Is Malayo-Polynesian a primary branch of Austronesian? A view from morphosyntax

Victoria Chen[1], Jonathan Kuo[2], Kristina Gallego[3], and Isaac Stead[4]

[1]Victoria University of Wellington
[2]National Taipei University of Technology
[3]Australian National University/University of the Philippines Diliman
[4]Max Planck Institute for Evolutionary Anthropology

An understudied morphosyntactic innovation, reanalysis of the Proto-Austronesian stative intransitive prefix *ma- as a transitive affix, offers new insights into Austronesian higher-order subgrouping. While Malayo-Polynesian – the linguistic subgroup comprising all Austronesian languages spoken outside Taiwan – is currently considered an Austronesian primary branch bearing no identifiably closer relationship with any linguistic subgroup located in the Austronesian homeland due to a lack of evidence indicating its precise origin (Ross 2005; Blust 1999, 2013), the fact it displays the same innovative use of *ma-* with four East Formosan languages distributed around the coastline of Taiwan (Amis, Siraya, Kavalan, Basay-Trobiawan) and shares the same merger (PAn *C/t) with this group suggests East Formosan and Malayo-Polynesian may share a common origin. This proposal points to a renewed subgrouping more consistent with a socio-historical picture where the out-of-Taiwan population descended from a seafaring community expanding to the Batanes and Luzon Islands after having developed a seafaring tradition. It also aligns with a major finding in recent genetic studies that the Amis have a significantly closer relationship with Austronesian communities outside Taiwan (Trejaut et al. 2005; Capelli et al. 2005). Future investigation of more potential shared innovations between Malayo-Polynesian and East Formosan could shed further light on their interrelation.

1 The puzzle

Malayo-Polynesian, the linguistic subgroup comprising all Austronesian languages spoken outside the homeland, Taiwan, is traditionally considered an independent language group defined by a series of innovations shared across extra-Formosan languages (e.g. Mills 1975; Dahl 1973; Blust 1977, 1999, 2001, 2013; Liao 2011; Ross 2002, 2005, 2012, a.o.).[1][2]

(1) *Major innovations defining Malayo-Polynesian*[3]
  a. PAn *C/t > PMP *t (merger)
  b. PAn *N/n > PMP *n (merger)

[1]List of abbreviations: ASP: aspect AUX: auxiliary; AV: actor voice; CEMP: Central-Eastern Malayo-Polynesian; CMP: Central Malayo-Polynesian; CN: common noun; EMP: Eastern Malayo-Polynesian; GEN: genitive; ID: indefinite INTR: intransitive; IRR: irrealis; LK: linker; LOC: locative; OBL: oblique; NPST: non-past; PAN: Proto-Austronesian; PL: plural; PMP: Proto-Malayo-Polynesian; PN: proper name; PRF: perfective; PST: past tense; PV: patient voice; RED: reduplication; REAL: realis; SG: singular; SHWNG: South Halmahera-West New Guinea; STAT: stative; TR: transitive

[2]See Bellwood (1984-5, 2005a, b, 2017) and Blust (1984, 1999) for specific arguments for Taiwan as the center of Austronesian diaspora.

[3]PAn and PMP in (1) stand for 'Proto-Austronesian' and 'Proto-Malayo-Polynesian,' respectively. Sources of each innovation: (1a-c): Mills 1975; Blust 1999, 2001, 2013; Ross 2005; (1b-c): Blust 1999, 2001, 2013; Ross 2005; (1d): Blust 2001; (1e): Blust 1977, 2011, 2013; Ross 2005; (1f): Ross 2002; Liao 2011; Blust 2013; (1g): Blust 2001, 2013; (1h): Blust 2001, 2013.

c. PAn *S/h > PMP *h (merger; with irregular loss of PAn *s > zero)

d. PAn *l > R/__Vj

e. PAn *mu '2PL.GEN' > PMP *mu '2SG.GEN'

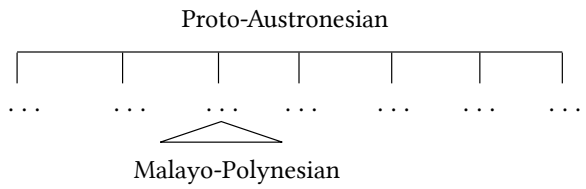
f. Morphological innovations: PMP *maN- 'Actor Voice morpheme', *paN- 'distributive', *paR- 'durative, reciprocal', *maR- 'Actor Voice morpheme'

g. Metatheses: PAn *-CV*S* > variation between *-hVC and *-CV*h* (with *S > *h*)

h. Irregular changes: PAn *biRbiR > PMP *bibiR 'lips'; PAn *Siwa > PMP *siwa 'nine'; PAn *paNudaN > PMP *paNdan 'pandanus' (among others)

While the innovations in (1) offer solid evidence that all Austronesian languages spoken outside Taiwan descend from a single origin, they do not provide direct evidence for the external affiliation of Malayo-Polynesian – namely, its relationship with the languages spoken in the Austronesian homeland, Taiwan. Does Malayo-Polynesian bear a closer relationship with a certain Formosan subgroup or subgroups than others? From a socio-historical point of view, the ex-Taiwan population would in principle belong to one of the indigenous groups located in Taiwan, speaking a language descended from a certain Formosan subgroup and giving rise to a subgrouping scenario such as that schematized in (2).

(2)



Such a connection may be adduced through exclusively shared innovations between Malayo-Polynesian and a particular Formosan subgroup. If no linguistic evidence indicates such a connection, we would have to assume either (a) the out-of-Taiwan event took place before any distinctive innovations had developed, or (b) the ancestor of Malayo-Polynesian originated as a distinct speech community when reaching Taiwan, then moved off the island without leaving any modern descendants in the homeland.[4] Given the relatively long pause in Taiwan before the settlement of Luzon (500 - 1000 years, see Bellwood 1979, 1980, 1983, 1984-5, 1988, 1995, 2007, 2017 for details), both scenarios are less likely, unless no positive evidence indicates other alternatives.

