

LEXICOSTATISTIC AND SOCIOLINGUISTIC SURVEY OF BALANTAK AND ANDIO

Robert L. Busenitz

Cooperative Program of Hasanuddin University
and
The Summer Institute of Linguistics

0. INTRODUCTION¹

The UNHAS-SIL Cooperative Program initiated work in Central Sulawesi with a survey of the languages of this province; this resulted in the publication in 1979 of *Languages of Central Sulawesi* by Donald F. Barr, Sharon G. Barr, and C. Salombe. This survey used the Swadesh 100-item wordlist to compare lexical similarity among twenty-two of some twenty-four languages of Central Sulawesi. An appendix of the publication contains the wordlists used in that study, and includes single wordlists for Balantak, Andio, and Saluan, three languages spoken in the Banggai *kabupaten* 'district'. Based on their lexical similarity, these three languages are classified as belonging to the Saluan subgroup of the eastern group of Central Sulawesi languages (1979:22f).

After beginning a field program and an indepth study of the Balantak language in the village of Dolom, 'Balantak *kecamatan*' 'subdistrict', we saw the need to investigate further the dialect variation within the Balantak language as spoken in Balantak and Lamala, the two subdistricts which are the homeland of the Balantak people, as well as its relationship to the neighboring languages of Andio and Saluan. We also wanted to know more about language use and language attitudes throughout the region and the implications for vernacular literature. This paper reports some of our initial efforts to investigate these areas.

The Andio language area, like a beachhead on the coastline of the Lamala subdistrict, is located near some fertile rice plains and surrounded by Balantak-speaking people. Though small in size, relatively speaking, some 1700 speakers in the two villages of Tangeban and Tauge' preserve and propagate this language. No dialect variation was reported and only one wordlist was elicited in Tangeban.

Data were also obtained from four locations in the eastern and central Saluan language area, the area nearest the Balantak and Andio language areas. This information in particular is preliminary and incomplete for the Saluan language. No data have yet been obtained from the western Saluan

and Kahumamahon areas. Nevertheless, these four Saluan wordlists are compared with Balantak and Andio and give us further indication of lexical similarity of these two languages with Saluan. We do not, however, include Saluan in the discussion of sociolinguistic factors.

Eight different Balantak villages representing the breadth of the Balantak language area were selected and visited during the survey (cf. map, Appendix I). Tangeban represented Andio. The four Saluan language datapoints are in the subdistricts of Kintom, Luwuk, and Pagimana, three of the five subdistricts which make up the Saluan language area. All thirteen villages are accessible by road, and were visited during August, 1988.

1. METHODOLOGY

The Sulawesi Umbrella Wordlist (SUW) with 488 items was used for a lexicostatistic survey.² All wordlists except one, Bahingan, were recorded by the author.³ In most instances, two or more people served as informants; if a larger group was present, one or two people usually served as spokesmen. The language of elicitation was Indonesian.

Of the 488 items, sixty were disqualified for comparison purposes in this study. The items disqualified can be grouped into several categories and are further analyzed in Appendix II.⁴ This left 428 items eligible for comparison in all thirteen wordlists.

Of the disqualified items, six were on the Swadesh 100-item wordlist (S100), a subset of SUW. These are listed separately in Appendix III. This left ninety-four items of S100 eligible for comparison. A comparison of the percentage of lexical similarity as based on both SUW and S100 is made and discussed later in this paper. The results are also briefly compared to Barr, Barr, and Salombe, 1979 (BBS).

In order to compare the thirteen wordlists, all responses from the thirteen locations for each item on the umbrella wordlist were listed together. This procedure (discussed in Sanders 1977:36-37) greatly facilitates forming lexical similarity sets and also results in more consistent decisions about similar and dissimilar items.

The inspection method (Sanders 1977:33-34) was used to determine lexical similarity sets. Words which are phonetically similar and have the same meaning generally cover the same semantic domain and are considered to be lexically similar. We follow here Grimes and Grimes' suggestion (1987:9) that the term 'cognate' be reserved for items that can be shown by the comparative method to have a similar origin and follow regular sound changes. In most instances, lexically similar words should also prove to be cognate; however, for this study no attempt has been made to determine and eliminate borrowings.

Where synonyms were elicited, they were considered fully similar if only one of the synonyms was lexically similar.

To compare lexical similarity sets and produce a matrix of percentage figures between different wordlists, the computer program described in *Penyelidikan Persamaan Bahasa Dengan Menggunakan Mikrokomputer* by Edgar W. Travis was used. After the 428 items from SUW had been compared, the lexical similarity sets for the ninety-four items from S100 were extracted and recomputed to make a second matrix based on those items alone.

