THE GREATER CENTRAL PHILIPPINES HYPOTHESIS[1]

ROBERT BLUST

UNIVERSITY OF HAWAII

A "Central Philippine" or "Meso Philippine" subgroup has been recognized by many scholars of Philippine languages, sometimes as a component of a larger "Southern Philippine" group. However, both the composition of this group and the criteria used to justify it have varied widely. In this paper I propose that Tagalog, Bikol, the Bisayan complex, South Mangyan (but not North Mangyan), the Palawanic languages (but not Kalamian), all of the languages of Mindanao except the South Mindanao group, and the Gorontalo-Mongondow languages of Sulawesi (but not the more northerly Sangiric and Minahasan languages) continue an immediate protolanguage, here called "Greater Central Philippines" (GCP). This view of the subgrouping of Philippine languages has general similarities with that of Zorc (1986), but differs from Zorc's position in several important respects. The evidence for GCP consists of a number of replacement innovations in the lexicon, lexical and semantic innovations which cannot be shown to involve replacement, and the merger of PAN (Proto-Austronesian) *g and *R. Speakers of PGCP (Proto–Greater Central Philippines) or one of its early descendants underwent a dramatic territorial expansion, probably from a homeland in northern Mindanao or the southern Visayas. This hypothesized expansion had at least the following consequences: (1) it reduced the linguistic diversity in the Visayas and Mindanao, (2) it led to a discontinuous geographical distribution of subgroup members, since one branch of GCP (Gorontalo-Mongondow) moved south past the territorially established Sangiric and Minahasan subgroups of Philippine languages, and (3) in varying degrees, it produced Conant's "stereotyped *g*" as a reflex of PPH (Proto-Philippines) *R in all non-GCP languages which bordered on GCP languages. More generally, the Greater Central Philippines hypothesis is inconsistent with a model of gradual linguistic differentiation *in situ*, but shows intriguing parallels with the model of speciation in evolutionary biology that Gould and Eldredge have called "punctuated equilibrium."

1. THE PHILIPPINE SUBGROUP OF AUSTRONESIAN LANGUAGES. For much of this century the expression "Philippine language" has been used ambiguously to refer: (1) to any language native to the Philippine Islands, without regard to its genetic affiliation, or (2) to any

Oceanic Linguistics, volume 30, no. 2 (Winter 1991)

member of a putative subgroup of Austronesian languages most members of which are located in the Philippine Islands. In this paper it will be convenient to confine "Philippine language" to the second (genetic) sense, and to use "language in the Philippines" with a purely geographical or political reference.

Ignoring the null set in which no Philippine language is spoken in the Philippines, there are four logically possible relationships between the sets "language in the Philippines" (LP) and "Philippine language" (PL). These arise from the three fundamental set theoretic relationships of identity, overlap, and inclusion.

1.1 IDENTITY. There is only one possible identity relation, which takes the form "All A are B and all B are A." Nineteenth-century writers appear to have used "Philippine language" only in its geographical or political sense. To my knowledge the earliest explicit use of the expression in its genetic sense is that of Blake (1906:318), who refers to languages in the Philippines as "a subdivision of the Malay branch of the Malayo-Polynesian family of speech." Although he proposed a genetic referent for the expression "Philippine language," Blake nonetheless defined the Philippine subgroup in purely geographical terms: its members included all and only the languages of the Philippine archipelago. Diagrammatically, his subgrouping claim can be symbolized by the complete overlap of two circles, LP and PL, which are concentric and coterminous.

Later writers who appear to have adopted essentially the same position are Llamzon and Martin (1976), who worked with lexical lists for 100 Philippine speech communities, and Walton (1979), who worked with lexical lists for 122 Philippine speech communities. In both publications the writers are concerned with the internal subgrouping of the languages considered, but are silent regarding the criteria used to delimit their data sample. It is thus not clear whether they accept the identity relation proposed by Blake, or whether they have simply reverted to the nineteenth-century practice of defining linguistic sample units by geographical or political boundaries.

1.2 OVERLAP. There is only one possible overlap relation, which takes the logical form "Some A are B and some B are A." Conant (1911:72) appears to have been the first to propose a linguistic subgroup of this form which includes languages spoken in the Philippines: "The Philippine Islands form the center of the speech territory in which the consonant of the RGH series appears as *g*. Hence it is customary to classify as belonging to the Philippine group, not only languages of that archipelago, but such other speech groups as show the *g* of that series. Among the non-Philippine languages of this category are the Duzon and Iranun of N. W. Borneo, the

Sinkan Formosan and the Favorlang of Formosa, the Ponosakan and Mongondou of North Celebes, and the Chamorro of the Marianas." Conant goes on to note that other languages spoken in the Philippines show *r*, *l*, or *y* for what (following Brandstetter) he terms "the RGH series." It follows that the latter are not, by Conant's "customary" method of classification, Philippine languages. The terms LP and PL in this classification are thus seen to have an overlap relation: "Some LP are PL and some PL are LP." Diagrammatically Conant's claim might be symbolized by the partial overlap of two circles, LP and PL.[2]

As will be seen, Zorc (1986) proposes a classification of Philippine languages which has the same logical form as that of Conant (but with significantly different content).

1.3 INCLUSION. Because it is a bilateral relationship (A can include B or B can include A), the relation of inclusion subsumes two subtypes. The first of these has the logical form "All A are B and some B are A." I will call this type of relation "extended inclusion." The earliest statement of extended inclusion that I have been able to find relating to the classification of the languages of the Philippines is that of Dyen (1965), who treats all of these languages together with the Dusunic and Murutic languages of northern Borneo and the Gorontalic languages of northern Sulawesi as members of a lexicostatistically defined subgroup which he calls the "Northwest Hesion." Within the Northwest Hesion Dyen recognizes a more restricted subgroup, the Philippine Hesion, which does not include Ilongot of the Philippines or the Gorontalic languages of Sulawesi. With regard to the languages of the Philippines, Dyen's Northwest Hesion thus has the logical form "All LP are PL and some PL are LP" (extended inclusion), while his Philippine Hesion has the logical form "Some LP are PL and some PL are LP" (overlap). Diagrammatically, Dyen's Northwest Hesion might be symbolized as a small circle LP totally contained within a much larger circle PL.

Charles (1974) adopts a classification of the languages of the Philippines which in its broad features is essentially identical to that of Dyen (1965).

The second type of inclusion relation has the logical form "Some A are B and all B are A." I will call this type of relation "restricted inclusion." The earliest statement of restricted inclusion that I have been able to find relating to the classification of the languages of the Philippines is that of Thomas and Healey (1962), who maintain that all languages of the Philippines except Tagabili and Bilaan of southern Mindanao form a genetic unit which they call the "Philippine Superstock." Tagabili and Bilaan are assigned to the "Southern Mindanao Family" which, on lexicostatistical grounds, is said to be coordinate with the Philippine Superstock and such control units as Malay and the Chamic languages of mainland Southeast

Asia. The classification of Thomas and Healey thus has the logical form "Some LP are PL, and all PL are LP." Diagrammatically, the Thomas–Healey hypothesis might be symbolized by the circle PL being totally included within the circle LP, which however barely encompasses it.

Reid (1982) proposes a classification of languages in the Philippines which has the same logical form as that of Thomas and Healey, but which differs significantly in content. Although Thomas and Healey exclude the South Mindanao (= Bilic) languages from their Philippine Superstock, and are silent regarding Samalan, they recognize the remaining languages of the Philippines as constituting a valid linguistic subgroup. Reid, on the other hand, makes two far more radical suggestions: (1) that the South Mindanao (= Bilic) languages form a primary subgroup of the Austronesian family, coordinate with Atayalic, Tsouic, Other Formosan, and Amis-Extra-Formosan, and (2) that the remaining languages of the Philippines fall into either of two groups, "Outer Philippines" (the languages of northern Luzon, and possibly North and South Mangyan, Manobo, and Danao) and "Malayo-Polynesian." From the standpoint of logical form, Reid's denial of the existence of a "Philippine" subgroup is not original, since Thomas and Healey also recognized a subgroup that consisted of some, but not all of the languages of the Philippines, and no languages external to the Philippine archipelago. The difference between the two proposals is rather one of degree. The Reid hypothesis can be symbolized by the inclusion of one circle within another which easily encompasses it.

The remaining issues relating to the existence of a Philippine subgroup concern its boundaries. There are three problem areas: (1) northern Borneo, (2) the southern Philippines, and (3) northern Sulawesi.

Charles (1974) believed that the languages of Sabah, and perhaps even some of the languages of northern Sarawak, are descendants of "Proto-Philippines." Arguments against this view are presented in Blust (1974), and have been accepted by Reid (1982) and Zorc (1986).[3] The boundary between Philippine and non-Philippine languages in northern Borneo thus appears to fall between Molbog, a Palawanic language spoken on Balabac Island and Banggi, a non-Philippine language of uncertain classification spoken on Banggi Island off the north coast of Sabah.

Although earlier classifications such as those of Thomas and Healey (1962), Dyen (1965) and Zorc (1977) do not mention the Samalan languages, Walton (1979) includes "Sama-Bajao" as one of three coordinate branches of his Southern Philippine group (the other two being Macro Meso Philippine and Sangiric). McFarland (1980) includes the "Sama languages" as one of six coordinate branches of his Philippine group, but Zorc (1986) excludes them from a Philippine group altogether. The primary divergence of Samalan from other languages in the Philippines and its convergence to neighboring Philippine languages through generations of contact is documented in detail by Pallesen (1985).

Finally, the southern boundary of Philippine languages in Sulawesi was drawn by Esser (1938) between the Gorontalic and Tomini languages. Since almost nothing was available in print on any Tomini language, this claim could not be evaluated until the appearance of Himmelmann (1990). Based on Himmelmann's material, it is now clear that some of the lexical and semantic innovations that Zorc (1986) has used to define the Philippine group do appear in one or more of the Tomini languages. Before the position of these languages can be worked out with greater certainty, however, two questions must be answered: (1) Are the Tomini languages a genetic unit, or simply a geographical unit? (2) How much of the seemingly distinctive Philippine material in the Tomini languages is due to borrowing from Gorontalo-Mongondow languages?

There is considerable variety in the viewpoints reviewed above, and this may give rise to discouragement as to whether any reliable and generally accepted results can be reached. In my view, however, Zorc (1986) has shown convincingly that there is a large genetic grouping which encompasses all of the languages of the Philippine Islands except Samalan, together with Yami of Botel Tobago Island (Taiwan), and the Sangiric, Minahasan, and Gorontalo-Mongondow groups in northern Sulawesi. The evidence that he cites for this grouping is by no means exhaustive, and is likely to be strengthened by further research. Whereas the position of Yami within the Philippine group is noncontroversial, the precise relationship of the three northern Sulawesi subgroups to one another and to languages in the Philippines is far less settled.

2. PHILIPPINE MICROGROUPS. This section recognizes 15 Philippine microgroups, relatively low-order and noncontroversial genetic groupings that have been independently acknowledged by more than one writer. Some of the groupings are less controversial than others, but in all cases I have endeavored to survey the range of published opinion so as not to conceal controversy where it exists. The following abbreviations are used in reference to the sources of classifications: TH62: Thomas and Healey (1962); D65: Dyen (1965); Z74: Zorc (1974); LM76: Llamzon and Martin (1976); Z77: Zorc (1977); W79: Walton (1979); McF80: McFarland (1980); R81: a personal communication from Reid cited in Zorc (1986); R82: Reid (1982); Z86: Zorc (1986).

2.1 BASHIIC. This is a close-knit, highly discrete, completely noncontroversial group which includes Yami of Botel Tobago Island (Taiwan), and Itbayaten and Ivatan of the extreme northern Philippines. The name is adopted from Reid (p.c.), who derives it from the Bashi Channel between the Batanes Islands and Taiwan. Tsuchida et al. (1987) prefer the name "Batanic," and record several additional dialects.

TH62 and D65 consider only Ivatan, but the former recognize it as one of four coordinate branches of their Philippine Superstock, and the latter as one of 11 coordinate branches of his Philippine Hesion. All subsequent writers recognize Bashiic as a highly discrete group, although W79 and McF80 do not consider Yami. Z86 suggests that Bashiic may group with Sambalic, Kapampangan, and North Mangyan (Mindoro) as one of two primary branches of "Northern Philippines" (the member languages sharing the merger of PAN *R and *y), but McF80 rejects this view.

The major source of lexical material is Yamada (1976) for Itbayaten, and Tsuchida et al. (1987) for the other languages.

2.2 CORDILLERAN. There is a large and fairly diverse collection of languages named after the primary mountain massif of northern Luzon (the Cordillera Central), but actually widely distributed through the mountains, valleys, and coastal strips all across the northern portion of the island. This group includes such representative languages as Agta, Atta, Balangaw, Bontok, Casiguran Dumagat, Gaddang, Ibanag, Ifugaw, Ilokano, Ilongot, Inibaloi, Isinay, Isneg, Itawis, Itneg, Kalinga, Kallahan, Kankanay, Pangasinan, Umirey Dumagat, and Yogad, as well as the recently discovered Arta (Reid 1989) and Alta (Reid 1991).

Lexicostatistical classifications have generally given somewhat different results from qualitative classifications in defining the membership of this group. TH62 include such languages as Kalinga, Ilokano, Tingguian, Isneg, Ibanag, Atta, Gaddang, Agta, Ifugaw, Kankanay, Bontok, Sagada, and Inibaloi, yet exclude Pangasinan (closely related to Inibaloi), Ilongot, and Baler Dumagat. D65 also excludes Ilongot and Baler Dumagat, as well as Casiguran Dumagat. LM76 recognize the Cordilleran group as a genetic unit, called "Northern Philippines," including Kapampangan (but not Sambalic) within it. W79 recognizes a subgroup essentially identical to that proposed here, which includes Ilongot as a branch coordinate with all other Cordilleran languages. McF80:11 recognizes a Cordilleran subgroup "which includes the Dumagat languages, the Northern Cordilleran languages, Ilokano, the Central Cordilleran languages and the Southern Cordilleran languages." He excludes Ilongot from this narrower group, but suggests that Cordilleran, Ilongot, and the Sambalic languages of central Luzon (Bolinao, Sambal, Botolan, Kapampangan, Sinauna) form three coordinate branches of a single "Northern Philippine" subgroup. Z86 recognizes essentially the same group, but questions the inclusion of Casiguran and Umirey Dumagat.

The languages that have been included in Cordilleran here, but excluded from the same group by one or more writers are, then: (1) Ilongot, (2) the Dumagat languages, especially Baler, and (3) Pangasinan. In a large, still unpublished study first reported in Blust (1981b), it was found that Ilongot

has the lowest basic vocabulary retention rate of any of the 25 Philippine languages tested to date. From this fact alone it can be concluded that its lexicostatistical percentages with other languages are deflated, and hence will produce a misleading impression of greater divergence from other languages than history would warrant.[4] I concur with Z86 that Ilongot probably is a Southern Cordilleran language. Reid (p.c.) informs me that "Baler Dumagat is actually Northern Alta, a Meso-Cordilleran language. Umirey Dumagat is probably not a Cordilleran language; at least its reflexes of schwa make it look as though it is relatable to Bikol, with which it is geographically contiguous at its southernmost extension."

The Cordilleran languages as a whole are well-represented by published dictionaries, although important lexicographical gaps remain. Major sources are Carro (1956) for Ilokano, Vanoverbergh (1972) for Isneg, Tharp and Natividad (1976) for Itawis, Headland and Headland (1974) for Casiguran Dumagat, Reid (1976) for Bontok, Vanoverbergh (1933) for Kankanay, Newell (1968) and Lambrecht (1978) for Ifugaw, Benton (1971) for Pangasinan and Reid (1971) for most of the other languages.

2.3 CENTRAL LUZON. This is a small group of languages in west-central Luzon, which includes Kapampangan and the three Sambalic languages Bolinao, Sambal, and Botolan (the latter spoken both by lowland Filipinos, and by a Negrito population in the Zambales Mountains).[5] To these Reid (p.c.) adds Sinauna, or Sinauna Tagalog of Tanay Province, which (despite its misleading name) "is clearly a distinct Central Luzon language, different from the Ayta groups of Zambales."

This group is not recognized by TH62, who list Kapampangan (= Pampangan) as especially close to Tagalog, and classify Sambal as a separate branch of their Southern Philippine Family, coordinate with Tagalog–Kapampangan, Bikol, the Bisayan languages, and various languages of Mindanao. D65 does not consider Sambal, but places Kapampangan in his Sulic Hesion, together with Tagalog, Bikol, the Bisayan languages, and various languages of Mindoro, Palawan, and Mindanao. As noted already, LM76 split Kapampangan from the Sambalic languages, and hence recognize nothing similar to the Central Luzon group. W79 includes the Sambalic languages and Kapampangan as two of the three coordinate branches in his Central Philippine group (the third branch including Tagalog, Bikol, the Bisayan languages, and many languages of Mindanao). As noted already, McF80 includes the Sambalic languages and Kapampangan as three coordinate branches of his "Northern Philippine" group. Z86 accepts this group, but calls it "Southern Luzon," and suggests that it belongs with Bashiic and North Mangyan in a Northern Philippine group. Z74 provides virtually the only published data on the languages of northern Mindoro. It may well prove true, as he

suggests, that Central Luzon and North Mangyan form a larger genetic unit, but I am aware of no convincing evidence that Bashiic is more closely related to it than to any other Philippine microgroup.[6]

Kapampangan is well represented lexically by the dictionaries of Bergaño (1860) and Forman (1971), but little lexical material is available for Sambalic apart from the Botolan Sambal wordlist in Reid (1971) and the Bolinao, Sambal, and Botolan materials in McFarland (1977).

2.4 INATI. This language, spoken by an isolated Negrito population on the island of Panay in the central Philippines, has been described briefly by Pennoyer (1986/87). Uniquely among languages in the Philippines, it has merged *R and *d in at least final position (intervocalically pre-Inati *d became /r/, and subsequently pre-Inati *R became /d/). The data provided by Pennoyer suggest that Inati does not subgroup closely with any neighboring language. Pending further information, it is best classified as a Philippine isolate.

2.5 KALAMIAN. This is a cluster of dialects, including Kalamian Tagbanwa and Agutaynon, spoken in the Calamian Islands between Palawan and Mindoro. Kalamian Tagbanwa is important for the reconstruction of Proto-Philippine *q, especially in initial position.