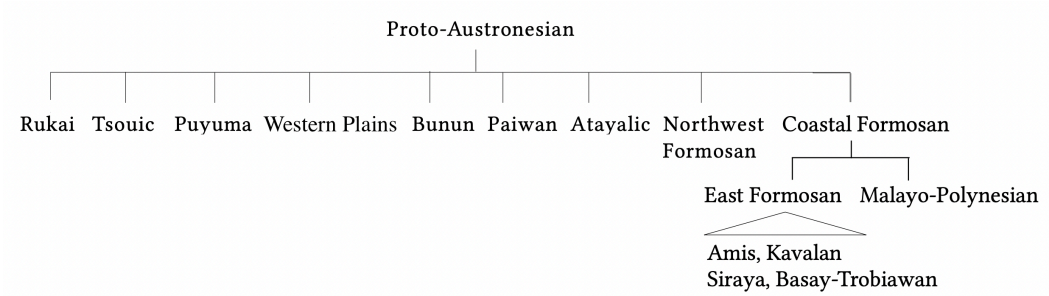
To date, the linguistic evidence for the origin of Malayo-Polynesian has remained vague. From the perspective of sound change, two Formosan subgroups are most likely to be the closest relatives of Malayo-Polynesian: (i) Bunun, which shares the mergers of PAn *C/t and *N/n with Malayo-Polynesian, and (ii) the East Formosan subgroup defined by the merger of PAn *j and *n (Blust 1999), which comprises four coastal Formosan languages (Amis, Siraya, Kavalan, Basay-Trobiawan) that reflect the same merger of PAn *C and *t with Malayo-Polynesian. However, both mergers are commonly treated as drifts (Blust 1999; Ross 2009), as neither is shared exclusively between these groups. Accordingly, Malayo-Polynesian is conventionally considered a higher-order subgroup of Austronesian that has no particularly close relation with any specific Formosan subgroup (e.g. Blust 1999; Ross 2009, 2012).

In this paper we present new evidence for an alternative view: Malayo-Polynesian is not an Austronesian primary branch, but a subgroup embedded under an Austronesian primary branch that also subsumes East Formosan languages. Evidence for this claim comes from an understudied functional variation of the Proto-Austronesian morpheme *ma- among higher-order Austronesian subgroups. While the stative intransitive function of this morpheme is attested across western Austronesian, an innovative use is exclusively shared between Malayo-Polynesian and East Formosan, both of whose modern descendants possess seafaring traditions. We propose accordingly that East

[4]This proposal builds on the (nonconventional) assumption that Proto-Austronesian split off prior to the settlement of Taiwan.

Formosan may be the closest relative of Malayo-Polynesian in Taiwan and together with it constitutes a single daughter of Austronesian with two children, as in (3). This proposal points to a subgrouping scenario more consistent with that expected from a socio-historical perspective whereby the out-of-Taiwan population descended from a speech community located in the linguistic homeland (2).

(3) *Austronesian higher-order subgrouping*



The remainder of this paper is organised as follows. In the next section, we review recent proposals for Austronesian higher-order subgrouping, focusing on the treatment of Malayo-Polynesian. Section 3 surveys two distinct uses of *ma- among higher-order Austronesian languages. Sections 4 and 5 present arguments for one of the two uses of *ma- as a true case of shared innovation. Section 6 presents further evidence for the proposed subgrouping and discusses consistent inferences from recent genetics research. Section 7 concludes. Except where otherwise indicated, the data presented in this paper comes from primary fieldwork over a period of five years from 2015 to 2020.

2 A note on Austronesian higher-order subgrouping

Before entering into the core discussion, we begin with an overview of the interrelationships of Malayo-Polynesian with other higher-order Austronesian subgroups. Current controversies in Austronesian higher-order subgrouping boil down to three issues, summarized in (4).

(4) a. Whether or not all Austronesian languages except Rukai, Tsou, and Puyuma constitute a single primary branch (Ross 2009, 2012).

b. Whether or not all Austronesian languages except Rukai constitute a single primary branch (Starosta 1995; Aldridge 2016).

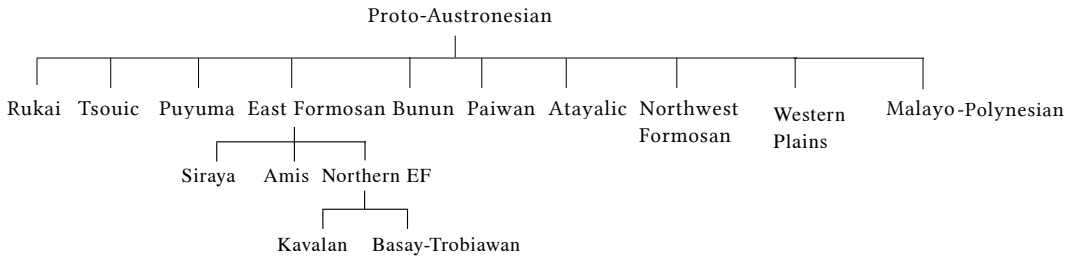
c. Whether or not Malayo-Polynesian is an independent primary branch (Blust 1999; Ho & Yang 2000; Ross 2020).

These proposals yield distinct interpretations of the position of Malayo-Polynesian (MP): (i) MP as an Austronesian primary branch (henceforth Proposal 1), (ii) MP as a first-order subgroup of an Austronesian primary branch (Proposal 2), and (iii) MP as a second or lower order subgroup of an Austronesian primary branch (Proposal 3).[5]

Proposal 1 was put forward in Blust (1999), which treats Malayo-Polynesian as a first-order offshoot of Proto-Austronesian along with nine other branches located in Taiwan, illustrated in (5). This proposal draws on the absence of exclusively shared phonological innovations between Malayo-Polynesian with any Formosan subgroups. Accordingly, Malayo-Polynesian is considered an independent primary branch.

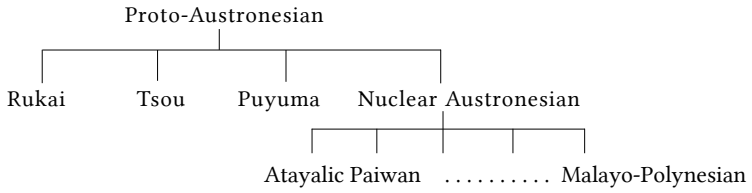
[5]This summary excludes earlier subgrouping proposals that did not present specific evidence (e.g. Dyen 1965; Ferrell 1969; Reid 1982; Harvey 1982) for the treatments and those that have been critically reviewed by recent work (e.g. Sagart 2004; Aldridge 2016). For a more detailed discussion of these proposals, see Ross (2012), Blust (2013), Blust (2015), and Blust & Chen (2017).

