



Project  
**MUSE**<sup>®</sup>  
*Scholarly journals online*

# Yet More on the Position of the Languages of Eastern Indonesia and East Timor

Mark Donohue and Charles E. Grimes

AUSTRALIAN NATIONAL UNIVERSITY AND SIL INTERNATIONAL

The line dividing the Austronesian languages into Western Malayo-Polynesian (WMP) and Central-Eastern Malayo-Polynesian (CEMP) is drawn east of Sulawesi and through the middle of Sumbawa. A number of phonological or semantic changes are claimed as forming the basis of this distinction, as well as the typological profile of the languages to the east being different from those to the west, and a number of lexical items being attested only east of the line. We examine the phonological and semantic innovations, as well as the erratic morphological ones, showing that none of them define the CEMP line, but indicate that (a) the Central Malayo-Polynesian (CMP)-area languages do not convincingly meet the criteria commonly accepted for a subgroup or even a linkage, (b) some of the WMP-area languages exhibit more of the same features found in at least some of the CMP-area languages than do others, and (c) many of the traits ascribed to the CMP- or CEMP-area languages can be found in more conservative WMP-area or Formosan languages as well.

**1. AIMS.**<sup>1</sup> In this article we wish to draw attention to data from the languages of central and eastern Indonesia and East Timor that call into question the nature of the division between the Western Malayo-Polynesian (WMP) and Central Malayo-Polynesian (and the Central-Eastern Malayo-Polynesian) languages (henceforth CMP and CEMP, respectively).<sup>2</sup> This division has been proposed with varying degrees of evidence over the years, particularly by Blust (1982), culminating in Blust (1993), which also contains a more comprehensive list of earlier work.

We intend to point out some problems with assuming that the Proto-Central-Eastern Malayo-Polynesian (PCEMP) and Proto-Central Malayo-Polynesian (PCMP) reconstructions define clear divisions in the Austronesian family. Blust himself describes PCMP as a linkage (using the terminology introduced by Ross 1988), rather than a single

1. Thanks to Beth Evans, David Kamholz, Dan Kaufman, Malcolm Ross, and an anonymous reviewer for comments that have substantially improved the presentation of this paper.
2. In addition to these three abbreviations, others used are: ABVD, Austronesian Basic Vocabulary Database; CAD, *Comparative Austronesian Dictionary*; CORE, core argument; EMP, Eastern Malayo-Polynesian; INAL, inalienable/plural marker; NONACT, nonactive voice; P-, Proto-; PAN, Proto-Austronesian; PMP, Proto-Malayo-Polynesian; R, realis; SHWNG, South Halmahera–West New Guinea; SUL, Sulawesi.

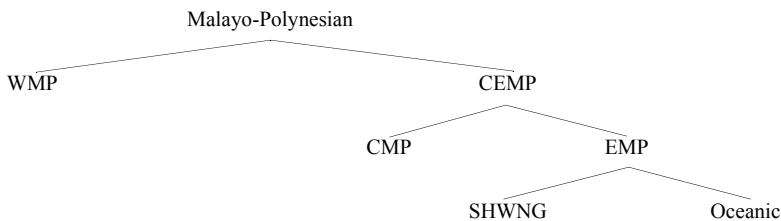
protolanguage, and many of the innovations proposed to characterize CMP in the body of the article are admitted not to apply to the whole range of CMP-area languages.<sup>3</sup> In this article we examine the problematic nature of the proposed phonological innovations, showing that they are both not sufficiently widespread in the eastern area (as Blust 1993 points out), and are sufficiently widely attested in the west.

We argue that because the few apparent innovations occur only sporadically within the CEMP area, the presence of these innovations in languages that are geographically contiguous to the CEMP-area languages suggests that any PCEMP node should, logically, include some languages that have been assumed to be classified as “WMP.” We conclude that, when considering the status of these innovations with respect to the Austronesian languages to their west, the putatively defining isoglosses do not add up to a subgrouping argument.

Second, we show that many of the innovations that Blust proposes for CMP and for CEMP are not, as he also points out, found across the whole CMP area (we leave aside the question of their attestation in EMP).<sup>4</sup> The fact that Blust claims to have reconstructed a linkage, rather than an undifferentiated protolanguage, means that we should expect some degree of incompleteness in the attestation of these forms, but that the incompleteness should be geographically defined, rather than random. The fact that we find many of the proposed innovations outside the CMP area, in the WMP area, argues for a greater fragmentation than can be assumed with a linkage that is still historically and geographically limited.

The tree popularly assumed for the Austronesian languages external to Formosa is shown in figure 1, following Blust’s grouping of all of the CMP-area languages as descendants of one protolanguage, and the grouping of this protolanguage with the protolanguage of the EMP languages as CEMP. This tree has been widely repeated in numerous publications.

**FIGURE 1. THE GENERALLY-CITED SUBGROUPING RELATIONSHIPS AMONG THE AUSTRONESIAN LANGUAGES FOUND IN INDONESIA**

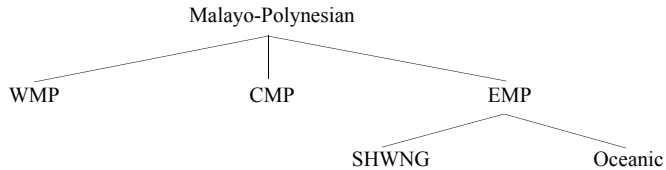


3. Here, and elsewhere, we use circumlocutions such as “CMP-area languages” and “CMP area” to refer to the languages that have been referred to as (in this case) Central Malayo-Polynesian and the area in which they are spoken, without prejudging the phylogenetic unity of those languages. Similarly, we use the term “Western Malayo-Polynesian” as a label of geographic convenience without implying belief in the linguistic status of “Western Malayo-Polynesian.”

4. This point was also made previously in Grimes (1991a) for a number of Blust’s proposed innovations for CMP and CEMP.

Earlier proposals for the subgrouping of the Malayo-Polynesian languages do not include a PCEMP node; figure 2 shows an earlier classification that treats WMP, CMP, and EMP as sisters.

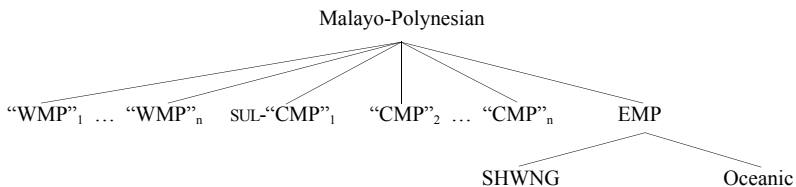
**FIGURE 2. AUSTRONESIAN SUBGROUPING FROM BLUST (1981)**



In this article we point out that the claims for PCEMP are not the most parsimonious explanation of the available data, and wish to raise the possibility that at least some of the CMP-area languages (from around Flores) share features with a subgroup of languages that have been treated as WMP, those of (Southeast) Sulawesi. Further, the lack of innovations that convincingly group the CMP-area languages together means that, accepting that there is no single “Proto Western Malayo-Polynesian” node (following Ross 1995), we are suggesting that further research over time may yield a structure something like that in figure 3.<sup>5</sup> Evidence for the link between (some of) the CMP-area languages and Southeast Sulawesi is scattered throughout the paper, as we discuss individual linguistic traits.

In this article we address the innovations that appear in Blust (1993), described by him as being “phonological” (both regular and irregular) or “semantic.” The lexical innovations are the subject of a separate study (Grimes and Donohue n.d.), and the morphosyntactic innovations can all be shown to be the result of contact-induced change (Donohue 2005a, 2007b, Donohue and Musgrave 2007, Donohue and Schapper 2007, Grimes 1991a, 2007, Grimes and Donohue n.d., Reesink 2002). While the detailed argumentation concerning lexical and morphosyntactic innovations is absent from this article, we summarize (and appeal to) those results in our discussion.<sup>6</sup>

**FIGURE 3. AUSTRONESIAN SUBGROUPING:  
A POSSIBLE REFLECTION OF THE DATA**



5. In this figure the labels “WMP” and “CMP” are used in a purely areal sense. Subscripts denote different genealogies in the same (roughly similar) area; thus CMP<sub>2</sub> is no more closely related to CMP<sub>n</sub> than to EMP.

While we have drawn on a number of sources, essentially as many as we have had access to, we have made extensive use of the *Comparative Austronesian Dictionary* (Tryon 1995) and the online Austronesian Basic Vocabulary Database (Greenhill, Blust, and Gray 2005–08). (See the appendix for a language-by-language list of data sources.)

**2. THE SETTING OF CMP AND CEMP.** The history of the CMP and CEMP proposals has been one of quiet acceptance in print, with the exception of Ross (1995); Grimes, Therik, Grimes, and Jacob (1997); and Adelaar (2005a). Grimes (1991a) and Nothofer (1992) provide additional criticisms that have not appeared in print. Blust (1993) published a series of innovations, some novel and some summarizing earlier work by him, that are claimed to characterize the languages of these two important subgroups of Austronesian. With a number of phonological, lexical, morphosyntactic, and semantic innovations characterizing CEMP, and a large number of innovations proposed to define CMP, the linguistic history of the Austronesian languages west of New Guinea appeared to be becoming as well established as that of the comparative picture of the (Oceanic) Austronesian languages in New Guinea and further east.

The “fuzziness” of the line drawn between the Western Malayo-Polynesian languages and those assigned to the Central-Eastern Malayo-Polynesian subgroup is also easily explained if we accept that many of the innovations that characterize the eastern languages are in fact the result of contact-induced change, with a long period of substratal influence resulting in the radically different typological profile that characterizes the eastern area.

The fact that a large number of the features that have been described as characterizing CEMP languages are also found in the WMP languages that border them is strongly suggestive of diffusion, because the features are not confined to a particular subgroup.<sup>7</sup> The question we must ask is where the source of that diffusion is, and the historical presence of a large Papuan population at and east of the WMP/CEMP border suggests strongly that Papuan languages should be considered to be the source. We touch on some of the evidence for this idea later in this paper, but do not focus on it here.

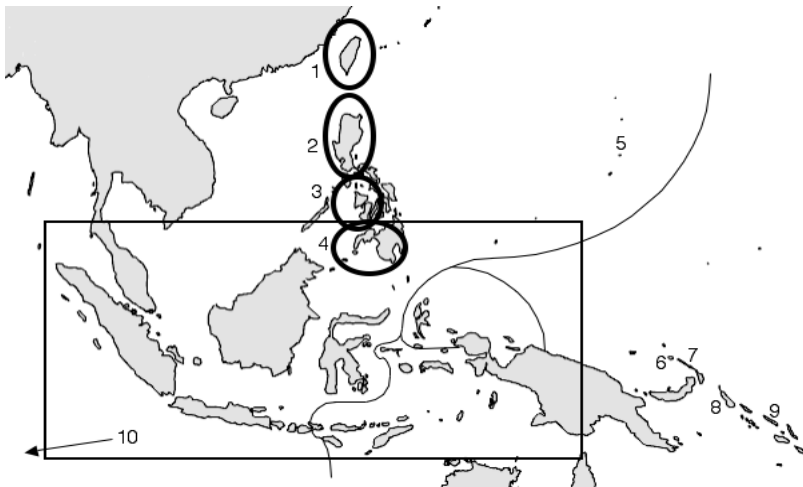
6. While we have no question about the distribution of forms such as \*kandoRa ‘cuscus’ and \*mans[aa]r ‘bandicoot’ being confined exclusively to languages near New Guinea (plus Fiji), we question whether they represent innovations. Because these terms refer to marsupials native to Sahul (New Guinea and Australia), not something in the experience of more western and northern Austronesian speakers, they could be (and likely are) borrowed from an as yet unidentified non-Austronesian source. Once established in any trade vernacular, they would be spread through all and any subgroups in contact. We find (Grimes and Donohue n.d.) only eight lexical items that are convincing for establishing the CMP area; in many cases the proposed innovations for CEMP are insufficiently attested across the CMP-area languages to allow for reconstruction, in the absence of supportable subgrouping theories for the CMP-area languages. Whether the eastern languages are the descendants of one or more than one ancestral language is immaterial, given the long-standing Papuan influence in the area (e.g., Donohue 2007a) and the likely contact between separate Austronesian linguistic traditions in the region. Finally, the value of proposed lexical innovations, in the absence of adequate lexical information for the vast majority of the languages of western Indonesia, is debatable; see section 3.

7. Alternatively, parallel development can be invoked, in which case the value of these changes for an eastern Indonesian subgroup must also be dismissed as parallel innovation, or else be provided with detailed justification for this explanation not applying. Whichever solution is adopted, the innovations do not define a CMP subgroup.

The immediate neighbors found along the western border of the proposed CMP/CEMP area are the languages of eastern Sulawesi (in three provinces: North Sulawesi, Central Sulawesi, and Southeast Sulawesi), and the Sumbawa language, in western Sumbawa.

To the south, the CMP area meets the Indian Ocean and the Arafura sea north of Australia; the east is bounded by the non-Austronesian languages of New Guinea, while to the north we find the westernmost members of the EMP group in South Halmahera; at least one other member of the EMP group, Irarutu, is in contact with CMP-area languages. The EMP group is bounded by CMP-area languages to the south of Halmahera, Papuan languages to the south and east of Cenderawasih Bay, wide stretches of ocean to the north, and a short stretch of Papuan-inhabited coastline followed by the westernmost Oceanic languages to the east of Warembori. East of Warembori all the Austronesian languages belong to the Oceanic subgroup. Maps 1 and 2 show the location of the CMP languages; the line running, erratically, from northeast to the central south in these maps marks the division between the “WMP” languages, to the west, and the “CEMP” languages in the east; the line running west from the Bird’s Head of New Guinea marks the northern extent of the “CMP” languages; between this line and the curved line stretching northwest from New Guinea are the EMP languages, in South Halmahera and Cenderawasih Bay, while to the east only Oceanic languages are found. Labels used for linguistically dense areas (e.g., 1–4 on Map 1, or the Flores area on Map 2) are purely geographic.

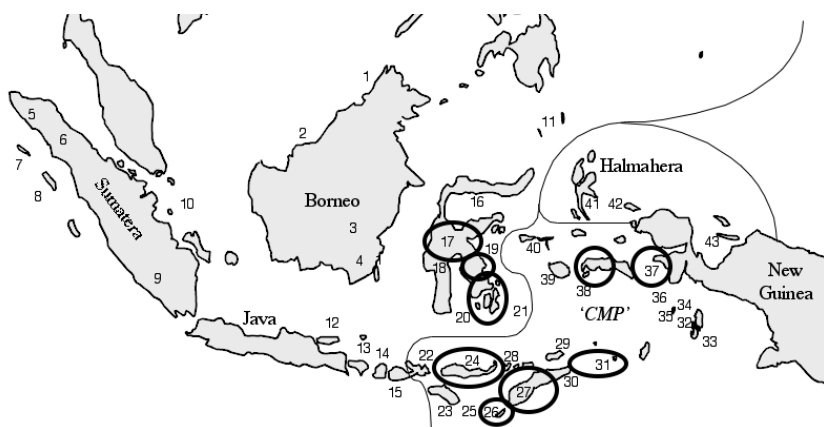
**MAP 1. LANGUAGES REFERRED TO IN THIS PAPER  
OUTSIDE THE REGION OF PRIMARY FOCUS**



- 1: Formosan languages: Bunun, Central Amis, Favorlang, Paiwan, Pazeh, Saisiyat, Thao, Yami, Iraralay  
 2: Northern Philippine languages: Arta, Ilokano, Kakiduge:n Ilogot, Tagalog  
 3: Central Philippine language: Hanunóo  
 4: Southern Philippine languages: Ba:ngingi Sama, Sarangani Blaan, Mamanwa, Maranao  
 5: Chamorro. 6: Tungak (Lavongai). 7: Nalik. 8: Taiof. 9: Varisi, Vaghua. 10: Malagasy (off map).  
 (Labels 1–4 are purely geographic.)

**3. CENTRAL MALAYO-POLYNESIAN.** Blust (1993) proposes between four and six phonological innovations (depending on whether we count glide truncation as one innovation or three) and two or three morphosyntactic innovations that define the Central Malayo-Polynesian group. These are listed in table 1.<sup>8</sup> The column headed “inclusive?” is based on Blust’s assessment of how much this is an inclusive innovation for the CMP-area languages. Thus, for instance, while postnasal voicing is explicitly acknowledged not to apply to all of the CMP-area languages, through his use of phrasing such as “in many of the CMP languages,” glide truncation is “one of the most distinctive characteristics of the phonological history of CMP languages,” implying a unified treatment; though see the comments in the introduction of this article, and his (1993:264–66) comments that show his acknowledgement of their noninclusiveness. Unless we find direct evidence to the contrary in Blust (1993), we assume that innovations proposed were intended to be changes that define all “CMP” languages, because listing them as innovations without comment as to their extent would be confusing for the reader.

**MAP 2. THE WMP/CMP DIVISION,  
AND LANGUAGES REFERRED TO IN THIS PAPER**



1: Iranun; 2: Melanau; 3: Tunjung; 4: Ma’anyan (also in Borneo: Kayan [Uma Juman], Modang, Punan Kelai); 5: Acehnese; 6: Gayo; 7: Simalur; 8: Nias; 9: Rejang; 10: Malay; 11: Sangir; 12: Madurese; 13: Balinese; 14: Sasak; 15: Sumbawa; 16: Gorontalo; 17: Central Suawesi languages: Da’a, Mori Atas, Mori Bawah, Napu, Padoe, Pamona, Uma’; 18: Mamuju, Topoiyo, Toraja (Woensdrecht); 19: Bungku-Tolaki languages: Bungku, Tolaki; 20: Insular Southeast Sulawesi languages: Busoa, Cia-Cia, Kaimbulawa, Kamaru, Kulisusu, Kumbewaha, Lasalimu, Moronene, Muna, Pancana, Taloki, Tolaki, Wawonii, Wolio; 21: Tukang Besi; 22: Bima; 23: Kambera; 24: Flores area languages: Lio, Manggarai, Nage, Ngadha, Palu’e, Sika, So’a; 25: Hawu; 26: Rote languages (Dela, Tii, Lole, Termanu, Rikou), Dhao; 27: Timor languages: Atoni (= Uab Meto), Kemak, Mambai, Tetun; 28: Kédang; 29: Tugun; 30: Fataluku; 31: Southwest Maluku: Erai (Wetar), Kisar, Leti, Oirata; 32: West Tarangan; 33: Batuley; 34: Banda (Elat); 35: Kei; 36: Watubela/Matabello; 37: Eastern Maluku: Arguni, Geser, Onin, Sekar; 38: Ambon area languages: Alune, Amahai, Hitu, Nuaulu, Paulohi; 39: Buru; 40: Soboyo (Taliabu); 41: Sawai; 42: Gebe; 43: Ansum

8. Blust does not claim that glide truncation applied to PCMP, but was rather “a product of independent changes in a number of languages” (1993:264), which should render its value in subgrouping nil. We present the changes in this table because glide truncation is “one of the most distinctive characteristics of the phonological history of CMP languages” (Blust 1993:264).

