

THE NORTH SULAWESI MICROGROUPS: IN SEARCH OF HIGHER LEVEL CONNECTIONS

J.N. Sneddon
Griffith University, Brisbane

There are three language groups in North Sulawesi: the Sangiric, Manahasan and Gorontalo-Mongondic groups. Each of these is referred to here as a microgroup, a small set of closely related languages whose relationships with each other can be accounted for in terms of generally precise statements of phonological change from a common parent language.

This paper examines whether there is any worthwhile evidence for linking the three microgroups, or any two of them, directly with one another. It is concluded that there is no phonological or grammatical evidence to link them and that uniquely shared lexical items are probably the result of borrowing.

Further, there is little evidence to connect any of the microgroups closely with other groups of Sulawesi languages. The evidence is that all three can be classified with the Philippine languages, although to date no study has established a close connection with any particular Philippine subgroup.

1 INTRODUCTION

The North Sulawesi Province of Indonesia is occupied by three lower order groups of Western Malayo-Polynesian language: the Sangiric, the Manahasan and the Gorontalo-Mongondic groups.

Identifying the three sets of North Sulawesi languages and assigning individual languages to their respective sets has presented few difficulties. Furthermore, with the exception of the Gorontalo-Mongondic subgroup, it has also been possible to determine the interrelationships between languages within each group. Determining the segmental phonemes of the parent languages has largely been possible and a number of lexical reconstructions have been made for each of the three parent languages.

Each of these three groups is here referred to as a microgroup, a small set of closely related languages whose common parent language can be reconstructed to a large extent and in which the phonological differences between member languages can generally be accounted for by precise statements of phonological change from that parent.

Following the successful identification of microgroups the next step should be the search for higher-level groupings or macrogroups - groups whose membership comprises two or more microgroups (along with possible language isolates). No detailed study has previously been undertaken to discover the macrogroup affiliations of the North Sulawesi microgroups. The present paper is offered as a contribution to the search for higher-level connections of these groups, in particular to see if any worthwhile evidence appears for linking them, or any two of them, directly with one another.

The three microgroups are discussed separately in Section 2, along with evidence for recognition of the groups and internal classification.

Section 3 is concerned with the external relationships of the North Sulawesi languages, and previous studies which attempt to group them, mainly with Philippine languages, are mentioned. Most of this section is devoted to examining, and rejecting, evidence for linking them with other groups of Sulawesi languages.

Section 4 examines the possibility of the North Sulawesi microgroups, or any two of them, linking directly with one another.

2.1 The Sangiric group

This group consists of Ratahan (Rth) and Bantik (Ban), both spoken in the Minahasa region of North Sulawesi, Sangir (San), spoken in the Sangir (Sangihe) Archipelago north of Minahasa, Talaud (Tal), spoken in the Talaud Archipelago to the north east of the Sangir Archipelago, and Sangil (Snl), spoken in a few areas in the extreme south of Mindanao and on near-by islands.

Adriani (1893:1) recognised the close relationship between San and Tal, referring to them as dialects of one language. Later (Adriani and Adriani 1908:2) he assigned Rth and Ban to the 'Sangir-Talaut' group. Maryott (1978 and earlier unpublished work) has identified Snl and its close relationship to San.

These languages have been the object of a comparative study (Sneddon 1984), in which interrelationships are established and the phonology and part of the lexicon of their parent language, Proto-Sangir (PSan), is reconstructed. Among the lexical and phonological evidence presented there for such a group the following uniquely shared innovations are particularly strong:

a. Lexical innovations (PSan reconstructions are given but not reflexes in the modern languages, for which see Sneddon 1984):

*moRoŋ 'mouth'
 *pakel 'heel'
 *putuŋ 'fire'
 *tolay 'tail'
 *binaba 'cloud'

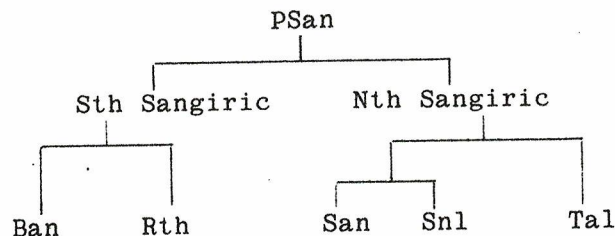
b. Semantic and irregular phonological innovations:

*ləgay 'laugh' (phonological alteration of PAN *geli())
 *bəŋaR 'molar' (metathesis of the last two consonants in PAN *baqRaŋ)
 *idun 'nose' (replacement of the final nasal of PAN *ijuSuŋ)
 *utak 'hair' (semantic alteration of PAN *qutak 'brain')
 *pisi 'skin' (semantic alteration of Proto-Philippine (PPh) *pi:si?
 'rope, bark')

c. Phonological innovations:

(i) Final *a became PSan *e but only in pronouns, e.g. PAN *kita 'we' > PSan *kite, PAN *-na 'he, she (genitive)' > PSan *-ne.
 (ii) PAN *a in closed final syllables became PSan *e if the preceding vowel was *a and the following consonant was an alveodental, e.g. PAN *Zalan > PSan *dalen 'road', PAN *habaRat > PSan *barei 'west wind'.
 (iii) PAN *e was replaced by another vowel in final syllables. In particular it was regularly replaced by *u before bilabial consonants, e.g. PAN *qitem > PSan *itum 'black', PAN *qatep > PSan *atup 'roof'.

The Sangiric languages subgroup as follows:



The northern languages share considerably higher lexicostatistical percentages with one another than they do with the southern languages and have undergone a number of phonological innovations, such as metathesis of t and s.

Rth and Ban also uniquely share a number of lexical innovations which,

due to their geographical distance from each other, are unlikely to result from borrowing.

The lowest lexicostatistical score within the microgroup is 47% for Rth and Tal.

2.2 The Minahasan group

This group contains five members, all spoken in Minahasa: Tondano (Tdn), Tonsea (Tse), Tombulu (Tbl), Tontemboan (Ttb) and Tonsawang (Tsw).

Adriani (1908, 1925) recognised these languages as forming a group. However, he incorrectly subgrouped Ttb and Tsw on the basis of a selective consideration of a few aspects of morphology and phonology. In his linguistic map, Esser (1938) followed Adriani in grouping the two together as dialects of one language, but later he wrote that the two languages were not as closely related as had previously been supposed (Noorduyn 1963:869). Sneddon (1970) presents lexicostatistical evidence which clearly places Tsw apart from the other Minahasan languages.

A detailed comparative study (Sneddon 1978) supports the lexicostatistical evidence. This work offers a reconstruction of Proto-Minahasan (PMin) phonology and a certain amount of morphology and lexicon.

Evidence for a separate Minahasan group includes:

a. Lexical innovations. Among a considerable number of shared lexical innovations are the following (for reflexes in the modern languages see Sneddon 1978):

*ente? 'strong'

*indo 'get, fetch'

*lar₂e 'just, new'

*ulit 'true'

*teles 'buy'

b. Semantic and irregular phonological changes, including:

*lehe? 'neck' (with metathesis of the last two consonants in PAN *ligeR).

(An earlier assumption that this change also occurred in the Sangiric languages is incorrect, see section 4.0 below.)

*inde? 'fear' (reflecting PPh *haldek 'fear', with phonological irregularities)

*ba'aŋ 'tooth' (semantic alteration of PAN *baRqaŋ 'molar')

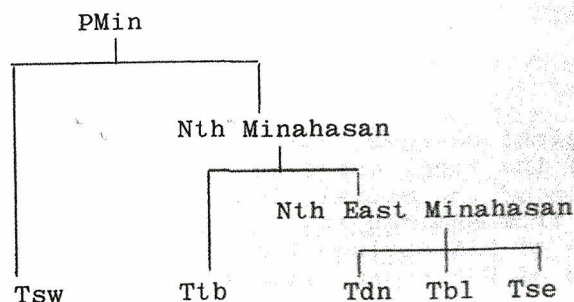
*kaməs 'wash clothes' (as well as 'squeeze, wring out', reflecting PAN *kemes 'held or squeezed in hand')

*sikep 'hawk; snatch, seize' (semantic alteration of PPh *sikep 'catch; grope for')

c. Phonological changes. Phonological evidence for a separate Minahasan group is far poorer than for the Sangiric group. The strongest piece of evidence is metathesis of initial PAN *R (> PMin *h) and a following vowel, e.g. PAN Ramut > PMin ahmut 'root', earlier *Riduq (based on evidence of Bontok gi'do 'earthquake') > PMin *ehdo 'earthquake'.

Another change, not reported in Sneddon 1978, is the replacement of PAN *e (schwa) by another vowel adjacent to PMin *h (reflecting PAN *R) or *? (reflecting (*q, *h or hiatus). It assimilated to a vowel separated from it by medial *h or *?, e.g. PAN *tugeD > PMin *tu'ud 'stump', PPh *keRaŋ > PMin *kahaŋ 'scab'. Elsewhere it became PMin *e (mid-front vowel): PAN *saler > PMin *saleh 'floor'.

The Minahasan languages group as follows:



Lexicostatistical evidence is unequivocal for this grouping, with cognacy figures between 69% and 72% among the North-East languages, from 57 to 61% for these languages with Ttb and percentages in the low 40s between these languages and Tsw. The lowest percentage is 41 for Tdn and Tsw.

2.3 The Gorontalo-Mongondic group

This can be conveniently divided into two subgroups: the Mongondic group, comprising Mongondow (Mdw), sometimes called Bolaang-Mongondow, and Ponosakan (Psk), and the Gorontalic group, comprising Gorontalo (Gtl), Buol (Bul), Kaiqipang (Kdp), Suwawa (SwW), Bolango (Blg), Bintauna (Bnt) and Lolak (Llk).

The choice of a single name for this group - as in the case of the Sangiric group, which has only one prominent language, San - would have been difficult. The name Gorontalo-Mongondic (hereafter GM) has more to do with the presence of two prominent languages, Gtl and Mdw, each with a large number of speakers, cultural influence over the other ethnic groups and a substantial linguistics literature, than with purely linguistic considerations (such as the occurrence of two primary subgroups, which is also the case in the other two microgroups).

On the basis of Charles' claim for a close connection between Gtl and Mdw (see below), Sneddon (1983) grouped these languages together as the Mongondow-Gorontalo Supergroup. The name Gorontalo-Mongondic was first used by Usup (1986 and earlier unpublished work).

Despite the double-barrelled label the languages can be recognised as constituting a single microgroup. The lowest cognate percentage for any two members of the group falls in the low 40s as is the case in the Minahasan group.

A close relationship between Mdw and Psk has long been recognised (e.g. Adriani and Adriani 1908:2). Adriani and Kruijt (1914:184) noted the close relationship between those Gorontalic languages which had been recorded, Gtl, Bul, Kdp and Blg (called Bulanga or Bulanga Uki by them).

However, a close relationship between the Gorontalic and Mongondic subgroups was not recognised until recently, apparently due to the fact that there was very little information available on Gorontalic languages apart from Gtl, which has undergone a relatively large number of sound changes.

