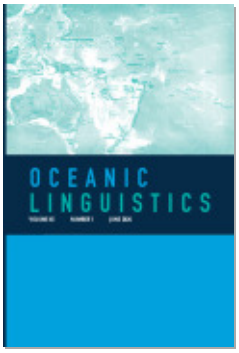
Contact or Inheritance? New Evidence on the Proto-Philippines Hypothesis



Victoria Chen, Maria Kristina Gallego, Jonathan Kuo, Isaac Stead, Benjamin van der Voorn

Oceanic Linguistics, Articles in Press, to appear in Oceanic Linguistics 65-1, (Article)

Published by University of Hawai'i Press
DOI: *https://doi.org/10.1353/ol.2026.a984857*

Contact or Inheritance? New Evidence on the Proto-Philippines Hypothesis

Victoria Chen,[†] Maria Kristina Gallego,[‡] Jonathan Kuo,[#] Isaac Stead,[§] and Benjamin van der Voorn[‖]

[†]VICTORIA UNIVERSITY OF WELLINGTON, [‡]UNIVERSITY OF THE PHILIPPINES DILIMAN, [#]NATIONAL TAIPEI UNIVERSITY OF TECHNOLOGY, [§]MAX PLANCK INSTITUTE FOR EVOLUTIONARY ANTHROPOLOGY, AND [‖]PAYAP UNIVERSITY

A central issue in Austronesian higher-order subgrouping concerns the position of the Philippine languages. Despite considerable debate, it remains unsettled whether these languages represent multiple intersecting Malayo-Polynesian primary branches forming a linkage, or whether they descend from a single shared ancestor distinct from Proto-Malayo-Polynesian. We present three new lines of evidence supporting the former view. First, the absence of a genuine *d/*z merger in Central Luzon and Minahasan languages undermines the sole phonological argument for a Proto-Philippine branch. Second, the geographical distribution of the purported Proto-Philippine-defining lexicon favors a diffusion-based explanation. Third, the distribution of an underexplored morphological innovation across Philippine languages suggests frequent contact among major island groups, excluding interior and peripheral areas. These findings not only highlight the absence of defining innovations for a cohesive Proto-Philippine but also point to extensive horizontal transmission across the archipelago, extending beyond lexical exchange. We further propose that the *d/*z merger is better understood as an areal drift rather than a diagnostic innovation. We conclude that Philippine languages are best analyzed as intersecting Malayo-Polynesian primary branches, in line with recent perspectives. Moreover, the high number of proposed lexical innovations attributed to Proto-Philippine likely reflects contact-driven processes—diffusion, borrowing, and linkage histories—rather than retention from a unified ancestor.

Keywords: Proto-Philippines; Austronesian Higher-Order Subgrouping; Lexical Innovation; Morphological Borrowing; Language Contact; Linkage

1. INTRODUCTION.[1] Although robust comparative evidence supports the view that all Austronesian languages spoken outside Taiwan descend from a

1. This project was supported by a Marsden Standard Grant (#VUW2309) from the Royal Society of New Zealand. We would like to thank Dan Kaufman, Malcolm Ross, Alex Smith, and Christina Truong; the audiences at SEALS 31 and ICAL 15; and the two anonymous *Oceanic Linguistics* reviewers for their helpful comments. All remaining errors are our own.

The authors are listed in alphabetical order and contributed equally to this paper. With deep respect and gratitude, we dedicate this work to Robert Blust, whose intellectual vision and generosity of spirit have left a lasting legacy.

single common ancestor (Dahl 1973; Mills 1975; Blust 1977, 1993; Ross 2002)—Proto-Malayo-Polynesian (PMP)—the first-order division of this branch remains contested in recent literature. A focal question concerns the linguistic position of the languages spoken in the Philippines, the first major landing site of the Austronesian diaspora. Whether these languages descend from a single offshoot of PMP or they make up multiple primary branches remains a point of contention in the literature (Reid 1982, 2020; Blust 2005, 2019 et seq.; Zorc 1986, 2019; Ross 2005, 2020; Liao 2020, among others).

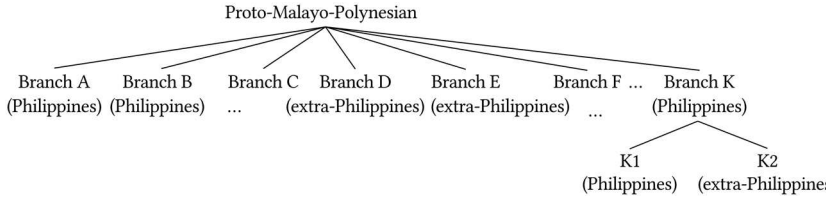
This debate highlights the limitations of the Neogrammarian approach in classifying closely related languages under mutual contact. Although there is a lack of comparative evidence that the Philippine languages subgroup with any other Malayo-Polynesian (MP) languages spoken elsewhere (except for the Gorontalo-Mongondow, Sangiric, and Minahasan languages of northern Sulawesi) (Blust 1995, 2005; Smith 2017), only a single phonological merger supports their potential common descent from a single ancestor distinct from PMP. It is therefore challenging to determine their affiliation using the traditional approach to linguistic subgrouping.

What makes Philippine languages particularly theoretically intriguing is the sheer volume of their purported lexical innovations on the one hand and the absence of phonological evidence for their affinity on the other—over 1,000 shared lexical items have been identified across these languages (albeit with varying patterns of distribution), none of which appear in MP languages outside the Philippines. Several researchers have thus argued for the existence of a common ancestor for all modern Philippine languages, Proto-Philippines (PPh), as a distinct primary branch of MP (Blust 2005, 2019, 2020, 2022; Zorc 1986, 2020). This affinity, as proposed in Blust (2005 et seq.), is the result of a historical leveling event that eliminated other PMP descendants in the Philippine archipelago. However, the lack of non-lexical evidence for this subgroup has led to the alternative view that the shared vocabulary is the outcome of lexical diffusion (Reid 2020; Ross 2020; Liao 2020; see also Smith 2017 for a similar view).

From a theoretical point of view, the challenge lies in identifying true cases of shared innovation among closely related languages spoken in communities with intense and long-standing social contact. Given Luzon's proximity to the Visayas and Mindanao island groups, one would expect a subgrouping scenario like figure 1, where PMP speakers gradually diversified into distinct linguistic communities through their expansion across the Greater Philippine region (alongside migrations out of the Philippine islands). In principle, these languages therefore represent several independent primary branches of PMP, parallel to ex-Philippine branches. It is also possible that some of the ex-Philippine branches derived from one of the branches that diversified within the Philippine islands, as exemplified by the hypothetical Branch K.

How solid is the evidence that the Philippine subgroups are subsumed under a common ancestor that excludes other MP languages? In this paper, we reappraise

FIGURE 1. HYPOTHETICAL SUBGROUPING SCENARIO FOR HIGHER-ORDER MP LANGUAGES.



previous arguments for and against the PPH hypothesis with new evidence concerning specific aspects of the phonology, lexicon, and morphosyntax of the Philippine languages. Three central questions (1a–c) are investigated.

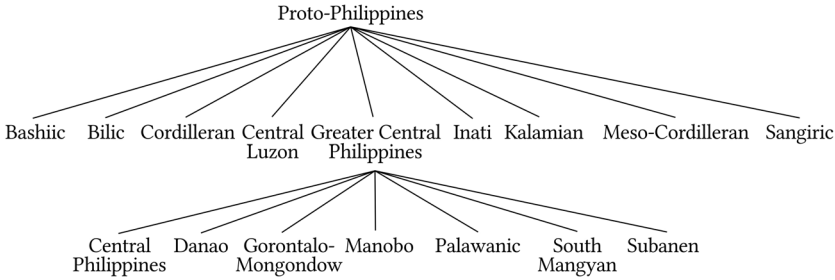
(1) The key questions for investigating the PPH hypothesis
   a. How robust is the purported phonological evidence supporting PPH?
   b. What (pre-)history can be reconstructed for the distribution of the proposed lexical innovations among the Philippine subgroups?
   c. Can the domain of morphosyntax tell us something new about the relationship of the Philippine languages?

This paper is structured as follows: we begin with a review of the PPH hypothesis in section 2. Section 3 reassesses the validity of PMP *d/*z merger as a defining innovation of PPH. Section 4 revisits the claimed lexical evidence for PPH, examining the geographical distribution of these innovations. Section 5 analyzes the derivations of Circumstantial Voice morphology in the Philippine languages and attempts to link it with the possible contact history of the region. Section 6 concludes with the claims that (i) the Philippine languages are best viewed as deriving from multiple PMP offshoots that interact as a linkage and (ii) the linguistic features present in modern Philippine languages reflect multiple layers of change, containing retentions from Proto-Austronesian (PAN)/PMP at the deepest level, overlapping innovations from a linkage history, and borrowed features from later contact among groups.

2. PROTO-PHILIPPINES: PUZZLES AND DEBATES. Since Blake (1906), linguistic similarities among Philippine languages have led many to view them as descending from a single protolanguage (Blake 1906; Scheerer 1918; Charles 1974; Llamzon 1975; Paz 1981; Blust 2005, 2019 et seq.; Himes 2012, a.o.). This proposed branch, illustrated in figure 2, includes several lower-order subgroups defined by various innovations (Blust 1991, 2005; Himes 2012; Lobel 2013; Reid 2018).

However, it remains unclear whether these subgroups share an ancestor distinct from PMP. Reid (1978, 1982) was the first to highlight the absence of exclusively shared phonological and morphosyntactic innovations for a

FIGURE 2. INTERRELATIONSHIPS OF PHILIPPINE LANGUAGES FOLLOWING BLUST (1991).



putative PPH branch, suggesting that these languages may not descend from a single source.[2]

In response, Zorc (1986) and Blust (2005) presented 327 lexical innovations as evidence supporting a single common ancestry post-PMP. However, Smith (2017:472) challenged this view further, citing the low quality of these proposed lexical innovations. Blust (2019) revisited the debate by expanding the list to 1,259 lexical innovations and one phonological innovation—the merger of PMP *z and *d. While Zorc (2020) concurs with Blust, Liao (2020), Reid (2020), and Ross (2020) argue that the overlapping distribution of these proposed innovations points to a linkage history rather than a neatly bifurcating family.

Ross (2020:370) notes that nearly all PPH-defining lexical innovations are not uniformly present in all Philippine languages, which is typical of a linkage history. Similar patterning may also arise from other scenarios, such as varying degrees of contact through trading along the coastlines (Zorc and Almarines 2022), further complicating the reconstruction of Philippine linguistic history. Recent Bayesian phylogenetic subgroupings (Gray, Drummond, and Greenhill 2009; King et al. 2024) further support a clustering of Philippine languages without strong evidence for PPH, reinforcing the linkage hypothesis.

This implies a rapid expansion of PMP speakers into island Southeast Asia, leaving behind an early dialect network in the Philippines. The swift population expansion precluded the development of clear innovation-defined subgroups, resulting instead in overlapping innovation patterns (e.g., François 2014:170–71; Ross 1995:45–46).

While Blust acknowledges the presence of linkage-like distribution of innovations (2019:183–84, 2020:453–54), the core disagreement in interpretation remains: Proponents of the linkage history (e.g., Ross 2020) view PMP as the immediate ancestor of an early Philippine dialect network that rapidly differentiated after other MP speakers exited the region, whereas Blust (2020:452)

2. Ross (2005:13) makes a similar argument that the phonological and grammatical (i.e., verbal) systems of PPH show little innovation and remain largely identical to PMP. However, note that this view is somewhat circular, as it is the grammatical systems of Philippine languages that play an important role in the reconstruction of PMP morphosyntax.

argues for a distinct, albeit internally diverse PPH, noting that its innovations are confined to Philippine languages and absent in those spoken south of the archipelago.

In what follows, we revisit the debate with three new lines of evidence: phonology, lexicon, and morphosyntax. In section 3, we highlight the previously overlooked absence of the *d/*z merger in several Philippine microgroups. We then examine the geographical distribution of PPH-defining lexical items, showing that they likely reflect contact rather than inheritance. Finally, we investigate an understudied variation in Circumstantial Voice morphology, presenting new evidence for multiple layers of borrowing across Philippine subgroups.

Ultimately, the goal of this paper is to evaluate the significance of shared lexical innovations for subgrouping, particularly in the absence of phonological evidence. This is crucial for Austronesian diachronic studies, as many primary branches of MP lack robust phonological support. Establishing clear criteria for using lexical or syntactic innovations in subgrouping is therefore essential.

3. IS THE *d/*z MERGER A DEFINING INNOVATION FOR PROTO-PHILIPPINES? A REAPPRAISAL.

We begin by reexamining the sole phonological innovation for the purported Philippine branch: the merger of PMP *d (e.g., ***d**uSa 'two') and *z (e.g., ***z**alan 'road'). Blust (2019) notes that no language in the Philippines maintains this distinction in stable forms, citing only one exception to this generalization: in Ayta Abellen (a Central Luzon language), *d and *z remain unmerged in two cases: *qu**d**ang > *u**l**ang* 'shrimp' versus *qu**z**an > *u**d**an* 'rain'. However, he also notes that this instance is ambiguous, because "other words [in Ayta Abellen] show a merger of *z and *d word-initially, as in PAN ***d**epa > ***d***epah* 'armsbreadth' and ***z**aRami > ***d***ayami* 'rice husk'" (Blust 2019:156–57).

