

## CHAPTER 7

# Historical linguistics of the Philippines

R. DAVID ZORC, JASON W. LOBEL, AND WILLIAM HALL

## 7.1 Introduction

The Philippines, a nation of over 7,000 islands, is located just south of Taiwan, northeast of Borneo, and due north of the Indonesian island of Sulawesi. Its 179 living, indigenous languages mostly belong to the Philippine subfamily (cf. Blust 2019b),<sup>1</sup> and 171 of the 190 members of the Philippine subfamily are spoken within its borders.

Philippine languages were among the first languages in Asia to be the subject of western linguistic study. The first documentation of a Philippine language took place in 1522 when Pigafetta (1525: 51v–53r) elicited an inconsistently transcribed list of Old Cebuano (some of which, however, is actually Malay). Less than a century later, the documentation and description of several major Philippine languages began with the work of a number of Spanish friar-linguists in the first decades of the Spanish occupation of the Philippines. These works provide not only invaluable data on archaic features and vocabulary items, which can aid in lexical and grammatical reconstruction, but also important clues about the migration history of certain languages. In some cases (e.g. Old Bikol and Old Ilonggo), this four-century-old data reveals ancient languages which are different enough to be considered separate from their modern counterparts.

The contact languages that have contributed the most to Philippine lexicons over the past millennium are Malay (primarily pre-1600), Spanish (from the 1500s to 1800s), and English (starting with the arrival of the Thomasites in 1901, but especially in the modern era with widespread access to English-language television, music, and social media). Borrowings from Arabic, largely borrowed via Malay, are

<sup>1</sup> The seven Sama-Bajaw languages (see Kaufman, this volume, chapter 26) spoken in the Philippines are not included in this number as they belong to the Greater Barito subgroup (Blust 2007c). Note that our count of Philippine languages differs from the Ethnologue (Eberhard et al. 2019) in (i) the inclusion of seven languages (Bulalakawnon, Northern Samarenyo, Samā IGaCOS, Tawlet Subanen/Kalibugan, Tigwa Manobo, Klata, and Tasaday) not listed therein, and (ii) the exclusion of several others which are either unattested, not genetically Philippine, or simply dialects of another listed language.

most common in the languages of Islamized populations in the southern Philippines and their neighbours, as well as in northern Sulawesi. Finally, several Chinese languages (Hokkien, Minan, Southern Min) have left their mark on Philippine lexicons in the areas of commerce and cuisine.

This survey will provide an overview of the Philippine microgroups (§7.2), some phonological developments relevant to Proto-Philippines and its various daughter languages (§7.3), and subgrouping issues and controversies (§7.4). It should be noted that the subgroups, languages, and axes mentioned through this chapter have been established based on not only phonological innovations but also significant replacement innovations for well-established PAN or PMP reconstructions: for example, the 8,139 cognate sets in Blust and Trussel's (2020) *Austronesian Comparative Dictionary* include over 720 high-level etymologies that each contain two or more pages of invaluable comparative evidence.

## 7.2 The Philippine languages

Following Blust (1991b, 2019b); Charles (1974); Reid (1989); Robinson and Lobel (2013); Lobel (2010, 2013a); and Zorc (1977, 2019), the languages of the Philippine subfamily can be assigned to the following thirteen primary branches, listed roughly from north to south:

1. Batanic/Bashiic
2. Northern Luzon ('Cordilleran')
3. Central Luzon
4. Umiray Dumaget
5. Manide-Alabat
6. North Mangyan
7. Greater Central Philippines (and Palawanic)
8. Kalamianic
9. Inati
10. Southwestern Mindanao
11. Southeastern Mindanao
12. Sangiric
13. Minahasan

ZORC, LOBEL, &amp; HALL

### 7.2.1 Batanic/Bashiic

The northernmost subgroup in the Philippines is Batanic (or ‘Bashiic’), consisting of Itbayaten, Ivatan, and Ibatan/Babuyan spoken on the Batanes Islands off the northern tip of Luzon, plus the Yami language of Orchid Island within the national borders of Taiwan. The internal and external relationships of these languages have been studied by Scheerer and Conant (1908); Dempwolff (1926); Tsuchida, Yamada, and Moriguchi (1987); Yang (2002); Ross (2005); and Blust (2017d, 2019b). Included among the numerous lexical innovations defining this subgroup are \*batah ‘say, tell’, \*bubun ‘bury’, \*bulək ‘belly’, \*dalmət ‘heavy’, \*hilak ‘white’, \*kadam ‘rat’, \*muhdan ‘nose’, \*makpahad ‘bitter’, and \*taur ‘heart’.

### 7.2.2 Northern Luzon

Formerly known as ‘Cordilleran’, the Northern Luzon subgroup consists of seven branches: Ilokano, Cagayan Valley, Northeastern Luzon, Central Cordilleran, Southern Cordilleran, Alta, and Arta, each of which is dealt with in the following subsections. Lawrence Reid (1979, 1989, 1991, 2006b) and Ronald Himes (1997, 1998, 2005) are the primary scholars who have addressed either the entire subgroup or various branches thereof.

#### 7.2.2.1 Ilokano

Spoken by nearly ten per cent of the Philippine population, Ilokano is now the major trade language in most of northern Luzon, a position once held by Ibanag, which may explain the widespread /g/ reflexes of \*R in Ilokano where /r/ would be expected in native Ilokano forms (cf. Tharp 1974). Its two-way history of borrowing makes it difficult to identify Ilokano lexical innovations, although some candidates include *dúrek* ‘earwax’, *isú* ‘he/she (3SG.NOM)’, *kaladkad* ‘climb’, *lapáyag* ‘ear’, and *punpun* ‘bury’. There are two main dialects of Ilokano: a northern dialect reflecting earlier \*ə as /e/ and having the form *saán* /saʔán/ for ‘no’, and a southern dialect in which \*ə is preserved as /ə/ and the word for ‘no’ is *haʔán* /haʔán/.

#### 7.2.2.2 Cagayan Valley

The Cagayan Valley subgroup includes Adasen, the three Atta languages (Atta Faire, Atta Pamplona, and Atta Pudtol), Central Cagayan Agta, Gaddang, Ga’dang, Ibanag, Itawit/Itawis, Isnag/Isnag, Malaweg, and Yogad. It was formerly proposed that the Cagayan Valley and Northeastern Luzon subgroups formed a ‘Northern Cordilleran’ node

within Northern Luzon (or ‘Cordilleran’), but Robinson and Lobel (2013) found no evidence to support such a grouping after a more thorough survey of the Northeastern Luzon languages. Cagayan Valley innovations include \*agól ‘liver’, \*agída ‘they (3PL.NOM)’, \*aŋ(ə)tiŋ ‘afraid’, \*dákəs ‘bad’, \*dapín ‘dirty’, \*əbíŋ ‘child’, \*po:ray ‘angry, brave’, \*saŋáw (temporal: now, recent past, a little later), \*təkaw ‘borrow’, \*to:lay ‘person’, \*ubo:bug ‘speak’, and \*ügaŋ ‘sweat’.

#### 7.2.2.3 Northeastern Luzon

This subgroup includes five Aboriginal (or ‘Negrito’<sup>2</sup>) Filipino languages—Dupanangan Agta, Pahanan Agta, Dinapigue Agta, Casiguran Agta,<sup>3</sup> and Nagtipunan Agta, all spoken on or near the Pacific coast of northeastern Luzon by groups self-identifying as ‘Agta’—plus the Paranan language of the ethnic-Austronesian inhabitants of Palanan town. Robinson and Lobel’s (2013) is the only historical-comparative study of these languages, although Headland (1975) published an earlier study based on lexicostatistics and mutual intelligibility. Innovations within this group include \*ləbbút ‘boil (water)’, \*ladúʔ ‘fever’, \*putát ‘full’, \*madəgáʔ ‘heavy’, \*dəmət ‘arrive’, \*sánig ‘hear, listen’, \*tóglað ‘push’, and \*bakál ‘stab’.

#### 7.2.2.4 Central Cordilleran

This large group of languages includes Balangao, Isinay, Luba, Manabo, Northern Kankanaey, Southern Kankanaey, and various languages known by the names ‘Itneg’ (Binongan, Inlaod, Maeng, Masadiit, Moyadan, and Banao), ‘Kalinga’ (Butbut, Limos, Lubuagan, Mabaka Valley, Majukayang, Southern, and Tanudan); ‘Bontok’ (Northern, Southern, Central, Eastern, and Southwestern) and ‘Ifugao’ (Amganad, Batad, Mayoyao, and Tuwali).<sup>4</sup> Innovations defining the Central Cordilleran subgroup include \*ʔákaw ‘steal’ (with unexplained loss of the \*t- of earlier \*takaw), \*ʔalməŋ ‘laugh’, \*baʔúd ‘tie, tether’, \*bagáŋ ‘neck’ (semantic shift from PAN \*baRqaŋ ‘molar’), \*tágu ‘person, human’ (with insertion of -g- into PAN \*Cau), \*tubu ‘leaf’ (semantic specialization of PMP \*tubuq ‘grow, sprout’), \*tuŋʔal ‘bone’

<sup>2</sup> The second author, as well as Mirante (2014) and Louward Zubiri (pers. comm., 5/27/2020) have observed that many Aboriginal Filipinos disapprove of the term ‘Negrito’, which will therefore be avoided in this chapter.

<sup>3</sup> Usually referred to as ‘Casiguran Dumagat’ in the literature, following its primary scholar Thomas Headland, this and other groups speaking NE Luzon languages actually refer to themselves as ‘Agta’, while *dumagat* is a generic Tagalog term referring to any of the Aboriginal Filipino groups near the Pacific coast of northern and central Luzon, including the Umiray Dumagat, the Alta, and the Agta, and other Aboriginal Filipinos on Alabat Island.

<sup>4</sup> Bayninan Ifugao and Kiangnan Ifugao, both included in Reid (1971) are actually dialects of Kallahan/Kalanguya and Tuwali Ifugao, respectively.

(a metathesis of PWMP \*tuqəlan), \*tupák ‘mouth’, \*warɣay ‘river’, and \*wasít ‘throw’.