D65 classifies Kalamian Tagbanwa along with the Tagalic Subfamily, the Dibabaic Subfamily, the Palawanic Subfamily, the Bukidnic Subfamily, Kapampangan, and Cotabato Manobo as seven coordinate branches of the Sulic Hesion. Z77 proposes that Kalamian is coordinate with South Mangyan, Palawanic, Subanun, and Central Philippines within his Meso Philippine group. W79 treats Kalamian as most closely related to the languages of Palawan, and somewhat more distantly to those of Mindoro (north and south), Sambalic, Kapampangan, and Tagalic. McF80 presents a similar picture, except that Sambalic and Kapampangan are excluded. Contrary to all other proposals Reid (1981; cited in Zorc 1986) tentatively places Kalamian, together with Bilic, Bagobo, and Tiruray, in a Southern Mindanao group. Z86 places Kalamian in a group with South Mangyan (but not North Mangyan), the languages of Palawan and Central Philippines. Several writers, then, have maintained that Kalamian subgroups with at least the languages of Palawan, South Mangyan, and the Tagalic languages. At the same time there has been a general recognition that Kalamian is a fairly distinct unit within any larger collection of languages in the central Philippines.

Little published evidence is available for Kalamian apart from the 372-word list of Kalamian Tagbanwa in Reid (1971), but what there is suggests a fairly marked Bisayan or other Central Philippine overlay. Contrary to all earlier proposals, I will suggest that Kalamian is a Philip-

pine isolate—a language which is no more closely related to other languages in the central Philippines than it is to, say, Bashiic or Cordilleran.[7]

2.6 BILIC. This small group of languages in southern Mindanao, including Bilaan, Tagabili (= Tboli), Tiruray, and Giangan Bagobo has been regarded by several writers as highly distinctive. TH62 assigned Bilaan and Tagabili to the South Mindanao Family, which they regarded on lexicostatistical grounds as no more closely related to other languages in the Philippines than to the Chamic languages of mainland Southeast Asia or to Malay. D65 treated the same two languages as members of the Bilic Subfamily, one of 11 coordinate branches of the Philippine Hesion. He treated Tiruray as another coordinate branch of the Philippine Hesion, but the evidence for a close genetic relationship between Tiruray, Bilaan, and Tagabili is overwhelming (Blust to appear). LM76 include the Bilic languages in their Southern Philippines group, while W79 regards Tiruray, Giangan Bagobo, Bilaan, and Tagabili as members of a South Mindanao subgroup within the Macro Meso Philippine branch of the Philippine language group. McF80 gives the same four languages (differing only in using the ambiguous term "Bagobo," which refers to two quite different languages in Mindanao) as members of a South Mindanao group. He treats the South Mindanao group as one of seven coordinate branches of the Philippine language group, thus—like TH62—emphasizing its divergence from neighboring languages. Reid (1982) takes by far the most extreme splitting position, proposing that the Bilic languages form one of five coordinate branches of the entire Austronesian language family. Z86 recognizes the Bilic group as constituted here, but assigns it to his Southern Philippines subgroup.

The relatively divergent position of the Bilic languages in relation to their immediate neighbors has been generally recognized, and is accepted here. Reid's exclusion of the Bilic languages from the Philippine group, however, is rejected for reasons discussed in Blust (to appear), and Zorc's inclusion of the Bilic languages in Southern Philippines is rejected for reasons that will be discussed in this paper.

The major sources of lexical material are Schlegel (1971) for Tiruray, Forsberg and Lindquist (1955) for Tagabili and Reid (1971) for Bilaan.

2.7 SOUTH MANGYAN. Zorc (1974) showed clearly that the languages of Mindoro (locally called "Mangyan") belong to two rather distinct groups: North Mangyan (Iraya, Alangan, Tadyawan) and South Mangyan (Hanunóo, Buhid). Tweddel (1970) reports a sixth Mangyan language, Batangan. From his general statements regarding its close similarity to Buhid we may tentatively infer that it is a South Mangyan language.[8]

TH62 do not mention the South Mangyan languages, but D65 assigned

Hanunóo and Buhid to the Hanunoic Subfamily which, together with the Tagalic Hesion, Bikol, the Mansakic Cluster, the Irayic Hesion, and Subanun, forms one of six coordinate branches of the Mesophilippine Hesion. LM76 include both North Mangyan and South Mangyan in their Central Philippine group. W79 recognizes a Mindoran subgroup which includes both North Mangyan and South Mangyan, as one of three coordinate branches of Meso Philippines, while McF80 recognizes North Mangyan, South Mangyan, and the Palawan languages as three coordinate branches of the group he calls Meso Philippines. Reid (1981, cited as a personal communication in Zorc 1986) also posits a Meso Philippine group with three coordinate branches: South Mangyan, Palawan and Central Philippines, and Z86 accepts this position.

The position adopted here is that North and South Mangyan are quite distinct, but that some convergence has occurred due to prolonged borrowing. This convergence is most problematic in lexicostatistical classifications, which generally neglect the distinction between directly and indirectly inherited forms.

The major source of lexical material is Conklin (1953) for Hanunóo. Zorc (1974) provides almost the only material available in print for Buhid or any of the North Mangyan languages.

2.8 PALAWANIC. This group consists of at least four languages, including Palawano, Aborlan Tagbanwa and Batak of Palawan Island, and Molbog of Balabac Island between Palawan and north Borneo. TH62 list Batak as a primary branch of their Southern Philippines Family without further detail, and D65 recognizes both Kalamian and the Palawanic Subfamily as coordinate branches of his Sulic Hesion. LM76 assign Palawanic to their Central Philippines, while Z77 proposes a Meso Philippine group with five coordinate branches: South Mangyan, Palawan, Kalamian, Subanun, and Central Philippine. W79 recognizes a grouping similar to that suggested here, while McF80 posits a Meso Philippine group with four coordinate branches: North Mangyan, South Mangyan, Palawan, and Central Philippine. Z86 recognizes a grouping similar to that proposed in Z77, but without a separate Subanun branch.

Thiessen (1977) has reconstructed the phonology of Proto-Palawan, together with a vocabulary of 165 words, but much work remains to be done.

The major sources of comparative material are Warren (1959) for Batak, and Reid (1971) for Batak and Aborlan Tagbanwa. The only published grammar is Revel-Macdonald (1979).

2.9 CENTRAL PHILIPPINES. This grouping includes several major languages of the central Philippines (Tagalog, Bikol, the Bisayan complex), and many minor languages of eastern Mindanao (Mamanwa, Mansaka,

Mandaya, Kalagan, Tagakaulu, etc.). The grouping recognized here is essentially identical to the Central Philippine group in Z77. TH62 include a number of these languages in their Southern Philippine Family, but also include Sambal, Kapampangan, Maranao, and Subanun, which are not included here. D65 recognizes a Tagalic Hesion with four coordinate branches: the Bisayan Cluster, Cagayanon, Mamanwa, and Tagalog. LM76 split the membership of this group between what they call "Central Philippines" (including Tagalog, Bikol, and Bisayan, among others) and "Southern Philippines" (including Mamanwa, Mansaka, and Mandaya, among others). W79 posits a Central Philippine subgroup that includes three coordinate branches: Sambalic, Kapampangan, and the rest (Tagalog, Bikol, Bisayan; Mamanwa, Mandaya, Mansaka, Kalagan). McF80, R81, and Z86 recognize a virtually identical grouping with the same name.

Major sources of lexical material include Panganiban (1966, 1973) for Tagalog, Mintz and Britanico (1985) for Bikol, Wolff (1972) for Cebuano Bisayan, Zorc (1969) for Aklanon, Motus (1971) for Hiligaynon, and Reid (1971) for the languages of eastern Mindanao. By far the most important comparative study is Zorc (1977).

2.10 MANOBO. An extensive collection of minor languages most of which are spoken in the mountains of central and eastern Mindanao. These include Binukid, Ilianen Manobo, Western Bukidnon Manobo, Ata Manobo, Tigwa Manobo, Dibabawon Manobo, Cotabato Manobo, Sarangani Manobo, Bagobo, and Tasaday on the island of Mindanao, Kinamigin on Camiguin Island in the Bohol Sea just north of Mindanao, and Kagayanen, a Manobo outlier spoken in the Cagayan Islands between the Bisayas and Palawan (cf. Harmon 1977). In D65, Manobo languages are scattered through several primary branches of the Sulic Hesion, apparently no more closely related to one another than they are to Tagalog, Kapampangan, Hanunóo, the North Mangyan languages, or Kalamian. However, a Manobo subgroup is explicitly recognized by TH62, Elkins (1974), LM76, Z77, W79, McF80, R81, and Z86, and important work on the reconstruction of Proto-Manobo has been done by Elkins (1974, 1982).

Relatively little lexicographic material is yet available, the major source being Elkins (1968) for Western Bukidnon Manobo. I have also referred to Molony and Tuan (1976) for material on Tasaday, and to Reid (1971) for a wider sampling of Manobo languages.

2.11 DANAW. This well-defined group of three languages (Maranao, Iranon, Magindanao) is spoken by a predominantly Moslem population in southwestern Mindanao. TH62 assign Maranao and Magindanao to a primary branch of their Southern Philippine Family. D65 treats Maranao as coordinate with the Sulic Hesion, and hence as no more closely related to the Manobo languages or Subanun than it is to Casiguran Dumagat,

Yakan (a Samalan language), Bashiic, or Cordilleran. LM76 include Danaw in their Southern Philippines. Z77 treats Danaw as one of five coordinate branches of Southern Philippines, while W79 recognizes a "Danao-Subanun" branch, coordinate with Manobo within a larger "Meso Mindanaoan" unity. McF80 classifies Subanun, Danao, and Manobo as three coordinate branches of Southern Philippines, a position essentially identical to that of Z86. R81 accepts a similar lower-level classification, but regards Southern Philippines as a primary branch of Western Malayo-Polynesian, rather than as part of a larger Philippine group.

Allison (1979) has reconstructed a phonological system, some grammatical morphemes, and a general vocabulary of 328 items for Proto-Danaw.

Major data sources are McKaughan and Macaraya (1967) for Maranao, and Juanmarti (1892) for Magindanaw.

2.12 SUBANUN. There is a set of two or three closely related languages (Subanun, with at least two rather divergent dialects, Kalibugan) spoken in the Zamboanga Peninsula of western Mindanao. A general recognition of the close relationship of Subanun to the Manobo and Danaw languages has been noted in the previous sections.

Published material on Subanun is fairly limited. Major sources are Finley and Churchill (1913) and Reid (1971).

2.13 SANGIRIC. This is a group of five languages spoken in the Sangir-Talaud Islands and on the northern peninsula of Sulawesi in Indonesia, with a relatively recent immigrant language or dialect (Sangil) spoken in the Sarangani Islands and on the southernmost tip of the Sarangani Peninsula in Mindanao. The definitive comparative study of the Sangiric languages is that of Sneddon (1984), who has reconstructed the phonological system of Proto-Sangiric together with a vocabulary of 750–800 lexemes. Because only Sangil lies within the political boundaries of the Republic of the Philippines, many writers on Philippine languages omit the Sangiric group. D65 treats Sangir on lexicostatistical grounds as not significantly closer to Tagalog (34.8%) than it is to Sasak (36.1%), Malay (36%), or even Sika (33%), a conclusion which is radically at odds with the qualitative evidence. LM76 include Sangir and Sangil in their Southern Philippines group, while W79 treats Sangiric as one of three coordinate branches of Southern Philippine, along with Sama-Bajao (= Samalan) and Macro Meso Philippine. McF80 mentions only Sangil, and treats it as one of six coordinate branches of the Philippine group as a whole. Based on statements in Sneddon (1978:10, 1984:11–12), Zorc (1986) recognizes a Sangiric-Minahasan group of languages, which he assigns to his Southern Philippine group.

Sneddon (1989a:93) has now rejected his earlier belief in a Sangiric-Minahasan subgroup of languages in northern Sulawesi. He believes that the Sangiric, Minahasan, and Gorontalo-Mongondow languages all belong to a Philippine group, but is noncommittal regarding their relations to one another or to other languages within such a group.

The only published dictionary for any of the languages is Steller and Aebersold (1959).

2.14 MINAHASAN. A group of five languages (Tonsea, Tombulu, Tondano/Toulour, Tontemboan, and Tonsawang) spoken in the general vicinity of Lake Tondano in the northern peninsula of Sulawesi. The definitive comparative study of the Minahasan languages is Sneddon (1978), in which the phonological system of Proto-Minahasan is reconstructed, together with a vocabulary of about 700 lexical items. D65 includes only Tontemboan, and treats it as one of seven coordinate branches of the Malayopolynesian Linkage (hence no closer to Sangir or Tagalog than it is to Paiwan or Fijian, for example).

A published dictionary is available only for Tontemboan (Schwarz 1908).

2.15 GORONTALO-MONGONDOW. A group of nine languages spoken in the central and western portions of the northern peninsula of Sulawesi. These are Ponosakan, Mongondow, Lolak, Atinggola-Bolango, Bintauna, Kaidipang, Suwawa, Gorontalo, and Buol. The most extensive comparative study of these languages is that of Usup (1986). Other important diachronic studies are Noorduyn (1982), Sneddon and Usup (1986), and Sneddon (1991). In addition Usup (1981) contains 881 cognate sets with reconstructions.

A substantial dictionary is available for Bolaang Mongondow (Dunnebier 1951), and a much smaller and more problematic dictionary for Gorontalo (Pateda 1977).

Because the subgrouping of the languages of the central and southern Philippines and northern Sulawesi involves a common core of agreement, but considerable variation in the composition of major groups, I have summarized the published positions on linking of microgroups in Table 1. Abbreviations are: TH62 : Thomas and Healey (1962) = Southern Philippine Family; D65 : Dyen (1965) = Sulic Hesion; LM76 : Llamzon and Martin (1976) = Central-Southern Philippines; Z77 : Zorc (1977) = Southern Philippine; W79 : Walton (1979) = Macro Meso Philippine; McF80 : McFarland (1980) = an unnamed group consisting of Meso Philippine plus Southern Philippine; R81 : Reid (1981, cited as a personal communication in Zorc 1986) = Meso-Philippines; Z86 : Zorc (1986) = Southern Philippine; B91 : Blust (this publication) = Greater Central Philippines.

TABLE 1. PUBLISHED POSITIONS ON THE SUBGROUPING OF THE LANGUAGES OF THE CENTRAL AND SOUTHERN PHILIPPINES AND NORTHERN SULAWESI

	TH62	D65	LM76	Z77	W79	McF80	R81	Z86	B91
1. Bashiic	NO	NO	X	X	NO	NO	NO	NO	NO
2. Sambalic	X	–	X	X	X	NO	NO	NO	NO
3. Kapampangan	X	X	NO	X	X	NO	NO	NO	NO
4. Tagalog	X	X	X	X	X	X	MP	X	X
5. Bikol	X	X	X	X	X	X	MP	X	X
6. Bisayan	X	X	X	X	X	X	MP	X	X
7. N. Mangyan	–	X	X	X	X	X	NO	NO	NO
8. S. Mangyan	–	X	X	X	X	X	MP	X	X
9. Kalamian	–	X	X	X	X	X	NO	X	NO
10. Palawanic	X	X	X	X	X	X	MP	X	X
11. Mamanwa	–	X	X	X	X	X	MP	X	X
12. Kalagan	X	–	–	X	X	X	MP	X	X
13. Mansaka	X	X	X	X	X	X	MP	X	X
14. Danaw	X	NO	X	X	X	X	SP	X	X
15. Manobo	X	X	X	X	X	X	SP	X	X
16. Subanun	X	X	X	X	X	X	SP	X	X
17. Bilic	NO	NO	X	NO	X	NO	NO	X	NO
18. Sangiric	–	NO	X	–	NO	NO	NO	X	NO
19. Minahasan	–	NO	–	–	–	–	NO	X	NO
20. G-M	–	NO	–	X	–	–	NO	X	X

Table 1 is to be read as follows. Item 1 shows that the Bashiic languages were not included by Thomas and Healey (1962) in their Southern Philippine Family, by Dyen (1965) in his Sulic Hesion, and so forth, but were included by Llamzon and Martin (1976) in their Central-Southern Philippines group, and by Zorc (1977) in his Southern Philippine Family (a position retracted in 1986). Item 2 shows that the Sambalic languages were included by Thomas and Healey (1962) in their Southern Philippine Family, by Llamzon and Martin (1976) in their Central-Southern Philippines group, by Zorc (1977) in his Southern Philippine Family, and by Walton (1979) in his Macro Meso Philippine, but that Dyen (1965) does not mention them, and writers on Philippine subgrouping since 1980 have excluded them from the subgroup that includes Tagalog, Bikol, the Bisayan languages, and many of the minority languages of Mindanao. Other items are to be read similarly, except that "MP" and "SP" refer to "Meso-Philippines" and "Southern Philippines," which R81 assigns to different primary branches of "Extra-Formosan," no more closely related to one another (according to him) than either is to Atayal or Tsou.

Since they are relatively noncontroversial, the foregoing microgroups can function as building blocks in the construction of a more inclusive subgroup that encompasses many, but not all of the languages of the

central and southern Philippines. Briefly, the establishment of a body of exclusively shared innovations between any pair of microgroups will be taken to link those microgroups in a larger genetic unit. The linking of microgroups will be logically transitive; thus, if evidence can be produced linking A with B and B with C, it will be inferred that A, B, and C are all members of a larger subgroup ABC even if direct evidence linking A with C is not yet to hand.

On both archaeological and linguistic grounds there is good reason to believe that Austronesian languages have been in the Philippines for 5,000 years or more (Bellwood 1985:122, Blust 1984/85:55). During such an extended period in which contact was unavoidable, a great deal of borrowing clearly must have taken place. In evaluating probable exclusively shared innovations, the factor of borrowing must constantly be kept in mind, and any clues to loan phonology must be carefully collected and used to distinguish directly from indirectly inherited material. Before it is possible to meaningfully examine loan phonology, however, we must have a basic picture of the Proto-Philippine system of phonemic contrasts, and the major reflexes of these phonemes in representative daughter languages.