Charles (1974:487) first suggested a close relationship between Gtl and Mdw, noting striking similarities in their vocabularies and phonological histories.

Noorduyn (1982) goes further in recognising a close relationship between Gtl and Mdw, although he comments that further comparative study is required to determine whether the Gorontalic and Mongondic languages belong in the same group. On the evidence available to him he was unable to determine whether SwW (or Bunda as he calls it, using an older name) was closer to Gtl or Mdw.

Usup (1986 and earlier unpublished work) has carried out a detailed comparative study, including reconstruction of Proto-Gorontalo-Mongondic (PGM) phonology and a word list.

Evidence for a GM group includes:

a. Lexical innovations. Among numerous innovations the following are a small sample:

- *gaŋu(t) 'dry'
- *litu? 'sit'
- *lanit 'sharp'
- *ligogow 'neck'
- *pampiq 'move'

b. Semantic and irregular phonological changes. Some examples are:

- *tulan 'bone' (with replacement of the final consonant in PAN *tulaŋ)
- *naŋka 'jackfruit' (with alteration of the medial cluster in PPh *naŋka)
- *buiŋ 'charcoal' (with loss of *j from PAN *buiŋ)
- *sulu? 'fire' (semantic alteration of PAN *suluŋ 'torch')
- *buŋa 'fruit' (semantic alteration of PAN *buŋa 'flower')

c. Phonological innovations. Phonological evidence for a separate group is not strong. Innovations are listed in Sneddon and Usup (1986), of which the strongest is replacement of PAN *e by *o, e.g. PAN *qatep > PGM *atop 'roof', PAN *?enem > *onom 'six'.

Replacement of *a by *o (via schwa) in the antepenultimate syllable can be added, because although it occurred in other groups of WMP languages, it did not occur in the Minahasan and Sangiric groups, e.g. PAN *balian > PGM *bolian (cf. PMin, PSan *balian).

Evidence for a Gorontalic subgroup, separate from the Mongondic languages (Mdw and Psk) is discussed in Sneddon and Usup (1986). Phonological differences distinguishing the two subgroups include:

(i) PAN *ay > Proto-Mongondic (PMdw) *oy, Proto-Gorontalic (PGtl) *e, e.g. PAN *balay > PMdw *baloy, PGtl *bale 'house'.

(ii) PAN *iw > PGtl *i, PMdw *uy, e.g. PAN *laRiw > PGtl *lagi, PMdw *laguy 'run'.

(iii) PGtl underwent metathesis of t and following s, while PMdw did not, e.g. PAN *Ratus > PGtl *gasut, PMdw *gatus 'hundred'.

(iv) The Gorontalic languages have a paragogic vowel following previous final consonants, although the vowel is now in the process of loss in Bul (see Sneddon and Usup 1986:417-8). This reflects an earlier paragogic vowel *o, which is reconstructed for PGtl, e.g. PAN *bukid > PGtl *bukido 'mountain' (PMdw *bukid), PAN *inum > PGtl *inumo 'drink' (PMdw *inum).

Although this principal subdivision is clear-cut, internal classification has presented difficulties not encountered in the Sangiric and Minahasan groups.

First, Llk shows a high lexicostatistical percentage with Mdw and had previously been classified as a dialect of Mdw (e.g. Wilken and Schwarz 1868, Sneddon 1983, Usup 1986). Sneddon (to appear) argues that on the basis of qualitative evidence Llk is a Gorontalic language which has been subjected to heavy influence from Mdw. For instance, Llk agrees with the Gtl languages in its reflexes of the first three changes listed above and there is good evidence for earlier occurrence of a paragogic vowel.

Within the Gorontalic subgroup there have been a considerable number of phonological changes, but it is not possible to subgroup the languages on the basis of these changes. A recent paper (Sneddon and Usup 1986) examines shared phonological innovations in the Gorontalic group and shows that the evidence for subgrouping provided by some apparently high quality shared changes conflicts with the equally impressive evidence of other changes. Further, some languages participated in a number of shared changes while not undergoing chronologically earlier changes. Thus many shared innovations must be the result of areal diffusion among the languages rather than common inheritance.

These changes are discussed in detail in Sneddon and Usup (1986). Here, a few of the changes are briefly described:

(i) Final *a became o in Gtl and Bul: PGtl *mata 'eye' > Bul, Gtl mato, cf. Kdp, Bnt, Blg, Sww, Llk mata.

(ii) *k became Ø initially and ? elsewhere, except after n, in Gtl, Sww, Blg, Bnt, e.g. PGtl *konuku 'fingernail' > Gtl olu?u (where *n > l regularly) ('hand'), Sww, Blg, Bnt onu?u, cf. Bul, Kdp, Llk konuku.

(iii) *s became t in Bul, Gtl, Sww, e.g. PGtl *sali 'buy' > Bul, Gtl, Sww tali, cf. Kdp, Blg, Bnt, Llk sali.

(iv) Voiced stops were lost after homorganic nasals in Bul, Kdp and Gtl with the exception that g was not lost in Gtl: PGtl *tondoko 'fence' > Bul tonuk, Kdp tonuku, Gtl tonu?o, cf. Sww, Blg, Bnt londo?o; PGtl *nanga 'jackfruit' > Bul, Kdp nao, cf. Gtl lange, Sww, Blg nanga, Bnt (where *g > k in all environments) naoka.

(v) Voiceless stops became voiced stops following homorganic nasals. This change occurred in Bul, Gtl, Kdp and Sww, e.g. PGtl *pampiqo 'move' > Bul pambiq, Gtl, Kdp, Sww pambiqa, cf. Blg, Bnt pampiqa.

Clearly borrowing between the languages has been an important factor in the distribution of these shared changes. For instance, the participation of Sww in change (v) above but not in chronologically earlier (iv) can only be a result of areal diffusion.

On the surface it is difficult to see how the random spread of sound changes among the Gtl languages can be reconciled with a tree model of

their interrelationships, and a wave model, in which overlapping circles depict geographically spread shared features, may be more appropriate.

The question of whether the Gorontalic languages emerged from a more-or-less homogeneous beginning through gradual differentiation and on-going cross-dialect borrowing or whether in fact they originally split in a way which could be captured by a tree diagram but with later extensive areal diffusion of innovations obscuring the picture of these interrelationships, cannot be answered at this stage.

The geographical distribution of some of the shared sound changes, which sometimes affected languages now geographically far apart while not affecting other languages geographically intermediate, suggests extensive movements of populations. There is historical evidence for some recent movements. For instance, the Bolango people, closely related to the Atinggola, had a kingdom in the Gorontalo region. In 1855, under pressure from Gorontaloese immigrants they began to move eastward to several localities in Bolaang-Mongondow, including Molibagu, where the language is presently located. In 1862 the kingdom of Bolango in Gorontalo was disbanded and the rest of the Bolangos followed east (Riedel 1870, Haga 1931).

It is possible that comparison of grammatical structures, for which information is largely or entirely lacking for all Gorontalic languages except Gtl, may help answer questions about their interrelationships when available. But for the present it is not possible to determine the relationships among the Gorontalic languages.

3 EXTERNAL CONNECTIONS

3.0 Background

Similarities between most of the North Sulawesi languages and the Philippine languages have long been recognised. However, their 'Philippine-like' character has always been referred to only in general terms and no attempts have been made to link them with particular Philippine languages.

Adriani (1893:2) listed San, Tal, Ban, Rth (which he called Bentenan), Psk and Mdw as belonging to the Philippine group. He assigned the five Minahasan languages to a 'sub-Philippine group'. He made no reference to the Gorontalic languages. Nor did he provide any information on why he made the distinction between the Minahasan group and the other two.

Later (Adriani and Kruijt, 1914:184) he again assigned Mdw and Psk to the Philippine group, regarding Mdw as marking the southern limit of Philippine languages. In this publication he recognised the Gorontalic languages as a separate group lying to the west of Mongondow.

Subsequently (1925:135) he assigned the Minahasan languages, along with the Sangiric and Mongondic languages, to the Philippine group, without reference to his earlier 'sub-Philippine group'. Here he again excluded the Gorontalic languages from the Philippine group. He states (1925:142) that travelling westward from Mdw the languages of Sulawesi become less Philippine-like, although the Gorontalic group and the Tomini group (spoken to the west of the Gorontalic group) possess some of the complexities of Philippine verbal systems.

No notable study of North Sulawesi languages followed Adriani's work for many years. In his linguistic atlas, Esser (1938) followed Adriani in assigning the Sangiric, Minahasan and Mongondic languages to the Philippine group. He placed the four Gorontalic languages known to Adriani in a separate group on their own.

3.1 Phonological evidence

Charles (1974) includes Gtl, Mdw, Ttb and San, the only North Sulawesi languages for which adequate data were available to him, in his consideration of reflexes of PPh phonology. He notes that Dyen (1965) excluded Gtl, Ttb and San from the Philippine group on lexicostatistical evidence (he did not use Mdw in his study). Charles includes these languages because

'Sangirese and related languages and the Minahasan languages have long been considered to belong to a Philippine subgroup' and because Mdw 'shows affinities with Meso-Philippine languages' (he does not elaborate). Gtl is included because 'comparison revealed striking similarity in the vocabularies and (to a point) in the phonological histories of Mdw and Gtl'. He notes: 'Whatever evidence for the remoteness of Sangirese and Tontemboan from the Philippine languages is lexical, not phonological'.

Charles says little on the classification of Philippine languages and offers no pointers for finer subgrouping of North Sulawesi languages within the Philippine group. Importantly, however, his study does not raise any difficulties to any of the North Sulawesi languages being derived from Proto-Philippine.

While Charles' study reveals many regular reflexes of Proto-Philippine phonology in the North Sulawesi languages, no phonological study has shown any comparable unity of Nth Sulawesi languages with languages in other areas of Sulawesi.

Mills (1981:60) states that San 'shares important sound changes with Makasarese'. One of these is that 'final continuants (are) preserved with a support-vowel'. Thus, San biwihə?, Makasarese (Mak) bibere? 'lip' < PAN *bibiR. (In San, and some other Sangiric languages, final voiced stops also add the syllable ə?, e.g. San laudə? 'sea' < PAN *laud.)

Mills' support vowel, sometimes called an echo vowel, is called a paragoge by Maryott (1977) and, following him, Sneddon (1984:25, note 29). It also occurs elsewhere in Sulawesi, including the Gorontalic languages (see Section 2.3 above). But the paragogic syllable cannot be regarded as a shared innovation wherever it occurs in Sulawesi. In the Sangiric group Rth does not have a paragoge, e.g. Rth nipis 'thin'. Thus its development in the other Sangiric languages occurred after their split with Rth.

A paragogic vowel also occurs in the Gorontalic languages but not in the Mongondic languages (see examples in Section 2.3 above); its occurrence in the Gorontalic languages therefore reflects a post-PGM development.

Mills (1975) reconstructs Proto-South-Sulawesi (PSS) forms with final consonants, e.g. *nipis 'thin' (Mak nipisi?), *pager 'enclosure' (Mak pagara?); development of an additional syllable in some South Sulawesi languages thus occurred subsequent to PSS.