(5) *Blust (1999)*



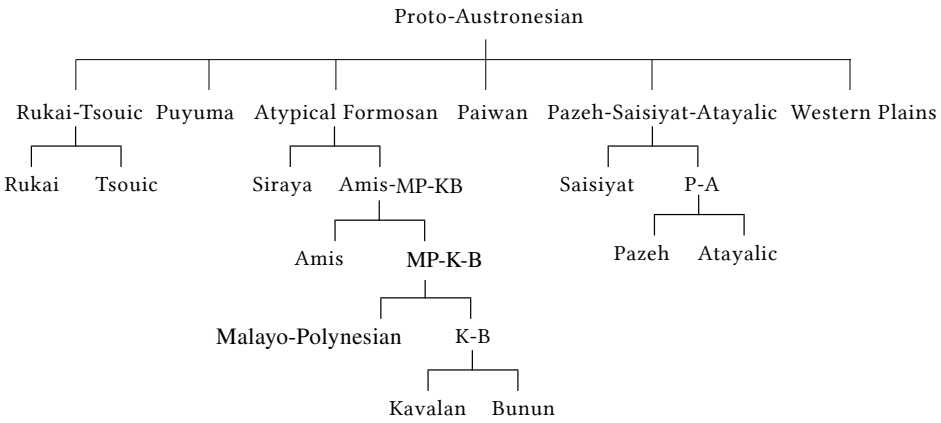
Ross (2009, 2012) argues instead that Malayo-Polynesian is a subgroup of Nuclear Austronesian, an Austronesian primary branch that comprises all Austronesian languages except Rukai, Tsou, and Puyuma, illustrated in (6). This view draws on the observation that Proto-Malayo-Polynesian, along with all Formosan languages except Rukai, Tsou, and Puyuma, displays a morphological paradigm that shows nominalizer/voice affix homophony. On the assumption that this trait reflects a single shared innovation, Malayo-Polynesian was grouped with the majority of Formosan languages under the primary branch defined by this alleged post-PAn innovation.[6]

(6) *Ross (2009, 2012)*



Yet a third proposal maintains that Malayo-Polynesian falls under an Austronesian primary branch that also subsumes three Formosan languages, Siraya, Kavalan, and Bunun (Ho & Yang 2000), illustrated in (7). This proposal relies critically on the assumption that the merger of PAn *C and *t reflected in all these languages reflects a single shared innovation, rather than independent drifts. Accordingly, Malayo-Polynesian is placed under a subgroup of this branch that also comprises Amis, Kavalan and Bunun, defined by the merger of *S/s attested in Malayo-Polynesian and all three languages.[7]

(7) *Ho & Yang (2000)*



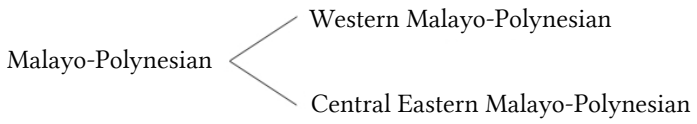
[6] Aldridge (to appear) puts forward a similar view whereby Malayo-Polynesian falls under a linguistic subgroup defined by nominalizer/voice affix homophony. As this proposal resembles Ross (2009, 2012) we do not discuss it separately.

[7] For more information on recent proposals for Austronesian primary-level subgrouping, see Blust & Chen (2017), Chen (2017), and Blust (2019) for details.

As seen above, neither nominalizer/voice affix homophony (6) nor the merger of PAn *C and *t (7) presents definite evidence for the origin of Malayo-Polynesian. The question whether the out-of-Taiwan population bears a closer relationship to a particular Formosan subgroup thus remains unclear.

Turning now to the internal subgrouping of Malayo-Polynesian, two competing proposals deserve a note. Under the traditional view (Blust 1983-84, 1991, 1999), Proto-Malayo-Polynesian underwent a binary split, giving rise to two primary branches, Western Malayo-Polynesian and Central-Eastern Malayo-Polynesian, as in (8). However, as Ross (1995:67) points out, given the standard view of Malayo-Polynesian migration patterns, Western Malayo-Polynesian as a group only reflects the speech of all of the communities that remained behind rather than joining the migration of the Central-Eastern speech community, and does not represent a single proto-language. Blust (1984) also acknowledges the lack of strong evidence defining WMP as a coherent subgroup. Building on this observation is a more recent proposal that Proto-Malayo-Polynesian is better viewed as having undergone a rapid split into 9 primary branches within a period of 500 years (Smith 2017), illustrated in (9). We adopt this new proposal for Malayo-Polynesian primary-level subgrouping as it has been shown to be more consistent with recent archaeological findings (Bellwood 1988, 1989, 2007; Kirch 2002; Ward, Athens, & Hotton 1998).[8]

(8) *Malayo-Polynesian primary-level subgrouping (Blust 1983-84, 1993, 1999)*



(9) *Malayo-Polynesian primary-level subgrouping (Smith 2017)*

With this background in mind, we are now ready to turn to the functional variation of the affix *ma- in western Austronesian, which casts new light on Austronesian higher-order subgrouping.

3 Two functions of *ma-* in western Austronesian

A comparative look at the morpheme *ma-* in higher-order western Austronesian languages reveals two distinct functions: (i) *ma-* as a stative prefix used in intransitives, and (ii) *ma-* as a Patient Voice-like affix used in transitives. While the former is attested across western Austronesian, the latter is

[8] Archaeological evidence shows that agriculturalists' settlement of Borneo, Java, Sumatra, Sulawesi, Timor, Halmahera, the Marianas, and Palau (between 4,000 and 3,500 BP) were not much later than that of the Philippines (4,500 BP). This contradicts the subgrouping proposal in (8) and suggests a rapid expansion after the settlement of Luzon. See Smith (2017) for details.

rare in Formosan languages and attested only in a subset of Malayo-Polynesian languages.[9] We begin with a summary of the distribution of each function, and proceed with a discussion of the implication of the distribution observed.

3.1 *ma-* as a stative intransitive prefix

The stative intransitive use of *ma-* is well-documented in the literature (see, e.g. Himmelmann 2004; Huang & Song 2006; Blust 2009, 2013; Ross 2015; Blust & Trussel ongoing, and the literature cited in these works). Across western Austronesian, this affix commonly combines with adjectival verbs in stative intransitive clauses in western Austronesian. In such constructions, the sole argument (usually a theme or an experiencer) bears subject-marking (labeled as PIVOT throughout the paper), analogous to other Actor Voice-marked intransitives. This case pattern is illustrated with the Tagalog examples (10a-c).