The degree of similarity between any two wordlists is expressed as a percentage figure. The degree of reliability for this figure or, in other words, the range of error that should be associated with this figure will depend on at least two things:

- 1) the reliability of the wordlist in giving words with the same meaning as the words from the wordlist with which it is being compared (This has to do with eliciting the wordlist and the skill and familiarity of the investigator with the languages.) and
- 2) the manner and consistency of the analyst in determining which words are similar and which are not (This has to do with determining lexical similarity sets--among other things, whether the analyst/inspector is a 'joiner' or 'splitter'.).

The greater the range of error associated with a percentage figure, the less useful that particular figure will be in comparing with other percentages.⁵

The two factors mentioned above have always created potential for decreasing the reliability of lexicostatistics.⁶ However, the potential is minimized when the same linguist performs all of either or both of the tasks; the probability of internal consistency, at least, is much greater. Except for one wordlist mentioned previously, the author elicited all of the data for this study and determined all of the lexical similarity sets. No attempt is made in this report, however, to give a range of error figure for the percentages of lexical similarity between two wordlists.

A sociolinguistic questionnaire was the primary instrument used to investigate language use and language attitudes. The questionnaire in various forms has been used by UNHAS-SIL for sociolinguistic surveys in South Sulawesi and covers areas such as livelihood and commercial orientation, education, social interaction, and the domain of various languages in a multilingual society. Usually the same people who assisted with the wordlists also answered these questions. Responses throughout the region were remarkably similar; these results are also discussed below.

2. LEXICOSTATISTIC RESULTS

2.1 Sulawesi Umbrella Wordlist

Table 1 is a matrix showing the percentage of lexical similarity based on SUW between all thirteen locations visited; the similarity with Indonesian is also included. Table 2 summarizes the highest and lowest percentages of any of the eight Balantak dialects with the other languages/dialects, including other Balantak dialects; Table 3 does the same for the four Saluan dialects.

LANGUAGE	VILLAGE												
Balantak	Tokuu												
	"	97	Kota										
	"	95	94	Dolom									
	"	93	92	96	Tombos								
	"	91	90	96	95	Sulu'bombong							
	"	89	88	91	91	94	Sobol						
	"	89	88	89	88	89	93	Poro'an					
	"	89	89	93	92	94	95	91	Eteng				
Andio	66	65	66	65	65	67	70	66	Tangeban				
Saluan	49	48	48	47	48	48	49	47	57	Lumpo'nyo			
	"	51	50	50	49	50	50	52	49	60	89	Sampaka'	
	"	49	48	48	47	48	48	49	47	57	84	90	Bahingan
	"	51	49	49	48	49	50	51	49	60	88	94	89 Kintom
Indonesian	24	25	24	23	24	24	24	23	25	26	25	26	26

Table 1
Percentage of lexical similarity based on SUW

	Balantak	Andio	Saluan	Indonesian
Highest	97	70	52	25
Lowest	88	65	47	23

Table 2
Lexical similarity of BALANTAK dialects
with other dialects/languages based on SUW

	Saluan	Andio	Balantak	Indonesian
Highest	94	60	52	26
Lowest	84	57	47	25

Table 3
Lexical similarity of SALUAN dialects
with other dialects/languages based on SUW

We can make several observations from this data:

- 1) Lexical similarity among all Balantak dialects varies from 88-97% (3-12 percentage points variation); from the perspective of lexical similarity and the commonly held 80% threshold, these speech communities representing the breadth of the Balantak language area should all be viewed as dialects of one Balantak language.
- 2) We see some expected chaining among the 'Balantak dialects. Dolom figures with respect to other Balantak points seem higher than expected in some instances, but may reflect a greater number of synonyms. If we follow the coast, Poro'an lies between Sobol and Eteng, but Sobol and Eteng have greater lexical similarity than either has to Poro'an. On the other hand, Eteng is geographically closest to Andio, but Poro'an stands several percentage points above all Balantak dialects in similarity with Andio.
- 3) Andio in this data is lexically more similar to Balantak (5-13 points higher) than to the Saluan dialects listed. It is certainly closer geographically to Balantak and these figures may also reflect significant borrowing, but this needs further investigation.⁷
- 4) Again, from the perspective of lexical similarity, the four speech communities in the Saluan area are dialects of one Saluan language, though variation is slightly greater than within Balantak, from 6-16 points. However, these data do not cover the full scope of the Saluan area, so the results are tentative.
- 5) We would not expect the Saluan dialects of Sampaka' and Kintom, which are the greatest distance from each other, to have the highest percentage of lexical similarity.
- 6) From this data, Balantak and Saluan only share a lexical similarity of around 50%.
- 7) All three languages of the subgroup are different only three percentage points or less in their lexical similarity with Indonesian.

