APPENDIX A. THE SOUTH SULAWESI LANGUAGE GROUP

A.1 Introduction

The concept of a South Sulawesi language group, introduced by Esser's (1938) broad classificatory maps of Netherlands Indies languages, has survived the advent of modern comparative linguistic studies initiated by Mills (1975a, 1975b). Its justification lies in a suite of generally shared qualitative characters, some of which are unique to the group, rather than in the lexicostatistical data which in fact would tend to remove the Makassar languages from the group (Sirk, 1989). Small bodies of lexicostatistical data may be found in several works (e.g. Mills, 1975a; Sirk, 1989) but none approaches the comprehensiveness of Grimes and Grimes (1987). Their study has since been constructively reviewed by Friberg and Laskowske (1989) who proposed several revisions (A.6) including alternative names to be preferred for some of the speech communities. However, because my primary original contribution will be to explore the implications of the Grimes' lexicostatistical data, I will review their scheme directly and retain their names for the speech communities.

Their database consists of a standardised 202-word list, slightly modified from the Swadesh 200-word list, recorded from 39 dialects and languages in South Sulawesi, with standard Indonesian as the 40th comparative group. Using the Grimes' abbreviations as spelled out in my later figures, the large map in Figure A-1 locates the 39 South Sulawesi speech communities based on the distribution maps of Grimes and Grimes (1987: Maps 2 and 7 to 11). Grimes and Grimes inspected the word lists to ascertain the percentages of words which appeared similar between each two speech communities. Following the guidelines of Simons (1977a), they presented these percentages in a 40 by 40 half-matrix sorted so as to place (i) the highest percentages of similar words along the diagonal, and (ii) the speech communities with the highest average lexicostatistical similarity (compared against the other speech communities) closest to the centre of the assortment. The half-matrix allowed them to classify

the speech communities into languages, subfamilies, families and stocks, i.e. groups whose members respectively show around 75-80%, 60-75%, 45-60% and 25-45% lexicostatistical similarity.

The two maps in Figure A-1 compare the Grimes' classification with the one suggested here by my formal statistical analysis. Note that the lexicostatistical boundaries can be read both externally and internally. For instance, in the Grimes' classification (the smaller map) the Makassar languages comprise a family when compared against the other families in the South Sulawesi stock, but internally they comprise a subfamily because the Grimes recognised languages as the highest internal division.[1] The differences between the two schemes will be reviewed in Sections A.4 and A.6.

A.2 Scatter Analysis of the Lexicostatistical Data

As a first analysis, let us disregard the issue of which speech communities group together so as to obtain a global picture. Here I chose multidimensional scaling with the similarities between the speech communities (hereafter, "operational taxonomic units" or OTUs) displayed on two-dimensional Euclidean space. I entered the half-matrix of similarities published by Grimes and Grimes into the OSIRIS MDSCAL package on the VAX computer at the Australian National University, and chose the appropriate options. The analysis juggled the OTUs until they found the configuration in which the closest OTUs are those with the highest similarity coefficients, and the most distant OTUs are those with the lowest similarity coefficients. That is, the rank order of Euclidean distances most fully reverses the rank order of inter-OTU similarities (Kruskal, 1964a), and the extent of mismatch from a perfect reversal can be measured as a percentage in "stress" (Kruskal, 1964b). The resulting configuration has 18.2% stress which, though technically verging towards poor, is in my experience a remarkably good

[1] In my formal analysis shown in the larger map, Makassar would be a stock when compared externally but a family when compared internally.

result considering the number of OTUs involved (cf. Bulbeck, 1981). Although the choice of three or more Euclidean dimensions would have resolved much of the stress, the results would have been harder to visualise, nor would it be possible to compare the configuration with the geographical distribution of the speech communities.

This is done in Figure A-2 which not only shows the two-dimensionally scaled configuration of the speech communities, but also fits a contorted map of South Sulawesi over the configuration. The exercise clearly shows the very strong influence of geography over lexicostatistical variation because, were these phenomena not strongly related, the large number of OTUs involved would render futile any map-fitting attempt. It is even possible to draw in *kabupaten* boundaries such that only one community, Mambi, cannot be placed in the correct *kabupaten*. Even though Kabupaten Luwu has speech communities from three "stocks" in Grimes and Grimes' classification, and Selayar and Mamuju each have representatives from two stocks, all three *kabupaten* retain their convex integrity in the map-fitting exercise (cf. Figure A-1).

Of course geography is not exercising direct control over lexicostatistical patterns so much as indirect control through its influence over population densities and avenues of communication. Thus the boundary between the Makassar and Bugis languages essentially follows the spine of the highest mountains of the western cordillera, while Bugis is demarcated from the languages of the "Northern South Sulawesi family" by a line running along the transition from prevailing lowlands to prevailing highlands (cf. Figures A-1 and 1-4). The main body of the South Sulawesi peninsula, a large area of well-populated lowlands inhabited by the Bugis (Figure A-1), has become greatly reduced in the "language map" (Figure A-2). Conversely the most sparsely populated region of South Sulawesi, corresponding to my "Northern Central Highlands" (Figure 1-4), has become grossly enlarged, except for a specific restriction corresponding to the Palu-Koro fault which is the main corridor through the region (Figure A-2).

All in all, the exercise highlights the point made by Grimes and Grimes (1987:11,13-14) - exchange of lexical items between communities in contact is a ubiquitous and constant process, regardless of the level of taxonomic distinctiveness between the speeches involved. Moreover the distortions to physical space themselves usefully highlight specific linguistic relationships. For instance, Campalagian sits on an imaginary peninsula jutting towards the Bugis dialects, with which Campalagian is most closely related, and away from the contiguous Mandar dialects. Selayar island has been "dragged" into the Gulf of Bone as a correct reflection of Laiyolo's specifically close relationships with Pamona, Padoe and especially Wotu.

A.3 Hierarchical Seriations and Lexicostatistical Data

Hierarchical seriations (or seriated dendrograms) represent a means of combining the best aspects of two important branches of statistical analysis, seriations and hierarchical clusters, in order to counteract weaknesses in either method used alone. Although seriation is most widely used to detect chronological ordering, what the exercise actually does is order items into a one-dimensional series, along a scale which need not be time (Irwin, 1985:122-125) and in this case will be social space. Hierarchical clustering, on the other hand, joins up OTUs into clusters via one of a variety of clustering algorithms (e.g. Sneath and Sokal, 1973; Kronenfeld, 1985). As usually applied, a weakness of hierarchical clustering is that, once an OTU joins a particular cluster, any specific relationships of the OTU with individual OTUs in other clusters is lost. This is because people believe that the sister branches of a hierarchical dendrogam can be twiddled at will without altering the significance of the results (e.g. Sneath and Sokal, 1973). However, here I will argue that the sister branches should be twiddled to reflect a seriated order.

The extra information available in the seriated dendrogam, compared to its unseriated counterpart, is the representation of the wider relationships of an OTU beyond those within its cluster. So for instance when two OTUs are similar, but one

(or both) has still closer links with a cluster which locks out the other OTU, the seriation may refind the temporarily lost inter-OTU similarity by placing the two OTUs next to each other at the adjacent peripheries of their respective clusters.

The extra information available for the seriation is that we have a structure for the OTUs over and above their ordering along a single dimension. It would of course be possible to produce the seriation first and then seek out the structure within the seriated OTUs, as for instance in the sorted and shaded matrices produced by Irwin (1985:125ff.) - a comparison which illustrates the point that seriations are clustered strings of clustered OTUs. In this respect the novelty of the present approach is that it produces the seriation from the structure rather than *vice versa*.

A further advantage of using hierarchical clustering to produce the seriation is the computational utility. Whereas the number of possible orders rapidly approaches a huge number as the number of OTUs increases, the superimposition of a hierarchical dendrogram drastically decreases the number of possible orders, and also allows whole suites of possible orders to be rejected merely by analysing how the higher-level blocks of OTUs relate to each other. True, given the size of the half-matrices with which I am dealing, up to 40 by 40, enough feasible orders still remain such that demonstration of the best order would require a specially written computer program. My efforts in computerisation do not go beyond writing a formula, using the Borland/Analytica (1987) "Reflex: The Analyst" PC programme, to assist in calculating the "goodness of fit" as measured here. However, applying this measure will show, at least for the data considered here, the degree to which seriated orders can vary with almost equally good results as long as the main axis of variation is observed.

As a final introductory point, I will cluster using the average linkage technique 'whereby admission of an item into a cluster is based on the average of the similarity of the item (or lot) with the other items in a cluster' (Irwin, 1985:131). Since our original coefficients of similarity are

a simple additive variable, i.e. the number of lexically similar items expressed as a percentage, the estimate of average similarity (as derived from average linkage clustering) equals the average percentage of similar words which the word lists in one cluster share with the word lists in the other cluster as would have been deduced from comparing the original word lists. For example, say one cluster has two word lists and the other has three, so that in comparing the clusters we are actually comparing two times three equals six word lists, then both average linkage clustering and comparing the original word lists use the same formula - (number of similar items across the word lists [first word] plus number of similar items across the word lists [second word]...plus number of similar items across the word lists [202nd word]) divided by (six [word lists] times 202 [words]) multiplied by 100 (to obtain a percentage). I raise the point because it is generally overlooked that once two OTUs join into a common stem in a dendrogram, it is the common stem which behaves thereafter, and logically the analysis should proceed as though the two OTUs had been pooled. The statistical relationships which prevail after the clustered OTUs are treated as a pooled sample need not always be accurately predictable from the relationships which had prevailed beforehand (B.4.1). However, this equivalence does hold true in the present case where (i) every original OTU has the same sample size, here 202 words, (ii) the mathematical relationship is expressed by a simple additive variable, and (iii) average linkage clustering is used.

Grimes and Grimes (1987) themselves combined variants of seriation and hierarchical clustering but in the reverse order from that applied here. After sorting their half-matrix (A.1) they followed the guidelines of Simons (1977b) to identify cases of "successive divergence", where values away from the diagonal decrease down the columns but not along the rows; "sporadic convergence", where a cell has an anomalously high value compared to the cells around it; and "chaining", where the values away from the diagonal decrease along the rows as well as down the columns. From their inspection they carved the sorted half-matrix into blocks, and represented

the inter-relationships with dendrograms in which the branches diverge to show successive divergence, converge to show sporadic convergence, and criss-cross to show chaining (Grimes and Grimes, 1987:10-11ff.). In my terms successive divergence is the classical hierarchical dendrogram where the ordering of the OTUs is of no significance; chaining is a hierarchical dendrogram whose OTUs need to be seriated to bring out the full relationships; and sporadic convergence represents cases where two OTUs show a specifically close relationship which cannot however be produced from seriating the dendrogram. These points will become clear by following the order in which the OTUs would cluster together.