So in all three groups mentioned above the paragoge occurs in only some languages within each group. It thus results from independent parallel developments.

Mills also states that in San and Mak all final stops were replaced by glottal stop. In San final voiced stops, which do not occur in Mak, actually add a paragoge (see examples above). In both languages voiceless stops reduce to ʔ. This change also occurred independently as it did not occur in other languages within their respective groups, e.g. San atu? but Rth atup 'roof' < PSan *atup; Mak bassi? but Sa'dan bassik 'splash' < PSS *bæssik.

Addition of a paragoge and reduction of final consonants to glottal stop, or their disappearance altogether, occur elsewhere in Sulawesi.

In Seko, a language of northern South Sulawesi, a paragogic vowel occurs in one dialect but not in the other, thus Seko Tengah etiki, Seko Padang etik 'brain'. The paragogic vowel would appear to be a recent development here.

In Besoa and Bada, Central Sulawesi languages belonging to the Kaili group, all final oral consonants have reduced to ʔ, while in closely related Napu they have been lost, e.g. Besoa; Bada laji?, Napu laji 'sky' < PAN *lanit. Loss of all final consonants, including nasals, occurs in most of the other Kaili-Pamona languages, such as Palu and Parigi, e.g. Palu, Parigi kuli 'skin' < PAN *kulit; Palu, Parigi inu 'drink' < PAN *inum.

In Muna all final consonants have been lost, e.g. kuli 'skin' < PAN *kulit; sala 'path' < PAN *Zalan.

These phenomena clearly are not a result of a common development but are an interesting case of an apparent drift towards simplification of word endings among Sulawesi languages.

Another important change which San shares with Mak, according to Mills, is replacement of final nasals by ŋ, e.g. PSS *luran 'load' > Mak luraŋ, PSan *ludan 'load' > San luraŋ; PSS *tanəm 'plant' > Mak tanaŋ, PSan

*dendum 'dark' > San danduŋ. But here also the change did not occur in other languages within their respective groups and so post-dates their respective parent languages.

This replacement of final nasals by ŋ is also frequent in Sulawesi languages. Thus m and n also became ŋ in some languages of the Tomini group, such as Dampalas and Ampibabo, e.g. Dampalas, Ampibabo iton 'black' < PAN *qitem; Dampalas, Ampibabo jalan 'road' < PAN *Zalan.

Final n has been lost completely in the eastern Gorontaloic languages and partly in the western Gorontaloic languages and is a case of areal spread of a change (see Sneddon and Usup 1986:413). Loss of final m and ŋ has also occurred to some extent in Sw and Gtl.

In the Badaic languages final nasals have been entirely lost, e.g. Bada, Besoa, Napu inu 'drink' < PAN *inum; Bada, Besoa, Napu *wula 'moon' < PAN *bulan.

Loss of final nasals or their reduction to a single phoneme also appears to be part of the widespread drift to phonological simplification of word endings among Sulawesi languages.

Presenting a stronger case for a shared innovation is the change of final *uq to o and *iq to e, which occurred in both PSS and PSan, e.g. PSS, PSan *pile 'choose' < PAN *piliq; PSS, PSan *buno 'kill' < PAN *bunuq. Here the change is reflected in all members of each group and thus is assigned to their respective parent languages. The change therefore cannot be shown to represent independent developments by its lack of occurrence in some members of each group.

This change also occurred in other language groups in different parts of the island. It occurred in the Badaic languages, Bada, Besoa and Napu, although the glottal stop still remains in the first two languages, e.g. Bada, Besoa pulo?, Napu pulo 'ten' < PAN *puluq, Bada, Besoa sule?, Napu sule 'return' < Proto-Kaili-Pamona *suliq 'return' (Michael Martens, personal communication).

These languages are members of the Kaili-Pamona group (see Martens' article, this volume). The change did not occur in other members of the group and so the change in the Badaic languages must have occurred independently, or else under influence of South Sulawesi languages. Because of Napu's close relationship with the other two languages, the loss of ? in Napu must be a recent, and therefore independent, development.

Ulo Sirk (personal communication) notes that the change also occurs in Lemolang. This is a language of uncertain genetic status, which has been placed tentatively in the South Sulawesi stock (see Friberg and Laskowske's article, this volume). Sirk regards it as belonging with the Badaic subgroup because of phonological developments, including this one. He notes that the change apparently occurs also in Tolaki, a language of Southeast Sulawesi, although the change in this language may have spread from Lemolang. Examples in Tolaki are: pile 'choose' < PAN *piliq, hulo 'lamp' < PAN *suluq.

Whatever evidence there is for a common origin for this sound change among the languages outside North Sulawesi, the possibility of borrowing through contact exists. But phonological borrowing between Sangiric and South Sulawesi languages, lying at opposite extremes of the Sulawesi linguistic area, is not plausible. Given that parallel developments do occur - and the paragoge and other changes to final consonants which have affected both San and Mak are clearly such - and the absence of any supporting evidence for a close genetic relationship, morphological and lexical (see below) as well as phonological, there is no reason to regard changes to final *uq and *iq as anything other than independent parallel developments.

3.2 Morphological evidence

Morphologically also, as noted by Adriani, the North Sulawesi languages are Philippine-like (although Adriani apparently underestimated the extent of this in Gtl). They all share in the typically Philippine 'focus' system of verbal morphology, a phenomenon lacking in other Sulawesi language groups. Mills (1981:60) comments that morphologically there is little resemblance

between North Sulawesi and South Sulawesi languages. However, knowledge of morphology and other aspects of grammar is still very poor for many language groups, such as the Tomini languages and most of the Gorontalo languages, and definite statements would be premature at this stage.

3.3 Lexical evidence

Apart from Charles' study, most attempts to include North Sulawesi languages in wider groupings have been based on lexical study, in particular on lexicostatistical comparisons.

Dyen (1965) included Gtl (and Sww), Ttb and San in his lexicostatistical classification. His study excluded all these languages from the Philippine group and showed no close relationship among the languages themselves (apart from Gtl and Sww).

A number of studies incorporate only languages of the Sangiric group, drawing on data in Reid (1971).

In his lexicostatistical study Walton (1979) finds San and Snl to branch directly from Proto-South Philippine, one of two first order branches of Philippine languages. Thus, while he recognises them as being within the Philippine group, he does not link them closely with any other Philippine languages.

Llamzon and Martin (1976), who include Tal along with San and Snl, also find these languages to form a separate group which branches directly from the South-Philippine node.

Zorc (1986:156) argues that exclusively shared lexical innovations suggest that languages of the Philippine archipelago form a single Austronesian subgroup. He places languages of the three North Sulawesi microgroups in the South Philippine branch.

Thus, just as Charles assigns all three microgroups to the Philippine group on phonological grounds, Zorc does so on lexical grounds. Charles' comment that any evidence for remoteness of these languages from the Philippine languages is lexical, refers only to the findings of Dyen's lexicostatistical study, which is at odds with the other studies, including other lexicostatistical studies. It is interesting that while all these studies (except Dyen) link the North Sulawesi microgroups with the South Philippine subgroup of Philippine languages (although Charles is less specific) none has produced any evidence for a close link to any particular Philippine language or microgroup.

Zorc's concern is lexical comparison within the Philippine group, but others have looked at lexical similarities within the Sulawesi region. Mills (1981) reconstructs a number of PSS words which have known cognates in other parts of Sulawesi, but apparently not elsewhere. However, he states (p.60) that the island contains 'at least three major subgroups, possibly four. No one group appears to be closely related to any other'. He suggests that all North Sulawesi languages can be classified as Philippine and that consequently cognate forms found throughout the island could theoretically be reconstructed at a higher level, except for the strong possibility of factors such as borrowing. Mills is thus clearly rejecting the possibility of a 'Sulawesi group'.

If the North Sulawesi languages belong to the Philippine group while other Sulawesi groups do not then lexical forms shared exclusively with other Sulawesi languages must result from borrowing (as Mills points out).

The number of loan words from non-Austronesian sources found in the North Sulawesi languages, even in basic vocabulary, hints at the extent to which borrowing has occurred.

The word for 'water' in PSan, *ake, is from a non-Austronesian North Halmaheran language (cf. Ternate aki, Galela ake). Occurrence of the form in all Sangiric languages but not surrounding languages, except for Tsw, which has borrowed considerably from adjacent Rth, enables the borrowing to be dated to PSan.

The word for 'dog' in Rth and Ban, kapuna, and southern Sangir dialects, kapuna?, is from a non-Austronesian source, being wide-spread in Moluccan and West Papua Phylum languages, e.g. Waioli kauna, Kilmuri kafuna. Here also the borrowing is an ancient one, if not as early as

PSan. The borrowing also occurs in Gtl, apula, but apparently not in other GM languages.

Many items dealing with kingship and administration were borrowed from Ternate, which had strong political influence in northern Sulawesi in the sixteenth century. For instance, Bul, Gtl (Limboto dialect), Kdp, Mdw saṅadi 'village head', from Ternate saṅaji; all GM languages except Gtl, bobato, in the meaning 'customary leader' or 'king's minister' from Ternate bobato; Bul, Kdp buki, 'woman of nobility', Mdw boke? 'princess' from Ternate bukiq.

If borrowing of words from non-Austronesian languages has been going on for a long time, clearly borrowing from Austronesian languages, both Sulawesi and Philippine, has also, although in such cases identification of forms as borrowings may be difficult or even impossible.

At the borders of the North Sulawesi languages borrowing has occurred with adjacent languages. Thus Bul shares numerous items with languages of the Tomini group, on which it borders at the western extreme of the North Sulawesi area, which apparently occur in no other North Sulawesi languages, e.g. Bul diuk, Tolitoli deuk 'dog'; Bul tanuk (replacing *sunay), Tolitoli tanduk 'horn'.

At the other geographical extreme, Snl shares, sometimes in common with San, numerous items with nearby Mindanao languages, which occur in no other Sangiric languages, e.g. Snl wilang 'to count' (replacing PSan *tiap), common in Mindanao languages; Snl utuḱ? 'brain' (the inherited uta? having shifted in meaning to 'hair'), frequent in Mindanao languages; Snl, San uba? 'monkey', related to ubal in various Mindanao languages.

Borrowing at a greater distance also occurs. Sometimes it is obvious, as with the semantic shift of PAN *ṅaṅa 'agape' to 'mouth', shared by the Gorontaloic languages with numerous Central Sulawesi languages, including Tomini and some Kaili-Pamona languages.

But it is not always immediately possible to identify shared forms as being the result of borrowing. Some items occurring in North Sulawesi languages and in other Sulawesi languages, but apparently not elsewhere are: the GM word for 'belly, stomach', reconstructed as PGM *kompon, which occurs widely in Central Sulawesi, in Tomini languages, as well as in Pamona, More, Balantak and so on; Tdn, Tbl kalipo?po? 'butterfly' and Wolio kalimpopo 'star', kalikalimpopo 'firefly'; PGM, PSan *kusay, PMin *kuse 'cuscus', as well as PSS *kuse, Pamona kuse; PSan *limbuṅ, Mdw, Dampalas (and other Tomini languages), Balantak, Andio limbuṅ 'round', with PSS *limbo 'gather around' also clearly related; PSS *libu(k)an, PGM *libuton 'island'; PGM lipu?, PSS, Wolio, Pamona lipu 'country, land, place'; PMin, PGM *paṅi, Uma, Pamona, Sa'dan, Mak paṅi 'tree with edible leaves (Pangium edule)'; PMin *r₂ondor₂, Bobongko rondor 'straight'; PMin, PGM *laga?, Uma laga? 'large red ants which live in trees', Parigi, Kaili laga?, Togian laga 'red ant'.