### 7.2.2.5 Southern Cordilleran

The Southern Cordilleran subgroup, for which Himes (1998) is the primary historical-comparative work, includes Ilongot/Bugkalot, Pangasinan, I-wak, Ibaloi, Kalanguya, Keley-i Kallahan, and Karao. Zorc (1979) has studied the historical development of contrastive accent in Pangasinan. Innovations defining this group include \*ʔəgás ‘intestines’, \*ʔaləgáy ‘stand’, \*baklaŋ ‘body’, \*balləg ‘big’, \*dálín ‘earth, soil’, \*ʔəsəl ‘speak’, \*sakáy ‘one’, \*təwən ‘sky’, and \*səlí ‘foot’.

### 7.2.2.6 Alta

This subgroup consists of Northern Alta and Southern Alta (Reid 1991; Garcia-Laguia 2018; Abreu 2018), spoken by Aboriginal Filipino populations living both along the eastern coast of north-central Luzon between the towns of Baler and Dingalan in Aurora Province, and upriver from those areas. Proposed innovations include \*bitlay ‘carry on shoulder’, \*dakəl ‘flood’, \*ibut ‘lost’, \*iʔə ‘this’, \*lanis ‘sweet’, \*lutit ‘mud’, \*mudəŋ ‘mountain’, and \*pənaŋ ‘hot (of weather)’.

### 7.2.2.7 Arta

This near-extinct language spoken by a small Aboriginal Filipino population living near Maddela, Quirino Province, was discovered by Reid (1989), who identified almost 150 unique forms (e.g. *binguət* /biŋuət/ ‘night’, *bukágan* /buka:gan/ ‘woman’, and *bunbun* ‘house’) and noted that its rate of retention of PMP lexicon was among the lowest figures (26.9%) known for any Philippine language, with Arta preserving reflexes of only fifty-one out of 189 PMP reconstructions. Arta currently has around a dozen first- and second-language speakers, and Kimoto (2017a, b) is the primary scholar on this language.

## 7.2.3 Central Luzon (and Remontado/Hatang-Kayi)

The Central Luzon subgroup consists of two branches, one of which is Kapampangan, and the other the Sambalic or Sambali-Ayta group, which includes Botolan Sambal, ‘Tina’ Sambal,<sup>5</sup> Bolinao, Ayta Mag-antsi, Ayta Mag-indi, Ayta

<sup>5</sup> ‘Tina’ is placed in quotation marks because at least some speakers consider it offensive, and the Ethnologue no longer includes it as the primary name of the language. However, the removal of this identifier leaves only ‘Sambal’, which becomes ambiguous as to which of the two Sambali languages is being referred to.

Abellen, Ayta Ambala, and Ayta Bataan/Magbukun. The primary study of this subgroup is Himes (2012). Kapampangan and Sambalic share a number of innovations, including \*ʔəmíʔ ‘urine’, \*gurut ‘back’, \*uɣut ‘coconut (generic)’ (< PPH \*huɣut ‘coconut shell cup’), \*taklaʔ ‘excrement, to defecate’ (also found in Remontado/Hatang-Kayi and Iraya), and \*tələk ‘deaf’ (borrowed into Pangasinan). Innovations unique to the Sambalic group include \*anag ‘termite’ (cf. PAN \*aNay), \*bəkraw ‘throat’, \*dalúnut ‘smooth’, \*dayi ‘still, yet’, \*dəbləm ‘dark’, \*duday ‘urine’, \*kudpal ‘thick’, \*láləʔ ‘deep’, \*maʔín ‘have, there is’, and \*rayʔəp ‘cold’.

Special mention should be made of the moribund Remontado/Hatang-Kayi language (often referred to in the literature by the unfortunate misnomer ‘Sinauna’ or ‘Sinauna Tagalog’ meaning ‘Archaic/Aboriginal Tagalog’) spoken by some 300 adult Remontados in the highlands of Tanay and General Nakar towns east of Manila (Lobel and Surbano 2019). Following Santos (1975), most scholars, including Blust (1991b) and Himes (2012) have included Hatang-Kayi in the Central Luzon group, but Lobel and Surbano (2019) question this based on the limited size of earlier data sets, much of which was ambiguous as to whether it was borrowed or inherited.

## 7.2.4 Umiray Dumaget

The Umiray Dumaget language—which has at least a northern/coastal and southern/inland dialect,<sup>6</sup> and possibly a third dialect on Polillo Island—is spoken over a considerable part of central-eastern Luzon, now primarily in inland and highland parts of the elaborate river networks of the area, but as late as the 1970s and 1980s, also along a stretch of the coastline of northeastern Luzon from near Baler in the north to near General Nakar in the south (MacLeod 1972). Other than a New Testament translation and various literacy materials developed by Tom and Pat Macleod of SIL-Philippines during their work among the Dumaget from the 1950s to the 1980s, little has appeared on this highly unique language whose numerous innovations include *sagú* ‘blood’, *órat* ‘water’, and *tapúk* ‘rain’ (Lobel, Andrada, et al. n.d.). Himes (2002) suggests assigning Umiray Dumaget to the Greater Central Philippine subgroup, but Lobel (2013a) rejects this, arguing that Himes’s analysis suffered from a lack of distinction between retentions, innovations, and borrowings. Instead, based on a much larger body of evidence

<sup>6</sup> In the southern dialect area, some Dumaget self-identify as *Bulus*, a term neither used nor recognized as an ethnic identifier by speakers from other areas, who only self-identify as *Dumaget* /dumagét/ (cognate with Tagalog *dumágat* /dumágat/). Finally, despite its misapplication in the literature, no Dumaget self-identify as ‘Agta’ (which simply means ‘person, human being’ in their language), nor are they referred to as such by Tagalogs.

## ZORC, LOBEL, &amp; HALL

including complete sets of functors, Lobel places Umiray Dumaget as a primary branch of the Philippine subfamily, not closely related to any of the other Philippine languages.

## 7.2.5 Manide-Alabat

The small Manide-Alabat subgroup consists of two languages spoken by Aboriginal Filipino populations in south-central Luzon: Inagta Alabat and Manide. Both languages were virtually unknown prior to work by Lobel over the past two decades, with Lobel (2010, 2013a) and Lobel et al. (2020) representing the only published linguistic analysis of these languages. These two closely related languages do not have any other close relatives, although related languages may once have been spoken by other aboriginal groups in Quezon Province who now speak only Tagalog as their native language, such as the ‘Katabagan’ of Catanauan and the so-called ‘Ayta’ of Tayabas town. Noteworthy features of these two languages include not only their large amounts of unique vocabulary, but also their vowel shifts which affect /a/ and /u/ after voiced stops and glides: Low Vowel Fronting (a shift shared with other Aboriginal Filipino languages along the Pacific coast of Luzon, cf. Lobel n.d.-a) and Low Vowel Backing in both languages, and Back Vowel Fronting in Manide. Proto-Manide-Alabat innovations include \*seŋul ‘sit’, \*panagbey ‘swim’, \*pála? ‘die, kill’, \*katlub ‘tongue’, \*seweŋ ‘ear’, \*gemes ‘rain’, \*peleŋjut ‘mosquito’, \*ma-lem?at ‘white’, \*su?eŋ ‘thorn’, \*hi?new ‘wind (n.)’, \*be?dis ‘faeces’, and over 200 others (Lobel n.d.-b).

## 7.2.6 North Mangyan

The North Mangyan subgroup consists of the Alangan, Iraya, and Tadyawan languages spoken in the northern half of Mindoro Island. Major works include Zorc (1974b) and Barbian (1977), both of which also cover the South Mangyan languages. Forms that appear to be unique to this group include \*apu ‘there is, exists, have’, \*dulaŋ ‘knee’, \*nakay ‘what?’, \*Rataŋ ‘hold’, and the unexplained addition of -y in \*duway ‘two’ from PAN \*duSa.

## 7.2.7 Greater Central Philippines

The Greater Central Philippine (GCPH) subgroup, first proposed by Blust (1991b), combines seven Philippine subgroups (Central Philippines, Manobo, Subanen, Danaw, South Mangyan, Mongondow–Gorontalo, and Palawan) on the basis of both shared lexical innovations and the \*R > /g/

shift. Of the ninety-four innovations that Blust presented to establish this group, it is clear from Blust and Trussel (2020) that twenty-eight are retentions from PAN, PMP, PWMP, or PPH. The sixty-six that remain include \*ʔəbúh ‘cough’, \*darág ‘yellow’, \*haldək ‘fear’, \*pispis ‘bird’, and \*púnu? ‘leader, chief’ (a semantic shift and metathesis of PMP \*puqun ‘base of a tree; cause; source, origin’). Zorc, however, questions the inclusion of the Palawan languages in this subgroup (cf. §7.2.7.7), given the small number of GCPH innovations reflected by them vis-à-vis other GCPH languages. Likewise, Lobel (2013a, 2016b) has raised the possibility that Molbog, the southernmost language included in the Palawan subgroup, may instead be more closely related to the Bonggi language of Sabah, Malaysia (cf. §7.4.4), a language which Blust (2010b) argues is most closely related to the Idaanic languages of Sabah.

## 7.2.7.1 Central Philippines

The approximately fifty members of this group, which consists of Tagalog, Mamanwa, and the various Bikol, Bisayan, and Mansakan languages, are spoken natively from southern Luzon to Sulu and southeastern Mindanao, and are spoken as a first language by over 60% of the Philippine population, primarily Tagalog (28%), the national language; Cebuano (13.1%); Ilonggo/Hiligaynon (7.5%), Bikol (6%),<sup>7</sup> and Waray (3.4%).

## 7.2.7.1.1 Tagalog

Chosen as the National Language or ‘Wikang Pambansa’ in 1937, Tagalog is the most widely studied Philippine language. Besides having been the donor of countless borrowings in dozens of proximate languages, Tagalog in pre-modern times was also the recipient of numerous loanwords, primarily from Malay, including forms ultimately from Arabic and Sanskrit (Wolff 1976); Kapampangan (whose territory was at certain times the centre of trade and power in the Philippines); and Chinese (Chan-Yap 1980). Due to its reciprocal borrowing relationships, as well as to its solitude (along with Kasiguranin) in its branch of the Central Philippine subgroup, it is exceedingly difficult to identify lexicon unique to Tagalog, but candidates include *búti* ‘good, well’, *dumí* ‘dirt, dirtiness’, *saán* /sa?án/ ‘where?’, *sagót* ‘answer’, *tagál* ‘long (of time)’, *tandâ* /tandá?/ ‘old (person)’, *tanóng* /tanún/ ‘ask’, *upô* /ʔupú?/ ‘sit’, and *úsok* ‘smoke’. Dialects outside of Manila often include historically important features such as the retention of post-consonantal glottal

<sup>7</sup> It is unclear whether this number includes all Bikol languages or only Central Bikol; and whether the three Bisayan languages in the Bikol Region—Masbatenyo, Northern Sorsoganon, and Southern Sorsoganon—may have also been inadvertently included.