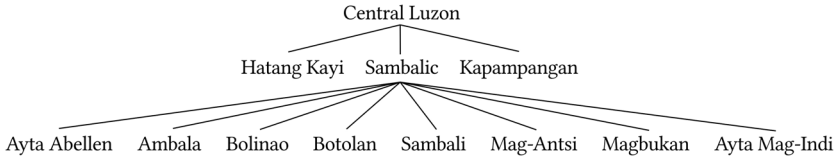
Our investigation into the Austronesian Comparative Dictionary (ACD) and recent literature reveals many more such exceptions not only in Central Luzon but also in the Minahasan languages. In what follows, we discuss how these cases constitute strong evidence for *d and *z remaining unmerged at the proto-level of these subgroups. We then revisit the absence of this merger in Chamorro—the so-called western Austronesian outlier that shares similar phonological innovations with Greater Central Philippine (GCP) languages—and discuss how the presence of a *d/*z distinction in the languages examined informs the linguistic prehistory of the Philippines.

3.1. UNMERGED *d AND *z IN THE CENTRAL LUZON LANGUAGES.

There are three lower-order branches within Central Luzon: Kapampangan, Sambalic, and Hatang Kayi, as seen in figure 3.[3] Out of those subgroups, there

3. Also known by the exonyms Sinauna and Remontado. Some basic information on this underdocumented language is found in Lobel and Vertudez Surbano (2019). Sambali is sometimes referred to by the pejorative exonym Tina.

FIGURE 3. CENTRAL LUZON SUBGROUPING (HIMES 2012; STONE 2008).



are four languages for which a useful amount of data are readily available: Kapampangan (KP), Ayta Abellen (AA), Ayta Mag-Antsi (AM), and Botolan (BT).[4] The interrelationship of these languages is illustrated in figure 3. The latter three languages are part of the Sambalic subgroup. Due to a lack of ready data, the Hatang Kayi branch is not represented in this paper.

Our examination of the ACD identifies unmerged *d and *z not only in the two Ayta Abellen examples mentioned in Blust (2019), but also in other Central Luzon languages.[5] Importantly, these exceptions are not sporadic or random but reveal a pattern. The key observations are summarized in (2).

(2) a. *z is realized as /d/ word-initially in all four languages. Intervocalically, it is realized as /d/ in Ayta Abellen and /r/ in the other three languages.
b. In many reflexes, across all four languages, *d is indeed realized identically to *z, as either /d/ or /r/ respectively.
c. Contra Blust's generalization, there are also many instances of exceptional *d > *l* found in stable forms. Examples from the four languages under discussion are PAN *daNum > AA *lanom* 'water', PAN *duSa > BT *lua* 'two', PMP *dateng > AA *lateng* 'to arrive; come', and PMP *badas > AM *balah* 'sand'.
d. The (unexpected) *d > *l* sound change occurs intervocalically in Kapampangan. In the Sambalic languages, it occurs more freely: intervocalically and word-finally in Ayta Mag-Antsi, and in all environments in Ayta Abellen and Botolan.

Table 1 summarizes the data. The third column lists regular sound change rules which affect both *z and *d in each language; the fourth column outlines instances where *d diverges and is realized as /l/. In the appendix, we present a hypothesis that the sound change *d > *l* / V_V occurred in Kapampangan and

4. More data would be valuable, particularly more instances of *z. A full dictionary of Ayta Mag-Antsi is already available at https://philippines.sil.org/resources/online_resources/sgb, produced by SIL. However, its contents have not yet been sorted and indexed to their protoforms.
5. The absence of a true *d/*z merger in Ayta Mag-Antsi has been briefly discussed in Liao (2020) as well as in Himes (2012). Smith (2017:471) also discussed a few cases in other languages: "multiple languages have different reflexes of *z and *d: for example, Itbayaten *doha* 'two' from *duha but *rahan* 'road, path' from *zalan, and Kakilingan Sambal *olang* 'shrimp' from *qudang but *odan* 'rain' from *quzan." Himes (2012) also independently concludes that *z and *d remain unmerged in Proto-Central Luzon, although not connecting the discussion to the PPH debate (see the appendix for more details). Our attempt here is to take a more comprehensive view at the absence of this merger within Central Luzon and across other Philippine subgroups.

TABLE 1. REFLEXES OF *d AND *z IN CENTRAL LUZON LANGUAGES.

Language	Total *d/*z	Identical *d/*z	Instances	*d realized as *l*
Kapampangan	64	*d > *r* /V_V	4	PMP *badas > *balás* 'sand'
		*d > *d*	46	PMP *ida > *ila* '3PL'
		*z > *r* /V_V	5	PWMP *ludem > *ma-lúlam* 'cloudy, about to rain'
		*z > *d*	5	PAN *tuduq > *tulu(P)* 'to drip, flow, spill'
Ayta Abellen	47	*d > *d*	30	PAN *daNum > *lanom* 'water'
		*z > *d*	5	PMP *dateng > *lateng* 'come'
				PWMP *sidem (or *silem) > *hilem* 'afternoon'
				PAN *dengeR > *lenge* 'hear'
				PMP *dingding > *lingling* 'wall of a house'
				PWMP *di hipaR > *lipay* 'the other side of a body of water'
				PAN *duSa > *lowa* 'two'
				PAN *qañud > *anol* 'carried on the current'
				PPH *sápad > *sapal* 'hand of bananas'
				PPH *tadék > *talek* 'dance'
				PAN *tuduR > *toloy* 'sleep'
				PWMP *pu(n)dut (or *pu(n)zut) > *polot* 'pick up'
Ayta MagAntsi	9	*d > *r*	1	PMP *badas > *balah* 'sand'
		*d > *d*	1	PMP *qalad > *alal* 'fence'
		*z > *r*	3	PAN *qudang > *ulang* 'squid'
				PWMP *tadu > *talo* 'beeswax'
Botolan	33	*d > *r* /V_V	4	PPh *agud > ágol 'moan, agonize'
		*d > *d*	17	PAN *daNum > *lánum* 'water'
		*z > *r* /V_V	2	PAN *duSa > *lua* 'two'
				PAN *SateD > *atel* 'escort'
				PPh *tadék > *talék* 'dance'
				PWMP *tadu > *tálo* 'beeswax'
				PMP *tidaq > *tílaP* 'remainder'
				PMP *tudung > *tolóng* 'head cover'
				PAN *tuduR > *túluy* 'tsleep'
				PAN *tuduS > *tóPol* 'knee' (metathesis)

the Sambalic languages while *d and *z were still unmerged, and the flow-on effects from this sound change resulted in the split reflexes of *d.[6]

At first glance, the realizations of *d and *z in the sampled Central Luzon languages seem mixed and unpredictable, but a closer look reveals that *d and *z are realized following their own distinct patterns. First, *z is always realized *consistently*, appearing as /d/ in one environment and /r/ in another (see table 2).[7] Second, *d is realized *inconsistently*—in an identical environment, it can either merge with *z, appearing as /d/ or /r/, respectively, or can appear as /l/. Now, if *d and *z were merged at the level of Proto-Central Luzon, we would expect to see the same pattern of realization in their reflexes throughout

6. All cited data are reproduced from the ACD entries, so some words will be listed as descending from PPH, despite the fact that this paper is arguing against its existence. Words listed under PWMP in the ACD are also listed as such, despite PWMP also being a highly questionable subgroup (cf. Smith 2017).

7. PWMP *puzut > AA *polot* seems to be an exception. But *puzut has a disjunct form, *pu(n)dut, both meaning 'to pick up with the fingers'. If *polot* is a reflex of *pu(n)dut there is no inconsistency.

all the Proto-Central Luzon descendant languages. However, that is not what the data suggest. In all four Central Luzon languages for which sufficient data are available, the reflexes of *d show an inconsistent pattern of realization, while the reflexes of *z show a consistent pattern. The conclusion is that *d and *z are in fact unmerged in these languages, strongly suggesting that PMP *z and *d were retained as separate phonemes in Proto-Central Luzon.

Two objections may be made against the data presented: first, that the inconsistency in the realization of *d is a result of loans from other Philippine languages; second, that the inconsistency is a result of sporadic change. The first of these objections is the stronger of the two, as the Central Luzon languages, particularly Kapampangan, have been influenced by surrounding languages—for example, Tagalog. Due to the overall phonological similarity of Philippine languages, it may also be difficult to identify early loans borrowed prior to the application of certain sound changes.[8] However, both objections fail in the face of the data on *z. If *z and *d were merged in Kapampangan, and the inconsistency in the reflexes of *d is solely a result of borrowing and/or sporadic change, probability demands we ought to see a comparable level of inconsistency in the realization of *z. In other words, if *z and *d had already merged into a united phoneme in Proto-Central Luzon, then we would expect words with *z or *d to be replaced by borrowing and/or sporadic change at roughly equal rates. However, all the reflexes of *z across the four languages studied are realized according to a regular rule, while 23% of the reflexes of *d (71% if only counting vowel-medial reflexes) appear as the divergent reflex /l/.[9]

It is clear that this difference in reflex patterning is not the outcome of borrowing or sporadic change but the result of an underlying separation between *z and *d in Proto-Central Luzon. This strongly undermines the viability of the *d/*z merger as a defining sound change for PPh.[10]

3.2. UNMERGED *d AND *z IN THE MINAHASAN LANGUAGES.

The Minahasan languages provide further evidence against the view that *d and *z are undistinguished across Philippine languages. Sneddon (1978, 1989) was the first to point out unmerged reflexes of PMP *d and *z in the Minahasan languages. Mead (2020) further investigates Sneddon's findings and identifies strong evidence that PMP *d and *z were originally unmerged in Proto-Minahasan (PMin)—the reflex of PMP *z is PMin *d in all environments, showing fourteen examples and no counterexamples. The reflexes of PMP *d, on the other hand, are split between PMin *d and *r, with the protophoneme being realized as PMin *d in thirty-eight of fifty-two instances

8. Though in other cases, it is possible to determine whether a word is borrowed or native—when it contains a phoneme which has undergone a unique sound change in a given language, for example, PMP *R > Ayta Abellen <y>.
9. There are twenty total instances of *z in the data. There are 131 instances of *d, 30 of which appear as /l/.
10. See the appendix for a detailed discussion of how the unmerged *d and *z in Central Luzon may be explained in the broader diachronic view of the linguistic prehistory of the Philippines.

(or *r in fourteen out of fifty-two instances). In the intervocalic environment, this split is particularly evident: PMP *d is realized as *r in seven of ten instances, the three exceptions being PAN *sdu 'hiccup' > *səduʔ* 'hiccup', PAN *ludaq 'saliva' > *ludaʔ* 'spit', and PMP *tudaq 'throw' > *tudaʔ* 'stab'.

Based on this high proportion of *r reflexes, Mead posits a systematic change rather than borrowing or sporadic change: PMP *d > PMIN *r intervocalically, and PMP *d > PMIN *d everywhere else, as also suggested in Sneddon's early work.

PMP *d and *z, therefore, have distinct reflexes intervocalically in PMIN but have merged in other environments. As there are no examples of PMP *z > PMIN *r, this merger of *d and *z must have occurred after the split of *d, so that *z and *d were still distinct phonemes at the time *d split.[11] With PMIN also showing evidence for unmerged *d and *z, the argument against a PPH *d/*z merger becomes stronger. That is, neither the Central Luzon nor the Minahasan subgroup aligns with the proposed sound change.

3.3. UNMERGED *d AND *z: IMPLICATIONS FOR LINGUISTIC PREHISTORY. Crucially, many of the languages discussed in this section are spoken on the periphery. The Minahasan languages are spoken in Sulawesi, located outside of the Philippine mainland. Ayta Abellen and Ayta Mag-Antsi are spoken by Negrito groups living in the highlands, and so are both historically separated from the lowland Philippine majority, both culturally and physically (in the case of Negrito languages, the archaeological evidence suggests that contact between the Negritos and Austronesian speakers happened from an early date (Reid 1994). This is consistent with the idea that *d/*z contrast was maintained in at least some early descendants of PMP in the Philippines, and its widespread loss is a later development, as suggested in Reid 1987). In being peripheral, both the Minahasan and Negrito languages are less likely to be subject to diffusionary forces through the dialect network and so are more likely to retain features which diffusion has obliterated from the rest of the Philippine languages. If the *d/*z merger has spread through diffusion, we have an explanation as to why *d and *z are merged throughout most of the Philippines while remaining incompletely merged in only these particular languages. On the other hand, no other peripheral language we have investigated has so far shown any split in *d and *z, except for Chamorro, which Blust does not include in his proposed PPH (Blust 2005:40).[12]

What we can see from the discussion so far is that the merger of *d and *z may better be understood as a common areal feature that can be observed in the

11. It is interesting to note the parallels between the situations of Proto-Central Luzon and PMIN with regard to *d and *z. In both cases, there is both a split in *d, significantly more prominent in the intervocalic environment, and a merger of one of the reflexes of *d with *z. It is possible that both languages followed similar pathways in the chronological change of these phonemes.