Two widespread forms are enclitics. Although these are bound forms they are mentioned here rather than under morphology as they are similar in some ways to free lexemes and function at the clause level rather than the word level.

One is a perfective marker, which generally indicates that an action has occurred, usually translatable 'already': PMin *-mo, Wolio, Sa'dan -mo. Anceaux (1952:47) writes of this form in Wolio: '-mo ... often denotes sure, ascertained facts and accordingly it often has the meaning 'already' ... an imperative with -mo denotes a stringent command.' This agrees very closely with its range of functions in the Minahasan languages. Michael Martens (personal communication) reconstructs *-mo 'perfective (already)' for Proto-Kaili-Pamona.

Most Minahasan languages contain an enclitic, reconstructed as *-pe? for PMin, which is an imperfective marker, indicating that an action is still in progress, translatable as 'still, yet'. It also indicates that an action precedes another, translatable 'first'. In imperatives it is a softener, translatable 'please'. Uma has two forms, -pe?, which functions like the identical Minahasan form in softening imperatives, translatable 'please', and -pi 'still, yet'. The Wolio enclitic -po has the same range of functions as Minahasan -pe?, while Sa'dan -po means 'still, yet'. Information for the GM languages is limited; however, Bul has -po 'still',

while Gtl has -po 'first'.¹² While these forms clearly do not represent chance similarities, differences in form and meaning prevent their being identified as direct cognates at this stage.

3.4 Summing up

The widespread occurrence of such forms as the above in Sulawesi languages suggests a common inheritance. However, if there were a Sulawesi 'super-group' then given the geographical proximity of the languages many more uniquely shared lexical items could be expected than have been recorded. Mills (1981:60) comments that there are very few North Sulawesi-South Sulawesi cognates which do not reflect already known etyma (i.e. which are not reconstructible at a higher level).

Furthermore, the North Sulawesi languages share many lexical items with Philippine languages which do not occur elsewhere in Sulawesi (e.g. see Zorc 1986).¹³

Also on the phonological and morphological evidence available the North Sulawesi languages are closer to the Philippine languages than to the other languages of Sulawesi. Thus, bearing in mind that lexical evidence for grouping is much less reliable than phonological or grammatical evidence, because of the greater prevalence of lexical borrowing, what lexical evidence there is for a Sulawesi 'supergroup' incorporating the North Sulawesi languages, in the absence of phonological and grammatical support, must be rejected.

4 CONNECTIONS BETWEEN THE THREE MICROGROUPS

4.0 Background

Only one claim has been made for a close connection between any two of the North Sulawesi microgroups. Sneddon (1978:10) suggested the Minahasan languages would prove to have their closest links with the Sangiric languages, as coordinate branches of a larger group, although offering no support for this. Later (Sneddon 1984:11-12) he stated that the two groups appeared to be immediately connected 'although this has yet to be established by a systematic comparative study'. He offered as evidence a list of lexical items reconstructed for PMin and PSan which were presumed to be shared innovations because 'with a few exceptions they refer to non-cultural items and are unlikely to be borrowings'.

However, some of the words on the list are now known to have external cognates, not recognised at that time. Thus PMin *səlet, PSan *səlet 'insert between two things' (with San səle? 'to stick between, insert; food which sticks between teeth'), cf. Balinese səlet-an 'remnants of food between teeth'; PMin, PSan *pədam 'feel', cf. PGM *pədam 'feeling'; PMin *təpeh, PSan *təpiR 'mat', cf. Mak təpəre?, Mandar, Sa'dan təpəre, PMP *təpiR 'mat'.

Although on the list there are certainly a number of forms unique to these two microgroups it cannot be assumed they reflect shared innovations; due to long contact between the groups borrowing of innovations much account for many of the items. This is taken up again below.

One item regarded as offering particularly high quality evidence for a grouping was Min *lehe?, PSan *leRe? 'neck', with shared metathesis of the last two consonants of PAN *ligeR. The evidence for the PSan reconstruction given in Sneddon 1984 was: San lehe?, Sni rerə?, Tal ulekka (with u explained as a fossilised prefix), Ban ehe? (with loss of l in this context being common). No cognate occurs in Rth.

However, the PSan form is probably an incorrect reconstruction; it now seems more likely that it was actually *leR.

PSan *R usually became Tal ž, e.g. *Ramut > žamutta. It became Tal k (i) after ə, which later became a, and (ii) finally, with subsequent development of paragogic a. In both situations k became geminate, except

under some circumstances not relevant here. The expected Tal reflex of *leRe? would be **(u)leže. The occurring form was explained as resulting from irregular loss of final syllable *e? (which appears to have been lost from a few other words) with subsequent addition of a paragogic vowel. Thus *leRe? > *leR > *(u)lek > ulekka. However, appeal to loss of *e? would not be necessary if the PSan form itself was recognised as ending in *R. Recognition of PSan *leR would allow derivation of the Tal form without any irregularity.

In Snl final ə? instead of expected e? was unexplained. But if the PSan form ended in *R then Snl final ə? would be the normal, expected paragogic syllable.

In Ban the paragogic vowel assimilates to the preceding vowel, e.g. PSan *likud > Ban likudu? 'back'. Thus an earlier *leR would regularly develop into Ban lehe?, just as would earlier *leRe?.

In the Tahulandang dialect of San the paragogic syllable is i?, not ə? as in the other San dialects, e.g. Tahulandang likuri?, Manganitu dialect likudə? 'back'. The Tahulandang form (not known at the time of writing Sneddon 1984) is lehi?, again pointing to a paragogic vowel following earlier final h (< *R).

This leaves only (the major dialects of) San lehe? with unexplained e?, instead of ə?, as an apparently irregular form.

Concerning the reduction of two syllable PAN *liqeR to one syllable PSan *leR, medial PAN *q was regularly lost in PSan, which would have left *liəR. Directly following another vowel *ə was lost in all Sangiric languages except San. This would have left *liR or *leR, which would explain the forms in all languages except San. Alternatively, the sequence iə has not otherwise been reconstructed for PSan and it is possible that e results from assimilation of i + ə, instances of reduction of vowel sequences being not uncommon in Sangiric languages (Sneddon 1984:49). Certainly a PSan reconstruction *liəR, later *leR, is thus a very strong probability and the case for shared metathesis with PMin collapses.

Thus the evidence presented in Sneddon 1984 for a Minahasan-Sangiric link is very much weakened. No further evidence for such a link has been discovered, and in view of further evidence for borrowing discussed below, the hypothesis must be abandoned.

4.1 Grammatical evidence

It has been mentioned that the three microgroups share Philippine-like verbal systems, absent from the rest of Sulawesi. However, there is nothing known that would distinguish them as a group from other Philippine-type languages. While a number of differences between the groups can be pointed out (e.g. infix -um- functions purely as an active voice marker in Minahasan languages, while occurring only with certain classes of verbs in languages of the other groups; *-an and *-ən have merged in the Sangiric languages), at present no uniquely shared grammatical features of the languages are known. Nevertheless, it must be added that more knowledge of the grammars of most North Sulawesi languages is required, followed by closer comparative study.

4.2 Lexical evidence

Numerous lexical items are shared, perhaps uniquely, among the North Sulawesi languages. A few of these, including some shared irregular developments from earlier forms, are: PMin *akəl, PSan *akel, PGM *akol 'sugar-palm (*Arenga saccharifera*)'
 PMin, PSan, PGM *uala 'canine tooth'
 PMin, PSan, PGM *undam 'medicine'
 PMin, PSan *kumi, Mdw kumi 'moustache' (data from other GM languages is lacking)
 PMin, PSan, PGM *lutam 'shoot'
 PMin, PSan *pəndam 'feel', PGM *pondam 'feeling'
 PMin *rentek 'work metal, forge', San hente? 'heated; melt (of iron)',

Mdw rente? 'very hot (of sun, glowing iron)' (data from other Sangiric and GM languages is not available).
PGM *watok₁₄ San uata?, Ttb tolo-atək, Tbl, Tse lolo-atək 'step(-father etc.)'.

However, there is risk in linking the three microgroups solely on such lexical evidence. First, it is always possible that outside cognates occur, as yet unnoted. Secondly, lexical borrowing is so prevalent that caution must be taken in ruling out borrowing for any item. The following is a good example: the form *buṅaŋ 'flower' looks like a shared innovation and has been reconstructed for the parent languages of all three microgroups. But it must now be treated as at least partially spread by borrowing. The word buṅa, unaltered in form from PAN *buṅa 'flower', also occurs in some languages, although with a change in meaning, e.g. San, Llk, Mdw buṅaŋ 'flower', beside San buṅa 'mushroom, fungus', Llk buṅa 'fruit', Mdw buṅa-i 'fruit'. In Gtl the form buṅa 'flower' is a borrowing, having irregular initial b and final a. This probably reflects a form borrowed as *buṅaŋ, loss of final ŋ occurring in a number of Gtl words. Directly inherited huṅo means 'fungus; fruit'. In Kdp also buṅaŋo 'flower' is a borrowing, with irregular initial b, beside directly inherited vuṅa 'fruit'. The evidence of Gtl and Kdp strongly points to buṅaŋ having been spread by diffusion and it cannot be regarded as reflecting an innovation in a common parent language of the three microgroups.

Borrowing between the North Sulawesi languages and other Sulawesi groups is discussed in Section 3. Here evidence for borrowing between different North Sulawesi microgroups is considered.

Borrowing between languages of different microgroups is, as would be expected, especially common among languages sharing common borders. Tsw has borrowed heavily from Mdw. Many borrowings are identified by their incorrect reflexes of PAN etyma. The following items are regular in Mdw but contain irregular g for expected h in Tsw: Mdw, Tsw aog 'bamboo' (PAN *qauR); Mdw, Tsw ibog 'saliva' (PAN *iber). Tsw has also borrowed from Rth. Thus Tsw ucah 'hair' from Rth utak (reflecting a Sangiric semantic innovation, cf. Tsw ucəh 'brain' < PMin *utək); Tsw ahe 'water' from Rth ake (reflecting a PSan borrowing from a non-Austronesian source).

Borrowing in both directions has occurred between Ban and adjacent northern Minahasan languages, especially Tbl, e.g. Tbl, Tse oyow 'swim' from Ban hoyow (PSan *Royaw); Tbl ule?, Tse ude? 'snake' from Ban ule?; Ban leoso 'meat' from Tbl, Tse le?os 'good (part)'; Ban kalipopo 'butterfly' from Tbl, Tdn kalipo?po?.