## HISTORICAL LINGUISTICS OF THE PHILIPPINES

stops (e.g. *gab-í* /gabʔi/ ‘night’, *big-át* /bigʔát/ ‘heaviness, weight’, and *ngay-on* /ŋayʔún/ ‘today, now’, vs. *gabí*, *bigát*, and *ngayón* in Manila Tagalog); and stress differences (e.g. *bitúin* /bitúʔin/ ‘star’ and *díyan* ‘there, near addressee’ vs. *bituín* /bituʔín/ ~ *bitwín* /bitwín/ ‘star’ and *diyán* ~ *dyan* in Manila Tagalog, respectively).

## 7.2.7.1.2 Bisayan

The primary study of the approximately forty languages belonging to the Bisayan subgroup (some of which are spoken at the southeastern corner of Luzon, in northeastern Mindanao, and in the Sulu archipelago) is Zorc (1977), in which thirty-six Bisayan speech varieties were compared on the basis of functor analysis, shared innovations, and lexicostatistics. Innovations defining the entire Bisayan subgroup include \*dakúʔ ‘big’, \*damgu ‘dream’, \*gəgma ‘love’, \*hibadú ‘know how to do, know (facts)’, \*lúʔuy ‘pity’, \*-naʔ ‘root of 2nd-position deictics’, and \*singit ‘shout’. Zorc’s five-way division of the Bisayan languages—Central, Western, Southern, Cebuano, and Asi/Bantoanon—has been adopted by all subsequent authors writing about these languages, with the sole exception of Gallman (1997), whose proposed Northeastern Mindanao group, containing Cebuano, South Bisayan, Mansakan, and Mamanwa, has not been accepted by other authors, and is best explained as the result of an East Mindanao axis (cf. §7.4.2).

The Western Bisayan subgroup consists of not only ‘standard’ Kalibonhon Aklanon and the ‘standard’ Kinaraya around San Jose de Buenavista in Antique, but also a number of minor languages including Bulalakawnon, Inonhan, Ratagnon, Malaynon, Buruanganon, Nabasnon, Pandananon, Libertadnon, Jamindanganon-Mambusaanon, Panayanon Binukidnon (sometimes referred to as ‘Sulod’), and numerous ‘non-standard’ dialects of Kinaraya in the provinces of Iloilo, Capiz, and Antique. Innovations defining this subgroup include \*ayán ‘go’, \*-gi (root of 1st-person deictic), \*həngəd ‘big, many’, \*hiləj ‘drunk’, \*hiŋgaʔ ‘lie down’, \*kasalpan ‘west’, \*libáyən ‘sibling’, \*ráhaʔ ‘cook’, \*tána ‘he/she, him/her (NOM)’, and \*sánda ‘they, them (NOM)’.

The Central Bisayan subgroup consists of the major languages Ilonggo/Hiligaynon, Waray-Waray, and Masbatenyo, as well as numerous minor languages including Romblomanon, Northern Sorsoganon, Southern Sorsoganon-Northwestern Samarenyo, Northern Samarenyo, Bantayanon, Utudnon/Baybayanon, Kinabalian, Porohanon and possibly also the Northern Binukidnon and Southern Binukidnon languages of Negros Island whose position has yet to be determined. Innovations defining this subgroup include \*irúy ‘mother’, \*kadáʔ ‘go

there (near addressee)’, \*kánam ‘play’, \*píraw ‘sleepy’, and \*sumat ‘say, converse’.

The Cebuano branch of the Bisayan subgroup consists of a single language, Cebuano, represented by numerous dialects throughout the central Visayan Islands. Dialects in the southern third of Cebu Island, the eastern coast of Bohol Island, the southwestern coast of Leyte Island, and the eastern coast of Negros Island retain more conservative phonological and grammatical features than the Cebuano spoken in Cebu City and northern Cebu Island, and throughout Mindanao as an immigrant or second language. As with Tagalog, finding unique forms for Cebuano is difficult due to its extended and complex history of both contributing and borrowing lexicon through contact with various Bisayan, Mansakan, and Manobo languages, but candidates include *daghan* ‘many’, *hagbung* /hagbuŋ/ ‘fall’, *húnung* /húnuŋ/ ‘stop’, *kámu* ‘cook’, *kubut* ‘hold’, and the genitive common noun case marker *ug*.

The Southern Bisayan subgroup consists of Surigaonon, Butuanon, Tandaganon/Tagon-on (called ‘Naturalis’ in Zorc 1977, after the term speakers used with Zorc to distinguish their language from the Cebuano of nearby towns) and Tausug. Innovations include \*bugáʔ ‘afraid’, \*hiram ‘mosquito’ (borrowed into Kamayo, Ata Manobo, and Dibabawon Manobo), \*kawáʔ ‘get, take’ (a reshaping of PPH \*kúhaʔ), \*kunsələm ‘tomorrow’, \*pisak ‘mud’, and \*yupúʔ ‘short’.

The Asi subgroup consists of a single language, Asi/Bantoanon, which is spoken in five towns, each with its own distinct dialect: Banton, Corcuera (on Simara Island), Concepcion (on the island known variously as Sibale, Concepcion, and Maestro de Campo), and Odiongan and Calatrava towns on Tablas Island. The Asi/Bantoanon language is distinguished phonologically by a triad of shifts: PPH \*d > /r/, \*y > /d/, and \*l > /y/, for example, *gador* ‘emphatic marker’ < PGCPH \*gayəd. The five dialects share a number of innovations including *ási* /ásiʔ/ ‘why?’ (from which the language gets its name), *bagúntor* ‘mountain’, *guyáh* ‘laugh’, *hidáit* /hidáʔit/ ‘love’ (other Bisayan \*gəgma), *insulíp* ‘tomorrow’, *kag* ‘nominative common noun case marker’ (most Central Philippine *ang* ~ *an*), *kumán* ‘earlier (in the same day)’, and *nak* ‘linker’ (other Philippine \*nga or \*na). A number of other forms hint at external contacts that speakers of this language may have had over the past millenium, sharing *bilá-bilá* ‘butterfly’ with Romblomanon; *dúta* /dútaʔ/ ‘earth’ with Cebuano *yúta* /yútaʔ/; *maádo* /maʔádu/ ‘good, well’ with Cebuano and Central Bisayan *maáyo* /maʔáyu/ ‘good’; *rampog* ‘raincloud’ with Central Bisayan \*dampəg; and *taybu* ‘dust’ with Rinconada and Bikol Libon *talbu*.

## ZORC, LOBEL, &amp; HALL

## 7.2.7.1.3 Bikol

The Bikol subgroup consists of eight languages native only to the Bikol Region in southeastern Luzon. The primary study of these languages is McFarland (1974), while Lobel (2004, 2005, 2013a) has done a considerable amount of additional work. The subgroups within the Bikol node are Northern Bikol (McFarland's 'Coastal Dialects'), consisting of a single language often called 'Central Bikol' or 'Bikol Naga', spoken in various dialects primarily along the northern coast of the Bikol Region from Daet through Naga, Partido, and Legaspi to the northern coast of Sorsogon, plus in the southern half of Catanduanes Island; Southern Bikol (McFarland's 'Inland Dialects'), consisting of the Rinconada ('Iriga'),<sup>8</sup> Buhiñon ('Buhi'), Bikol Libon ('Libon'), West Albay Bikol ('Oas'), and Bikol Miraya ('Daraga') languages; Northern Catanduanes Bikol ('Pandan') consisting of a single language spoken throughout the northern half of Catanduanes Island; and Inagta Bikol, a language spoken by the Agta of Mt. Isarog and Mt. Iriga/Asog in Camarines Sur province (Lobel n.d.-c). Not included in the Bikol subgroup are the three Bisayan languages that are native to the Bikol Region, Northern Sorsoganon ('Sorsogon'), Southern Sorsoganon ('Gubat'), and Masbatenyo ('Masbate'), the first two of which are most often simply called 'Bikol' by their speakers in spite of actually belonging to the Central Bisayan subgroup (Zorc 1977). Innovations defining the Bikol subgroup include \*bayúŋ 'bird', \*ʔəsád ~ \*sarúʔ 'one', \*gədaʔan 'die, kill', \*həlay 'long (time)', \*payú 'head', \*rahay 'good', and \*sadáy 'small'.

## 7.2.7.1.4 Mansakan

The Mansakan subgroup consists of nine languages split between three branches: Kamayo (with a northern and southern dialect) in the Northern branch; Mansaka, Davawenyo, Samā IGaCOS<sup>9</sup> and the various dialects of Mandaya<sup>10</sup> in the Central branch; and Kāgan, Kalagan<sup>11</sup> and Tagakaulo in the Southern branch. Zorc (1977) was the first scholar to propose a Mansakan subgroup, which Gallman (1979) later referred to as a 'South-East Mindanao' node within what he would later (1997) propose as an East Mindanao subgroup (a grouping which no other scholars have accepted). Proto-Mansakan innovations include \*atulun 'fire', \*daʔig 'many', \*hambun 'afternoon', \*hikəl 'laugh', \*kamayu 'to

<sup>8</sup> The remaining alternate names in parentheses in these two paragraphs are those used by McFarland (1974).

<sup>9</sup> This little-known language, spoken on the Island Garden City of Samal (IGaCOS) near Davao City and sometimes referred to as 'Samal', should not be confused with the only distantly related Sama languages.

<sup>10</sup> Not including the group identifying as Mandaya in Monkayo town, who in fact speak Dibabawon, not Mandaya.

<sup>11</sup> While *Kāgan* /kaagan/ is simply the native pronunciation of *Kalagan*, the latter is usually used in the Davao region to refer to the Islamic ethnolinguistic group, while the former is used to refer to the non-Muslim group.

you (2PL.OBL)', \*kisələm 'tomorrow', \*kulkulhun 'fingernail', \*lumun 'sibling', \*tiyayuʔ 'cry, weep', and \*yaʔan '3SG.NOM'.