2.2 Swadesh 100-item wordlist

Table 4 is similar to Table 1 except that the percentages are based on ninety-four items from S100. These ninety-four items are a subset of the 428 items used from SUW. In Tables 5 and 6 are summarized the highest and lowest percentages for Balantak and Saluan, respectively, as in Tables 2 and 3.

LANGUAGE	VILLAGE												
Balantak	Tokuu												
"	99	Kota											
"	93	94	Dolom										
"	91	93	99	Tombos									
"	93	94	100	99	Sulu'bombong								
"	91	93	98	97	99	Sobol							
"	93	95	93	91	94	94	Poro'an						
"	91	93	98	97	99	98	93	Eteng					
Andio	66	65	63	61	63	63	68	62	Tangeban				
Saluan	48	47	48	48	48	48	48	47	59	Lumpo'nyo			
"	52	51	50	50	50	50	52	49	63	90	Sampaka'		
"	52	51	51	51	51	51	52	50	63	86	98	Bahingan	
"	54	53	51	51	51	51	54	50	65	88	99	97	Kintom
Indonesian	33	32	31	31	31	32	30	31	33	34	35	36	37

Table 4
Percentage of lexical similarity based on S100

	Balantak	Andio	Saluan	Indonesian
Highest	100	68	54	33
Lowest	91	61	47	30

Table 5
Lexical similarity of BALANTAK dialects
with other dialects/languages based on S100

	Saluan	Andio	Balantak	Indonesian
Highest	99	65	54	37
Lowest	86	59	47	34

Table 6
Lexical similarity of SALUAN dialects
with other dialects/languages based on S100

We can make some additional observations and raise some questions by comparing these tables with those based on the larger wordlist:

- 1) NonIndonesian comparisons except Balantak-Andio are 0-5% higher, while those with Indonesian are 7-11% higher.⁸

- 2) Why are Balantak-Andio comparisons for S100 2-4% lower instead of higher like the rest? Why does exclusion of noncore vocabulary show Saluan to be closer to the same lexical relationship with Andio than Balantak is, or, conversely, why does exclusion of core vocabulary show Andio to be significantly closer to Balantak than Saluan?⁹
- 3) Now the variation of lexical similarity among the three languages with Indonesian is as great as seven percentage points.

Whatever we conclude for 1) and 2), it is clear that, in this case at least, lexical items are not random in their probability to be similar with another word.¹⁰ Certain items are more likely to be similar than others, and the items chosen or eliminated from a wordlist will affect the percentage of similarity. We can not arbitrarily disqualify items without skewing results to some degree and making comparisons with other wordlists which do not have the same items less valid.¹¹

Finally, we compare these results with the study that was done earlier.

2.3 Comparison with Barr, Barr, and Salombe, 1979

Tables 7 through 10 attempt to present a summary of our results for Balantak, Andio, and Saluan in such a way that they can be compared with the results from BBS.

Table 7 shows the highest and lowest percentages for any Balantak dialect with any Saluan dialect, and both with Andio as based on 428 items from SUW. Table 8 does the same, based on ninety-four items from S100. Table 9 shows the results of the author's own comparison of the wordlists in BBS (1979:102-4), using the same ninety-two items which they used. Table 10 shows BBS (1979:26) results for the ninety-two items they selected from S100.

Balantak	Balantak	Balantak	Balantak
70 Andio	65 Andio	68 Andio	61 Andio
52 60 Saluan	47 57 Saluan	54 65 Saluan	47 59 Saluan

Table 7
Highest and Lowest Percentages,
428 items (SUW)

Balantak	Balantak	Balantak	Balantak
78 Andio	78 Andio	78 Andio	78 Andio
66 74 Saluan	66 74 Saluan	66 74 Saluan	66 74 Saluan

Table 8
Highest and Lowest Percentages,
94 items (S100)

Balantak
74 Andio
61 71 Saluan

Table 9
My Percentages, 92 items (S100)
Data in BBS (1979:102-4)

Balantak
78 Andio
66 74 Saluan

Table 10
Percentages on 92 items (S100)
from BBS (1979:26)