3. PROTO-PHILIPPINE PHONOLOGY. Charles (1974) reconstructed Proto-Philippines with the four vowels *i, *u, *e (schwa), *a, and 17 consonants as follows:

FIGURE 1. PROTO-PHILIPPINE CONSONANTS (after Charles 1974)

p	t		k	q
b	d		j	
m	n	ñ	ŋ	
	s			h
	l			R
w		y		

Charles does not discuss the phonetic interpretation of these symbols, except in passing (1974:fn. 4), but *q probably was a pharyngeal stop, *j a palatalized velar stop, and *R a uvular trill. Conspicuously absent from this inventory is *g, for which excellent evidence is available (sometimes, as with PPH *tageRaŋ 'rib', in the same morpheme as *R), and *r, for which somewhat more problematic, but nonetheless good, evidence is to hand.

Charles noted (1974:fn. 4) that "the only Philippine language reflecting PPh *ny unmerged with *n is Kapampangan," and in support of this statement he cites two etymologies, one showing *ñ > /y/ and the other *ñ > /ñ/

(written /ny/): (1) *ñamuk > /yamuk/ 'mosquito' and (2) *qañud > /anyud/ 'carried away on the current'. Additional evidence for the retention of this PAN distinction in Kapampangan is available in Bergaño (1860) and Forman (1971), where we find that the regular reflex of PAN *ñ is Kapampangan /ñ/: (3) /kañaw/ 'to rinse; rinsing, washing' (with reflex of root *-ñaw 'wash, bathe, rinse'); (4) /kuñat/ 'flexible, tough, leathery' (with reflex of root *-ñat 'stretch');[9] (5) *laña > /laña/ 'vegetable oil'; (6) *leñep > /lañap/ 'disappear, fade away'; (7) *ñilu > /liñu/ (< Met.) 'setting teeth on edge'; (8) *ñaman > /ñaman/ 'delicious' (Forman 1971); and perhaps (9) *buñi > 'sound, noise' > /buñi/ 'celebrated, applauded, acclaimed'.

Ferguson (1966) has claimed that no attested language has more orders of nasals than of stops. While the PPH inventory reconstructed by Charles can be reconciled with this claim (since p, t, k and q appear to have involved four different points of articulation), in Austronesian languages as a whole we can make the somewhat stronger claim that for each nasal order there is a corresponding order of stops (or, better, obstruents). If this relationship holds for attested languages we would expect it to hold no less for their reconstructed ancestors. As noted already, the Proto-Philippine phoneme that Charles reconstructed as *j probably was a palatalized velar (with reflexes [d], [g], [l], and [r] in Philippine languages), rather than a palatal. There is, then, a clear typological implication that, next to *ñ, PPH also had a palatal obstruent. The evidence for retention of the distinction between PAN *z and *d in Proto-Philippines is, if anything, even more tenuous than that for PPH *ñ.[10] Kapampangan itself has a palatal affricate which Forman (1971) writes /j/, but it is rare, and found mostly in loanwords. Of more direct interest is the citation in Pennoyer (1986/87) of Inati [udyan], [odyan] < PAN *quzaN 'rain', and of [tinudyuʔ] 'fingernail', possibly an infixed reflex of *tuzuq 'point, indicate'. Too little data are given to determine whether Inati regularly retains the distinction of PAN dental and palatal obstruents, and further research on this point would be of real interest to the reconstruction of PPH phonology. In place of Charles's PPH inventory of 17 consonants I propose the following inventory of 20:

FIGURE 2. PROTO-PHILIPPINES CONSONANTS (REVISED)

p	t			k	q
b	d	z	j	g	
m	n	ñ		ŋ	
	s				h
	l				
	r			R	
w		y			

The major developments of this system that will be relevant to the following discussion are: (1) *k disappeared in native Kalamian forms, and became /g/ intervocalically in native Bilic forms; (2) *q merged with zero in initial position in most Philippine languages, and became /ʔ/ elsewhere.[11] However, in Kalamian *q became /k/ in all positions following the loss of PPH *k, and in Tagabili *q became /k/ in initial and final positions; (3) *d merged with *z virtually everywhere in the Philippines, and with *j virtually everywhere outside Cordilleran (*j merged with *g in some, but not all Cordilleran languages); intervocalically the result of this merger is a continuant /r/ or /l/ in many of the languages of the central and southern Philippines, including Inati, Kalamian, Bilic, Hanunóo, Palawanic apart from Molbog, most or all Central Philippine languages, many (but not all) Manobo languages, Danaw, Subanun, and Gorontalo-Mongondow. Zorc (1977:211) points out that although all modern Bisayan dialects have a continuant reflex of PPH *-d-, *-z-, and *-j-, internal Bisayan evidence requires that the merger of these phonemes be reconstructed as Proto-Bisayan *-d-; (4) *n normally disappeared in final position in the Gorontalic languages; (5) *s became /h/ unconditionally in Botolan Sambal, and /t/ unconditionally in Kalamian; in the Bilic languages *s became /h/ when noninitial, disappearing entirely in final position in Sarangani Bilaan; in Gorontalo, Buol, Suwawa, and Bolaang Mongondow *s became /t/ (then *t from any source became BM /s/ before a high front vowel); (6) *R became Bashiic /y/, Northern Cordilleran /g/, Ilokano /r/ or /g/, Central and Southern Cordilleran /l/ or /g/, Central Luzon /y/, Inati /d/ (after the change *d > /r/ intervocalically), Kalamian /l/, South Mangyan /g/, Palawanic /g/, Central Philippines /g/, Manobo /g/, Danaw /g/, Subanun /g/, Proto-Sangiric *R (with reflexes /r/, /h/, and zero), Proto-Minahasan *h, and Proto-Gorontalo-Mongondow *g; (7) *-i, *-ay, *-uy, and *-iw became -/ey/, and *-u, *-aw became -/ew/ in Tiruray.

Of the foregoing protophonemes the one with greatest diagnostic value for distinguishing directly inherited from indirectly inherited vocabulary over a wide geographical area is PPH *R. A quick overview of the reflexes of PPH *R will thus provide a convenient, if somewhat makeshift and overgeneralized profile of the more important borrowing relationships that have characterized the past millennium or two of contact between Austronesian languages in the Philippines.

4. THE "RGH LAW IN PHILIPPINE LANGUAGES" REVISITED. In an early classic of Austronesian comparative linguistics Conant (1911) described the reflexes of *R in a wide variety of Philippine languages. He divided the languages into four types: (1) "r" languages, (2) "l" languages, (3) "y" languages, and (4) "g" languages, and noted (1911:74–75) that

TABLE 2. REFLEXES OF *R IN PHILIPPINE AND NEIGHBORING LANGUAGES

	DIRECT	INDIRECT	LOAN PERCENT
1. Itbayaten	y (19)	g (3), l (1), zero (1)	12.5
2. Gaddang	g (17)	zero (2)	
3. Ilokano	r (11)	g (21)	
4. Bontoc	l (10)	g (10)	
5. Pangasinan	l (20)	g (12), r (1), y (1)	26.0
6. Sambal	y (20)	g (9)	31.0
7. Kapampangan	y (12)	g (20)	62.5
8. Tagalog	g (53)	y (5)	8.5
9. Bikol	g (83)	r (5), y (2)	5.5
10. Hanunóo	g (46)	y (10)	18.0
11. Kalamian Tagbanwa	l (20)	g (9), y (2)	29.0
12. Palawan Batak	g (24)	y (2)	6.5
13. Inati	d (13)	g (28)	68.0
14. Cebuano Visayan	g		
15. Molbog	g (6)	r (2), dz (1), h (1)	20.0
16. Mansaka	g (30)		
17. Western Bukidnon Manobo	g		
18. Maranao	g		
19. Sindangan Subanun	g (28)		
20. Tiruray	r (44)	g (36)	45.0
21. Samal	h (8)	g (6), zero (5), l (4)	26.0
22. Banggi	g (10)	r (8), zero (2)	40.0
23. Proto-Sangiric	R (86)	g (4), zero (1)	4.5
24. Proto-Minahasan	zero/h (30)	g (3), zero (1)	9.0
25. Bolaang Mongondow	g		
26. Gorontalo	h		

"unlike the Tagalog, or pure *g* type, the *r*, *l*, and *y* languages show some irregularities, their characteristic consonant often interchanging with *g*." Conant called this irregular /g/ reflex of *R a "stereotyped Philippine *g*."

Table 2 presents a sample of *R reflexes in languages representing all of the 15 microgroups of Section 2. Time has not permitted an exhaustive survey of lexical reflexes, but there is no reason to believe that the data examined are not representative of the total picture. Data for Itbayaten, Gaddang, Bontok, Sambal, Kalamian Tagbanwa, Palawan Batak, Mansaka, Sindangan Subanun, and Samal were taken from the 372-word list in Reid (1971). For these languages all reflexes of *R in the material were sought out and tabulated. For Ilokano, a check was made only of forms in Carro (1956) that begin with /a/, /b/, and /d/. More thorough research is reported in Tharp (1974). For Pangasinan, Kapampangan, Tagalog, Bikol, and Hanunóo, I searched the available dictionaries rather hastily, and undoubtedly missed a number of reflexes of *R in final position. For

Cebuano, Western Bukidnon Manobo, Maranao, Bolaang Mongondow, and Gorontalo, the available dictionaries were not scanned at all, since there appeared to be little if any evidence for double reflexes of *R apart from obvious and relatively recent (past three or four centuries) Malay loanwords. The Sangir dictionary was scanned more carefully, as were the Proto-Minahasan and Proto-Sangiric wordlists in Sneddon (1978) and Sneddon (1984). The Inati data are taken from Pennoyer (1986/87), and the Molbog data from Thiessen (1977, 1981). The Tiruray data are compiled in Blust (to appear), and the Banggi material was collected from Schneeberger (1937).

The results of this bird's-eye survey are of considerably more interest than might initially be supposed.

In Itbayaten some 79% of the known reflexes of *R are /y/, with the possibility that conditioned change accounts for single instances of /l/ and zero. The three known examples of /g/ (12.5%) are best attributed to borrowing from North Cordilleran languages, thus suggesting a relatively light and apparently unilateral borrowing relationship, with North Cordilleran as the donor and Bashiic as the recipient languages.

As noted by Tharp (1974), the reflexes of *R in Ilokano present a complex picture, and suggest fairly detailed conditioning rather than borrowing as the primary source for the split of *R into /r/ and /g/. In Bontok and other Central Cordilleran languages, /l/ and /g/ appear to be about equally common reflexes of *R; the best explanation for the velar reflexes, as suggested by Reid (1973) for Kankanay, probably is borrowing.

In Pangasinan, although conditioning may also have played a part in the split of *R, both geographic and linguistic considerations favor the hypothesis that *R > /g/ is found largely in Tagalog loanwords, and the far rarer *R > /y/ in Kapampangan loans. If so, over 26% of the Pangasinan vocabulary which contains a reflex of *R may be a product of borrowing from Tagalog.

Very similar to the situation in Pangasinan is the situation in Sambal, where perhaps 31% (9/29) of the reflexes of *R are probable Tagalog loans.

The reflexes of *R in Kapampangan and Tagalog show clearly that the two languages have borrowed extensively from one another, and that Tagalog has been by far the more important donor language. Up to 62.5% of all Kapampangan lexical items which contain a reflex of *R may be Tagalog loans, whereas less than 9% of Tagalog reflexes of *R can be categorized as likely loans from Kapampangan.

The picture in Hanunóo shows unambiguous evidence of borrowing from North Mangyan languages, with about 18% of the known reflexes of *R being likely loans. The data available to me for North Mangyan unfortunately do not permit a similar calculation of possible South Mangyan influence in North Mangyan languages.

Kalamian Tagbanwa shows a strong borrowing influence from some Central Philippine source (9/31, or about 29% of known reflexes), and a weak borrowing influence from some Central Luzon source (2/31, or about 6.5%).

Palawan Batak shows a weak borrowing influence from some Central Luzon source (2/26, or just under 8%), while Inati, as Pennoyer notes, has been heavily influenced by the Bisayan language Kinarayʔa, with over two-thirds of its known reflexes of *R being likely Bisayan loans.

The situation in Molbog, and in Banggi to the south of Molbog in Malaysia, is somewhat puzzling. Both languages appear to regularly reflect *R as /g/, but also have /r/, and occasionally other reflexes. Most instances of *R > /r/ in Banggi are best explained as Malay loans, but this does not appear to be possible with PMP *zuRuq > Molbog /duruʔ/ 'liquid', PMP *Ramut > Molbog /ramut/ 'root', or PMP *deŋeR > Banggi /ki-doŋor/ 'hear'. Similarly, although *R > /h/ in Molbog suggests borrowing from Samal, PMP *zaRum > Molbog /dohum/ 'needle' is not easily explained as a Samalan loan in light of Samal /jalum/ 'needle'.

In Tiruray, fully 45% of all *R reflexes are likely loans, principally from Danaw sources (Blust to appear). The situation appears to be similar for Tagabili and Bilaan.

In Samal, about 26% (9/23) of all *R reflexes appear to be loans from languages in the central or southern Philippines. *R > /l/ may be due to the assimilation of Malay loanwords, and zero reflexes may be conditioned.

The Proto-Sangiric wordlist in Sneddon (1984) shows 4 of 91, or about 4.5%, irregular *R > *g reflexes. This suggests that the Sangiric languages, particularly Sangir and Talaud, with their relatively isolated geographical position, have been less susceptible to borrowing influences than have most languages in the Philippines proper.

Finally, Proto-Minahasan shows likely borrowing in 3 of 34, or just under 9%, of its *R reflexes, almost certainly from a Gorontalo-Mongondow source language or languages.

What are we to make of Table 2? First, the number and percentage of irregular reflexes which are most simply explained as products of borrowing provides information about the *intensity* of contact. It seems clear, for example, that languages such as Inati, Kapampangan, and to a somewhat lesser extent, Tiruray, have been in intensive contact with languages in which *R is regularly reflected as /g/, whereas the Sangiric and Minahasan languages show far less evidence of borrowing from such sources. Second, the direction of borrowing can be used to support inferences about the *relative prestige* of languages that were in contact during the prehistoric period. Kapampangan and Tagalog, for example, have borrowed from one another, but it is clear from Table 2 that Tagalog has been by far the more important donor language.

If we set aside Bashiic, and Northern and Central Cordilleran, which for reasons of geography are best treated as outside the scope of our concern, there are three general patterns of borrowing reflected in Table 2: (1) languages which regularly reflect *R as something else have borrowed from languages which regularly reflect *R as /g/; (2) languages which regularly reflect *R as /g/ have borrowed from languages which reflect *R as /y/; (3) languages which regularly reflect *R as /g/ have borrowed from languages which regularly reflect *R as /r/. Pattern (1) (Conant's "stereotyped Philippine *g*") is easily the most widespread of the three, extending from Pangasinan in the north to Minahasan in the south. Pattern (2) is restricted to central and southern Luzon, Mindoro, the Calamian Islands and northern Palawan, or the region surrounding the languages of the Central Luzon subgroup (Sambalic, Kapampangan, North Mangyan). Pattern (3) is at least partly a product of borrowing from Malay in many Philippine languages, but appears to involve non-Malay source languages in some areas (e.g. Molbog /duruʔ/ 'liquid' and /ramut/ 'root' cannot be explained as loans from Malay or from any language currently spoken in the region of Palawan).

The simplest explanation of pattern (2) clearly is that languages of the Central Luzon group were occasional lexical donors to Palawan Batak, Kalamian Tagbanwa, Hanunóo, and Tagalog (possibly also Bikol, although Bikol loans from such sources may have been transmitted through Tagalog). The area of greatest known borrowing from Central Luzon sources is southern Mindoro (Hanunóo), but the less extensive evidence of borrowing from Central Luzon languages in Kalamian Tagbanwa and Palawan Batak is in many ways more interesting, since it suggests that languages which reflect *R as /y/ may once have extended over a wider territory than they presently occupy.

Pattern (3) is more complex, involving some patent Malay loanwords, but possibly also substrate languages in southern Palawan which regularly reflected *R as /r/. As noted already, similar irregularities which do not appear to be ascribable to borrowing from Malay occur in Banggi, as with *ki-doŋor* 'hear'. There are some indications, then, that languages which regularly reflected *R as /r/ were once spoken in the region of southern Palawan and the Banggi archipelago.

Pattern (1) is by far the most important of the three. With only one clear exception, languages which show the regular change *R > /g/ have been predominant lexical donors to neighboring languages which do not show this change. At least eight of the dyads that can be extracted from Table 2 exhibit this pattern (X > Y = X is predominant lexical donor to Y): (1) Tagalog > Pangasinan, (2) Tagalog > Sambal, (3) Tagalog > Kapampangan, (4) Bisayan > Inati, (5) Bisayan > Kalamian Tagbanwa, (6) Danaw > Bilic, (7) Gorontalo-Mongondow > Sangiric, (8) Gorontalo-

Mongondow > Minahasan. In two other cases, those of Tiruray > Kalamansig Cotabato Manobo and North Mangyan > Hanunóo, there has been significant borrowing by an *R > /g/ language from a donor language which does not show this change, but the available data are insufficient to determine whether Tiruray or the North Mangyan languages are in fact predominant donor languages. As seen already, Tiruray has borrowed heavily from an *R > /g/ language. In Blust (to appear) this donor language is assumed to be Danaw, but many Tiruray loans may in fact be from Manobo sources. In any event, the partial lexical convergence of Tiruray and Kalamansig Cotabato Manobo appears to be a relatively recent event in comparison to the widespread and apparently earlier dissemination of loans from *R > /g/ languages throughout much of the central and southern Philippines. In the second case, there simply is too little published data available on the North Mangyan languages to determine whether North Mangyan lexical borrowing on Hanunóo (and other South Mangyan languages) exceeds lexical borrowing in the opposite direction.

The one other case that appears to be a clear exception to the general pattern of borrowing *from* (rather than *into*) *R > /g/ languages in the central and southern Philippines is that of Sama-Bajaw (a non-Philippine language, with *R > /h/) and Taosug (a Central Philippine language, with *R > /g/). According to Pallesen (1985), Taosug convergence to Sama-Bajaw has been far more extensive than Sama-Bajaw convergence to Taosug. As will be seen, the apparently exceptional character of the Taosug-Samalan case (in which the *R > /g/ language is the recipient rather than the donor of most loan features) actually conforms to a more general pattern.