**3.1 PHONOLOGICAL INNOVATIONS FOR CMP.** Blust proposes four phonological innovations attested in CMP-area languages: the loss of antepenultimate syllables not beginning with an oral consonant; the truncation of diphthongs; the voicing of stops following nasals; and the change of the antepenultimate vowel in \*baqəRu ‘new’ to a schwa, thus \*bəqəRu.

Blust (1993:264) admits that the loss of antepenultimate \*qV and \*hV cannot have been a feature of PCMP, based on evidence from CMP-area languages such as we present in the following section. He does state that “prepenultimate initial vowels that were initial (not preceded by a laryngeal) evidently have disappeared in all CMP languages, and hence presumably in PCMP,” but offers no examples of this class of words, and does offer \*anaduq ‘long’ as a Proto-Central Malayo-Polynesian reconstruction (1993:283), with a prepenultimate initial vowel.

Regarding the truncation of diphthongs, Blust (1993:264) notes that “as with the loss of prepenultimate initial vowels, we are forced to conclude that this common CMP merger—widespread and distinctive as it is—was also a product of independent changes in a number of languages.” In the following sections we examine the claims that the different phonological features put forward by Blust have for exclusiveness, and inclusiveness, in terms of defining the CMP-area languages as a genetic entity.<sup>9</sup>

**3.1.1 \*hVσσ / \*qVσσ / \*Vσσ > \*σσ.** The loss of an antepenultimate syllable when the onset of the first syllable is \*q, \*h, or  $\emptyset$  is erratically present outside the purported CMP/CEMP region in Southeast Sulawesi. Table 2 shows the development of antepenultimate \*qV and \*hV in Tukang Besi (no reflexes of \*Vσσ are known); while \*qV is preserved (with a reduced vowel), \*hV is lost (recall that PAN \*S > PMP \*h). Note that an initial \*q in an antepenultimate syllable is erratically lost; this is significant, given that it is regularly preserved in the final two syllables of a word. (The term for ‘egg’ in Tukang Besi is not cognate, but has only been included for completeness.)

The loss of the antepenultimate syllable in ‘far’ and ‘west’ is also present in a number of Bungku-Mori-Tolaki languages (also bordering on the CMP/CEMP region), where \*habaRat is reflected as *bara*, and \*Sa-Lawid as *lai*. Note that in the words that do *not*

**TABLE 1. BLUST’S PROPOSED CMP-DEFINING INNOVATIONS**

	REGULAR PHONOLOGICAL INNOVATIONS	INCLUSIVE?
1	PMP *hVσσ, *qVσσ, *Vσσ > PCMP *σσ	no
2	Glide truncation	no
2a	PMP *ay > *a	no
2b	PMP *aw > *a	no
2c	PMP *uy > *u	no
3	Postnasal voicing: NT > ND	no
IRREGULAR PHONOLOGICAL INNOVATIONS		
1	PMP *baqəRu > PCMP *bəqəRu ‘new’	no

9. Blust also proposes 28 lexical innovations that define the CMP group. Grimes and Donohue (n.d.) show that only eight of these, attested in (on average) two or three languages each, stand up to scrutiny. Given the paucity of lexicographic materials from the WMP-area languages, or from the EMP languages, and the consequent problems that these innovations face for verification, these lexical data will not be further discussed here.



TABLE 2. THE FATE OF \*qVσσ, \*hVσσ, AND \*Vσσ  
IN TUKANG BESI

	PAN/PMP	TUKANG BESI
‘egg’	*qatəluR	(goraʔtu)
‘sun’	*qaləjaw	ʔoloo
‘spleen’	*qəpəju	hoʔou (via *pəqə(j)u)
‘ghost’	*qanitu	ʔonituu
‘centipede’	*qalipan	__oliha
‘far’	*Sa-Lawid	(me-)_lai
‘west’	*habaRat	__βaha

show loss of the antepenultimate syllable, the vowel is completely predictable. The languages of the Southeast Celebic group regularly show the reduction of antepenultimate vowels to schwa, followed by the regular \*ə > o change.<sup>10</sup> The reduction of a vowel to schwa can be considered a step in the direction of the loss of that vowel (and any associated consonants), and so we find the roots of the eastern Indonesian antepenultimate syllable loss in Southeast Sulawesi (and further west, as described below). The fact that the actual loss of syllables is not well advanced in Southeast Sulawesi indicates that this cannot be considered to be a borrowing from the east, but rather is a related change.

In at least one form we can see a similar process operating at the PMP level, where PAN \*Suai (also \*Sua[n]ji) ‘younger sibling’ developed into PMP as \*wai, rather than the expected \*\*huai. The \*w in PMP presumably reflects the reanalysis of an earlier epenthetic glide (presumably \*huai > \*[huwai] > \*wai). Additionally, many other languages scattered throughout the Formosan and WMP areas also show reduction of the antepenultimate syllable in \*qVσσ, \*hVσσ, and \*Vσσ roots. A small sample is shown in table 3. (The phonetic glottal stop in, for instance, Tagalog *araw* [ʔaraw] does not reflect the earlier \*q, but can be argued to be purely epenthetic synchronically, and certainly innovative diachronically, because it also appears on words with no \*q etymology.)

Crucially, we also see the retention of the antepenultimate syllable in a number of CMP-area languages. It is these very retentions that led Blust (1993: 280–84) to posit the retention of these syllables in PCMP. We agree with this assessment, and so dismiss this as a change that defines the CMP-area languages. Rather, the change from a trisyllabic root (with nonoral initial coda) to a disyllabic root has been in process at least from PMP times, with the change proceeding geographically and lexically (see table 4).<sup>11</sup>

While we have no attestations of the antepenultimate syllable of \*qasiRa ‘salt’ in a CMP-area Austronesian language, we do have the forms *asi* ‘salt’ in Makasai, *asir* ‘salt’ in Oirata, and *asir(u)* ‘salt’ in Fataluku, three Papuan languages from the eastern end of

10. Of course, from a synchronic phonological perspective this means that antepenultimate vowels are unspecified for any features other than syllabicity, because the [o] articulation is completely predictable. Nonetheless, from a diachronic standpoint, they show a change to o, most likely via ə. Within the CMP region itself, Grimes has also noted predictable behavior of the antepenultimate vowel in Buru (1991b), Hawu and Dhao (forthcoming), and for Rote languages such as Dela, Lole, Rikou, Termanu, and Tii (fieldnotes). All are reflected in different ways (either template-driven as /e/ or /a/, or as a vowel copy of the penultimate vowel), but share with the Southeast Sulawesi languages the synchronic feature of being unspecified for any features other than syllabicity.

Timor. The existence of these forms in the non-Austronesian languages, clearly loans from an Austronesian source (or sources) reflecting \*qasiRa, suggests a source that reflects at least the vowel of the original antepenultimate syllable.

**TABLE 3. REFLEXES OF \*qVσσ, \*hVσσ, AND \*Vσσ  
IN LANGUAGES WEST OF THE CMP-AREA**

	PAN/PMP		LANGUAGE
'egg'	*qatəluR	<b>qetsilu</b>	Paiwan
		<b>qaricuy</b>	Thao
		<b>__ateluy</b>	Ma'anayan
		<b>__təluR</b>	Malay
		<b>__taluh</b>	Balinese
		<b>__təloq</b>	Sasak
		<b>__talo</b>	Tunjung
'salt'	*qasiRa	<b>?asila?</b>	Bunun
		<b>qatia</b>	Paiwan
		<b>__siə</b>	Sasak
		<b>__sira</b>	Malagasy
		<b>__sia</b>	Melanau
'hand, arm'	*[qal]ima	<b>qalima</b>	Mamanwa
		<b>?ali:mah</b>	Aklanon-Bisayan
		<b>__lima</b>	Paiwan
		<b>__rima</b>	Favorlang
		<b>__lima</b>	Wawonii
		<b>__lima</b>	Balinese
'shoulder'	*qabaRa	<b>?afala</b>	Central Amis
		<b>?abaxán</b>	Pazeh
		<b>__abaga</b>	Ilokano
		<b>__aβaa</b>	Wolio
		<b>__pala</b>	Balinese
		<b>__waga</b>	Maranao
'sun, day'	*qaləjaw	<b>qadaw</b>	Paiwan
		<b>γoleo</b>	Muna
		<b>__áraw</b>	Tagalog
		<b>__dəa</b>	Modang
		<b>__daw</b>	Punan Kelai
'long'	*inaduq	<b>?inaro?</b>	Saisiyat
		<b>__laɖuq</b>	Paiwan
		<b>[ma-]naro</b>	Iraralay
		<b>__nandu</b>	Sangir
		<b>__naru</b>	Gayo
		<b>[a-]nau</b>	Nias
		<b>__aru</b>	Kayan (Uma Juman)
'west'	*habaRat	<b>__barat</b>	Malay

11. We assume the morpheme division as shown in the Kemak form *m-alaru* 'long' in table 4. An alternative interpretation of this form is that it represents the 'stative' prefix \*ma- attached to \*anaduq, which has reduced to \*naduq via antepenultimate syllable loss; the morpheme breaks would then be *ma-laru*, not *m-alaru*. Antoinette Schapper (pers. comm.) reports *monaru* for this lexeme (and *lelo* 'sun', *meta* 'black'—see table 9), which suggests either changes in antepenultimate vowels in some dialects, or that \*ma- has separately developed to \*mo-, as attested in many Sulawesi languages.

Just as antepenultimate laryngeal syllables are erratically lost in Southeast Sulawesi (and elsewhere in the WMP world— see, for example, Grimes and Grimes [1987:174–75]), so too are they erratically retained in the CMP area.<sup>12</sup>

**3.1.2 \*ay > a, \*aw > a, \*uy > u.** Similar to the facts of antepenultimate reduction, glide truncation is present to various degrees in different languages. The AN languages of eastern Indonesia and East Timor truncate original diphthongs in several different ways: retaining the vowel and dropping the glide, dropping the vowel and syllabifying the glide, or merging the two. This is shown in abstract form in table 5.

**TABLE 4. REFLEXES OF THE INITIAL SYLLABLE OF \*qVσσ, \*hVσσ, OR \*Vσσ IN CMP-AREA LANGUAGES**

	PMP		LANGUAGE
'egg'	*qatəluR	<b>kat</b> lu	Watubela
		__atulú	Matabello
		__t[ə]llo	Palu'e
		__dəl:u	Hawu
'sun'	*qaləjaw	<b>hal</b> a	Kemak
		__olēr	Matabello
		__/[ə]ra	Palu'e
		__lea	Buru
'shoulder'	*qa-baRa	<b>ka</b> fár	Tugun
		kbaa-s	Tetun
		__faha-n	Buru
		__var	Kei
'west'	*habaRat	__aφara	West Tarangan
		__aφár	Batuley
		__φaha	Kola
		__fahat	Buru
		__va	Palu'e
'long'	*anaduq	m- <b>al</b> aru	Kemak
		__naru	Bima
		ka-__naru-k	Tetun
		ma-__nalú	Termanu

**TABLE 5. PATTERNS OF GLIDE-TRUNCATION IN THE CMP-AREA**

	RETAIN V	SYLLABIFY GLIDE	COALESCENCE
*-uy	u#	i#	?
*babuy	babu	babi	—
*-ay	a#	i#	e#
*pajay	paja	paji	paje
*-aw	a#	u#	o#
*namaw	nama	namu	namo
*-iw	i#	u#	?
*laRiw	lari	laru	—

12. The final vowel in Fataluku is erratic; in addition to *asiru*, we also find *asira* and *asir*. These and other aspects of Fataluku grammar are still under investigation. It might be argued that Makasai, Oirata, and Fataluku borrowed their terms from a pre-PCMP Austronesian language. The problem with this interpretation is that the evidence is compelling that these languages arrived in East Timor after the Austronesian languages, in a back-migration from the Bomberai peninsula (see, e.g., O'Connor 2003). This makes inferences about the putative pre-PCMP Austronesian language hard to evaluate.

As table 6 shows, \*ay and \*aw are reflected as *e* and *o* in the Southeast Sulawesi languages, but the fate of \*uy shows consistent glide truncation in some languages: Tukang Besi, Kamaru (closely related to Wolio), and erratically in Cia-Cia.

Blust (1993:266) objects to attempts to equate the suite of changes that is his glide truncation with the occurrence of sporadic glide truncation in other languages on the grounds that it is the entire suite of diphthongs that must be attested in order to establish glide truncation: “it would be necessary to demonstrate not only the change \*-uy > /u/, but also the changes \*-ay > /a/, \*-aw > /a/ and (if examples can be found) \*-iw > /i/.” This is indeed a high benchmark to set, one that none of the CMP languages he presents have achieved. The point we are making is that similar patterns of glide truncation are also widely attested in WMP, just as they are in a number of CMP area languages. The only area in which glide truncation for \*uy, \*ay, and \*aw is universally attested for all of the words cited by Blust (1993:265) is in central Maluku (no data on \*iw is presented in Blust 1993, though forms such as Palu’e *kadɟu* ‘wood’, Hawu, Dhao *afu*, Buru *kau*, Soboyo *kayu*, ‘wood’ indicate that the high back rounded element of \*iw was preserved widely across the CMP region).<sup>13</sup> Rather than there being a single process of “glide truncation,” it seems much more likely that there are (at least) three separate processes that can occur and that result in the glide being lost without obvious trace: \*aw > a, \*ay > a, and \*uy > u.<sup>14</sup> As can be seen in map 3, there is no reason to believe that these are connected (though note the interesting fact that the truncation of \*ay > a implies the existence of \*aw > a: all languages for which \*ay > a has applied also show \*aw > a, but the converse is not true).<sup>15</sup> The history of \*uy appears to be independent of the history of the other two glides considered here, as does that of \*iw. In short, there is no evidence that the different truncation processes are related, and the degree, if not the manner, to which it is found in the CMP-area languages is mirrored in their relatives to the west in Sulawesi. (No attempt has been made to map the vowels of the languages of Sumatera, Borneo, Java, Malaya, the Philippines, or New Guinea.)

While glide truncation is found in a number of CMP-area languages, it is not found in all lexical items displaying all three diphthongs investigated except in central Maluku; at the same time, glide truncation is found outside the CMP area in the \*uy diphthong, and a

TABLE 6. THE FATE OF \*ay, \*aw, AND \*uy IN SOUTHEAST SULAWESI

	*ay	*aw	*uy	*uy	*uy
	‘chin’	‘sun’	‘fire’	‘pig’	‘swim’
PMP	*qaZay	*qaləjaw	*hapuy	*babuy	*naŋuy
Muna	ase	ɣoleo	ifi	βeβi	leni
Cia-Cia	hae	holeo	api	βaβi	pika-naŋu
Tukang Besi	ase	ʔoloo	ahɯ	βaβɯ	naŋɯ
Kamaru	ase	eu	apu	baβu	po-naŋu

13. An analysis in which PMP \*kahiw > \*kayu following the loss of \*h and the subsequent resyllabification of \*\*kaiw cannot hold, because Soboyo at least generally has a nonzero reflex of PMP \*h (Blust 1981). In any event, the languages cited do not reflect \*kahi, indicating that any truncation of \*-iw must have been a relatively late, and independent, process.

14. There are, of course, many languages in which the glide has been lost, but has left traces in the remaining vowel, such as \*ay > i, or \*ay > e.

15. We thank Malcolm Ross for this observation.

great number of languages in the CMP area clearly require the positing of all three diphthongs, \*ay, \*aw, and \*uy.

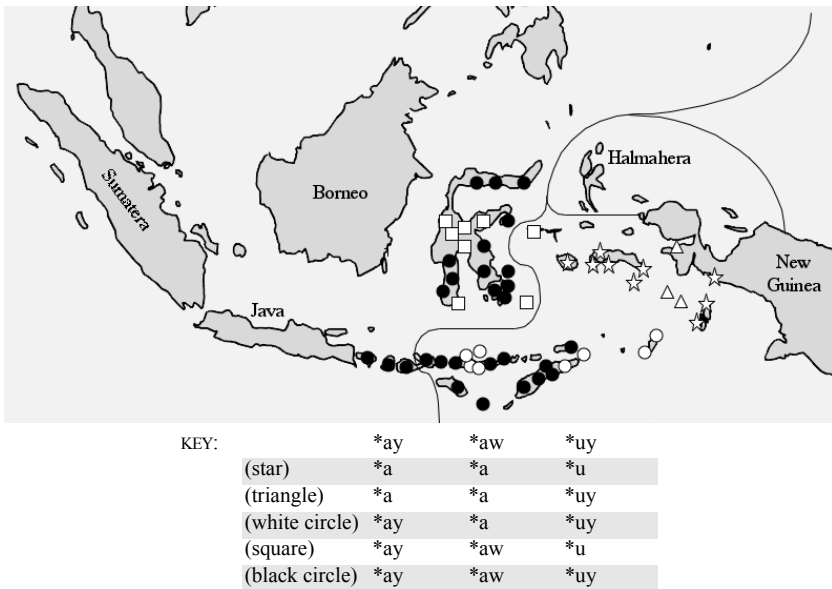
**3.1.3 Postnasal voicing.** The claim is that “stops have been voiced after nasals” (Blust 1993:266). A case for postnasal voicing can only be made convincingly with respect to originally \*NT clusters,<sup>16</sup> because the stop in a cluster such as \*mb being reflected as *b* says nothing about a process of postnasal voicing applying or not, and because the appearance of a \*ND cluster as D can simply reflect the simplification of syllable structure. On the other hand, \*mp > *mb* (for instance) does demonstrate a change; and the appearance of \*NT as D, on the other hand, does imply some degree of fusion.

There are at least six logical ways to resolve a cluster of the form \*NT:<sup>17</sup>

1. the NT sequence is preserved as NT;
2. voicing spreads to produce ND;
3. the nasal is lost, leaving T;
4. the plosive is lost, leaving N;
5. the feature [voice] from the nasal is merged in the plosive, leaving D; and
6. the consonant cluster is preserved as gemination, leaving TT.