12. Another Negrito language, Inati, as presented in Pennoyer (1987), possibly has unmerged *d/*z, for example, PAN *huzan > *odʒen*, but the data are not clear enough to come to any conclusions. Thanks to the anonymous reviewer for bringing this language to our attention.

majority of the MP branches spoken in the Philippines. This observation aligns well with the parallel discussions in the literature on the western Austronesian languages (Blust 1999; Liao 2020). Therefore, the merger is not a particularly convincing innovation for the purposes of subgrouping. Table 2 summarizes attested mergers in various western Austronesian languages. The wide geographical and subgrouping representation suggests that this merger is not an innovation that can reliably support specific subgroups.

As table 2 shows, the Formosan languages that have undergone this merger span five of the ten first-order subgroups of the Austronesian family. This further indicates that the *d/*z merger is a widespread and independent development rather than a single innovation shared among the subgroups.

In terms of the Central Luzon subgroup, specifically, it appears increasingly phonologically distinct among the other Philippine languages, not only in preserving the distinction between *d and *z but also between *n and ñ. This is demonstrated in languages such as Kapampangan (e.g., PMP **buñi* > *buñi* 'celebrated, acclaimed') and perhaps Hatang Kayi (e.g., PMP **ñamuk* > *yamuk* 'mosquito'). Proponents of a unitary PPH might raise the possibility of a subgroup defined by the *d/*z and *n/*ñ mergers that would include all Philippine languages except the Central Luzon subgroup. However, this scenario is highly improbable based on two crucial points. First, the two phonological mergers are arguably weak in terms of quality due to their frequency of occurrence and are thus insufficient to define the other members of the proposed Philippine family in opposition to the Central Luzon subgroup. Second, the Central Luzon languages share some of the proposed lexical innovations supporting PPH and are not otherwise lexically deviant from other Philippine languages.

Finally, the absence of the *d/*z merger in Chamorro is noteworthy. There has been a proposal in the literature that the speakers of Chamorro originated in the Philippines. The main evidence, as discussed in Blust (2001), lies in the

TABLE 2. SAMPLE LIST OF WESTERN AUSTRONESIAN LANGUAGES MANIFESTING *d/*z MERGER OUTSIDE THE PHILIPPINES (DATA PARTIALLY FROM THE ACD, PARTIALLY FROM LIAO 2020).

	Name of branch	Affiliation	Geographic region
1	Atayalic	Atayalic	Taiwan
2	Basay	East Formosan	Taiwan
3	Puyuma	Puyuma	Taiwan
4	Papora	Western Plains	Taiwan
5	Saaroa	Tsouic	Taiwan
6	Kelabit	North Sarawakan, Malayo-Polynesian	Borneo
7	Kiput	North Sarawakan, Malayo-Polynesian	Borneo
8	Toba Batak	Sumatran, Malayo-Polynesian	Sumatra
9	Rejang	Western Indonesian (?), Malayo-Polynesian	Sumatra
10	Balaesang	Celebic, Malayo-Polynesian	Sulawesi
11	Javanese	Western Indonesian, Malayo-Polynesian	Java
12	Tetun	CEMP, Malayo-Polynesian	Lesser Sunda
13	Rotinese	CEMP, Malayo-Polynesian	Lesser Sunda

presence of a regular reflex, *pakyo*, of the PAN word for 'typhoon,' *baRius, which suggests that the speakers migrated to the Marianas from within the typhoon belt. This hypothesis aligns with the fact that Chamorro shares the *R/*g merger with GCP languages (Blust 1991). Assuming that the Philippines was the homeland of Chamorro speakers and that Chamorro was part of the GCP subgroup (setting aside the fact that Chamorro is rarely classified as a Philippine language despite its potential affinity with the Philippine languages), the absence of the *d/*z merger in Chamorro suggests that the *R/*g merger occurred after the departure of Chamorro from the Philippines.

To conclude, the absence of the *d/*z merger in at least two of the major Philippine microgroups (Central Luzon and Minahasan) challenges its status as a defining phonological innovation for PPH. Moreover, the fact that some, but not all, Central Luzon languages exhibit this merger suggests that it may represent a recurrent drift—or potentially the result of diffusion—common among Philippine languages. The key implication of this finding is that the proposed PPH branch lacks a definitive phonological basis. Consequently, this branch remains defined solely by lexical innovations, the validity of which is explored in the following section.

4. REVISITING THE LEXICAL EVIDENCE FOR A PHILIPPINE BRANCH. We turn now to a reconsideration of the lexical evidence for PPH. A key part of this evidence is the list of 1,259 lexical items with no external cognates outside the Philippines (Blust 2019, 2020). It is claimed that (i) these lexemes are present in high quantity, so as to imply descent from a single protolanguage, and (ii) that a core of them (thirty-seven items) represents the strongest type of lexical evidence for common descent, that is, replacement innovations of PMP equivalents. Therefore, Blust (2019) proposes that the languages that share these innovations must be descended from a distinct common ancestor, that is, PPH, which spread throughout the archipelago (and, as Blust proposed, replaced the descendant languages of PMP, which had been spoken there since the initial MP expansion). Much robust discussion has already taken place on the lexical evidence for PPH, leading to the alternative proposal that these lexical innovations may be the outcome of lexical diffusion (see, e.g., Smith 2017; Liao 2020; Reid 2020; Ross 2020; Zorc 2020). We summarize below the key arguments against the lexical data as evidence for PPH and expand on these together with some new findings and remarks.

The geographical distribution of the proposed PPH cognate sets does not support PPH as a protolanguage. In order to test this, we conducted a basic search for geographically defined clusters of cognate sets supposedly reconstructed to PPH as follows: first, the full list of PPH protoforms and their reflexes was retrieved from https://acd.clld.org/; second, language location data for each language containing a reflex of each set was attached, sourced from Glottolog. Next, two geographical regions were defined (somewhat arbitrarily),

northern Luzon and northern Sulawesi, as potential geographically defined spheres of interaction (Zorc and Almarines 2022).

Northern Luzon was defined geographically as a polygon extending northward from a line beginning roughly at latitude 15′5″ N. Ninety of the cognate sets argued by Blust (2019) to be reconstructible to PPH are found only within this region. Furthermore, many of these cognate sets have reflexes in Ilocano and then in some languages with fewer speakers which have had a historical relationship with Ilocano. Even given that the ACD data are probably not exhaustive, it seems improbable that this many cognate sets of the total could have been retained only in this geographical area. We suggest that for such cases, it cannot be ruled out that these cognate sets are instead the result of overlapping innovations at the level of the PMP linkage or in the 3000 years since the divergence of the Philippine languages began, rather than common inheritance, especially in the absence of phonological evidence.

For example, several lexemes show a distributional pattern like that for the proposed PPH protoform *labas 'pass by, overlook something when searching; to pass by (of time)'. Consider map 1, where black markers represent the reflexes of the protoform. The original data from ACD are presented in table 3.

In this type of distributional pattern, the reflexes of a proposed PPH lexeme are shared by languages spoken in geographically contiguous areas. For *labas, its reflexes are shared across three distinct subgroups, namely Bashiic, Northern Luzon, and Central Luzon. No other languages to the south of these subgroups share reflexes of this lexeme. The region is also characterized by its high degree of multilingualism and contact.[13] It is likely through such mechanisms that linguistic features spread through this part of the Philippine linkage.[14]

Such a pattern of lexical distribution also involves a large language, typically a regional lingua franca such as the present-day status of Ilocano in northern Luzon as described in Rubino (2000), spoken alongside smaller languages. These bigger languages are typically part of the language inventory of bi- or multilingual speakers of smaller groups in the region, suggesting a sociological gradient through which loanwords spread from this bigger source language into the smaller languages.[15] Loanwords can spread independently across different languages through trade, as observed in the lexeme *angang 'large earthenware jar' shared between the Bashiic languages Itbayaten and Ibatan and the Northern Luzon languages Isneg and Ifugaw. Blust (2019) treats the form as a PPH innovation. However, as it pertains to material culture, which can easily be transferred across geographically contiguous groups, the similarity across

13. Maree (2007) and Gallego (2020) have discussed, for example, how Ibatan, a Bashiic language, shows contact-induced features driven by Ilocano.
14. Alternatively, such patterning seen in the northern Philippine subgroups may actually represent PMP retentions that have been lost elsewhere, as PMP must have developed in the area. These multiple explanations demonstrate the complexities of teasing apart linguistic features brought by retention, diffusion across the linkage, and later contact.
15. Other unidentified lingua francas have surely existed in the past, and the extent to which they influenced other languages remains an open question.

MAP 1. REFLEXES OF PROTOFORM *labas.

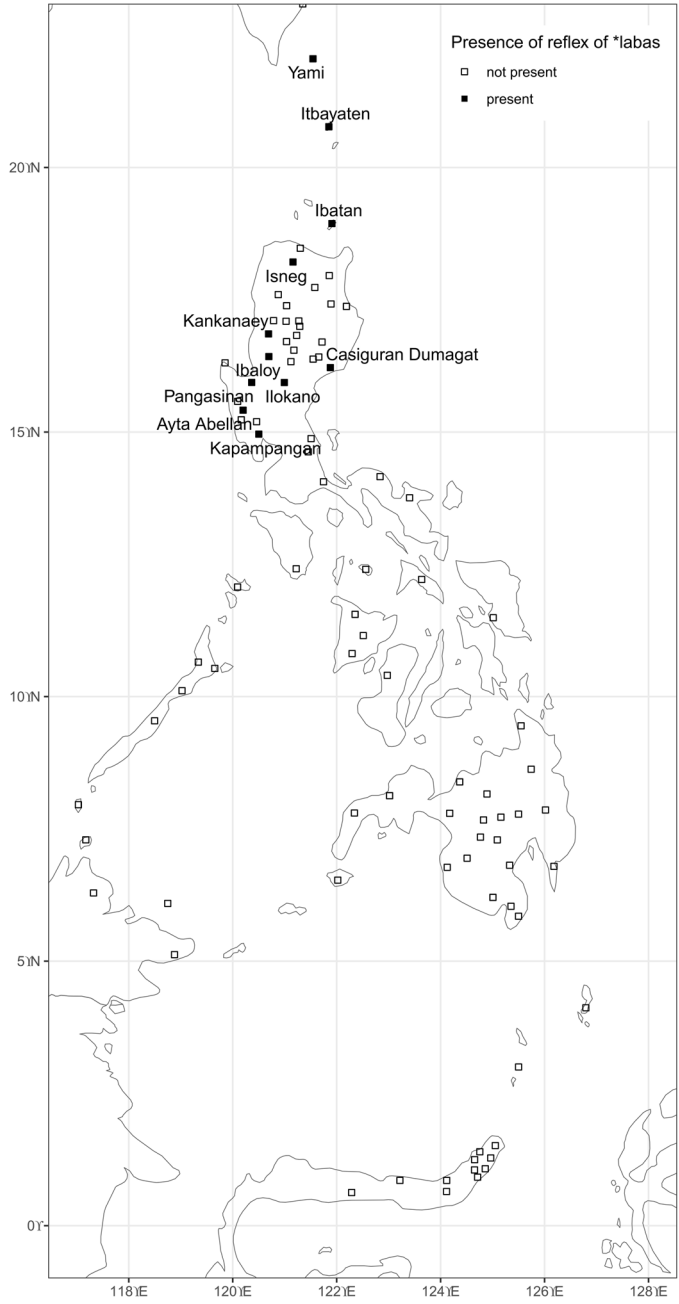


TABLE 3. REFLEXES OF *labas REPORTED IN THE ACD.

	Reflex	Gloss	Language	Approximate number of speakers
1	labah	to pass by	Ayta Abellen	3,500
2	lábis	excessive, more than enough (as a pole that is longer than it needs to be); excess	Casiguran Dumagat	610
3	debas-en	to make, take, etc. something too far (as house dimension beyond the specifications, bananas beyond proper ripeness)	Ibaloy	116,000
4	h<om>abas	to pass by (object, day, or time)	Ibatan	33,000
5	ag-pa-lábas	to let pass; tolerate; be understanding	Ilocano	8,100,000
6	na-lábas	past	Isneg	40,000
7	pa-xavas-en	to let pass	Itbayaten	3,500
8	na-labás	gone; gone away; passed; passed on	Kankanaey	240,000
9	labas	pass by, pass through (in the process of leaving, going out)	Kapampangan	2,800,000
10	on-labás	to go beyond, pass through; surplus, excess above requirements	Pangasinan	1,800,000
11	ni-mi-avas	passed by	Yami	4,000

the four languages is likely the result of borrowing rather than shared inheritance.[16]

Furthermore, in situations of intense social contact, we can also observe that "less borrowable" categories such as basic vocabulary and closed word classes can also be affected by lexical replacement (Thomason and Kaufman 1988; Haspelmath and Tadmor 2009). For example, the form *bunguq 'skull' is shared between the Central Luzon languages Sambal (Tina), Ayta Abellen, and Sambal, and the Central Philippine languages Tagalog, Aklanon, and Bikol. The influence of Tagalog in the region is well documented, and the languages in contact with it exhibit contact-induced features even in domains that are not typically affected by contact. Such phenomenon will be discussed further in the following section.