PAN *R probably was a voiced uvular trill. It has become /g/ (generally merging with PAN *g) in at least the following cases: (1) Atayal of northern Taiwan (but not in the closely related Sediq); (2) the extinct Favorlang and Siraya of southwestern Taiwan; (3) the Northern Cordilleran languages of northeastern Luzon; (4) the South Mangyan languages; (5) the Palawanic languages; (6) the Central Philippine languages; (7) the Manobo languages; (8) the Danaw languages; (9) the Subanun languages; (10) the Gorontalo-Mongondow languages; (11) some of the Tomini languages (possibly conditioned); (12) the languages of Sabah; (13) Berawan of northern Sarawak (but not the closely related Kiput or Narum); (14) Chamorro of western Micronesia. There can be no doubt that the changes in Formosan languages, Northern Cordilleran, the languages of Sabah, Berawan, and Chamorro were independent of those in the central and southern Philippines. However, the question remains whether the change *R > /g/ in the latter languages is historically the product of several parallel changes, or of a single change in one language ancestral to the entire group.

Before attempting to answer this question it will be worthwhile first to consider another set of observations which initially may appear to be unrelated, but which I believe are a product of the same historical event which produced Conant's "stereotyped Philippine *g*" throughout the central and southern Philippines.

5. LINGUISTIC DIVERSITY IN THE CENTRAL PHILIPPINES. Historical linguists of various persuasions (Sapir 1916, Dyen 1956) have used the "index of highest diversity" (together with the "principle of least moves") as an indicator of relative time depth for the presence of related languages in different geographical areas. As with the similar principle in botany, the basic assumption is that greater divergence (however this is measured) translates into a longer chronological sequence *in situ*.

If we apply the index of highest diversity to the languages of the Philippines we arrive at rather puzzling results. On the island of Luzon at least three distinct groups of languages (Bashiic, Cordilleran, Central Luzon) are represented north of Manila Bay, and apparently have been in their present locations for a considerable period of time. South of Manila Bay only one Philippine subgroup is represented: Central Philippines. The same is essentially true through the Visayan Islands: with the minor exception of Inati, spoken by a few hundred Negritos in the mountains of Panay, all languages of the Visayas belong to Central Philippines. In Mindoro the situation is somewhat more complex, since here two rather distinct subgroups abut. Similarly, in Palawan and the adjacent Calamian Islands two rather distinct subgroups are found. Finally, the linguistic situation in Mindanao is comparable to that in Luzon north of Manila Bay: the Samalan languages appear to fall outside the Philippine subgroup altogether, and the remaining relatively noncontroversial groups (East Mindanao, Manobo, Danaw, Subanun) all appear to be rather closely related to one another. A consideration of the distribution of Philippine subgroups, then, clearly indicates that southern Luzon and the Visayan Islands are an area of low linguistic diversity in comparison with northern Luzon and Mindanao.

The same conclusion is reached if we consider the number of languages in relation to space occupied, without reference to subgrouping. Luzon is approximately 40,000 sq. mi., of which some three-quarters are north of Manila Bay. Within this area, according to Wurm and Hattori (1981), there are some 48 distinct languages, or one language per 625 sq. mi. South of Manila Bay, on the other hand, there are (according to the same source) only five languages, or one language per 2,000 sq. mi. Table 3 maps the relation of number of languages to area (in square miles) for the entire archipelago.

TABLE 3. LANGUAGE DENSITIES IN VARIOUS REGIONS OF THE PHILIPPINES

ISLAND(S)	AREA	NO. LGS.	RATIO
N. Luzon	30,000	48	1/625
S. Luzon	10,000	5	1/2000
Visayas	23,000	15	1/1533
Mindoro	3,922	7	1/560
Palawan	5,697	10	1/570
Mindanao	36,537	44	1/830

In general, when we have reason to believe that two geographical areas have been settled for an equivalent time by speakers of related languages, but differ considerably in linguistic diversity, it is safe to assume that the area of lesser diversity has undergone an episode of prehistoric language levelling. Such an assumption was made in Blust (1978) for southern Halmahera, which shows very little linguistic diversity despite indirect evidence for the presence of Austronesian speakers in northeastern Indonesia for upwards of 4,000 years. Historical examples of the process of language levelling are, of course, also known. During the nineteenth century the Iban, who were earlier concentrated in the region of the upper Kapuas River along the Kalimantan–Sarawak border, began to expand into southern Sarawak, whence they continued to spread northward during the twentieth century as far as the border of Brunei (Sandin 1967). In the process of this expansion some weaker groups such as the Seru Dayaks were exterminated.

To a certain extent the above presentation is, admittedly, arbitrary. If we include the Kalamian group, and Inati of Panay, the central Philippines loses a good deal of its otherwise surprising linguistic homogeneity. But in some ways the presence of these groups actually reinforces the argument that a significant episode of prehistoric linguistic levelling affected the central Philippines. It is difficult to see Inati as anything other than a remnant of what once was a more widely distributed language group, and the Kalamian group gives every indication of having been settled in its present location for a lengthy time in comparison with its Bisayan neighbors.

6. THE GREATER CENTRAL PHILIPPINES HYPOTHESIS. In this section it is argued that the *R > /g/ languages of the central and southern Philippines and the Gorontalo-Mongondow languages of Sulawesi are descended from a single language that merged PPH *R and *g. Because

its politically most important member languages belong to the Central Philippine group, I call this wider subgroup "Greater Central Philippines."

The essential body of evidence supporting the Greater Central Philippines group is not phonological, but lexical. In certain important respects, the views that I will develop in this section were adumbrated by Zorc (1974). At the conclusion of his paper "Internal and external relationships of the Mangyan languages," Zorc (1974:593) pointed out that "although the forms *Danum 'water' and *DaRaq 'blood' are spread throughout the Philippines and are inherited from Proto-Austronesian, both forms have undergone replacement in the central and southern Philippine area (going as far south as Mongondow on Celebes). At some point in the mutual history of these languages, there must have been competition when *wahiR and *tubig began to replace *Danum, and when *Duruq began to replace *Daraq. In this regard there may be a kind of relative chronology."[12] Presumably based on observations such as these, Zorc (1986) has proposed a "Southern Philippines" subgroup that includes: (1) Bilic, (2a) Manobo, (2b) Danaw, (2c) Subanun, (3a) Central Philippine, (3b) South Mangyan, (3c) Palawanic, (3d) Kalamian Tagbanwa, (4) Mongondow, (5) Gorontalo, and (6) Sangiric plus Minahasan.

Zorc is one of the preeminent scholars in the Philippine and general Austronesian field, whose work is characterized by meticulous attention to detail. Nonetheless, in the passage that I have quoted, I believe he missed an opportunity to develop an important idea which has major repercussions for the subgrouping of Philippine languages, and fundamental implications for Philippine prehistory.

Both *wahiR 'fresh water; stream, river' and *danum 'fresh water' can be reconstructed for Proto-Malayo-Polynesian, and although the former survived in the Manobo and Danaw languages, *tubig was innovated alongside it with no presently determinable difference of meaning. Similarly, PMP *daRaq and *zuRuq can be reconstructed with the meanings 'blood', and 'sap, juice, gravy, soup' respectively, but the former item has been lost and the latter shifted to the meaning 'blood' in many of the languages of the central and southern Philippines. What is significant about these comparisons is the set of languages which they both delineate. As can be seen in Appendix 1, reflexes of the lexical innovation *tubig appear in the following microgroups: (1) Central Philippines, (2) Danaw, (3) Subanun, and (4) Gorontalo-Mongondow. Reflexes of the semantic innovation *duguq 'blood' appear in: (1) South Mangyan, (2) Palawanic, (3) Central Philippines, (4) Danaw, (5) Subanun, and (6) Gorontalo-Mongondow. It is noteworthy that neither change appears in Bilic, Sangiric, or Minahasan, and although Kalamian Tagbanwa has *dugu?* 'blood', its reflex of *R reveals it to be a loan. Moreover, a reflex of *daRaq

'blood' is found in Tiruray (*daraʔ*), Proto-Sangiric (*daRa), and Proto-Minahasan (*dahaʔ). Evidently, then, it is too general a statement to say that "both forms have undergone replacement in the central and southern Philippine area."

What is most noteworthy about these two lexical distributions is the support they lend to a hypothesis that the Gorontalo-Mongondow languages of Sulawesi subgroup immediately with what can conveniently be described as the *R > /g/ languages of the central and southern Philippines, whereas the Sangiric and Minahasan languages do not. If this were all the evidence that supported such a view it would be intriguing, but inconclusive.

The Appendix presents 94 proposed lexical or semantic innovations which appear at the present time to be exclusively shared by Gorontalo-Mongondow languages with one or more of the *R > /g/ microgroups of the central and southern Philippines. Most of these sets were initially identified through a comparison of Bolaang Mongondow (Dunnebier 1951) with one or more of the *R > /g/ languages of the central and southern Philippines. Some additional comparisons were uncovered through a comparison of Gorontalo (Pateda 1977) or of the Proto-Gorontalic reconstructions in Usup (1986) with the same set of languages in the Philippines. Many more sets undoubtedly remain to be discovered when time and fuller lexical resources for the Gorontalo-Mongondow languages permit.

Initially the set of proposed Greater Central Philippines (GCP) innovations numbered over 150, but this set was reduced to its present number through a fairly careful search for external cognates in the following languages: Itbayaten, Ilokano, Isneg, Ifugaw, Bontok, Kankanay, Itawis, Casiguran Dumagat, Pangasinan, Kapampangan, Botolan Sambal, Kalamian Tagbanwa, Tiruray, Sangir, and Proto-Minahasan. It is possible that cognates will ultimately be found in other languages, or even in the above set of languages, but if so these should be few and of only minor significance for the argument presented here.

As is often the case in subgrouping, borrowing has proven to be a complicating factor in testing the GCP hypothesis. Among the types of controls often invoked as tests of probable borrowing are: (1) basic vs. nonbasic vocabulary, (2) diagnostic reflexes, and (3) distribution type. As shown in Blust (to appear), basic vocabulary can be borrowed in far greater amounts than is commonly believed to be the case. What, then, of diagnostic reflexes and distribution type?

If the membership of the GCP group as established by the material in the Appendix is valid, PPH *R and *g merged in Proto-Greater Central Philippines. Kalamian Tagbanwa, which regularly reflects *R as /l/ and *g as /g/, cannot be a descendant of this protolanguage. The fact that it

contains a reflex of PGCP *duguq 'blood' (PMP *zuRuq 'sap, juice, gravy, soup') must therefore be attributed to borrowing. In this case a geographically contiguous external witness with an irregular reflex of what otherwise appears to be an exclusively shared innovation does not affect the conclusion that the innovation originated in Proto-Greater Central Philippines. The reverse of this situation is seen in reflexes of PGCP *taŋag, but Tiruray *taŋar* 'elope'. Since the latter form (with *R > /r/) must be native, we are left no choice but to reconstruct PPH *taŋaR 'elope' based on cognates that to date have been observed only in GCP languages and the geographically contiguous Tiruray.

In the foregoing examples we have a clear phonological basis by which to identify a form as borrowed or native, but this is not always the case. The most serious subgrouping problems created by lexical borrowing arise where the known distribution of a form is confined to members of a putative subgroup, together with geographically contiguous non-members to which they contributed early loanwords. If the item in question does not have a diagnostic reflex, a rough estimate of the probability that it has been borrowed can be derived from the general intensity of borrowing as inferred from cases where diagnostic reflexes are available. Thus Inati *bobol* 'feather, hair' almost certainly is a Bisayan loan, since reflexes are otherwise confined to GCP languages (and to Tiruray *bubul*, where diagnostic reflexes mark it as a loan). Our relative certainty in this matter stems from the data presented in Table 2, which implies that over two-thirds of the Inati lexicon may consist of Bisayan loans. But what are we to do with lexical distributions that appear to be confined to GCP and Sangir, where less than 5% of the vocabulary shows irregularities indicative of borrowing from a GCP source? Surely there is a lower probability that such distributions contain loanwords, although borrowing cannot be entirely ruled out. Nonetheless a number of comparisons have been uncovered which appear to be confined to GCP and Sangir.

Since borrowing is a source of "noise" in establishing true innovations, it might be suggested that the Gorontalo-Mongondow forms in the Appendix could also be loans from languages in the Philippines. Zorc (1982:313) has ably drawn attention to the pitfalls of subgrouping on the basis of lexical evidence, but he believes that given the proper precautions such data may be used with some measure of confidence. Among the precautions that he advises is to: "Consider the character and quality of each lexical innovation, including its geographical and linguistic distribution, potential spread, etc." In general languages which are not geographically contiguous are less likely to be (or to have been) in a borrowing relationship than languages which are geographically contiguous. There are, of course, additional factors which play a part in determining whether or not borrowing is likely to occur, how intensive it will be, and so forth,

but continuous and noncontinuous distributions are fundamentally distinct in the evaluation of explanatory hypotheses, whether they relate to language or to nonlinguistic culture (Blust 1981*a*). In the case at hand, it is far less likely that the Gorontalo-Mongondow languages have borrowed the forms in the Appendix from languages in the Philippines than that, for example, the Tomini languages (especially those immediately bordering the Gorontalo and Buol language areas) have borrowed such words as Tomini *olat*, Boano *oat* 'to wait', or Lauje, Tialo *toga* 'lamp' from a Gorontalo-Mongondow (G-M) source.

Perhaps the most striking feature of the GCP hypothesis is the linkage that it proposes between geographically noncontiguous microgroups. As noted above, it is this distributional feature which enables us with some confidence to dismiss borrowing as a likely explanation for the exclusively shared innovations that have been collected so far. The same feature is difficult to reconcile with the view that the GCP language group acquired its present distribution through a process of gradual splitting without significant migration. If they do indeed subgroup immediately with languages in the central and southern Philippines, how could the G-M languages have reached their present location, to the south and west of both the Sangiric and the Minahasan microgroups, via a process of change which involved only gradual splitting *in situ*? It is difficult to avoid the conclusion that at some point in their history, the G-M languages were transported several hundred miles from the GCP homeland through the migration of their speakers southward past the Sangiric and Minahasan groups which preceded them in northern Sulawesi.

There is nothing revolutionary about the suggestion that the G-M languages reached their present location as the result of a migration. Much longer migrations of Austronesian-speaking peoples have been inferred from linguistic evidence (e.g. the Malagasy migration from Southeast Borneo). What makes the G-M migration of greater than ordinary interest is that it appears to be only part of a larger population movement that affected much of the central and southern Philippines.

At least two types of observations support this view. First, as shown in Table 2, all languages in the central and southern Philippines which do not regularly reflect *R as /g/ have some /g/ reflexes of *R. Second, apart from a few small remnant linguistic groups such as Kalamian and Inati, there appears to be considerably greater linguistic diversity in the far north and the far south of the Philippine archipelago than in the more central region of the Visayas and southern Luzon.

The simple existence of Conant's "stereotyped *g*" could be explained as the product of borrowing between geographically contiguous languages, many of which had independently undergone the change in question. However, the language which regularly underwent the change *R > /g/ almost invariably is the predominant donor language. From this observa-

tion we can reasonably infer that languages showing the regular change *R > /g/ tended to have superior prestige in these prehistoric contact situations. If the merger of *R and *g occurred repeatedly rather than only once in the linguistic history of the central and southern Philippines we must ask why the languages which underwent this change appear to have acquired prestige with it.

The one clear counterexample to the pattern observed in Table 2 perhaps provides a clue to the mystery. Pallesen (1985) notes that Sama-Bajaw has exerted greater contact influence on Taosug (a Central Philippine language) than the reverse. His reconstructed culture-historical scenario shows the Sama-Bajaw as an intrusive population that arrived in northeast Mindanao from somewhere in Indonesia somewhat over 700 years ago, ultimately inducing the Taosug themselves to emigrate from this area to the region of the Sulu Sea. In this case it was the dynamic, migratory Sama-Bajaw who tended to dominate the (initially) stationary Taosug. The most forceful apparent counterexample to our thesis thus appears to confirm the thesis in a more general form: it is not languages that underwent the merger of *R and *g which tended to have greater prestige in prehistoric contact situations in the central and southern Philippines, but rather the language of a population expanding beyond its earlier borders.

Finally, if the G-M languages do belong to a GCP subgroup, what is their place within this group? A quick tabulation of exclusively shared innovations in the Appendix shows that the G-M languages most often share proposed GCP innovations with the Central Philippine group (65), then with Maranao (43), Western Bukidnon Manobo (40), Hanunóo (26), Subanun (16), and Palawan Batak (11). Similar results are obtained in considering only innovations that appear to be exclusively shared *within* the proposed GCP group: Central Philippine (17), Maranao (6), Western Bukidnon Manobo (6), Hanunóo (3), Subanun (1), and Palawan Batak (none). The difficulty with using these figures for subgrouping purposes is that they closely match the richness of the lexicographical resources for each microgroup in question. The Central Philippine languages are represented by several extensive dictionaries, Maranao by a single large dictionary, Western Bukidnon Manobo and Hanunóo by somewhat smaller dictionaries, and Subanun and Palawan Batak only by wordlists. On present evidence, then, there are no strong grounds for considering the G-M languages anything other than a primary branch of Greater Central Philippines.

7. WAS THERE A PROTO-SOUTHERN PHILIPPINES? The classification of languages in the central and southern Philippines proposed by Zorc (1986) differs in three important respects from that proposed in this paper: (1) the position of Kalamian, (2) the position of Bilic, and (3) the

position of the languages of northern Sulawesi. Whereas Zorc includes all of these languages in a proposed Philippine subgroup that he calls "Southern Philippines," I exclude Kalamian and Bilic, and include only the Gorontalo-Mongondow languages among the languages of North Sulawesi in the group I call "Greater Central Philippines."

Space will not allow me to treat the classification of Kalamian here, but the position of the Bilic languages is discussed at some length in Blust (to appear). The aim of this section is to examine, at least briefly, the evidence for including all three North Sulawesi microgroups in a "Southern Philippines" group.