All of these options are attested. Table 7 shows data illustrating the different changes in a selection of languages (no attempt has been made to be exhaustive) for the lexemes

**MAP 3. DISTRIBUTION OF PATTERNS OF GLIDE TRUNCATION IN EASTERN INDONESIA AND EAST TIMOR**



16. In this section we use the conventions T, D, and N to refer to any voiceless stop, voiced stop, and nasal, respectively, regardless of place of articulation.

17. Iranun, which possibly reflects \*dumpul as *tebpul*, could conceivably be described as showing \*NT > DT, a seventh pattern. The correspondences are not by any means clear, however.

\*dumpul ‘blunt’ (with the exception of Bugis *utti*, Rote *huni*, Tetun *hudi*, Kowiai *funu*, and Imroing *ut* ‘banana’, which reflect \*punti, because the forms for ‘blunt’ in these languages are not cognate).

Examples (not exhaustive) of the different changes described above have applied as follows; CMP-area languages have been shown in italics.

1. \*mp > mp: Malay, Gayo, Madurese, *Manggarai*
2. \*mp > mb: Malagasy, *Kambera*, *Arguni*
3. \*mp > p: Maranao, Rejang, *Imroing*, *Helong*, Varisi
4. \*nt > n: *Rote*, *Kowiai*, *Amarasi*, *Baikeno*
5. \*mp > b: *Palu’e*, *Selaru*, *Buru*, *Tetun*, *Dhao*, *Hawu*
6. \*nt > tt: Bugis, Kalumpang, Seko, Masenrempulu

Note that the reduction of an NT sequence to (N)D is not restricted to the CMP-area languages, and that strategies other than \*NT > D are attested in CMP-area languages as well. While the \*NT > (N)D process is dominant in the CMP area, it is by no means exclusive to that area, nor is it inclusive of all languages in the CMP area. Note also that the CMP area shows the greatest diversity of means of resolving \*NT sequences, showing all of the hypothesized strategies.

We should note that there are languages, such as Malay, in which \*NT is preserved as NT in lexical roots, such as *tumpul* ‘blunt’, but is not allowed as the result of morphophonemic processes. Thus a morpheme-internal NT sequence is tolerated, as in (1), but not when it is the result of the active prefix *məŋ-* being added to a labial-initial root. If the root-initial consonant is voiced the nasal assimilates to the place of articulation of the fol-

**TABLE 7. THE RESOLUTION OF \*NT SEQUENCES IN AUSTRONESIAN LANGUAGES: \*dumpul ‘BLUNT’ AND \*punti ‘BANANA’**

PMP	LANGUAGE	AREA	
*dumpul	Hanunóo	Philippines	dúmpul
	Maranao	Philippines	tepol
	Rejang	Sumatera	topoa
	Malay	Sumatera	tumpul
	Gayo	Sumatera	tumpul
(*punti	Malagasy	Madagascar	dómbo
	Madurese	CMP-area	tompol
	Bugis	Sulawesi	utti)
	Palu’e	CMP-area	tubu
	Kambera	CMP-area	katumbul
(*punti	Manggarai	CMP-area	dumpul
	Rote	CMP-area	huni
	Tetun	CMP-area	hudi
	Kowiai	CMP-area	funu
	Imroing	CMP-area	ut)
*dumpul	Arguni	CMP-area	tumbie
	Selaru	CMP-area	tubal
	Varisi, Vaghua	Solomons	ndɔpu

lowing stop, as in (2), because ND sequences are permitted. If the initial consonant is voiceless the assimilation takes place, but the stop is lost, as in (3).<sup>18</sup>

- |     |             |                        |               |
|-----|-------------|------------------------|---------------|
| (1) | /tumpul/    | [tumpul]               | ‘blunt’       |
| (2) | /məŋ+buaŋ/  | [məmbuaŋ], *[məmuaŋ]   | ‘throw away’  |
| (3) | /məŋ+panah/ | [məmanah], *[məmpanah] | ‘shoot arrow’ |

The fact that even in languages that preserve \*NT sequences the constraint against producing nonlexical NT sequences is operative implies that this is a very ancient and pervasive feature of the family. The fact that \*NT > D and \*NT > T are attested in the Philippines indicates that the dispreference for NT clusters is present as a PMP-level phenomenon, though the degree to which it is prominent in any given language or area varies greatly.

The other environment in which N+T sequences are found as D crosses morpheme boundaries. As Blust points out, for instance, Buru, Kemak, and Mambai reflect \*matuq ‘white’ as *boti*, *buti*, and *buti*, respectively, involving the reduction of \*ma- to \*m-, and subsequent merger of features to yield *b* from \*m-p. This is a rare change in CMP-area languages, and possibly even rarer outside the area (but see Kakiduge:n Ilongot *budək* ‘white’).<sup>19</sup>

**3.1.4 \*baqəRu vs. \*bəqəRu.** Blust (1993) claims that “a number” of the CMP languages do not reflect PMP \*baqəRu ‘new’, but require the positing of an intermediate form, \*bəqəRu, with a reduced vowel in the antepenultimate syllable. The data on which he bases this reconstruction (Blust 1993: 266) are shown in table 8, supplemented with additional relevant forms.<sup>20</sup> We have added reflexes of \*təlu ‘three’ in these languages to demonstrate the fact that, in addition to reflecting \*u finally, the other vowel in ‘new’ reflects a schwa, and not an \*a.

The data are clear, but the motivation for an irregular phonological development relies on the identification of the vowel following the initial labial consonant as reflecting \*ə, and not \*a, following this consonant. Some CMP-area data clearly reflect \*a and cannot be reflexes of \*\*bəqəRu, such as Helong *balu*, Luang *war-waru*, Selaru *har-har*<sup>20</sup>.

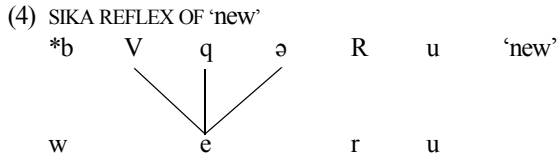
For those that clearly do not reflect \*a, the obvious problem is in identifying which vowel is retained. So while a form such as Sika *weru-ŋ* ‘new’ clearly reflects \*b as *w*, and \*R as *r*, thus assigning the final *u* of the root as a regular reflex of \*u, the *e* is all that is left of \*aqə (or \*əqə): does this *e* reflect \*ə, \*a, a generic, undifferentiated antepenultimate vowel, or the sequence \*aqə (or \*əqə)?

18. Compare with Arta, from the northern Philippines (Reid 1989), as an example of a language in which the constraint \*NT does not feature prominently in the language at all, as evidenced by the fact that [mp] sequences are permitted both lexically and as a result of morphological processes.

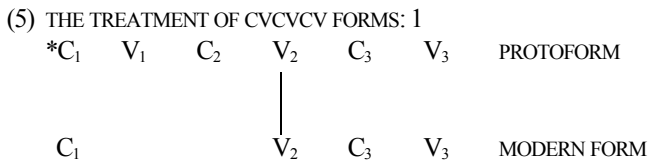
- |       |            |                      |              |
|-------|------------|----------------------|--------------|
| (i)   | /tumpu/    | [tumpu]              | ‘swim’       |
| (ii)  | /maŋ+alla/ | [maŋalla]            | ‘laugh’      |
| (iii) | /maŋ+bayú/ | [mambayú], *[mamayú] | ‘pound rice’ |
| (iv)  | /maŋ+pili/ | [mampili], *[mamili] | ‘choose’     |

19. The transfer of voicing is also reminiscent of similar processes found in Vanuatu languages such as Raga and Southeast Ambrym (Crowley 2002a, b).

20. Blust lists the Tetun form incorrectly as *foa-n*. The correct form is *foun* (see also Morris 1984:36) in all dialects of the vernacular Tetun, as well as in Tetun Dili (a Tetun-based creole).



When we examine other words with trisyllabic reconstructions, we frequently find reductions of the form shown in (5): the initial consonant is preserved, but the vowel associated with the antepenultimate syllable in the protoform is not. The second consonant is lost, but the vowel of the penultimate syllable is preserved. This pattern is clear in the Manggarai, Sika, and Kédang forms for 'black', < \*ma-qitVm via \*mitVm, and in the reflexes for 'alive' in Manggarai and possibly Kemak, < \*ma-qucip via \*mudip, shown in table 9 (the languages were selected on the basis of their having been used by Blust [1993] to illustrate the \*baqəRu > \*bəqəRu change, with additional languages appended at the end to give a fuller picture).



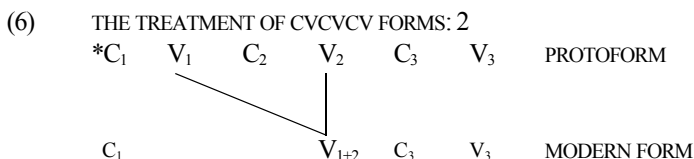
A second pattern sees the initial vowel of the modern disyllable reflecting the fusion of the original antepenultimate vowel with the original penultimate vowel, as schematized in

**TABLE 8. BLUST'S EVIDENCE FOR \*baqəRu > \*bəqəRu (SUPPLEMENTED)**

LANGUAGE	AREA	'new'	'three'
Bima	east Sumbawa	ʔbou	tolu
Manggarai	west Flores	wera	telu
Sika	east Flores	weru-ŋ	telu
Kédang	Solor	weru-n taʔe-n	telu
Hawu	west of Timor	viu	təl:u
Dhao	west of Timor	hiu	təl:u
Rote	west of Timor	beu-k	telu
Dela	west of Timor	feu	telu
Termanu	west of Timor	beu	telu
Amarasi	West Timor	feʔu	tenu ~ teun
Baikeno	West Timor	feʔu	tenu ~ teun
Uab Meto	West Timor	feʔu	tenu
Helong	West Timor	balu	tilu ~ tiul
Tetun	Central Timor	fou-n	tolu
Kemak	Central Timor	he:u	telu
Mambai	East Timor	heu	teul
Erai	Southwest Maluku	ha-heru-n	telu
Kisar	east of Timor	woru-woru	wo-kelu
Elat	east Maluku	feru-no	telu
Paulohi	central Maluku	heru-i	toru
Hitu	central Maluku	helu	telu
Alune	central Maluku	belu-ke	telu
Buru	west of Timor	fehu-t	telu



(6). This is possible in the forms for ‘black’ in Uab Meto (= Atoni), Tetun, Kemak, Mambai, Erai, Kisar, Elat, Paulohi, Hitu, and Alune, depending on whether we assume the protoform to contain an \*i or an \*e, and in the reflexes for ‘alive’ in Bima, Manggarai, Sika, Rote, Uab Meto (= Atoni), Tetun, Erai, and Kisar. (In many of these another possible interpretation is that *o* reflects \*u, the penultimate vowel of the original form, with no fusion, thus exemplifying the type seen in [5] above.) Note that the forms illustrating these different treatments of vowels are not uniform in the one language: while Manggarai preserves the \*i unchanged in *miteŋ*, it shows fusion of \*a and \*u in *mosé*.



With this information we are better able to evaluate the proposed innovative form for ‘new’. While the penultimate vowel in the modern form clearly reflects a schwa, not an \*a, do we have any evidence that the schwa that is reflected was the vowel of the antepenultimate syllable? In short, no. We have an entirely plausible, independently motivated, and elsewhere attested process in which the antepenultimate vowel is lost, and only the penultimate vowel preserved. There is no reason to posit an innovative lexical form.<sup>21</sup>

It is worth noting that similar surface behavior is found in the languages of Southeast Sulawesi. Table 10 parallels table 8, but with data from Southeast Sulawesi.<sup>22</sup>

**TABLE 9. THE TREATMENT OF REPRESENTATIVE \*CVCVCV ROOTS IN CMP-AREA LANGUAGES**

LANGUAGE	‘black’ *ma-q[ie]təm	‘alive’ *ma-quɔp
Bima	meʔe	mōri
Manggarai	miteŋ	mosé
Sika	mitaŋ	moret
Kédang	miteŋ	(bitan)
Termanu	(ŋgèò)	moli
Amarasi	meta-n	an-mōni ~ an-mōin
Tetun	meta-n	mōris
Kemak	meta-ma	mwasā
Mambai	meta-n	mwir
Erai	meta[m]	mōri
Kisar	memekem	mōri
Elat	metemeten	munuli
Paulohi	mete	mahaïti
Hitu	mete	mahai
Alune	metene	(ʔʷana)
Buru	mite-t	—
Hawu	—	muri

21. Grimes (1991a) also concluded that the alternate is unnecessary: all the data can be accounted for by \*baqəRu.

22. Italics indicate noncognate forms; the antepenultimate *u* in Kumbewaha, Lasalimu, Cia-Cia, and Muna represents a semi-regular development of \*ə following a bilabial. The South Sulawesi data (see Grimes and Grimes 1987:138–39) overwhelmingly reflect \*baqəRu.

Superficially it appears that the same innovation has taken place in Southeast Sulawesi, but in fact this simply reflects the regular process of antepenultimate vowel reduction to ə, followed by the regular \*ə > o sound change that characterizes all of these languages (in all cases \*a is regularly reflected as a in these languages). This path of development is shown in (7), describing the development generally; (8) describes the protoform and contemporary forms of ‘new’ in Tukang Besi. Note particularly that, as a result of neutralization of vowels in antepenultimate positions, we cannot decide whether the modern Southeast Sulawesi forms are descended from \*baqəRu or \*bəqəRu. But because they *can* be derived from \*baqəRu, parsimony requires us not to posit an alternative form.

## (7) THE TREATMENT OF CVCVCV FORMS: 1

*C <sub>1</sub>	V <sub>1</sub>	C <sub>2</sub>	V <sub>2</sub>	C <sub>3</sub>	V <sub>3</sub>	PROTOFORM
C <sub>1</sub>	ə	C <sub>2</sub>	V <sub>2</sub>	C <sub>3</sub>	V <sub>3</sub>	INTERMEDIARY FORM
C <sub>1</sub>	o	C <sub>2</sub>	V <sub>2</sub>	C <sub>3</sub>	V <sub>3</sub>	MODERN FORM

## (8) TUKANG BESI REFLEX OF ‘new’

*b	V	q	ə	R	u	‘new’
β	o	ʔ	o	—	u	

In Southeast Sulawesi, then, we see the reduction in quality distinctions in antepenultimate syllables. This appears to be related to the patterns seen earlier in both the CMP-area languages and the Southeast Sulawesi languages. We have seen in 3.1.1 a process of antepenultimate reduction in the form of the sporadic loss of entire syllables. The evidence shown in this section suggests that a general tendency towards disyllabicity (that is, preservation of the canonical Austronesian root shape as a word shape) acted to reduce trisyllables. This began with the neutralization of vowel contrasts in such syllables, and proceeded with the loss of parts of those syllables, or else the entire syllable (if \*qV-, \*hV-, or \*V- initial), with whole syllable loss proceeding lexically and geographically but also sporadically.

**TABLE 10. PSEUDO-EVIDENCE FOR \*baqəRu > \*bəqəRu  
IN SOUTHEAST SULAWESI LANGUAGES**

LANGUAGE	AREA	‘new’	‘three’ (< *təlu)
Tukang Besi	southeast edge	βo’ou	(to)tolu
Kumbewaha	east Buton	βukou	totolu
Lasalimu	east Buton	βukou	totolu
Cia-Cia	central Buton	βukou	totolu
Muna	Muna	buyou	tolu
Pancana (Kapontori)	west Buton	bu’ou	totolu
Kaimbulawa	southwest Buton	bohoh	totolu
Busoa	southwest Buton	βohoh	toluayo
Taloki	northwest Buton	tonia	otolu
Kulisusu	north Buton	tonia	otolu
Wawonii	Wawonii island	(sarai)	otolu
Moronene	southwest mainland	(tonia)	otolu
Tolaki	southern mainland	βo’ohu, βu[ʔ]ohu	tolu
Bungku	northeast mainland	sarai	otolu
Mori Atas	central Sulawesi	βo’u	otolu
Mori Bawah	central Sulawesi	βo’ohu	otolu

**3.2 INTERIM SUMMARY: THE PHONOLOGICAL EVIDENCE FOR CMP.** Table 11 summarizes the discussion in this section. As can be seen, none of the phonological features proposed are exclusive to the languages of the CMP area, and none of them are inclusive of all of these languages. In short, when combined with information about the location of the languages that do reflect various of these proposed innovations, none of them argue for the existence of a CMP protolanguage ancestral to all of the languages currently considered to be CMP languages, either as a single language or as a differentiated linkage. By “sporadic” we mean that a feature does not occur in all (or necessarily even most) languages in the area; or it does not occur in every region within the area; or there are known exceptions; or the data in favor of a feature may occur side by side with exceptions to that feature, rather than removed by great distances. Parentheses around “yes” and “no” mean that the claim cannot be made unequivocally, or that there are known problems with making the stronger claim.

**TABLE 11. THE PHONOLOGICAL EVIDENCE FOR CMP**

	WEST OR NORTH?	SULAWESI?	CMP-AREA?	EAST?
Antepenultimate reduction (3.1.1)	sporadic	*hVσσ	sporadic	sporadic
Glide truncation (3.1.2)	(no)	(yes)	sporadic	(no)
Postnasal voicing (3.1.3)	sporadic	sporadic	sporadic	sporadic
*baqəRu > *bəqəRu (3.1.4)	no	(yes)	(no)	no

**3.3 CMP: MORPHOSYNTAX AND LEXICON.** Although we do not discuss Blust’s proposed morphosyntactic innovations, we do not believe that any of them are convincingly suggestive of a single innovation such as would characterize a single ancestral protolanguage. The data on preverbal pronominal prefixes (with reconstructions \*ku- ‘1SG’, \*mu- ‘2SG’, \*na- ‘3SG’, \*ma- ‘1PL.EXCL’, \*ta- ‘1PL.INCL’, \*mi- ‘2PL’, and \*ra- ‘3PL’) is more complex than Blust describes, being neither inclusive of nor exclusive to the CMP-area languages (Grimes and Donohue n.d.).<sup>23</sup> Many languages in the CMP area have no preverbal pronominal prefixes inflecting any verbs at all. In table 12 we see a small but representative selection of languages showing the realization of Blust’s proposed PCMP prefixes, with likely cognate forms shown in bold. While Kambera and Sika show perfect matches for all seven person/number combinations, Leti to the east of Timor has an innovative 3PL, Kei shows only four out of seven cells matching the reconstructed forms, and Amahai (from central Maluku) shows only three. The noncognate forms are in most cases easily interpreted as reductions from a free pronoun form, though for some languages (e.g., Buru) this is not a plausible explanation. Sawai and Ansus, languages typifying the SHWNG branch of EMP, show radically differing affixal forms, with the Sawai forms (from Halmahera, immediately north of the CMP-area) agreeing closely with Blust’s PCMP prefixes, and the Ansus forms, from Cenderawasih Bay, more separated from the CMP-area languages, showing completely different forms. The evidence from the CMP-area languages is that prefixal agreement was innovated a num-

23. Particularly note Wolff (1996), building on earlier work by Starosta, Pawley, and Reid (1981), who points out that, with the genitive clitics appearing in second position and auxiliaries preceding verbs, all that was required was the loss of overt forms for many of the auxiliaries for the genitive enclitics to become verbal proclitics, and later prefixes.

ber of times; the formal similarities can be accounted for by the fact that in most cases the prefixes adopted the form of the genitive clitics of PMP.