## 7.2.7.1.5 Mamanwa

The Mamanwa language is spoken by an Aboriginal Filipino population of the same name native to the northeastern Mindanao provinces of Surigao del Norte, Surigao del Sur, and Agusan del Norte, with small but long-established migrant communities on the eastern Visayan islands of Samar, Leyte, and Biliran. In spite of the fact that the vast majority of its lexicon appears to be borrowed from neighbours such as Kamayo, Surigaonon, Agusan Manobo, and Cebuano, Mamanwa retains a number of functors which point to a non-Central Philippine origin. Most of our knowledge about this language is the product of the decades that missionary sisters Helen Miller and Jeanne Miller spent working among the Mamanwa (Miller 1964, 1973; Miller and Miller 1969, 1976, 1991), with only Lobel (e.g. 2013a) having done additional work on this group in the decades since. Unique Mamanwa forms include *kamahan* 'monkey' and *nao* /naʔo/ 'I, my (1SG.GEN)', and a handful of borrowings hint at past interactions with Mansakan (e.g. \*atmuʔ 'full, replete', \*kamayu '2PL.OBL', \*kulkulhun 'fingernail', \*lumun 'sibling'); South Bisayan (e.g. \*dəkag 'itch', \*laʔas 'old, of person', \*ləpəs 'rope', \*taʔəd 'many'); and Manobo (e.g. \*bubun 'mountain', \*buhiʔ 'full, sated', \*ʔimpis 'egg').

## 7.2.7.2 Manobo

The Manobo subgroup consists of at least nineteen languages spoken throughout central and eastern Mindanao. Elkins (1974, 1984) has addressed the entire group in great detail, while Zorc (1974a) and Harmon (1977) have addressed the position of Kagayanen (a geographically distant member of the Northern Manobo subgroup) and Burton (1996) presents a study of borrowing relationships between various Manobo and Mansakan languages. Following Elkins plus Lobel's subsequent work on the entire group (including languages for which data was unavailable to Elkins), the nineteen Manobo languages can be classified into five branches: Northern, including Talaandig-Higaonon (a.k.a. Binukid), Kinamiging, Banwaon, and Kagayanen; Southern, including Tagabawa, Sarangani Manobo, Cotabato-Kalamansig Manobo, and Tasaday; Core-Western, including Obo Manobo, Ilianen Manobo, Western Bukidnon Manobo, and Pulangiyan Manobo; Core-Central, including Ata Manobo, Matig Salug Manobo, and Tigwa Manobo; and Core-Eastern, including Agusan Manobo, Rajah Kabungsuwan Manobo, Umajamnon, and Dibabawon. Among the hundreds of Proto-Manobo lexical innovations reconstructed by Elkins (1974, 1984) are \*ʔahaʔ 'see', \*ʔaram

## HISTORICAL LINGUISTICS OF THE PHILIPPINES

‘choose’, \*bakasan ‘snake’, \*din ‘3SG.GEN’, \*rimusəng ‘sweat’, \*gətək ‘belly’, \*laŋsə ‘blood’ (although Southern Manobo has \*dipanug), \*lipədəŋ ‘sleep’, \*pinənuʔu ‘sit’, and \*tabək ‘answer’.

7.2.7.3 *Subanen*

The Subanen subgroup consists of seven languages spoken in the Zamboanga Peninsula in western Mindanao: Northern Subanen, Southern Subanen, Eastern Subanen, Central Subanen, Western Subanon, Western Kolibugan, and Salug-Godod Subanen (including Tawlet/Kalibugan). The primary division in this subgroup is between a Western branch which contains Western Subanon and Western Kolibugan, and the Nuclear branch which contains the other five. Early work on these languages includes Christie (1909) and Finley and Churchill (1913), while more recent work has been done by Hall (1987); Daguman (2004); Lobel and Hall (2010); Lobel (2013a); and Estioca (2020). This subgroup is defined by several phonological innovations and over seventy lexical innovations (Lobel 2013a), including \*dupiʔ ‘rain’, \*gəbək ‘run’, \*gəŋaj ‘gills’, \*ləgdəŋ ‘straight’, \*m[a]-ikaʔ ‘small’, \*tapuk ‘lungs’, and \*tərawan ‘spear’.

7.2.7.4 *Danaw*

The Danaw (or ‘Danao’) subgroup includes Maranao, the various dialects of the Maguindanaon and Iranun languages in central and western Mindanao, and the Eastern Sabah and Western Sabah dialects of Iranun in Sabah, Malaysia. While the late Howard McKaughan contributed substantially to our knowledge of these languages (1958, 1959, 1962, 2002a, b, c; McKaughan and Macaraya 1967, 1996), very little comparative work has been done on this subgroup other than a handful of articles (Allison 1979; Fleischman 1981; and Lobel and Riwarung 2009, 2011).<sup>12</sup> Innovations defining the group are both lexical (e.g. \*mayaw ‘hot’, \*agag ‘dry in sun’, \*idtuŋ ‘throw’) and phonological, including complex rules reducing earlier consonant clusters to the eleven clusters permissible in Proto-Danaw.

7.2.7.5 *South Mangyan*

The South Mangyan subgroup consists of the Hanunoo, Buhid, Eastern Tawbuwid, Western Tawbuwid, and Bangon languages spoken in central and southern Mindoro Island. Zorc (1974b) has treated the internal subgrouping of

<sup>12</sup> Many authors have asserted that there was a difference in endonyms between the Iranun of Mindanao and the Iranun of Sabah. In fact, both groups refer to themselves exclusively as ‘Iranun’ (sometimes spelled ‘Iranon’), while names such as ‘Ilanun’ and ‘Illanun’ are exclusively exonyms used in Malay, Maguindanaon, and English.

the Mangyan languages in general, while Pennoyer (1980) addressed the relationship between Buhid and the Tawbuwid.<sup>13</sup> Very few lexical innovations have been identified for this subgroup so far (e.g. \*labuŋ ‘leaf’, \*sirəŋ-sirəŋ ‘daily’), while a handful of forms are shared exclusively with Bisayan (e.g. \*badás ‘sand’) or Bikol (e.g. \*túkaw ‘sit’), likely due to contact with these two groups at some point in the past.

7.2.7.6 *Mongondow-Gorontalo*

One of three Philippine subgroups located outside of the geographical Philippines, along with Sangiric (§7.2.12) and Minahasan (§7.2.13), the Mongondow-Gorontalo subgroup consists of nine languages spoken in central northern Sulawesi, Indonesia: Mongondow, Lolak, Bintauna, Bolangitang-Kaidipang, Bolango, Suwawa, Buol, Gorontalo, and the now-extinct Ponosakan. Recent work on these languages includes that of James Sneddon and late Indonesian scholar H. T. Usup (Usup 1984, 1986; Sneddon and Usup 1986; Sneddon 1991; Lobel 2011, 2015, 2016a; and Lobel and Papatungan 2017).<sup>14</sup> Following Charles (1974), Blust (1991b) groups the Mongondow-Gorontalo languages in the Greater Central Philippine subgroup. The 140 Proto-Mongondow-Gorontalo lexical innovations identified by Lobel (n.d.-d) include \*boyod ‘rat’, \*buloy ‘spouse’, \*gogoyon ‘hungry’, \*mo-lanit ‘sharp (of edge)’, \*lituʔ ‘sit’, \*liyoyŋ ‘forget’, \*oykag ‘river’, \*porok ‘smoke (n.)’, and \*utas ‘sibling’.

7.2.7.7 *Palawan*

The Palawan subgroup is generally defined as consisting of Batak, Aborlan Tagbanwa, Molbog, and various languages spoken by groups identifying as Palawan ~ Pala-wan,<sup>15</sup> including Brooke’s Point Palawan, Central Palawan, Southwest Palawan, and the closely related languages of the Panimusan (long-Islamized coastal groups who do not identify as ‘Palawan’ but whose language is nevertheless largely identical to that of the non-Muslim Palawan people). Thiessen (1981) is the only study of the entire subgroup. Scebald (2003) later introduced the Central Tagbanwa language, for which data was not previously available. Blust (1991b) argues that these languages form one of seven nodes of the Greater Central Philippine macrogroup, but Zorc questions

<sup>13</sup> Barbian (1977) contains an invaluable expanded wordlist, but the subgrouping analysis contained therein is highly problematic, based on shared features and ignoring innovations and important differences between the languages. Likewise, Barbian treats Ratagnon as a Mangyan language, when it clearly belongs to the Western Bisayan group.

<sup>14</sup> Lobel is currently working on a dictionary and grammar of the now-extinct Ponosakan and the moribund Lolak.

<sup>15</sup> The latter reflecting the conservative reflex /palaʔwan/ found in the Palawan dialects that preserve /ʔC/ clusters.

## ZORC, LOBEL, &amp; HALL

this since, in spite of sharing the \*R > \*g shift, the number of PGCPH lexical innovations found in the Palawanic languages is minimal, a fact that could also indicate that Proto-Palawanic was the earliest to split from Proto-GCPH. The Palawanic languages share a number of lexical innovations including \*bakal ‘throw’, \*bəgit ‘bird’, \*gərəŋ ‘back’, \*rayak ‘pull’, and \*tabuk ‘smoke’. The Northern and Southern branches of Palawanic are distinguished by forms such as \*kəʔdəŋ ‘dog’ in the former corresponding to \*idəŋ in the latter, both replacing PAN \*asu. The Southern branch is further distinguished by innovations such as \*dələk ‘rain’, \*kəsit ‘laugh’, and \*tipusəd ‘sibling’.

## 7.2.8 Kalamianic

The Kalamianic subgroup consists of the Agutaynen, Calamian Tagbanwa, and Karamianen languages, all located on islands between northern Palawan, southern Mindoro, and western Panay. Himes (2007) is the primary historical-comparative work on these languages, which are distinguished by the phonological shifts of PMP/PPH \*R > \*l and \*q > \*k. Lexical innovations defining the Kalamianic subgroup include \*aliŋət ‘near’, \*aniŋ ‘say’, \*ələd ‘fear’, \*gəʔəd ‘bolo’, \*guʔuy ‘call’, \*kandas ‘liver’, \*kult ‘white’, \*tan- ‘3rd-person pronominal formative’, and \*yawaʔ ‘you (2SG.NOM)’.