Before we continue, a definition of the types and relative strengths of lexical evidence for subgrouping will be useful. Smith (2023) divides lexical innovations into three types: exclusively shared replacement innovations; novel concept innovations; and homonymic innovations. In line with previous authors, Smith (2023:334) considers exclusively shared replacement innovations to be the strongest type of lexical evidence for subgrouping. This is because they involve the replacement of a well-supported reconstruction at an earlier level of subgrouping by a new lexeme in a well-defined group of succeeding languages, and as such the most parsimonious interpretation of this fact is that they represent an innovation in a protolanguage which was inherited by the successor languages. Smith further observes that novel concept innovations, which appear when a migrating

16. Also note that the four languages share exactly similar forms, and the opaqueness of phonology makes it virtually impossible to distinguish between shared innovation and borrowing.

group enters a new area, may be widely borrowed from substrate populations and so cannot be relied on for subgrouping. Homonymic innovations, where a homonymic term appears and persists alongside the original term, are even less interpretable in terms of inheritance or directionality.

Consider the case of *lánut, for which Blust's reconstructed PPH meaning is "abaca fiber, fiber-yielding plant, tree with a bark which yields a vine-like fiber." All of the attested reflexes of this form are present only within the Philippine archipelago, with none in the GCP languages of northern Sulawesi such as Tondano and Gorontalo, whereas reflexes of this term do in fact appear in the GCP languages Central Subanon, Western Subanon, Aklanon, Mansakan, and Bisayan. Furthermore, this is an example of a novel concept innovation, which would be weak evidence for subgrouping even if it were more widely distributed. In fact, a large number of the etyma proposed to be reconstructible to PPH share this pattern of attestation: they are present in the Philippines proper but not in the Philippine languages of northern Sulawesi. Notably, only seven proposed PPH etyma are found in these languages: *usauR 'to go downstream', *láyug 'tall, of trees', *liqed 'footprints, tracks', *iqit 'scraped or filed down', *habél 'weave cloth', *buál 'uproot a tree', and *butí 'smallpox; swine disease'.[17]

If it was indeed the case that these etyma were present in PPH, we are forced to assume that all reflexes of these cognate sets were lost or replaced in the GCP languages of northern Sulawesi. A more parsimonious explanation is that these etyma represent loanwords, which spread via the Philippine maritime interaction sphere but did not reach the GCP languages in northern Sulawesi further across the ocean. As for the seven PPH etyma which do appear in northern Sulawesi, these could either represent forms which spread throughout the early PMP dialect network or retentions from PMP for which there are no longer any Philippine-external witnesses. We consider this to be counterevidence for the PPH hypothesis and evidence for an interaction sphere postdating the departure of the GCP North Sulawesi languages for their current locations. We will discuss below an example of probable morphosyntactic borrowing which appears to show the same distribution, further reinforcing this interpretation.

Several authors (Smith 2017; Ross 2020) have already mentioned short- and long-distance trading links as a vector for the diffusion of loanwords through the archipelago. Smith (2017:464), in his review of the strength of PPH etyma, points to the word for Manila hemp (*Musa textilis*), an important resource for the production of fiber for rope and weaving, as being widely distributed (in the Batanes, Northern Luzon, Central Luzon, and GCP subgroups) and as fitting Blust's criteria for a strong witness to PPH. However, Smith suggests that since Manila hemp is a widely traded item, this may be a loan. Blust (2019:215) counterargues that since Manila hemp is native to the Philippines, PPH speakers presumably had a word for this item, and as such it cannot be a loanword. This

17. However, a closer examination of the Dutch dictionaries of Sulawesi may add to this list. We set this aside for future investigation.

issue is also highlighted in Liao (2020). In any case, this is another example of a novel concept innovation, and as such has less value for subgrouping.

This one example is representative of the debate over the lexical evidence for PPH as a whole. Even if speakers of some ancestral Philippine protolanguage had a term for Manila hemp which was retained in its daughter languages, in the absence of diagnostic reflexes[18] and the apparent phonological conservatism of the Philippine languages, it is in principle not possible to distinguish whether this language was PPH or whether these lexemes arose in and spread through the PMP dialect network.

In essence, we are dealing with the difficult challenge of distinguishing loanwords from common inheritance, and some mention of this is already made in each of the papers cited above. We argue that it is highly likely that a dialect linkage developed during the initial MP expansion, and even widely distributed "PPH" etyma may in fact be items which spread through this dialect network early on.

This ties into the second aspect of our argument, which highlights that dismissing the likelihood of diffusion via trade so readily may not be a prudent approach. Reid (2018) argued along these lines, but Blust (2020:185) countered that lexical borrowing in trade should favor the types of items which are exchanged in commercial transactions, and that the majority of items in the 1,259-item list are not of this type. We believe this point is unpersuasive, as the term "trade" is not necessarily limited to commercial transactions. Long-distance trade-type relationships which do not involve narrowly commercial exchange are attested for in many linguistic areas, such as the Kula ring in western Melanesia (Malinowski 1922). Long-distance interaction of other types, such as intermarriage and multilingualism, is also an important factor in loanword diffusion and may give rise to borrowings across wider domains (Haspelmath and Tadmor 2009).

Archaeological evidence and early Chinese documents have shown that long-distance interaction networks are of great antiquity in the Philippines, that chiefdoms of considerable political complexity were already well developed in the eleventh century CE, and that the initial period of complex society formation spanned the period from the late first millennium BCE to the early first millennium CE (Junker 1998). Spanish accounts from the early colonial period (Loarca 1903; Junker 1998) also give us a glimpse into a world of great political complexity, with smaller and larger polities, speaking mutually unintelligible languages, engaging in constant warfare, alliance building, and tributary relationships. In addition, maritime trade within the archipelago is well established in the archaeological record for the Philippines, both during and prior to the

18. It is important to note that while diagnostic reflexes are a good basis for establishing the etymology of a word, they may not entirely rule out instances of borrowings. That is, loanwords may be heavily adapted into the recipient language to the point that the expected reflexes of diagnostic protophonemes may be observed in the borrowed forms. Reid (2020:387–88) gives Isinay as an example, citing the speakers' level of bilingualism driving the high degree of adaptation of Tagalog loanwords into the language.

initiation of "foreign" trade late in the first millennium AD (Hutterer 1977; Junker 1990a,b).

Given this complex sociopolitical history, we would argue that the presence of a large amount of loanword diffusion within the Philippine archipelago after the breakup of a higher-level dialect network is more likely than not, and that due to the difficulty alluded to above of distinguishing borrowings from common inheritance in Philippine lexicons, we should be circumspect about accepting arguments for PPH which turn on the quantity of lexical evidence.

Although we cannot completely discount the possibility of common inheritance as the source of these etyma, their semantic properties seem to indicate that they could just as easily be loanwords, and in the absence of diagnostic shared sound change we cannot discard either possibility. In our view, this renders the proposed lexical evidence for PPH problematic.

To conclude, rather than pointing to a unitary PPH which replaced PMP daughter languages in the Philippines, the lexical evidence reveals an alternative scenario, one of a rich and complex history of contact and cultural change. The compilation of the 1,259-item list of Philippine-only etyma and their proposed reconstructions is a major resource for uncovering more about Philippine prehistory, and further careful examination should reveal many more interesting details about the Philippine interaction sphere.

5. HORIZONTAL TRANSFER ACROSS THE PHILIPPINE ISLANDS: INSIGHTS FROM MORPHOSYNTAX.

We have discussed in sections 3–4 how closer examination calls into question the proposed phonological and lexical evidence for a distinct Philippine branch. We now turn to an underexplored case of morphosyntactic variation that further supports the role of extensive contact across the Philippines. As shown below, this variation is widespread among languages spoken on the major islands but absent from outlier languages within the same subgroups—reinforcing the view that the lexical items reported as exclusive to Philippine languages likely arose through contact.

5.1. CIRCUMSTANTIAL VOICE BASICS.

Many Philippine languages retain a complex voice system from PAN, known as *Philippine-type voice*, which features a four-way affixal morphology on the verb governing argument-marking patterns and relativization restrictions (see, e.g., Himmelmann 2002; Chen and McDonnell 2019). This system is illustrated with the Tagalog examples in (3), where a change in voice morphology correlates with a change in the syntactic pivotal phrase, italicized in the translation. To remain analysis-neutral, the case markers (CM) and prepositions (P) are glossed with abstract labels (CM_1; CM_2; P_1; P_2), as their specific identity is not relevant to the key point under discussion.[19]

19. CM_1 corresponds to the case marker conventionally labeled as 'genitive' or 'ergative'; CM_2 refers to the one traditionally glossed as 'oblique' or 'accusative'. The precise case value of these markers is irrelevant to our discussion.

(3) TAGALOG

a. B<**um**>ili si nina ng sinigang mula kay tala
buy<**AV**> PN.PIVOT Nina INDF.CM_2 sour.soup P_1 PN.CM_2 Tala
para Kay rica.
P_2 PN.CM_2 Rica
'*Nina* bought sour soup from Tala for Rica.' (ACTOR VOICE)

b. bi-bilih-**in** ni nina ang sinigang mula kay tala para kay rica.
CONT-buy-**PV** PN.CM_1 Nina PIVOT sour.soup P_1 PN.CM_2 Tala P_2 PN.CM_2 Rica
'Nina will buy *sour soup* from Tala for Rica.' (PATIENT VOICE)

c. bi-bilih-**an** ni nina ng sinigang si tala para kay rica.
CONT-buy-**LV** PN.CM_1 Nina INDF.CM_2 sour.soup PN.PIVOT Tala P_2 PN.CM_2 Rica
'Nina will buy sour soup from *Tala* for Rica.' (LOCATIVE VOICE)

d. **i**-bi-bili ni nina ng sinigang mula kay tala si rica.
CV-CONT-buy PN.CM_1 Nina INDF.CM_2 sour.soup P_1 PN.CM_2 Tala PN.PIVOT Rica
'Nina will buy sour soup from Tala for *Rica*.'
(CIRCUMSTANTIAL VOICE)

As is well known, this four-way voice system is found not only in Philippine languages but also in various other higher-order Austronesian languages spoken in Taiwan, northern Borneo, and northern Sulawesi. Consider below the same set of voice alternations in the Formosan language Paiwan (4).

(4) PAIWAN

a. q<**m**>alup a caucau tua vavuy i gadu tua vuluq.
<**AV**>hunt PIVOT man CM_2 pig LOC mountain CM_2 spear
'*The man* hunts wild pigs in the mountains with a spear.'
(ACTOR VOICE)

b. qalup-**en** nua caucau a vavuy i gadu tua vuluq.
hunt-**PV** CM_1 man PIVOT pig LOC mountain CM_2 spear
'The man hunts *wild pigs* in the mountains with a spear.'
(PATIENT VOICE)

c. qalup-**an** nua caucau tua vavuy a gadu tua vuluq.
hunt-**LV** CM_1 man CM_2 pig PIVOT mountain CM_2 spear
'The man hunts wild pigs *in the mountains* with a spear.'
(LOCATIVE VOICE)

d. **si**-qalup nua caucau tua vavuy i gadu a vuluq.
CV-hunt CM_1 man CM_2 pig LOC mountain PIVOT spear
'The man hunts wild pigs in the mountains *with a spear*.'
(CIRCUMSTANTIAL VOICE) (Ferrell 1979:202)

In both languages, the four voices each exhibit a distinct set of argument-marking patterns. In Circumstantial Voice (CV), the syntactically prominent phrase is typically an instrument or a benefactive, marked with a special label—'pivot'—which signals its eligibility for relativization, demonstrated in (3d) and (4d).

There is solid comparative evidence that this four-way morphology can be traced back to PAN or to a stage shortly after its initial divergence

(Ross 2009; Blust and Chen 2017). Each of the four voices is indexed by a single, monomorphemic affix, as illustrated in (5).

(5) Early Austronesian voice morphology (Ross 2009; Blust and Chen 2017)
   a. Actor Voice: *<um>
   b. Patient Voice: *-en
   c. Locative Voice: *-an
   d. Circumstantial Voice: *Si-/Sa-

5.2. INNOVATIVE CV MORPHOLOGY IN PHILIPPINE LANGUAGES. The focus of the section is an underexplored variation in CV morphology found in Philippine languages. In the majority of Austronesian primary branches represented by Formosan languages, Circumstantial Voice exhibits a one-to-many correspondence with the possible semantic roles of the pivot—that is, it can freely combine with pivot phrases denoting a wide range of roles, including instrument, theme, benefactive, reason, purpose, and cause. This pattern is exemplified by Seediq (6) and Paiwan (7). Each represents a distinct Austronesian primary branch.