A.4 Hierarchical Seriation of the Data From South Sulawesi

Let us start with what is the most tightly clustered block of OTUs, the "Mandar language" of Grimes and Grimes. The seriated dendrogram appears as given below, with the values at which the OTUs cluster by average linkage shown at the point of joining:

```
DENDROGRAM A-1                    80.5|
                  88|                        |88
        95|                |                |
         |        |        |                |        |
       BAL      MAJ      PBG              SEN      MAL
```

The sorted half-matrix of lexicostatistical similarities is:

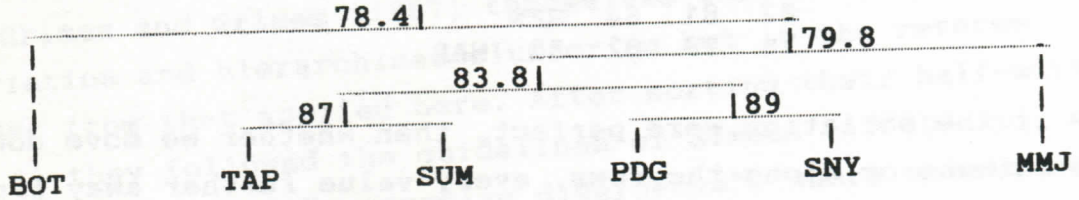
BAL				
95	MAJ			
88	88	PBG		
82	81	84	SEN	
76	78	82	88	MAL

Now if the seriation were perfect, then whether we move down the columns or along the rows, every value further away from the diagonal should become smaller or at least stay the same. All cases of larger values further away from the diagonal than smaller values represent stress in the seriation. Hence we consider the sum of all the possible differences between any cell closer to the diagonal and any cell further away, and the sum of all these possible differences but ignoring

the sign of the difference. Considering the first value as the numerator and the second as the denominator, and expressing the result as a percentage, measures the "goodness of fit" of the seriation compared against a perfect seriation. In this case, (95-88) + (95-82) + (95-76) + (88-82) + (88-76) + (82-76) + (88-88) + (88-81) + (88-78) + (81-78) + (84-81) + (84-82) + (81-82) + (84-82) + (88-82) + (88-78) + (88-76) + (82-78) + (82-76) + (78-76), all divided by the sum |95-88| + |95-82| + |95-76| + |88-82| + |88-76| + |82-76| + |88-88| + |88-81| + |88-78| + |81-78| + |84-81| + |84-82| + |81-82| + |84-82| + |88-82| + |88-78| + |88-76| + |82-78| + |82-76| + |78-76| (where |x| signifies sign ignored), equals 129/131 or 98.5%. This order, also found by Grimes and Grimes, is easily spotted as the best, although Sendana (SEN) and Malunda (MAL) could have been reversed with almost equally satisfactory results. Mandar is however a clear case of a dialect chain (Grimes and Grimes, 1987:36) with, as shown here, the two subgroupings of Balanipa/Majene/Pamboang (BAL/MAJ/PBG) and SEN/MAL. Geographically speaking we are looking at a steady transition along the Mandar Peninsula coastline, from Balanipa in the Gulf of Mandar, to Malunda to the north (see Figure A-1).

The block of sampled speech communities which Grimes and Grimes identify as the Mamuju language forms an almost equally tight cluster. Grimes and Grimes (1987:39) also found the best order for this dialect chain whose goodness of fit is 158/186 or 85.0%:

DENDROGRAM A-2

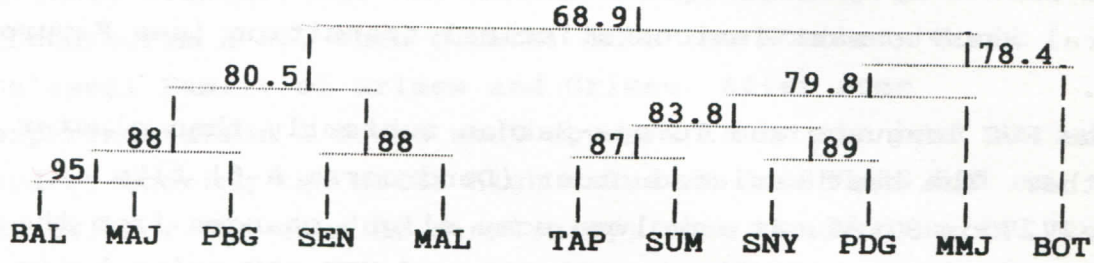


As represented here we are looking at a south-to-north lexicostatistical transition along the Mandar Peninsula north of the Mandar language area, although with Botteng (BOT) an isolate to some degree (see Figure A-1). As with Dendrogram A-1, and Dendrogram A-4 which follows, Grimes and Grimes give

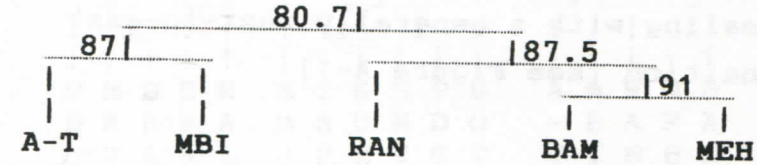
a counterpart dendrogram which contains very much the same information, but using the conventions proposed by Simons (1977b).

Botteng acts as an isolate in the wider context as well, as seen in the cluster formed by the Mandar and Mamuju languages. The best possible order is obtained by simply juxtaposing the two clusters as given above, except that Botteng swings around to fall beyond Mamuju (MMJ). This order which, excluding Botteng, suggests a south-to-north dialectical transition encompassing both languages from Balanipa to Mamuju, scores a goodness of fit of 2511/2751 or 91.3%.

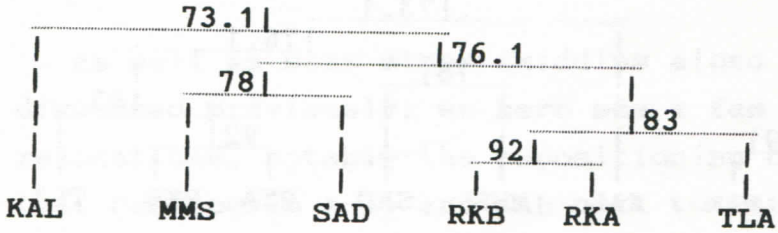
DENDROGRAM A-3



Let us move eastwards into the Southern Central Highlands (see Figure 1-4), considering first the block of speech communities which the Grimes identify as the Pitu Uluna Salu (PUS) language, and then their Toraja-Sakdan subfamily.



DENDROGRAM A-4
PUS language
"Fit":
114/116 = 98.3%



DENDROGRAM A-5
Toraja-Sakdan
subfamily
"Fit":
247/279 = 88.5%

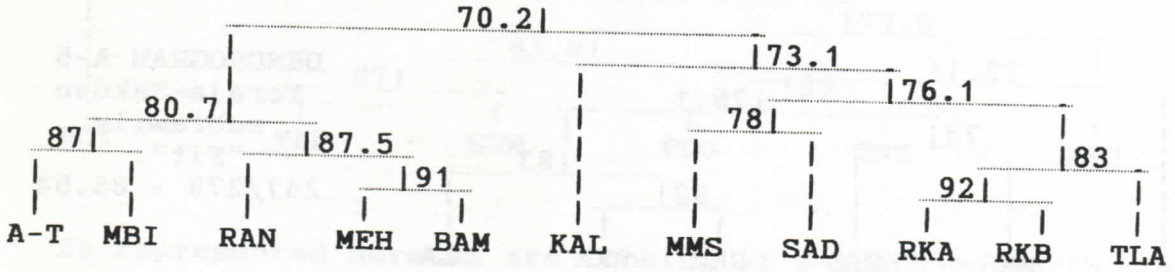
The PUS dialect chain as shown here was also identified by Grimes and Grimes (1987:42) and shows a south-to-north transition except that Bambang (BAM), which is located north

of Mambi (MBI), shows a specific affinity with Mehalakan (MEH) to the south (see Figure A-1).

As regards the Toraja-Sakdan subfamily, Grimes and Grimes (1987:46) placed Rongkong, i.e. Rongkong Bawah (RKB) and Rongkong Atas (RKA) treated as a single entity, between Mamasa (MMS) and Sakdan (SAD), and identified a sporadic convergence involving Rongkong and Toala. If we keep RKB and RKA separate then the Grimes' order can score no better than 205/269 = 76.2%. In the half-matrix which treats Rongkong as a single entity (Grimes and Grimes 1987:46), the counterpart to the order given here scores 129/141 = 91.5%, while the Grimes' order scores 115/141 = 81.6%. Clearly the former is more satisfactory, even though Kalumpang (KAL) and Rongkong would show a slight sporadic convergence, and indicates a general west-to-east lexicostatistical transition (see Figure A-1).

The PUS language and Toraja-Sakdan subfamily then cluster together. The best seriated order (Dendrogram A-6) fits at 2533/2727 or 92.9%. It involves some slight changes from the orders shown above, with the RKA-RKB and MEH-BAM pairs being twiddled around, in both cases involving small and quite insignificant improvements to the seriated order. Further, the placement of RAN (Rantebulahan) with respect to the MEH-BAM pair is also of little significance. For instance, if we changed the RAN-MEH-BAM order given below to MEH-BAM-RAN, the goodness of fit would be measured as 2478/2670 or 92.8%. In all cases we are dealing with a generally west-to-east lexicostatistical transition (see Figure A-1).