The unexceptional nature of the CMP-area correspondences becomes obvious when we examine prefixal agreement in other Malayo-Polynesian languages, a selection of which are shown in table 13, with Chamorro, two languages of Sumatera, two Sulawesi languages, and Sumbawa, from just west of the WMP/CEMP line (the locations of these languages are shown in map 2). The Nias (realis) prefixes match Blust's PCMP prefixes better than many CMP-area languages, and correspondences are similarly easy to find in the other languages shown here. Da'a (in Central Sulawesi), in the process of developing prefixes, shows similarly cognate forms where it has prefixes. Clearly the development of prefixal agreement is wider than the CMP area, and does not represent a single innovation within it.<sup>24</sup> (The presence of a distinct 3PL form in Toraja recorded by Woensdrecht, and reported in Haaksma 1933, is exceptional among the South Sulawesi languages, which generally have no number distinction in the third person.)

**TABLE 12. SOME PRONOMINAL PREFIX SETS FROM EASTERN INDONESIA: "CMP" AND EMP**

	PCMP	LESSER SUNDAS					
		KAMBERA	SIKA	DHAO	DELA	AMARASI	
1SG	*ku-	ku-	ʔ-	k-	ʔU-	ʔu-	
2SG	*mu-	(m)u-	m-	m-	mU-	mu-	
3SG	*na-	na-	n-	n-	na-	na-	
1PL.EXCL	*ma-	ma-	m-	ŋ-	mi-	mi-	
1PL.INCL	*ta-	ya-	t-	t-	ta-	ta-	
2PL	*mi-	(m)i-	m-	m-	mi-	mi-	
3PL	*ra-	da-	r-	r-	ra-	ra-	

	PCMP	MALUKU				EMP	
		LETI	KEI	AMAHAI	SELARU	SAWAI	ANSUS
1SG	*ku-	u-	u-	u-	kw-	k-	e-
2SG	*mu-	mu-	um-, mu-	a-	mw-	m-	<bu>
3SG	*na-	na-	en-, na-	i-	y-, ky-	∅	<di>
1PL.EXCL	*ma-	ma-	am(u)-	ma-	aramy my-	k-	amat-
1PL.INCL	*ta-	ta-	it-	i-	t-	t-	tat-
2PL	*mi-	mi-	im-, mi-	mi-	my-	f-	met-
3PL	*ra-	ka-	er-, hir-	si-	r-	r-	et-

**TABLE 13. SOME PRONOMINAL PREFIX SETS FROM WESTERN INDONESIA AND CHAMORRO: "WMP"**

	PCMP	Marianas	Sumatera		Sulawesi		Sumbawa
		Chamorro	Nias	Simalur	Da'a	Toraja	Sumbawa
1SG	*ku-	hu-	u-	u-	ku-	ku-	ku-
2SG	*mu-	u(n)-	(m)u-	mu-	mu-	mu-	mu-
3SG	*na-	ha-	na-	ni-	∅	na-	∅
1PL.EXCL	*ma-	in-	ma-	mai-	∅	ki-	tu-
1PL.INCL	*ta-	ta-	ya-	ta-	∅	ta-	tu-
2PL	*mi-	en-	(m)i-	mi-	∅	mi-	mu-
3PL	*ra-	ma-	da-	da-	∅	ra-	∅

24. We note that Ross reconstructs three distinct prefix paradigms for Proto-Oceanic, lower down in the Austronesian tree, implying that nominative agreement had not stabilized by the Proto-Oceanic stage.

Leaving aside agreement, the derivation-like monoconsonantal suffixes that characterize many of the CMP-area languages can be attributed to inherited Austronesian morphology and parallel developments (as Adelaar 2005a notes). The widely attested attributive *-n* most likely derives from the Malayo-Polynesian clitic *\*-ña* ‘3SG.GEN’. (Ross [1998] describes the functional changes associated with similar developments of the same clitic in Oceanic languages.) The suffix *-k* appears to be related to the applicative morpheme found in Southeast Sulawesi with forms including *-aka*, *-ako*, *-yoo*, reconstructing locally as *\*-akə*, and clearly related to *-kan* and *akan* in Malay, suggesting a form *\*-akən* in much of the Indonesian area and almost certainly etymologically related to the Oceanic “long transitivizer” *\*-aki[ni]*.

Table 14 summarizes the morphosyntactic basis for CMP. Evidence showing that the lexical innovations proposed by Blust cannot be used to support an argument for the close relatedness of the CMP-area languages will be presented in Grimes and Donohue (n.d.). For now, we simply note that the overwhelming majority of the innovations proposed are either not inclusive in the CMP-area languages, or else are not exclusive of languages external to the region. In short, there is no linguistic basis for positing CMP as a subgroup that withstands a close scrutiny of the data.

**TABLE 14. THE MORPHOSYNTACTIC BASIS FOR CMP**

	West of CMP?	CMP	East of CMP?	West Papuan
PREFIXAL AGREEMENT	yes	sporadic	sporadic	yes
V-n, V-k	yes?	yes	yes	sporadic

**4. CENTRAL-EASTERN MALAYO-POLYNESIAN.** Table 15 presents the innovations that Blust identifies as defining CEMP. He divides the innovations into four groups: phonological, lexical, morphosyntactic, and semantic. By far the largest number of innovations are in the lexical field; we find this significant, for reasons that are explained below, but we do not discuss these innovations in detail in this article. Otherwise, there are eight phonological innovations, five morphosyntactic innovations, and six semantic innovations. For these, as well as for the lexical innovations, Blust admits that some are more convincing than others. For instance, discussing the second irregular phonological innovation, Blust states that while the innovative form is characteristic of CEMP languages, the conservative form, *\*i-sai*, is also found throughout the CEMP area. In table 15 any feature that does not apply to all CEMP languages, and is only found sporadically within the region, we flag as “no.” We have assumed that Blust takes an innovation as defining CEMP with his statement (inferred or explicit) about how widespread the innovation is.

In total we find that there are claimed to be five diagnostic phonological innovations, two diagnostic morphosyntactic innovations, and one diagnostic semantic innovation. We examine all of these putative innovations below.

**4.1 THE PHONOLOGICAL EVIDENCE FOR CEMP.** In this section we examine the changes that Blust identifies as phonological innovations (regular and irregular) for PCEMP.

TABLE 15. BLUST'S PROPOSED CEMP-DEFINING INNOVATIONS

	REGULAR PHONOLOGICAL INNOVATIONS	DEFINES CEMP?
1	Cluster reduction: C1V1C2C1V1C2 > C1V1C1V1C2	yes
2	Cluster reduction: CVNiCjVC > CVNjCjVC	yes
IRREGULAR PHONOLOGICAL INNOVATIONS		
1	PMP *uliq > PCEMP *oliq 'return'	yes
2	PMP *i-sai > PCEMP *i-sei 'who'	no
3	PMP *ma-qitəm > PCEMP *ma-qetəm 'black'	no
4	PMP *maRi > PCEMP *mai 'come'	yes
5	PMP *tudan 'sit' > PCEMP *todan 'sit'	yes
6	PMP *inum > PCEMP *unum	no
MORPHOSYNTACTIC INNOVATIONS		
1	Prefixal / proclitic agreement on verb	no?
2	Alienable/inalienable possession distinction	no?
3	Frozen morphology	
3a	PMP *həpat > PCEMP *həpat, pat, pati 'four'	no
3b	PMP *ma-huab > PCEMP *mawab 'yawn'	yes
3c	PMP *ma-hiaq > PCEMP *mayaq 'shy'	yes
SEMANTIC INNOVATIONS		
1	PMP *t-ina 'mother' > PCEMP *t-ina 'big'	no
2	PMP *m-udəhi 'behind' > PCEMP *mudi 'back (of body)'	no
3	PMP *ma-qitəm > PCEMP *ma-qetəm 'dirty'	no?
4	PMP *tuqəla[nŋ] > PCEMP *zuRi 'bone'	yes
5	PCEMP *daun ni qulu 'head hair'	no
6	PCEMP *daləm 'mind, feeling'	no?

**4.1.1 Cluster reduction, 1: \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>.** The process of cluster reduction is widely attested in the CMP area, but is also an expected feature of any language that approximates a CV syllable structure. Outside the CEMP area, in Southeast Sulawesi, for instance, we find that roots of the original form \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>3</sub>V<sub>2</sub>C<sub>4</sub> are realized as C<sub>1</sub>V<sub>1</sub>C<sub>3</sub>V<sub>2</sub>, showing an even more complete adherence to the cross-linguistically unmarked CV template, in examples such as \*tuktuk > *Tukang Besi tutu* 'pound' (Donohue 1999).

While the pattern of cluster reduction is a characteristic of a large number of languages in the CMP area, it is not restricted to them. Furthermore, it is not the sole means of resolving codas in C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> roots. Examples (9)–(11) present illustrative data on the resolution of historical clusters in Buru, a language of central Maluku (presented in Grimes 1991a, 1991b). In (9) we see examples of the sort of cluster reduction described by Blust, in which a reduplicated root preserves the onsets, but not the codas, in medial clusters. In (10), however, we see that other lexical items appear to follow the opposite pattern, with the coda, and not the onset, of the medial cluster preserved.<sup>25</sup> Example (11)

25. Grimes (1991a) proposes that forms such as *gepe* are derived from an unreduplicated monosyllable, \*kep, and an epenthetic vowel echoing the quality of the vowel in the preceding syllable. This interpretation is equally valid; regardless of which is chosen, the cluster reduction pattern described by Blust does not account for the Buru facts.

shows that two distinct reflexes of \*kiskis in Buru show the two different strategies at work resulting in two distinct modern reflexes.

- (9) Buru examples of  $*C_1V_1C_2C_1V_1C_2 > C_1V_1C_1V_1C_2$
- a. \*t u k t u k ‘pound’  
 t u k t u \*VC#>V#  
 t u t u \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>  
 t o t o \*u > o  
*toto* ‘pound, thrust downward’
- b. \*b i R b i R ‘lips’  
 b i R b i \*VC#>V#  
 b i b i \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>  
 f i f i \*b > f  
*fifi-n* ‘mouth’
- (10) Buru examples of  $*C_1V_1C_2C_1V_1C_2 > C_1V_1C_2V_1C_2$
- a. \*b a s b a s ‘cut’  
 b a s b a \*VC#>V#  
 b a s a \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>1</sub>C<sub>2</sub>  
 f a s a \*b > f  
*fasa* ‘cut, decide’
- b. \*k e p k e p ‘hold’  
 k e p k e \*VC#>V#  
 k e p e \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>1</sub>C<sub>2</sub>  
 g e p e \*k > g (semi-regular)  
*gepe* ‘hold tightly in hand’
- (11) Buru split reflexes of \*kiskis ‘scrape, grate’
- a. \*k i s k i s  
 k i s k i \*VC#>V#  
 k i k i \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>  
*kiki* ‘scrape, scratch’
- b. \*k i s k i s  
 k i s k i \*VC#>V#  
 k i s i \*C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>C<sub>1</sub>V<sub>1</sub>C<sub>2</sub> > C<sub>1</sub>V<sub>1</sub>C<sub>2</sub>V<sub>1</sub>C<sub>2</sub>  
*kisi* ‘stroke with fingertip’

Examples (9)–(11) show clearly that, even within the one language or the one lexical item, there is no one simple process of cluster reduction that explains all of the modern forms. We have no doubt that this is not a unique feature of Buru, but one that is likely to be repeated in more languages as more detailed accounts of their diachronic phonologies become available (see discussion in Grimes 1991a).

**4.1.2 Cluster reduction, 2:  $*CVN_iC_jVC > CVN_jC_jVC_2$ .** This change is very similar to the change described earlier in 3.1.3, when discussing postnasal voicing in CMP-

area languages. We have nothing further to add to the discussion found there, other than to note that the assimilation of nasals to the place of articulation of a following stop is hardly rare either within Austronesian or more widely cross-linguistically. Examples (12)–(14) show the productive synchronic process of nasal assimilation to the place of the following stop in Malay (see also [1]–[3] from 3.1.3). (See also Newman 1984, and many since.)

- (12) /məŋ+buəŋ/[məmbuəŋ],      \*[məŋbuəŋ]      ‘throw away’  
 (13) /məŋ+doɾoŋ/[məndoɾoŋ],      \*[məŋdoɾoŋ]      ‘push’  
 (14) /məŋ+dʒəuh + i/[məŋdʒəuhi],      \*[məŋdʒəuhi]      ‘become distant’

Similar changes of  $N_iT_j > N_jT_j$  are not hard to find beyond the CMP area, particularly in the central Indonesian area that forms the western “border” of the CMP area. PMP \*dəmdəm ‘dark’ is reflected in *Tukang Besi* as *mo-rondo* ‘night’, and not \**mo-romdo* with a heterorganic nasal-stop sequence. (The requirement for homorganic NC sequences is not, of course, confined to the “border” area; we find many examples in Malay such as \**dindij* ‘cold’ > Malay *dindij*.)<sup>26</sup>

**4.1.3 \*uliq > \*oliq ‘return’.** Many of the innovations that Blust ascribes to CMP are also sporadically attested in Southeast Sulawesi. The change of \*uliq > \*oliq might, as Blust (1993:247) (citing Nothofer 1992) notes, be attested in Muna, where the form *oli* would not be a regular reflex of \*uliq (the *d-* in the form cited by Nothofer and Blust is possibly a frozen third person plural prefix). This etymology remains, however, speculative.

More significantly, \*uliq is reflected unchanged in at least some CMP languages, such as Palu’e *ph-uli* ‘return (parallel speech form)’ (*ph-* is erratically attested on a number of etymologically vowel-initial verb roots, such as *ph-ana* ‘right’ < \*wanan, via \*anan).<sup>27</sup> The appearance of \*oliq most likely represents the sporadic lowering of \*u to \*o due to the presence of the postvelar \*q. Such long-distance effects are not unknown, as in Palu’e *ola* ‘snake’ < \*ulaR with the final postvelar \*R causing lowering of the \*u in the first syllable (similar examples involving \*q can be found in, e.g., Malay *oleh* < \*uliq).

**4.1.4 \*i-sai > \*i-sai, \*i-sei ‘who?’** Blust notes that the change from \*i-sai > \*i-sei is a phonetically natural one, so its subgrouping value must be minimal, especially because forms reflecting \*i-sai are also found east of the CMP area in Oceania. *Tukang Besi*, from Southeast Sulawesi, shows both *emai* and *ie’ei* for ‘who’, and the second of these might be etymologically \*ia ‘3SG’ + \*ei ‘who?’ (the loss of \*s is erratically attested elsewhere: \**tasik* ‘sea’ > *tai* ‘deep sea’, \**si-ia* ‘3SG’ > *ia*). Regardless of this (rather speculative) etymology, the fact that forms reflecting \*sai are found throughout the CEMP area indicate that this (phonetically natural) innovation cannot be taken as a diagnostic change. Further, and more convincingly, the Kaili-Pamona subgroup of Celebic, from

26. Similarly, not all CMP-area languages preserve both elements in an NC sequence. The same etymon, \**ma-dəmdəm* ‘dark’, is reflected in Palu’e, from Flores in the CMP-area, as *m(ə)re* ‘dark, night’, not \**m(ə)de*, which might be expected if the nasal was preserved, just as with \**tumbuq* ‘grow’ > *th(ə)bu* (if the source of this Palu’e word was \**tubuq*, rather than \**tumbuq*, we would expect \**th(ə)vu*, which is not attested).

27. We also note that not all of the forms Blust cites in support of \*oliq are accurate. Wijngaarden (1986:89) lists *oli* in Hawu as meaning ‘not stay somewhere consistently’, and gives the sentence *peoli la Wa ma Hawu* PE-*oli* OBL west and Hawu ‘now on Sumba, and then again on Sawu’, which does not suggest the meaning ‘return’. Grimes (fieldnotes) shows *vari* as the normal Hawu word for ‘return’, whereas the doublet *holi-hoo* (with variant *oli-hoo*) is used to mean ‘back-and-forth, both ways’.



Central Sulawesi, has the protoform \*i-sei-ma ‘who’ (Martens 1989), which is strong evidence that the irregular development of \*i-sai > \*i-sei was found earlier than would be suggested by the borders of “CMP” languages.

**4.1.5 \*ma-qitəm > \*ma-qetəm ‘black’.** An irregular lowering of \*i to \*e causes PAN \*ma-qitəm ‘black’ to be reflected as \*ma-qetəm in some, but by no means all, of the CEMP languages. The first problem with this etymology is, as pointed out by Blust, the fact that the entire southwestern ‘arm’ of the CMP area, all of Flores, Sumba, Hawu, and the Solor archipelago, reflect \*ma-qitəm with no lowering.

Blust further acknowledges that Nothofer (1992) points out a number of WMP languages with an erratic *e* in this lexeme, but offers no explanation for the fact that the two WMP languages with irregular lowering of \*i in this lexeme are spoken near the putative WMP/CMP border. Table 16 shows that the lowering of \*i > *e* in this example is irregular in Wolio, one of the languages that Nothofer cites. PMP \*i is regularly retained as *i* when adjacent to \*q, and this is also true of the near relatives of Wolio, namely Kamaru, Kalao, Laiyolo, and Wotu. The irregular lowering of \*i is, therefore, a feature of Proto-Wotu-Wolio (Donohue 2005b) and, given that these languages are sisters of the Kaili-Pamona languages Bare’e (drawing on Grimes and Grimes 1987:124–25), Pamona, and Uma’ (from Tryon 1995) which also display *maeta* or (Uma’) *mo/eto*, it appears highly likely that the variant \*ma-(q)eta was present in Proto-Celebic, as suggested by Martens’ (1989:203) reconstruction of \*mV-’eta for Proto-Kaili-Pamona (see table 17; figure 4 clarifies the relationships between these languages).