(6) SEEDIQ (Formosan)
   a. **s**-hanguc=mu Ø sari ka dakis.
      CV-cook=1SG.CM$_1$ CM$_2$ taro PIVOT Dakis
      'I cooked taro for Dakis.' (benefactive pivot)

   b. ga=na **s**-sebuc Ø ricah ka qreti.
      PROG=3SG.CM$_1$ **CV**-hit CM$_2$ plum PIVOT stick
      'He/she is knocking down plums (from the trees) with the stick.' (instrument pivot)

   c. **s**-k<n>-narux na temi ka knrudan=na.
      CV-STAT<PFV>sick CM$_1$ Temi PIVOT age=3SG.POSS
      'Temi got sick because of her age.' (reason pivot)

   d. wada=mu **s**-paadis Ø dakis ka tigami.
      PFV=1SG.CM$_1$ **CV**-send CM$_2$ Dakis PIVOT letter
      'I sent Dakis a/the letter.' (theme pivot in ditransitives)

   e. **s**-p-seeliq=mu Ø robo ka rodux nii.
      CV-CAUS-butcher=1SG.CM$_1$ CM$_2$ Robo PIVOT chicken this
      'I asked Robo to butcher this chicken.' (theme pivot in causatives)
      (Chen 2017:101, 99, 79, 121, 107)

(7) PAIWAN (Formosan)
   a. 'u-**s**<in>**i**-pangul sa a'-pungdjuq ta kasiw ti kapi.
      1SG.CM$_1$-**CV**<PFV>hit LK STAT-broken CM$_2$ wood PIVOT Kapi
      'I hit the wood broken for Kapi.' (Wu 2013:192) (benefactive pivot)

   b. **si**-tekel ni kapi ta vava a kupu.
      CV-drink PN.CM$_1$ Kapi CM$_2$ wine PIVOT cup
      'Kapi drinks wine with the cup.' (Wu 2013:31) (instrument pivot)

c. **s**<in>**i**-kan ni zepul ta ci'aw a za vengaLay nimadu.
CV<PFV>eat PN.CM$_1$ Zepul CM$_2$ fish PIVOT that pregnancy 3SG.CM$_1$
'Zepul ate fish because of her pregnancy.'
(Chang 2006:73) (reason pivot)

d. **si**-pavai ni cemedas a zua hana tjay zepul.
CV-give PN.CM$_1$ Cemedas PIVOT that flower CM$_2$ Zepul
'Cemedas gave that flower to Zepul.'
(Chang 2006:297) (theme pivot in ditransitives)

e. 'u-**si**-pa-veli=anga tjay kapi a vatu.
1SG.CM$_1$-**CV**-CAUS-buy=COS CM$_2$ Kapi PIVOT dog
'I have caused Kapi to buy the dog.'
(Wu 2013:33) (theme pivot in causatives)

In contrast to the conservatism observed in Formosan languages, many Philippine languages exhibit a set of innovative morphemes that co-occur with the reflex of the PAN CV affix *Si-. This set of morphemes functions to further specify the thematic role of the pivot phrase. For example, the sequence *i-pag-* in Tagalog marks benefactives (8a), *i-paN-* marks instruments (8d), and *i-ka-* indicates reason phrases (8c).[20]

(8) TAGALOG (MP)
a. i-p<in>ag-luto ni kyla ng adobo si juan.
CV-PAG<PFV>-cook CM$_1$ Kyla INDF.CM$_2$ adobo PIVOT Juan
'Kyla cooked adobo for Juan.' (CV + *pag-* for benefactive pivot)

b. i-p<in>ang-ka-kain ni kyla ang kutsara.
CV-PANG<PFV>-RED-eat PN.CM$_1$ Kyla PIVOT spoon
'Kyla is eating with the spoon.' (CV + *pang-* for instrument pivot)

c. i-k<in>a-matay ni juan ang kanser.
CV-KA<PFV>-die CM$_1$ Juan PIVOT cancer
'Juan died of cancer.' (CV + *ka-* for reason pivot)

On the other hand, it remains possible in these languages to use the reflex of *Si- as a stand-alone CV affix. For example, CV-marked ditransitives and causatives that feature a pivot-marked theme argument do not employ any of the three additional affixes *pang-*, *pag-*, or *ka-*, as shown in (9).[21]

(9) TAGALOG (MP)
a. i-b<in>igay ni juan ang pera kay aya.
CV-<PFV>give PN.CM$_1$ Juan PIVOT money PN.CM$_2$ Aya
'Juan gave Aya the money.' (CV for theme pivot in ditransitives)

20. There are also specific Tagalog verbs that do not take the innovative affixes even when combining with a corresponding instrument, benefactive, or reason phrase. The conditions governing the use of these affixes are beyond the scope of this paper.
21. An anonymous reviewer agreed with this generalization but pointed out several exceptions where *pag-* is attached to non-benefactive pivots: *i-pag-kait* 'withhold theme,' *i-pag-tanggol* 'defend theme,' and *i-pag-daos* 'celebrate theme.' This indicates that the sequence *i-pag-* is not fully productive in Tagalog and that its use is subject to native speaker intuition. Our consultation with native speakers did not reveal any clear conditions predicting the presence or absence of *pag-*.

b. i-p<in>a-kanta=ko kay ivan ang kanta.
CV-CAUS<PFV>sing=1SG.CM$_1$ PN.CM$_2$ Ivan PIVOT song
'I asked Ivan to sing a song.' (CV for theme pivot in causatives)

The data above thus reveal four patterns of Circumstantial Voice verbal morphology not attested in Formosan languages, as outlined in (10):

(10) Four subtypes of CV morphology found in some Philippine languages
a. CV
b. CV + *pag-*
c. CV + *pang-*
d. CV + *ka-*[22]

This pattern is not unique to Tagalog. Itbayaten, a Bashiic language with no particularly close relationship to Tagalog, exhibits the same three-way thematic subdivision in CV morphology. Three examples are given in (11).

(11) ITBAYATEN (Bashiic, MP)
a. i-cha-hakey
CV-KA-like
'X is the cause of liking.'
(Yamada 2014:50) (CV + *ka-* for reason pivot)
b. i-pang-among
CV-PAN-fish
'to fish with X'
(Yamada 2014:50) (CV + *paN-* for instrument pivot)
c. i-pang-xap
CV-PAN-get
'to get for X'
(Yamada 2014:50) (CV + *paN-* for benefactive pivot)

Crucially, as in Tagalog, not all CV-marked verbs in Itbayaten employ the innovative affix: while some instrument-selecting CV verbs take the affix *paN-*, as in (8b), others with similar semantics do not require it. Consider below two reported examples that lack the innovative affix (12a,b).

(12) ITBAYATEN
a. i-chali
CV-dig
'to dig with X' (Yamada 2014:50) (CV for instrument pivot)
b. i-'inom
CV-drink
'to drink with X' (Yamada 2014:50) (CV for instrument pivot)

22. An anonymous reviewer suggests that Formosan languages also use the CV morphological sequence with the prefix *ka-*, implying that the pattern *i-ka-* is neither innovative nor unique to Philippine languages. However, this similarity is only in form and not in function. In Formosan languages, the sequence CV-*ka-* is obligatorily used for the CV form of stative stems; that is, the function of *ka-* is stative rather than marking a reason pivot and is therefore distinct from the innovative function discussed here. Readers are encouraged to consult the reference grammars of Formosan languages for further details.

Our survey of 44 better-documented Philippine languages reveals a similar use of the additional, innovative morphology across five of the fifteen subgroups defined in Blust (2001): Bashiic, Central Luzon, Kalamian, Cordilleran (=Northern Luzon), and GCP.[23] Consider below two pairs of examples from Asi (GCP) and Pangasinan (Meso-Cordilleran). In both languages, the same innovative morphemes are used in CV constructions with different types of pivot phrases.

(13) Asi (GCP)

a. i-pang-limpyo nako kag suka it mga gaha.
CV-PANG-clean 1SG.CM$_1$ PIVOT vinegar G PL window
'I'll clean the windows with vinegar.'
(CV + *pang-* for instrument pivot)

b. i-sandrok nako sida it suya.
CV-dish.food 1SG.CM$_1$ 3SG.PIVOT G viand
'I'll dish up some of this viand for her.'
(plain CV for benefactive pivot)

c. Kag i-k<in>a-matay nida ay kanser.
PVIOT CV-KA<PFV>-die 3SG.CM$_1$ AY cancer
'What he died of was cancer.' (CV + *ka-* for reason pivot)
(Hendrickson and Kilgour 1985–1989:39–40)

(14) Pangasinan (Meso-Cordilleran)
i-mpan-katli nen Mark iyan katli ed samay papel.
CV-PANG.PFV-cut CM$_1$ Mark DEM.PIVOT scissors CM$_2$ DET paper
'Mark used this scissors to cut that paper.'
(CV + *paN-* for instrument pivot) (Rosario 2017:81)

Importantly, although this innovative pattern is found across multiple Philippine subgroups, it is not attested in all members within these groups. This is illustrated by the data below from Yami (15) and Karao (16). Both languages either employ a single (retentive) affix *i-* for CV clauses—similar to what is found in Formosan languages—or make use of other affixes to mark specific types of CV constructions (e.g., *i- . . . -an* for clauses with a benefactive pivot, as in Karao (16b)).

(15) Yami (Bashiic)

a. i-tangtang
CV-smash
'to smash with X' (CV for instrument pivot)

b. i-vaod
CV-tie
'to tie up with X'
(Rau and Dong 2006:540, 547, 554) (CV for instrument pivot)

23. We remain agnostic about the subgrouping affiliation of Iraya (a language spoken by Mangyans on the island of Mindoro) and tentatively classify it as part of Central Luzon.

(16) KARAO (Meso-Cordilleran)

a. 'i-tegteg na nga'ngi-'i batho-cha 'aramdi.
CV-flatten CM_1 child-PIVOT rock-CM_2 wire
'The child will use the rock to flatten the wire.'
(CV for instrument pivot)

b. 'i-tongkal-an to'o-'i nga'nga na 'amayo.
CV-buy-AN CM_1 person-PIVOT child CM_2 toy
'The person will buy the child the toy.'
(CV + *-an* for benefactive pivot) (Brainard 1997:101, 100)

The distribution of languages exhibiting this innovation is illustrated in map 2. All language names shown on map 2 correspond to those included in the survey.

The specific CV morphology observed in the 44 languages examined is summarized in table 4, where the second column indicates whether or not the morphology is an innovation. The criteria used to determine the presence or absence of the innovation are as follows: if a language possesses a reflex of the PAN CV affix *Si- that appears in combination with one of the three affixes *pang-*, *pag-*, or *ka-*, we classify it as innovative. Asterisks (1*) indicate cases that display additional types of innovation beyond the CV morphology under discussion. For example, in some languages, the CV affix *i-* that co-occurs with *pang-*, *pag-*, or *ka-* has become optional or entirely lost, such as the use of *paN-* for instrumental pivots in Mag-Antsi. We still count such cases as reflecting the innovation, albeit with an additional step of morphological loss, and mark them with an asterisk. Some languages exhibit the relevant extra morphemes but do not show a clear functional distinction among them—for instance, the use of *i-pang-* for benefactive pivots in Aklanon. These are also considered innovative, as our primary goal is to trace the distribution of *i*-X- sequences across the Philippine archipelago. However, if a language shows no evidence of the affix *i-* for Circumstantial Voice and instead only displays *pang-*, *pag-*, or *ka-*—as in the case of Tawbuid—we do not consider it to exhibit the innovation under discussion.

A closer look at the distribution of the twenty Philippine languages (out of the forty-four examined) that exhibit these innovative affixes reveals an important fact: while the innovation is found across the three major islands—Luzon, Visayas, and Mindanao—including four Negrito languages (Arta, Agta, Mag-Antsi, and Southern Alta) that are in extensive contact with either Ilocano or Tagalog, it is not attested in the Sangiric, Minahasan, or GCP languages spoken in northern Sulawesi and the islands between Sulawesi and Mindanao. Consider the examples below from Talaud (spoken in the Talaud Islands) and Tondano (spoken at the northern tip of Sulawesi). Both languages retain a reflex of the PAN CV affix *Si- as CV morphology and do not employ the innovative morphology (17)–(18).

MAP 2. DISTRIBUTION OF LANGUAGES WITH AND WITHOUT THE INNOVATIVE CV MORPHOLOGY.