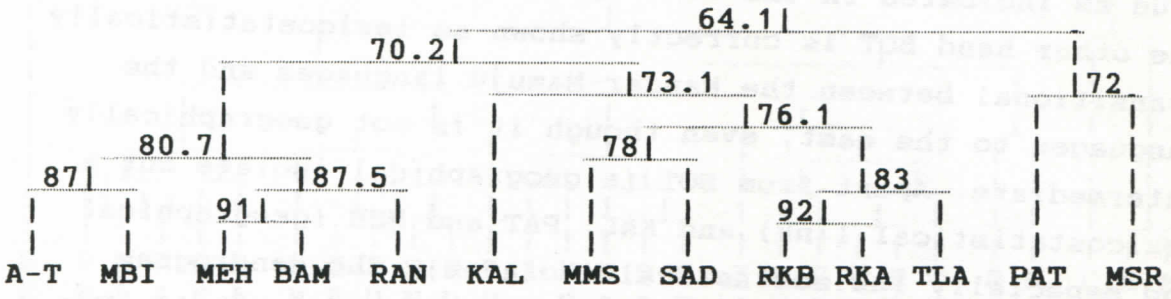
DENDROGRAM A-6



The next event involves Pattaek (PAT) clustering with Masenrempulu (MSR) and their common stem joining Dendrogram A-6 to the right-hand side of Toala (TLA). The best seriated

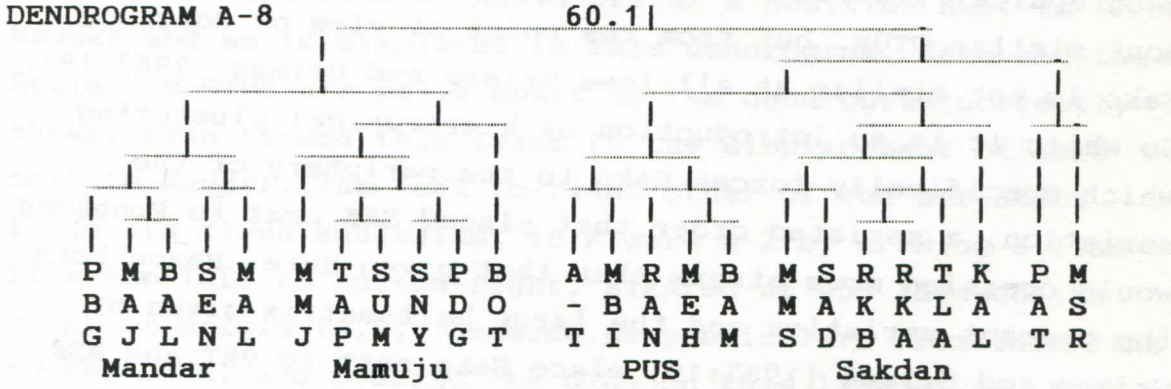
order, which also involves some fiddling with RAN-MEH-BAM and RKA-RKB, scores 4537/4845 or 93.6%, and places PAT and MSR in a position homologous to that in the sorted half-matrix of Grimes and Grimes (1987:19).

DENDROGRAM A-7



This cluster, and the Mandar-Mamuju cluster given above (Dendrogram A-2), then cluster into the Northern South Sulawesi Family of Grimes and Grimes. After some experimentation I found what appears to be the best seriated order, scoring the still high value of 34,887/38,523 or 90.6%. For the sake of simplicity, this dendrogram and most of those which follow will dispense with the values for the lower-level joins.

DENDROGRAM A-8



As well as some minor twiddles along the same lines discussed previously, we here see a few more prominent relocations, notably the repositioning of MMJ next to Mandar (cf. Dendrogram A-3) and KAL next to Pattaek (cf. Dendrogram A-7). These switches have occurred because both MMJ and KAL are to some degree isolates which do not fit in with the wider lexicostatistical patterns. They have been pushed to the furthest peripheries which the higher-level structure of

the seriated dendrogram allows because this involves the smallest possible stress in the overall seriation, at least as gauged here. It would however be quite wrong to read this diagram as implying that MMJ is close to Mandar and KAL is close to Pattaek-Masenrumpulu, because quite the reverse is true as indicated in the dendrograms given previously.[2] On the other hand BOT is correctly shown as lexicostatistically transitional between the Mandar-Mamuju languages and the languages to the east, even though it is not geographically intermediate. Apart from BOT (a geographical isolate but a lexicostatistical link) and KAL, PAT and MSR (geographical and especially lexicostastical isolates), the dendrogram shows a clear west-to-east transition (see Figure A-1). The great separation between Mandar and PAT in the dendrogram, even though they are geographically adjacent, points to a sharp distinction at their border.

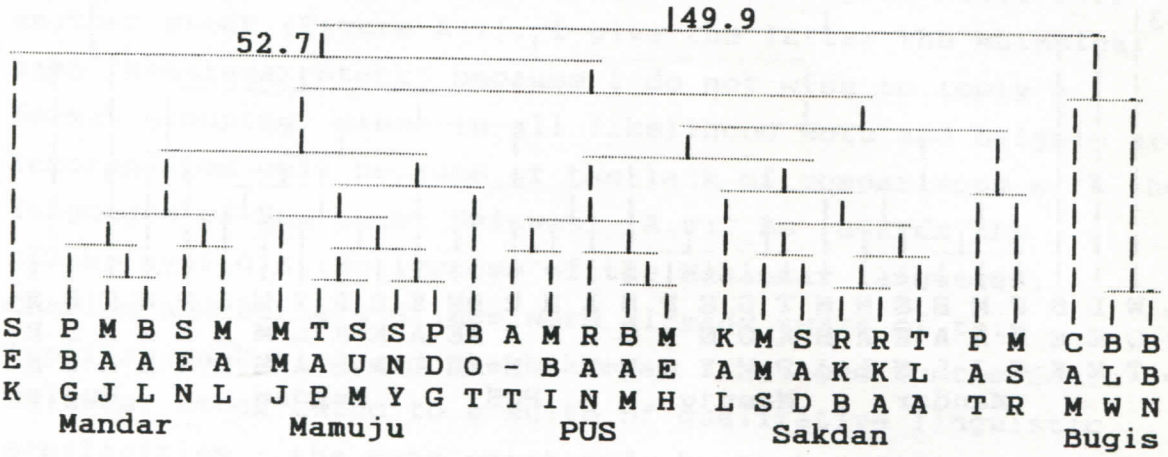
Seko Tengah (SEK) then joins Dendrogram A-8 with an average 52.7% lexicostatistical similarity. It falls to the far right of MSR, leaving the rest of the dendrogram's order unchanged, but lowering the goodness of slightly, to 40,195/44,432 or 90.5%. From the point of view of Seko the geographically close RKA and RKB (see Figure A-1) are the most similar OTUs, but from the point of view of Rongkong, Seko is not similar at all (see Grimes and Grimes, 1987:19). So while it is my introduction of hierarchical clustering which specifically forces Seko to the periphery of the seriation, a seriated order that placed SEK next to Rongkong would occasion more stress than that given here. Hence both the present seriation and the large half-matrix given by Grimes and Grimes (1987:19) place Seko next to PAT and MSR, even though Seko is specifically dissimilar from the latter OTUs, as a compromise towards bringing Seko as close as reasonable to Rongkong and the other Toraja-Sakdan OTUs.

Our next cluster consists of Bugis, represented by the two dialects Luwuk Bugis (BLW) and Bone Bugis (BBN), plus Campalagian (CAM) which joins the Bugis even though its highest individual similarities are with contiguous Mandar

[2] One of the advantages of building up the dendrogram in the step-wise procedure followed here is that it pre-empts these sorts of erroneous misreadings.

dialects. This cluster joins the main dendrogram after Seko had joined. After some experimentation the best order appeared to be:

DENDROGRAM A-9

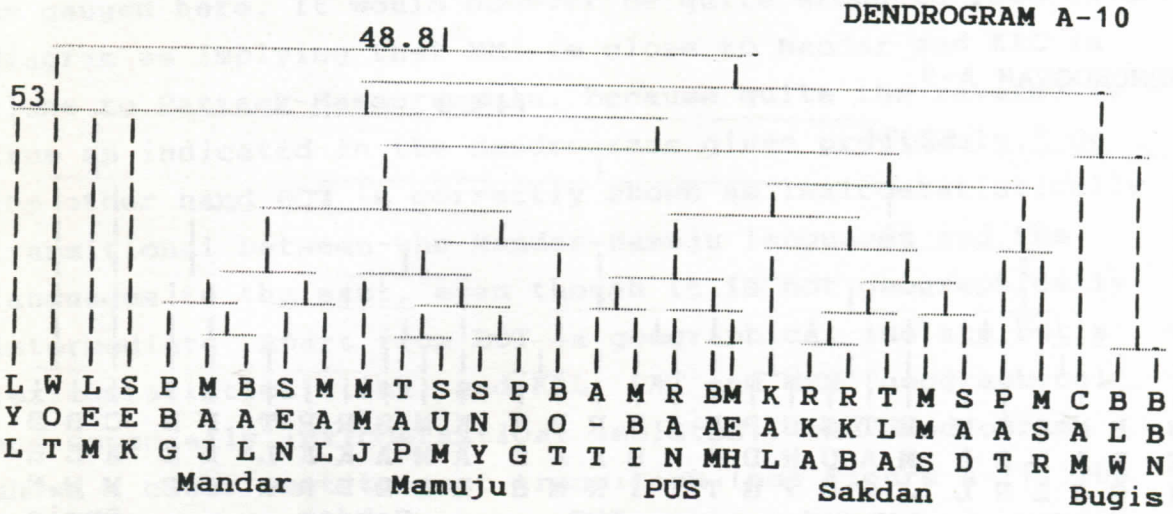


The dendrogram's goodness of fit, 53,302/62,957 or 84.7%, shows that adding the Bugis Family creates considerable stress. Bone and especially Luwuk Bugis show quite strong lexicostatistical similiarities with Pattaek (PAT) and Toala (TLA), and to a lesser degree with MSR, Rongkong and Sakdan (SAD), causing the Bugis Family to fall at the right of the dendrogram. CAM however would prefer a position next to Mandar and so is misplaced in this dendrogram. Seko and the Bugis-CAM OTUs are poles apart as the dendrogram correctly shows, even though this leads to the misplacement of Seko next to Mandar. The leFt-to-right order of the seriation finds its close equivalent in Figure A-2 by drawing a line from top-left to bottom-right, slanted at 45° through the point of origin (0,0), drawing perpendiculars from this line to the OTUs, and placing the OTUs on this line at the base of the perpendiculars. Finally, note that the isolate KAL, which is lexicostatistically dissimilar from the Bugis OTUs, has been switched back towards PUS (cf. Dendrogram A-8).

The next cluster to join the main cluster is the triplet which includes Lemolang, Wotu and Laiyolo (LEM-WOT-LYL). Wotu

shares 53% lexicostatistical similarity with both LEM and LYL, hence we are dealing with a three-way join at 53%.[3]

DENDROGRAM A-10



The goodness of fit of this order is 76,646/91,075 or 84.2%, showing that the placement of LYL-WOT-LEM at the Mandar end of the seriation adds yet further stress, because they (like Seko) would prefer the Toraja-Sakdan end of the seriation which has been pre-empted by the Bugis Family. Note that switching the extremes would create further stress, e.g. the order BBN-BLW-CAM-MSR-PAT-PBG-MAJ-BAL-SEN-MAL-MMJ-TAP-SUM-SNY-PDG-BOT-A-T-MBI-RAN-MEH-BAM-RKB-RKA-TLA-MMS-SAD-KAL-SEK-LEM-WOT-LYL scores 73,146/92,130 or 79.4%. Also note that the Grimes, and by Friberg and Laskowske (1989), place Wotu and Laiyolo in the Muna-Buton stock, whereas they place the speech communities from LEM to Bugis in the South Sulawesi stock.