Perhaps the most obvious observation to be seen from these tables is that the results vary to a fair degree, dependent at least on the length of the wordlist, the analyst, and, in the case of Balantak and Saluan, the dialect which is being compared (reflected by highest and lowest scores). It is clear that the difference between Tables 7 and 8 is because of data. The same is true to some extent for Tables 8 and 10. (Eighty-seven items selected from S100 were identical for the two studies; cf. Appendix III.) The difference between Tables 9 and 10 is because of the analyst. We expect language variation in the Balantak and Saluan language areas (shown by the highest and lowest percentages for both Tables 7 and 8) which is not reflected by the earlier study, but even if we consider only the highest percentages from Table 8 the results of this study are 9-12 points below the earlier study. The differences between Tables 9 and 10 are 3-5 points.¹²

3. SOCIOLINGUISTIC FACTORS

We were particularly interested in language use and language attitudes because of the implications for vernacular literature. We knew that language variation was present and viewed lexicostatistics as a first attempt to predict intelligibility; intelligibility is the first prerequisite, of course, for the intelligent use of literature. However, as Cooper (1976) has noted, the acceptance of literature may not correlate directly with the intelligibility of oral speech. We felt reasonably certain, based on lexical similarity and general public opinion, that a high level of intelligibility existed among all Balantak dialects, but, if they were present, we wanted to uncover any other factors that might otherwise influence the acceptability of vernacular literature based on any of the various dialects of the language.¹³

The questionnaire used covered a variety of areas, but here we only attempt to summarize the information most closely related to language use and language attitudes.¹⁴ Most of the information that we gathered was reported to us and not what we directly observed. The discussion here is limited to Balantak and Andio.

Education. Approximately one-half of the places visited reported using the vernacular as well as Indonesian in the first year of elementary school. This is to assist those new students who do not yet know Indonesian. Other places reported using only Indonesian.

There are two or three state and private junior high schools (SMP) in both subdistricts. If students go on to senior high school (SMA), they must go to Luwuk, the district capital, where they will encounter high contact with Indonesian as well as other vernacular languages. There are plans to begin a senior high school in the town of Balantak (Balantak Kota), and there is a private senior high school in Tangeban.

We did not collect information on the background of schoolteachers which I now regret; however, my informal observation is that well over half of them in the Balantak area were themselves Balantak.

Social Interaction. The large majority of Balantak marriages were with other Balantaks; Andio reported more mixed marriages. Outsiders residing in the Balantak language area, an estimated 10-20% of the population, include those from the most closely located language groups as well as Gorontalo, Bugis, Balinese, Javanese, and Chinese people.¹⁵

Social interaction for economic or government purposes may or may not be conducted in the vernacular (cf. 'Language Situation' below). Village leaders and heads of households would be most likely to engage in interaction with outsiders where Indonesian is required.

Radio and more recently television, particularly in coastal locations, while not providing direct social interaction, do provide a growing exposure to Indonesian.

Reported Dialect Differences. Apart from linguistic evidence, what were popularly held opinions about Balantak dialects? No clear pattern emerged except for a consensus that there are differences between the Balantak and Lamala subdistricts; however, everyone considered them one language. We heard no dissent on this, and felt this showed significant cultural unity and solidarity among Balantaks.

Various places reported Batubiring, the Dolom-Talima area and our field residence, to be the place of origin for the Balantak culture and language, and that the dialect used there was the *halus* 'proper' Balantak language. It is not clear to me whether these remarks were made in deference to those on the survey team who reside there. The Dolom-Talima area is more culturally conservative and traditional in outlook than coastal areas of Balantak, but it does not seem to otherwise enjoy social or economic prestige.

Coastal vs. Noncoastal. This distinction did not grow out of our questionnaire results since only one location, Dolom, was noncoastal; nevertheless, some points seem worth making here. All interior villages, which make up about a fifth of the total, are less than a day's walk from the coast, and only a few are not yet accessible by road. However, as noted above for Dolom-Talima, these people tend to be more culturally conservative and traditional in outlook. For a handful of people found mostly in isolated places, Balantak is their only language. Coastal people could be characterized as having greater cultural assimilation and are generally better educated. Most outsiders to the Balantak area live in coastal communities.

The implication of this feature of Balantak society for the language variety of prestige, particularly with regard to vernacular literature, is not known. Does any cultural prestige that might exist in noncoastal locations as noted above outweigh or balance lack of economic or social prestige? What is the relative significance of these factors for the prestige of any particular dialect? These factors in particular should be further observed and studied, and may, in the case of vernacular literature, only become apparent as actual literature is produced and used.