## 7.2.9 Inete/‘Inati’ and the Ata of Negros

While the existence of the Ete/‘Ati’<sup>16</sup> people of Panay Island had been known since the Spanish occupation of the Philippines, it wasn’t until Pennoyer (1986–1987) that any data from, or description of, their language surfaced, and little Inati data has appeared in the literature outside of that included by Pennoyer and in Lobel (2013a). The language appears to form a primary branch of the Philippine subfamily (Blust 1991b: 80), with Blust noting that “[u]niquely among languages in the Philippines, it has merged \*R and \*d in at least final position (intervocally pre-Inati \*d became /r/, and subsequently pre-Inati \*R became /d/)”, for example, *kadat* ‘bite’, *kiturud* ‘sleep’, *paridus* ‘bathe’. There are also a large number of unique forms and features in the functor subsets, but the vast majority of its lexicon has been borrowed from neighbouring Bisayan languages on Panay, primarily Kinaray-a and Aklanon, but also Ilonggo. Unique lexical items include *sapiw* ‘house’, *awuy* ‘yes’, *nalang* ‘no’,

<sup>16</sup> Note that *Ete* /ete/ (likely underlying \*ata, as /a/ raises to [ɛ] ~ [æ] in the Inete language) and *Inete* /inete/ are the endonyms for the group and its language, respectively, while *Ati* and *Inati* are the Bisayan corruptions thereof.

*umê* ‘arrive’, *dugúk* ‘go’, *ngadin* ‘don’t know’, *pegek* ‘chicken’, *betleng* ‘put’, *da-it* ‘rain’, *miyá* ‘what?’, *ki-ara* ‘where?’, *mesned* ‘far’, *himpun* ‘fire’, and *gine* ‘also’.

It should be noted that the Ete/‘Ati’ of Panay, Guimaras, and Boracay islands are distinct from the Ata people of Negros Island (who in turn should not be confused with the linguistically and ethnically distinct Ata Manobo of central Mindanao, who are closely related to the neighbouring Tigwa and Matigsalug Manobo). The Ata of Negros Island are descendants of a once widespread population encountered by the Spanish in the sixteenth century (Rahman and Maceda 1955; Scott 1984, 1992), thus compelling them to name the island ‘Negros’ (from the Spanish plural for ‘black’, or, in this context, ‘Black person’). The most important linguistic distinction is that while the Ati and the Ata have distinct ethnic identities, only the Ati retain a complete language distinct from those of their neighbours, whereas Lobel, in multiple visits to Ata communities throughout the island, has found only three Ata who can remember even fragments of an Inata language distinct from the neighbouring Binukidnon languages native to the mountains of Negros Island. Unique Inata forms where these few individuals were in linguistic agreement during these separately collected elicitation include *din-ay* ‘there (far from speaker and addressee)’, *kan-ay* ‘go there’, *dihna* ‘here’, *bangut* ‘man’, *tairan* ‘woman’, *kukuyuban* ‘house’, and *tukub* ‘eat’ (Lobel n.d.-e).

## 7.2.10 Southwestern Mindanao

The Southwest Mindanao subgroup consists of Tboli (sometimes called ‘Tagabili’), Koronadal Blaan, Sarangani Blaan, and the more distantly related Tiruray (more correctly ‘Teduray’). Innovations defining the Southwest Mindanao subgroup include \*bakuŋ ‘deaf’, \*butəŋ ‘night’, \*dawin ‘loincloth’ (vs. PMP \*bahaR), \*deʔe ‘many’, \*kodog ‘boil (water)’, \*isaq ‘break open, hatch’ (vs. PAN \*pəcəq), \*kuwah ‘oar, paddle’, \*lakay ‘tail feather of rooster’ (vs. PMP \*lawi), and \*Rasan ‘skinny’. The Tboli-Blaan node is defined by innovations such as \*boŋ ‘big’, \*fayah ‘tomorrow’, \*kahuŋ ‘swim’ (vs. PAN \*Nanguy), \*kahiʔ ‘salt’ (vs. PAN \*qasiN), and \*litəʔ ‘blood’ (cf. PAN \*Nitəq ‘sap’).

## 7.2.11 Southeastern Mindanao

This group consists of the various dialects of the Klata language (also known as Bagobo Klata, Giangan, or Diangan), spoken in Magpet town in Cotabato province, and in the Baguio, Calinan, Marilog, and Tugbog districts of Davao City. This moribund language was only recently discovered to be distinct from all other southern Philippine languages

(cf. Evans 2017 and Zorc 2019). It is no longer generally being passed on to children, largely due to intermarriage with Cebuanos and other local groups. A few of its unique vocabulary items include the plural marker *bɛ* (vs. Proto-Southern-Philippines \**maŋa*), *benne* /*benneʔ*/ ‘cry, weep’ (vs. PAN \**Caŋis*), *byoo* /*byoʔo*/ ‘year’ (vs. PMP \**taqun*, PPH \**dagʔun*), *klammaŋ* ‘star’ (vs. PAN \**bituqən*), *kulung* ‘back’ (vs. PAN \**likud*), *lammi* ‘new’ (vs. PAN \**baqəRuh*), *ongob* /*oŋob*/ ‘fingernail’ (vs. PMP \**kukuh*), and *paya* ‘big’ (vs. PAN \**Raya*, PMP \**laba*, PPH \**dakəl*).

### 7.2.12 Sangiric

Separated from one another by the maritime border between the Philippines and Indonesia, the five Sangiric languages have been recognized as a primary branch of the Philippine subgroup since Sneddon (1984), the primary work on this subgroup. Three of these languages (Sangir/Sangihe, Talaud, and Sangil) are spoken on a series of small islands between southeastern Mindanao and northeastern Sulawesi, while the other two (Ratahan/Pasan and Bantik) were traditionally spoken in the Minahasa subprovince on the mainland of northeastern Sulawesi. Like the vast majority of languages native to northeastern Sulawesi, the Sangiric languages within the national borders of Indonesia are all moribund and highly endangered, seldom spoken, and not being learned by anyone younger than the ‘grandparent generation’, as the entire population has switched to Manado Malay. Innovations identified by Sneddon (1984) include \**akeʔ* ‘water’, \**babəlaw* ‘afternoon’, \**busak* ‘banana’, \**iaʔ* ‘I (1SG.NOM)’, \**payarŋ* ‘thigh’, \**pepe* ‘urine’, \**siŋaʔ* ‘know a person’, \**tanak* ‘live, dwell’, \**təbay* ‘old (of object)’, \**təmbuʔ* ‘head’, \**tipu* ‘smoke’, \**timbonan* ‘head’, \**t[io]ŋkaRia* ‘ear’, \**tətuR* ‘hot coals, embers’, and \**tolay* ‘tail’. Note that the Sangiric and Minahasan groups, bordering on one another in some areas, share a handful of innovations, including \**dou* ~ \**r<sub>2</sub>eʔo* ‘thirst’, \**paluka* ‘shoulder’, \**həŋisəʔ* ~ \**rəŋis* ‘burn’, \**dirihəʔ* ~ \**ririh* ‘yellow’, \**tagas* ‘low tide, ebb’, \**təkəl* ‘sleep’, \**tumpa* ‘descend, alight’, \**tunay* ‘thorn’, and \**utak* ‘hair’ (< PMP \**utək* ‘brain’).

### 7.2.13 Minahasan

The Minahasan languages are spoken exclusively on the mainland of northeastern Sulawesi, in what was formerly the subprovince or ‘regency’ of Minahasa (now broken up into several smaller regencies). Sneddon (1978) assigned these five languages—Tontemboan, Tonsawang, Tonsa, Tondano, and Tombulu—to the Philippine subfamily, a

position also adopted by Blust (1991b). Similar to the Sangiric languages, all of the Minahasan languages except Tonsawang are moribund, with even elderly Minahasans now speaking Manado Malay much more frequently than their ancestral languages, which virtually no children are currently learning. Proto-Minahasan innovations identified by Sneddon (1978) include \**baŋkoʔ* ‘big’, \**bə<sub>2</sub>ən* ‘eye’, \**bisa* ‘which?, where?’, \**datə* ‘cold’, \**ələp* ‘drink’, \**əŋah* ‘cough’, \**biaʔi* ‘here’, \**kəʔkəʔ* ‘laugh’, \**pəntuʔ* ‘bitter’, \**tələb* ‘fly (away)’, and \**tiəy* ‘pig’.

## 7.3 Phonological developments

Much has been written on both the synchronic phonologies of Philippine languages, and on various aspects of the historical phonology of the Philippine subfamily or various branches thereof, starting with such noteworthy early works as Conant (1911, 1912, 1916) and Dempwolff (1925). This section will therefore present only a very brief overview of Philippine phonology and phonological developments.

### 7.3.1 Consonants

Proto-Philippines can be reconstructed with a phonological system consisting of twenty consonants and four vowels (Blust, pers. comm., 9/28/2020). Synchronically, the vast majority of Philippine languages have either three- or four-vowel systems, and most commonly a sixteen-member consonant inventory, as illustrated in Table 7.1.

The following relatively rare consonants are also known to occur in some languages:

- a) /β/ (\*-b-) in Central Cagayan Agta and some Manobo and Sangiric languages
- b) /f/ (< \*p) in Koronadal Bilaan, Sarangani Bilaan, Tiruray, Tawbuwid, and Bangon
- c) voiced velar approximant /w/ (< \*l) in Buhi-non Bikol and Aklanon (written <e> in the latter)
- d) interdental approximant /ʃ/ (< \*l and sometimes \*r) in Southern Catanduanes Bikol, Kagayanen, the Cajiocan dialect of Romblomanon, and some Mandaya dialects
- e) retroflex tap [ɽ] (< \*l preceding /i/ or /e/ if not following another /i/ or /e/), in Mongondow, Lolak, and certain other Mongondow-Gorontalo languages
- f) retroflex lateral approximant [ɭ] (< \*l adjacent to any combination of the vowels /a o u/ whether word-initial, word-final, or intervocalic) in Mongondow, Lolak, and certain other Mongondow-Gorontalo languages (Lobel and Paputungan 2017)

**Table 7.1** The phoneme system of Proto-Philippines (Blust, pers. comm., 9/28/2020) and the most common Philippine phoneme system

THE PHONEME SYSTEM OF PROTO-PHILIPPINES						THE MOST COMMON PHILIPPINE PHONEME SYSTEM					
CONSONANTS					VOWELS	CONSONANTS					VOWELS
*p	*t	*k	*q	*ʔ	*i	*u	p	t	ʔ	i	u
*b	*d	*j	*g		*ə		b	d	ʔ	(ə/i/o)	
	*s			*h	*a		s	h		a	
*m	*n	*ñ	*ɲ				m	n	ɲ		
	*l						l				
	*r			*R			r				
*w	*y						w	y			

- g) a series of heavy voiceless obstruents /p' t' k' s'/ in Maranao (< Proto-Danaw \*bp, \*dt, \*gk, and \*ds, respectively) (Lobel and Riwarung 2009, 2011)
- h) a series of aspirated voiceless obstruents /p<sup>h</sup> t<sup>h</sup> k<sup>h</sup> s<sup>h</sup>/ in Southern Subanen (< Proto-Subanen \*kp, \*kt, \*gk, and \*ks, respectively) (Lobel and Hall 2010).