(10) *Tagalog*

a. **Ma**-taba **ang babae**.
STAT.AV-be.fat PIVOT **woman**
'The woman is fat.' (*ma-clauses*)

b. K**<um>**anta **ang babae**.
<AV>sing PIVOT **woman**
'The woman sang.' (Actor Voice clauses)

c. D**<um>**ating **ang babae**.
<AV>arrive PIVOT **woman**
'The woman arrived.' (Actor Voice clauses)

According to our survey, this function of *ma-* is attested in all Austronesian primary branches regardless of which subgrouping proposal is adopted (Ho 1998; Blust 1999; Starosta 1995; Sagart 2004; Ross 2009). Accordingly, *ma-* as a stative intransitive prefix can be uncontroversially reconstructed to Proto-Austronesian, as standardly assumed (e.g. Blust 2013; Ross 2015; Blust & Trussel ongoing). The data below demonstrates this function in 12 selected languages, each representing a different higher-order subgroup of Austronesian. Square brackets following the language name indicate the language's subgrouping affiliation under Blust's (1999) subgrouping and Smith's (2017) proposal for Malayo-Polynesian primary-level subgrouping.[10]

(11) a. **Ma**-liyay na bangsaran.
AV.STAT-be.drunk DEF.PIVOT young.man
'The young man is drunk.' (Puyuma [Puyuma])

b. Na=**ma**-Leva-Leva=mun?
PRF=AV.STAT-RED-be.joyful=2PL.PIVOT
'Have you been happy?' (Chang 2006:269) (Paiwan [Paiwan])

c. **Ma**-diqla bahi.
AV.STAT-be.bad dream.PIVOT
'The dream is bad.' (de Busser 2009:328) (Bunun [Bunun])

d. Hae:wan **ma**-skes.
night AV.STAT-be.cold
'At night it is cold.' (Zeitoun & Chu 2015:514) (Saisiyat [Northwestern Formosan])

e. **Ma**-pteng qo tao qandelak.
AV.STAT-be.hungry PIVOT man tomorrow
'The man will be hungry tomorrow.' (Reid 1966:127) (Ivatan [Batanic, MP])

[9] In some western Austronesian languages, the morpheme *ma-* bears other functions. We do not discuss them here as these functions are not directly relevant to our subgrouping argument. See Ross (2015) and Himmelmann (2004) for details.

[10] We label the subgrouping affiliation of Ivatan as 'Batanic' (and not 'Philippines') in (10e), as the linguistic position of the Batanic languages has remained controversial (see Ross 2005, 2020; Blust 2019, 2020; Liao 2020 for the discussion.

f. **M**-sibus beyuq ka walu ga.
AV.STAT-sweet very PIVOT honey that
'That honey is very sweet.' (ODFL) (Seediq [Atayalic])

g. **Ma**-adraw kay awlru-su.
AV.STAT-be.big PIVOT head-2SG.POSS
'Your head is big.' (ODFL) (Rukai [Rukai])

h. **Ma**-ta-talino ang=mga-bata-ng Intsik.
AV.STAT-PL-be.smart PIVOT=PL=child=LK Chinese
'The Chinese children are bright.' (Kroeger 1991:24) (Tagalog [Philippines, MP])

i. **Ma**-guf i geftåo na biha.
AV.STAT.be.happy the generous LK old.woman
'The generous old woman was happy.' (Chung 2020:159) (Chamorro [Chamorro, MP])

j. Ng kmal mle **me**-rau.
3SG.S very AUX.PST AV.STAT-be.rich
'He was very rich.' (*Chedaol Biblia*, *Matthew 19:22*) (Palaun [Palauan, MP])

k. Glas mo **ma**-bila.
glass 3SG STAT.INTR-shatter
'The glass is shattered. (Tamambo [CEMP, MP])

l. No-**mo**-nini.
3R-STAT.INTR-cold
'They are getting cold.' (Donohue 1999:157) (Tukang Besi [Western Indonesian])

Given this wide distribution we can conclude that the stative intransitive use of *ma-* is best viewed as a retention from Proto-Austronesian inherited by the majority of higher-order subgroups. See Blust (2013), Ross (2015), and Blust & Trussel (ongoing) for the same assumption.

3.2 *ma-* as a Patient Voice-like affix used in transitives

Yet an understudied function of *ma-* is attested in a subset of western Austronesian languages, where the affix appears in transitive clauses characterized by a genitive-marked initiator and a theme in subject-marking.[11] Consider the data below from Amis (Formosan) (12) and Tagalog (Malayo-Polynesian) (13). Examples (12a) and (13a) illustrate the intransitive use of the affix; (12b) and (13b) demonstrate the transitive use with a genitive-marked initiator.

(12) a. **Ma**-curah ku lumaq.
MA-burn PIVOT house
'The house burned.'

b. **Ma**-curah **ni** **Kulas** ku lumaq.
MA-burn GEN **Kulas** PIVOT house
'Kulas burned the house.' (Amis [East Formosan])

(13) a. **Na**-sunog ang bahay.
MA.REAL-burn PIVOT house
'The house burned.'

b. **Na**-sunog **ni** **Ivan** ang bahay.
MA.REAL-burn PN.GEN **Ivan** PIVOT house
'Ivan accidentally burned the house.' (Tagalog [Philippines, Malayo-Polynesian])

[11] As the genitive-marked argument is not restricted to the agent, we adopt the term 'initiator' here to refer to noun phrases bearing either the agent or causer role.

In both (12b) and (13b), the presence of the genitive initiator alters the sentence from stative to eventive/dynamic. The GEN-PIVOT case frame is reminiscent of canonical Patient Voice (PV) constructions, which feature the same argument-marking pattern. Consider the examples in (14)-(15).[12]

(14) a. **Ma**-curah **ni** **Kulas** ku lumaq.
MA-burn GEN **Kulas** PIVOT house
'Kulas burned the house.'

b. Curah-**en ni** **Kulas** ku lumaq.
burn-PV GEN Kulas PIVOT house
'Kulas burned the house.' (Amis [East Formosan])

(15) a. **Na**-sunog **ni** **Ivan** ang bahay.
MA.REAL-burn PN.GEN **Ivan** PIVOT house
'Ivan (accidentally) burned the house.'

b. S<in>unog ni Ivan ang bahay.
burn<PV.PRF> PN.GEN Ivan PIVOT house
Ivan burned the house.' (Tagalog [Proto-Philippines, Malayo-Polynesian])

As seen above, both the *ma*-marked clauses and the two-place constructions marked with a canonical Patient Voice affix (*-en* in Amis and *-in* in Tagalog) bear a genitive initiator and a pivot-marked theme, and the two constructions differ only in some subtle semantic interpretations. In Amis, *ma*-marked clauses often denote less volitional events with an emphasis placed on the affectedness of the undergoer (Tsukida 2005; Wu 2006; see also Huang & Sung 2008 for a similar description for Kavalan), in contrast to the *en*-marked Patient Voice construction, which usually features a volitional initiator (Wu 2006:39–41, 269). Similarly, the *ma*-marked construction in Tagalog often bears an accidental or abilitative reading (Himmelmann 2004, 2006), in contrast to the PV-marked construction, which denotes volitionality.[13] Given their structural similarities, much literature has described the *ma*-marked transitive clauses as a type of Patient Voice construction (e.g. Amis: Tsukida 2005; Wu 2006; Kavalan/Amis: Huang & Song: 2006; Ivatan: Reid 1966; Proto-Paitanic: Lobel: 2013; Itbayaten: Yamada 2014; Yami: Rau & Dong (2006); Cebuano: Tanangkingsing (2013); inter alia).