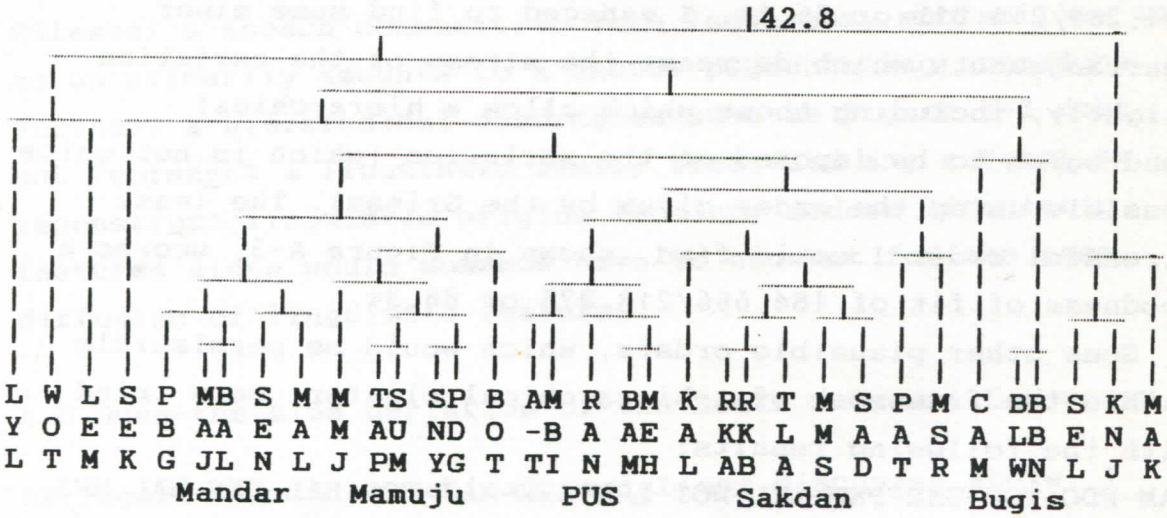
Our next addition to the main dendrogram involves the Makassar subfamily of Grimes and Grimes (1987), represented by Lakiung Makassar (MAK), Coastal Konjo (KNJ) and Selayar (SEL). Even though the Grimes (see Figure A-1) and Friberg and Laskowske (1989) place the Makassar languages in the South Sulawesi stock, the present analysis shows that they cluster with the other OTUs belonging to the South Sulawesi stock only after Wotu and Laiyolo (Muna-Buton stock) had joined the main cluster. All three Makassar languages,

[3] In the next two dendrograms, no space will be allowed between some lexicostatistically similar OTUs, so that the whole dendrogram can fit on the page.

especially MAK, are lexicostatistically aberrant within the South Sulawesi context. If we rigidly apply the Grimes' lexicostatistical classifications then their data would have the Makassar languages form their own stock, while the other peninsular languages including Wotu and Laiyolo would form another stock (Figure A-1). I give the latter the whimsical name "Remainder Stock" because I do not wish to imply a formal grouping, since in all likelihood Wotu and Laiyolo are incorporated only because of the lack of comparisons with the languages of Southeast Sulawesi (A.6). As regards the stock-level distinctiveness of the Makassar languages, Charles and Barbara Grimes were already aware of this implication but placed the Makassar languages in the South Sulawesi stock owing to a suite of qualitative linguistic similarities - the same point made by Sirk (1989) and others.

Joining the Makassar stem with the main cluster places it to the far right, beyond the Bugis. The fit of the seriation, 105,563/125,044 or 84.4%, is a slight improvement over the previous dendrogram, as a reflection of the great lexicostatistical dissimilarity between LYL-WOT-LEM-SEK and the Makassar languages.

DENDROGRAM A-11



At this stage it is worth pointing out two "sporadic convergences", one involving LYL and SEL, and the other involving WOT and BLW (respectively, 54% and 56% lexicostatistical similarity). LYL shows only a low lexicostatistical similarity with the non-SEL Makassar

languages, as does WOT with respect to BBN and CAM, resulting in the placement of LYL and WOT at the other extreme from BLW and SEL in all the seriated dendrograms which involve these OTUs. The extreme placement of these OTUs should not be regarded as a shortcoming of the seriated order, rather the concept of "sporadic convergence" usefully emphasises that certain individual similarities are anomalous within the context of overarching trends.

The sporadic convergences considered here would reflect lexical hybridisation resulting from the geographical proximity of SEL with LYL, and BLW with WOT and the other languages of the Luwuk Coastal Plain (see Figures 1-4 and A-1). Accordingly, the Selayar language might have resulted from a fairly recent expansion of a Makassar language (Konjong) on to Selayar island, and the Luwuk dialect of Bugis might represent a fairly recent expansion of Bugis into the Gulf of Bone.

To complete the exercise, standard Indonesian then joins the main cluster, to be followed by the five languages of the Central Sulawesi stock included by the Grimes in their large half-matrix. The order given by Grimes and Grimes (1987:19) is close to the best, with its goodness of fit calculable as 184,289/214,015 or 86.1%. I managed to find some minor rearrangements which decrease the stress of the seriation slightly, including those which allow a hierarchical dendrogram to be imposed on the seriation (which is not quite possible using the order given by the Grimes). The least stressful order I could find, shown in Figure A-3, scores a goodness of fit of 184,666/213,876 or 86.3%.

Some other plausible orders, which would be permissible within the framework of a hierarchical cluster, were tried with the following results:
RAM-PDO-TOP-SAR-PMN-LYL-WOT-LEM-SEK-MAJ-PBG-BAL-SEN-MAL-MMJ-SUM-TAP-SNY-PDG-BOT-A-T-MBI-RAN-BAM-MEH-RKB-RKA-TLA-MMS-SAD-KAL-PAT-MSR-CAM-BLW-BBN-SEL-KNJ-MAK-BI (i.e. an extension of what had been the best order before Indonesian and the Central Sulawesi languages joined): goodness of fit 182,502/213,082 or 85.6%;

BI-MAK-KNJ-SEL-BBN-BLW-CAM-PBG-MAJ-BAL-SEN-MAL-MMJ-SNY-PDG-SUM-TAP-BOT-MSR-PAT-A-T-MBI-MEH-BAM-RAN-MMS-SAD-RKB-RKA-TLA-KAL-SEK-LEM-WOT-LYL-PMN-SAR-TOP-PDO-RAM (i.e. the best order I could find which placed MSR and PAT centrally within the Northern South Sulawesi language family): goodness of fit 183,506/215,320 or 85.2%.

These involve radical rearrangements of the Northern South Sulawesi OTUs compared to the order shown in Figure A-3, yet the seriation's goodness of fit changes by 1.1% or less. All three seriations correctly strike the main axis of variation which, as shown in Figure A-2, places the block of primarily southern OTUs (Indonesian, Makassar, Bugis) at one extreme, the block of primarily northern OTUs (Seko, Lemolang, Wotu, Laiyolo, Central Sulawesi languages) at the other extreme, and the OTUs of the Northern South Sulawesi family in the middle. By recognising this main axis of north-to-south variation, and sensibly arranging the OTUs within the three main blocks, 85 to 86% of the variation in the original 40 by 40 half-matrix can be correctly represented in a suite of seriations. This point highlights the ubiquitous diffusion of lexical items across South Sulawesi's landscape, and suggests that the use of lexicostatistical data to classify South Sulawesi's speech communities into languages, subfamilies and so on primarily amounts to a useful pigeonholing procedure. Further, a hierarchical dendrogram such as Figure A-3 does not represent a linguistic family tree, and any attempt to reconstruct linguistic origins from any modern linguistic features alone would somehow have to account for the past diffusion of linguistic features.

A.5 Down-the-Line Seriation of the South Sulawesi Languages

As regards the hierarchical seriations discussed above, I have found guidelines to help determine the order of the branches, but no hard and fast rules, except the *post hoc* rule that the order should maximise goodness of fit. Appendix B, which deals with the tradewares, uses a rule-bound seriating procedure which can be called "down-the-line" seriation. The procedure is designed specifically for those clustering exercises where the relationships of a common stem

cannot be accurately predicted from the relationships which the member OTUs had before they clustered. In these cases the original half-matrix of coefficients need not be an accurate guide to how the OTUs should cluster, and so they do not provide a fair test of the validity of the seriated order which results. However, as explained above (A.3), lexicostatistical data are one of the few data sets where the correct hierarchical clustering can be deduced from the original coefficients with perfect accuracy. So it is of some methodological interest to investigate whether down-the-line seriation is applicable to lexicostatistical data.

We will first try out "nearest next-to-sister" seriation. Here the argument runs as follows: if hierarchical clustering always places sister stems adjacent (at their level in the dendrogram), then the stems should be twiddled such that the nearest next-to-sister stems are also adjacent. The seriation proceeds in the reverse order of the linkages, from the top of the page down (see Figure A-4). The last two clusters to join, the "Central Sulawesi stock" and "Other", are the first to split. The second-to-last join, Rampi with the other Central Sulawesi languages, becomes the second split in the seriation. At this point Rampi is lexicostatistically the less like "Other" languages, so the common stem of the non-Rampi Central Sulawesi languages is placed next to the "Other" common stem, and Rampi is placed at an extreme. The third-to-last join, Padoe (PDO) with Saruda-Topoiyo-Pamona (STP), becomes the third split. At this stage the clustering half-matrix was:

"Other"			
34.77	Padoe		
38.25	40.33	STP	
35.89	36	40	Rampi

Excluding the 40.33 value at which Padoe and STP joined, the highest lexicostatistical similarity is between STP and Rampi. Hence STP is placed next to Rampi as a reflection of their next-to-sister relationship.

With Padoe now fixed as the left-hand member of the Central Sulawesi languages, the next split, BI from "Other", is gauged with respect to Padoe. At this stage the relevant

lexicostatistical similarities were "Other":PDO = 34.8 and BI:PDO = 33. BI's lower similarity results in BI taking an extreme position to the left hand of the dendrogram (Figure A-4). Hence the next split, between the Makassar languages (MKL) and my "Remainder stock", is gauged with respect to the immediate outliers, BI and Padoe. At this stage the relevant half-matrix used during the clustering procedure was:

BI			
40.8	"Remainder"		
38.67	42.78	MKL	
33	35.13	31.67	Padoe

The "Remainder stock" is positioned next to BI, because they share the highest lexical similarity after that which had clustered MKL and the "Remainder stock", so the Makassar stem swings over towards Padoe. And so on, with the seriating decisions shown in small figures in Figure A-4.