The glottal fricative /h/ is found in many Philippine languages, sometimes as a continuation of PMP \*h (< PAN \*S), other times as a reflex of other PMP/PPH phonemes such as \*s, \*r, or \*R, and, in some languages, only in forms borrowed from Spanish, Malay, or English. PMP \*h itself is continued as /h/ in Itbayaten, Manide, Inagta Alabat, most Central Philippine languages (including Tagalog, Mamanwa, and at least some members of the Bikol, Bisayan, and Mansakan subgroups), and several mostly northern Manobo languages (Talaandig-Higaonon, Kinamiging, Banwaon, and Umajamnon).<sup>17</sup> Other sources of /h/ in Philippine languages include:

- a) \*s > /h/ in Amganad Ifugao, Ayta Abellen, Ayta Ambala, Ayta Bataan, Ayta Mag-anchi, Balangaw, Batad Ifugao, Bayninan Ifugao, Botolan Sambal, Kayapa Kallahan, Keley-i Kallahan, Koronadal Blaán, Sarangani Blaán,

<sup>17</sup> A number of other languages such as Casiguran Agta have /h/ < \*h in borrowings but not in inherited forms.

- and Tboli; and sporadically in all of the Northeastern Luzon languages except Casiguran Agta
- b) \*r > /h/ in Pahanan Agta, Casiguran Agta, Nagtipunan Agta, Dinapigue Agta and the southern dialect of Dupanangan Agta
- c) \*l > /h/ in Ivatan, and sporadically in Tagalog and some dialects of Bantayanon
- d) \*R, \*g > /h/ in Ponosakan, Suwawa, and Gorontalo (plus \*b > /h/ in some environments in Gorontalo)
- e) \*R > /h/ in Minahasan, and in a handful of etyma in Mamanwa
- f) \*k > /h/ in Southern Subanen and Buhid
- g) \*p, \*q > /h/ in Tboli
- h) \*b, \*d, \*j, \*R, \*r > /h/ in Central Cagayan Agta.

The occurrence of word-final /h/ is limited to Itbayaten, Ayta Abellen, Tina Sambal, Mamanwa, Aklanon, Surigaonon (primarily in dialects in Surigao del Sur), Bantayanon (at least in rural dialects), Northern Samarenño, and Eastern Samar Waray.<sup>18</sup> Thus PAN \*CiṛaS, PMP \*tiṛah 'food stuck in teeth' > Itbayaten *tiñah*, Aklanon, Mamanwa *tiṛáh*. Likewise,

<sup>18</sup> There are differences in the retention of word-final /h/ between reduplicated monosyllables (e.g. \*muhmuh 'rice crumbs') and words of other forms (e.g. \*təbuh 'sugarcane'), with some languages such as Bantayanon, Surigaonon, and some dialects of Waray-Waray preserving word-final \*h in the former but not the latter.

other than in words of the shape CVhCVh in the aforementioned languages, few others preserve /h/ as a reflex of PMP \*h in non-word-final coda position, such as Inata *dihna* ‘here’, Northern Binukidnon *dihni* ‘here’, Manide, Inagta Alabat *beh-en* /beh?en/ ‘sneeze’. Finally, an underlying word-final /h/ is posited on orthographically vowel-final words in certain other languages like Tagalog and Cebuano, but it is only pronounced when inflected, for example, *natiyahán* ‘happened to have food stuck in one’s teeth’ (although this morphophonemic /h/ is far more regular in Tagalog than in Cebuano).

### 7.3.2 Vowels

Proto-Philippines had a four-vowel system (\*a \*i \*u \*ə), and most of its daughter languages have three or four phonemic vowels. Languages that have three-vowel systems have merged PPH \*ə with either \*a, \*i, or \*u. Those that have four-vowel systems are split between those that continue \*ə as a tense high central vowel /i/, a lax mid central vowel /ə/, a mid back vowel /o/ (phonetically [ɔ] in some languages), and a mid front vowel /e/. Languages with more than four vowels include:

- a) Casiguran Agta, described in Headland and Headland (1974) as having an eight-vowel system consisting of /a e ε i o ɔ u i/
- b) The Oas dialect of Miraya Bikol, which has monophthongized sequences of \*au (> /o/), \*ai (> /e/), and \*aə (> /u/) while maintaining a distinct reflex of \*ə, producing a seven-vowel system consisting of /a e i o u i u/
- c) A number of Mongondow-Gorontalo languages, in which a non-etymological /e/ appears alongside an /o/ reflex of \*ə, yielding a five-vowel system consisting of /a e i o u/.

Furthermore, educated and/or urban speakers of many other Philippine languages have acquired a non-native distinction between /e/ and /i/, and /o/ and /u/, even though these two vowels (/e/ and /o/) are not otherwise contrastive in the native lexicon of their languages. However, Spanish and English loanwords are adapted to the native phonology for most speakers of Philippine languages, even major languages like Tagalog, Cebuano, Ilokano, and Ilonggo.

Finally, the Maranao language of central Mindanao is unique in having developed a system of voice register (Lobel and Riwarung 2009, 2011) reminiscent of that found in certain Mon-Khmer languages, in which earlier consonant

clusters have developed into the unitary heavy consonants mentioned earlier in this section, which have a raising and tensing effect on the following vowel. Although acoustic analysis is still lacking, it is clear that what initially appear to be eight distinct vowels in Maranao are instead two complementary sets of four vowel allophones, with each set occurring after a separate set of consonants.

## 7.4 Issues and controversies

By far the largest controversy in the historical linguistics literature on Philippine and Philippine-type languages has been the question of whether or not there is sufficient evidence to posit a ‘Proto-Philippines’, that is, whether the languages of the Philippines (minus the Sama-Bajaw languages) plus Yami in Taiwan and the Sangiric, Minahasan, and Mongondow-Gorontalo languages of Sulawesi in Indonesia, are more closely related to one another than to any other languages. One of the major issues that have been resolved (see Zorc and Almarines CHECK 2022) is the appearance of innovations that cross-cut genetic boundaries established by the comparative method which are *axis relationships*, geographic and sociolinguistic unities among languages based on subsequent trade and cultural ties. These will be discussed in the following sections.

### 7.4.1 Proto-Philippines

As Blust (2019b) notes, the existence of a Philippine subfamily was largely assumed for much of the twentieth century, and rarely debated until Reid (1982) argued against it in a now-retracted paper (Reid 2020) based primarily on the presence or absence of what he claimed was an “intrusive nasal” (in spite of providing neither a list of such forms nor thorough evidence to support his claim). In response, Zorc (1986) presented ninety-eight lexical Proto-Philippine innovations that had not previously appeared in the literature. Little more appeared on the issue for almost two decades until Ross (2005) revisited the issue in his reanalysis of the position of the Batanic languages, pointing out the lack of evidence for a Proto-Philippines beyond what he characterized as a relatively small number of lexical items which he believed could have been borrowed through contact. However, while working on his Austronesian Comparative Dictionary (Blust and Trussel 2020), Blust has identified hundreds of lexical items unique to the Philippine languages, ultimately publishing a comprehensive treatise (Blust 2019b) defending the existence of a Proto-Philippines based on over

ZORC, LOBEL, & HALL

a thousand lexical innovations and the merger of proto-phonemes \*z and \*d (as opposed to the retention of both \*ñ and \*n). While less than a year has passed since the appearance of Blust (2019b), reaction papers scheduled to appear in the 2020 issue of *Oceanic Linguistics* can be summarized as follows:

- (1) Zorc (2020), generally supportive of Blust’s arguments, proposes that the complex accent patterns exclusive to the Philippines (cf. Smith, chapter 2, this volume) and the reconstruction of an initial PPH \*y- may provide additional supporting evidence for a Proto-Philippines, identifying nine minimal pairs for accent reconstructed for PPH within Blust’s article. Zorc also deals with the lack of true vowel sequences and the need to reconstruct glottal stop in all positions (initial, intervocalic, preconsonantal, post-consonantal, and word-final) both in most microgroups and for PPH itself. He proposes that forms shared exclusively between Bashiic and Ilokano were most probably the product of a Bashiic-Ilokano Axis, and disputes the value of Casiguran Agta as any more than a ‘witness’ language due to evidence of heavy borrowing from Tagalog and/or Kasiguranin.
- (2) Liao (2020) and Reid (2020) are more critical, with Liao pointing out that none of Blust’s proposed PPH innovations are retained in all of its daughter branches, and argues that “[i]ssues with. . . negative evidence cannot be eased simply by drastically increasing the number of lexical innovations. . . not established through. . . bottom-up reconstruction”. Similar to Zorc (2020), Reid notes problems with Blust’s reconstruction of \*q instead of \*ʔ in certain cases, including thirty-nine etyma with no Tboli or Kalamianic evidence; six etyma where either Tboli or Kalamianic justify \*q; and three etyma where Agutaynen has irregular reflexes. Reid also revisits prenasalization (-NC- clusters), unknown in Formosan languages except in reduplicated monosyllables (Dahl 1976: 128), and difficult to reconstruct for PPH except as the product of syncope of words infixes with \*-<um> or \*-<in>.

It remains to be seen what further discussion will surface on this issue. Combined with the merger of \*d and \*z and the apparent innovation of phonemic accent, the sheer number of lexical innovations presented by Blust (2019b) would seem to complicate any arguments that they are simply the result of borrowing. However, it is likely that the lack of grammatical innovations (understandable in light of the relative conservatism of the Philippine languages as a whole) will

motivate at least some authors to continue to question the validity of this grouping.