Crucially, in languages that employ only the canonical stative function of the *ma*-morpheme, an initiator can never be incorporated as a genitive-marked constituent, hence the sentence remains intransitive. To the best of our knowledge, in these languages, oblique case-marking is the only option for incorporating an initiator. Consider the examples below from the Formosan languages Paiwan and Puyuma (16a-c).[14] Each represents a distinct Austronesian primary branch.[15]

(16) a. **Ma**-takeDus a za kupu ni 'aLuay **tay** **Kalalu**.
AV.STAT-touch PIVOT that cup PN.POSS 'aLuay PN.OBL **Kalalu**
'Kalalu (unintentionally) touched 'aLuay's cup'.' (Chang 2006:214) [Paiwan]

[12] In Tagalog and many other Malayo-Polynesian languages of the Philippines and northern Borneo the sequence *ma-* inflects for mood and surfaces as *na-* in realis environments. See Himmelmann (2004) for details.

[13] In some Philippine languages the transitive use of this morpheme also bears an abilitative reading and the actual interpretation of the sentence is determined by context (abilitative vs. accidental). Note, however, that the accidental/abilitative reading is not attested with available descriptions of *ma-* (and its equivalent) in Chamorro and Palauan (two Malayo-Polynesian primary branches). This suggests that these two readings may be secondary innovations not reconstructable to Proto-Malayo-Polynesian. Further inter-language variation in lexical semantics of this morpheme deserves future investigation, and is beyond the scope of this paper. Here, we focus on the diachronic implications of the morpheme's reanalysis from an intransitive affix to a transitive affix.

[14] Genitive NPs in Puyuma are obligatorily encoded as a pronominal proclitic. The proclitic can be optionally cross-referenced by a third-person full NP, as seen in (16c). According to primary fieldwork on Nanwang Puyuma and available descriptions (Teng 2008), neither structures is possible for *ma*-marked clauses, illustrated in (16c).

[15] Note that the [PIVOT - OBLIQUE] case frame is also available to languages that display the transitive use of *ma-*. Consider: Tagalog *Na-inis* ***ng/sa*** *(GEN/OBL)* ***bata*** *ang ale* 'The woman was annoyed by the child.' Amis: *ma-'esam kura tamdaw* ***tura*** *(OBL)* ***lalangaw*** 'The person was annoyed with the fly'.

b. **Ma**-binga=ku **dra** **tu-nirengayan kan Siber**.
AV.STAT-annoy=1SG.PIVOT ID.OBL **3.POSS-word** LK **Siber**
'I was annoyed by Siber's words.' [Puyuma]

c. *$\textbf{Tu}_i$=**ma**-binga=ku (kan Siber/kana tu-nirengayan kan Siber)$_i$.
3.GEN$_i$=**MA**-annoy=1SG.PIVOT (SG.GEN Siber/CN.GEN 3.POSS-word LK Siber)$_i$
(intended: 'I was annoyed {by him/her/Siber/Siber's words}.') [Puyuma]

Within Formosan, the transitive use of *ma-* is attested only in four coastal languages: Amis (17), Kavalan (18), Basay-Trobiawan (19), and Siraya (20).[16] These languages constitute all members of Blust's (1999) East Formosan primary branch defined by the merger of PAn *j and *n.

(17) **Ma**-corah **ni** **Kulas** ku lumaq.
MA-burn GEN **Kulas** PIVOT house
'Kulas burned the house.' [Amis, East Formosan]

(18) **Ma**-bedung **ni** **Buya** peRasku 'nay.
MA-break GEN **Buya** bottle that
'Buya broke the bottle.' (Sung & Huang 2006:8) [Kavalan, East Formosan]

(19) **Ma**-unu=**isu** **ma**-tavan=na tama-isu?
MA-what=2S.GEN MA-head.hunt=ASP father-2S.POSS
'Why have you beheaded your own father?' (Li 2014:26) [Basay-Trobiawan, East Formosan]

(20) **Ma**-i-riney=**eta** hia tu-turo ki rata.
MA-LOC-make=1PL.GEN here RED-three PIVOT tabernacle
'Let's make here three tabernacles' (Adelaar 2011:88) [Siraya, East Formosan]

This observation allows for several generalizations. First, given ma_{TR}-'s limited distribution in Formosan, this function is likely to be innovative, rather than retentive. Second, given its presence in all members of East Formosan it is more economical to view this function as reconstructable to the common ancestor of the four languages. This suggests that ma_{TR}- is likely to be an innovation at the Proto-East-Formosan level, which, at the same time, lends further support to East Formosan as an independent linguistic subgroup (note that this group was previously defined only with a single merger (PAn *j/n) (Blust 1999)).[17]

Crucially, outside Formosan, a similar function of *ma-* is attested across multiple Malayo-Polynesian subgroups. According to our survey, the transitive use of *ma-* is observed in at least four Malayo-Polynesian primary branches under Smith's (2017) subgrouping: Philippines, Western Indonesian, Palauan, and Chamorro. Examples (21a-f) present a sample list of data that illustrate this distribution.[18] See also the table in (22) for a sample of our survey results.[19]

(21) a. **Ma**-yamak i batalan **ni** **napu**.
MA-break the plank GEN **wave**
'The plank was broken by the waves.' (Chung 2020:213) [Chamorro, Chamorro, MP]

b. A ngikel a **me**-ka **er** a ngalek.
DL fish DL MA-eaten GEN DL child
'The child is eating the fish.' (Gibson 1993:143) [Palauan, Palauan, MP]

[16] All Basay data cited in this paper comes from the Trobiawan dialect. Following previous work (Blust 1999; Li 2004, Sagart 2005), we adopt the term 'Basay-Trobiawan' to refer to this language throughout the paper.