Rth has borrowed heavily from Minahasan languages, e.g. Rth unto?, PMin *ənto? 'live'; Rth, Minahasan languages karis 'streak, scratch' (Rth r directly reflects only earlier *d and so not the *r of PAN *karis). Rth has also borrowed from Mdw, e.g. Mdw, Rth loben 'big' (the word is often regularised to Rth lowen, intervocalic b not being part of the Rth phonological system); Mdw, Rth puliŋ 'full'.

Borrowing at a greater distance presents more of a problem as it is more likely, in the absence of phonological irregularities, to be taken as evidence for genetic relatedness.

Sometimes borrowing at a distance is obvious because, like borrowing between adjacent languages, it occurs in one language but not in other languages of the same microgroup. Thus Ban has a number of items in common with Mdw which reflect PGM reconstructions but which occur in no other Sangiric language, e.g. Mdw, Ban natu? 'egg'; Mdw ulid, Ban ulidi? 'lie down'. At least one Ban-speaking village occurs in Bolaang-Mongondow district and Bantik people may once have been in close contact with the Bolaang-Mongondow kingdom.

Usually, however, borrowing is only obvious when there are phonological indicators. San has a number of items with phonological irregularities pointing to borrowing from a GM source, e.g. San owolə? 'burn and send up a lot of smoke etc.', cf. Psk owol, Kdp, Atg obulo 'smoke', Gtl wobulo 'billow (of smoke)' < PPh *qebel 'smoke'; San buro? 'albino', as well as Rth wuro? and Tsw budo? 'white', cf. Mdw budo? 'white', Kdp vudo? 'albino' < PAN *budeq 'white'. In these two cases the occurrence of o in San, instead of ə or e, indicates borrowing from a GM source. Likewise Tdn, Tbl roŋit 'mosquito' is from a GM source, cf. Mdw yoŋit, Kdp, Bnt hoŋito < PAN

*re it.

Some borrowing is ancient. PSan *togas 'hard, strong', reflected in all Sangiric languages, beside directly inherited *tiRas 'hard' < PAN *teRas, must be from a GM source, cf. PGM *togas, directly reflecting the PAN form. San bogasə?, Tal bohassa 'husked rice' shows the same borrowing pattern and may well have been borrowed from the same source at the same time, cf. PGM *bogas. Here also a regular reflex occurs: PSan *biRas 'husked rice' < PAN *beRas. For this item, however, absence of cognates in the other Sangiric languages prevents the borrowing being assigned to the PSan period.

PMin *ipag 'brother/sister-in-law', for expected **ipah, is also an old borrowing, probably from PGM *ipag, reflecting PAN *sipar. San ipagə?, Tal ipaga also look like borrowings from the same source. However, the borrowing cannot be assigned to PSan; Rth, Ban ipa? reflect a separate borrowing, apparently from Mdw ipa?, which is the result of an irregular innovation.

Further evidence for borrowing between Minahasan and Sangiric languages, previously undetected, comes from problems with certain sound correspondences. The correspondence PMin *h - PSan *R reflects PAN, PPh *R, e.g. PMin *daha?, PSan *daRa 'blood' < PAN *daRaQ. The correspondence PMin *r - PSan *R reflects PAN *r, e.g. PMin *ribu, PSan *Ribu < PAN *ribu 'thousand'. However, in the modern languages the correspondence Minahasan languages r - San, Ban h frequently occurs, where there is a PAN or PPh reconstruction with *R. Such cases probably always result from borrowing.

Since PAN *R had become PMin *h, any borrowings by Minahasan languages of Sangiric words with *R would replace this with the nearest equivalent, which was r. Thus we can expect a Minahasan r - San, Ban h correspondence in words borrowed by Minahasan languages from Sangiric languages at a time before *R became San, Bah h (see Section 4.3b below).

One important instance is Ttb, Tdn, Tbl, Tse wəru 'new', where r is not a correct reflex of *R in PAN *bageRu, PPh *baqRu (nor is a correct reflex of *a). This can be treated as a borrowing from a Sangiric source (PSan *bəRu, which correctly reflects the earlier form), although PMin *beru was reconstructed in Sneddon 1978.¹⁵

Another example which must be treated as a borrowing is Tsw lahuc, North Minahasan languages rakut 'bind, tie up; bundle', cf. PSan *Rakut 'tie up, tie together' < PAN *Raku(Ct) 'tie, fasten', PPh *Rakut 'bind; bundle'. This was reconstructed as PMin *r₂akut in Sneddon 1978; PMin *ahkut 'tie, bind with rope', which correctly reflects the PAN form, was not yet reconstructed at that time.

Words in Minahasan languages with r corresponding to PAN/PPh *R, such as the above, were queried in Sneddon 1978 (pp.15-16) but were nevertheless reconstructed for PMin. These must all now be regarded as probable borrowings from a Sangiric source.

Borrowing by Sangiric languages also has resulted in r - h correspondences where they would not be expected. Some Ban and San words with h are borrowings of words with r in European languages or Malay; such borrowings probably occurred at a time before *R became h in Ban and San and before d developed an r variant intervocalically in San. Thus in the absence of an r phoneme, r in borrowed words was replaced by the nearest equivalent, *R, which subsequently became h e.g. San gaheda 'church' from Portuguese, probably via Malay gareja; San kasahə? 'rough, coarse', from Malay kasar; Ban hekeŋ 'count' from Dutch reken; Ban kahatasa? 'paper' from Arabic via Malay kertas.

These clearly establish that borrowings from an 'r source' can have h in present-day Ban and San and justify recognising many items in Ban as loans from Minahasan languages. Thus Ban kuhe? 'cooking pot' from one of the Minahasan languages, which all have kure?, via earlier Ban *kuRe?, not from PSan *kudin; Ban louhu? 'lake' from a Minahasan language, cf. Tdn, Ttb lour 'lake', beside laodo? 'ocean' < PSan *laud. Sneddon 1984 reconstructs PSan *seRam 'ant' with reflexes Ban sahaŋ, Rth saam (19th century saham) reflecting PPh *sejem. This must now be treated as a borrowing (cf. PMin *serəm), presumably occurring before Ban and Rth separated. The directly inherited form would be **sədum, occurrence of a vowel other than u before m in the final syllable strongly supporting its

status as a borrowing. The limited amount of lexical data available for Ban suggest that the extent of such borrowing is quite large.

There are a great many related words where the Minahasan languages have r and San has h. Considering the limited number of PAN reconstructions with *r many of these items are unlikely to be directly inherited in both groups. But in the absence of other evidence (such as other phonological irregularities or a PAN etymon with *r) it is not possible to clearly identify an item as a borrowing, let alone state in which direction borrowing has occurred, since Minahasan borrowings from a Sangiric source will reflect *R as r while Sangiric (San and/or Ban) borrowings from a Minahasan source will reflect Minahasan r, from earlier *r or *d, as h, each resulting in an r - h correspondence. These correspondences hint at the extent of lexical borrowing between Minahasan and Sangiric languages, while highlighting the difficulty in positively identifying particular instances of it.

4.3 Phonological evidence

If close genetic relationship exists between any two of the North Sulawesi microgroups we could expect this to be reflected in one or more shared phonological innovations. In Section 2 are listed some of the innovations which characterise the groups and help distinguish them from one another. Here reflexes of a number of sounds and sound sequences are examined to see if they provide any worthwhile evidence for a linking of any of the microgroups.

a. Reflexes of PAN *d, *D, *z, *Z, *j.
According to Charles (1974:480), PAN *d, *D, *z, *Z merged in PPh *d. He states that apparently all languages descending from PPh, except Cordilleran languages of northern Luzon, have merged PPh *d and *j in a voiced apical obstruent. However, there are difficulties for recognising the reflexes in the Minahasan languages as developments of a single phoneme at an earlier stage. Here the various PAN phonemes are looked at separately for each microgroup.

(i) Reflexes in PMin

Data in this section are drawn from the PMin word list in Sneddon 1978. PMin *r₁ was reconstructed on the basis of the occurrence of r in all Minahasan languages. The reconstruction *r₂ was made on the basis of l in Tsw and the Makelai dialect of Ttb and r elsewhere. Where evidence from Tsw or Makelai was lacking, r (without subscript number) was reconstructed, as an abbreviation for (*r₁ or *r₂).

PAN *z and *Z are always reflected as PMin *d, although the number of recorded examples, especially for *z, is not great, e.g. PAN *qazay 'jaw' > PMin *ade, PAN *ke(zZ)ut 'pinch' > PMin *kedut, PAN *ta(zZ)em 'sharp' > PMin tadəm, PAN *quZan 'rain' > PMin *udan, PAN *Zalan 'road' > PMin *dalan.

PAN *D is reflected as *r or *r₂ after a vowel other than *ə and in a few items initially: PAN *siDa 'they' > *sir₂a, PAN *Da-DaRa 'maiden, young girl' > PMin raraha. Otherwise it is reflected as *d initially, finally and medially after ə: PAN *DuRi 'thorn' > PMin *duhi, PAN *likuD 'back' > PMin *likud, PAN *seDaq 'fish' > PMin *səda?. While there is not complete complementarity in distribution, *r/r₂ and *d both occurring initially, the general rule is that *r/r₂ occur intervocally, except after *ə, while *d occurs elsewhere. There are a very few exceptions: PAN *bayaD 'pay' > PMin *baer, PAN *teDug 'shelter' > PMin *teruŋ. Reflex *r₁ occurs in PAN *kuDen 'cooking pot' > PMin *kur₁e? (although borrowing may be involved here, considering the irregular final consonant).

PAN *d is reflected as *r or *r₁ initially, medially and possibly finally, although few examples have been recorded in any of these environments: PAN *dadaŋ 'to heat' > PMin *r₁ar₁aŋ, PAN *ubu(dj) 'palm heart' (ambiguous for *d and *j) > PMin *ubur. There are no recorded examples of *d becoming *r₂. Usually *d became *d in all positions, although there are very few recorded cases medially: PAN *dilaq 'tongue' > PMin *dila?, PAN *ludASaq 'spit' > PMin *luda?, PAN *tumid 'heel' > PMin *tumid.

PAN *j became PMin *d finally and following *ə: PAN *gapejuSu 'gall, bile' > PMin *apədu, PAN *pusej 'navel' > PMin *pusəd, PAN *palaj 'palm' > PMin *palad. Medially, after a vowel other than *ə, it became *r₂: PAN *ajej 'charcoal' > PMin *ar₂əj, PAN *ijuSuŋ 'nose' > PMin *ŋir₂uŋ. In three items it became *r. Although Tsw/Makelai evidence is lacking, *r in these items can now be refined to *r₂, on the assumption that PAN *j > PMin *r₂ regularly, as no cases of *j > *r₁ are known, e.g. PAN *pija > PMin (Sneddon 1978) *pira 'how many', now *pir₂a.¹⁶

In summary, PAN *z and *Z are always reflected as PMin *d. PAN *D and *j have *r/r₂ reflexes intervocalically, except after *ə, and *d elsewhere, although *j is the more regular in its reflexes. Neither is reflected as *r₁, except in one known instance. PAN *d tends to be reflected as *d in all environments, although some cases of *r/r₁ occur. It is not reflected as *r₂.