#### 7.4.2 The complicating factors of axis relationships

In writing his dissertation in 1972–1973, Zorc was confronted with a number of innovations that were spread across various otherwise well-established subgroups, for example, the replacement of PAN \*bəli ‘to buy’ (retained in Tagalog, Tausug, Mansaka, and Bilic) by \*bakál in Aklanon *bakáe* /bakáw/, Asi, Romblomanon *bakáy*, Bulalakawnon, Kinaray-a, Caluyanan, Ilonggo/Hiligaynon, Masbatenyo, Hanunoo, and all Bikol languages *bakál*. As this form \*bakál cuts across four separate discrete subgroups (West Bisayan, Central Bisayan, Bikol, and South Mangyan), Zorc posited forms with this distribution as evidence of a North Bisayan ‘Axis’.<sup>19</sup> It is clear from evidence in Reid (1971) and Zorc (ongoing) that analogous replacements have occurred throughout the Philippines: Waray, Cebuano, Surigaonon, Mamanwa, Kamayo, Kagayanen, and several Mansakan dialects have *palít* (from PPH \*palít ‘exchange’ [ACD], possibly related to Dempwolff’s \*palit ‘return gift’), suggesting an Eastern Mindanao Axis; Kalamianic and Palawanic have \*alaŋ, supporting a Palawan–Kalamian Axis; Northern Philippine languages have either \*gátaŋ or \*lákuʔ, suggesting a northern Luzon Axis; South Cordilleran has \*tuŋgal, which is unique to that subgroup; and the Danao languages, Dibabawon and Western Bukidnon Manobo share *pamasa* ‘buy’ (possibly from Persian *bāzār* ‘market’ via Malay *pasar* + \*paN-). Each of these replacements represents ‘leakage’ (in the terminology of Blust 2019b for loans across genetic boundaries) from one well-established subgroup into other neighbouring languages where significant trade or social networks once existed. These post-split innovations give the false impression of a genetic subgrouping, whereas what they actually indicate is a significant sociolinguistic replacement phenomenon.

Thus far, Zorc has uncovered evidence for the following eight axis relationships (note that some languages such as Ilokano, Tagalog, and Hanunoo are included in two or more such axes):

<sup>19</sup> Zorc coined the term ‘axis’ in 1972, since which similar phenomena have been described as ‘network’ by Milroy and Milroy (1985), and as ‘linkage’ by Pawley and Ross (1995) and Ross (1988). Note that this phenomenon could also be accounted for by the German terms ‘Sprachbund’ and ‘Sprechbund’.

## HISTORICAL LINGUISTICS OF THE PHILIPPINES

- 1) **Bashiic-Ilokano axis** [Ilokano and the Batanic/Bashiic languages]  
 \*dúyuR ‘coconut-shell receptacle for food or water’: Ilokano *dúyog*, Yami, Itbayaten, Ibatan *royoy*, Ivatan *duyuy*.  
 \*káláʔat ‘sudden, abrupt’: Ilokano *kelláʔat*, Itbayaten *akxat*.  
 \*laŋlaŋ ‘eat together as a group’: Ilokano *ag-la-laŋlaŋ* ‘to eat together’, Itbayaten *xayxay* ‘eating in a group (at least two)’, Ibatan *haŋhaŋ* ‘two people eat together’.  
 \*Rábat ‘flotsam’: Ilokano *gábat* ‘flotsam, debris, stray, straggler, loot’, Isamorong Ivatan *yavat* ‘driftwood’.  
 \*RaRáj ‘large marine mollusk (Turbo marmoratus)’: Ilokano *raráj* ‘kind of large, elongated mollusk with a pointed shell; mother-of-pearl’, Itbayaten *yayaŋ* ‘seashell with a shutter or lid: Turbo marmoratus (larger of the two Turbo varieties)’, Ivatan *yayaŋ* ‘turbo shell’, Ibatan *yayaŋ* ‘kind of large sea snail’.
- 2) **Northern Luzon axis** [Ilokano, Cagayan Valley, and Central Cordilleran]  
 \*layús ‘flood’: Ilokano, Gaddang, Manabo, Luba, Itneg and Isinai.  
 \*lukmæg ‘fat’: Ilokano, Luba, Bontok, and Isneg.  
 \*sabáli ‘other, different’: Isneg, Ilokano, Manabo, Itneg, and Balangaw.  
 \*salʔit ‘lightning’: Isneg, Ilokano, Itneg, and Kalinga.  
 \*suʔpit ‘narrow’: Isneg, Malaweg, Kalinga, and Manabo.
- 3) **Central Luzon axis** [Central and South Cordilleran]  
 \*bətík ‘run’: Amganad Ifugao, Kiangan Ifugao, Ibaloi, Kallahan, and Pangasinan.  
 \*bútəŋ ‘drunk’: Isinai, Kiangan Ifugao, Northern and Southern Kankanaey, Inibaloi.  
 \*dagəm ‘wind’: Isinai, Northern and Southern Kankanaey, Ibaloi, Kayapa Kallahan, and Pangasinan.  
 \*imuk ‘mosquito’: Isinai, Ibaloi, and Kayapa Kallahan.  
 \*taláw ‘star’: Balangaw, Bontok, Luba, Kankanay, Ibaloi, Kayapa Kallahan.
- 4) **Southern Luzon axis** [Tagalog, Sambalic, Kapampangan, Remontado/Hatang-Kayi, Pangasinan, Casiguran Agta, Bikol, and Hanunoo]  
 \*alikabúk ‘dust’: Tagalog, Kapampangan, Botolan Sambal, Ayta Mag-Indi, Bulalakawnon, and Casiguran Agta.  
 \*buláti ‘earthworm’: Tagalog, Kapampangan, Botolan Sambal, Remontado/Hatang-Kayi, and Masbatenyó.
- \*damúlag ‘carabao’: Kapampangan, Sambalic, Northern Bikol, and Southern Bikol.  
 \*páwəs ‘sweat’: Kapampangan *páwas*, Tagalog *páwis* (borrowed by Remontado/Hatang-Kayi and Casiguran Agta).  
 \*tiʔris ‘urine’: Northern Bikol, Tadyawan, and Pangasinan; Hanunoo ‘millipede secretion’.
- 5) **North Bisayan axis** [West Bisayan, Central Bisayan, Asi, Bikol, and Hanunoo]  
 \*bahól ‘big’: Aklanon, Kinaray-a, Looknon, Bulalakawnon, Datagnon, Kuyonon, and Romblo-manon.  
 \*bəʔól ‘take’: Aklanon, Kinaray-a, Bulalakawnon, Datagnon, Kuyonon, Asi, Romblo-manon, and Hanunoo.  
 \*hambal ‘say, speak’: Aklanon, Kinaray-a, Bulalakawnon, Caluyanan, Asi, Ilonggo/Hiligaynon, Romblo-manon, Masbatenyó, and Kagayanen.  
 \*indu ‘your (2PL.OBL)’: Caluyanan, Kuyonon, Northern Bikol, Romblo-manon, and Asi/Bantoanon.  
 \*isará ‘one’: Aklanon, Kinaray-a, Pandananon, Bulalakawnon, Caluyanan, Kuyonon, Hanunoo, and Kagayanen.  
 \*taʔú ‘give’: Aklanon, Kinaray-a, Pandananon, Bulalakawnon, Caluyanan, Romblo-manon, Asi, Northern Bikol, and Rinconada Bikol.
- 6) **Palawan-Kalamian axis** [Kalamianic and Palawanic]  
 \*alaŋ ‘buy’: Calamian Tagbanwa, Aborlan Tagbanwa, Batak (note also Palawan *ələn*).  
 \*bəlag ‘not so’ [NEG]: Agutaynen, Karamianen, Palawan, and Molbog; Aborlan Tagbanwa ‘different’.  
 \*kumba ‘lungs’: Calamian Tagbanwa, Karamianen, Central Tagbanwa, and Aborlan Tagbanwa.  
 \*luwak ‘plant, dibble (v.)’: Calamian Tagbanwa, Agutaynen, Karamianen, Batak, Aborlan Tagbanwa, and Palawan.  
 \*tagak ‘blood’: Calamian Tagbanwa, Karamianen, Batak, and Central Tagbanwa.
- 7) **Palawan-Mindoro axis** [North and South Mangyan, Kalamianic, and Palawanic]  
 \*[h]abuət ‘long’: Kalamianic, Palawanic, North Mangyan, and South Mangyan.  
 \*aŋbəʔ ‘rat’: Aborlan Tagbanwa, Batak, Hanunoo, Buhid, and West Bisayan.  
 \*bílug ‘body’: Northern Palawanic, North Mangyan, and South Mangyan.

ZORC, LOBEL, & HALL

- \*hamparj ‘say, speak’: Batak, Aborlan Tagbanwa, Palawan, and Hanunoo (cf. PBIS \*hamparj ‘play’).
- \*kawa ‘you (2SG.NOM)’: Kalamianic \*yawa? (< \*i-kawa), Central Tagbanwa and Tadyawan *kawa*.

- 8) **Eastern Mindanao axis** [South Bisayan, Mansakan, Mamanwa, Manobo, Danaw, and Subanen]
- \*allaj ‘slave’: Mansakan, Dibabawon, and Sarangani Manobo.
  - \*baʔal ‘make’: Western Bukidnon Manobo, Subanen, Maranao, and Maguindanaon.
  - \*dayaw ‘good’: South Bisayan, Mansakan, Mamanwa, Ata Manobo, and Dibabawon.
  - \*sidan ‘they (3PL.NOM)’ (with addition of final nasal to PMP \*sida): Mamanwa, Kamayo, Mansakan, and Subanen.

#### 7.4.3 Relationship of Bashiic/Batanic, Central Luzon, and North Mangyan

In his treatment of the Mangyan languages of Mindoro, Zorc (1977: 34) pointed out the division between the North Mangyan (Iraya, Alangan, and Tadyawan) and South Mangyan (Hanunoo, Buhid, Western and Eastern Tawbuwid, and Bangon) languages, and suggested the possibility of a ‘North Extension’ containing not only the North Mangyan languages but also Batanic/Bashiic and Central Luzon (i.e. Kapampangan and the Sambali-Ayta languages). This ‘North Extension’ was based on the merger of PAN \*R with \*y, not generally found elsewhere among Philippine languages, as well as a handful of putative lexical innovations, including \*dagul ‘big’, \*udi ‘left (side)’, and \*dimla? ‘cold’. Neither McFarland (1980) nor Blust (1991b, 2019b) accept the inclusion of Batanic/Bashiic in this grouping, but far more research and documentation is still needed for all of these languages. Note that the presence of forms reflecting \*R as /y/ instead of /g/ or /l/ in members of the South Mangyan, Palawan, and Kalamianic subgroups may turn out to be evidence that an ancient member of Zorc’s North Extension may have once been a prestige language in the area: for example, Calamian Tagbanwa, Karamianen, some Batak *ikuy* ‘tail’ (< \*ikuR); Agutaynen *ki-yuy* /kiʔyuy/ ‘egg’ (< \*qitəluR); Agutaynen *niyuy* ‘coconut’ (< \*niyuR); Batak, Aborlan Tagbanwa *punyanan* ‘parent-in-law’ (< \*tuRaŋ); and Kalamianic \*wai? ‘water’ (< \*wahiR).