Blust (1993:248) dismisses Simalur *étem* ‘black’ as more likely reflecting the contraction of \*a-i, following the loss of \*q, citing the closely related Sichule and Nias forms *a-itə* and *a-itō*, respectively. He notes that “it is not clear that this [*e* resulting

**TABLE 16. REFLEXES OF PROTO-AUSTRONESIAN \*i AND \*u NEXT TO \*q IN WOLIO AND ITS NEAR RELATIVES**

	PAN	WOLIO, KAMARU, ETC.
‘black’	*ma-qitəm	ma-eta
‘bitter’	*ma-paqit	ma-pa’i
‘feces’	*taqi	ta’i
‘choose’	*piliq	pili
‘white’	*ma-putiq	ma-puti
‘return’	*uliq	mb-uli ? (not a regular morpheme division)
‘snake’	*qulaR	ulo
‘rain’	*quZan	(Kalao: uda)
‘wash’	*DiRuq	(Kalao: pan/diu)
‘ten’	*puluq	sapulu, sapuluaju

**FIGURE 4. RELATIONSHIPS AMONG THE CELEBIC LANGUAGES**

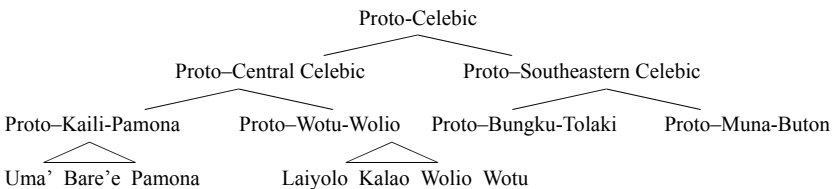


TABLE 17. \*qitəm VS. \*qetəm IN MAINLAND CELEBIC LANGUAGES

	LANGUAGE	LOCATION	'black'
Southeast Sulawesi	Koroni	northeast coast	mo <sup>ʔ</sup> ito
	Bungku	northeast coast	mo <sup>ʔ</sup> ito
	Mori Bawah	north, inland	moito
	Padoe	north, inland	me <sup>ʔ</sup> eto
	Tolaki	south, inland	meeto
	Mori Atas	north, inland	moeto
	Tomadino	northeast coast	mo <sup>ʔ</sup> eto
Central Sulawesi	Uma <sup>ʔ</sup>	central highlands	maeta
	Pamona	eastern highlands	maeta
	Napu	central highlands	maeta
	Proto–Kaili-Pamona	central highlands	*mV- <sup>ʔ</sup> eta

from \*a-i—MD&CG] is the case for the problematic Bare'e and Wolio words." We suggest that this explanation cannot possibly be the case for the Sulawesi data, because the \*a is still preserved as *a* in the Celebic forms, leaving no possibility that the *e* results from the coalescence of \*a+i.

Mead (1999:152) presents additional data from Bungku-Tolaki languages of the northeast coast of mainland Southeast Sulawesi for languages that appear to have irregularly lowered \*i in 'black' (alongside many languages with \*i > i). Koroni and Bungku clearly show regular reflexes of \*ma-qitəm, as does Mori Bawah, just to the northwest. Padoe might be analyzed as showing \*a<sup>ʔ</sup>i > e<sup>ʔ</sup>e, with the glottal stop being transparent for the purposes of vowel merger (see Borroff 2007); Tolaki represents a simple case of \*ai > ee, following the loss of \*q (\*ay > e is regular in all languages native to Southeast Sulawesi).

With Mori Atas and Tomadino, however, the antepenultimate vowel of \*ma-qitəm is regularly reflected as *o*, via reduction to schwa. The fact that the \*i is reflected as *e* cannot be explained, except by positing an irregular intermediate form \*ma-qetəm. Given these alternations in the Bungku-Tolaki languages, as well as in Proto-Central Celebic, we must reconstruct \*qetəm as an alternant in Proto-Celebic.<sup>28</sup>

All of these data indicate that the sporadic change of \*qitəm > \*qetəm had begun at least as far back as Central Sulawesi,<sup>29</sup> becoming increasingly common in Southeast Sulawesi and areas further east—but not south, in the Bima-Sumba-Flores area, as Blust (1993:248) notes, where forms such as *mite* 'black' (from Palu'e) are found, showing no evidence of a \*qetəm origin. Additionally, identical forms are also found further east (e.g., *mite-t* 'black' in Buru; see table 9). In other words, the lowering of \*i > \*e in this lexeme is neither restricted to the CEMP area, nor is it universal within that area.

**4.1.6 \*maRi > \*mai 'come'.** The irregular loss of \*R in PCEMP \*mai 'come' < PMP \*maRi is similarly attested in Southeast Sulawesi. Table 18 shows data from Tukang Besi (outside the CEMP area) and Buru (inside CEMP/CMP), making it clear that, while \*R is regularly retained as *h*, including in environments preceding and following *i*, it is lost in 'come'.

28. We also note Ba:ngingi Sama *etom*; not enough is known of the historical phonology of this language to be able to claim that this represents an irregular lowering of \*i or not.

29. The existence of Punan Kelai (Borneo) *maèdang* suggests either an even earlier source for the alternations, a diffusional spread of the innovative form, or possibly a relationship between Punan and the Celebic languages.

TABLE 18. PROTO-AUSTRONESIAN \*R IN TUKANG BESI AND BURU

	PAN	TUKANG BESI (WMP-area)	BURU (CMP-area)
'blood'	*DaRaq	raha	raha-n
'thorn'	*DuRi	ruhi	rohi-n
'wash'	*DiRu	he/rihu	ep-riho
'stingray'	*paRi	pahi	pahi
'come'	*maRi	mai	mahi

These data show that the irregular loss of \*R in this lexeme is not restricted to the CEMP languages, but is also found in Southeast Sulawesi. Other languages in Southeast Sulawesi similarly showing an irregular loss of \*R in 'come' are Kaimbulawa, Cia-Cia (of Batu Atas), and Muna, all of which display *mai*. Elsewhere in Sulawesi we find Proto-South Sulawesi \*mai, Uma' (from Central Sulawesi) *mai*, and Gorontalo (from northern Sulawesi) *mayi*, and further afield even Yami, from Formosa, shows *mai*. In none of these languages is \*R normally lost, and so the fact that the language reflects \*mai, and not \*maRi, is irregular for this one lexeme.

This strongly indicates that there is a long history of \*maRi alternating with \*mai. Equally important is the fact that there are CEMP languages in which the \*R is *not* lost in reflexes of \*maRi. Table 19 shows that, while there are (many) languages (in addition to those cited by Blust) in which the \*R is lost, there are at least a few languages, not geographically contiguous, in which \*R is retained, with the individual languages' regular reflex of \*R. As with the change in 'black', the irregular loss of \*R in \*maRi is neither restricted to the CEMP area, nor is it universal within that area.<sup>30</sup>

TABLE 19. \*maRi AND \*mai IN CMP LANGUAGES

LANGUAGE	LOCATION	'come'
Palu'e	Flores	ma_i
Tetun	Timor	ma_i
Nuauulu	Central Maluku	ma_i
Buru	Central Maluku	mahi
Watubela	Central-east Maluku	go/mari
Geser	Central-east Maluku	mari
Onin	Bomberai (NG)	ma__

4.1.7 \*tudān > \*todān 'sit'. As presented by Blust, the only CMP reflexes of \*todān are from eastern and southeastern Maluku: Yamdena *na-m-toran*, Geser *ma-toran*, Sekar *m-tonag*, and Koiwai *ma-toran*, a group of languages that he himself proposes as a subgroup of CMP (1993:279), and so this evidence should not be taken as supporting a CMP-level reconstruction, or the reconstruction of PCMP. Further, as noted by Blust, this form is not found in Oceania; this restriction makes it highly likely that, rather than being an innovation for CEMP, it is a lexeme that has spread geographically.

Speculatively, we note that the proposed innovation of PMP \*tudān 'sit' > PCEMP \*todān is possibly reflected in Tukang Besi *torae* 'place'. The word is suspicious because

30. Grimes (1991a) presents data with forms *mai* and *ma/maa* from sixteen languages within the CMP area, all of which normally retain \*R and \*i. So while it is certainly well attested within the CMP area, it nevertheless also occurs widely outside the CMP area.

of its trisyllabicity in a language that otherwise overwhelmingly has disyllabic roots; the *-e* might erratically reflect a frozen third person agreement marker *-ʔe* (simply *-e* in closely related languages), regularly forming a bivalent verb ‘seat it’. The affixation of *-ʔe* on otherwise monovalent verbs to form causative derivations is a regular morphological process, as shown in table 20. (The other correspondences in the word, *\*todan* > *tora*, are regular.) While *\*tora* is not attested in the modern language meaning ‘sit’, the lexeme with this meaning, *keḌe*, is clearly innovative, and so might well have replaced an earlier reflex of *\*[tʰuo]dan*.

TABLE 20. TUKANG BESI VALENCY ALTERNATIONS  
MONITORED BY THE USE OF *-ʔe*

MONOVALENT USE		BIVALENT USE	
ḃuti	‘fall’	ḃutiʔe	‘drop something’
pono	‘full’	ponoʔe	‘fill’
mente	‘be surprised’	menteʔe	‘surprise someone’
like	‘awaken’	likeʔe	‘wake s.o. up’

**4.1.8 *\*inum* > *\*inum*, *\*unum* ‘drink’.** The innovation of forms reflecting *\*unum* (alongside the continuing *\*inum*) is well attested in Oceania and southern Cenderawasih Bay, but the only attestations in the CMP area that Blust lists are Paulohi *umu*, Ngaibor *n-un*, and Kambera *ŋ-unuŋu*. Paulohi (from central Maluku) shows a plausible reflex (though dialectally *inu*, *ninu*, and *nin* are also attested), but Ngaibor reflects an intermediate form *\*(n)[i]nu(m)*, not *\*unum*. The accretion of *n-* (probably a frozen 3SG nominative prefix) to a small number of verbs is widespread in the CMP region, including the Aru islands where Ngaibor is from: West Tarangan, for instance, reflects *nun*; the regular 3SG prefix in this language is *i-*, while *\*n* is found with only three verbs: *nun* ‘drink’, *nal* ‘fetch’, and *nar* ‘do’ (Rick Nivens pers. comm.). In north-central Flores Paluʔe shows *ninu* ‘drink’ and *nala* ‘fetch’; there is no agreement for 3SG in this language, and synchronically *n-* must be analyzed as part of the root. The same is true of Hawu *ŋinu* and Helong *ninu* ~ *nium*, neither of which inflect their verbs for person. Rather than being evidence for a phonological innovation, it appears that there is evidence for the sporadic fossilization of *n-* onto a number of otherwise vowel-initial roots in eastern Indonesia.

Further we note, following Lynch (2002), that reflexes of *\*um-inum* are likely to show *\*(m)unum* forms in Oceanic languages, reflecting the transference of rounding in the original <um> to the bilabial nasal, and then subsequent transfer of the rounding from the nasal to the following vowel.

(15) AN ALTERNATIVE PATH TO *\*unum*

<i>*u</i>	<i>m</i>	-	<i>i</i>	<i>n</i>	<i>u</i>	<i>m</i>	‘drink’
	<i>m<sup>w</sup></i>	-	<i>i</i>	<i>n</i>	<i>u</i>	<i>m</i>	<i>*um</i> > <i>*m<sup>w</sup></i> / #__
	<i>m</i>	-	<i>u</i>	<i>n</i>	<i>u</i>	<i>m</i>	<i>*m<sup>w</sup>i</i> > <i>*mu</i>
			<i>u</i>	<i>n</i>	<i>u</i>	<i>m</i>	loss of original affix

The fact that this change is not found in the WMP-area languages is explained by the same appeals to areal trends, and borrowing from Papuan languages, that are invoked when discussing the appearance of labiovelars in Proto-Oceanic.<sup>31</sup> As with

other changes that are more regular in the east of the Austronesian area (such as \*mai ‘come’ < \*maRi) the uneven distribution simply reflects different founder effects in different parts of the range.

## 4.2 SUMMARY OF THE PHONOLOGICAL EVIDENCE FOR CEMP.

Table 21 summarizes the discussion in this section; the Sulawesi region shows great variation, and here we only report on the existence of commonalities with the proposed CEMP innovations, regardless of the presence of differences in other areas. The terms “sporadic,” “yes,” and “no” are used with the same meanings as in table 11. As can be seen, none of the phonological features proposed are exclusive to the languages of the CEMP area, and none of them are inclusive of all of these languages. In short, when combined with information about the location of the languages that do reflect various of these proposed innovations, none of them argue for the existence of a CEMP protolanguage ancestral to all of the languages currently considered to be CEMP languages, either as a single language or as a differentiated linkage.

TABLE 21. SUMMARY ASSESSMENT OF THE PHONOLOGICAL EVIDENCE FOR CEMP

	West or north?	Sulawesi?	All of CMP-area?	All EMP-areas?
Cluster reduction 1 (4.1.1)	sporadic	yes	no	(yes)
Cluster reduction 2 (4.1.2)	sporadic	sporadic	sporadic	(yes)
*uliq > *oliq (4.1.3)	no	(no)	no	(yes)
*i-sai > *i-sei (4.1.4)	no	(yes?)	no	no
*ma-qitəm > *ma-qetəm (4.1.5)	no	yes	no	no
*maRi > *mai (4.1.6)	(yes)	yes	no	(yes)
*tudan > *todan (4.1.7)	sporadic	sporadic	?’	yes
*inum > *unum (4.1.8)	no	(yes)	no <sup>†</sup>	(no)

\* This depends on the credence given to the etymologies given for 4.1.7, and the degree to which the number of undocumented languages in western Indonesia can be discounted.

† The data supporting this item are extremely limited both geographically and numerically. The changes that have applied to this item are more complex than a simple replacement innovation would suggest, and indicate multiple independent, but converging, changes.

**4.3 MORPHOSYNTACTIC EVIDENCE FOR CEMP.** As with the discussion of the CMP evidence, we are not focusing on morphosyntactic innovations in this article, but due to the lexical/phonological nature of many of the proposals for CEMP we treat them briefly here. Our discussion focuses on the CMP-area languages, because they are the languages that share a border with the WMP-area languages, some of which, we claim, share similar features with some CMP-area languages.

**4.3.1 Proclitic agreement on the verb.** While it is true that agreement prefixes are widespread in the CEMP languages, there are many CMP and Oceanic languages

31. To add to this explanation, we note that there are a large number of non-Austronesian languages with labiovelar phonemes along the coast of North-Central New Guinea and its immediate hinterland, along the route that must have been followed by the Austronesians as they headed east towards the Pacific. These languages come from a variety of families in this area (Kwerba, Tor, Lakes Plains, Greater Skou, Ndu, and the Molof isolate), and are categorically absent outside this region, suggesting that labiovelar phonemes were areal in their distribution, and contact with this area perhaps influenced the pre-Proto-Oceanic population as it passed by.

without such agreement, and equally a number of languages, in Sulawesi (particularly Southeast Sulawesi), that do show similar agreement prefixes. The data for this have already been presented in summary form in section 3.3, and need not be repeated here. We note that there are no proposals to reconstruct a paradigm (or paradigms) of CEMP-level prefixes, which indicates that the occurrence of prefixal agreement in the CEMP-area languages is the product of multiple innovations, rather than one single event. As can be seen in table 12, the agreement forms in EMP languages are not 100% cognate with those found in even the more apparently conservative CMP-area languages, nor are they cognate with any of the reconstructed Oceanic agreement paradigms (Lynch, Ross, and Crowley 2002).

**4.3.2 Alienable/inalienable possession distinction.** Distinctions between alienable and inalienable possession in terms of the basic morphosyntactic coding possibilities are found in the CEMP area, but it is only in EMP that they become the norm, and where cognate morphology can regularly be found. In the CMP area a contrast in possessive classes is sometimes found, but the morphology (and even the basic constructions) used to mark the contrast is not cognate across the CMP-area languages, or with the EMP languages. Furthermore, a contrast between alienable and inalienable possession is present in a number of “WMP” languages. While peripheral to the discussion of the coding of basic possessive marking, it is worth stressing the fact that this sort of semantic opposition is not unique to the CEMP-area languages within Austronesian.

Tukang Besi makes no distinction in the forms used to mark possession, as can be seen in (16), but does nonetheless distinguish two kinds of possession. At the clausal level we note that external possession, in which the possessor adopts the grammatical function of the possessum, is grammatical only for inalienable items, such as body parts, but not for alienable things.

- (16) a. te ana=su  
CORE child=1SG.GEN  
'my child'
- b. te beka=su  
CORE cat=1SG.GEN  
'my cat'
- (17) a. No-topa='e na бага=su.  
3R-slap=3P NOM cheek=1SG.GEN  
'They slapped my cheek.'
- b. No-topa=aku na бага=su.  
3R-slap=1SG.P NOM cheek=1SG.GEN  
'They slapped my cheek.'
- (18) a. No-topa='e na beka=su.  
3R-slap=3P NOM cat=1SG.GEN  
'They slapped my cat.'
- b. \*No-topa=aku na beka=su.  
3R-slap=1SG.P NOM cat=1SG.GEN

A similar pattern is found in Indonesian:

- (19) a. Saya di-angkat ovari (saya).  
1SG NONACT-take ovary 1SG  
'I had my ovaries removed.'

- b. \*Saya di-angkat topi (saya).  
 1SG NONACT-take hat 1SG  
 ‘I had my hat removed.’

The two means of expressing clausal possession in *Tukang Besi* similarly show an uneven distribution across different kinds of possessive relationships. Inalienable relations are expressed with an existential construction (effectively showing “possessor lowering”), exemplified in (20), while the same construction cannot be used to indicate possession in an alienable relationship. These nouns can appear with the existential verb, but only to describe their location, as in (21).

- (20) a. Ane ke ana=su. b. \*ane ke beka=su  
 exist with child=1SG.GEN exist with cat=1SG.GEN  
 ‘I have a child.’
- (21) Ane ke beka=su di ito.  
 exist with cat=1SG.GEN OBL.R there:higher  
 ‘My cat is up there.’

The other possessive construction involves incorporation of the possessed nominal into the verb *hoto* ‘have’ (discussed in detail in Donohue forthcoming). While this is grammatical with alienable possessive relationships, it is not used with inalienable relationships, as seen in (22); compare the grammaticality judgments here with those in (20).

- (22) a. \*ku-hoto-ana b. Ku-hoto-beka.  
 1SG-have-child 1SG-have-cat  
 ‘I have a cat.’

At the phrasal level, *Tukang Besi* allows the option of coding a possessed inalienable item with the NP-final word *mai*: compare (23) with (16). This is never obligatory, but taken together we have substantial morphosyntactic evidence for the relevance of the category ‘(in)alienability’ in the language, even though the pronominal indication of possession does not vary.