In his passing comments on Philippine subgrouping, Zorc (1986) separates Mongondow and Gorontalo, but conjoins Sangiric and Minahasan (thus: Mong., Gor., Sn. + Mn.). The existence of a Gorontalo-Mongondow subgroup seems now to be well-established (Usup 1981, 1986; Noorduyn 1982; Sneddon and Usup 1986), and will not be considered further. The linking of Sangiric and Minahasan was initially proposed by Sneddon (1978:10, 1984:11–12), who clearly is the source for the position taken by Zorc. However, as noted earlier, Sneddon (1989*a*) now rejects this linkage, maintaining that none of the three North Sulawesi microgroups is particularly close to any other. He nonetheless defers to Philippinists regarding the existence of a "Southern Philippine" group (1989*a*:103):

> 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 Southern Philippines branch, but no suggestions have been offered as to which languages within this 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.

Evidence has already been presented for separating the G-M languages from the other North Sulawesi microgroups, but it is perhaps worth asking whether any other type of evidence can be brought to bear on the question. I have been, and remain a critic of lexicostatistics, which can give a seriously distorted picture of subgrouping relationships because it fails to distinguish innovations from retentions. However, to satisfy the curiosity of those who might wonder, Table 4 presents a calculation of the lexicostatistical percentages linking Itbayaten (Bashiic), Tagalog (Central Philippines), Maranao (Danaw), Tagabili (Bilic), Sangir (Sangiric), Tondano (Minahasan), and Bolaang Mongondow (Gorontalo-Mongondow):

TABLE 4. LEXICOSTATISTICAL PERCENTAGES LINKING SEVEN LANGUAGES OF THE PHILIPPINES AND NORTHERN SULAWESI

	Itbayaten	Tagalog	Maranao	Tagabili	Sangir	Tondano
Bolaang Mongondow	56/198 (28.3)	63/198 (31.8)	70/198 (35.4)	41/196 (20.9)	50/199 (25.1)	40/199 (20.1)
Tondano	36/198 (18.2)	47/198 (23.7)	39/198 (19.7)	33/196 (16.8)	50/199 (25.1)	
Sangir	51/198 (25.8)	53/200 (26.5)	56/198 (28.3)	40/196 (20.4)		
Tagabili	35/196 (17.9)	48/195 (24.6)	50/195 (25.6)			
Maranao	54/198 (27.3)	78/197 (39.4)				
Tagalog	50/198 (25.3)					

Because the structure of a lexicostatistical classification depends in critical respects on the *linking* of percentages, a sample of say, 70 languages in the Philippines and northern Sulawesi undoubtedly would give somewhat different results than those found here. Nonetheless, the one clear subgroup indicated by the percentages in Table 4 includes Tagalog, Maranao, and Bolaang Mongondow, the first two languages linked by a common percentage of 39.4, and this group linked with Bolaang Mongondow by an averaged percentage of 33.6 (half of 31.8 + 35.4). The next highest percentage found between any pair of languages in Table 4 is 28.3. In addition, although Tondano scores somewhat higher with Sangir than with any other language in the sample, the percentages of Sangir with most languages in the Philippines are slightly higher than those with Tondano. The general lexicostatistical picture, then, is completely consistent with the view that the Sangiric and Minahasan microgroups are primary branches of the Philippine language group, whereas the Gorontalo-Mongondow languages subgroup immediately with such central and southern Philippine microgroups as Central Philippines and Danaw.[13]

To the extent that lexicostatistical percentages can be made to yield reliable estimates of separation times (and this is certainly questionable), the averaged percentage 35.5% linking the pairs Bolaang Mongondow: Tagalog, Bolaang Mongondow: Maranao and Tagalog: Maranao in Table 4 translates into a separation time of approximately 500 B.C. Whether this data is wholly accurate or not, it does provide a general linguistic indication of a major event in Philippine prehistory, one that should be visible in the archaeological record.

To summarize, around 500 B.C., for reasons unknown, speakers of Greater Central Philippines began to expand outward from a center some-

where in northern Mindanao or the southern Visayas. Through conquest and absorption of weaker populations, they reduced the linguistic diversity of the Visayas, Palawan, and southern Luzon. Only a few remnant populations, such as the Inati of Panay and the Kalamian group survived. Other languages in southern Palawan which reflected *R as /r/ disappeared, but not without leaving loans in the languages of their successors. The ancestral Central Luzon group, which may have extended to northern Palawan (as suggested by occasional *R > /y/ reflexes in Palawan Batak) held its ground, but contracted its territory. Last, but not least, one branch of Greater Central Philippines migrated southward past the Sangiric and Minahasan peoples to establish a foothold on the northern peninsula of Sulawesi.

Before concluding this section two final comments are perhaps in order. First, many of the lexical reconstructions that appear in Zorc (1971) are attributed to Mathew Charles, whose comparative work was based to a very large extent on the comparison of Bolaang Mongondow with various languages in the Philippines. Since many of these languages are assigned here to a Greater Central Philippine subgroup, it appears likely that a number of the "Proto-Philippine" reconstructions proposed by Charles actually are GCP innovations. More work needs to be done to determine the extent to which this may be true.

Second, the culture-historical scenario outlined above turns critically on the notion that languages may disappear if their speakers are overwhelmed by technically superior populations that speak a different language. Where historical documentation is available we know that this has been the case throughout recorded history. In 500 B.C. at least five languages belonging to two different language families were spoken in the Italian Peninsula (the non-Indo-European Etruscan; the Indo-European Oscan, Umbrian, Latin and Faliscan). As a result of the military success of the Roman Empire a single language is spoken over much of the Italian Peninsula today (and closely related descendants of Latin over a much larger area). As I hope to have shown, there are grounds for inferring a similar kind of prehistoric linguistic levelling in the central Philippines, although the type of "conquest" here need not have been military. More generally, the Philippine archipelago as a whole shows much less linguistic diversity than one would expect for a region that must have been settled very early in the history of the Austronesian expansions. There thus are grounds for inferring an even earlier episode of linguistic expansion and extinction, one which preceded the dispersal of Proto-Greater Central Philippines by perhaps a millennium. The homeland of Proto-Philippines and the linguistic situation in the Philippine archipelago prior to the dispersal of this language remain subjects of speculation, although it appears likely, as Reid (1987) suggests, that some non-Austronesian languages were spoken by Philippine Negrito

hunter-gatherers prior to intensive contact with the incoming waves of Austronesian speakers.

8. EPILOGUE: BIOLOGICAL MODELS IN HISTORICAL LINGUISTICS. At the conclusion of a paper to which I have now referred repeatedly because of its importance, Zorc (1986:156) expressed a reluctance to draw a Philippine language tree since in his opinion "Ph developments were more like amoebic colonizations than absolute splits." This view of language splitting has an analogy in the Neogrammarian view that sound change is gradual, and just as saltatory sound changes (such as metathesis) are difficult to accommodate in such a theory, so are abrupt language cleavages difficult to accommodate under the view that language splitting consists exclusively of "amoebic colonizations."

In this paper I have not only proposed a new subgrouping of the languages of the central and southern Philippines, but I have in addition proposed a different model of linguistic change than is often assumed by comparativists. As it turns out (post hoc) it bears a striking resemblance to certain recent models of the process of speciation in evolutionary biology.

Biological models have been fashionable in historical linguistics at least since Schleicher's invention of the family tree. The model of change that I have been forced to assume in order to explain the data considered in this paper has interesting general parallels to what Gould and Eldredge (1977) have called "punctuated equilibrium." As noted by Gould and Eldredge, in classical Darwinian evolution, speciation is the result of a slow and steady process of environmental change acting adaptively on inherent genetic variation. In this view all evolutionary change should appear as a linking of infinitely graduated transitional forms. The fossil record of life on earth, however, reflects no such graduated linkage. Instead, it reflects eons of statis punctuated by (geologically) brief episodes of explosive change. The triggering event in such episodes appears to be the mass extinction of species through environmental cataclysm, followed by the rapid spread and modification of some surviving species as it radiates into a variety of vacated microenvironments. The classical event of this kind is the great extinction at the end of the Cretaceous period, which witnessed the abrupt disappearance of the dinosaurs and the sudden efflorescence of a plethora of new mammalian species.

When languages expand at the expense of their contemporaries the triggering event in the reduction of diversity is not some external change, but the expansion and influence of the dominant population itself. Moreover, unlike biological change, which in the currently dominant view is essentially an opportunistic response to periodic environmental alteration, language change appears to be a continuous process. In this way the

comparison with the mechanisms of speciation is inexact. There is nonetheless enough general similarity in the outcome of extinction and repopulation in the two cases to warrant a comparison between them as a reminder that for all its variety Nature is filled with haunting parallels and recurrences in domains as diverse as the origin of species and the history of languages.[14]

NOTES

1. Lawrence A. Reid, James N. Sneddon, and R. David Zorc provided valuable critical commentary and relevant data which led to improvements in an earlier version of this paper. Any surviving deficiencies are my responsibility alone.
2. In fact, Conant does not explicitly indicate whether the merger of *R and *g is taken to define the category "Philippine language" or only a subset of Philippine languages. If the former, it follows that those languages in the Philippines which have not merged *R with *g are not, by Conant's "customary" definition, Philippine languages. If the latter, it follows that some languages in the Philippines (e.g. Ibanag, with *R > *g*) were regarded as more closely related to some languages outside the Philippines (e.g. Chamorro, with *R > *g*) than to other languages within the Philippines (e.g. Ilokano, with *R > *r*). Rather than an overlap relation, this would be a relation of limited inclusion. Nothing in Conant (1911) nor in any of his other publications on the languages of the Philippines states or implies the latter view, and I consequently disregard it here.
3. Charles's inclusion of the languages of northern Borneo in a Philippine subgroup was motivated at least in part by the presence of similar systems of focus-marking in both areas. A similar reasoning led Topping, Ogo, and Dungca (1975:ix) to suggest that Chamorro is most closely related to languages in the Philippines. As noted in Blust (1974), however, arguments of this type beg the question whether the similarities used as subgrouping evidence are innovations or retentions.
4. The figure obtained for Kakidugen Ilongot is 68/199 = 34.2%. An even lower retention percentage (26.9) has since been obtained for Arta (Reid 1989:48).
5. Wimbish (1986) reports that there are at least six different Negrito languages in the Zambales Mountains which are called "Ayta," and which previously were believed to constitute a single language with Sambal. Further data collection and comparative work clearly is needed in this area.
6. The inclusion of Bashiic with Central Luzon appears to be based largely on the merger of PAN *R with *y. This merger has occurred in scattered groups of languages, including Bashiic and Central Luzon in the Philippines, Southeast Barito in Borneo (and Malagasy), Gayō and Lampung in Sumatra, and Sundanese in west Java. Although striking because of its rarity, an innovation of this type by itself has only limited subgrouping value. Reid (p.c.) accepts Zorc's inclusion of the Bashiic group with Central Luzon and North Mangyan, basing his views on the reportedly distinctive use of cross-referencing third person pronominal markers in Ivatan and Kapampangan.
7. For additional unpublished data on Kalamian Tagbanwa, I am indebted to R. David Zorc, who has generously provided me with a vocabulary of approximately 1,500 words.

8. According to Zorc (p.c.), Batangan is also known as Taubuid. Reid (p.c.) informs me that Batangan data "is included in Karl-Josef Barbian's 1977 English–Mangyan Vocabulary, and The Languages of Mindoro, both University of San Carlos, Cebu City." I have been unable to obtain copies of either publication.
9. For the use of "roots" (submorphemic recurrent sound–meaning associations) in conjunction with the standard comparative method see Blust (1988).
10. I recognize only a single PAN palatal affricate, written *z. Dempwolff's distinction of alveolar and retroflex stops (Dyen's *d and *D) is reconstructible only in final position, but no Philippine language is known to reflect *-d and *-D differently.
11. In writing Philippine languages, I have generally adopted the practice of omitting morpheme-initial glottal stop, since it does not contrast with zero. Reid (p.c.), however, has reminded me that the initial glottal stop of the stem often remains under prefixation, and in some cases metathesizes to preconsonantal position. I regard these as cogent arguments for writing morpheme-initial glottal stop, particularly when describing the synchronic phonology of the languages in question. However, few dictionaries of Philippine languages indicate the glottal stop in this position, and to alter the material of my sources in violation of my own longstanding practice would simply invite confusion. Whether one treats the initial glottal stop as part of underlying representations or as the output of phonological rules, it remains the case that PPH *q and zero merged in initial position in the great majority of Philippine languages.
12. Due to a printing error *Duruq and *Daraq appear for intended *DuRuq and *DaRaq.
13. In evaluating the percentages given in Table 4, it should be kept in mind that languages may vary considerably in the retention rate of basic vocabulary over long periods of time (Blust 1981*b*). Tentative retention *percentages* (from which rates can be calculated) for the seven languages in question have been computed as follows: Itbayaten: 40.2, Tagalog: 45.5, Maranao: 48.7, Tagabili: 37.4, Sangir: 49.7, Tondano: 36.7, Bolaang Mongondow: 40.5. These figures represent calculated percentages of Proto-Malayo-Polynesian basic vocabulary retained in the modern languages; percentages of Proto-Philippine basic vocabulary retained in the modern languages may not be isomorphic, and have yet to be calculated.
14. Byron W. Bender (p.c.) has reminded me of other cases of linguistic expansions (e.g. the Bantu expansion into eastern and southern Africa), and notes that such events in general appear to be made possible by the acquisition of "a new tool, resource, or something else that gives demographic advantage" to the territorially expanding population. He suggests further that "external factors triggered both the disappearance of the dinosaurs and the spread of the (Greater Central Philippines) languages: external to organisms and external to languages. . . . I would say that the major difference between the biological and the linguistic examples . . . is that in the biological case, it was the demise of one group (triggered by an external climatic event) that permitted the expansion of another/others—the filling of a vacuum, as it were, whereas inthe linguistic cases it is the shifting of advantage to the expanding group (by an external, cultural event) that enables them to overrun neighbors who haven't come into possession of the new trait yet."

REFERENCES

Allison, E. Joe. 1979. Proto-Danaw: A comparative study of Maranaw, Magindanaw, and Iranun. Papers in Philippine Linguistics, no. 10. *Pacific Linguistics* A-55:53–112.

Antonissen, A. 1958. *Kadazan–English and English–Kadazan dictionary*. Canberra: Government Printing Office.

Appell, George N., and Laura W. R. Appell. 1961. A provisional field dictionary of the Rungus Dusun language of North Borneo. Mimeographed.

Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian archipelago*. London: Academic Press.

Benton, Richard A. 1971. *Pangasinan dictionary*. PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Bergaño, Diego, 1860. *Vocabulario de la lengua Pampanga en Romance*. Manila.

Blake, Frank R. 1906. Contributions to comparative Philippine grammar (Part One). *Journal of the American Oriental Society* 27:317–396.

Blust, Robert. 1974. The Proto–North Sarawak vowel deletion hypothesis. Unpublished Ph.D. dissertation, University of Hawaii.

———. 1978. Eastern Malayo-Polynesian: A subgrouping argument. *Pacific Linguistics* C-61 (Fascicle 1): 181–234.

———. 1981*a*. Linguistic evidence for some early Austronesian taboos. American Anthropologist 83.2:285–319.

———. 1981*b*. Variation in retention rate among Austronesian languages. Paper presented at the Third International Conference on Austronesian Linguistics, Den Pasar, Bali, January 1–11, 1981. Ms., 83 pp.

———. 1984/85. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26:45–67.

———. 1988. *Austronesian root theory: An essay on the limits of morphology*. Amsterdam: John Benjamins.

———. to appear. On speech strata in Tiruray. *Pacific Linguistics*.

Carro, Andrés. 1956. *Iloko–English dictionary*, translated, augmented, and revised by Morice Vanoverbergh. Manila.

Charles, Mathew. 1974. Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics* 13:457–509.

Conant, Carlos Everett. 1911. The RGH law in Philippine languages. *Journal of the American Oriental Society* 31:74–85.

Conklin, Harold C. 1953. *Hanunóo–English vocabulary*. Berkeley and Los Angeles: University of California Press.

Dunnebier, W. 1951. *Bolaang Mongondowsch–Nederlandsch woordenboek*. The Hague: Nijhoff.

Dyen, Isidore. 1956. Language distribution and migration theory. *Language* 32: 611–626.

———. 1965. *A lexicostatistical classification of the Austronesian languages*. International Journal of American Linguistics, Memoir 19. Baltimore: The Waverly Press.

Elkins, Richard E. 1968. *Manobo–English dictionary*. Oceanic Linguistics Special Publication No. 3. Honolulu: University of Hawaii Press.

———. 1974. A Proto-Manobo word list. *Oceanic Linguistics* 13:601–641.

———. 1982. The Proto-Manobo kinship system. In *GAVA': Studies in Austronesian languages and cultures dedicated to Hans Kähler*, ed. by Rainer Carle et al. Berlin: Reimer.

Esser, S. J. 1938. Talen. In *Atlas van Tropisch Nederland.* Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Ferguson, Charles A. 1966. Assumptions about nasals: A sample study in phonological universals. In *Universals of Language*, 2d ed., ed. by Joseph Greenberg, pp. 53–60. Cambridge, Mass.: The M.I.T. Press.

Finley, John Park, and William Churchill. 1913. *The Subanu.* Studies of a Sub-Visayan mountain folk of Mindanao. Publication No. 184. Washington: Carnegie Institution.

Forman, Michael L. 1971. *A Kapampangan dictionary.* PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Forsberg, Vivian, and Alice Lindquist. 1955. Tagabili vocabulary. Manila: Summer Institute of Linguistics. Typescript.

Gould, S. J., and Eldredge, N. 1977. Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* 3:115–151.

Harmon, Carol W. 1977. Kagayanen and the Manobo subgroup of Philippine languages. Unpublished Ph.D. dissertation, University of Hawaii.

Headland, Thomas N., and Janet D. Headland. 1974. *A Dumagat (Casiguran)–English dictionary.* Pacific Linguistics C-28. Canberra: The Australian National University.

Himmelmann, Nikolaus P., ed. 1990. Sourcebook on Tomini-Tolitoli languages. Typescript, 336 pp. Department of Linguistics, University of Köln.

Juanmarti, Jacinto. 1892. *Diccionario Moro–Maguindanao–Español y Diccionario Español–Moro–Maguindanao.* Manila.

Lambrecht, Frans Hubert. 1978. *Ifugaw–English dictionary.* Baguio City, Philippines.