#### 7.4.4 The position of Molbog and Bonggi

The languages of the Philippine province of Palawan have been demonstrated by a number of authors (e.g. Zorc 1977;

Thiessen 1981; Blust 1991b, 2010b; Himes 2007) to be split between a Kalamianic group and a Palawanic group, both belonging to the Philippine subfamily. Likewise, the majority of the languages of Sabah (as well as a handful of others spoken in Sarawak, Brunei, and the Indonesian province of Kalimantan Utara) are known to belong to two subgroups, Southwest Sabah and Idaanic/Northeast Sabah (Blust 1998b, 2010b; Lobel 2013a), neither of which immediately subgroups with the languages of the Philippines (Blust 1998b) in spite of their Philippine-type features. However, one controversy remains with regard to these two groups: the position of Molbog and Bonggi. Molbog is spoken primarily on the Philippine island of Balabac off the southern tip of Palawan, neighbouring minor islands, and a handful of communities near the southern tip of Palawan plus two perhaps century-old communities on Banggi Island off the northwestern tip of Sabah. Bonggi, on the other hand, is spoken on the aforementioned Malaysian island of Banggi plus Balamangan Island to its immediate west. Thiessen (1981) treats both languages as members of the Palawanic group, while Blust (2010b) argues that Bonggi subgroups with the Idaanic languages (Idaan, Begak, Sungai Seguliud, Subpan, and the elusive Buludupi) in a ‘Northeast Sabah’ subgroup within his Greater North Borneo grouping. Lobel (2013a), on the other hand, notes striking similarities (but admittedly not shared innovations) suggesting a closer connection between Molbog and Bonggi in a Molbog-Bonggi subgroup whose external relationships have yet to be determined. While Blust’s data linking Bonggi to the Idaanic languages appear to be quite strong, the similarities Lobel notes between Molbog and Bonggi still warrant further investigation: The Palawan-Sabah area is clearly at the border of Philippine and non-Philippine languages, but what remains to be understood is whether Molbog and Bonggi were once two closely related languages that came under mutually exclusive influences, one (Molbog) from languages to its north, the other (Bonggi) from the languages to its south. If not, then the similarities shared by these two languages, one each from the Philippine and Greater North Borneo macrogroups, are the result of a contact-induced convergence whose further investigation may shed light on the undocumented social history of this little-studied border zone.

#### 7.4.5 The position of the languages of Sabah

Much less a ‘controversy’ among scholars of Philippine and Philippine-type languages than simply an ‘issue’ warranting mention is the position of the languages of Sabah in light of their Philippine-type grammatical characteristics. In very few places outside of the geographical Philippines

and northern Borneo is the Philippine-type focus and case-marking system retained, with large pockets in Taiwan and northern Sulawesi (the latter belonging to the Philippine subfamily), and a small number of more distant holdouts in Madagascar (Malagasy), Java (Old Javanese), and Sumatra (Batak). However, as Blust (1998b, 2010b, 2013a) points out, evidence from the Formosan languages in Taiwan clearly and uncontroversially indicates that the ‘Philippine-type’ structure is in fact a retention from Proto-Austronesian (or, at the very least, a protolanguage forming a primary branch thereof), and the lack of shared phonological and lexical innovations indicates that none of the languages originally native to Borneo belong to the Philippine subfamily. This was not immediately apparent in the early years of Austronesian scholarship, prior to the emergence of larger amounts of data on Malagasy and the Formosan languages of Taiwan. Today, however, no ‘controversy’ remains in this regard, and subsequent authors (e.g. Lobel 2013a, 2016b; Smith 2017a) have accepted Blust’s separation of the languages of northern Borneo from the Philippine subfamily.

#### 7.4.6 Migration and historical levelling

Blust (1991b, 2005a) calls attention to the relatively low level of diversity found among modern Philippine languages in comparison to their presumed length of time in the Philippines after the departure of speakers of Proto-Malayo-Polynesian from what is now the country of Taiwan. To explain this, Blust proposes two periods of language levelling: the first in which speakers of Proto-Philippines expanded throughout the Philippines at the expense of speakers of other Malayo-Polynesian subgroups that were presumably present in the area at that time; and a second, during which speakers of Proto-Greater Central Philippines expanded throughout not only the central and southern Philippines but also northern Sulawesi, levelling non-GCPH languages that had previously been spoken in those areas.

A number of other levelling episodes could also be added to Blust’s list. Inete/Inati, the language of the aboriginal Ete/Ati of Panay, appears to be a primary branch of the Philippine subfamily (Pennoyer 1986–1987; Blust 1991b, 2019b), and was present in its current location prior to the expansion of speakers of Bisayan languages into the western Visayan Islands. On neighbouring Negros Island, the various Bukidnon peoples are clearly remnants of earlier Bisayan-speaking populations who fled into the mountains to maintain their freedom during the Spanish occupation (George Largado, pers. comm., 2006) and became minoritized by the massive influx of Cebuanos and Ilonggos in the mid-nineteenth century (Scott 1984, 1992). In Luzon,

many parts of the Pacific coast of central Luzon were inhabited almost exclusively by the Umiray Dumaget when Canadian missionaries Thomas and Pat Macleod arrived in the area in the 1960s (as evidenced by numerous reports and thousands of photographs archived by SIL-Philippines); half a century later, virtually no Dumaget communities can be found along the coast, with Tagalogs having now taken over the land that the Dumagets’ ancestors had lived on since time immemorial (Lubita Andrada, pers. comm; Salvador Cruz, pers. comm.). What has happened with the Dumagets over the past fifty years has occurred many times over the past 500 years or so, as many native populations, now largely confined to upland areas in Luzon, Palawan, and Mindanao, report that their ancestors once moved freely between coastal and interior areas prior to the arrival of the groups now living along the coast, who, based on linguistic evidence, are clearly much more recent arrivals.

In other cases, however, remnants of levelled populations still exist even in coastal areas: for example, early Spanish documents make no mention of Cebuano speakers on Leyte Island, and appear to suggest a somewhat wider distribution of Waray-Waray on that island. Today’s Baybayanon (Rubino 2005b) and Kinabalian languages (Lobel 2013a) appear to represent remnants of Warayan dialects that existed prior to the expansion of Cebuano along the western and southern coasts of Leyte Island. In northeastern Luzon, the still-coastal Agta have yet to be fully displaced from their ancestral waters, although the expansion of Tagalogs from the south and Ilokanos from the north are slowly minoritizing both them and the natives of the town centre of Casiguran (note that the Kasiguranin language spoken by the latter largely developed from a mix of Tagalog and Casiguran Agta).

#### 7.4.7 Aboriginal Filipinos

In addition to the ‘ethnic Austronesian’ population, an estimated 15,000 Aboriginal Filipinos<sup>20</sup> are also native to the Philippines, and the population of Filipinos with at least one aboriginal grandparent may be as large as 100,000. Most of these populations live on the fringes of modern Philippine society, and a number of authors have written about the violence, discrimination, and other abuse that these populations regularly suffer (see discussion in Lobel 2013a). Thomas Headland, Lawrence Reid, and Jason Lobel

<sup>20</sup> This number could be as high as 75,000, if we include the 10,000 Iraya Mangyans and 50,000 Ata, Tigwa, and Matigsalug Manobos of Mindanao, large portions of whose populations clearly have ‘Negrito’-like physical characteristics.

ZORC, LOBEL, & HALL

are among the most widely published scholars on the languages of aboriginal Filipinos, several of which form primary branches of the Philippine macrogroup (e.g. Inati, Umiray Dumaget, Manide, and Inagta Alabat) and preserve highly conservative features lost in all other Philippine languages. Headland and Blood (2002) and Lobel (2013a: 55–102) give a more extended overview of Aboriginal Filipino populations and their languages.

## 7.5 Conclusion

Although an impressive amount of historical-comparative work has been done on Philippine languages over the past century, much work remains, and is even more urgent due to the rate at which the numbers of fully fluent native speakers of many of these languages have been diminishing over the past several decades. For various reasons—including intermarriage, the rise of the internet and social media, and the expansion of electricity and mobile phone networks into rural areas—tens of millions of Philippine youth are no longer growing up competent in their parents' language(s). It is therefore imperative that as much documentation as possible be completed before these languages disappear, especially since an unfortunately large number of Philippine languages still lack even basic documentation, such as a sketch grammar, dictionary, or text collection, let alone a reference grammar (see Ewing and Kimoto, chapter 15, this volume).

In planning future fieldwork, scholars should strive to go far beyond the relatively short wordlists of the past (e.g. the Swadesh 100- and 200-item lists, and the SIL 372-item list) and collect not only longer wordlists (e.g. those developed by Zorc and Lobel, each with well over 1,500 items) and extensive sentence lists, but also recordings of spontaneous speech. In particular, collections of texts, ide-

ally accompanied by audio recordings, allow researchers to collect invaluable information about such topics as local history, customs and traditions, flora and fauna, recipes, biographies of important people, and origin myths. Similarly, historical-comparative work must move beyond simply reconstructing lexical items, and pay more attention to the reconstruction of morphology, grammar, and discourse. A considerable amount of morphology marking verbs, nouns, pronouns, deictics, numerals, and modifiers remains to be reconstructed, and grammatical paradigms particular to each language need to be mapped alongside those reconstructed for Proto-Philippines. Additional research is also needed on Philippine accent patterns, which may prove to be a highly significant innovation unique to the Philippine subfamily (Zorc 2020; Smith, chapter 2, this volume), absent from both Formosan languages and non-Philippine Malayo-Polynesian languages.

Finally, additional research will also hopefully allow us to determine the genetic position of certain languages like Umiray Dumaget, Inati, Manide, Inagta Alabat, Remontado/Hatang-Kayi, Molbog, and Bonggi. If it turns out that some or all of these languages are, in fact, primary branches of the Philippine subfamily, then this will also help revise our understanding of Proto-Philippines and even possibly of higher-level nodes in the Austronesian family tree.

## Acknowledgements

We wish to express our heartfelt gratitude to all of our native speaker friends and consultants who provided the data on the languages covered herein. We are likewise grateful for the invaluable feedback from Robert Blust, Lawrence Reid, and Hsiu-chuan Liao, as well as from the late John Lynch, to whose memory this chapter is dedicated.