Language Situation. This was probably the most interesting part of the questionnaire, showing the domains of three languages in use: the vernacular, Indonesian, and, in Muslim communities, Arabic.

Four broad domains of language use emerge here:

- 1) home, work, and traditional cultural events; the vernacular was usually used,
- 2) religious activities in the church or mosque; Indonesian or Indonesian and Arabic were primarily used,
- 3) government functions; Indonesian was used almost exclusively for formal occasions, but informal activities in offices, etc., might transpire in the vernacular, and
- 4) health clinics, businesses, shops; the language used here depended on the personnel involved. If the shopkeeper or health attendant was native to the area or had learned the vernacular, the vernacular was used; otherwise, Indonesian.

There were indications that the use of Indonesian is increasing. Several places reported that the parents spoke Indonesian to their children in the home in order to prepare the children for attending school, and that the children used Indonesian when playing with each other. Most places did not report this, and our observations agree with this.

To summarize the language situation, we found vigorous use of the vernacular in the daily life of the people throughout the region. It is the first language for the majority of the children as well as adults, and social interaction and work activities are clearly the domain of the vernacular. Indonesian is used in the domain of formal education, government affairs, and religious activities. Arabic is used in Muslim contexts.

4. SUMMARY

Our lexicostatistic results show a high level of lexical similarity among Balantak dialects. They show Andio to be more similar to Balantak than

Saluan, but this is only based on a comparison of the lexicon; further grammatical comparisons and reconstruction are necessary to establish genetic relationships.

In support of our conclusions from lexical similarity, Balantak people consider their ethnic group to have only one language: Balantak. This cohesion would support the acceptance of vernacular literature from anywhere in the language area, but it may be the case that certain dialects would enjoy more prestige than others; in this case, a clear concensus as to preference may emerge. We would expect 'standard' Balantak to evolve if vernacular literature continues to develop.

Among some community leaders and educators there appears to be a growing interest in vernacular literature for various reasons:

- 1) the language is viewed as an important expression of their culture;
- 2) vernacular literature is viewed as an appropriate way to record and preserve the language and culture; and
- 3) vernacular literature is seen as an important tool for the development of the people and the region, particularly in educational aspects.

The degree of this interest and support, both locally and nationally, will likely prove to be a key factor in the further development of vernacular literature by the Balantak and Andio people.

ENDNOTES

¹This paper was written under the auspices of the Hasanuddin University-Summer Institute of Linguistics Cooperative Program, and the Balantak Language Field Program.

I am grateful to numerous local officials who facilitated this survey. Mr. H. M. Yunus, *bupati* 'district chief' for the Banggai district, together with his staff provided letters of introduction for all *camat-camat* 'subdistrict chiefs'. The subdistrict chiefs from the five subdistricts of Balantak, Lamala, Luwuk, Kintom, and Pagimana which were visited in this survey were also very helpful. Finally, the village leaders for each of the thirteen villages often hosted us on our overnight visits and either helped us themselves or secured help for obtaining the information we needed. We made many new acquaintances and friendships during these few short weeks. I am grateful for the kind help we received, and I trust that our interaction for this study has also been of some benefit to them and the people they represent.

I also wish to thank colleagues Donald Barr, Timothy Friberg, Michael Martens, Philip Quick, Ronald Snell, and Roland Walker for their helpful comments on earlier drafts of this paper.

²The Sulawesi Umbrella Wordlist is a composit of various other wordlists: Reid's 372-item Philippine wordlist less two items not relevant to Sulawesi; Blust's 200-item Proto-Malayo-Polynesian wordlist; various South and Central Sulawesi SIL survey wordlists, including the Swadesh 100-item and Swadesh 200-item wordlist; and several additional items.

³The Bahingan Saluan wordlist was recorded by Mr. Dago Molintas, a native of Bahingan, and resident school teacher in Dolom, Balantak.

⁴Martens and Hanna (1988:5) disqualified forty-four items from SUW for their study of the Badaic languages, but they did not detail the items eliminated.

⁵Cf. Simons (1977) for a full discussion of range of error and applying significance decisions to lexicostatistic data. He suggests grading the reliability of data according to five levels, ranging from data as a result of many years of field work down to survey data obtained by monolingual elicitation. Also, 'the greater the number of words compared, the greater the significance between different values' (1977:81). Based on these assumptions, he sets up tables of significance.