The resulting order differs radically from those considered previously.[4] The comparison cited immediately above causes the whole block of Sulawesi languages not belonging to the Central Sulawesi stock, from Laiyolo to Makassar, to be reversed with respect to the outliers Indonesian (BI) and the Central Sulawesi languages. As shown by comparison with Figure A-2, this is a less satisfactory order, and the goodness of fit comes to 157,912/207,183 or only 76.2%.

Although "nearest next-to-sister" analysis produces a comparatively poor seriation, it does highlight two connected points which are disguised in the main seriations. Firstly, it is the "Remainder stock" rather than the "Makassar stock" which is lexicostatistically closer to Indonesian. Indeed, the "Remainder stock" relates more to Indonesian than to the Central Sulawesi languages despite the diffusion of lexical items between the "Remainder" and Central Sulawesi stocks.

[4] Note that the method would have produced the same result even if one of the Central Sulawesi languages other than Padoe had taken up the left-hand position. This is because the highest average lexicostatistical similarity which the "Remainder stock" shares with any Central Sulawesi language, 40.23% (with Saruda), falls below the 40.8% average lexicostatistical similarity recorded between Indonesian and the "Remainder stock".

Secondly, the aberrant lexicon of the Makassar languages within the South Sulawesi context cannot be attributed to borrowings from Malay (Indonesian), nor have any other specific relationships between Makassar and any non-South Sulawesi language been seriously proposed.

A variant which I tried on the above procedure involves seriating by the least stress within the immediate comparisons. For instance, in the comparison involving "Other", Padoe, STP and Rampi, some stress is clearly involved because Padoe (compared against STP) is the less like both of the outliers (i.e. "Other" and Rampi). We might argue that Padoe should be placed next to "Other", since its lexicostatistical similarity with "Other" is only 3.48% less than the comparable value of STP, as opposed to 4% when Padoe and STP are compared against Rampi. At this stage, and the stage where Indonesian hives off from "Other", the same order results as in the "nearest next-to-sister" seriation. But when we come to placing the Makassar and "Remainder" stocks with respect to Indonesian and Padoe, less stress (40.8 minus 38.67 or 1.13%) results from placing the Makassar stem next to Indonesian than from placing it next to Padoe (35.13 minus 31.67 or 3.46%). That is, the procedure correctly picks up the point that, although the Makassar languages are generally aberrant, they are less so with respect to Indonesian than with respect to the Central Sulawesi languages. However, lower in the dendrogram the procedure reverses the order of the PUS, Toraja-Sakdan and MSR-PAT blocks of OTUs compared to the best seriations (cf. Figures A-3 and A-5). This reversal lowers the goodness of fit slightly to 178,111/213,965 or 83.2%, even though the fit with the main axis of variation in Figure A-2 appears equally good or even slightly better.

Down-the-line seriation disconsiders relationships beyond the immediate comparisons and so, not surprisingly, fails to produce the best fitted order overall. By the same token it is more sensitive to immediate relationships. For instance, the within-ordering of the Central Sulawesi languages suggested by Figure A-2 is better represented by down-the-line seriation (Figures A-4 and A-5) than in the seriations which in effect order the Central Sulawesi

languages with respect to those further south (e.g. Figure A-3). However, one problem with down-the-line seriation is that OTUs which are anomalous within a cluster fall to the extremities of the cluster, even if they are generally anomalous rather than representive of a transition towards a sister cluster. For instance, Saruda or Pamona would appear to be a better candidate than Padoe as the Central Sulawesi language by which to seriate the other languages (see Figure A-2). This potential weakness is not counteracted at all by "nearest next-to-sister" seriation and so would appear to explain its poor performance. In contrast, "least neighbours' tension" seriation overrules the main anomalies and so appears to perform quite well.

A.6 Other Reviews of the Languages of South Sulawesi

The Summer Institute of Linguistics, which undertook the survey of South Sulawesi speech communities studied by Grimes and Grimes (1987), has since carried out further surveys. Based on the results, Friberg and Laskowske (1989:2ff.) suggested various revisions to the Grimes' distribution map and classification of South Sulawesi's languages, principally involving the fringes of the Northern South Sulawesi family. However, since Friberg and Laskowske substantiate their argument with only a selection of immediate lexicostatistical comparisons, which falls far short of the 40-by-40 half-matrix provided by Grimes and Grimes (1987:19), their suggestions cannot be rigorously tested here.

For instance Ulumanduk, which the Grimes could not include in their study, proved to be lexicostatistically (and geographically) central within the Mandar Peninsula speech communities, and apparently also extensive. Thus Friberg and Laskowske reclassify Tappalang (TAP), Botteng (BOT) and Sondoang as dialects of Ulumanduk rather than dialects of Mamuju, and find that much of the northern Mandar-speaking area (in the Grimes' study) is really Ulumanduk territory. Ulumanduk belongs to a Pitu Uluna Salu subfamily, so compared to the situation shown in Figures A-1 to A-5, PUS would cut in between Mandar and Mamuju (Friberg and Laskowske, 1989). This significant rearrangement may be correct, because the

relations shown between groups depend on which particular OTUs are sampled, but Friberg and Laskowske nowhere provide a half-matrix comparing all the Mandar, Mamuju and PUS OTUs whereby the breakup into particular groups could be formally demonstrated. Their point that Mamasa, PUS, Ulumanduk, Mamuju and Mandar form one linguistic continuum, on account of the ubiquity of transitional dialects (Friberg and Laskowske, 1989:6), would certainly be supported by the present study (e.g. Figure A-3), or even extended to include almost the entire Northern South Sulawesi family (cf. Friberg and Laskowske, 1989:8-9).

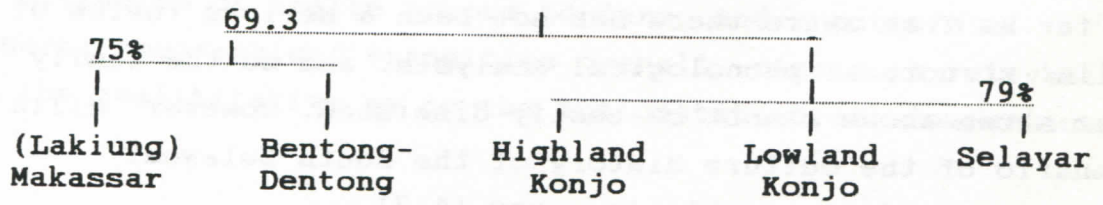
Friberg and Laskowske (1989:8-9) further propose the following minor amendments to the Northern South Sulawesi family: reclassification of Pattaek as a dialect of Mamasa; recognition of Rongkong and Luwuk (Toalak) as dialects of the same language rather than as separate languages; and relabelling the Grimes' three dialects of the Masenrempulu language - Enrekang, Maiwa and Duri - as the three languages of the Masenrempulu subfamily (while the Grimes' other Masenrempulu language, Pattinjo, becomes a dialect of Enrekang). Friberg and Laskowske's second point is strictly correct (cf. Dendrogram A-5); they may also have enjoyed the advantage of sounder information in proposing their first and third points, but they do not provide any supporting data.

As regards the Muna-Buton stock, specifically the Grimes' finding of a 53% lexicostatistical similarity between the Barang-Barang dialect of Laiyolo (LYL) and Wotu, Friberg and Laskowske (1989:13) relate a Barang-Barang folk history in which a Barang-Barang prince sailed away to settle in Wotu. If this passage is meant to imply a relationship of sporadic convergence between LYL and Wotu, it would do so misleadingly, because LYL shows comparable lexicostatistical similarities (38-50%) with the Central Sulawesi languages sampled by Grimes and Grimes (1987:19), similarities which are consistently higher than those involving Wotu. Further, the relationships extend to Sulawesi's southeast tip given the 58% lexicostatistical similarity between LYL and Buton (Grimes and Grimes, 1987:63). Sirk (1988:11) confirms the last point to the extent of showing that Wolio, a dialect of

Buton, shares a 76% similarity with LYL and a 56% similarity with Wotu, as calculated from a 100-item Swadesh word list. In the terminology of Grimes and Grimes, Wotu, Laiyolo and Buton would comprise a family, one which would belong in the same stock with Muna based on the lexicostatistical comparisons given by Sirk (1988:11). In his study of Muna, Van den Berg (1989:8) also accepts a Muna-Buton grouping and points out the need to investigate its wider relationships with the languages of Central and Southeast Sulawesi.

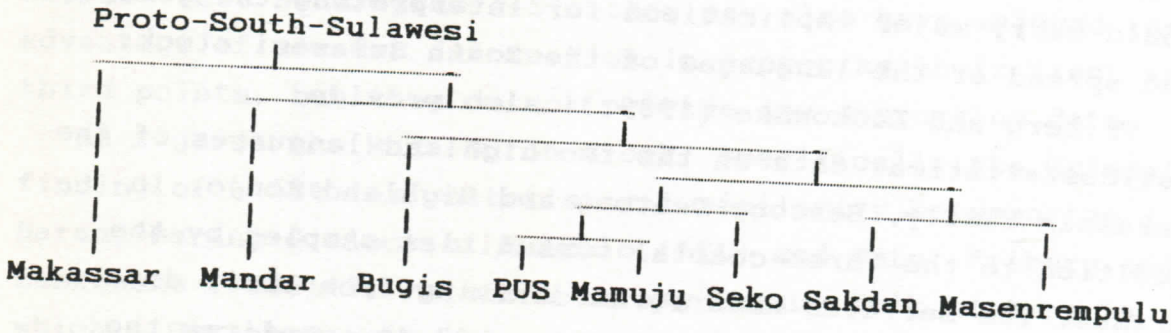
The above paragraph might be interpreted to suggest that LYL, and the related Kalao language on Kalao island near Selayar, represent a linguistic continuum once widespread in South Sulawesi, stretching from Selayar across the Gulf of Bone, and then further east to Sulawesi's southeastern tip. In this scenario the South Sulawesi segment of the continuum has been interrupted by the impingement of the Makassar, Bugis and Toraja-Sakdan languages. As another possible interpretation, LYL and Kalao could have derived from the coasts of the Gulf of Bone or Sulawesi's southeast tip, or at least have been heavily influenced by thoroughgoing contact from these regions. Choosing between these interpretations could carry major implications for interpreting the genesis and spread of the languages of the South Sulawesi stock.