[17] See Ross (2012) and Sagart (2014) for specific critiques of this proposal.

[18] While this innovative function is common in western Austronesian, it is not attested in many CEMP languages, where reflexes of *ma-* have lost the ability to license an optional genitive initiator. See, for example, Evans & Ross (2002) and work cited there.

[19] Note importantly that whether or not Philippine languages constitute a single (or multiple) Malayo-Polynesian primary branch remains controversial (Reid 1982; Blust 2017, 2020; Smith 2017; Liao 2020; Ross 2020; Zorc 2020). If they do represent multiple branches, the fact that ma_{TR}- is attested in Philippine languages under different lower-level subgroups will further strengthen the view that ma_{TR}- can be traced back to Proto-Malayo-Polynesian.

c. Saya [**ma**-alap **didi** **ama=mi=ti**] a laman.
that [MA-get PL.GEN **father=1PL.GEN.SPC**] LK boar
'That is the wild pigs which our fathers can get.' (Kimoto 2017:215) [Arta, PPh, MP]

d. Pai' to'o' dinó, niyá' **ma**-akan kadaingan.
NEG put.OV-IMP there later **MA.NPST**-eat children.(GEN)
'Don't put it there, the children might accidentally eat it. [Papar, WI, MP]

e. Kai' bulii' tiyo' ginó, indák am **ma**-akan **nu dalaing**.
NEG put.LV-IMP that there or.else **MA.NPST**-eat **3SG children**
'Don't put it there, it might (accidentally) get eaten by the child.'[Murut Nabaay, WI, MP]

f. Racun ino **na**-akan **nu anak titinó**.
poison that **MA.NPAST**-eat **GEN child that**
'That poison, it got accidentally eaten by that child.' [Bisaya, WI, MP]

g. Ino **na**-akán **nu bogók** dodinái.
that **MA.PST**-eat **GEN child** earlier
'That got (accidentally) eaten by the child earlier.' [Tatana, WI, MP]

h. Racun diri **na**-akan **di tanák**.
poison that **MA.PST**-eat **GEN child**
'That poison, it accidentally got eaten by the child.' [Sungai Karamuak, WI, MP]

i. Racun sití, **na**-akan **ni anak-ku**.
poison that **MA.PST**-eat **GEN child-1SG.GEN**
'That poison, it got (accidentally) eaten by my son/daughter.' [Sinabû, WI, MP]

(22) *Two functions of ma- in selected western Austronesian languages*

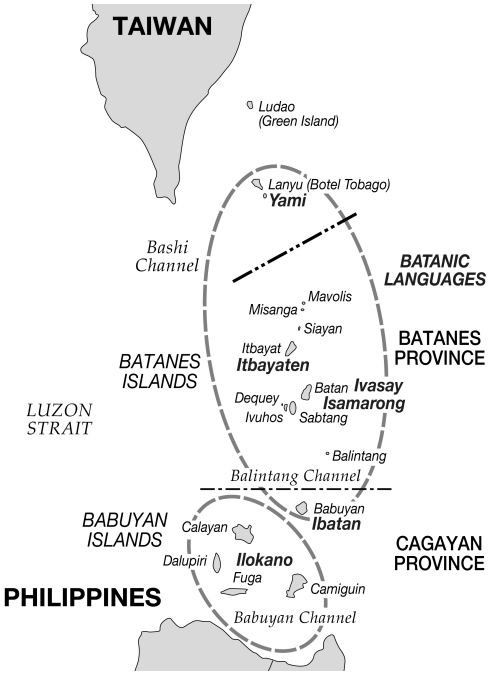
	subgrouping affiliation	intransitive *ma-*	transitive *ma-*	source(s)
Puyuma	Puyuma	YES	NO	Teng 2008, Cauquelin 2015, primary data
Paiwan	Paiwan	YES	NO	Chang 2000, 2006
Saisiyat	Northwestern Formosan	YES	NO	Zeitoun et al. 2015, Zeitoun 2018 &
Atayalic	Atayalic	YES	NO	Huang 2000, 2018
Pazeh	Western Plains	YES	NO	Lin 2000, 2018
Rukai	Rukai	YES	NO	Zeitoun 2000, 2007, 2018
Bunun	Bunun	YES	NO	De Busser 2008, Li 2018
Tsou	Tsouic	YES	NO	Zeitoun 2000, Chang & Pan 2018
Saaroa	Tsouic	YES	NO	Pan 2012
Kanakanavu	Tsouic	YES	NO	Song 2018, Wild 2018
Amis	East Formosan	YES	YES	Huang & Song 2006, Wu 2006, primary data
Kavalan	East Formosan	YES	YES	Li & Tsuchida 2006
Siraya	East Formosan	YES	YES	Adlaar 2011
Basay-Trobiawan	East Formosan	YES	YES	Liu 2007, Li 2014
Yami	Batanic(/Philippines), Malayo-Polynesian (MP)	YES	YES	Rau & Chu 2006
Ibatan	Batanic(/Philippines), MP	YES	YES	Yamada 2014
Ivatan	Batanic(/Philippines), MP	YES	YES	Reid 1966
Ilocano	Philippines, MP	YES	YES	Rubino 1997
Arta	Philippines, MP	YES	YES	Kimoto 2017
Tagalog	Philippines, MP	YES	YES	Himmelmann 2004, 2006; primary data
Cebuano	Philippines, MP	YES	YES	Tanangkingsing 2008
Proto-Paitanic	Greater Dusunic, West Indonesian, MP	YES	YES	Lobel 2013, pers. com.
Sinabu	Paitanic, Greater Dusunic, West Indonesian, MP	YES	YES	Lobel pers. com.
Bisaya Sabah	Bisaya-Lotud, Greater Dusunic, West Indonesian, MP	YES	YES	Lobel pers. com.
Sungai Karamuak	Dusunic, Greater Dusunic, West Indonesian, MP	YES	YES	Lobel pers. com.
Papar	Papar, Greater Murutic, West Indonesian, MP	YES	YES	Lobel pers. com.
Murut	Murutic, Greater Murutic, West Indonesian, MP	YES	YES	Lobel pers. com.
Tatana	Tatana, Greater Murutic, West Indonesian, MP	YES	YES	Lobel pers. com.
Tidung Sumbol-Dungusan	Tidung, Greater Murutic, West Indonesian, MP	YES	YES	Lobel pers. com.
Old Balinese	West Indonesian, MP	YES	YES	Beratha 1992
Chamorro	Chamorro, MP	YES	YES	Chung 2000, Gibson 1980
Palauan	Palauan, MP	YES	YES	Gibson 1993