Llamzon, Teodoro A., and Ma. Teresita Martin. 1976. A subgrouping of 100 Philippine languages. In *South-East Asian Linguistic Studies*, vol. 2, ed. by Nguyen-Dang Liem. Pacific Linguistics C-42:141–172. Canberra: The Australian National University.

McFarland, Curtis D. 1977. *Northern Philippine linguistic geography.* Study of Languages & Cultures of Asia and Africa Monograph Series, no. 9. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

———. 1980. *A linguistic atlas of the Philippines.* Study of Languages & Cultures of Asia and Africa Monograph Series, no. 15. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

McKaughan, Howard P., and Batua A. Macaraya. 1967. *A Maranao dictionary.* Honolulu: University of Hawaii Press.

McManus, Edwin G. 1977. *Palauan–English dictionary*, edited and expanded by Lewis S. Josephs with the assistance of Masa-aki Emesiochel. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Mintz, Malcolm Warren, and José del Rosario Britanico. 1985. *Bikol–English dictionary.* Quezon City: New Day.

Molony, Carol, and Dad Tuan. 1976. Further studies on the Tasaday language: Texts and vocabulary. In *Further Studies on the Tasaday* ed. by D. E. Yen and John Nance, pp. 13–96. Panamin Foundation Research Series, no. 2. Makati, Rizal, Philippines: Panamin Foundation.

Motus, Cecile. 1971. *Hiligaynon dictionary.* PALI Language Texts: Philippines. Honolulu: University of Hawaii Press.

Newell, Leonard E. 1968. *A Batad Ifugao vocabulary.* HRAFlex Book OA19-001. New Haven: Human Relations Area Files.

Noorduyn, J. 1982. Sound changes in the Gorontalo language. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2, ed. by

Amram Halim, Lois Carrington, and S. A. Wurm. Pacific Linguistics C-75: 241–261. Canberra: The Australian National University.
Pallesen, A. Kemp. 1985. *Culture contact and language convergence*. Manila: Linguistic Society of the Philippines.
Panganiban, Jose Villa. 1966. *Talahuluganang Pilipino–Ingles*. Manila.
———. 1973. *Diksyunaryo Tesauro Pilipino–Ingles*. Quezon City: Manlapaz Publishing Co.
Pateda, Mansoer. 1977. *Kamus Bahasa Gorontalo–Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
Pennoyer, F. Douglas. 1986/87. Inati: The hidden Negrito language of Panay, Philippines. *Philippine Journal of Linguistics* 18/19:1–36.
Reid, Lawrence A. 1973. Kankanay and the problem *R and *l reflexes. In *Parangal kay Cecilio Lopez: Essays in honor of Cecilio Lopez on his seventy-fifth birthday*, ed. by Andrew B. Gonzalez, pp. 51–63. Quezon City: Linguistic Society of the Philippines.
———. 1976. *Bontok–English dictionary*. Pacific Linguistics C-36. Canberra: The Australian National University.
———. 1982. The demise of Proto-Philippines. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2, ed. by Amram Halim, Lois Carrington, and S. A. Wurm. Pacific Linguistics C-75: 201–216. Canberra: The Australian National University.
———. 1987. The early switch hypothesis: Linguistic evidence for contact between Negritos and Austronesians. *Man and Culture in Oceania* 3 (Special Issue) 41–59.
———. 1989. Arta, another Philippine Negrito language. *Oceanic Linguistics* 28:47–74.
———. 1991. The Alta languages of the Philippines. In *VICAL 2, Papers from the Fifth International Conference on Austronesian Linguistics*, ed. by Ray Harlow, pp. 265–297. Auckland: Linguistic Society of New Zealand.
———, ed. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication No. 8. Honolulu: University of Hawaii Press.
Revel-Macdonald, Nicole. 1979. *Le Palawan (Philippines). Phonologies. Catégories. Morphologie*. Paris: SELAF.
Sandin, Benedict. 1967. *The Sea Dayaks of Borneo before white rajah rule*. Ann Arbor: Michigan State University Press.
Sapir, Edward. 1968 [1916]. Time perspective in aboriginal American culture: A study in method. In *Selected Writings of Edward Sapir in Language, Culture and Personality*, ed. by David G. Mandelbaum, pp. 389–467.
Schlegel, Stuart A. 1971. *Tiruray–English lexicon*. University of California Publications in Linguistics, no. 67. Berkeley.
Schneeberger, W. F. 1937. A short vocabulary of the Banggi and Bajau language. *Journal of the Malaysian Branch of the Royal Asiatic Society* 15:145–64.
Schwarz, J. Alb. T. 1908. *Tontemboansch–Nederlands woordenboek met Nederlandsch–Tontemboansch register*. Leiden: Brill.
Sneddon, J. N. 1978. *Proto-Minahasan: Phonology, morphology and wordlist*. Pacific Linguistics B-54. Canberra: The Australian National University.
———. 1984. *Proto-Sangiric and the Sangiric languages*. Pacific Linguistics B-91. Canberra: The Australian National University.
———. 1989*a*. The North Sulawesi microgroups: In search of higher level connections. In *Studies in Sulawesi linguistics*, part 1, ed. by J. N. Sneddon, pp. 83–107. NUSA: Linguistic studies of Indonesian and other languages in Indonesia, vol. 31. Jakarta.

———. 1991. The position of Lolak. In *VICAL 2, Papers from the Fifth International Conference on Austronesian Linguistic*, ed. by Ray Harlow, pp. 299–318. Auckland: Linguistic Society of New Zealand.

———, ed. 1989*b*. *Studies in Sulawesi linguistics, part 1*. NUSA: Linguistic studies of Indonesian and other languages in Indonesia, vol. 31. Jakarta.

———, and Hunggu Tajuddin Usup. 1986. Shared sound changes in the Gorontalic language group: Implications for subgrouping. Bijdragen tot de Taal-, Land- en Volkenkunde 142(4):407–426.

Steller, K. G. F., and W. E. Aebersold. 1959. *Sangirees–Nederlands woordenboek met Nederlands–Sangirees register*. The Hague: Nijhoff.

Tharp, James A. 1974. Notes on the Ilokano reflexes of Proto-Austronesian *R. *Working Papers in Linguistics* 6(6):47–51. Honolulu: Department of Linguistics, University of Hawaii.

———, and Mateo C. Natividad. 1976. *Itawis–English wordlist, with English–Itawis finderlist*. HRAFlex Books OA29-001. New Haven: Human Relations Area Files.

Thiessen, Henry Arnold. 1977. The phonemic consequences of two morphophonemic rules in Molbog. *Studies in Philippine Linguistics* 1(2):1–26.

———. 1981. Phonological reconstruction of Proto-Palawan. *Anthropological Papers*, no. 10. Manila: National Museum of the Philippines.

Thomas, David, and Alan Healey. 1962. Some Philippine language subgroupings: A lexicostatistical study. *Anthropological Linguistics* 4(9):21–33.

Topping, Donald M., Pedro M. Ogo, and Bernadita C. Dungca. 1975. *Chamorro–English dictionary*. PALI Language Texts: Micronesia. Honolulu: The University Press of Hawaii.

Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi. 1987. Lists of selected words of Batanic languages. Tokyo, Department of Linguistics, Faculty of Letters, University of Tokyo.

Tweddell, Colin E. 1970. The identity and distribution of the Mangyan tribes of Mindoro, Philippines. *Anthropological Linguistics* 12(6):189–207.

Usup, Hunggu Tajuddin. 1981. *Rekonstruksi fonem proto kelompok bahasa Gorontalo sebelah timur*. Research report. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

———. 1986. Rekonstruksi Protobahasa Gorontalo-Mongondow. Unpublished doctoral dissertation, University of Indonesia.

Vanoverbergh, Morice. 1933. *A dictionary of Lepanto Igorot or Kankanay*. Anthropos Linguistic Monograph, no. 12.

———. 1972. *Isneg–English vocabulary*. Oceanic Linguistics Special Publication No. 11. Honolulu: University Press of Hawaii.

Walton, Charles. 1979. A Philippine language tree. *Anthropological Linguistics* 21(2):70–98.

Warren, Charles P. 1959. A vocabulary of the Batak of Palawan. Philippine Studies Program, Transcript No. 7. University of Chicago.

Wimbish, J. 1986. The languages of the Zambales Mountains: A Philippine lexicostatistic study. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 30:133–142.

Wolff, John U. 1972. *A dictionary of Cebuano Visayan*. Special Monograph No. 4, Manila: Linguistic Society of the Philippines.

Wurm, S. A., and S. Hattori, eds. 1981. *Language atlas of the Pacific area*, part 1, *New Guinea area, Oceania, Australia*. Canberra: The Australian Academy of the Humanities in collaboration with the Japan Academy.

Yamada, Yukihiro. 1976. A preliminary dictionary of Itbayaten. Ms., 404 pp. Kochi, Japan.
Zorc, R. David. 1969. *A study of the Aklanon dialect*, vol. 2, *Dictionary*. Kalibo, Aklan: Public Domain.
———. 1971. Proto-Philippine finder list. Typescript, Cornell University.
———. 1974. Internal and external relationships of the Mangyan languages of Mindoro. *Oceanic Linguistics* 13:561–600.
———. 1977. *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Pacific Linguistics C-44. Canberra: The Australian National University.
———. 1982. Micro- and Macro-subgrouping: Criteria, problems and procedures. In *GAVA': Studies in Austronesian languages and cultures dedicated to Hans Kähler*, ed. by Rainer Carle et al. Berlin: Dietrich Reimer.
———. 1986. The genetic relationships of Philippine languages. In *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, ed. by Paul Geraghty, Lois Carrington, and S. A. Wurm. Pacific Linguistics C-94:147–173.
———. n.d. Unpublished fieldnotes on Kalamian Tagbanwa.

APPENDIX: PRELIMINARY DATA SETS

The languages from which data is cited in the appendix are segregated by microgroup and numbered as in Section 2. This format serves to highlight the linking of Greater Central Philippines microgroups into a larger genetic unit. For practical reasons I have listed only sets that include data both from the Philippines and from northern Sulawesi, although many more sets could be cited which link GCP microgroups only within the Philippines. Language abbreviations and sources of data are as follows.

Akl.: Aklanon (Zorc 1969)
Banggi (Schneeberger 1937)
Bkl.: Bikol (Mintz and Britanico 1985)
Bkd.: Binukid (Reid 1971)
Boano (Himmelmann 1990)
BM.: Bolaang Mongondow (Dunnebier 1951)
Btk.: Palawan Batak (Warren 1959, Reid 1971)
Ceb.: Cebuano Visayan (Wolff 1972)
GCP: Greater Central Philippines
Gtl.: Gorontalo (Pateda 1977)
Han.: Hanunóo (Conklin 1953)
Hlg.: Hiligaynon (Motus 1971)
Kdp.: Kaidipang (Usup 1981)
Klg.: Kalagan (Reid 1971)
Lauje (Himmelmann 1990)
Mar.: Maranao (McKaughan and Macaraya 1967)
MbAd.: Dibabawon Manobo (Reid 1971)
MbS.: Sarangani Manobo (Reid 1971)
Mmn.: Mamanwa (Reid 1971)
Msk.: Mansaka (Reid 1971)

Pal.: Palauan (McManus 1977)
PGTL: Proto-Gorontalo (Usup 1981)
PPH: Proto-Philippines
Sbl.(Bt.): Botolan Sambal (Reid 1971, McFarland 1977)
Sub: Subanun, dialect unstated (Finley and Churchill 1913)
Sub.(S.): Sindangan Subanun (Reid 1971)
Sub.(Sc.): Siocon Subanun (Reid 1971)
Tag.: Tagalog (Panganiban 1966, 1973)
Tao.: Taosug (Reid 1971)
Tbw.(A.): Aborlan Tagbanwa (Reid 1971)
Tbw.(K.): Kalamian Tagbanwa (Reid 1971, Zorc n.d.)
Tialo (Himmelmann 1990)
Tir.: Tiruray (Schlegel 1971)
WBM.: Western Bukidnon Manobo (Elkins 1968)

(1) *anda 'where?'
2.11 Mar. anda 'where?, when?'; anda on 'where?, which?'
2.15 BM. onda 'where?, which?, what?'
NOTE: The initial vowel correspondence is irregular, Mar. reflecting PPH *a and BM. reflecting PPH *e. The Mongondow form may have been cliticized to most following words, resulting in the normal reduction of prepenultimate *a to schwa, which regularly became /o/.

(2) *alut 'shave off'
2.9 Ceb. alút 'shave the head, cut hair; haircut'
2.10 WBM. alut 'cut hair'
2.11 Mar. alot 'haircut'
2.15 Gtl. waluto 'whittle, shave off (as wood, rattan)'

(3) *apid 'stacked up, in layers'
2.9 Tag. ápid 'illicit coitus'
Ceb. ápid 'arrange things of approximately the same size in a neat stack'; apid-ápid 'be stacked'
2.10 WBM. apid 'of flat objects with considerable surface or of layers of clothing, to be one on top of the other'
2.11 Mar. apid-apid 'successive generations; something arranged by layers'
2.15 Gtl. wapidu 'put in stacks or layers'
NOTE: Evidently distinct from PWMP *apid 'braid'.

(4) *baketín 'piglet, pig (said in anger)'
2.7 Han. baktín 'suckling pig'
2.9 Bkl. baktín 'pig, said in anger'
Ceb. baktín 'piglet'
2.10 WBM. beketin 'baby pig'
2.11 Mar. baktiŋ 'baby pig, piglet'
2.15 PGTL *bokotiŋo 'pig'
Kdp. bokotiŋo 'pig (said in anger toward s.o. who is compared to a pig)'
NOTE: Mar. -ŋ is irregular. A final velar nasal in the Gorontalic languages can reflect either *-ŋ or (rarely) *-n.

(5) *bakús 'tie, bind'
2.9 Akl. bakós 'tie around the waist'
Ceb. bakús 'belt'
2.10 WBM. bakus 'bind or tie up a person or animal'
2.15 BM. bakut 'pack, wrap up'

NOTE: With root *-kus 'wind around; bundle'. Sneddon (p.c.) points out that those Gorontalic languages which distinguish *s from *t reflect a *t in this form. This item may therefore have undergone an irregular change in Proto-Gorontalo-Mongondow, or its similarity to Bisayan and Manobo forms may be purely a product of chance.

(6) *balaŋas 'fruit tree: *Nephelium* sp.'
2.10 WBM. belaŋas 'an uncultivated tree, *Nephelium mutabile*, which bears an edible fruit which has a fatty seed'
2.11 Mar. balaŋas 'wild fruit tree with edible fruit'
2.15 BM. boḷaŋat 'the rambutan: *Nephelium lappaceum*'

(7) *baraw 'argue with'
2.10 WBM. barew 'deny an accusation'
2.15 BM. bayow 'insolent, rude, not give a damn; to contradict, argue with'

(8) *baruy 'pandanus'
2.7 Han. báriw 'a pandanus or screw pine (probably *Pandanus copelandii* Merr.)'
2.10 WBM. baruy 'generic for the various species of the genus pandanus which occur locally'
2.15 BM. bayui 'pandanus'

NOTE: Reid (p.c.) suggests a relationship with Agta *bidiyu* 'pandanus', but the phonological correspondences appear to be without parallel.

(9) *bátaq 'child'
2.9 Tag. bátaʔ 'child'
Akl. bátaʔ 'childish, young, juvenile; immature'
Hlg. bátaʔ 'child, baby, young girl or boy usually below the age of ten'
Ceb. bátaʔ 'child; son or daughter; mistress, concubine; bodyguard, protege of someone of high rank'
2.10 WBM. bataʔ 'child'
2.11 Mar. wataʔ 'child, baby'
2.12 Sub.(Sc.) bataʔ 'child'
2.15 Gtl banta 'child'

NOTE: The ordinary Gtl. word for 'child' is *walaʔo*. Pateda (1977) gives *banta* as 'anak' in the expression *ti pu:tiri banta lo longiya* 'sang puteri anak raja' = 'the daughter (honorific) of the king'. As Sneddon (p.c.) has reminded me, if the Gorontalo form reflects *bataq it should be **bota or **bata. I nonetheless retain it here in the belief that it is cognate, and that some explanation for the irregularity eventually will be found.

(10) *beŋkel 'tie around'
2.11 Mar. beŋkel 'tie around'
2.15 BM. boŋkoḷ 'kerchief, girdle, belt'

(11) *bi(n)tanʔág 'a tree: *Calophyllum* sp.'
2.9 Ceb. bitanʔág 'tree of wet areas'
2.11 Mar. bitanag 'tree used for lumber'
2.15 BM. bintanag 'large tree'
PGTL *bintanago 'a tree: *Calophyllum* sp.'

(12) *bítek 'intestinal worm'
2.7 Han. bítuk 'earthworm, intestinal parasitic worm, particularly *Ascaris*'
2.9 Ceb. bítuk 'roundworm in the digestive tract'
2.10 WBM. bituk 'intestinal worms'
2.15 BM. bitok 'intestinal worm'
NOTE: Also Bkl. *pítok* 'pinworm'. WBM. *bituk* is assumed to be a loan.

(13) *butíti 'pufferfish; incipient frog'
2.9 Bkl. butíti 'tadpole; a large-bellied fish: *Tetradon lunaris*; large-bellied'
Akl. botíti 'poisonous fish with bloated stomach'
Ceb. butíti 'general name for pufferfishes'
2.12 Mar. botiti 'fish with air-filled entrails'
2.15 PGTL *Butiti 'kind of small fish whose stomach inflates'
BM. busisiʔ 'kind of marine fish; also used for a constellation, the Stingray, and for a half-formed frog (with new head, and tail from its tadpole stage)'
NOTE: PPH had *butiti (reflected also in Isneg, Sangir, and Tontemboan), but its reconstructible meaning is 'swollen, of the belly'. The meaning 'pufferfish', and the perhaps unrelated meaning 'incipient frog' apparently are confined to GCP languages. I take Tontemboan *wutitiʔ* 'pufferfish' to be a loan.