Charles states that PAN *d, *D, *z, *Z merged in PPh *d and that PPh *d and *j subsequently merged everywhere but in Cordilleran languages. While some amount of variation in the PMin reflexes would not be surprising (because of factors such as pre-PMIn borrowing, together with some erroneous reconstructions because of factors such as undetected post-PMIn borrowing from external sources or among the Minahasan languages themselves), one would expect such variation to be minor. But the differences in reflexes here are quite marked.

Nevertheless, it is premature at this stage to conclude that PMin retains distinctions lost from PPh, together with all that this would entail. But a detailed reexamination of the situation in the Minahasan languages is certainly called for as a priority.

(ii) Reflexes in PSan

PAN *d, *D, *z, *Z and *j are all reflected as PSan *d, e.g. PAN *qazay 'jaw' > PSan *aday 'chin', PAN *ZaRum 'needle' > PSan *daRum, PAN *DuRi 'bone' > PSan *duRi, PAN *bukid 'hill' > PSan *bukid, PAN *qulej 'worm' > PSan *ulid.

(iii) Reflexes in PGM

The array of PGM phonemes reflecting the PAN phonemes has not yet been satisfactorily accounted for and represents the most difficult problem in the reconstruction of PGM segmental phonology. As this subject has had no more than fleeting mention in previous publications it is worth considering in some detail here.

In this context it is convenient to make a primary distinction between PGM *d and continuants. The difficulty in reconstructing the PGM continuants results from the large number of correspondence sets occurring in the languages and the lack of correlation between these sets and PAN phonemes.

In the following discussion the GM languages can conveniently be assigned to five groups on the basis of regular correspondences: (Group 1) Bul, Gtl; (2) Kdp, Sww, Blg, Bnt; (3) Llk; (4) Mdw; (5) Psk. Members of a group consistently show the same reflexes as one another except in a few cases where borrowing is presumably involved.¹⁷

Below are presented five correspondence sets, (a) to (e). The number of occurrences of that correspondence set in a comparative word list of more than 800 items (Usup 1986) is indicated. A number of items in this corpus lack cognates in all languages; those where the incompleteness prevented unequivocal assignment to a particular correspondence set are not counted here.

Set a: (1) l, (2) h, (3,4) r, (5) h (50 occurrences), e.g. (1) Bul gulaŋ, Gtl hula, (2) Kdp, Blg guhāŋo, Sww guha, Bnt kuhāŋo, (3) Llk guraŋ, (4) Mdw guraŋ, (5) Psk guhaŋ (PAN *guDaŋ 'old, adult').

Set b: (1) l, (2,3) h, (4,5) y (13 occurrences), e.g. (1) Bul tolom, Gtl tolomo, (2) Kdp tohomō (with irregular initial t), Sww toho, Bnt sohomo, (3) Llk sohom, (4) Mdw toyom, (5) Psk soyom 'ant' (PAN *sejem).

Set c: (1) l, (2,3,4,5) y (11 occurrences), e.g. (1) Bul, Gtl polu, (2) Kdp peo, Sww peu, Blg, Bnt poyu, (3) Llk opoyu, (4) Mdw opoyu, (5) Psk opoyu 'gall' (PAN *qa(N)pejuSu).

Set d: (all groups) y (14 occurrences), e.g. (1) Bul layaŋ, Gtl layaŋo, (2) Kdp layaŋo, (3) Llk layaŋ, (4) Mdw layaŋ 'throw away' (PAN *layaŋ 'fly, hover').

Set e: (1) l, (2) y, (3) h, (4,5) y (12 occurrences), e.g. (1) Bul, Gtl

lopo, (2) Sww yo:pa, Kdp, Bnt, Blg yopa, (3) Llk hopa, (4) Mdw yopa, (5) Psk yopa 'fathom' (PAN *Depa). A number of other correspondence sets occur but with membership limited to one, two or three words.¹⁸ In some words earlier *ayV or *oyV has developed into eV sporadically among the languages, as in the Kdp and Sww words in set (c) above. In such cases, and also in other sequences like iu, phonetic [y] is interpreted as non-phonemic.

In Mdw doublets sometimes occur, one item with r the other with y, e.g. darag, dayag 'yellow', pura, puya 'red', or r and phonemic zero where the preceding vowel is i, e.g. guru, niu 'nose'. Blust (1983) discusses this phenomenon, pointing out that such variants generally reflect earlier forms with *r. But also, in what Blust argues is a result of hypercorrection, some forms with [y] (either /y/ or Ø) have developed r variants, e.g. pia, pira 'good' from earlier *pia.

Such variation in Mdw forms is not considered above because cognates in the other languages indicate which is the original Mdw form.

A considerable amount of overlap occurs in the above sets such that no set is entirely different in membership from any other set; no language has more than two phonemes occurring in these sets, except Llk, which has three.¹⁹

Nevertheless, it is possible to make a basic distinction between set (a) and the other sets. Set (a) correlates with Pan/PPh *r; with the exception of PAN *regit 'mosquito', which is reflected by set (b), all instances in the corpus of earlier *r, such as in PAN *ribut 'storm', *buruk 'rotten', are reflected by set (a). Set (a) also differs from the other sets in that it occurs word-finally (ignoring the paragogic vowel in the Gorontalic languages), e.g. Bul gugul, Gtl huhulo ('cold'), Kdp, Sww, Blg guguho, Llk, Mdw gugur 'shake' < PWMP *gurgur. This correspondence set is recognised as reflecting a PGM phoneme *r. Thus for the above cognate set PGM *gugur is reconstructed.

Although sets (b) to (e) occur in smaller numbers of words, together they account for half the correspondences which could be assigned to a particular set. Recognising sets (b) to (e) as each reflecting a separate PGM phoneme would require reconstruction of four phonemes. But with the separation of set (a), reflecting PGM *r, these sets involve only two phonemes in each language, except in Mdw and Psk where they involve only one, y. In Bul and Gtl one of the two phonemes is l, which is also the regular reflex, as it is in all GM languages, of another PGM phoneme, *l. The only correlation between these sets and PAN phonemes is that set (d) is the regular reflex of PAN *y, as in the example given above.

In this paper the problem of how to account for the various correspondence sets in PGM can be avoided, although this is hardly a satisfactory solution; probably the best solution, at least provisionally, would be the setting up of subscript forms *y₁, *y₂ etc. to account for these correspondence sets.

Reflexes of the PAN phonemes can now be examined. As no PGM phonemes are set up, except for *r, continuant reflexes are here identified by the sets (a) to (e) and reflexes in selected modern languages are given.

PAN/PPh *j became PGM *d finally: PAN *pusej 'navel' > PGM *pusod, PPh *silaj 'palm sp. (Corypha sp.)' > PGM *silad. Medially *j is reflected by a continuant, e.g. PAN *bujag 'foam' is reflected by set (a), e.g. Gtl lolombula, Kdp vuha, Llk, Mdw bura?, PPh *sejem 'ant' by set (b) (reflexes given above) and PAN *qa(N)pejuSu 'gall, bile' by set (c) (reflexes given above).

PAN *D became PGM *d finally and usually also initially: PAN *Dalem 'deep' > PGM *dalom, PAN *tugeD 'stump' > PGM *tu'od. It became a continuant medially and sometimes initially: PAN *Depa 'fathom' > set (e): Gtl lopo, Kdp yopa, Llk hopa, Mdw yopa; PAN *seDaq 'fish' > set (c): Gtl tola ('pike fish'), Kdp sea, Llk sea?, Mdw toya?, PAN *puDa 'red' > set (a): Gtl pulo ('whitish'), Kdp puha, Llk, Mdw pura, Psk puha. A *d reflex occurs in PAN *se(Dd)u 'hiccup' (ambiguous for *D and *d) > PGM *sodu.

PAN *d is reflected as *d initially and finally: PAN *damaR 'resin' > PGM *damag, PAN *bukid 'mountain' > PGM *bukid. Reflexes of PAN words with medial *d are too few to allow a clear statement. It is reflected as *d in PAN *budeq 'white' > Mdw budo?, Psk buro? (no cognates known; Psk r reflects earlier *d) and as set (b) continuants in PAN *ludASaq 'spit' >

Gtl lula, Sww duha, Llk duha?, Mdw duya?.

PAN *D and *d are reflected as PGM intervocalic *d where this results from loss of the preceding consonant in an RM: PAN *DapDap 'tree sp.' > PGM *dodap, PAN *diqdiq 'boil' > PGM *didi?.

PAN *Z and *z have only been recorded in a small number of items. PAN *z is only noted once in the corpus: PAN *azay 'jaw' > set (b): Gtl wale ('temple'), Kdp, Llk ahe, Mdw ayoy. PAN *Z is reflected as *d initially, e.g. PAN *Zalan 'road' > PGM *dalan. There is one exception in the corpus: PAN *Zaget 'bad' > set (a): Gtl leeto, Llk ra?at, Mdw ra?at, Psk ha?at. It became a continuant medially: PAN *quZan 'rain' > set (b): Bul ulan, Kdp uha, Llk uha, Mdy uyan.

In general then the PAN consonants are reflected finally and usually initially as *d. They are reflected as a continuant intervocalically and sometimes initially.²⁰ In a very limited number of cases in the corpus the medial reflex is *d.

As distinct from PMin, where *j and *D are reflected medially as *d after *a and as a continuant (*r/r₂) after other vowels, no such distinction occurs in the GM languages; cf. the reflexes of PAN *se(Dd)uq and *seDag cited above, the former with *d, the latter with a continuant.

There is therefore no evidence that PGM maintained any of the PAN distinctions; like PSan it reflects the PPh merger of these sounds in *d. This phoneme later, apparently prior to PGM, developed continuant allophones intervocalically. Sometimes merger with the reflexes of PAN *r and *y occurred, sometimes not, and apparently by the time of PGM there were several contrasting continuant phonemes.

Summarising this section it can be said that while the reflexes of PAN *d, *D, *z, *Z and *j point to a remoteness of the Minahasan microgroup from the Philippine languages, their merger in *d offers no evidence for a GM-Sangiric link as it merely reflects changes at an earlier stage in the Philippine group.

b. Reflexes of PAN *R

The reconstructed reflexes are PGM *g, PMin *h and PSan *R. Thus there is no immediate evidence here for linking any two of the microgroups.

One uncertainty is the phonetic nature of PSan *R. The Sangiric dialects are distinguished by a northern group which have apical reflexes and a southern group which have h.

The northern group includes Tal, where the reflex in most environments is z, a voiced retroflexed fricative. Snl and the Taruna dialect of San have r. In Snl this is phonetically [r^g], an alveolar flap with simultaneous velar friction, thus having both apical and dorsal features.

In southern San dialects, such as Manganitu, Tabukang and Tahulandang, and in Ban, h occurs. In Rth *R was reflected as h until recently, as shown by 19th century word lists. This has since been lost.