Friberg and Laskowske (1989:3) also provide lexicostatistical data on the two highland languages of the Makassar family, Bentong-Dentong and Highland Konjo, in addition to the three coastal communities sampled by the Grimes. The seriated dendrogram deriving from their data, with a goodness of fit of 147/151 or 97.4%, confirms the smooth, west-to-east lexicostatistical transition already suggested by the Grimes' smaller data set.



The linguists working with the Summer Institute of Linguistics' programme in South Sulawesi have desisted from

interpreting their classificatory system as a family tree. This is not the case with Mills (1975a, 1975b) whose goals were to reconstruct Proto-South-Sulawesi through a structural phonological analysis and to suggest a family tree for the South Sulawesi languages. Two difficulties with Mills' work, noted by Sirk (1989:55-56), are the poor quality of his word lists (compiled either from dictionaries or largely by informants familiar with the orthography of Bahasa Indonesia), and the numerous exceptions to his reconstruction of Proto-South-Sulawesi which Mills was unable to explain. A third difficulty is that Mills (1975a:497-498) claims to have considered his lexicostatistical evidence in reconstructing the two family trees he offers - yet not only is his evidence very incomplete, based on only nine OTUs (Mills, 1975a:490, 492), but also the implications support the Grimes' scheme (and hence Figure A-3) more than they support either family tree offered by Mills (cf. Grimes and Grimes, 1989:202-203). Given that his lexicostatistical efforts are demonstrably inadequate, his more critical family tree is the one which makes less recourse to lexicostatistics. It can be shown as follows:[5]



Mills' structural phonological analysis confirms the aberrance of Makassar but suggests that Mandar, rather than Bugis, was the next linguistic taxon to leave the main stem. As far as I am aware there has not been a serious review of Mills' structural phonological analysis, and so the family tree shown above cannot be easily dismissed. However, Mills' scenario of the culture history of the South Sulawesi languages is demonstrably in error (A.7).

[5] Note that the only difference between his family trees involves the position of Bugis and Mandar.

A very interesting and possibly crucial development within recent years is Adelaar's demonstration that the Tamanic languages, which are located almost centrally in the island of Kalimantan, also belong to the South Sulawesi group. Adelaar notes that the phonological evidence tends to relate Tamanic specifically with Bugis, even if the frequencies of shared lexical innovations would slightly favour a specific relationship between Tamanic and Toraja-Sakdan.[6] The nature of the phonological similarities and the geographical isolation of the Tamanic languages rule out the possibility of similarities through borrowing. Hence the possibilities boil down to whether Tamanic could have resulted from an ancient migration from South Sulawesi to Kalimantan, whether the South Sulawesi languages (excluding Tamanic) could have resulted from an ancient migration from Kalimantan, or whether there was some other, as yet unidentified homeland. Of these, the first scenario seems the most feasible (Adelaar, in press).

Finally, while Sirk (1989) concludes that the qualititative data support the concept of a South Sulawesi language group as conventionally understood to include Makassar, he also notes the existence of contrary trends. For instance the manner of forming the numeral "four hundred", and a few lexical roots, distinguish the non-Makassar languages of the South Sulawesi group from Makassar. Accentuation patterns, and the lowering of high vowels before *-q, are areal features also shown by the Badaic languages of Central Sulawesi province, while the lexical item *beluak unites the Central Sulawesi languages specifically with the non-Makassar members of the South Sulawesi language group (Sirk, 1989:74-75). So while different borders are suggested by the qualitative and the lexicostatistical data, the issue of diffusion of traits across borders, and a pattern of general geographical transition overall, remain as relevant to the qualititative as to the lexicostatistical comparisons.

[6] The specific relationships of Tamanic with both Bugis and Toraja-Sakdan recall Mills' suggestion (1975a:498) that Bugis must have passed through a Sakdan- or Masenrempulu-like stage.

A.7 A Tentative Overview

According to the dominant paradigm in island Southeast Asia's later Holocene prehistory, the colonisation of the archipelago by Austronesian speakers is related to the appearance of a Neolithic assemblage - especially pottery - in the archaeological record (Bellwood, 1985, 1988; Spriggs, 1989). Recalibration of the radiocarbon dates which Glover (1978) associates with a ceramic Toalean phase at Ulu Leang 1 would date the arrival of Austronesian speakers at Leang-Leang, Maros, to approximately 5000 years ago (Spriggs, 1989). However, my understanding of the Leang-Leang sites would date the initial appearance of pottery to less than 3500 years ago (Bulbeck, in prep. a). This would seem to be rather late given the overall chronology, as currently understood, of the expansion of Austronesian within the Indo-Malaysian archipelago (e.g. Blust, 1988). It would imply a back-migration into South Sulawesi from further west, as assumed by Mills (1975a), and possibly supported by the lexicostatistical data (especially Figure A-4). Given the demonstration by SSPHAP's survey of Toalean sites in open (non-cave) contexts (e.g. Chapters 7 and 10), and the assumption that the "Toaleans" were hunter-gatherers who did not speak an Austronesian language, it is possible that the Toaleans had occupied the South Sulawesi peninsula at sufficiently high densities to resist substantial Austronesian colonisation for centuries or millennia.

A point of almost universal agreement is the relatively great distinctiveness of the Makassar languages from those surrounding it. Probably this does not merely reflect the comparative geographical isolation of the Makassar languages, nor the possible replacement of former intervening languages owing to the expansion of Bugis and Makassar. It probably also implies that "proto-Makassar" was the first language to split off the main line of descent of the South Sulawesi languages (e.g. Mills' family tree depicted above). To help explain the lexical aberrance of Makassar, Mills (1975a:513-514) suggests that Makassar may have mixed with an indigenous substratum related to the Central Sulawesi languages. However, as shown above, the Makassar languages

are lexically aberrant not only with respect to the other languages belonging to the South Sulawesi group, but also the Central Sulawesi languages, Wotu and Laiyolo (apart from the sporadic convergence between Selayar and Laiyolo), and even standard Indonesian. If the lexical aberrance of Makassar is to be accounted for by the inclusion of a substratum, on current evidence that substratum could have been provided by the original, "Toalean" inhabitants.

Leaving aside that speculation for other speculations, we can note that the seriated dendrograms of the languages of the "South Sulawesi stock" read like a succession of splits involving Makassar first, then Lemolang, then Bugis, then Seko, then the languages of the Northern South Sulawesi family (Figures A-3 to A-5). Given that Lemolang and Seko are so clearly transitional towards other "stocks", we can ignore them and simplify the structure to a successive three-step ramification - Makassar, then Bugis, then the Northern South Sulawesi family. Interpreting this structure as a family tree - it is in essence the family tree of Mills (1975a) which he did not prefer - strikes an appealing correspondence with South Sulawesi's rural population statistics which show the highest densities in the Makassar strongholds, intermediate densities in the Bugis strongholds, and the lowest densities to the north (Figure 1-3). The comparison lends some support to the model of a Proto-South-Sulawesi "homeland" on or near the peninsula's south coast, followed by a northwards colonisation as population densities built up.

However, we have Mills' preferred family tree, based on his structural phonological analysis, where Mandar, not Bugis, is the next stem after Makassar to split off the common line of descent. In this case our scenario would have to be adjusted to show an early branch colonising the Mandar coastline prior to the general northwards thrust during which the Bugis and Northern South Sulawesi (barring Mandar) stems divided. We also have Adelaar's reconstruction of a specific link between Tamanic and Bugis (and possibly Toraja-Sakdan), suggesting a migration to Kalimantan from the Bugis heartland after the Mandar stem had split off. This scenario would indicate that a major linguistic expansion, resulting not

only in the most numerous and geographically extensive of the South Sulawesi languages (Bugis), and not only in the majority of these languages within South Sulawesi and the colonisation of most of the province, but also in a migration to central Kalimantan, had occurred in comparatively recent times. Be that as it may, the presence of Austronesian languages in South Sulawesi since at least 3000 years ago, and our virtual ignorance of events until the last 800 years or so, obviously leave ample scope to postulate a range of possible scenarios.

Nonetheless at least one scenario, Mills', can be rejected. Spurred on by his structural phonological analysis which suggested that the Mandar stem had split off the main line of descent before the Bugis stem, Mills posited that the Proto-South-Sulawesi homeland was close to the Mandar area, and more specifically at the mouth of the Sungai Sakdang. He hypothesised that the prospects for trade might have drawn these Proto-South-Sulawesi speakers to an area already inhabited by people speaking languages related to the Central Sulawesi languages. To get the South Sulawesi languages to their present-day positions, Mills developed an elaborate migratory scenario. Firstly, proto-Makassar split off and colonised the Sakdan valley, then migrated to the Luwuk Coastal Plain and thence to the main body of the peninsula, finally pitching a last stand on the peninsula's south coast - pursued all the time by the proto-Bugis. Before proto-Bugis split off, proto-Mandar had split off and migrated westwards. After proto-Bugis split off, the PUS, Mamuju, Seko, Toraja-Sakdan and Masenrempulu language stems ramified and colonised the Central Highlands and Mamuju area (Mills, 1975a, 1975b).

This scenario seems to be unnecessarily defiant of the principle of parsimony (e.g. Blust, 1988:47), involving as it does long and complex migrations by proto-Makassar and proto-Bugis, and a succession of language replacements. In addition it is difficult to see how the mouth of the Sungai Sakdang could have held such extraordinary benefits for population growth that it remained the static centre for four successive migratory waves. Moreover Mills (1975a:500) makes

the extraordinary claim that South Sulawesi lacks suitable harbours south of the Sakdang delta (see Allied Geographical Section [1945] for example). He also invokes an inland sea separating the southwest peninsula from the Central Highlands during the period in question (Mills, 1975a:502-503,533-534), which is now known to have existed only between 7100 and 2600 years ago, and even then extended only from the mouth of the Sungai Cenrana to the ridge of Celebes molasse immediately east of the Tempe Depression (Gremmen, 1990). Then there is Mills' strange argument that the Central Highlands would have been cooler and therefore more densely populated than the southwest peninsula (Mills, 1975a:500-501), even though pleasantness of climate to a westerner is not usually invoked by human geographers as a factor influencing carrying capacity. Mills' argument further involves the presence of a few Makassar toponyms in the area now inhabited by the Bugis, explaining away the Makassar toponym of Majene which does not fit his scenario (Mills, 1975a:502,534), and ignoring the comparable presence of Bugis toponyms (e.g. Sanrabone, Bone-Bone, Pattiro - see Chapters 6 to 10) in the area now inhabited by the Makassar. Finally there is Mills' fanciful conflation of isolated statements in the Bugis chronicles and local origin myths into an undated scenario - at least one of these folk stories, Luwuk's reputation as the oldest Bugis kingdom, is almost certainly false (13.4.2).