(14) *dalíq 'haste, speed'
2.7 Han. dalíʔ 'haste, speed'
2.9 Tag. dalíʔ 'quickness, promptness'; dáli-dáliʔ 'hurriedly, hastily'
Bkl. dalíʔ 'hurriedly, quickly, rapidly'; mag-dalíʔ-dáliʔ 'be in a hurry'
Akl. dalíʔ 'hurry up, go quickly'
Hlg. dalíʔ 'immediately, hurriedly, quickly'
Ceb. dáliʔ 'easy, quick, immediate; do something quickly'
2.15 BM. mo-daliʔ-daliʔ 'quick, fast, hasty; do with speed'
NOTE: I take Pgs. *dalí* 'continue, proceed, go on; hurriedly' to be a loan from Tagalog.

(15) *darág 'yellow'
2.9 Bkl. darág 'yellowed; yellowish, as old clothes'
Klg. ma-lalag 'yellow'
2.10 Mar. rarag 'old leaf—one to be shed off'
2.12 Sub. dalag 'yellow'
2.15 BM. darag/dayag 'yellow'
NOTE: Sneddon (p.c.) reconstructs Proto-Gorontalo-Mongondow *darag 'yellow'.

(16) *delém 'dark; night'
2.7 Han. dulúm 'darkness'; ma-dlúm 'dark'
2.8 Btk. dələm 'night'

2.9 Tag. dilím 'darkness'; muláŋ dilím 'first night of the decrease of the moon'
Bkl. dulóm 'the dark of the moon, the period with no moonlight'
Akl. dueóm 'get dark(er)'
2.10 MbS. delem 'night'
2.11 Mar. delem 'moonless; night which is dark; black'
2.15 BM. doḷom 'evening, night'; mo-doḷom 'dark'

NOTE: With root *-lem 'dark'. Reid (p.c.) adds Northern and Southern Alta *dalam* 'night', noting that the vowel correspondences are not regular. These forms may be loans, but a plausible source language remains to be found.

(17) *dugúq 'blood'
2.7 Han dugúʔ 'blood'
2.8 Tbw.(A.) duguʔ 'blood'
2.9 Tag. dugóʔ 'blood'
Bkl. dugóʔ 'blood'
Ceb. dúguʔ 'blood'
Mmn. dogoʔ 'blood'
Msk. duguʔ 'blood'
2.11 Mar. rogoʔ 'blood'
2.12 Sub. duguʔ 'blood'
2.15 BM. duguʔ 'blood'

(18) *dúmul 'touch with the face'
2.9 Ceb. dúmul 'put something near the face or the face near something'
2.11 Mar. domol 'touch with lips, as horse smelling something'
2.15 Gtl. dumulo 'touch with the face while bowing the head'

(19) *duŋaq 'bow the head, nod'
2.11 Mar. doŋaʔ 'bow the head'
2.15 PGTL *duŋaʔ 'nod the head in agreement'
BM. duŋaʔ 'bow the head'

NOTE: Also WBM. *dunkaʔ* 'bow the head'.

(20) *ebú 'cough'
2.9 Tag. ubó 'cough'
Bkl. ábo 'cough'
Akl. ubó(h) 'cough up'
Ceb. ubú 'cough'
2.12 Sub.(Sc.) mog-obu 'to cough'
2.15 PGTL *oBu 'cough'
BM. obuʔ (-ʔ unexpl.) 'a cough'; mog-obuʔ 'to cough'

NOTE: MbAd. *ubu* 'cough' probably is a Bisayan loan.

(21) *ebul 'opaque, cloudy'
2.10 WBM. evul 'of a diseased eye or of water, to become cloudy or whitish in color
2.15 Gtl. wobulo 'clouds of smoke; gray'

(22) *emqem 'suck on something'
2.9 Akl. úmʔum 'suck on, dissolve in the mouth (without swallowing, as is done to medicine or candy'

Ceb. umʔúm 'put something partly or wholly into the mouth or between the lips'
2.10 WBM. emʔem 'put something in the mouth'
2.15 Gtl. womomo 'suck, be sucked on (as candy)'

(23) *-enu 'what?'
2.7 Han. únu 'what, what for?'
2.9 Tao. unuh 'what?'
2.10 WBM. m-enu 'how?, what?'
2.15 BM. onuʔ (-ʔ unexpl.) 'what?'
NOTE: Tbw.(K.) *unu* 'what?' is regarded as a loan from a Central Philippine language. The interrogative element *-nu descends from Proto-Austronesian; it is the form with initial *e which appears to be innovative.

(24) *gebáŋ 'sway'
2.9 Tag. gibáŋ 'swaying with tendency to fall'
2.15 BM. gobaŋ 'hold arms around a toddler to prevent him/her from falling'

(25) *gebáq 'swidden'
2.9 Ceb. gúbaʔ 'cultivated virgin land'
2.10 WBM. gevaʔ 'finish felling all the trees when making a swidden'
2.11 Mar. gebaʔ 'clearing, slash and burn farming'
2.15 BM. gobaʔ 'swidden, dry rice field'
NOTE: This form clearly reflects PWMP *Rebaq 'collapse, crash down (as a house, or falling trees)', a meaning which still is reflected in Ceb., Mar., WBM., and other GCP languages. The semantic innovation to 'swidden', however, appears to be unique to this group.

(26) *giman 'snare trap for birds or small animals'
2.10 WBM. giman 'rattan bow trap for birds, wild chickens, rats, snakes, etc.'
2.11 Mar. giman 'trap for water birds and rats; fish net'
2.15 BM. giman 'snare trap for birds, jungle fowl, etc.'

(27) *ha(m)beŋ 'block or obstruct; dam'
2.10 WBM. haveŋ 'block an opening or a path so that nothing can pass; a constructed barrier'
2.15 BM. amboŋ 'dam, dyke, obstruction'
NOTE: With root *-beŋ 'block, stop, dam'.

(28) *helát 'wait'
2.8 Btk. ɨlat 'to wait'
2.9 Bkl. halát 'to wait, wait for'
Akl. hueát 'to wait (for)'
Ceb. hulát 'to wait, wait for'
2.12 Sub.(S.) miɡ-ɨlat 'to wait'
2.15 BM. oḷat 'wait, wait for'
NOTE: Also Tbw.(A.) *ɨlat* 'to wait'. Tbw.(K.) *ɨlat* 'to wait' is assumed to be a loan from a Central Philippine language, while Totoli *olat*, Boano *oat* 'to wait' are regarded as loans from some Gorontalo-Mongondow source.

(29) *haldek, *hendek 'fear, fright'
2.9 Tag. hindík 'continuous and agonizing hard breathing; last agonizing breath of the dying'
Mmn. haldɨk 'to fear'
Msk. allɨk 'to fear'
2.10 WBM. handek 'to fear, to frighten'
2.11 Mar. lek (< earlier **edlek) 'fear, afraid'
2.12 Sub.(Sc.) m-ondok 'to fear'
2.15 BM. ondok 'fear, anxiety, fright'
NOTE: Sneddon (p.c.) adds that "Northern Minahasan languages *indeʔ* 'fear' is related but the initial vowel and glottal stop are irregular." Tentatively I assume that the similar form in Minahasan languages is a product of borrowing.

(30) *hibet 'oath'
2.10 WBM. hivet 'make sure of something; a self-directed curse invoked as a certification of the veracity of one's word'
2.11 Mar. ibet ibet 'light oath, assurance'
2.15 BM. ibot 'oath'; moŋ-ibot 'swear an oath'

(31) *hin-anak-an 'family'
2.9 Ceb. hin-aŋ-kan (< anák 'son, daughter') 'hen that has raised a brood'
2.15 BM. in-anak-an 'family'

(32) *hitaq 'groin'
2.7 Han, hítaʔ 'crotch, groin'
2.9 Tag. hítaʔ 'thigh'
Bkl. hítaʔ 'groin'
Akl. hítaʔ 'crotch, groin'
Hlg. hítaʔ 'thigh, crotch'
2.15 BM. itaʔ 'groin'
NOTE: Kap. *itaʔ* 'thigh' is assumed to be a Tagalog loan.

(33) *húlas 'sweat'
2.7 Han. húlas 'perspiration, sweat'
2.9 Tag. húlas 'sweat'
Mmn. holas 'sweat'
Tao. hulas 'sweat'
2.12 Sub.(S.) gulas 'sweat'
2.15 BM. uḷat 'sweat'
NOTE: Sneddon (p.c.) reconstructs Proto-Gorontalo-Mongondow *ulas 'sweat'.

(34) *intaluq 'defecate'
2.9 Klg. min-ta:lun 'defecate'
Msk. intarun 'defecate'
Tao. mag-ʔintauʔ 'defecate'
2.15 BM. moŋ-intaḷuʔ 'defecate'
NOTE: Strictly speaking, the Mongondow form can be compared only with Tao. *mag-ʔintauʔ*, but all four forms must surely be the product of a single innovation.

(35) *írek 'armpit'
2.9 Ceb. íluk 'armpit'
2.10 WBM. irek 'armpit'

2.11 Mar. irek 'armpit; carry under arm'
2.15 PGTL *ihoko 'armpit'
BM. iyok 'armpit'

NOTE: Also Kelabit *ilek* 'armpit', a form which cannot be reconciled with *irek. Sneddon (p.c.) adds Lolak, Ponosakan *iyok*, and notes that Kaidipang *ihoko* 'armpit' unambiguously reflects earlier *r.

(36) *itis 'drip or pour out'
2.10 WBM. itis 'pour a liquid into a container; pour over something'
2.11 Mar. itis 'dehydrate, squeeze dry'
2.15 BM. isit 'drip, trickle out, as blood from a cut'

NOTE: Also Tir. *itis* 'pour out a liquid', with diagnostic -/s/ pointing to indirect inheritance (Blust to appear).

(37) *kalawag 'turmeric'
2.10 WBM. kelawag 'turmeric'
2.11 Mar. kalaoag 'Curcuma longa L.'
2.15 PGTL *kolawago 'turmeric'
BM. koḷawag 'Curcuma, turmeric'

NOTE: Tir. kelawag 'turmeric, Curcuma longa L.' is regarded as a loan from a Danaw language.

(38) *kílid 'side, edge'
2.9 Akl. kílid 'side'; ta-kilíd 'turn to one's side'
Hlg. kílid 'beside, on the brink'
Ceb. kílid 'side'
2.10 WBM. kilid 'the side or edge of an object or person'
2.11 Mar. kilid 'side, edge'
2.15 BM. kilid 'side, edge of a yard or garden'

(39) *kupíq 'fold'
2.9 Tag. kupíʔ 'folded, doubled'
2.15 PGTL *kupiʔ 'to fold'

NOTE: With root *-piq 'fold'.

(40) *kútad 'barren land'
2.9 Tag. kútad 'barren land'; kutád 'barren, sterile'
2.15 BM. kutad 'swell up, of a corpse; expansion of seed in the ground before it germinates and dies; to die'

(41) *lábis 'excess, excessive'
2.7 Han. lábis 'an excessive amount; more, excessive, surpassing'
2.9 Tag. lábis 'excessive, more than enough'
2.15 Gtl. labito 'leftover, remainder'

NOTE: Reid (p.c.) adds Bontok, Ilokano *lábes* 'exaggerate, be excessive', but the last vowel fails to correspond.

(42) *lámbuŋ 'blouse'
2.7 Han. lámbuŋ 'woman's waist or blouse; a seven-piece, close-seamed, slipover, homespun cotton garment with a flare bottom'
2.9 Bkl. lambóŋ 'ankle-length tunic or robe' (archaic)
Akl. eámboŋ 'skirt, blouse, dress'
2.15 PGTL *lambuŋo 'jacket, blouse'

NOTE: Sneddon (p.c.) adds Tontemboan *lambuŋ* 'ankle length blouse formerly worn by men and women', Tondano *lambuŋ* 'clothing, shirt', Tonsea *dambuŋ* 'jacket, blouse', Ratahan *lambuŋ* 'blouse'. He considers these probable loans.

(43) *lantuŋ 'fence, corral'
2.10 WBM. lantuŋ 'pole which is placed in the floor of a house, dividing it into two equal areas'
2.11 Mar. lantoŋ 'stop, obstruct; fence'
2.15 BM. ļantuŋ 'pen, cage, fenced enclosure for cattle'

NOTE: Sneddon (p.c.) cites forms from a number of other languages and reconstructs Proto-Gorontalo-Mongondow *lantuŋ.

(44) *láyug 'to fly'
2.8 Btk. layug 'to fly'
Tbw.(A.) layug 'to fly'
2.9 Bkl. láyog 'to fly, to take off (as an airplane); a witch that flies upright with its arms outstretched and its eyes gazing at the full moon'
Akl. eáyog-eáyog 'bounce up and down (like a bamboo bridge or narrow wooden walkway when people cross)'
Ceb. láyug 'to fly (usually said of fowls that keep to the ground)'
Msk. layug 'to fly'
2.12 Sub.(Sc.) l-um-ayug 'to fly'
2.15 BM. ļayug 'to fly, glide, float on the air'

NOTE: Sneddon (p.c.) reconstructs Proto-Gorontalo-Mongondow *layug 'to fly, glide'. Tbw.(K.) *layug* 'to fly' is assumed to be a loan from a Central Philippine language.

(45) *lilíd 'to roll'
2.7 Han. lilíd 'rolling over and over'
2.10 WBM. lilid 'of an object, to roll on the floor or ground'
2.11 Mar. lilid 'roll'
2.15 BM. lilid 'to roll (as over the ground), roll away; roll up; roll over'

(46) *ŋiŋíq 'cry, fuss'
2.9 Bkl. ŋiŋíʔ 'to cry (babies)'
2.15 Gtl. ŋiŋi 'peevish, unenthusiastic toward something because it is difficult, etc.'

(47) *páŋkaq 'to hammer, hit'
2.9 Ceb. páŋkaʔ 'bump against something'
2.15 BM. paŋkaʔ 'to hammer, forcefully strike with e.g. the sago beater, so that the sound can be heard from afar'

(48) *panqés 'acrid stench'
2.7 Han. panús 'stench of urine'
2.9 Tag. panís 'stale and spoiled (said of cooked left-over or neglected foods)'
Ceb. panʔús 'for food, wash, or the body to get a rancid smell from having been wet and not allowed to dry out well'
2.11 Mar. panos 'sour'
2.15 PGTL *paŋuso 'foul, pungent smell'

NOTE: Mar. *panos* shows /o/ for expected /e/, and may be a loan. PGTL *paŋuso shows /ŋ/ for expected /n/, but irregularities involving /n/ and /ŋ/ are not uncommon in the G-M languages.

(49) *pantáw 'lookout'
2.10 WBM. pantew 'to look out over'
2.15 BM. pantow 'climb a lookout tree (as in the middle of a garden) or height to observe the vicinity'
NOTE: Bontoc *pantew* 'the yard in front of a house', Bkl. *pantáw* 'a washing porch constructed of bamboo slabs, usually found at the back of the house attached to the cooking area', Ceb. *pantáw* 'porch which may or may not be roofed and/or walled—usually in the back of a house' may be connected. If so, the above forms in WBM. and BM. can be cited only as evidence of a semantic innovation.

(50) *pateŋ 'stare, look fixedly'
2.10 WBM. pateŋ 'go and investigate something you are curious or suspicious about'
2.11 Mar. pateŋ 'stare'
2.15 BM. patoŋ 'look at intently'
NOTE: With root *-teŋ 'stare, look fixedly'.

(51) *pelek 'short, small'
2.11 Mar. pelek 'bit; little, minute, small'
2.15 BM. poḷok 'short'

(52) *pelús 'slip off'
2.9 Bkl. palós 'get off, dismount'
2.11 Mar. pelos 'slip off'
2.15 BM. poḷut 'get loose (as flower petals), get loose (as an animal from a noose trap)'
NOTE: With root *-lus 'slip off'.

(53) *pigíq 'buttocks'
2.9 Tag. pigíʔ 'hips, rump'
Klg. pigiʔ 'buttocks'
2.10 MbAd. pigiʔ 'buttocks'
2.12 Sub.(Sc.) pigiʔ 'buttocks'
2.15 BM. pogiʔ 'buttocks'
NOTE. BM. *pogiʔ* reflects *pegiq; Tag. *pigíʔ* can reflect either form.

(54) *piŋí 'cluster, bunch'
2.9 Bkl. piŋí 'be side-by-side'
2.11 Mar. piŋi 'cluster, bunch'
2.15 PGTL *piŋi 'joined together, as two bananas that are fused'

(55) *pirík 'shake off, as water from the hand'
2.9 Bkl. pirík-pisík 'to spatter, splatter; to flip (as a fish out of water); to shake dry (as a wet dog)'
Ceb. pilík 'spatter liquid by shaking something or flipping'
2.10 WBM. pirik 'shake water off the hands (intr.); shake water on something with the hand'

2.11 Mar. pirik 'sow, scatter, throw'
2.15 BM. pirik 'throw away something in the hand'

(56) *pispis 'baby bird'
2.7 Tag. pispís 'the young of doves and other birds'
Akl. píspis 'bird (generic)'
Hlg. píspis 'bird (generic)'
2.10 WBM. pispis 'a baby rat or bird'
2.15 PGTL *pipiso 'baby bird'
BM. pipit 'baby chick'

(57) *púgad 'nest'
2.7 Han. púgad 'nest, as of a bird'
2.8 Btk. pugád 'bird's nest'
2.9 Tag. púgad 'nest'
Akl. pugád 'nest'
Hlg. pugád 'nest'
2.11 Mar. pogad 'take off the hen's nest, remove from nest'; pogad-aʔ 'hen's nest'
2.15 BM. pugad 'nest'

(58) *pugítaq 'octopus, squid'
2.9 Tag. pugítaʔ 'cuttlefish, octopus'
Bkl. pugíta (loss of *-q unexpl.) 'octopus'
2.15 BM. pugitaʔ 'polyp, kind of octopus'
NOTE: Palauan *bokitáŋ* 'octopus, squid', the only reflex known outside the GCP group, contains several irregularities which mark it unmistakably as a loan.