⁶Noorduyn's essay (pp. 9f) on Sulawesi languages also includes an interesting discussion of the strengths and weaknesses of lexicostatistics. Martens (1985) discusses potential error on survey wordlists.

⁷From a cursory overview of the data, Andio also appears to be closer phonologically to Balantak than Saluan: it also has the voiced alveolar trill /r/ which is absent in Saluan; Saluan, on the other hand, has the voiced alveopalatal affricate /j/, the alveopalatal nasal /ñ/, and the voiceless laryngeal fricative /h/, none of which occur in Balantak and Andio. All three languages have long vowel phones; they also have nasal plus consonant phones, but it is not clear how they function in Andio and Saluan. The relationship of stress to long vowels in Saluan appears different in some instances from the Balantak pattern. (Cf. Busenitz and Busenitz (1990), for further details on Balantak phonology.)

⁸Martens and Hanna (1988:5-6) also found higher percentages for S100 as compared to 444 items from SUW and suggested that this was because S100 contains core vocabulary, items that tend to remain stable and resist change/borrowing. Perhaps the significantly higher percentage with Indonesian in particular, as noted here, supports their view.

⁹Michael Martens via personal communication has suggested that a higher percentage of similarity on noncore vocabulary may reflect significant

social interaction and borrowing--the core vocabulary still resisting change and therefore showing greater dissimilarity.

In this case, a historical period of divergence when both core and noncore vocabulary become dissimilar from neighboring language varieties--noncore vocabulary more quickly so--is followed by a historical period of convergence when noncore vocabulary more quickly becomes similar with neighboring languages.

¹⁰Perhaps these findings provide as good support as any for the distinction of 'core' and 'noncore' vocabulary. Note, however, as mentioned in endnote 9, that factors of divergence and convergence still might make things difficult to sort out. Lexicostatistic studies should deal with the same lexical items to be fully comparable. The subsets of SUW allow for these wordlists to be extracted and compared with other studies.

¹¹We attempt to avoid this by comparing only items which exist for all 14 dialects/languages in this study; nevertheless, note that the number of synonyms for any particular list could have a significant effect on the comparisons. I have not tabulated the synonyms in the wordlists.

¹²Martens (1990:55-6) compares some of his Kaili-Pamona data with Barr, Barr, and Salombe and notes differing results. Noorduyn (p. 15) also notes lower percentages for his own comparison of Barr, Barr, and Salombe data; in his case, Buol and Toli-Toli. This underscores the problem behind the question, how much phonetic similarity is 'enough' to establish lexical similarity? Others have tried to answer this problem with phonostatistics.