- (23) a. te ana=su mai b. ?/\*te beka=su mai  
 CORE child=1SG.GEN INAL CORE cat=1SG.GEN INAL  
 ‘my child(ren)’ ? ‘my cat(s)’

The existence of an alienable/inalienable contrast can readily be ascribed to Papuan contact, given the widespread presence of this feature in the Papuan languages of insular eastern Indonesia (including the distinction between direct and indirect possession) and the lack of any reconstructible forms for the contrast in the CMP-area languages (in contrast to EMP).

**4.3.3 Frozen morphology.** Blust cites a number of pieces of “frozen morphology” as evidence for the CEMP group. We examine these below.

**4.3.3.1 \*həpat > \*həpat, \*pat, \*pati ‘four’.** It is claimed that the loss of the initial syllable of \*əpat ‘four’ is an erratic feature of the CEMP-area languages. The loss of the initial syllable in the lexeme for ‘four’ is well attested in other languages from Southeast Sulawesi. Illustrative examples are shown in table 22.

Understanding of this item has been complicated by many written sources for CMP-area languages not indicating sequences of two like vowels. But when we compare Southeast Sulawesi data from table 22 (Pancana *popaa*, Muna *paa*, Wolio *apa*) with Buru *paa*, Amarasi, Dela, and Termanu *haa*, Baikeno *hadʔ*, Tetun *haat*, Helong *aat*, and Dhao and Hawu *əp:a*, then **\*\*pat** is inadequate to account for the data. On the other hand a form closer to PMP, like **\*(h)əpat** (or, for some, an alternate **\*\*apat**), is adequate to account for the data, given the metathesis and loss of historical final consonants widely documented in the CMP region (Grimes 1991a).<sup>32</sup>

Other examples of languages in South Sulawesi that apparently reflect **\*\*pat** can be found in Grimes and Grimes (1987:128–29): Mamuju *pataʔ*, Topoiyo *patta*, Wotu *pataŋo* (cf. Wolio, Kamaru). Browsing further afield in the *Comparative Austronesian Dictionary* also reveals Sarangani Blaan *fɔ*, Acehnese *puɔ*, and Balinese *pat*, indicating that the erratic loss of the initial vowel in this lexeme is found much further north and west than could be accounted for by assuming that it is an innovation in the CEMP area.

**TABLE 22. PMP \*həpat REFLEXES SUGGESTING AN INTERMEDIATE \*\*patV IN SOUTHEAST SULAWESI**

LANGUAGE	'four'	
Tukang Besi	<u>h</u> ato-hulu	'forty'
Tomia, Binongko	<u>pa</u> ʔa	'four'
Lasalimu, Kumbewaha	<u>popa</u> ʔa	'four'
Cia-Cia (Masiri)	<u>popa</u> ʔa	'four'
Cia-Cia (Sampolawa)	opaʔa	'four'
Pancana	<u>popaa</u>	'four'
Muna	<u>fato</u> , <u>paa</u>	'four'
Muna (Kadatua)	<u>fato</u> ʔono	'four'
Busoa	<u>fato</u> -aʔo	'four'
Wolio	<u>pata</u> -pulu (but apa 'four')	'forty'
Kamaru	<u>pata</u> -ajo	'four'

**4.3.3.2 \*ma-huab > \*mawab 'yawn'.** Forms such as Palu'e *ŋoa* 'yawn', which reflects an earlier **\*\*oa** with an epenthetic *ŋ*, clearly reflect **\*huab** (without **\*ma-**), not **\*mawab**, showing that Palu'e, from Flores, did not participate in the innovation proposed. Grimes (1991a) similarly points out that, while Buru *mawa* 'yawn' can derive from either **\*ma-huab** or **\*mawab**, the form *duba* 'sleepy, yawn' can only be derived from **\*ma-huab** (through metathesis) and not from **\*mawab**.<sup>33</sup>

We note in passing that, because the vast majority of the languages of eastern Indonesia reflect **\*h** as  $\emptyset$ , the change from **\*ma-huab** > **\*mauab** > **\*mawab** is primarily one of transcription, and not of phonemicization (see Clynes 1997, 1999 on the status of "diphthongs" in Austronesian reconstructions).

32. We note parallel behavior for **\*ənəm** 'six' in CMP-area languages, with Buru, Amarasi, Baikeno, Dela, Termanu *nee*, Tetun *nee-n*, Dhao, Hawu *ən:a*, and Helong *enɛŋ*.

33. The Buru *uba* sequence that is derived from a historical **\*uab** sequence through metathesis follows a broader pattern in the language and in the region triggered by **\*p** and **\*b** (see Grimes 1991a, 1991b). A possible basis for the *d* in *duba* is as a fossilized agreement prefix (cf. *da-* '3SG', and see the discussion of *n-* in 4.1.8).



**4.3.3.3 \*ma-hiaq > \*mayaq ‘shy’.** Soboyo, one of the few languages in eastern Indonesia to have an overt reflex of \*h, reflects *mahi* ‘shy, ashamed, embarrassed’ (Fortgens 1921:75). As Soboyo is part of the proposed northern “Sula-Buru” group in Maluku, this means that we must reconstruct a form with \*h for any language ancestral to Soboyo, and so this does not support the \*mayaq form at the high PCEMP level that Blust proposes.

Because most languages in the CMP area lose historical \*h and \*y, it is tricky to try to argue that a form derived from \*ma-hiaq, but not \*mayaq, or vice versa. Consider, for example, Buru *em-gea*, Tetun *moe*, Helong *mae*, Dela, Termanu *-mae*, Dhao *makaē*, Hawu *mekae*, all meaning ‘shy, ashamed, embarrassed’. What is clear for languages that lose medial \*y (such as in \*layaR ‘sail’, yielding the predictable Buru and Helong *laa*) is that \*mayaq should result in forms looking more like \*\*maa (which are already taken for PCEMP \*maya ‘tongue’, which yields Buru *maa-n* and Helong *mee-n*, and for a few languages like Helong, *maa* < \*ma(R)i ‘come’). It is more likely historically that the vowels in the forms listed above for ‘shy, ashamed, embarrassed’ came from a historical form with the vowel \*i rather than one with a \*y.

**4.3.4 Semantic innovations.** A number of lexemes are claimed to have shifted semantic reference at the CEMP level. We examine these below.

**4.3.4.1 \*t-ina ‘mother’ > ‘big’.** Many attestations of *ina* ‘mother’ in the sense of ‘larger member of a set’ are found in Southeast Sulawesi. One example from Muna (van den Berg 1996) is given in (24).

(24)	ina:	inano pana	‘bow’ (opposed to ‘arrow’: anano pana) (pana: ‘bow and arrow, or shoot arrow’)
		inano katumbu	‘mortar’ (opposed to ‘pestle’)
		inano ati	‘main part of the breastbeam (of the loom); (ati: ‘breastbeam [part of the loom]’)

The first two sets are known to obtain in Tukang Besi as well, where *ina* is also ‘mother’, and *ana* is ‘child’:

(25)	a.	pana	‘shoot arrow; bow and arrow set’
		ina nu pana	‘bow’
		ana nu pana	‘arrow’
	b.	tumbu’a	‘mortar and pestle’ (tumbu ‘pound’)
		ina nu tumbu’a	‘mortar’
		ana nu tumbu’a	‘pestle’

Further west, examples such as Malay *ibu jari* mother (of) finger ‘thumb’ (i.e., the biggest finger) (similarly *ibu kaki* ‘big toe’, where *kaki* = ‘foot, leg’), *ibu kota* mother (of) cities ‘capital city’ (the biggest city), *ibu sungai* mother (of) river ‘main branch of a river’, *ibu pasir* mother (of) sand ‘pebble’, *ibu tentara* mother (of) army ‘main body of the army’ are not hard to find. It is clear that the ‘mother = important’ metaphor is not confined to eastern Indonesia.<sup>34</sup> Further, it is not found as widely within the CEMP area as claimed by Blust. In the sources that we were able to check, we note that Tetun

34. Note also English constructions such as ‘mother of battles’, ‘mother of (all) storms’, etc. Matisoff (1991) discusses this metaphor, and shows how widespread it is. See also Ross (2003:195–96) on the ‘mother = big, child = small’ metaphor in Oceanic; we note that One, a Torricelli language from northwestern Papua New Guinea, also has ‘mother = big’ and ‘son = small’.

*inan* does not have a ‘big’ sense. Morris (1984:90) lists the following two senses, which do not mention ‘big’ at all:<sup>35</sup>

1. ‘female (of animals)’
2. ‘mother; maternal aunt’; *inan susun* ‘one’s own mother’; *inan hasusu* or *inan susu* ‘a wet nurse’; *inan boot* or *inan kauaik* ‘maternal aunt older than the mother’; *inan klaran* or *in’lala* ‘middle maternal aunt’; *inan ikus* or *in’iku* ‘maternal aunt younger than the mother’

For Rote, Jonker (1908:202) lists *ina* as meaning ‘1. wife, female (of animals, when they are fully grown). 2. female (mare or chicken). 3. a woman, a female human being. 4. big. 5. very’. Blust cites Rote *ina* as meaning ‘female, of adult animals, thus big, in contrast to an immature female animal’, which we find to be a mistranslation of the Dutch.<sup>36</sup>

For Hawu, Wijngaarden (1896:98) lists *rena* as meaning ‘female; wife’, and offers the subentry *djara rena* (horse female) ‘mare’, and further *rena ae* ‘a big wife; also big woman, wide, heavy’, which would seem to support Blust’s claim that *r-ena* means ‘female, of animals; large woman; broad, heavy’, but Wijngaarden adds the note ‘In den zin van groot komt ‘t niet voor’ (“it [*rena*] does not appear in the sense *large*”). The apparent contradiction posed by *rena ae* is simply explained by examining the entry for *ae*: ‘very, many’, and so by extension ‘large’. The normal Hawu reflex is *ina* ‘mother’.

**4.3.4.2 \*m-udəhi > \*mudi ‘back (of body)’.** Palu’e *kuri* ‘behind’ reflects \*likuD via a metathesized form \*kuliD (or, with more convolutions, \*kuDil). This clearly indicates that PMP \*likuD was not replaced by \*mudi, as Blust (1993: 261) suggests, and that (at least in the west of the CMP area) \*likuD was retained. Additional data that confirm this can be found in Buru *mori-n* ‘back, behind (place or person)’; Dhao *limuri* ‘back, rear, last in a sequence’, *li’u* ‘1. behind (location), 2. outside’; Helong *liku-n* ‘behind, outside (e.g., a house)’; Termanu *muli-ha’i-n* ‘backbone’; Tetun *moli-n* ‘outside, back (location or anatomy), *liku* ‘carry a person on the back’, *kidu-n* ‘1. lower part, backside, buttocks, 2. back (of house), *ha-kidu-k* ‘move backwards, retreat’, also geographically/dialectally limited *li’ur* ‘outside, behind’; Amarasī, Baikeno *koti-n* ‘1. behind, back, rear, stern, 2. outside’.

Southern Tukang Besi *mburi* ‘behind’ shows the same fusion of \*m- with the root seen in the CEMP-area languages, and the same loss of the \*-eh-. At the same time the northern Tukang Besi form *taliku* ‘behind, back’ obviously reflects \*ta-likuD, and not \*likuD alone, indicating that this innovative form is not confined to CEMP-area languages, and suggesting a fused preposition \*ta, not otherwise present in the language.

**4.3.4.3 \*ma-qitəm > \*ma-qetəm ‘dirty’.** We find very little data to support this reconstruction, and note that the great majority of CEMP-area languages known to us have separate lexemes for ‘black’ and ‘dirty’. Languages justifying this assertion include Palu’e *mite* ‘black’, *raki* ‘dirty’; Wetan *metma* ‘black’, *kapri, kupa* ‘dirt(y)’; Buru *mite-t* ‘1. black, dark, 2. dirty, soiled’, *mede-t* ‘black, dark’, *raki* ‘bodily dirt, filth’; Amarasī

35. Hull (1999), discussing a variety of expatriate East Timorese Tetun, similarly does not include ‘big’ in any glosses.

36. The original definitions are: ‘wijffe, vrouwelijk van dieren, wanneer zij volwassen zijn’; ‘eene teef (merrie, kip)’; and for *inak* ‘eene vrouw, een vrouwelijk menschelijk wezen’; ‘groot’; ‘heel, zeer’.

*metan* ‘black, dark (color)’, Tetun *metan* ‘black, dark (color), *ma-kuku-n* ‘1. dark (of weather), 2. dark (metaphysical)’, *fo’er* ‘dirty, filthy’, *kdoor* ‘dirty, filthy, defiled’; Helong *mitan/mitəŋ* ‘1. black, dark (color), 2. dirty, soiled’, *hbabut* ‘dark (of weather)’, *hmomos* ‘1. dirty, filthy, 2. menstruation’; Dhao *mæfi* ‘black, dark (color)’, *maroga* ‘dark (weather)’, *kafalu* ‘dirty, filthy, unclean’; Hawu *karəb:a* ‘dark (weather)’, *ra’i* ‘body filth, dirty’; Dela <sup>u</sup>*geo* ‘black’, *ma-<sup>o</sup>a-hatu* ‘dark (of weather)’; Termanu <sup>u</sup>*geo* ‘black’, *ma-ka-ha-hatu* ‘darkness’, *ma-<sup>u</sup>ga-<sup>u</sup>ganu-k* ‘dirty, filthy’, *ma-<sup>u</sup>ge-<sup>u</sup>geo* ‘dirty, filthy’.

**4.3.4.4 \*zuRi/\*DuRi ‘bone’.** Blust proposes that PCEMP innovated a change of PMP \*zuRi ‘fish bone’ to PCEMP ‘bone’. We find this not to be a strong proposal, for a number of reasons.

Mahdi (1994: 454) proposes that a form that would be reconstructable as \*ZuRi ‘fish bone’ (though he does not assign it regular Austronesian status) can be found through the Philippines, citing forms such as Kapampangan *dwi* ‘fish spines’, Siokon Subanon *dugi* ‘fishbone’, among others (including CEMP-area languages). Given this pre-“CEMP” attestation and the fact that the same semantic range is also found with the CEMP-area languages (e.g., Kambara *ri* ‘bone, fishbone; vegetables, food; thorn, bristle; wife, spouse’ [‘been, bot, graat; groente, bijspijs; doorn, stekel; vrouw, echtgenote’; Onvlee 1984:439, *ri*]), we believe that a more parsimonious explanation of the facts is that, as attested in the northern Philippines and northern Sulawesi (both in the WMP area), ‘thorn’ came to take on ‘fishbone’ as an additional function; this sense continued and, possibly due to founder effects, was exaggerated in the CEMP area, with the additional erratic, but logical, extension to ‘bone’.

Even if we accept the Proto-Oceanic innovation \*suRi, another issue is whether the CMP-area language forms reflect the innovation \*zuRi, or whether they are erratic semantic extensions (not innovations) on the earlier PMP \*DuRi ‘thorn’. Given the scanty knowledge we have of the phonological histories of most of the CMP-area languages, this possibility cannot be discounted; we are not aware of a single case of a ‘bone’ term in a CMP-area language that can be plausibly related to \*zuRi, but not \*DuRi. Indeed, given what we know of the historical phonologies of the languages, the following forms (at least) can only reflect \*DuRi, and not \*zuRi: West Tarangan *tul* ‘bone’, Buru *rohi-n* ‘bone, fishbone’, Amarasi, Baikeno *mui* ‘bone’, Tetun *rui-n* ‘bone’, Dhao, Hawu, Dela *rui* ‘bone’.<sup>37</sup>

Even if \*zuRi must be reconstructed as a separate lexeme, the fact that languages such as Manggarai, Ngadha, So’a, Lio, and Nage (Flores) have *toko* ‘bone’, apparently reflecting PMP \*tuqəla[nŋ], shows that this cannot have been an inclusive replacement.<sup>38</sup>

**4.3.4.5 \*daun ni qulu ‘hair’.** A number of languages near New Guinea have the compound expression ‘leaf of head’ to refer to (head) hair. Blust (1993:262) notes that “this striking semantic innovation ... may be a product of contact with Papuan languages,” because the semantic extension ‘hair/leaf/grass’ is well attested in New Guinea

37. Other forms, such as Termanu *dui* ‘bone’, could reflect either \*zuRi or \*DuRi. As stated above, we are not aware of any cases that could reflect only \*zuRi, and not \*DuRi.

38. PMP \*tuqəla[nŋ] is also possibly reflected in Soboyo *talaŋ* and Gebe *kalonŋ*, from the northern Maluku area, and also in the Oceanic languages Tunggak *tuani*, Nalik *ruan* (New Ireland), and Taiof *tuana-naih* (Bougainville).

(e.g., Laycock 1986). Blust goes on to note that “the change is known principally from languages that are not in such a contact situation” (1993:262), because five of the six languages he cites with reflexes of this construction are not in contact with Papuan languages.

We note that there are two conditions under which contact-induced change can plausibly be posited: either the language in question is in contact with a plausible donor language, or the language in question has been in contact with a plausible donor language. It is true that only one of Blust’s six cited languages (Sekar) is *currently* in contact with a Papuan language, but we note that Fijian, Tongan, Samoan, and Māori all belong to the Oceanic subgroup of Austronesian, and that Proto-Oceanic underwent extensive contact with Papuan languages prior to dispersal, while it was forming in New Guinea. The remaining language, Bonfia, is found on Seram, in Central Maluku, where no Papuan languages are spoken. Seram is, however, only 400 km from the Bomberai peninsula, where Papuan languages are spoken and where a trading sultanate was based (in Onin), and lies directly between the Papuan languages of North Halmahera and those of the Timor region. Given the presence of Papuan languages to the north, east, and south, it would not be unreasonable to posit an earlier Papuan presence in Central Maluku as well.<sup>39</sup> The evidence is clearly in favor of \**daun ni qulu* being a calque from Papuan languages, and not an innovation of CEMP. The vast majority of languages in our data have different lexemes for ‘hair’, ‘leaf’, and ‘head’ (e.g., Palu *e lolo* ‘hair’, *vunu* ‘leaf’, *taba* ‘head’; Buru *folo-n* ‘hair, fur, feather’, *omo-n* ‘leaf’, *olo-n* ‘head’). The one exception is Dhao, which has *rəu kət:u* ‘hair (of head)’ (lit. leaf of head), *rəu* ‘leaf’, and *kət:u* ‘head’. Note the parallel pattern with nearby non-Austronesian (Papuan) languages of Alor-Pantar in Pura *ong va* ‘hair’ (lit. head leaf), *va* ‘leaf’, *ong* ‘head’. The very scarcity of the expression in the Austronesian languages lends to the argument that this was a contact-induced change affecting some, but not all, of the languages proposed to be CEMP, and not an innovation defining the group.