(59) *pugpúg 'to shed, as fur or feathers'
2.9 Bkl. pugpóg 'to shed (feathers, fur)'
2.10 WBM. pugpug 'shake or brush something off of something else'
2.11 Mar. popog 'shake off, as ash from cigarette'
2.15 BM. pupug 'ichthyosis, a scaly, dry and flaky skin rash'

(60) *púnuq 'leader, chief'
2.7 Han. púnuʔ 'leader, head, chief, esp. with reference to people other than the Hanunóo'
2.9 Tag. púnoʔ 'chief; officer, official; trunk of tree, beginning'
Bkl. púnoʔ 'chief, headman, leader' (arch.)
Akl. púnoʔ 'leader, head'
Hlg. púnuʔ 'tree trunk; leader, chief'
Ceb. púnuʔ 'officials in charge of an office'
2.10 WBM. punuʔ-an 'ruler, chief, head'
2.12 Sub. poon 'leader'; pono-an 'governor'
2.15 BM. punuʔ 'title of nobility; lord, prince'
NOTE: This item is a metathesis of PMP *puqun 'base of a tree, beginning, source, origin', a term which has important applications to the mother's brother in some of the societies of the Lesser Sunda Islands of Indonesia, where matrilateral cousin marriage was traditionally practiced. In both cases where the term is applied to a social role, the role in question is one of leadership, or one commanding respect.

(61) *remuq 'dirty'
2.12 Sub.(S.) mi-limuʔ 'dirty (clothes)'
2.15 BM. romuʔ, yomuʔ 'dirty, filthy'
Gtl. lomu-lomu 'dirty'
NOTE: Also Tao. *lummiʔ* 'dirty (clothes)'.

(62) *renek 'gentle, peaceful'
2.11 Mar. renek 'be at peace, keep silent or quiet'
2.15 BM. yonok 'gentle, tame, good-natured'

(63) *runtáy 'pound fiber'
2.9 Tag. lutáy 'broken or rent into shreds'
2.10 WBM. runtey 'to pound or rub abaca fiber to soften it'
2.11 Mar. rontai 'bludgeon, beat, whip, club'
2.15 BM. yuntai 'pound fibers from a sago trunk, pound until soft or fine'

(64) *sagay 'stinging nettle'
2.11 Mar. sagai 'tiny black ants whose bite causes pain like poison ivy; feel restless or uncomfortable; poison shrub: *Urophyllum* sp.'
2.15 BM. tagoi 'stinging nettle tree: *Laportea crenulata*'

(65) *ságinŋ 'banana'
2.7 Han. ságinŋ 'banana, plantain; a generic term including all edible varieties of the two species of *Musa*'
2.9 Tag. ságinŋ 'banana'
Mmn. saginŋ 'banana'
Klg. saginŋ 'banana'
2.10 WBM. saginŋ 'generic for the various species of banana'
2.11 Mar. saginŋ 'banana'
2.12 Sub. saginŋ 'banana'
2.15 BM. tagin 'banana plant and fruit'
PGTL *sagin 'banana'
NOTE: Also Kankanay *ságinŋ* 'variety of banana with yellow skin', Kap. *ságin*, Bilaan, Samal *saginŋ* 'banana', Tir. *saginŋ* 'a generic term for banana', Dampelas *saginŋ*, Totoli, Boano *sagin* 'banana'. I regard the forms of this word in Kapampangan, Bilaan, Samal, Tiruray, and the Tomini languages as GCP loans. Kankanay *ságinŋ* is more difficult to explain as a loanword, though it is unattested in any other Cordilleran language.

(66) *sápaq 'creek, brook, stream'
2.7 Han. sápaʔ 'river'
2.8 Btk. sapá (loss of *-q unexpl.) 'river, lake'
2.9 Tag. sápaʔ 'brook, rivulet'
Bkl. sápaʔ 'stream, rivulet'
Akl. sapáʔ 'pond, small lake'
Ceb. sápaʔ 'brook or creek'
Mmn. sapaʔ 'river'
2.12 Sub. sapa-sapa 'brook, rivulet'
2.15 BM. tapaʔ 'brook, creek'
NOTE: I regard Alangan, Iraya *sapaʔ* 'water', Samal *sapaʔ* 'river' as GCP loans.

(67) *sapuq 'meat'
2.10 WBM. sapuʔ 'flesh, meat'
2.11 Mar. sapoʔ 'meat, flesh, muscle'
2.15 BM. tapuʔ 'meat, flesh'
PGTL *sapuʔ 'meat, flesh'

(68) *sarak 'meaning; translate'
2.11 Mar. sarak 'translate, interpret, explain or give meaning, commentary, exegesis'
2.15 BM. tayak 'look for, seek; sometimes: meaning, what one seeks to reach'

(69) *segáq 'lamp, light'
2.9 Tag. sigáʔ 'blaze of burning garbage or weeds; bonfire'
Bkl. sagá (loss of *-q. unexpl.) 'brilliant, glaring, shining'
Akl. sugáʔ 'to burn, set on fire'
Hlg. sugáʔ 'light, electric bulb'
Ceb. súgaʔ 'lamp'
2.15 PGTL *soga 'lamp'
NOTE: Dempwolff (1938) compared Tag. *sigáʔ* with Ngaju Dayak *seha* 'burn' under an etymon *seRaq, but this reconstruction now appears to have been erroneous. Lauje, Tialo *toga* 'lamp', with irregular *s > /t/ are almost certain loanwords from one of the Gorontalic languages in which the change *s > /t/ is regular.

(70) *séŋkeb 'lie prone'
2.7 Han. súkub 'facing down, pronation'
2.15 PGTL *soŋkob 'fall prone, fall on one's face'
NOTE: With root *-keb 'face downward'.

(71) *sekepu 'lap, hold in the lap'
2.10 WBM. sekepu 'hold something in one's lap'
2.15 PGTL *sokopu 'lap, hold in the lap (as a small child)'

(72) *seki 'lower leg of a quadruped'
2.10 WBM. seki 'specific for the bottom of a pig's foot, but also used of the whole foot'
2.11 Mar. seki 'leg'
2.15 BM. toki 'part of an animal's foot just above the hoof or claws'
NOTE: Compare PPH *sikí 'foot and leg of an animal'. Tir. *sekey* 'leg' is regarded as a Danaw loan.

(73) *seŋa 'blow the nose'
2.8 Btk. súŋa 'blowing one's nose'
2.9 Tag. siŋá 'mucus expelled from the nose; act of expelling mucus from the nose'
Bkl. suŋá 'blow the nose'
2.10 WBM. seŋa 'blow the nose'
2.11 Mar. seŋa 'blow the nose'
2.15 PGTL *soŋa 'blow the nose'

(74) *sikil 'stick out'
2.9 Tag. síkil 'crowding out by pushing backward with arms or shoulders'
Ceb. sikíl 'for something long to jut out'
2.15 BM. sikil 'point or flap of a wrapped-around sarong which sticks out'

(75) *sílay 'see, look'
2.9 Tag. sílay 'brief appearance; get a brief glance of'
2.15 PGTL *sile 'see, look at'
NOTE: Sneddon (p.c.) has drawn my attention also to Proto-Sangiric *selay, Tontemboan *sere* 'to look, see'. Both sets of forms exhibit at least one irregularity, and may be borrowed.

(76) *súgba 'cook'
2.7 Han. súgba 'roasting'
2.9 Bkl. sugbá 'cook or roast something over an open fire or coals; to grill, barbecue'
Akl. súgba(h) 'put over embers'
Hlg. súgba 'roast, broil directly on top of live coal'
Ceb. sugbá 'broil over hot coals'
2.15 BM. tubaʔ (-ʔ unexpl.) 'roast over or in the fire'

(77) *sukud 'measure; measurement'
2.9 Ceb. sukúd 'take a measurement'
2.10 WBM. sukud 'a measure; to measure something'
2.11 Mar. sokod 'measure'
2.15 Gtl. tuʔudu 'measurement (of length)'

(78) *sulúg 'cock, rooster'
2.7 Han. sulúg 'young rooster with spurs that are still small'
2.9 Bkl. súlog 'young rooster'
Akl. sueóg 'rooster, cock (young)'
2.15 PGTL *sulugo 'rooster'

(79) *suti 'clean, pure'
2.11 Mar. soti 'clean, pure, sterile; purge, purify'
2.15 BM. susi 'clean, pure'; bulawan susi 'pure gold'
BM. tusi 'pure, complete'; bulawan tusi 'pure gold'

(80) *tádu 'beeswax'
2.7 Han. táru 'beeswax produced by the bee known as *putyúkan*'
2.8 Btk. tádo 'wax'
2.9 Bkl. táro 'beeswax'
Ceb. tálu 'tallow, the hard fat in animals or the wax from beehives, used for making candles, soap, etc.'
2.10 WBM. tazu 'honeycomb; beeswax; the cosmetic preparation of beeswax and coconut oil which is put on the lips to make them shine'
2.11 Mar. taro 'wax'
2.15 BM. tayu 'beeswax'

NOTE: Ceb. *tálu* may show contamination with English *tallow*. Banggi (*tāru* 'wax') is regarded as a GCP loan. Sneddon (p.c.) notes that related forms occur in a

number of other languages outside the GCP group "e.g. Napu (Kaili-Pamona group, Central Sulawesi) *taru*, also Tondano, Tonsea, Tombulu *taru*, although Tontemboan *taʔndu* is odd and Tonsawang *tayu* is definitely from Mongondow." I appreciate the force of this observation, but am divided in my feelings regarding how best to explain the known distribution. PMP *lilin 'beeswax' was retained in many daughter languages, including Proto-Philippines. Unless there was some still undetermined difference of meaning between the two reconstructed forms, *tádu would appear to be a replacement innovation. The known distribution of its reflexes includes languages in the GCP group, and languages that border them or that are close enough to justify an inference that prehistoric trade could have taken place between the two. It is admittedly difficult to see why beeswax, which is readily available in most environments, would become an article of trade. Nonetheless, the facts available to me do suggest that reflexes of *tádu outside the GCP group are a product of borrowing.

(81) *takták 'shake out one by one'
2.9 Tag. takták 'act of shaking out the contents of a container by pounding its open mouth against some hard object'
Bkl. takták 'tap, rap, dislodge something by tapping or rapping'
Ceb. takták 'for small things or something fastened or stuck to something to detach and drop; cause to drop'
2.10 WBM. taktak 'of an object or container with a hole in it, to slam it down in order to dislodge something inside'
2.11 Mar. tatak 'drop, fall, let fall like fruits'
2.15 PGTL *totako 'dribble away, be lost in driblets'

(82) *taŋál 'mangrove tree'
2.7 Han. taŋál 'type of tree: *Ceriops* spp.'
2.15 PGTL *taŋalo 'mangrove tree'

(83) *taqil 'conserve, use sparingly'
2.10 WBM. taʔil 'prepare and set aside something for a special purpose'
2.15 Gtl. tailo 'conserve, use s.t. of value sparingly so that it will last longer'

(84) *tebél 'constipation'
2.9 Tag. tibí 'constipation, costiveness'
Bkl. tuból 'constipated'
Akl. tubóe 'hard, dry stool (feces)'
Ceb. tubúl 'hard stool; have constipation'
2.15 PGTL *tobolo 'constipation'
NOTE: Bontoc *tóbel* 'be constipated' appears to reflect a doublet.

(85) *tebtéb 'cut down'
2.9 Bkl. tubtób 'crop or cut the hair'
Ceb. tubtúb 'cut something at its base, very close to the surface'
2.15 PGTL *totobo 'chop down'

(86) *tibas 'cut or hack off'
2.7 Han. tíbas 'cutting away'
2.9 Msk. tibas 'strike with a bolo'
2.10 WBM. tibas (expected **tivas) 'strike with an axe or bolo'
2.15 BM. sibat 'cut off, hack off (as bamboo)'

(87) *tibtíb 'cut off small pieces'
2.9 Tag. tibtíb 'sugarcane tips used for planting'
Ceb. tibtíb 'chip off, cut off small pieces with repeated strokes'
2.15 BM. sisib 'cut into small pieces, cut into slices or layers'

(88) *tigtíg 'jerky movement'
2.9 Tag. tigtíg 'jerkiness, shakiness (on vehicle or the like)'
2.15 BM. sisig 'winnowing of pounded rice or ground corn'
NOTE: Also Ilokano *tiktík* 'to winnow'.

(89) *tiqél 'leg, foot'
2.9 Ceb. tiʔíl (< Assim.) 'foot, leg'
2.15 BM. siʔoḷ 'foot, leg, paw'

(90) *túbig 'water'
2.9 Tag. túbig 'water'
Klg. tubig 'water'
Msk. tubig 'water'
Tao. tubig 'water'
2.11 Mar. tobig 'cool with water'
Mar. toig 'container to take water from boat; bale water'
2.12 Sub. tubig 'water'
2.15 BM. tubig 'water, liquid, juice'

(91) *tunaq 'mud puddle, buffalo wallow'
2.9 Tag. tunáʔ 'submerged, sunk'
Ceb. túnaʔ 'wallow'; tunaʔ-an 'puddle for water buffalos to wallow in'
2.10 WBM. tunaʔ 'of a carabao or pig, to wallow in the mud'
2.11 Mar. tonaʔ 'wallow in mud, watery mud'
2.15 BM. tunaʔ 'mud puddle where pigs and water buffalos wallow'

(92) *tuntún 'lower with a rope'
2.7 Han. tuntún 'dangling, hanging through (and down)'
2.9 Bkl. tuntón 'to suspend something; to lower something by suspending it from something else (as when lowering something by rope)'
Akl. tónton 'to lower, let down; plumb bob'
Hlg. túntun 'to lower down (as a rope), to dangle'
Ceb. tuntún 'lower something; sag loosely; string or rope used to lower something'
2.11 Mar. tonton 'to lower, as with rope'
2.15 BM. tuntun 'lower a rope, let fall; also: throw or fling down'
NOTE: Proto-Minahasan *tonton 'lower with a rope' is tentatively assumed to be a loan, though Sneddon (p.c.) believes this to be unlikely.

(93) *úpas 'sheath of the banana stalk'
2.7 Han. úpas 'sheathing of banana, plantain or similar plants'
2.9 Bkl. úpas 'abaca plant'
Akl. úpas 'trunk of the banana plant'
Ceb. úpas 'banana leaf stalk'
2.10 WBM. upas 'the heart of a banana stalk, which is eaten as a vegetable'
2.11 Mar. opas 'stalk of banana'

2.15 PGTL *wupaso 'sheath of the banana stalk'
BM. upat 'damp banana fiber'

NOTE: Sneddon (p.c.) reconstructs Proto-Gorontalo-Mongondow *upas 'sheath of the banana stalk'. Bontoc *ópas* 'the wrapping of bundled tobacco; to wrap a bundle of tobacco' may be connected.

(94) *uyun 'agree(able) with'
2.7 Han. úyun 'fitting, suitable'
2.9 Bkl. úyon 'in accordance with; in harmony with'; mag-úyon 'agree with one another'
Akl. úyon 'to like, adjust to, react favorably to; conform to; approve of; be parallel, be agreeable (to)'
Ceb. úyun 'something long, parallel to, alongside; stay, place oneself parallel to something'
2.10 WBM. uyun 'to agree; an agreement'
2.15 PGTL *wupaso 'sheath of the banana stalk'
BM. upat 'damp banana fiber'

NOTE TO APPENDIX. Most GCP languages show relatively few changes from Proto–Greater Central Philippines. The most notable exceptions to this statement are the Gorontalo-Mongondow (G-M) languages, particularly Gorontalo (Gtl.) itself. The major sound changes relating these languages to Proto–Greater Central Philippines can be summarized as follows: (1) Gtl. and Suwawa added /w/ before words that began with *a, *u, or *e (which became /o/), and usually /y/ before initial *i; (2) *e became /o/ in all G-M languages; (3) prepenultimate *a became /o/ (presumably through earlier schwa) in all G-M languages; if initial this /o/ was retained in Bolaang Mongondow (BM.), but lost in Gtl.; (4) the diphthongs *-ay and *-aw were monophthongized to -/e/ and -/o/ in Gtl. but became -/oy/ and -/ow/ in BM.; (5) *h was lost in all G-M languages except Gtl., where it is sometimes preserved; (6) *-q disappeared in all of the Gorontalic languages, but not in BM.; as shown by Sneddon and Usup (1986:414, 419) the loss of this phoneme occurred after the breakup of Proto-Gorontalic; (7) apart from homorganically prenasalized stops medial consonant clusters were reduced in all G-M languages; (8) a supporting vowel /o/ was added after final consonants in the Gorontalic languages, but not in BM.; (9) *s became /t/ in BM., Gtl., Suwawa, and Buol, but not in the other Gorontalic languages; (10) *n became /l/ in Gtl. before the change *nd > /n/ and loss of *-no; (11) *d became /r/ or /y/ in BM.; Usup (1986) writes Proto-Gorontalic *h as the reflex of PPH *d or *r in many forms, but the supporting evidence sometimes suggests that the PGTL reconstruction should be *r; (12) final *a became /o/ in Gtl. and Buol, but not in the other Gorontalic languages; (13) *k became /ʔ/ in all of the Gorontalic languages except Kaidipang and Buol; (14) *g became /h/ in Gtl., (15) *b became /h/ before /u/ in Gtl.; (16) *r became /l/ in Gtl.; (17) *mb and *nd were reduced to the corresponding simple nasal in Gtl., but *ŋg remained; in Buol and Kaidipang *mb, *nd and *ŋg all reduced to the corresponding simple nasal; (18) *o following *b, *mb, *d, *nd and *g became /u/ in Gtl., (19) in Buol and Kaidipang *a usually became /o/ after a voiced stop: in Gtl. *a usually became /o/ after *b; (20) *a following *mb, *d, *nd, *g, and *ŋg became /e/ in Gtl.; (21) prenasalized voiceless stops became prenasalized voiced stops in Buol, Kaidipang, Gtl., and Suwawa. For further details of the historical phonology of the Gorontalic languages, the reader is referred to Sneddon and Usup (1986).

Two problem areas in the comparative phonology of G-M languages come to notice especially often in the data of this Appendix. First, BM. sometimes has a final glottal stop when none is expected (items 20, 23, 76). Second, *n is sometimes reflected as ŋ, and *ŋ is sometimes reflected as /n/ in various of the G-M languages (items 4, 65, and such additional etymologies as PPH *dahun > Gtl., Kaidipang *duŋo* 'leaf'). Regarding the first type of irregularity Sneddon (p.c.) notes that final glottal stop often appears in Lolak where it is not expected. He adds "I wouldn't be surprised if it is spreading. Kaidipang, like Sundanese, has all final vowels closed by glottal stop." Sneddon regards the second type of irregularity in the Gorontalo-Mongondow languages as quite marginal, but it occurs sufficiently often in the data I cite to justify at least passing notice.