Note in particular that the transitive use of *ma-* is attested across a group of closely related languages scattered between Taiwan and Luzon known as the Batanic languages: Yami, Ivatan, Ibatan, and Itbayaten (23), as seen in (24).[20] The fact that ma_{TR}- is attested in these languages (23a-d) has

[20] According to primary fieldwork, *ma-* can but not always denote non-volitional or abilitative reading in Batanic languages. The actual interpretation of a sentence is still context-based. This is why (24a-d) do not bear non-volitional reading.

important implications for interpreting the chronology of ma_{TR}, as Proto-Batanic was possibly the first offshoot of Proto-Malayo-Polynesian (Ross 2005). We will revisit this in section 4.[21]

(23) *Figure 1: Distribution of the Batanic languages*



(24) a. **Na**-boyaw **ni** **Adod** ∅ bago saya.
MA-chase GEN **Adod** PIVOT boar those
'Those pigs were chased away by Adod.' (Maree 2007:223) [Ibatan, Batanic]

b. **Ma**-voyaw=**mo** qo manok.
MA-chase=**2SG.GEN** PIVOT chicken
'The chicken is being chased by you.' (Reid 1966:125) [Ivatan, Batanic]

c. **Na**-tta=**ko** si Orsing dawi.
MA-see=**1SG.GEN** PIVOT Orsing there
'I saw Orsing there.' (Yamada 2014:72) [Itbayaten, Batanic]

d. **Ma**-kala=**ta** o mogis nio?
MA-find=**1PL.INCL.GEN** PIVOT rice 2PL.GEN
'Can we find your rice?' (Rau & Dong 2006:115) [Yami, Batanic]

4 Proposal: ma_{TR}- as a single, shared innovation of EF and MP

The distribution of these two functions of *ma-* raises an important question: is the transitive use of *ma-* (henceforth ma_{TR}-) reconstructable to Proto-Austronesian, as is the stative intransitive function?; if not, does it reflect a single innovation or multiple drifts? As seen earlier in (17)-(20), this function is attested in four coastal Formosan languages that belong to the same subgroup and four Malayo-Polynesian primary branches. Possible interpretations of this distribution are summarized in (25). Here, we use the term 'drift' in its accepted sense, i.e. parallel and independent innovations (Sapir 1921; Jespersen 1922; Andersen 1990; McMahon 1994; Croft 2000; Baxter et al. 2006; Steels & Szathmary 2018; inter alia.).[22]

[21] We acknowledge that this view remains unsettled; see recent debates in Ross (2005, 2020); Blust (2019, 2020); Liao 2020; Reid (1982, 2020); Zorc (2020).

[22] Lavidas (2007) and Kulikov (2003) discuss two similar grammaticalization processes in Greek and Vedic, in which an originally intransitive/anticausative construction was reanalyzed as transitive-causative. This suggests that the reanalysis under concern here is not a unique case.

(25) a. **Scenario I (retention)**: Both functions of *ma-* were retentions from PAn; the stative function was more stable, and the transitive function has been lost in most Austronesian primary branches.

b. **Scenario II (single innovation)**: The transitive use of *ma-* reflected a single innovation that took place in the common ancestor of East Formosan and Malayo-Polynesian.

c. **Scenario III (two independent innovations, i.e. drift)**: The transitive use of *ma-* reflected two independent innovations that took place in Proto-East-Formosan and Proto-Malayo-Polynesian.

d. **Scenario IV (multiple independent innovations, i.e. drift)**: The transitive use of *ma-* reflected multiple innovations occurring in various East-Formosan and Malayo-Polynesian subgroups.

In what follows, we present specific arguments for Scenario II.

4.1 Against Scenario I ($ma_{\text{TR-}}$ as a PAn retention)

Interpreting ma_{TR}- as a retention from Proto-Austronesian is disfavored given its absence in any other higher-order Austronesian subgroup except East Formosan and Malayo-Polynesian. Interpreting this function as a Proto-Austronesian retention thus forces an undesirable assumption: ma_{TR} has been lost independently in the majority of Austronesian primary branches, whereas the stative intransitive function of the same morpheme remains stable across all these primary branches.

This scenario entails one other disfavored assumption: the transitive function of *ma-* is unstable and prone to loss, but being highly resistant to change in East Formosan and Malayo-Polynesian subgroups. Finally, the fact that *m*-initial morphemes in western Austronesian canonically denote an Actor Voice case frame suggests that ma_{TR}– which denotes a Patient Voice case frame – is more likely to be innovative.[23]

Accordingly, Scenario I is disfavored.

4.2 Against Scenarios III and IV (ma_{TR}- as the outcome of drift)

Two other potential accounts for the distribution of ma_{TR}- are to analyze it as (i) the outcome of two independent innovations occurring in Proto-East-Formosan and Proto-Malayo-Polynesian, or as (ii) the result of multiple innovations occurring at lower-level subgroups of East Formosan and Malayo-Polynesian.

Neither proposal is ideal. As shown in section 3, the transitive use of *ma-* is attested in all members of East Formosan, indicating that this innovation is best reconstructed to Proto-East Formosan. Similarly, the presence of ma_{TR}- in four of the nine Malayo-Polynesian primary branches — in particular in two isolated branches Chamorro and Palauan and all Batanic languages (possibly the immediate descendent of Proto-Malayo-Polynesian) — strongly suggests that this function can be traced back to Proto-Malayo-Polyesian. This argues against interpreting ma_{TR}- as the outcome of multiple independent innovations in EF and MP subgroups (Scenario IV).

Analyzing ma_{TR}- as two independent developments in PEF and PMP (Scenario III) is also disfavored for similar reasons, as this proposal forces the undesirable assumption that PAn *ma- underwent two highly similar innovations in two of its immediate descendants (and before further split of both branches) but not within any of the other eight branches. Moreover, given the presence of PAn *C/t merger in both East Formosan and Malayo-Polynesian we must assume assume PEF and PMP underwent two identical innovations independently: innovation of ma_{INTR}- into a transitive affix and the merger of PAn *C and *t into *t*.

Not only do both scenario go against the Principle of Economy, but they also leave a few independent pieces of evidence for an EF-MP connection unexplained – including a bunch of potential lexical innovations shared between EF and MP. See section 6 for details.