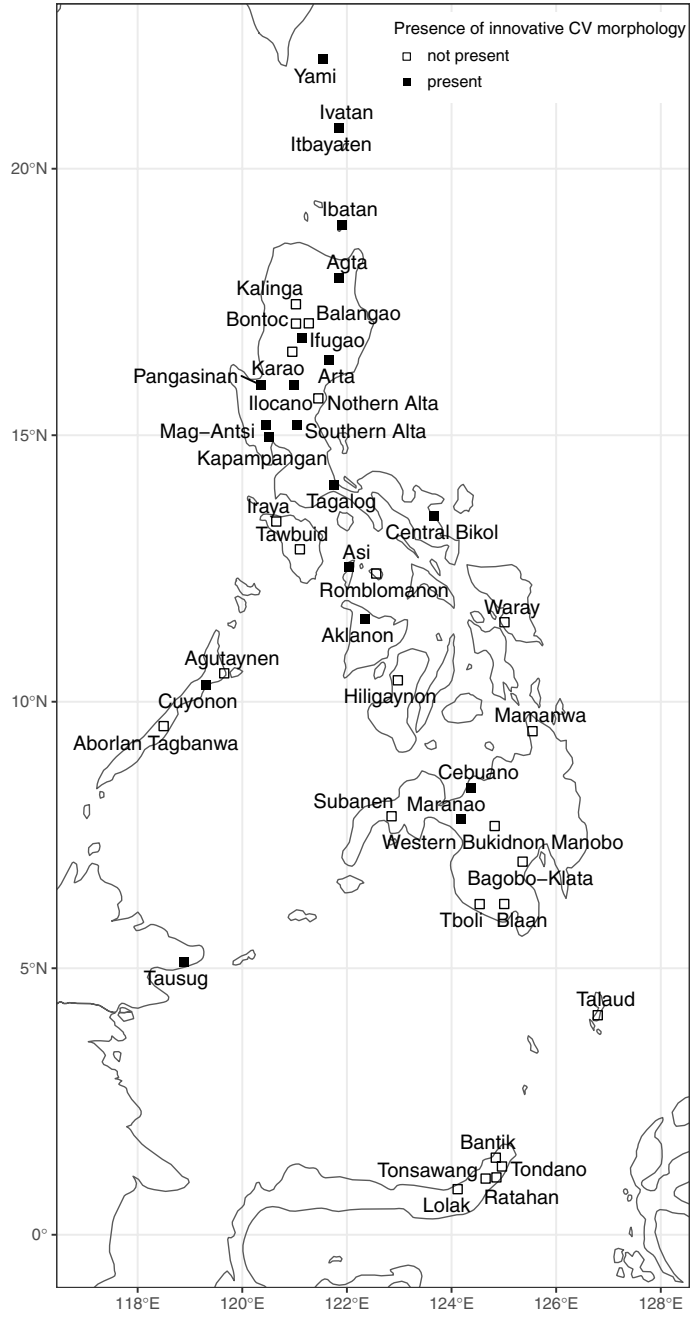


TABLE 4. LIST OF PHILIPPINE LANGUAGES SURVEYED IN THIS PAPER.

	Language	Inn.†	Voice morph.‡	Pivot type	Source
	Bashiic				
1	Yami	0	i-	INST	Rau and Dong (2006)
		1	i-ka-	reason, manner	Rau and Dong (2006)
2	Itbayaten	0	i-	INST	Yamada (2014)
		1	i-cha-	reason, cause	Yamada (2014)
		1	i-paN-	BEN, INST	Yamada (2014)
3	Ivatan	1	qi-paN-	INST	Reid (1966)
4	Ibatan	0	i- . . . -an	BEN	Maree (2007)
		1*	paN-, pay-	INST	Maree (2007)
	Bilic				
5	Tboli	0	zero	INST	Forsberg (1992)
6	Blaan	0	zero	INST	Bondoc (2015)
7	Bagobo-Klata	0	pɔ-, p-	INST	Estrera and Rajik (2022)
	Central Luzon				
8	Iraya	0	pinaN-	INST	Or (2018)
		0	-an	BEN	Or (2018)
9	Kapampangan	1*	i-, pan-, pag-	BEN	Forman (1971)
		1*	pi-, pipag-, pipan-	INST, location	Forman (1971)
		1*	pag-, pan-, i-	INST	Forman (1971)
10	Ayta Mag-Antsi	1*	paN-	BEN, INST	Javier et al. (2018)
		1	ika-	reason	Javier et al. (2018)
	Greater Central Philippines				
11	Mamanwa	0	i-, zero	INST, BEN	Miller (1969)
12	Western Bukidnon Manobo	0	i-	INST, BEN	Elkins (1967)
13	Maranao	1	i-, i-paN-	INST	McKaughan and Macaraya (1967)
		1	-an, paN-, p-i-ang-	referential	McKaughan and Macaraya (1967)
14	Hiligaynon	0	-an	BEN	Wolfenden (1971)
		0	i-	INST	Wolfenden (1971)
15	Romblomanon	0	ging-, in-, gina-, i-	INST	Law (1997)
		0	ging- . . . -an, ging- . . . -i	BEN	Law (1997)
16	Central Bikol	1*	i-, paN-	INST	Mintz (1971)
		1	i-, i-pig-, i-ka-	BEN	Mintz (1971)
17	Tausug	1	hipaN-, hipag-	INST	Estrera and Rajik (2022)
18	Cebuano	1	i-, gi, i-pag-, gika-	INST	Tanangkingsing (2009), Bunye and Yap (1971)
		1*	i-, -an, pag- . . . -i, paga- . . . -an, paN- . . . -an	BEN	Tanangkingsing (2009)
19	Waray	0	i-	INST	Payne and Oyzon (2020)
		0	i-	BEN	Payne and Oyzon (2020)
20	Aborlan Tagbanwa	0	i-	INST	Hussey (1965)
21	Tawbuid	0	faN-	INST	Fleming (2022)
		0	na- . . . -an, fag- . . . -an	reason	Fleming (2022)
22	Subanen	0	pig-, pog- . . . -on	INST	Estioca (2020)
		0	pig- . . . -an, pog- . . . -an	BEN	Estioca (2020)

	Language	Inn.†	Voice morph.‡	Pivot type	Source
23	Lolak	0	pog-, poN-	INST	Lobel and Paputungan (2017)
24	Cuyonon	1*	ing-paN-	INST	Castro et al. (2023)
		0	ing- . . . -an	BEN	Castro et al. (2023)
25	Asi	1	i-, ipaN-	INST	Hendrickson and Kilgour (1985–1989)
		1	i-ka-	cause/reason	Hendrickson and Kilgour (1985–1989)
		0	a- . . . -an, i-	BEN	Hendrickson and Kilgour (1985–1989)
26	Tagalog	0	i-	theme	primary data
		1	i-paN-	INST	primary data
		1	i-pag-	beneficiary	primary data
		1	i-ka-	reason	primary data
27	Aklanon	1*	-in- . . . -an, pa- . . . -an, ipaN-	BEN	de la Cruz and Zorc (1968)
		0	-in-, i-, iga-	INST	de la Cruz and Zorc (1968)
	Kalamian				
28	Agutaynen	0	-an	beneficiary	Quakenbush, Henrickson, and Edep (2010)
		0	i-	INST	Quakenbush, Henrickson, and Edep (2010)
	Cordilleran				
29	Nothern Alta	0	i-	INST, causative theme	García Laguía (2018)
30	Balangao	0	i-	INST	Shetler (1976)
		0	i- . . . -an	BEN	Shetler (1976)
31	Bontoc	0	i-	BEN	Fukuda (1997)
		0	i- . . . -an	INST	Fukuda (1997)
32	Karao	0	i- . . . -an	BEN, recipient	Brainard (1997)
		0	i-	INST	Brainard (1997)
33	Southern Alta	1*	paN-, i-	INST	Abreu (2018)
		0	-en	BEN	Abreu (2018)
34	Ifugao	0	iN- . . . -an	BEN	Hohulin and Burquest (2011)
		1*	i-, impuN-, iN-, puN-	INST	Hohulin and Burquest (2011)
35	Kalinga	0	i- . . . -an, in- . . . -an	BEN	Ferreirinho (1993)
		0	i-, in-	INST	Ferreirinho (1993)
36	Pangasinan	0	i-	INST	Benton (1971)
		0	impama-	reason, cause	Benton (1971)
		1	pan-, inpan-, inpan-, pangi-, inpangi-	INST	Benton (1971)
		0	i- . . . -an, in- . . . -an	BEN	Benton (1971)
37	Agta	1*	i-, paN-	INST	Robinson (2008)
		0	i- . . . -an	BEN	Robinson (2008)
38	Ilocano	1*	para-, paN-, pag-, pangi-	INST	Rubino (2000)
		0	i- . . . -an	BEN	Rubino (2000)
39	Arta	1*	paN-, i-	INST	Kimoto (2017)

	Language	Inn.†	Voice morph.‡	Pivot type	Source
	Minahasan				
40	Tonsawang	0	i-	INST, BEN	Hayes (2019)
41	Tondano	0	i-	INST, BEN	Sneddon (1975)
	Sangiric				
42	Bantik	0	zero	ditransitive theme, BEN, INST	Utsumi (2012, 2013)
43	Ratahan	0	zero	theme, INST, BEN	Himmelmann and Wolff (1999)
44	Talaud	0	i-	ditransitive theme, BEN, INST	Utsumi (2013)

† Inn. = innovation.
‡ Voice morph. = voice morphology.

(17) TALAUD (Sangiric)[24]

a. lama'a i-saraing ngimangitou.
dish CV-dance 3PL.CM$_1$
'Dishes will be used by them in a dance.'
(CV with an instrument pivot)

b. inassa n-i-laha=ku huna m-maria.
fish PST-CV-cook=1SG.CM$_1$ for CM$_1$-Maria
'I cooked fish for Maria.' (CV with a benefactive pivot)

c. ana itou n-i-luass i-piteres [uauggu na-ʔangkat=te huru].
child 3SG.CM$_1$ PST-CV-be.pleased CM$_1$-Peter [because AV.PST-promote=COMP teacher]
'Her/his child made Peter pleased because (her/his child) became a teacher.' (Utsumi 2022:20, 24, 25) (CV with a reason pivot)

(18) TONDANO (Minahasan)

a. tali i-keong ni tuama roda waki pasar.
rope CV-will.pull ERG.AN.SG man cart to market
'The man will pull the cart with the rope to the market.'
(CV with an instrument pivot)

b. se oki' i-lutu' ni mama sera'.
AN.PL child CV-will.cook ERG.AN.SG mother fish
'Mother will cook fish for the children.'
(CV with a benefactive pivot)

c. lelongkotan i-wareng ni tuama waki wale.
ladder CV-will.return ERG.AN.SG man to house
'The man will return the ladder to the house.'
(CV with a theme pivot) (Sneddon 1975:63)

This distribution raises important questions about the prehistory of Philippine languages: Where is the center of dispersal for the innovative CV morphology, and how does its distribution relate to the PPH debate? We suggest that pinpointing the precise locus of this innovation is challenging. The preliminary database summarized in (16) shows that the three-way differentiated CV

24. The translation of (17c) is minimally adapted from the original source to align with the data.

morphology is primarily attested in languages of the GCP, suggesting that this branch may be both the origin of the innovation and a source of borrowing. However, the core significance of this case study lies not in identifying its origin but in understanding its distributional pattern—one that mirrors the phonological and lexical innovations discussed earlier. Specifically, the innovation appears across multiple Philippine microgroups, but only in a subset of languages within each group. Moreover, it is absent from languages spoken on remote islands and inland areas.

We propose that this empirical picture is best understood as the result of extensive contact, leading to morphosyntactic borrowing across the three main islands of the Philippines while largely bypassing offshore and interior communities. This pattern supports a borrowing/linkage scenario rather than one of retention. In other words, it is unlikely that the innovation was present in PPH and was subsequently lost in some languages—a scenario that is implausible given both its total absence in geographically isolated areas and the presence of the more conservative CV morphology in various Philippine languages from distinct subgroups as well as in languages outside the Philippines, such as those in Formosa.[25]

This empirical pattern is difficult to explain unless one assumes that the innovation in CV morphology emerged and was lost independently in each Philippine subgroup. Nor is the pattern consistent with a drift-based analysis, given that the languages exhibiting the innovation use the same set of morphological markers to divide thematic roles. This distribution is crucial to understanding the (pre-)histories of Philippine languages. While tracing the exact source and pathways of this innovation lies beyond the scope of the present paper, it remains a promising area for future research. Ultimately, the evidence from morphosyntax reinforces the presence of extensive contact at the morphosyntactic level.