If we accept Mills' family tree based on structural phonological analysis, the most parsimonious scenario would posit a Proto-South-Sulawesi homeland on or near the peninsula's south coast, and a general north-to-south colonisation thrust preceded by Mandar as a precocious offshoot. The other main scenarios would also suggest a homeland within the southwest peninsula. From the point of view of the present thesis, the fugitive status which Mills would ascribe to the Bugis and especially the Makassar can be safely ignored. The linguistic line leading to modern Bugis would appear to have been based in the present Bugis heartland for several thousand years, and the counterpart propositions for the Makassar is even more certain.

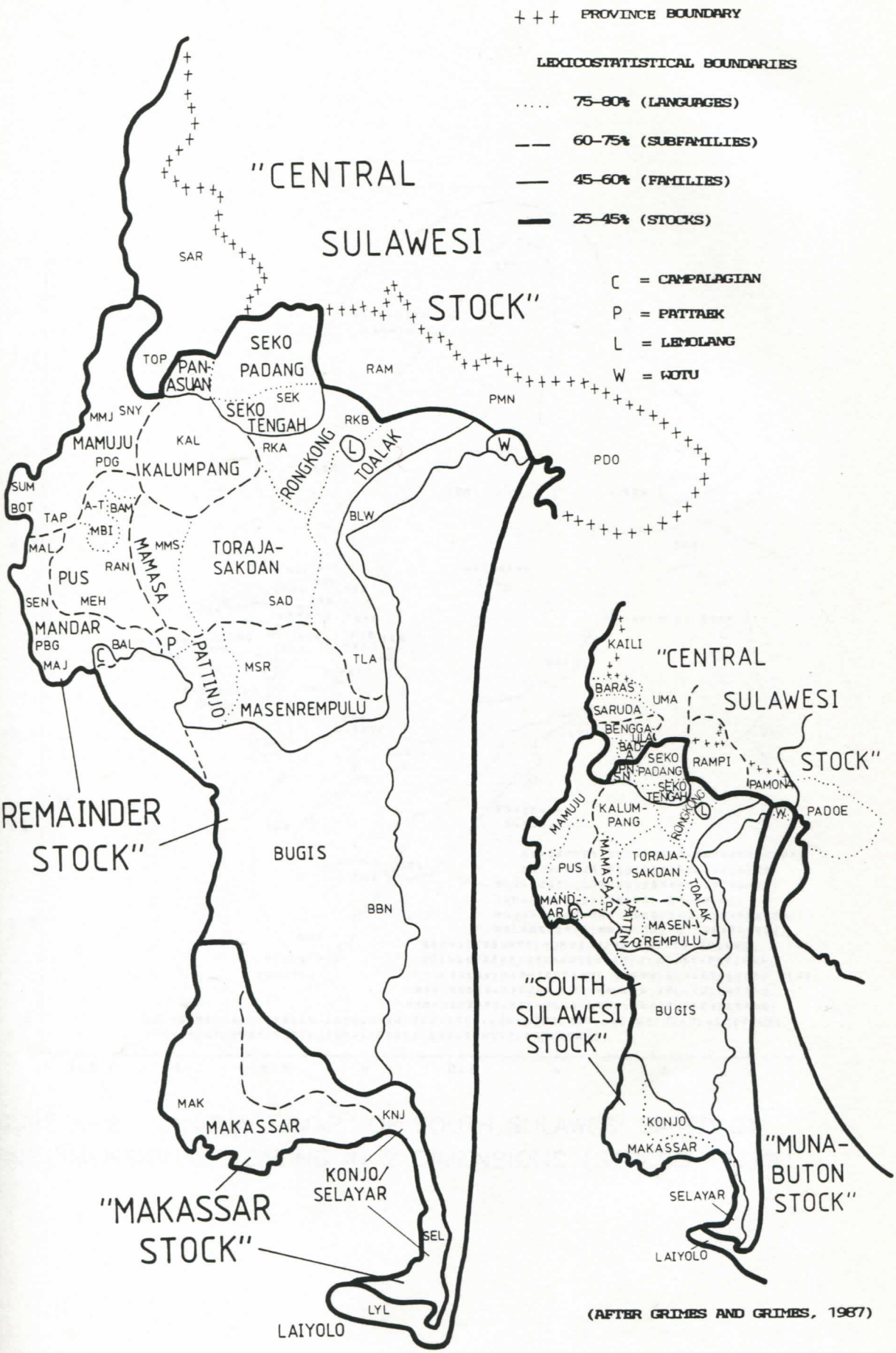


FIGURE A-1. FORMAL LEXICOSTATISTICAL RELATIONSHIPS OF SOUTH SULAWESI LANGUAGES

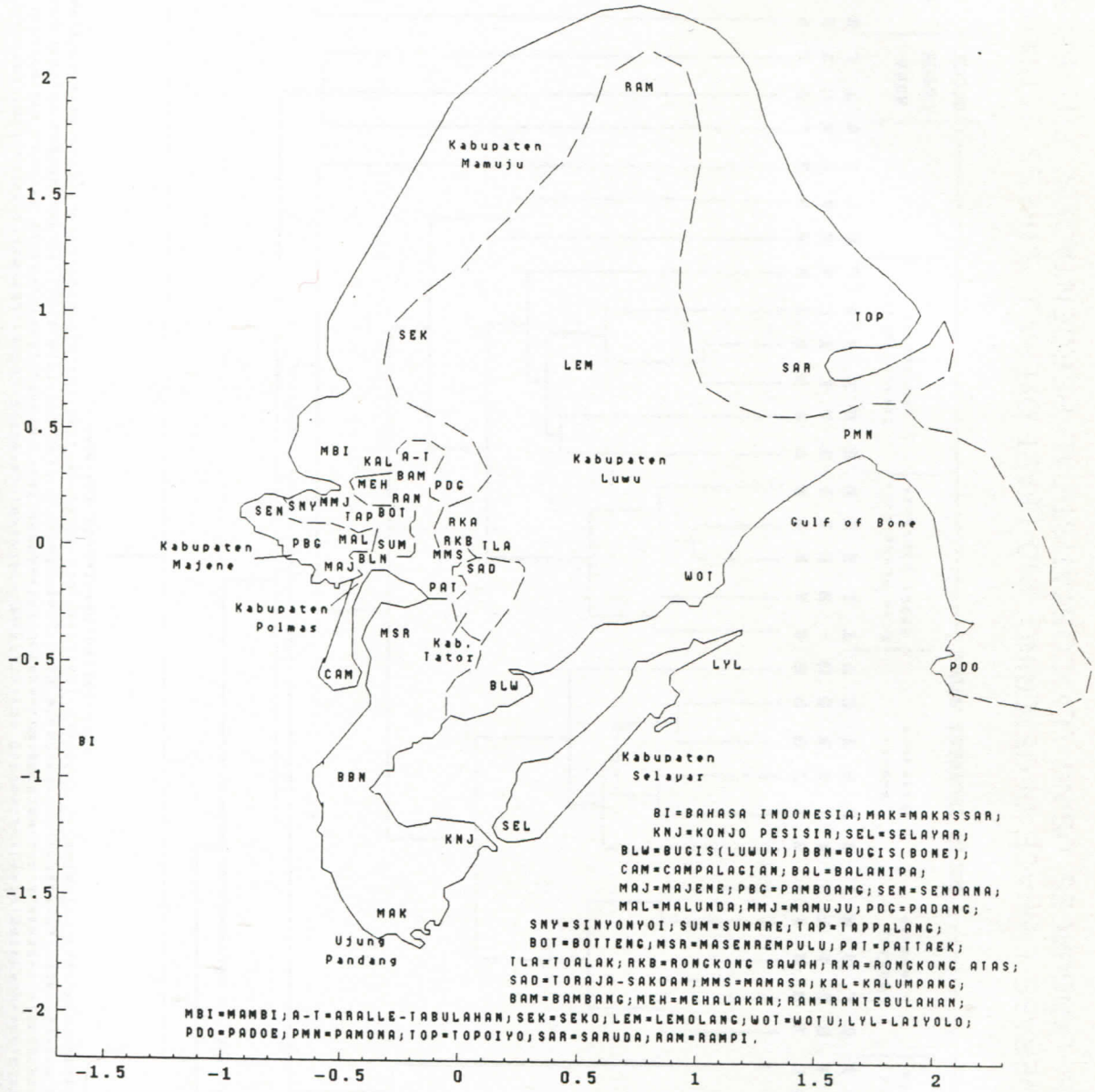


FIGURE A-2. "LANGUAGE MAP" OF SOUTH SULAWESI BASED ON MULTIDIMENSIONAL SCALING IN 2 DIMENSIONS (STRESS 18.2%)