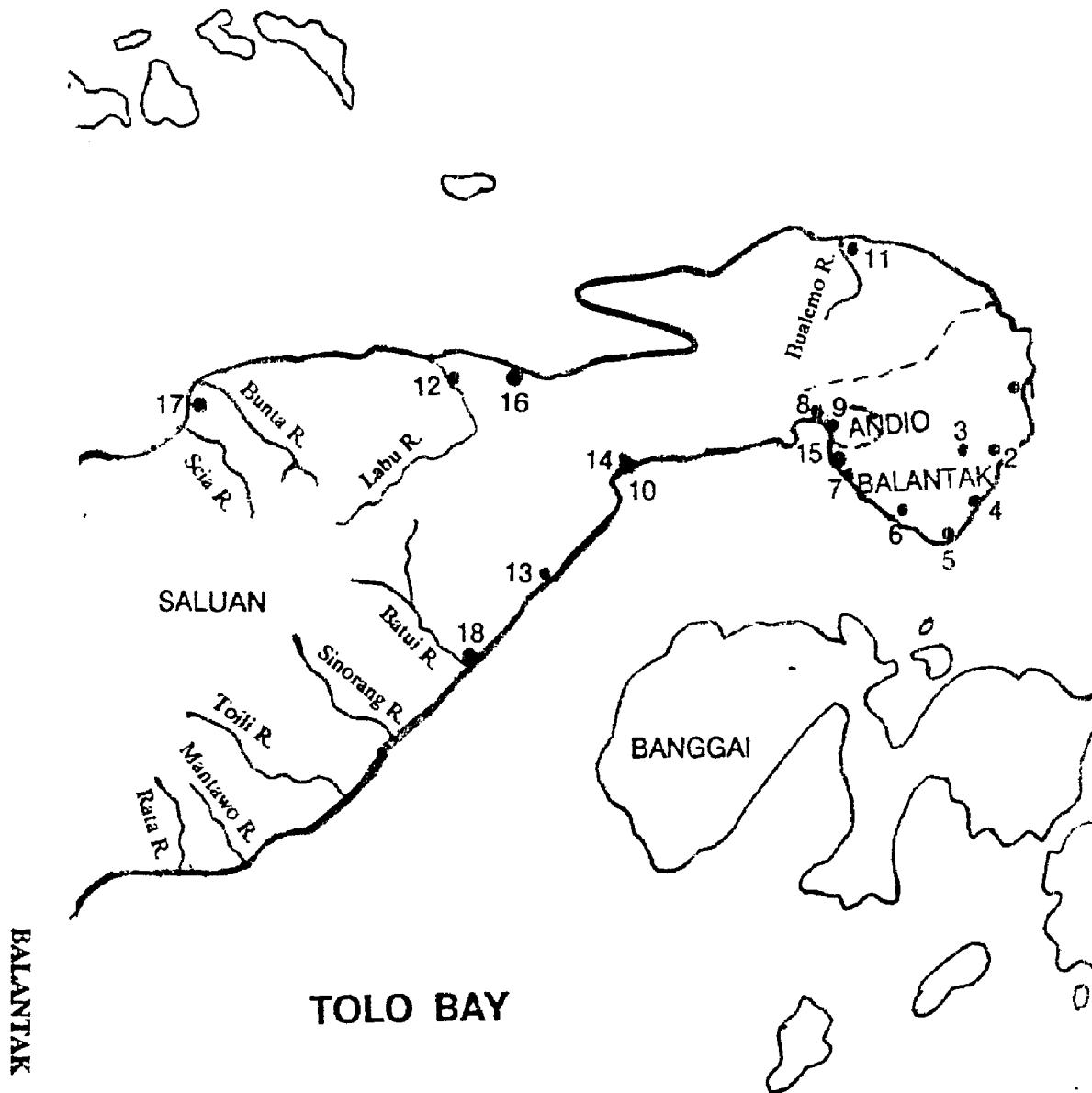
Barr, Barr, and Salombe did' not detail how they determined lexical similarity. If the procedure described above (Sanders 1977:36-37) was used where all words for one item are listed together, it is sometimes possible to see 'cognate chains' (Sanders 1977:34) where intermediate words clearly show a progression so that words quite dissimilar phonetically can be considered lexically similar. This, then, has the effect of increasing lexical similarity percentages.

This dissimilarity of results underscores the validity of the Simons (1977) discussion of tables of significance for lexicostatistics (see endnote 5).

¹³By 'acceptability' we do not mean anything having to do with the content or format of the literature, but rather, the linguistic form or language variety in which it is presented.

¹⁴Cf. Busenitz and Martens (1979:14f) for further discussion of acceptability of literature and sociolinguistic surveys.

¹⁵Cf. Busenitz, M. (1989), for a more complete overview of Balantak society.



APPENDIX I

Survey Locations

BALANTAK

1. Tokuu, Balantak subdistrict
2. Kota Balantak, Balantak subdistrict capital
3. Dolom, Balantak subdistrict
4. Tombos, Balantak subdistrict
5. Sulu'bombong, Lamala subdistrict
6. Sobol, Lamala subdistrict
7. Poro'an, Lamala subdistrict
8. Eteng, Lamala subdistrict

ANDIO

9. Tangeban, Lamala subdistrict

SALUAN

10. Lumpo'nyo, Luwuk subdistrict
11. Sampaka', Pagimana subdistrict
12. Bahingan, Pagimana subdistrict
13. Kintom, Kintom subdistrict capital

OTHER

14. Luwuk, Luwuk-Banggai district capital
15. Bonebobakal, Lamala subdistrict capital
16. Pagimana, Pagimana subdistrict capital
17. Bunta, Bunta subdistrict capital
18. Batui, Batui subdistrict capital

APPENDIX II

The items disqualified from the Sulawesi Umbrella Wordlist for this study can be categorized as follows:

1. The elicitation of the item was inadequate, either because it became clear that the responses did not all cover the same semantic domain or because, for one reason or another, no response was obtained in some locations. With careful analysis and rechecking, this category could probably be reduced substantially. The following twenty-eight items fell in this category: 13, 16, 52, 117, 118, 119, 157, 174, 186, 202, 208, 262, 263, 264, 290, 302, 304, 321, 322, 323, 324, 325, 326, 374, 376, 377, 438, and 458.