5.3. IMPLICATIONS: HORIZONTAL TRANSFER AT THE MORPHOSYNTACTIC LEVEL. As discussed above, the observed distribution of the innovative CV morphology makes it unlikely that these patterns result from inheritance from the treetop (i.e., retention from a putative common ancestor that subsumes all Philippine languages). Such a scenario would require the innovation to have been lost in (i) all languages spoken outside the three major

25. An anonymous reviewer noted the presence of the affix *pang-* (and similar *p*-initial instrument-indexing affixes) in several Bornean languages (e.g., Bonggi, Kelabit, Kimaragang, Tatana' Murut, Lun Bawang, and Totoli), where *pang-* (or its equivalent) functions similarly to the CV affix *i-* by selecting instrumental pivots. We acknowledge that such cases may reflect the original loss of *i-* from the sequence *i-paN-*. However, it remains unclear whether *i-paN-* ever existed in these languages, as the presence of *paN-* (or its equivalent) could reflect inheritance from a shared ancestor or direct borrowing. To our knowledge, Kimaragang Dusun is the only language where *i-* and *pang-* co-occur and can be used interchangeably to index instrumental pivots (Kroeger 2010). Contact effects are likely at play, given that Kimaragang is spoken at the northern tip of Borneo, geographically close to the Philippines. Since we are not proposing *i-X-* as a defining feature of Philippine languages but rather using its distribution to support the presence of extensive contact among the Philippine islands, the occurrence of *paN-* in CV constructions in Borneo does not undermine our central argument.

islands and (ii) the majority of languages within the branches that do exhibit the change. Instead, the distribution supports the presence of horizontal transfer across the main islands of the Philippines.

This innovative CV morphology not only ranks low on Field's (2002) hierarchy of borrowability (19)—suggesting the likelihood of extensive contact—but also aligns well with Type III structural change (20), as defined by Heine and Kuteva (2005).

(19) Hierarchy of category borrowability (Field 2002)
content item > function word > agglutinating affix > fusional affix

(20) Type III Structural effect of contact-induced grammaticalization (Heine and Kuteva 2005:124)
The new and the old categories coexist side by side, but the structure of the old category is redefined as a result of the presence of the new category (differentiation).

A related question raised by both anonymous reviewers concerns the source language in which this innovation first emerged and the directionality of the hypothetical borrowing. One reviewer also questioned whether the current distribution of these three affixes might instead reflect extensive loss in extra-Philippine languages. Given the corroborating pieces of evidence observed in the domains of phonology, lexicon, and morphosyntax, it is apparent that the Philippines is a highly interactive zone, with varying patterns of multilingualism among the speakers, and so, linguistic resources are easily transferred to and from languages, making it difficult to determine a single source of borrowing or diffusion. Structural borrowing also presupposes extensive lexical borrowing (e.g., Thomason and Kaufman 2001; Heine and Kuteva 2005; Aikhenvald 2006; Matras 2009, inter alia.), and so, the presence of structural borrowing in the innovative CV morphology strongly implies widespread lexical borrowing across the main islands of the Philippines. This understudied locus of morphosyntactic variation, along with the argument for lexical borrowing and diffusion in the previous section, provides strong evidence against the PPH hypothesis.

6. CONCLUSION. We have reexamined recent arguments for PPH as the sole surviving PMP descendant on Luzon (Blust 2019, 2020, 2022; Zorc 1986, 2020) and presented new evidence in support of an alternative view. We argue that there is little justification for positing this alleged ancestor, as the proposed evidence for PPH can be more plausibly explained through multiple layers of diffusion and borrowing resulting from extensive contact, as suggested by Ross (2020). Three lines of new evidence support this view. First, the absence of the PMP *d/*z merger in the Central Luzon subgroup undermines its status as the single defining phonological innovation of PPH. Second, the geographical distribution of the purported lexical innovations strongly suggests contact effects rather than inheritance. Third, the distribution of an underexplored pattern of Circumstantial Voice (CV) morphology reveals extensive

contact across Philippine subgroups, providing further empirical support for the argument that the major Philippine islands comprise a highly interactive zone, resulting in the widespread transfer of linguistic features. We therefore conclude that there is no compelling motivation to assume the existence of PPH.

APPENDIX

Himes (2012) proposes an account of the Central Luzon languages using regular sound change rules. One of the proposed rules is *d > *l* / V_V in Kapampangan and the Sambalic languages. In addition to this, he proposes a number of other rules to explain the split realization of *d, but these are lacking in that they are highly targeted rules which often lack phonetic motivation and leave much of the data still unexplained. For example, Himes suggests that in the Sambalic languages, *d is reflected as /l/ word-initially, except when the next consonant is a liquid. However, there are many stable forms, stretching back to Proto-Western-Malayo-Polynesian (PWMP) and earlier, which violate this rule (e.g., PWMP *dapít > AA **dapít* 'part, place, direction', PAN *depah > AA *depah* 'armsbreadth', or PAN *daqaN > BT *daan* 'old (of objects)'). In addition, there is no clear phonetic motivation for *d to change to /l/ in word-initial but not word-final position. We suggest that the sound change *d > *l* / V_V is sufficient to explain not only the vowel-medial instances of *d > *l*, but also the word-initial and word-final instances, once the possibility of back-formation from affixed forms or forms in connected speech is taken into account.

First, we examine the straightforward instances of *d > *l* / V_V. There were seventeen instances of this sound change found in the data, across all four languages, and seven exceptions. The exceptions are PWMP *ledek > AA *ledek* 'to pound grain by mortar and pestle'; PPH *pidek (with metathesis)$_2$ > BT/AM *kirp* 'eyelash'; PAN *kuden > KP *kúran* 'large cooking pot for rice', PPH *dúdun > KP *durún* 'locust', and PPH *katúday > KP *kature* 'a plant: Sesbania grandiflora'; and PMP *pudul > BT *poról* 'blunt, dull', plus one exception found in Himes (2012) but not the ACD data: PMP *ludaq > BT *ludá?* 'spit'. For six of these exceptions, we have good reason to suspect that the *d > *l* / V_V sound change did not apply to them: AA *ledek* is likely a new coinage from the root *-dek$_2$. BT/AM *kirp*, KP *durún*, and KP *kature* are all words unique to the Philippines and may be considered as loanwords or new coinages which entered into the languages in question at a time after the proposed sound changes took place and so were not affected by them. BT *ludá?* may actually reflect PWMP *luzaq 'saliva; to spit', which was likely originally *luCzaq, as suggested by the medial geminate in Samal *lujja* 'to spit'. So the medial *z would not be affected by the sound change *z > *r* / V_V and would later merge with *d. BT *poról* can be explained in terms of the OCP (obligatory contour principle) effect, a universal cognitive tendency in human language where repeated identical segments are dispreferred (McCarthy 1981). Austronesian languages seem to have a specific constraint against syllables

with an identical liquid in the onset and coda of a syllable—the only such instance in all the ACD is POC *nguRuR 'to moan, groan, grunt'. If the medial /d/ in *pudul 'blunt, dull' changed to /l/, the result would be **polol, and so to avoid this highly dispreferred sequence, *d irregularly changes to /r/. If the given explanations hold, only the Kapampangan reflex *kúran* from PAN *kuden remains as an unexplained exception to the rule *d > *l* / V_V, leaving it standing on solid ground.

In the Sambalic languages (but not Kapampangan), *d can also appear as /l/ in word-initial or -final position, though it shows up less frequently than in the vowel-medial position. In Ayta Abellen, for example, PAN *daRaq > *daya* 'blood' but PAN *daNum > *lanom* 'water'. We hypothesize that such a split realization of *d could be the result of a secondary effect of the rule *d > *l* / V_V, that is, the result of back-formation from affixed forms, or even from forms in connected speech. In the case where *d is in word-initial or -final position, the presence of prefix or suffix, respectively, would place *d in a vowel-medial position, making it subject to the rule *d > *l* / V_V. This would result in two competing forms of the same stem being present in the language, one (an *l*-stem) contained within the affixed form, the other (a *d*-stem) within the unaffixed stem. For example, if Proto-Central Luzon *dateng 'to come' were affixed to become *ka-dateng-an, the affixed form would change to *kalatengan under the rule *d > *l* / V_V, and later that form could be reinterpreted as consisting of a stem *lateng affixed with *ka- . . . -an. If this process occurred in Proto-Sambalic, it would mean that in some cases, the innovative *l*-stem would replace the original, and in other cases, the original *d*-stem would remain and be reaffixed to push out the innovative affixed form. A similar scenario could play out if a word without a final consonant precedes a word beginning with *d in connected speech, or a word with a final *d is followed by a word without an initial consonant. Phonetically, *d would be in an intervocalic position and thus be realized as /l/. Again, the result would be two competing forms in the language, one with a *d*-stem and one with an *l*-stem. Such a scenario would lead to the inconsistent realization of *d in word-initial and -final positions that we see in the Sambalic languages. After this process, the remaining *d*-stem reflexes would merge with *z, splitting with the *l*-stems in their realization. An anonymous reviewer informs us that a similar process is currently active in Tagalog, where /d/ regularly becomes /r/ between vowels: "In roots like *damdam* 'feel', we now find a variant *ramdam*, which arose from frequent prefixation in its verbal form *ma-damdam-an* 'to feel'." See also Blust (2013:623).

REFERENCES

Abreu, M. 2018. *A Reference Grammar of Southern Alta (Kabuloan Dumagat)*. DLSU, BAGCED-DEAL. Manila: De La Salle University.

Aikhenvald, A. 2006. Reflections on language contact, areal diffusion, and mechanisms of linguistic change. *Sprachbund in the West African Sahel*, 23–36.

Benton, R. 1971. *Pangasinan Reference Grammar.* University of Hawai'i Press.

Blake, F. 1906. Contributions to Philippine grammar. *Journal of the American Oriental Society* 27:317–96.

Blust, R. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *Working Papers in Linguistics* 9(2): 1–15. Honolulu: Department of Linguistics, University of Hawai‘i.

———. 1991. The Greater Central Philippine hypothesis. *Oceanic Linguistics* 30(2): 73–129.

———. 1993. *s metathesis and the Formosan/Malayo-Polynesian language boundary. In *Language, a Doorway Between Human Cultures: Tributes to Dr. Otto Chr. Dahl on His Ninetieth Birthday*, ed. by D. Øyvind, 178–83. Oslo: Novus.

———. 1995. The position of the Formosan languages: method and theory in Austronesian comparative linguistics. In *Austronesian Studies Relating to Taiwan*, ed. by P. Li et al., 585–650. Symposium Series of the Institute of History and Philology, Academia Sinica, No. 3. Taipei: Academia Sinica.

———. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by E. Zeitoun and P. Li, 31–94. Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 1. Taipei: Academia Sinica.

———. 2001. Malayo-Polynesian: new stones in the wall. *Oceanic Linguistics* 40:151–55.

———. 2005. The linguistic macrohistory of the Philippines: Some speculations. In *Current topics in Philippine Linguistics and Anthropology, Parangal kay Lawrence A. Reid*, ed. by H. Liao and C. Rubino, 31–68. Linguistic Society of the Philippines.

———. 2013. *The Austronesian Languages* (Revised). The Australian National University.

———. 2019. The resurrection of Proto-Philippines. *Oceanic Linguistics* 58(2): 153–256.

———. 2020. Response to comments on “The resurrection of Proto-Philippines.” *Oceanic Linguistics* 59(1–2): 450–79.

———. 2022. Proto-Philippine addenda: mind, theory, and data. *Oceanic Linguistics* 61(1): 322–404.

Blust, R., and V. Chen. 2017. The pitfalls of negative evidence: ‘Ergative Austronesian’, ‘Nuclear Austronesian’ and their progeny. *Language and Linguistics* 18(4): 577–621.

Bondoc, I. 2015. A grammar sketch of Blaan. MA Thesis, University of the Philippines Diliman.

Brainard, S. 1997. Ergativity and grammatical relations in Karao. In *Grammatical Relations: A Functionalist Perspective*, ed. by T. Givón, 85–154. Amsterdam/Philadelphia: John Benjamins.

Bunye, R., and E. P. Yap. 1971. *Cebuano Grammar Notes*. Honolulu: University of Hawai‘i Press.

Castro, J., C. Comeo, R. Bautista, and C. Mamucud. 2023. The Verbal Morphology of Cuyonon: Voice, Aspect, and Modality. Paper presented at The 32nd Meeting of the Southeast Asian Linguistics Society (SEALS32).

Chang, A. 2006. A reference grammar of Paiwan. PhD diss., The Australian National University.

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

Chen, V. 2017. A reexamination of the Philippine-type voice system and its implications for Austronesian primary-level subgrouping. PhD diss., University of Hawai‘i at Mānoa.