As the phonetic nature of *R in PSan was not known, the symbol <R> was chosen in Sneddon 1984 as it was a continuation of the phoneme represented by this symbol in PAN and PPh.²¹

Sneddon (1984:40) suggests that PSan *R is unlikely to have been an alveolar r, which occurs as a reflex of PAN *R in some other Sulawesi languages, such as languages of the South Sulawesi group (Mills 1975), the occurrence of a reflex h in southern dialects arguing against this.

One problem with this argument is that PSan *R represents a merger of PAN *R and *r. Thus, since h is the reflex of PAN *r in these languages, e.g. PAN *ribu 'thousand' > San hiwu, Ban hibu, there is no reason why it cannot be the reflex of a PSan [r]. Further, there are other examples of h developing from earlier r. Thus Kdp, Sww, Blg, Bnt have h as a reflex of PGM *r and earlier *r has become h in some dialects of Muna.

Nevertheless, the fact that PSan *R is a reflex of both apical *r and dorsal *R and that these two features are blended in its reflex in Snl (/r/ [r^g]), prevents any firm statement on its phonetic nature at this stage.

PMin *h is reflected as h in Tsw, ? in Ttb and h in Tbl except for one dialect which has ?. Nineteenth century word lists show Tse and Tdn to have had h, which has since been lost.

PGM *g represents a merger of PAN *R and *g. It is reflected as g in all languages except Gtl where it is h, Bnt where it is k and Psk where it

is h (via nineteenth century g) following a vowel and g elsewhere. Thus reflexes of PAN *R show a marked degree of instability, which highlights the difficulties for speculating about the history of their phonetic nature prior to the parent languages of the microgroups.

Consequently it is not possible to show that any two of PGM *g, PMin *h, PSan *R reflect a period of uniquely shared change in their development from PAN *R. However, merger of PAN *r and R in PSan *R and merger of PAN *g and *R in PGM *g almost certainly rule out any uniquely shared period of development, while the retention of separate *g and *r in PMin argues against any shared period with either of the other two.

c. Changes to high vowel+semivowel

Charles (1974:493) comments on the process: PPh high vowel (a front)+ semi-vowel (-a front) > vowel (-a front). He illustrates the process for San and Mdw but makes no comment on the possibility of this being a shared innovation or on whether it is unique to these languages.

His examples are, with additional data: (i) PPh *hiwaq > San ua 'eviscerate and dress a food animal or large fish', Mdw ua? 'wound (from knife or weapon)'. PSan *ua 'to disembowel' and PMin *ua? 'slice open, disembowel' have been reconstructed. Information on the GM languages is lacking, except for the Mdw item. Here the change from *iw to u has occurred in all three microgroups. (ii) PPh *liwan > San luaq 'to exchange, buy', Mdw luan 'to exchange'. PSan *luan 'exchange' has been reconstructed. No other reflexes are known.

Blust (1983:46) adds two further examples: (iii) earlier *guyud > (pre-Mdw) *giud > Mdw girud 'drag, pull'. PSan *Riud 'to pull', reflecting PPh *Ruyud, is reconstructed. No forms are known for the Minahasan languages. Blust comments that 'the contraction of original -uy- to -i- in Mdw ... must have occurred prior to r-insertion in girud since the change *u > i is otherwise unreported'. This is presented with problems by the forms in other GM languages: Llk girud, Psk gihur,²² Sww, Blg gihudo, Kdp gihuru, Bnt kihuro, Bul gilud. Apart from Gtl hiidu, which reflects giud,²³ all languages reflect PGM *girud (with set (a) correspondences), which is reconstructed by Usup (1986:310). Blust also gives: (iv) Mdw diug 'dugong' < PAN *duyuy. Known cognates are PMin *duyuy, San duluy (reflecting PSan *duyuy), in which the change did not occur.

In the GM languages some variation occurs between -uyu- and -iu- forms, e.g. Llk, Psk, Bul piuy, Kdp, Bnt piuy, Mdw puyuy 'hair bun' (< PAN *puyuy 'd₃ something around a centre'); Llk piut, Sww, Blg piuto, Kdp, Bnt piito,²³ Mdw piut,²⁴ puyut, Bul pulut, Gtl puluto 'pick up' (PPh *pu₂ut); Llk, Psk piuy, Mdw piuy (as well as piruy), Blg iu₂o, Bul ilu₂u, Gtl wulino (with metathesis), Kdp, Bnt uyuyo 'nose' (PAN *iju₂u₂); Llk, Psk iup, Mdw iup (as well as irup), Sww iupo, Gtl hiipo (with unexplained initial h), Kdp uyuyo (PAN *Seyuy).

Thus there are cases where earlier *uy(u) became i(u), in accordance with the process pointed out by Charles, but also cases of the reverse, both changes occurring in an apparently random fashion.

Because of the limited number of recorded examples of the phenomenon described by Charles, and the irregularity in its occurrence, these changes must be regarded at this stage as offering no evidence for a shared innovation.

d. Changes to consonant clusters

If the first member of a cluster in a repeated monosyllable (RM) was a nasal it assimilated to the following consonant in PSan and PGM, but not in PMin, e.g. PAN *kemkem 'handful' > PGM *ko₂kom, PSan *ke₂kum, PMin *ke₂mke₂m; PAN *dindin 'wall' > PGM *dindin, PSan *de₂din; PAN *Ti₂Ti₂ 'ring, clank' > PGM *tintin, PMin *ti₂ti₂.

However, this is extremely weak evidence for a grouping of the GM and Sangiric languages as the same change has occurred independently in numerous Western Austronesian languages. For instance, it occurred in Sasak and Sumbawa after their separation from Balinese, e.g. Sasak dindin, Sumbawa dinin (with regular loss of *d), Balinese dindin 'wall'. The same change occurred in Malay, e.g. dindin 'wall', and in the Barito languages of Kalimantan, e.g. Kahayan dindin, Maanyan rini₂ (with regular loss of

*d) 'wall'. Thus the chances of independent development in PGM and PSan are very high.

If the cluster in an RM was of two oral consonants the first was lost in PSan and PGM while again no change occurred in PMin: PPh *bulbul 'body hair' > PSan *bēbul, PGM *bubul, PMin *bulbul; PAN *kiskis 'scrape' > PSan *kəkis, PGM *kikis, PMin *kiskis.

Again the change is a common one in Western Austronesian languages. It occurred in Sasak and Sumbawa after their separation from Balinese: Sasak, Sumbawa kikis, Balinese kiskis 'to scrape'. It also occurred in the Barito languages: Maanyan sesep, Lawangan səsəp 'suck' < PAN *sepsep; Maanyan, Dusun Deyah kikit 'bite' < PAN *kitkit.

Thus while the GM and Sangiric languages underwent the same changes to consonant clusters listed here, while the Minahasan languages did not, the changes are common ones and there is a strong possibility they were independent developments.

Charles (1974:458) discusses medial clusters in PPh involving *r, *l and *h and the changes which occurred to them in a wide variety of languages.

These changes affected the North Sulawesi languages, usually by loss of the first consonant, e.g. PPh *arta 'slave' > PGM, PMin *ata; PPh *baqRu 'new' > PGM *bagu, PSan *baru.

However, the changes were not entirely regular. Thus in PMin *baʔkəs 'to tie', the *R in PPh *baRkes is reflected by *ʔ, not Ø. Sometimes a cluster-final *q was lost, e.g. PPh *beRqat 'heavy' > PMin *bahat (PGM *bogāt, PSan *bēRat), but in PPh *baRqaŋ 'molar' > PMin *baʔaŋ (PGM *bagaŋ, PSan bəŋaR (with metathesis)) cluster-initial *R was lost instead. In the case of PPh *aljaw 'day', replacement of the first consonant by a nasal occurred in PGM *ondow and PMin *əndo but not in PSan *əlaw. But the same change occurred in numerous other languages, e.g. Western Bukidnon Manobo gandew, Sindangan Subanon gendaw, and cannot be regarded as a shared GM-Minahasan innovation.

Because of the widespread distribution of such changes among Philippine languages and their somewhat unsystematic reflexes (see Charles 1974: 459-465 for examples) they presumably occurred early in the history of the Philippine languages and do not offer evidence for a grouping of the North Sulawesi microgroups. In fact, some of the changes occurred in other languages, such as Malay, and could represent a number of parallel developments.

e. Changes to vowels

The change of *a to *ə occurred in the first syllable of RMs in PGM and PSan but not in PMin, e.g. PAN DapDap 'tree sp.' > PSan *dədap, PGM *dodap (where *o reflects earlier *ə), PMin *dapdap.

Here also PSan and PGM appear to share an innovation which did not occur in PMin. However, a shared innovation can be discounted; it is actually part of a process of weakening of all vowels to *ə in the first syllable of RMs in PSan, while high vowels did not change in PGM: PAN *kiskis 'scrape' > PSan *kəkis but PGM *kikis; PPh *bulbul 'body hair' > PSan *bēbul but PGM *bubul.

In this section have been examined a number of sound changes which might offer evidence for a direct link between North Sulawesi microgroups. But in each case closer inspection has shown that any possible evidence is too weak to be seriously considered.

5 CONCLUSION

The North Sulawesi languages share numerous lexical items. But the evidence these provide for a close genetic link is weakened by the fact of geographical proximity and the amount of lexical borrowing which can be shown to result from this proximity. Further, lexical evidence for a close link must be supported by other evidence and this is lacking. Related languages in geographical proximity are more likely to retain shared features than languages isolated from one another. Thus the absence of any

known uniquely shared phonological or grammatical features is even more conspicuous.

Although there has not been a detailed comparison with other Sulawesi languages, what has emerged so far gives no support for an immediate link between any of the North Sulawesi microgroups and any other Sulawesi group. Possible general or widespread Sulawesi characteristics probably result from areal diffusion.

Charles, Zorc and others have presented evidence that the North Sulawesi languages belong in the Philippine group, and in general phonological, lexical and grammatical character there is nothing to suggest they do not derive from PPh, with the exception of the problem of PMin reflexes of *d, *D, *z, *Z and *j, which requires further study. The three microgroups are usually placed in a South Philippine branch but no suggestions have been offered as to which languages within that branch any of the three might tie in with.

Thus a close link between any two of the microgroups is rejected. No alternative classification has yet been offered but the evidence is that the search for close affinities must be directed northward, to the languages of the Philippines.