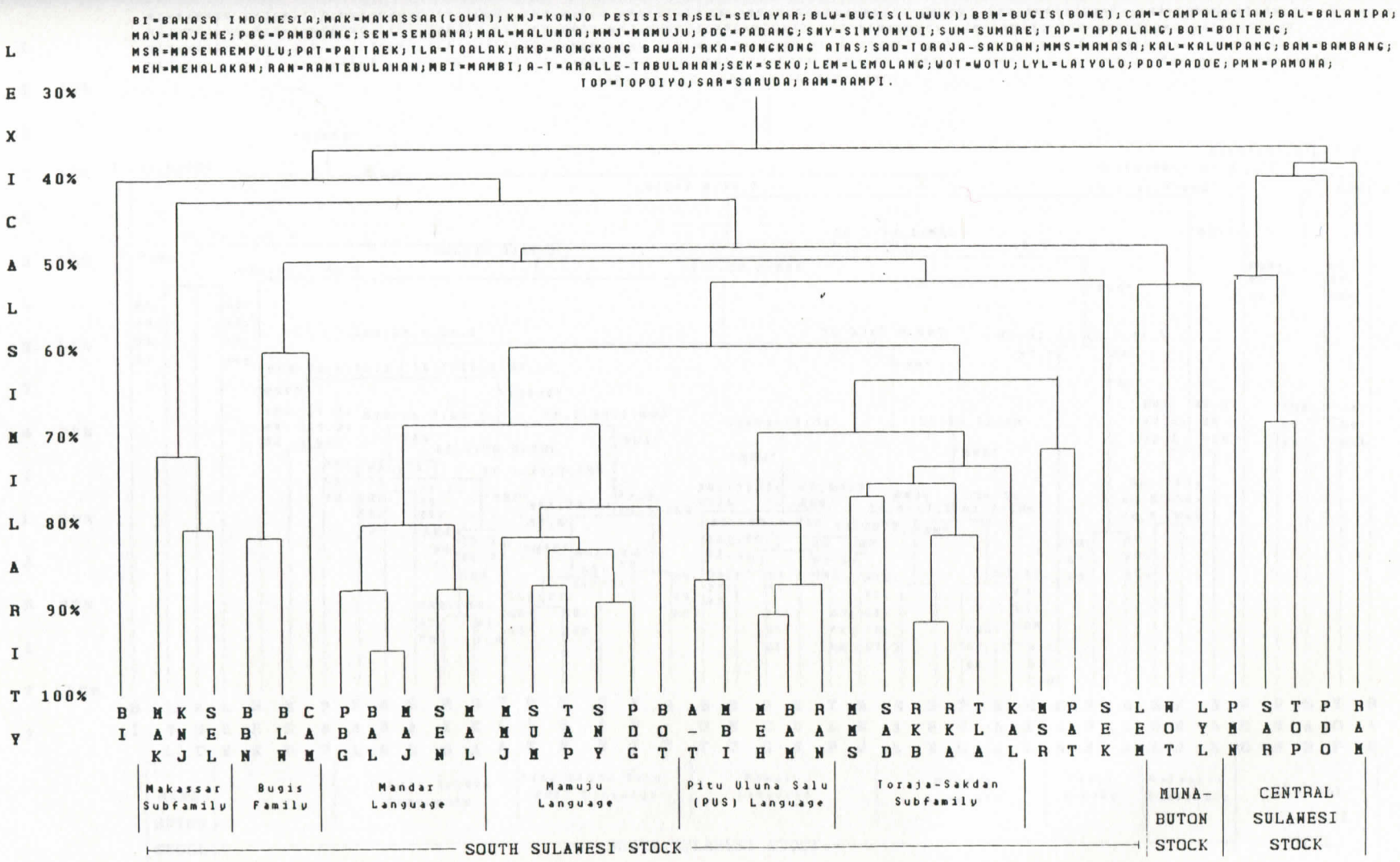


FIGURE A-3. AVERAGE LINKAGE CLUSTERING AND LEAST OVERALL STRESS SERIATION OF SOUTH SULAWESI LANGUAGES USING LEXICOSTATISTICAL PERCENTAGES. FIT: 86.3%.

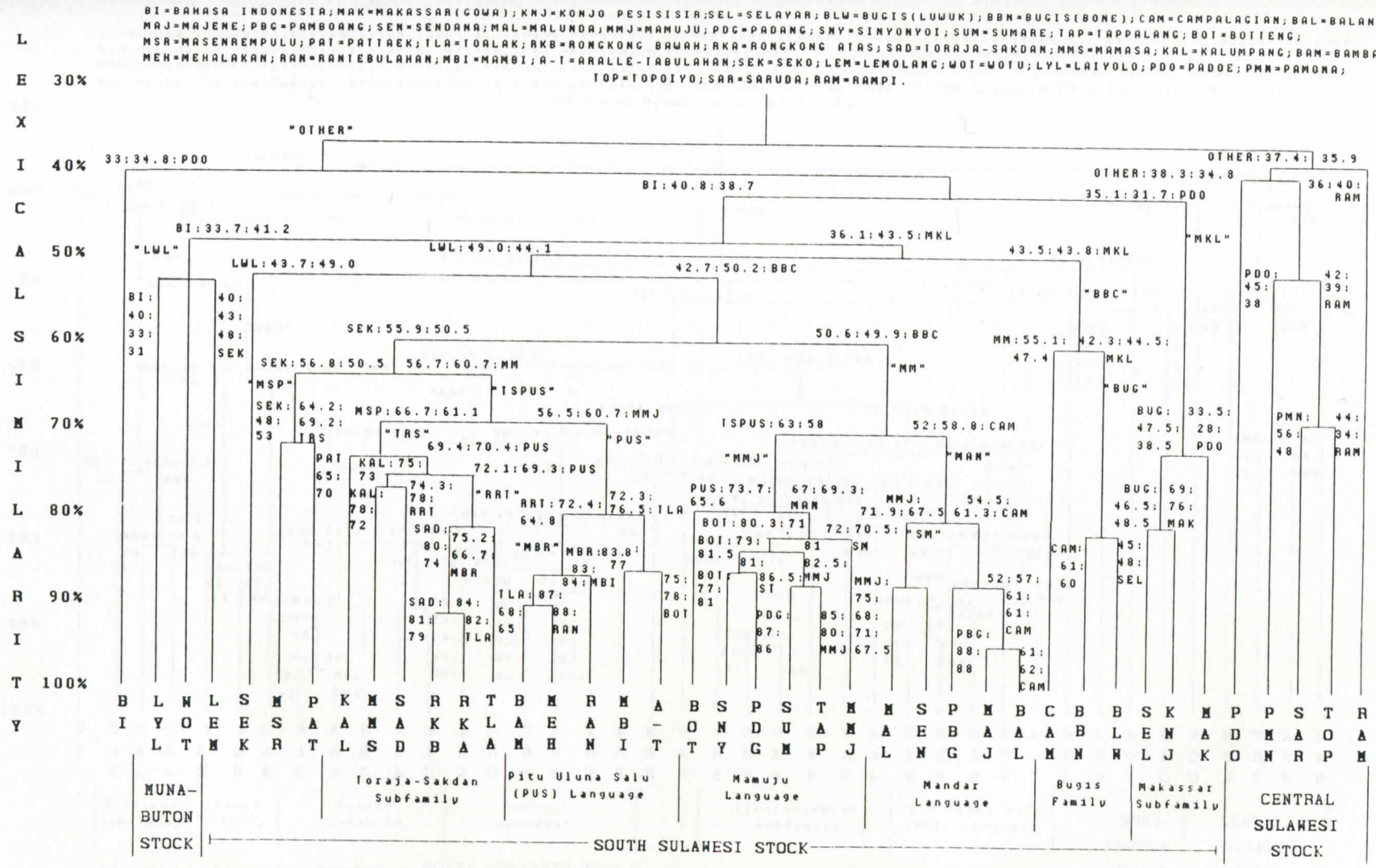


FIGURE A-4. AVERAGE LINKAGE CLUSTERING AND NEAREST NEXT-TO-SISTER SERIATION OF SOUTH SULAWESI LANGUAGES USING LEXICOSTATISTICAL PERCENTAGES. FIT: 76.2%.

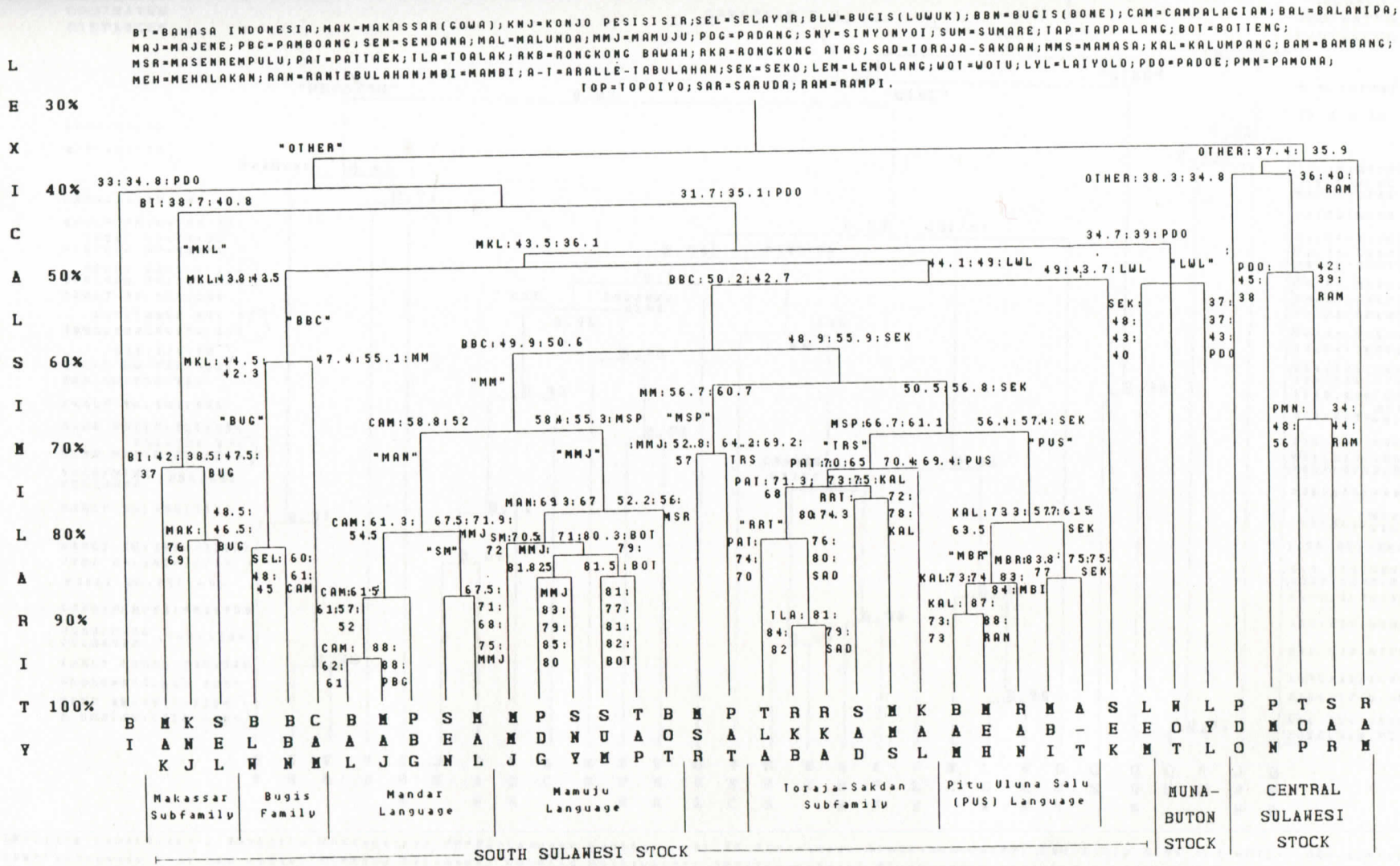


FIGURE A-5. AVERAGE LINKAGE CLUSTERING AND LEAST NEIGHBOURS' TENSION SERIATION OF SOUTH SULAWESI LANGUAGES USING LEXICOSTATISTICAL PERCENTAGES. FIT: 83.2%.