Chen, V., and B. McDonnell. 2019. Western Austronesian voice. *Annual Review of Linguistics* 5:173–95.
Dahl, Otto Chr. 1973. *Proto-Austronesian. Scandinavian Institute of Asian Studies*, Monograph Series No. 15. Lund: Studentlitteratur.
de la Cruz, B. A., and R. D. Zorc. 1968. *A Study of the Aklanon Dialect*. Volume One: Grammar. Peace Corps.
Elkins, R. 1967. *Major Grammatical Patterns of Western Bukidnon Manobo*. Norman: Summer Institute of Linguistics of the University of Oklahoma.
Estioca, S. 2020. A grammar of Western Subanon. PhD diss., University of Hawai‘i at Mānoa.
Estrera, E., and J. Rajik. 2022. Modality in Bagobo-Klata and Tausug. *The Archive* 3:1–86.
Ferreirinho, N. 1993. *Selected Topics in the Grammar of Limos Kalinga, the Philippines*. Canberra: Pacific Linguistics.
Ferrell, Raleigh. 1979. Construction markers and subgrouping of Formosan languages. *Southeast Asian Linguistic Studies* 3:199–212.
Field, F. 2002. *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins.
Fleming, H. 2022. *A Grammar of Eastern Tawbuid*. SIL International.
Forman, M. 1971. *Kapampangan Grammar Notes*. University of Hawai‘i Press.
Forsberg, V. 1992. A pedagogical grammar of Tboli. *Studies in Philippine Linguistics* 9:1–110.
François, A. 2014. Trees, waves and linkages: Models of language diversification. In *The Routledge Handbook of Historical Linguistics*, ed. by C. Bowern and B. Evans, 161–89. Routledge.
Fukuda, T. 1997. A discourse-oriented grammar of Eastern Bontoc. *Studies in Philippine Linguistics* 10(1): 1–116.
Gallego, M. K. S. 2020. Ibatan of Babuyan Claro (Philippines) - Language contexts. *Language Documentation and Description* 17:87–110.
García Laguía, A. 2018. Documentation of Northern Alta: grammar, texts and glossary. PhD diss., Universitat de Barcelona.
Gray, R. D., A. J. Drummond, and S. J. Greenhill. 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913): 479–82.
Haspelmath, M., and U. Tadmor. 2009. The Loanword Typology project and the World Loanword Database. In *Loanwords in the World's Language: A Comparative Handbook*, ed. by M. Haspelmath and U. Tadmor, 1–34. De Gruyter.
Hayes, R. 2019. A morphosyntactic description of verbal constructions in Tonsawang, an endangered language of northern Sulawesi. MA thesis, SOAS, University of London.
Heine, B., and T. Kuteva. 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
Hendrickson, G., and H. J. Kilgour. 1985–1989. Bantoanon (Asi) grammar essentials. Retrieved from http://tagmi.net/en/grammar.
Himes, R. 2012. The Central Luzon Group of Languages. *Oceanic Linguistics* 51(2): 490–537.
Himmelmann, N. 2002. Voice in Western Austronesian: An update. In *The History and Typology of Western Austronesian Voice Systems*, ed. by F. Wouk and M. Ross, 7–15. Canberra: Pacific Linguistics.
Himmelmann, N., and J. Wolff. 1999. *Toratán (Ratahan)*. Lincom Europa.
Hohulin, E. L., and Donald A. Burquest. 2011. *The Cross-referencing System of Tuwali Ifugao: Form and Function*. Manila: SIL Philippines.
Hussey, S. C. 1965. Aborlan Tagbanwa verbal system and related topics. MA thesis, The Hartford Seminary Foundation.

Hutterer, K. 1977. Prehistoric trade and the evolution of Philippine societies: A reconsideration. In *Economic Exchange and Social Interaction in Southeast Asia: Perspectives from Prehistory*, ed. by K. L. Hutterer, 177–96. Ann Arbor: University of Michigan Center for South and Southeast Asian Studies.

Javier, J. R., F. Rosario, S. Benosa, and S. Klimenko. 2018. Ang Mag-antsi Sentence. *The Archive* 15:73–109.

Junker, L. 1990a. The organization of intra-regional and long-distance trade in prehispanic Philippine complex societies. *Asian Perspectives* 29:167–209.

———. 1990b. Long-distance trade and the development of socio-political complexity in Philippine chiefdoms of the first millennium to mid-second millennium A.D. PhD diss., University of Michigan, Ann Harbor.

———. 1998. Integrating history and archaeology in the study of contact. *International Journal of Historical Archaeology* 2:291–320.

Kimoto, Y. 2017. A grammar of Arta: A Philippine Negrito language. PhD diss., Kyoto University.

King, B., S. GreenHill, L. Reid, M. Ross, M. Walworth, and R. D. Gray. 2024. Bayesian phylogenetic analysis of Philippine languages supports a rapid migration of Malayo-Polynesian languages. *Scientific Reports* 14:149–67. doi: 10.1038/s41598-024-65810-x.

Kroeger, P. 2010. The grammar of hitting, breaking, and cutting in Kimaragang Dusun. *Oceanic Linguistics* 49:1–20.

Law, R. S. 1997. *The Romblomanon Grammar Essentials Sketch*. Summer Institute of Linguistics. Available at: https://romblomanon.net/en/study-of-language/grammar.

Liao, H. 2020. A reply to Blust (2019) "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2): 426–49.

Llamzon, T. A. 1975. Proto-Philippine phonology. *Archipel* 9:29–42.

Loarca, M. 1903–1909 (1582). Relacion de las Islas Filipinas. In *The Philippines. 1493–1898*, ed. by E. Blair and J. Robertson, Vol. 5, 32–187. Cleveland: Arthur H. Clark.

Lobel, J. 2013. Philippine and north Bornean languages: Issues in description, subgrouping, and reconstruction. PhD diss., University of Hawai'i.

Lobel, J., and A. Paputungan. 2017. Notes from the field: Lolak: another moribund language of Indonesia, with supporting audio. *Language Documentation and Conservation* 11:328–66.

Lobel, J., and O. Vertudez Surbano. 2019. Notes from the Field: Remontado (Hatang-Kayi): A Moribund Language of the Philippines. *Language Documentation and Conservation* 13:1–34.

Malinowski, B. 1922. *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagos of Melanesian New Guinea*. London: Routledge & Kegan Paul.

Maree, R. 2007. *Ibatan: A grammatical sketch of the language of Babuyan Claro Island*. Linguistic Society of the Philippines Special Monograph Issue 53. Manila: Linguistic Society of the Philippines.

Matras, Y. 2009. Defining the limits of grammatical borrowing. In *The Indo-European Language Family: Questions about its Status (Journal of Indo-European Studies Monograph Series 55)*, ed. by A. Marcantonio, 1–25. Washington, D.C.: Institute for the Study of Man.

McCarthy, J. 1981. A prosodic theory of non-concatenative morphology. *Linguistic Inquiry* 12:373–418.

McKaughan, H. P., and B. A. Macaraya. 1967. *A Maranao Dictionary*. Honolulu: University of Hawai'i Press.

Mead, D. 2020. Did Proto Malayo-Polynesian *d and *z merge in Proto Minahasan? - A review of the fate of PMP *d, *z, *j, *r and *l in Proto Minahasan and the Minahasan languages. Unpublished manuscript.

Miller, H. 1969. Mamanwa morphology. *Papers in Philippine Linguistics* 2:11–32. Canberra: Pacific Linguistics.
Mills, Roger F. 1975. Proto South Sulawesi and Proto Austronesian phonology. PhD diss., Department of Linguistics, The University of Michigan. Ann Arbor, MI: University Microfilms International.
Mintz, M. 1971. *Bikol Grammar Notes*. Honolulu: University of Hawai'i Press.
Or, E. 2018. A grammar of Iraya [Unpublished master's thesis]. University of the Philippines Diliman.
Payne, T., and V. Oyzon. 2020. Transitivity, modality, and voice in Waray. *Philippine Journal of Linguistics* 51:1–29.
Paz, C. 1981. *A reconstruction of Proto-Philippine phonemes and Morphemes*. Cecilio Lopez Archives of Philippine Languages and the Philippine Linguistics Circle.
Pennoyer, D. 1987. Inati: The hidden Negrito language of Panay, Philippines. *Philippine Journal of Linguistics* 18(2)/19(1): 1–36.
Quakenbush, J. Stephen, Gail R. Henrickson, and Josenita L. Edep. 2010. *A Brief Overview of Agutaynen Grammar.* Manila: SIL Publishing.
Rau, V., and M. Dong. 2006. *Yami Texts with Reference Grammar and Dictionary.* Institute of Linguistics, Academia Sinica.
Reid, L. 1966. *An Ivatan Syntax*. Oceanic Linguistics Special Publication. Honolulu: Pacific and Asian Linguistics Institute, University of Hawai'i.
———. 1978. Problems in the reconstruction of Proto-Philippine construction markers. In *Second International Conference on Austronesian Linguistics: Proceedings*, ed. by S. Wurm and L. Carrington, Vol. 1: 33–36. Department of Linguistics, Research School of Pacific Studies, The Australian National University.
———. 1982. The demise of Proto-Philippines. In *Papers from the 3rd International Conference on Austronesian Linguistics*, ed. by A. Halim, L. Carrington, and S. Wurm, vol. 2: *Tracking the Travelers*, 201–16. 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.
———. 1994. Unravelling the linguistic histories of Philippine Negritos. In *Language Contact and Change in the Austronesian World*, ed. by T. E. Dutton and D. T. Tryon, 443–75. Berlin: Mouton de Gruyter.
———. 2018. Modeling the linguistic situation in the Philippines. In *Let's Talk about Trees*, ed. by R. Kikusawa and L. Reid, 91–105. Osaka: Senri Ethnological Studies, Minpaku.
———. 2019. Reassessing the Position of Isinay in the Central Cordilleran Language Family. In A. Lee and V. C. R. Santiago (eds.), The Archive Special Publication No. 16, 1–32. Manila: University of the Philippines Diliman.
———. 2020. Response to Blust "The resurrection of proto-Philippines." *Oceanic Linguistics* 59(1–2): 374–93.
Robinson, L. 2008. Dupaningan Agta: Grammar, vocabulary, and texts. PhD diss., University of Hawai'i at Mānoa.
Rosario, F. 2017. Subcategorization of Pangasinan verbs. *Diliman Review* 61(1): 58–87.
Ross, M. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian Dictionary*, 45–120. Mouton de Gruyter.
———. 2002. Final words: Research themes in the history and typology of western Austronesian languages. In *The History and Typology and Western Austronesian Voice Systems*, ed. by F. Wouk and M. Ross, 451–74. Canberra: Pacific Linguistics.
———. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2): 1–24.
———. 2009. Proto Austronesian verbal morphology: A reappraisal. In *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, ed. by K. A. Adelaar and A. Pawley, 295–326. Canberra: Pacific Linguistics.

———. 2020. Comment on Blust "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2): 366–73.
Rubino, C. 2000. *Ilocano dictionary and grammar: Ilocano-English, English-Ilocano*. Honolulu: University of Hawai'i Press.
Scheerer, O. 1918. Outlines of the history of exploration of the Philippine languages and their relatives in East and West. *The Philippine Review* 3(1): 59–67.
Shetler, J. 1976. *Notes on Balangao Grammar*. Huntington Beach, CA: Summer Institute of Linguistics.
Smith, A. 2017. The Western Malayo-Polynesian problem. *Oceanic Linguistics* 56(2): 435–90.
———. 2023. Evidence and models of linguistic relationships: Subgroups, linkages, lexical innovations, and Borneo. *Oceanic Linguistics* 62:324–65.
Sneddon, J. N. 1975. *Tondano Phonology and Grammar*. Canberra: Pacific Linguistics.
———. 1978. *Proto-Minahasan: Phonology, Morphology, and Wordlist*. Canberra: The Australian National University.
———. 1989. The North Sulawesi microgroups: In search of higher level connections. *Nusa* 31:83–107.
Stone, R. 2008. The sambalic languages of central Luzon. *Studies in Philippine Languages and Cultures* 19:158–83.
Tanangkingsing, M. 2009. A Functional reference Grammar on Cebuano. PhD diss., National Taiwan University.
Thomason, S., and T. Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
———. 2001. *Language Contact*. Edinburgh: Edinburgh University Press.
Utsumi, A. 2012. Applicative verbs and applicative construction in the Bantik language. *Journal of the Southeast Asian Linguistics Society* 5:107–25.
———. 2013. Talaud verbs: Paradigm of basic verbs. *Tokyo University Linguistics Papers* 33:319–61.
———. 2022. Semantic typology of voice systems in Western Malayo-Polynesian languages. *STUF - Language Typology and Universals* 1:129–60.
Wolfenden, E. P. 1971. *Hiligaynon reference grammar*. Honolulu: University of Hawai'i Press.
Wu, J. 2013. The syntax of linking constructions in Mayrinax Atayal and Sinvaudjan Paiwan. PhD diss., National Tsing Hua University.
Yamada, Y. 2014. A grammar of the Itbayat language of the Philippines. Himeji, Japan.
Zorc, D. 1986. The genetic relationships of Philippine languages. In *FOCAL II: Papers from the 4th International Conference on Austronesian Linguistics*, ed. by P. Geraghty, L. Carrington, and S. Wurm, 147–73. Pacific Linguistics.
———. 2019. Klata/Giangan—A new southern Philippine subgroup. *The Archive: Special Publication* 16:33–51. Diliman: University of the Philippines.
———. 2020. Reactions to Blust's "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2): 394–425.
Zorc, D., and A. Almarines. 2022. Axis relationships in the Philippines: When traditional subgrouping falls short. *The Archive* 17:1–112.

victoria.chen@vuw.ac.nz
msgallego@up.edu.ph
kuojona@gmail.com
isaac_stead@eva.mpg.de
bjvdv.13@gmail.com