NOTES

1. The term microgroup is borrowed from Zorc (1982), who uses the terms microsubgroup and macro-subgroup. While the North Sulawesi languages fit neatly into separate microgroups, it is obvious that it will not always be possible to apply the concept comfortably, e.g. where extended dialect chaining occurs. There could also be disagreement on what constitutes a 'close relationship'; for instance, what lexicostatistical percentage would be required. Minimum scores between members of any North Sulawesi microgroup are above 40%.
2. The reconstruction is by Zorc (personal communication). Most PPh reconstructions are from Charles (1973, 1974 and in Zorc 1971).
3. Blust (1984:44) details exceptions to phonological change, confined to particular grammatical categories, particularly pronouns, in Rejang. He cites Bloomfield's (1933:362-4) assertion that sound change is unaffected by meaning. Blust states (42-3) 'The burden of proof that the Rejang pronouns do not constitute a meaning-based exception to phonological change rests with those who, like Bloomfield, would insist that such types of exception do not occur in natural language.' The phonological change in PSan pronouns appears to be an example of a sound change confined to a particular grammatical class, thereby constituting a sound change affected by meaning.
4. Sneddon (1984:57) gives San and Snl 82% cognates on the full lexicostatistical list. Llamzon and Martin (1976) and Walton (1979), using the northern Tabukang dialect of San presented in Reid (1971), give higher percentages, 88 and 90 respectively. The Sangil migrated to Mindanao from the Sangir Archipelago several hundred years ago and on purely linguistic grounds San and Snl constitute a series of dialects of a single language. However, on other grounds, especially political and social (including religious), they are treated as separate languages.
5. Bolango and Atinggola are spoken on opposite sides of the North Sulawesi peninsula but nevertheless are closely related dialects of one language. The Blg dialect is here chosen to represent this language, following Usup (1986).
6. Usup (1986:196) gives a lowest score of 37% for Gtl and Psk. For this study, using Usup's lists with very few changes, a Gtl-Psk score of 42% was obtained.

7. Numerous items can be reconstructed for PGM with antepenultimate *a, but very few of these items are known to have a PAN or PPh etymon. It is therefore likely that the change *a > *e (> *o) occurred in pre-PGM and that *a was later returned to this position in innovations and borrowings.

8. I am grateful to David Henley of the Australian National University who informed me of this, in considerably more detail than presented here. His studies show earlier migrations of the Atinggola-Bolango people and a possible link with the Lolak group.

9. Although assigning Kdp to the Gorontalic group, Adriani (Adriani and Kruijt, 1914:184) also stated that Kdp is closely related to Mdw and the Philippine language (p.193) and thus marks a transition between the Philippine and Gorontalic groups.

10. Manado Malay also appears to have commenced this process with loss of final t, e.g. saki (earlier sakit) 'sick', mulu (earlier mulut) 'mouth' and replacement of final m and n by ŋ (although not regularly), e.g. makaŋ 'eat' (earlier makan), malan 'night' (earlier malam).

11. The same borrowing occurs with the meaning 'crocodile' in Bada, kapuna, and Besoa, kapuna?, in Central Sulawesi.

12. The Bul, Gtl forms ambiguously reflect *-po or *-pa, cf. PAN *pa 'incompletive article', Mdw -pa 'still, yet'. Sa'dan also has -pa, which reflects PSS *-pa 'still, yet'.

13. There also occur isolated related forms over greater distances, i.e. between North Sulawesi languages and languages outside Sulawesi and the Philippines. Whether these represent selective retentions or shared innovations or borrowing (the last two possibilities would require some explaining) cannot be considered here. These include forms like PMin *ber₂en, Sula belen and Soboyo beleŋ (both Taliabu Island) 'eye'; PSan *kinās, Maanyang (Barito Group, South Kalimantan) kenah, Benoa (Mahakam area, East Kalimantan) kinas 'fish'; Mdw, Psk kolikip, Dusun Deyah (Barito group, South Kalimantan) kelekeb 'wing'; Tondano, Penihing (Mahakam area, East Kalimantan) oki? 'little' (although this may well be a case of convergence); PMin *pola 'sugarcane', Toba, Karo (North Sumatra) pola 'sugarpalm'.

14. This item is provided by Noorduyn (to appear).

15. PMin *tar₂e 'newly, just now', based on Tsw tale, North Minahasan tare, possibly was also the earlier form for 'new', cf. Tsw taleya 'new'.

16. The phoneme *r₂ has been reconstructed word-finally in PMin *r₂ondor₂ 'straight', where it may result from assimilation. There is no PAN etymon.

17. For instance, Sww has come under considerable influence from Gtl. Thus the occurrence of l in Sww, instead of y or h, corresponding to Gtl l is indicative of borrowing, e.g. Gtl poliama ('star'), Sww poliama, cf. Kdp puhiana, Mdw pariama, 'Pleiades'.

18. For instance, there are a few cases of Llk h, all other languages y, e.g. Llk kahu, Bul, Kdp, Mdw, Psk kayu, Gtl, Sww, Blg, Bnt ayu 'wood'.

19. Mdw has h in a few items, e.g. ha?at, as well as ra?at, ya?at 'bad'. This is as yet unexplained, although h is frequent in Llk and Psk.

20. Although intervocalic *d is reconstructed for PGM it is not common and most items have no known PAN etymon.

21. The same situation occurs in the Bali-Sasak-Sumbawa microgroup, where PAN *R is reflected by h in Balinese (> ∅ in some environments in the major dialect) and by r in Sasak and Sumbawa, which form a subgroup. Thus: PAN

*bibiR 'lip' > Balinese bibih, Sasak, Sumbawa biwir. Here also a phoneme *R of indeterminate phonetic nature must be reconstructed for the parent language of the microgroup.

22. Psk has the prefixed forms mo -gihur, with g retained after a nasal, and mo-hiur, where g → h after a vowel. Possibly loss of the following h is dissimilative.

23. The change *iu > ii is common in some Gorontalic languages, e.g. PGM *tiukan 'bee' > Gtl tii[?]a.

24. This form is not in Dunnebier's Mdw dictionary (1951) but was given by an informant as the more common form.

BIBLIOGRAPHY

- Adriani, N.
1893 Sangireesche Spraakkunst. Leiden: Nederlandsch Bijbelgenootschap.
- 1925 "De Minahasische Talen". Bijdragen tot de Taal-, Land- en Volkenkunde 81:134-164.
- Adriani, N. and M.L. Adriani
1908 Hoofdstukken uit de spraakkunst van het Tontemboansch. The Hague: Martinus Nijhoff.
- Adriani, N. and A.C. Kruijt
1914 De Bare'e-sprekende Toradja's van Midden-Celebes. Part 3: Taal- en letterkundige schets der Bare'e taal en overzicht van het taalgebied: Celebes-Zuid-Halmahara. Batavia: Landsdrukkerij.
- Anceaux, J.C.
1952 The Wolio Language. The Hague: Martinus Nijhoff.
- Bloomfield, Leonard
1933 Language. London: Allen and Unwin.
- Blust, Robert
1983 'A note on hypercorrection in Mongondow'. Bijdragen tot de Taal-, Land- en Volkenkunde. 139/4:459-464.
- 1984 'On the history of the Rejang vowels and diphthongs'. Bijdragen tot de Taal-, Land- en Volkenkunde. 140/4:422-450.
- Charles, Mathew
1973 "Additions to Zorc (1971)". typescript, Cornell University.
- 1974 "Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages". Oceanic Linguistics 13/1-2:457-509.
- Dunnebier, W.
1951 Bolaang Mongondowsch-Nederlandsch Woordenboek. The Hague: Martinus Nijhoff.
- Dyen, Isidore
1965 A Lexicostatistical classification of the Austronesian Languages. International Journal of American Linguistics, Memoir 19. Baltimore: The Waverly Press.

- Esser, S.J.
1938 Languages (sheets 9, 9b) in: Atlas van Tropisch Nederland. The Hague: Martinus Nijhoff.
- Haga, B.J.
1931 "De lima-pahalaa (Gorontalo): volksordening, adatrecht en bestuurpolitiek". Tijdschrift voor Indische Taal-, Land- en Volkenkunde 71:186-314.
- LLamzon, Teodoro A. and Ma. Teresita Martin
1976 "A subgrouping of 100 Philippine languages". in Nguyen-Dang Liem ed. South-East Asian linguistic studies. Vol. 2. pp 141-172. Pacific Linguistics Series C, No. 42. Canberra: Australian National University.
- Maryott, Kenneth R.
1977 "The Phonemes of Sarangani Sangire". Studies in Philippine Linguistics 1/2:264-279.
1978 "Pre-Sangir *l, *d, *r and associated phonemes". Studies in Philippine Linguistics 2/1:115-138.
- Mill, Roger F.
1975 Proto South Sulawesi and Proto Austronesian Phonology. unpublished dissertation, University of Michigan.
1981 "Additional addenda". in Robert A. Blust ed. Historical linguistics in Indonesia. Part I, NUSA Vol 10:59-82.
- Noorduyn, J.
1963 ed. "Mededelingen uit de verslagen van Dr. S.J. Esser". Bijdragen tot de Taal-, Land- en Volkenkunde 119:329-370.
1982 "Sound changes in the Gorontalo language". in Amran Halim, Lois Carrington and S.A. Wurm eds. Papers from the Third International Conference on Austronesian Linguistics. Vol 2, Tracking the Travellers, pp241-261. Pacific linguistics Series C, No. 75. Canberra: Australian National University.
to appear "The languages of Sulawesi". in Hein Steinhauer ed. Papers in Austronesian Linguistics No. 1, Pacific Linguistics, Canberra.
- Reid, Lawrence A.
1971 Philippine minor languages: word lists and phonologies. Oceanic Linguistics special publication No. 8. Honolulu: University of Hawaii Press.
- Riedel, J.G.F.
1870 "De landschappen Holontalo, Limoeto, Bone, Boalemo en Kattinggola, of Andagile, geographische, statistische, historische en ethnographische aanteekeningen". Tijdschrift voor Indische Taal-, Land- en Volkenkunde 19:46-153.
- Sneddon, J.N.
1970 "The languages of Minahasa, North Celebes". Oceanic Linguistics 9/1:11-36.
1978 Proto-Minahasan: phonology, morphology and wordlist. Pacific Linguistics Series B, No. 54. Canberra: Australian National University.
1983 Maps 43, 44. in: Wurm, S.A. and Shiro Hattori eds. Language atlas of the Pacific area. Canberra: Australian Academy of the Humanities.

- 1984 Proto-Sangiric and the Sangiric languages. Pacific linguistics Series B, No. 91. Canberra: Australian National University.
- to appear "The position of Lolak". in Proceedings of the Fifth International Conference on Austronesian Linguistics.
- Sneddon, J.N. and Hunggu Tadjuddin Usup
1986 "Shared sound changes in the Gorontalic language group: implications for subgrouping". Bijdragen tot de Taal-, Land- en Volkenkunde 142/4:407-426.
- Usup, Hunggu Tadjuddin
1986 Rekonstruksi Protobahasa Gorontalo-Mongondow. Unpublished dissertation, University of Indonesia.
- Walton, Charles
1979 "A Philippine language tree". Anthropological Linguistics 21/2:70-98.
- Wilken, N.P. and J.A. Schwarz
1868 "De taal in Bolaang-Mongondou". Mededeelingen vanwege het Nederlandsche Zendelinggenootschap 12:189-247.
- Zorc, R. David
1971 Proto Philippine finder list. Typescript, Cornell University.
- 1982 "Micro- and macro-subgrouping: criteria, problems and procedures". in Rainer Carle et al., eds. Gava: Studies in Austronesian languages and cultures dedicated to Hans Kahler. pp307-320. Berlin: Dietrich Reimer Verlag.
- 1986 "The genetic relationships of the Philippine languages". In Paul Geraghty, Lois Carrington and S.A. Wurm eds. Focal II: Papers from the Fourth International Congress on Austronesian Linguistics. Pacific Linguistics Series C, No. 94. pp47-173, Canberra: Australian National University.