**4.3.4.6 \**daləm* ‘mind, feelings’.** CEMP-area languages exhibit one of two common strategies for the seat of emotions. The first mirrors that found in Malay and many other WMP languages: \**qatay* ‘1. liver, 2. seat of character, emotions, and values’. The second semantically mirrors that found in many Papuan languages throughout eastern Indonesia and East Timor linking to the Austronesian \**daləm* ‘1. inside’, but adding the secondary sense of ‘2. insides, seat of character, emotions, and values’. But as with many other essentially lexical innovations, the use of reflexes of \**daləm* with the sense of ‘mind, feelings’ is attested in “WMP” languages from Southeast Sulawesi. In Muna *lalo* ‘inside’ also has the meaning ‘heart, seat of emotions’, as seen in the dictionary entries listed in (26). In Tukang Besi *laro* is not regularly used to refer to emotions or the mind, but does appear in one (known) expression, in (27), in which the reference is clearly to a mental or emotional state, and not to a physical one.

39. One of the significant contributions of B. D. Grimes (1993) is demonstrating the many Melanesian-like features pervading Buru culture (in west-central Maluku), and anthropological evidence suggesting contact with a pre-Austronesian population. Donohue (2007a) presents evidence that a Papuan presence must be posited as far west as central Sumbawa, exactly at the point that the Austronesian languages begin to be classified as CEMP. As argued in Donohue (2007a), this is too fortuitous to be coincidental.

- (26) MUNA  
*lalo* ‘heart, seat of emotions’; *felalo* ‘be inside’
- (27) TUKANG BESI  
*laro* ‘inside’; *ja’o laro* ‘angry (bad inside)’

There are additional complicating factors involved in this proposed innovation. First, the pattern involving ‘(self’s) [body.part]’ to be linguistically equated with the subject of an emotion or experiential predicate is extremely common cross-linguistically. The use of “inside” to refer to ‘mind, feelings, emotions’ as this subject is more restricted, but is especially common among the Papuan languages of New Guinea (e.g., McElhanon 1977).<sup>40</sup> Away from New Guinea itself we have rich attestation from Oirata, an East Timor language spoken in Southeast Maluku. De Josselin de Jong (1937) gives the list of senses associated with Oirata *isa* shown in (28).

- (28) *isa* ‘heart, core, contents’  
*isa ahara he* ‘hopeless’ (*ahara* only appears in this compound; *he* ‘NEG’)  
*isa arutu* ‘greedy’ (*arutu* only appears in this compound)  
*isa elewe* ‘dejected’ (*alewe* only appears in this compound)  
*isa hanate* ‘compassionate’ (*hanate* ‘distress’)  
*isa huhule* ‘loathe, be sick’ (*huhule* ‘disease’)  
*isahuna* ‘in the middle’ (*huna* ‘calf [of leg]’)  
*isa iliare* ‘grow faint-hearted’ (*i-liare* ‘REFL-transformed’)  
*isa kahare* ‘craving’ (*kahare* ‘spoil, bad’)  
*isa lolo he* ‘anxious, worrying’ (*lolo* ‘good, true’; *he* ‘negative’)  
*isa malare* ‘angry, jealous’ (*malare* ‘sour, bitter, hot’)  
*isa eme halu* ‘repent, regret’ (*eme* ‘get, cause’; *halu* ‘remorse’)  
*isa muduni* ‘keep a secret’ (*muduni* ‘within’)  
*isa seile* ‘hold out, constrain oneself’ (*seile* ‘draw, pull’)  
*isatapu* ‘breast, heart’ (*tapu* ‘kernel, pit, seed’)  
*isatapu anaje* ‘think over’ (*anaje* ‘try, fetch’)  
*isatapu nanate* ‘abhor, shudder’ (*nanate* ‘APPL-stand’)  
*isatapu pai* ‘make a keepsake’ (*pai* ‘cause’)  
*isatapu ruru* ‘be moved’ (*ruru* ‘throb, shake’)  
*isa tutu* ‘like, want’ (*tutu* ‘drink’)  
*isa umumu* ‘forget’ (*umu* ‘die’)  
*isa wale* ‘gift (out of charity)’ (*wale* ‘walk, travel’)  
*isa wara* ‘at ease, content’ (*wara* ‘clear, clean, evident’)  
*isa pai wara* ‘move one’s heart, inspire with sympathy,  
satisfy’ (*pai* ‘cause’)

The Pura language of Alor-Pantar (part of a large cluster of Papuan languages) exhibits a similar pattern of polysemy: *omi* ‘1. inside (body), 2. place of belief, feelings, and emotions’. Thus *omi veŋariak* ‘like, want’, *veŋ ii-omi* ‘think’, *i-omi metma ... veŋ-aana-*

40. Most Indo-European languages associate feelings with the organ that pumps blood. Think of the metaphors, love songs, and poetry associated with English *heart*, Spanish *corazón*, Portuguese *coração*, French *coeur*, Greek *καρδία*, and so forth. Think also of English expressions such as “gut feeling,” “butterflies in X’s stomach,” “POSSESSOR’s heart jumped,” “POSSESSOR’s heart reaches out to Y,” etc., to see how common (internal) body-part metaphors are in the expression of feelings.

*maŋ* ‘believe firmly’, *omi maŋ iipi sehi* ‘confused (lit. his insides were confused)’, *omi maŋ dirat etatabi* ‘sad (lit. his insides were very sick)’, and so forth.

Support for the idea that this feature has diffused, rather than been innovated, comes from those (CMP-area) Austronesian languages that show the same semantic extension of ‘inside’ > ‘experiential subject’, but without the term being a reflex of \**dalem*, as in the Palu’e example (29). Here *une* (usually possessed) is the generic noun indicating ‘inside’, with extensions to ‘straits (the water between two land masses)’. It is an obligatory part of the expression for happiness, occurring possessed (the possessor of *une* is the experiencer) with *khata*, a lexeme not attested independently.<sup>41</sup>

- (29) *une(-n)* ‘inside; mind, feelings(-3GEN)’  
*nua unen* ‘in the house’  
*une-n khata* ‘(third person is) happy’  
 (cf. *ate* ‘liver’; *ʔbu* ‘heart’; *kəʒe* ‘stomach’; *taʔi* ‘intestines’;  
*nra* ‘think, thought’)

This second pattern is geographically widespread in the CMP-area languages. So, for example, Buru has *lale-n* ‘1. inside, 2. insides, seat of character, emotions, and values, 3. fatty tissue above the pancreas, 4. deep’, *dae huma lale-n* ‘inside the house’, *bele-k lale-n* ‘confused’, *dela-k lale-n* ‘grieve’, *dola-k lale-n* ‘upset with someone’, *odo lale-n* ‘ponder, mull over’, *lale-n haa-t* ‘happy, generous’, and so forth. It uses *nena-n* for ‘liver (organ)’. Buru does have a reflex of \**qatay* ‘liver’, *ata-n* ‘spleen (of deer)’, but that word does not have the secondary sense of seat of emotions. Dhao has *dara* ‘1. inside, 2. insides, seat of character, emotions, and values’, *ət.u dara əm.u* ‘inside the house’, *dara pəd:a* ‘sad, bitter (lit. sick, pained insides)’, *dara kateme* ‘wholeheartedly’, *dara maqera* ‘patient, thoughtful (lit. long insides)’, *dara na karefe* ‘his insides were happy’, and so forth.<sup>42</sup>

Just as there are multiple ‘inside’ centers for emotion in the CMP-area languages, so too do we find that the secondary sense of ‘liver’ is also maintained in the area. Amarasi, Baikeno, and other varieties of the Uab Meto chain have *neka-n* ‘1. liver, 2. seat of character, emotions, and values’, *neka-n a-reko-t* ‘goodness, favor, grace (lit. good liver)’, *n-taam neka-n* ‘fall in love (lit. enter his liver)’, *neka-f mese-ʔ ma an-sao-f meseʔ* ‘like-minded, in accord, living in harmony (doublet; lit. one livered and one hearted)’, *atoin neka mneʔo* ‘honest, just, fair person (lit. straight-livered person)’. The Amarasi reflex of \**daləm* ‘inside’ is *nana-n*, as in *neu in uum je nana-n* ‘(go) inside his house’; it does not carry the secondary sense. Amarasi has a low frequency reflex of \**qatay* ‘liver’, *ate-f* ‘1. liver, 2. source of life’, but both of these senses occur quite rarely in contrast to forms of *neka-n*, which can be used by a single individual dozens of times a day, mostly in its secondary sense. How the primary senses of *ate-f* and *neka-f* differ as physical organs (both described as ‘liver’ and pointing to the same part of the anatomy) is unclear.

41. Similarly, we note the presence of a large number of metaphors in Philippine languages involving ‘inside’, such as the use of *loob* ‘inside’ in Tagalog, which indicates that this is not a CEMP-specific innovation.

42. Williams-van Klinken (2007) presents a fascinating glimpse of recent developments in Tetun Dili (a Tetun-based creole). There is a wealth of long-standing traditional metaphors built around *lara-n* ‘1. inside, 2. seat of character, emotions, and values’. However, among the small Portuguese-speaking elite, mostly in speeches by high-level government and church officials, there are recent artificial coinages of a few of the same metaphors using *fua-n* ‘heart, the organ that pumps blood’.

Hoskins (1993:170) notes of the language of Kodi, in western Sumba, that “the liver is the seat of emotion, knowledge, and intentionality in Sumba. As in many Austronesian languages, emotion is described in terms of the liver (generosity is ‘wide livered,’ cowardice ‘small livered,’ and happiness a ‘beautiful liver’). The liver is the organ of sincerity, where it is not possible to dissimulate. A man’s conscience and self-awareness are found ‘inside his liver’ (*ate dalo*), and a speaker’s true intentions, veiled in public declarations, reside ‘beneath the liver’ (*pa kambu ate*), only slowly finding their way up to the lips.” The use of ‘inside’ as the seat of the emotions is found sporadically in WMP-area languages to the west and north of the CEMP-area, and is found in a large number of non-Austronesian languages in the CMP-area. There is, in short, no subgrouping value for this variable.

**4.4 CEMP REEXAMINED.** In the preceding sections we have seen that the features that are proposed as defining CEMP as a subgroup do not stand up to the accumulating data. The results of this examination are summarized in table 23; table 24 summarizes the distribution of these “innovations” with respect to the CMP-area languages, which form the westernmost region of the CEMP area.

We have little doubt that as further lexical data become available from the “WMP” languages, the last few “innovations” will also be removed from the CEMP or CMP lists. The emerging picture is that, rather than being a clearly definable subgroup, CEMP and CMP simply represent an accumulation of innovations, phonological and lexical, that build up as we look further and further to the east. The morphosyntactic innovations can all be ascribed to Papuan influence, as detailed in Donohue (2005a, 2007b).

**TABLE 23. THE CEMP INNOVATIONS RECONSIDERED**

	PHONOLOGICAL INNOVATIONS	DEFINES CEMP?
1	PMP *uliq > PCEMP *oliq ‘return’	no
2	PMP *i-sai > PCEMP *i-sei ‘who’	no
3	PMP *ma-qitam > PCEMP *ma-qetəm ‘black’	no
4	PMP *maRi > PCEMP *mai ‘come’	no
5	PMP *tudan ‘sit’ > PCEMP *todan ‘sit’	(yes)
6	PMP *inum > PCEMP *unum	no
	MORPHOSYNTACTIC INNOVATIONS	
1	Prefixal / Proclitic agreement on verb	no
2	Alienable/inalienable possession distinction	no
3	Frozen morphology	
3a	PMP *həpat > PCEMP *həpat, pat, pati ‘four’	no
3b	PMP *ma-huab > PCEMP *mawab ‘yawn’	no
3c	PMP *ma-hiaq > PCEMP *mayaq ‘shy’	no
	SEMANTIC INNOVATIONS	
1	PMP *t-ina ‘mother’ > PCEMP *t-ina ‘big’	no
2	PMP *m-udəhi ‘behind’ > PCEMP *mudi ‘back (of body)’	no
3	PMP *ma-qitam > PCEMP *ma-qetəm ‘dirty’	no
4	PMP *tuqəla[nj] > PCEMP *zuRi ‘bone’	no
5	PCEMP *daun ni qulu ‘head hair’	no
6	PCEMP *daləm ‘mind, feeling’	no

**5. CONCLUSIONS.** There is no phonological basis for either the CMP or CEMP groups as they are currently described. The morphosyntactic features proposed can all be ascribed to contact, and the lexical innovations do not withstand the scrutiny imposed by more detailed data from WMP-area languages and by excluding geographically proximal languages. Further, the lack of detailed lexical materials on most of the WMP-area languages in Indonesia means that any subgrouping based on proposed lexical innovations must be received tentatively, at best.<sup>43</sup> Of the semantic innovations proposed, we note only that the semantics involved are either not widely attested, not exclusive of WMP-area languages, not surprising cross-linguistically, or attested in at least some of the Papuan languages of the CMP-area. In short, they are not convincing enough to be counted as subgrouping evidence.

The linguistic macro-history of eastern Indonesia, where the WMP/CEMP border is said to be found, requires much more detailed investigation. The material we have presented shows clearly that some of the languages of Sulawesi share more features with at least some of the CMP-area languages than do most other languages in the WMP area. Before any further discussion of the status of “CMP” or “CEMP,” more detailed basic materials in the CMP-area and EMP languages are required, allowing us to pursue bottom-up subgroupings that take into account the complex role that diffusion has played over the years.

Methodologically, it is clear that pursuing historical linguistics—in the absence of a very clear sense of the constraints imposed by geography on the level to which those

**TABLE 24. DISTRIBUTION OF THE BASES FOR CEMP**

	West of CMP?	CMP	East of CMP?	Western Papuan	Alor-Pantar
*uliq > *oliq ‘return’	no	(yes)	yes	n/a	n/a
*i-sai > *i-sei ‘who’	(yes)	(yes)	(yes)	n/a	n/a
*ma-qitəm > *ma-qetəm ‘black’	yes	(yes)	(yes)	n/a	n/a
*maRi > *mai ‘come’	yes	(yes)	yes	n/a	n/a
*tudan ‘sit’ > *tədan ‘sit’	no	(yes)	(yes)	n/a	n/a
*inum > *unum	no	(yes)	(yes)	n/a	n/a
Prefixal agreement	yes	(yes)	(yes)	yes	yes
(In)alienability	(no)	(yes)	(yes)	yes	yes
*həpat > *həpat, *pat, *pati ‘four’	yes	(yes)	(yes)	n/a	n/a
*ma-huab > *mawab ‘yawn’	no	(yes)	(yes)	n/a	n/a
*ma-hiaq > *mayaq ‘shy’	no	(yes)		n/a	n/a
*t-ina ‘mother’ > PCEMP *t-ina ‘big’	(yes)	(yes)		n/a	n/a
*m-udəhi ‘behind’ > *mudi ‘back’		(yes)		n/a	n/a
*ma-qitəm > *ma-qetəm ‘dirty’	no	(no)		n/a	n/a
*tuqəla[nŋ] > *zuRi ‘bone’	(yes)	(no)		n/a	n/a
*daun ni qulu ‘head hair’	no	(yes)	(yes)	yes	yes
*daləm ‘mind, feeling’	no	(yes)	(yes)	yes	yes

43. Reliance on reconstructions for floral or faunal species that were not found in the WMP area are doubly suspect, because it is highly likely that the term would have been borrowed from a pre-existing, non-Austronesian source familiar with the species. The distribution of such a pseudo-reconstruction can then tell us much about the social and linguistic history of the area, but not about the subgrouping of the languages.



reconstructions can apply, and the degree to which contact and borrowing, rather than shared history, are responsible for the spread of a term—is fraught with dangers. As we have shown in the discussion of “glide truncation” in 3.1.2, the distribution of the “innovation” is highly constrained, and does not spread over the whole area (in this case, the “CMP” area). To account for the lack of inclusivity of the innovation, Blust appeals to the notion of a linkage, a chain of related speech varieties that form a network, rather than a single protolanguage. We agree with this model, but suggest that rather than there having been “CMP” linkages or “CEMP” linkages, the linkage that explains the distribution of “innovations” that are not exclusive to the CMP-area languages or the CEMP-area languages is a much larger one that should include languages ancestral to those not just in the CMP area, but also in the WMP area. The geography of the linguistic variables, as well as a model of the social environment into which the Austronesians expanded, are all factors that must be taken into account when attempting to model Austronesian linguistic history in the Melanesian area, an area that (like most of the areas where Austronesian languages are currently spoken) had a large non-Austronesian speaking population prior to the arrival of the Austronesians, which growing evidence suggests could have been maritime (e.g., O’Connor and Veth 2005). The fact that there are some discontinuous subgroups within Malayo-Polynesian (e.g., Adelaar 2005b) indicates that simply finding two languages at a “safe” distance (that is, far enough away from each other that contact and borrowing are less likely), and proposing a high-level reconstruction on the basis of the presence of an apparently cognate lexeme in the two, is not a safe policy. Without the low-level reconstructions and subgroupings, we cannot hope to fast-track our way to high-level conclusions in a way that withstands the scrutiny of the data over time.

## APPENDIX: DATA SOURCES

The sources for the languages cited here are:

- CAD: Acehnese, Aklanon-Bisayan, Ba’ngingi Sama, Balinese, Central Amis, Da’a, Ilokano, Ma’anyan, Madurese, Malagasy, Manggarai, Ngadha, Paiwan, Rote, Sarangani Blaan, Sasak, Sawai, Sika, Thao, Uma’, Yami.
- ABVD: Alune, Banda Elat, Bima, Bunun, Erai (Wetar), Favorlang, Gebe, Gorontalo, Hanunóo, Hitu, Iralalay (dialect of Yami), Iranun, Kakiduge:n Ilongot, Kayan (Uma Juman), Kédang, Kei, Kemak, Kisar, Mamanwa, Mambai, Manggarai, Maranao, Melanau, Modang, Nage, Nalik, Pazeh, Punan Kelai, Rejang, Saisiyat, Sangir, Sumbawa, Taiof, Tunjung, Uab Meto (Atoni).
- Donohue (fieldnotes): Ansus, Arguni, Busoa, Cia-Cia, Geser, Kaimbulawa, Kamaru, Kulisusu, Kumbe-waha, Lasalimu, Oirata, Onin, Palu’e, Pancana, Sekar, Tukang Besi, Wawonii, Wolio.
- Grimes (fieldnotes): Amarasi, Baikeno (Ambeno), Buru, Dhao (Ndao), Hawu (Sabu, Sawu), Helong, Lole, Termanu, Tetun, Tii, Uab Meto (Atoni). (The fieldnotes for Buru, Tetun, and Termanu include contributions from Barbara Dix Grimes, those for Helong from Misriani Balle and Stuart Cameron.)

