19, Baltimore.

- DYEN, Isidore
1978a "The Position of the Languages of Eastern Indonesia." In: Proceedings of the Second International Conference on Austro-nesian Linguistics, Fascicle 1. Pacific Linguistics C-61:235-254.
- EIJBERGEN, H.C. van
1865 "Korte Woordenlijst van de Taal der Aroe- en Kei-Eilanden." Tijdschrift voor Indische Taal-, Land-, en Volkenkunde 14:557-68.
- GEURTJENS, H.
1921 Woordenlijst der Keieesche Taal. Verhandelingen van het Koninklijk Bataviaasch Genootschap 53:1-196.
- GRIMES, Charles E. and Barbara D.
to appear Languages of South Sulawesi.
1984 "Languages of the North Moluccas: A Preliminary Lexicostatistic Classification." Ms.
- GUDSCHINSKY, Sarah
1956 "The ABCs of Lexicostatistics (Glottochronology)." Word 12:175-210. Reprinted in: Language in Culture and Society, Dell Hymes, ed., 1964. New York: Harper and Row.
- LOVING, Richard and Gary SIMONS, eds.
1977 "Language Variation and Survey Techniques," Workpapers in Papua New Guinea Languages, Vol. 21, Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- McKINNON, Susan
1983 Hierarchy, Alliance and Exchange in the Tanimbar Islands. Chicago. (PhD Thesis).
- SALZNER, Richard
1960 Sprachenatlas des indopazifischen Raumes. Wiesbaden, Otto Harrassowitz.
- SANDERS, Arden G.
1977 "Guidelines for Conducting a Lexicostatistic Survey in Papua New Guinea." In: Language Variation and Survey Techniques. Loving and Simons eds. 21-43. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- SIMONS, Gary
1977 "Recognizing Patterns of Divergence and Convergence in a Matrix of Lexicostatistic Relations." In: Language Variation and Survey Techniques, Loving and Simons eds. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

SMITH, Kenneth D.

1984 Languages of Sabah: A Survey Report. Julie K. King and John Wayne King, eds. Pacific Linguistics C-78:1-49. Canberra: Australian National University.

VOEGELIN, C.F. and F.M.

1977 Classification and Index of the World's Languages. New York: Elsevier.

WURM, S.A. and B. WILSON

1975 English Finderlist of Reconstructions in Austronesian Languages (Post-Brandstetter). Pacific Linguistics C-33. Canberra: Australian National University.

WURM, S.A. and Shiro HATTORI, gen. eds.

1981 Language Atlas of the Pacific Area. Pacific Linguistics C-66. Canberra: Australian National University.