2. The elicited response in some locations was a descriptive phrase, not a single word. This criterion was not used to eliminate Indonesian items, e.g. 233, 394, 455. The following fifteen items fell in this category: 3, 9, 11, 59, 77, 78, 85, 86, 87, 88, 203, 315, 345, 346, and 360.

3. An item was the same word or had the same root in all instances of the dialects/languages compared as another item. Where one set of dialects was the same as a previous item, but another set was not, the item was not eliminated. Note that including these items would raise percentages. The following thirteen items fell in this category: 44 (cf. 28), 90 (cf. 89), 92 (cf. 91), 113 (cf. 97), 148 (cf. 48), 176 (cf. 175), 178 (cf. 175), 179 (cf. 175), 370 (cf. 369), 410 (cf. 407), 434 (cf. 411), 440 (cf. 244), 452 (cf. 63).

4. No generic word exists in some locations. Conceivably this could overlap with 1. above in that respondents may not give a word because it does not exist in their dialect, though one would expect it if it exists in a closely related dialect. The two items in this category are 136 and 225.

5. The survey context was inappropriate for elicitation. The two items are 64 and 65.

The total number of disqualified items is sixty.

APPENDIX III

The following six items are from the Swadesh 100-item wordlist and are included in the sixty disqualified items as listed in Appendix II. This reduces the items eligible for comparison from that list to ninety-four.

113,	'feather'	322,	'that'
148,	'bark'	376,	'to say'
321,	'this'	334,	'to kill'

Barr, Barr, and Salombe disqualified eight items from the Swadesh list for their study, leaving ninety-two eligible for comparison (1979:21-22). Only one of these overlaps with the six above. They disqualified:

5,	'hair'	295,	'all'
70,	'man'	376,	'to say'
144,	'tree'	460,	'to walk'
173,	'seed'	463,	'to lie down'

This leaves only eighty-seven items that are the same for the Swadesh list in this study and the Swadesh list in Barr, Barr, and Salombe. This may account for some of the discrepancy in percentages between the two studies.

REFERENCES

- ABAS, Husen, and T. David ANDERSEN, eds.
1990 *Bahasa-bahasa Daerah Sulawesi Dalam Konteks Bahasa Nasional. Prosiding Konferensi dan Seminar Nasional Ke-5, Masyarakat Linguistik Indonesia*. Ujung Pandang: Proyek Kerjasama UNHAS-SIL.
- BARR, Donald F., Sharon G. BARR, with C. SALOMBE
1979 *Languages of Central Sulawesi*. Ujung Pandang: Hasanuddin University-Summer Institute of Linguistics.
- BUSENITZ, Marilyn J.
Balantak Overview. MS.
- BUSENITZ, Robert L., and Marilyn J. BUSENITZ
Balantak Phonology and Morphophonemics. To appear.
- BUSENITZ, Robert, and Michael MARTENS
1979 Considerations for Language Identification Surveys. In *Notes on Linguistics*, No. 10, 10-24 (April 1979). Dallas, TX.: Summer Institute of Linguistics.

- COOPER, Robert L.
1976 Review of Dialect Intelligibility Testing, by Eugene H. Casad.
Language 52:738-41.
- GRIMES, Charles E. and Barbara D. GRIMES
1987 Languages of South Sulawesi. *Pacific Linguistics*, D-78.
- LOVING, Richard, ed.
1977 Language Variation and Survey Techniques. *Workpapers in Papua New Guinea Languages*, Vol. 21. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- MARTENS, Michael
1985 A Comparison of Two Uma Wordlists. In *Bits & Pieces: Bulletin of Indonesia Branch Technical Studies*, October 1985, pp. 27-29.
- 1990 Rekonstruksi Bahasa Purba Kaili-Pamona. In Abas and Andersen, eds., 1990:54-68.
- MARTENS, Michael, and Roger HANNA
UNHAS-SIL Sociolinguistic Survey: Bada, Besoa, and Napu Language Groups. MS.
- NOORDUYN, J.
The Languages of Sulawesi. MS.
- SANDERS, Arden
1977 Guidelines for Conducting a Lexicostatistic Survey in Papua New Guinea. In Loving, ed., 1977:21-41.
- SIMONS, Gary
1977 Tables of Significance for Lexicostatistics. In Loving, ed., 1977:75-106.
- TRAVIS, Edgar W.
1986 *Penyelidikan Persamaan Bahasa Dengan Menggunakan Mikrokomputer*. Ambon, Maluku: Universitas Pattimura and The Summer Institute of Linguistics.