Other sources used are:

- |  |                                       |
|--|---------------------------------------|
| Amahai: Haaksma (1933)                     | Buru: Grimes (1991a, 1991b)           |
| Arta: Reid (1989)                          | Chamorro: Topping (1975)              |
| Batuley: Nivens (pers. comm.)              | Dela: Tamelan (2007),                 |
| Bungku: Mead (1999)                        | Tamelan and B. D. Grimes (fieldnotes) |
| Bungku-Mori-Tolaki languages: Mead (1999), | Favorlang: Ferrell (1969)             |
| Ferrell (1969)                             | Gayo: Eades (2005)                    |

- Hanuó: Conklin (1953)  
 Imroing: Taber (1993), Tsuchida, Yamada,  
 and Moriguchi (1987)  
 Kambéra: Onvlee (1984)  
 Kisar: Rinooy (1886), Stresemann (1927)  
 Kola: Nivens (pers. comm.)  
 Kulisusu: Mead (1999)  
 Leti: van Engelenhoven (2004), Hudson (1967)  
 Malay: authors' own knowledge  
 Mamanwa: Reid (1971)  
 Manggarai: Verheijen (1967)  
 Maranao: ABVD, McKaughan & Macaraya (1967)  
 Matabello (a variety of Watubela?): Wallace (1962)  
 Mori Atas, Mori Bawah: Mead (1999)  
 Moronene: Mead (1999)  
 Muna: van den Berg (1996)  
 Napu: Martens (1989)  
 Nias: Sundermann (1905)  
 Nuauulu: Bolton (1990)  
 Oirata: de Josselin de Jong (1937)  
 Padoe: Mead (1999)  
 Pamona: Martens (1989)  
 Paulohi: Stresemann (1927)  
 Pazah: ABVD, Ferrell (1969)  
 Pura: K. Berry and C. Berry (fieldnotes)  
 Rejang: Jonker (1908)  
 Saisiyat: ABVD, Ferrell (1969)  
 Sangir: Steller and Aebersold (1959)  
 Simalur: Haaksma (1933)  
 Soboyo (Taliabu): Fortgens (1921)  
 Tagalog: English (1965)  
 Taloki: Mead (1999)  
 Tetun: Morris (1984), van Klinken (fieldnotes)  
 Tolaki: Mead (1999)  
 Toraja: (Woensdrecht, cited in Haaksma 1933)  
 Tugun: Hinton (2000)  
 Tukang Besi: Donohue (1999)  
 Tungak (Lavongai): Stamm (1988)  
 Tunjung: Hudson (1967)  
 Varisi, Vaghua: Tryon and Hackman (1983)  
 Watubela: Blust (1993)  
 Wawonii: Mead (1999)  
 West Tarangan: Nivens (pers. comm.)  
 Wolio: Anceaux (1987) and own notes

## REFERENCES

- Adelaar, Alexander. 2005a. The Austronesian languages of Asia and Madagascar: A historical perspective. In *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 1–42. London: Routledge.
- . 2005b. Malayo-Sumbawan. *Oceanic Linguistics* 44:357–88.
- Anceaux, J. C. 1987. *Wolio dictionary (Wolio-English-Indonesian)*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde. Dordrecht: Foris.
- Balle, Misriani. 2007. Multiple words (homonymy) or multiple senses (polysemy)? Exploring the semantics of Helong. Paper presented to the Fifth East Nusantara Conference, Universitas Nusa Cendana, Kupang, Nusa Tenggara Timur, Indonesia, August 1–3.
- Blust, Robert. 1981. The Soboyo reflexes of Proto-Austronesian \*S. In *Historical Linguistics in Indonesia*, part 1, ed. by Robert A. Blust, 21–30. Jakarta: NUSA.
- . 1982. The linguistic value of the Wallace Line. *Bijdragen tot de Taal-, Land- en Volkenkunde* 138 (2/3): 231–50.
- . 1993. Central and Central-Eastern Malayo-Polynesian. *Oceanic Linguistics* 32:241–93.
- . 2001. Malayo-Polynesian: New stones in the wall. *Oceanic Linguistics* 40:151–55.
- Bolton, Rosemary. 1990. A preliminary description of Nuauulu phonology and grammar. MA thesis, University of Texas at Arlington.
- Borroff, Marianne L. 2007. A landmark underspecification account of the patterning of glottal stop. PhD thesis, SUNY Stony Brook.
- Brandes, Jan Laurens Andries. 1884. *Bijdragen tot de Vergelijkende Klankleer der Westersche afdeeling van de Maleis-Polynesische taalfamilie*. Utrecht: P. W. van de Weijer Stoomdrukkerij.

- Clynes, Adrian. 1997. On the Proto-Austronesian “diphthongs”. *Oceanic Linguistics* 36:347–61.
- . 1999. Rejoinder: Occam and the Proto-Austronesian “diphthongs.” *Oceanic Linguistics* 38:404–08.
- Conklin, H. 1953. *Hanunóo-English vocabulary*. University of California Publications in Linguistics, Volume 9. Berkeley: University of California Press.
- Coward, David F., and Naomi E. Coward. 2000. A phonological sketch of the Selaru language. In *Spices from the East: Papers in languages of eastern Indonesia*, ed. by Charles E. Grimes, 9–54. Canberra: Pacific Linguistics.
- Crowley, Terry. 2002a. Raga. In *The Oceanic languages*, ed. by John Lynch, Malcolm Ross, and Terry Crowley, 626–37. Richmond: Curzon.
- . 2002b. Southeast Ambrym. In *The Oceanic languages*, ed. by John Lynch, Malcolm Ross, and Terry Crowley, 660–70. Richmond: Curzon.
- De Josselin de Jong, J. P. B. 1937. *Studies in Indonesian culture I: Oirata, a Timorese settlement on Kisar*. Amsterdam: Noord-Hollandsche Uitgeverij.
- Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- . 2005a. Word order in New Guinea: Dispelling a myth. *Oceanic Linguistics* 44:527–36.
- . 2005b. The pretenders to the Muna-Buton group. In *Papers in Austronesian subgrouping and dialectology*, ed. by John Bowden and Nikolaus Himmelmann, 21–35. Canberra: Pacific Linguistics.
- . 2006. Pronouns, clitics, orders and grammaticalisation: Theoretical issues in Tukang Besi. Paper presented at the Tenth International Conference on Austronesian Linguistics, Puerto Princesa, Philippines, January 17–20.
- . 2007a. The Papuan language of Tambora. *Oceanic Linguistics* 46:520–37.
- . 2007b. Word order in Austronesian: From north to south and west to east. *Linguistic Typology* 11(2): 351–93.
- . Forthcoming. Obligatory incorporation and ‘have’ in Tukang Besi. In *Language and text in the Austronesian world: Studies in honour of Ülo Sirk*, ed. by Yury Lander. Munich: Lincom Europa.
- Donohue, Mark, and Charles E. Grimes. n.d. Contact and the changes in Southern Austronesian languages. In preparation.
- Donohue, Mark, and Simon Musgrave. 2007. Typology and the linguistic macro-history of island Melanesia. *Oceanic Linguistics* 46:348–87.
- Donohue, Mark, and Antoinette Schapper. 2007. Towards a morphological history of the languages of Timor, Alor, and Pantar. Paper presented at the Fifth East Nusantara Conference. Universitas Nusa Cendana, Kupang, Nusa Tenggara Timur, Indonesia, August 1–3.
- Eades, Domenyk. 2005. *A grammar of Gayo, a language of Aceh, Sumatra*. Canberra: Pacific Linguistics.
- English, Leo James. 1965. *English-Tagalog dictionary*. Manila: Congregation of the Most Holy Redeemer.
- Ferrell, R. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Monograph 17: Institute of Ethnology, Academia Sinica.
- Fortgens, J. 1921. *Bijdrage tot de kennis van het Sobojo (eiland Taliaboe, Soelagroep)*. The Hague: Martinus Nijhoff.
- Greenhill, S. J., Robert Blust, and R. D. Gray. 2005-08. The Austronesian Basic Vocabulary Database. <http://language.psy.auckland.ac.nz/austronesian>.
- Grimes, Barbara Dix. 1993. The pursuit of prosperity and blessing: A study of social life and symbolic action on Buru Island, eastern Indonesia. PhD thesis, Australian National University.

- Grimes, Charles E. 1991a. Central Malayo-Polynesian: An interim evaluation. Paper presented at the Sixth International Conference on Austronesian Linguistics. Honolulu, Hawai'i, May 20–24.
- . 1991b. The Buru language of eastern Indonesia. PhD thesis, Australian National University.
- . 2000. Introduction: New information filling old gaps in eastern Indonesia. In *Spices from the East: Papers in languages of eastern Indonesia*, ed. by Charles E. Grimes, 1–8. Canberra: Pacific Linguistics.
- . 2007. Building the jigsaw puzzle: Seeing the big picture still depends on solid descriptive work for individual languages and cultures. Keynote address at the Fifth East Nusantara Conference, Universitas Nusa Cendana, Kupang, Nusa Tenggara Timur, Indonesia, August 1–3.
- . Forthcoming. Hawu and Dhao in eastern Indonesia: Revisiting their relationship. In *Typological and areal analyses: Contributions from East Nusantara*, ed. by Michael Ewing and Marian Klamer. Leiden: KITLV Press.
- Grimes, Charles E., and Mark Donohue. n.d. *The linguistic history of Eastern Indonesia*. In preparation.
- Grimes, Charles E., and Barbara D. Grimes. 1987. *Languages of South Sulawesi*. Canberra: Pacific Linguistics.
- Grimes, Charles E., Tom Therik, Barbara Dix Grimes, and Max Jacob. 1997. *A guide to the people and languages of Nusa Tenggara*. Paradigma B–1. Kupang: Artha Wacana Press.
- Haaksmā, Rémy. 1933. Inleiding tot de studie der vervoegde vormen in de Indonesische talen. PhD thesis, Leiden.
- Hinton, Bryan. 2000. The languages of Wetar: Recent survey results and word lists with notes on Tugun grammar. In *Spices from the East: Papers in languages of eastern Indonesia*, ed. by Charles E. Grimes, 105–30. Canberra: Pacific Linguistics.
- Hoskins, Janet. 1993. Violence, sacrifice, and divination: Giving and taking life in eastern Indonesia. *American Ethnologist* 20(1): 159–78.
- Hudson, Alfred B. 1967. Padju Epat: The ethnography and social structure of a Ma'anyan Dajak group in Southeastern Borneo. PhD diss., Cornell University.
- Hull, Geoffrey. 1999. *Standard Tetum–English dictionary*. Sydney: Allen and Unwin.
- Jonker, J. C. G. 1908. *Rottineesch–Hollandsch woordenboek*. Leiden: Brill.
- Laycock, D. C. 1986. Papuan languages and the possibility of semantic classification. *Papers in New Guinea linguistics* No. 24, 1–10. Canberra: Pacific Linguistics.
- Lynch, John. 2002. The Proto-Oceanic labiovelars: Some new observations. *Oceanic Linguistics* 41:310–62.
- Lynch, John, Malcolm Ross, and Terry Crowley. 2002. *The Oceanic languages*. Richmond: Curzon.
- Mahdi, Waruno. 1994. Some Austronesian maverick protoforms with culture-historical implications—II. *Oceanic Linguistics* 33:431–90.
- Martens, Michael P. 1989. Proto-Sulawesi-Pamona: Reconstruction of the protolanguage of a language subgroup in Sulawesi. MS, Hasanuddin University and The Summer Institute of Linguistics.
- Matisoff, James A. 1991. The mother of all morphemes: Augmentatives and diminutives in areal and universal perspective. In *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society (SEALS)*, ed. by Martha Ratliff and Eric Schiller, 293–349. Tempe: Arizona State University, Program for Southeast Asian Studies.
- McElhanon, Kenneth A. 1977. Body image idioms in Irianese and Papua New Guinean languages. *Irian* 6(3): 3–27.
- McKaughan, Howard P., and B. A. Macaraya. 1967. *A Maranao dictionary*. Honolulu: University of Hawai'i Press.

- Mead, David E. 1999. *The Bungku-Tolaki languages of South-Eastern Sulawesi, Indonesia*. Canberra: Pacific Linguistics.
- . 2003. Evidence for a Celebic supergroup. In *Issues in Austronesian historical phonology*, ed. by John Lynch, 115–41. Canberra: Pacific Linguistics.
- Mills, Roger F., and John Grima. 1980. Historical developments in Lettinese. In *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*, ed. by Paz Buenaventura Naylor, 273–84. Ann Arbor: University of Michigan.
- Morris, Cliff. 1984. *Tetun–English dictionary*. Canberra: Pacific Linguistics.
- Newman, John. 1984. Nasal replacement in western Austronesian: An overview. *Philippine Journal of Linguistics* 15(2): 1–17.
- Nothofer, Bernd. 1992. CEMP and CMP – two Austronesian subgroups? A critical evaluation of Blust (1990) and Grimes (1991). Paper presented at the Second International Maluku Studies Conference, Honolulu, Hawai‘i, July 30–August 2.
- O’Connor, Sue. 2003. Nine new painted rock art sites from East Timor in the context of the western Pacific region. *Asian Perspectives* 42(1): 96–128.
- O’Connor, Sue, and Peter Veth. 2005. Early Holocene shell fishhooks from Lene Hara Cave, East Timor, establish complex fishing technology was in use in Island South-east Asia five thousand years before Austronesian settlement. *Antiquity* 79:1–8.
- Onvlee L., U. H. Kapita, and P. J. Luijendijk. 1984. *Kambaraas (Oost-Soembaas)–Nederlands Woordenboek*. Dordrecht: Foris/Koninklijk Instituut voor Taal-, Land- en Volkenkunde.
- Pawley, Andrew, and Medina Pawley. 1994. Early Austronesian terms for canoe parts and seafaring. In *Austronesian terminologies: Continuity and change*, ed. by A. K. Pawley and M. D. Ross, 329–61. Canberra: Pacific Linguistics.
- Reesink, Ger P. 2002. Clause-final negation: Structure and interpretation. *Functions of Language* 9(2): 239–68.
- Reid, Lawrence A. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication no. 8. Honolulu: University of Hawai‘i Press.
- . 1989. Arta, another Philippine Negrito language. *Oceanic Linguistics* 28:47–74.
- Rinnooy, N. 1886. Maleisch–Kissersch woordenlijst. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 31:149–213.
- Ross, Malcolm. 1995. Some current issues in Austronesian linguistics. In *Comparative Austronesian dictionary*, ed. by Darrell Tryon, 45–120. Berlin: Mouton de Gruyter.
- . 1998. Possessive-like attributive constructions in the Oceanic languages of Northwest Melanesia. *Oceanic Linguistics* 37:234–76.
- . 2002. The history and transitivity of western Austronesian voice and voice-marking. In *The history and typology of western Austronesian voice systems*, ed. by Fay Wouk and Malcolm Ross, 17–62. Canberra: Pacific Linguistics.
- . 2003. Properties of inanimate objects. In *The lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, vol. 2: *The physical world*, ed. by Andrew K. Pawley, Malcolm Ross, and Meredith Osmond, 197–220. Canberra: Pacific Linguistics.
- Stamm, Joseph, Tamsin Donaldson, Clive H. Beaumont, and M. J. Lloyd. 1988. *Lavongai materials*. Canberra: Pacific Linguistics.
- Starosta, Stanley, Andrew K. Pawley, and Lawrence Reid. 1982. The evolution of focus in Austronesian. In *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and Stephen Wurm, 145–70. Canberra: Pacific Linguistics.
- Steller, K. G. F., and W. E. Aebersold. 1959. *Sangirees–Nederlands Woordenboek met Nederlands–Sangirees Register*. The Hague.
- Stresemann, E. 1927. Die Lauterscheinungen in den ambonischen Sprachen. *Zeitschrift für Eingeborenen-Sprachen*, Vol. 10. Berlin: Dietrich Reimer.
- Sundermann, H. 1905. *Niassisch–Deutsches Wörterbuch*. Moers: J. W. Spaarmann.

- Taber, Mark. 1993. Toward a better understanding of the indigenous languages of southwestern Maluku. *Oceanic Linguistics* 32:389–441.
- Tamelan, Thersia. 2007. An overview of Dela-Oenale verb morphology: A language spoken in western Rote. Paper presented to the Fifth East Nusantara Conference, Universitas Nusa Cendana, Kupang, Nusa Tenggara Timur, Indonesia, August 1–3.
- Topping, Donald M. 1975. *Chamorro–English dictionary*. Honolulu: University of Hawai‘i Press.
- Tryon, Darrell, ed. 1995. *Comparative Austronesian dictionary*. Berlin: Mouton de Gruyter.
- Tryon, D. T., and B. Hackman. 1983. *Solomon Islands languages: An internal classification*. Canberra: Pacific Linguistics.
- Tsuchida, S., Y. Yamada, and T. Moriguchi. 1987. *Lists of selected words of Batanic languages*. Department of Linguistics, Faculty of Letters, University of Tokyo.
- van den Berg, René. 1996. *Muna–English dictionary*. Leiden: KITLV Press.
- . 2003. The place of *Tukang Besi* and the Muna-Buton languages. In *Issues in Austronesian historical phonology*, ed. by John Lynch, 87–113. Canberra: Pacific Linguistics.
- van Engelenhoven, Aone. 2004. *Leti: A language of southwest Maluku*. Leiden: KITLV Press.
- Verheijen, J. A. J. 1967. *Kamus Manggarai. I, Manggarai–Indonesia. II, Indonesia–Manggarai*. The Hague: Martinus Nijhoff.
- Wallace, Alfred Russell. 1962 [1869]. *The Malay archipelago*. New York: Dover.
- Wijngaarden, J. K. 1896. *Sawuneesch woordenlijst*. Koninklijk Instituut voor Taal-, Land- en Volkenkunde. The Hague: Martinus Nijhoff.
- Williams-van Klinken, Catharina. 2007. Is he hot-blooded or hot inside? Expression of emotion and character in Tetun Dili. Paper presented to the Fifth East Nusantara Conference, Universitas Nusa Cendana, Kupang, Nusa Tenggara Timur, Indonesia, August 1–3.
- Wolff, John U. 1996. The development of the passive verb with pronominal prefix in western Austronesian languages. In *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, ed. by Bernd Nothofer, 15–40. Hamburg: Abera.