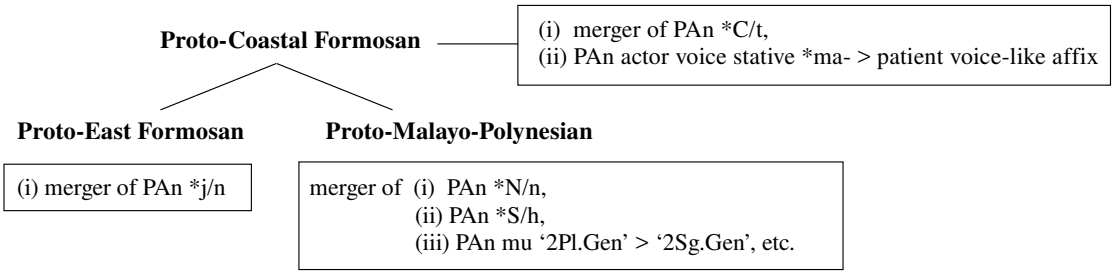
[23] See Ross (2015) for a comprehensive discussion of *m*-initial morphemes and their function in Proto-Austronesian.

5 East Formosan as the closest relative of Malayo-Polynesian

We argue accordingly that the distribution of ma_{TR}- is best analyzed as an innovation occurring prior to the split of the common ancestor of Proto-East Formosan and Proto-Malayo-Polynesian, and inherited by all its modern descendants.

This proposal points to the subgrouping scenario illustrated in (26), whereby Proto-Malayo-Polynesian descended from a speech community, here labeled *Coastal Formosan*, from which modern East Formosan languages are also derived. Malayo-Polynesian moved out of Taiwan and developed its own traits, after which Proto-East-Formosan underwent the merger of PAn *j and *n and split into the four modern languages.[24] The fact that Siraya, Basay-Trobiawan, Kavalan, and Amis are distributed in different coastal regions of Taiwan suggests that this proposed common ancestor of the East Formosan and the Malayo-Polynesian people may have been a seafaring community that was spread along the coastlines of the island. The current proposal that one group of its descendants expanded further to the Batanes and Luzon Islands follows from this inference.

(26) *Working hypothesis: EF and MP as sisters under a single Austronesian primary branch*



The current proposal yields two important broader implications. First, the East Formosan languages are the closest relatives of Malayo-Polynesian in Taiwan. Second, all Austronesian primary branches are represented on Taiwan, with Malayo-Polynesian being a subbranch of one of the nine Formosan branches.[25] This proposal is consistent with the fact that East Formosan constitutes the only Austronesian language group in Taiwan distributed around the coastline of the island, as seen in Figure 2.[26]

[24] This proposal partially coincides with an earlier insight that Amis falls under an Austronesian primary branch that also subsumes extra-Formosan langauges (Harvey 1982) and Reid's (2016:132) view that PMP was a language spoken by multiple small groups of migrant seafarers from southeastern Taiwan.

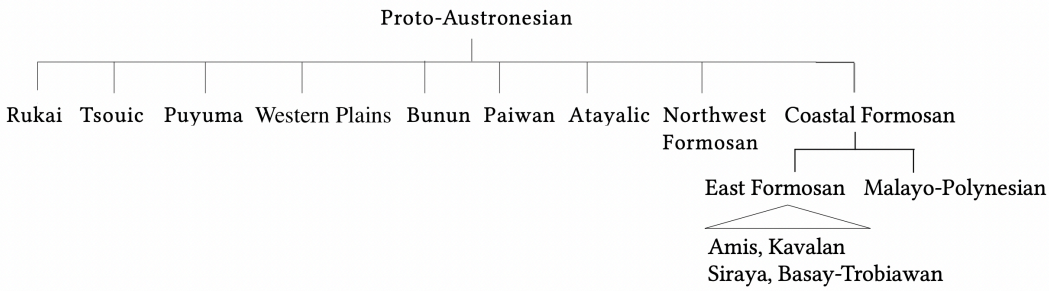
[25] This innovation shared by EF and MP also provides evidence against including Bunun in the EF-MP language group (e.g. Ho & Yang 2000), as the transitive use of *ma-* is not attested in Bunun. We remain agnostic about an alternative scenario whereby Bunun forms an Austronesian primary branch with the EF-MP language group defined by the merger of PAn *N and *n – which is reflected in all three groups of languages.

[26] Several authors have noted that East Formosan speaking communities share certain cultural similarities, such as the matrilineal kinship system shared by the Amis, Kavalan and Siraya (Li 2004:372, note 17; Shepherd 1993:458, note 88, Ferrell 1969:56). This may represent tentative evidence for a common ethnic origin for speakers of these language groups.

(27) *Figure 2: Distribution of Formosan subgroups*

As seen in (27), the fact that the four East Formosan languages are scattered around the coasts of Taiwan suggests that EF speakers may have expanded along the coastline of the island by sea. This distribution and the known seafaring traditions of the East Formosan groups (Mabuchi 1960; Ferrell 1969; Li 2004, 2012) offers a reasonable origin story for Malayo-Polynesian as a speech community that was enabled to split off and spread to new environments by the seafaring technology already possessed by the parent community. This proposal points to a nine-branch tree for Austronesian higher-order subgrouping (28).

(28) *Revised Austronesian higher-order subgrouping*



6 Further evidence for EF-MP connection

Before concluding, we present additional support for the current proposal. Section 6.1 discusses potential shared lexical innovations between East Formosan and Malayo-Polynesian; section 6.2 sums up recent findings in genetics in line with the proposed EFMP connections.

6.1 Potential lexical innovations defining Proto-Coastal-Formosan

If East Formosan and Malayo-Polynesian indeed share a common origin, we expect to see lexical innovations exclusively shared between the two groups. One piece of evidence comes from the cognacy between Amis's existential negator *awa* with that in four Malayo-Polynesian languages spoken in the Philippines. As seen in (29), neither East Formosan nor Malayo-Polynesian languages possess a reflex of the Proto-Austronesian existential negator **uka*, whose reflexes are attested across Formosan language (29) (Blust & Trussel ongoing; Lin 2011). Moreover, an innovative form is attested in the East Formosan language Amis and a few Malayo-Polynesian languages of the Philippines (30), demonstrating a potential case of *replacement innovation* occurring in the shared ancestor of EF and MP.

This cognacy is difficult to explain if Amis and Philippines languages belong to two distinct primary branches, while expected from the proposed EF-MP connection.

(29) *Reflexes of Proto-Austronesian existential negator *uka (Austronesian Comparative Dictionary (ACD), Lin 2011)*

Saisiyat	Northwestern Formosan	*'oka*
Seediq	Atayalic	*uka*
Thao	Western Plains	*uka*
Bunun	Bunun	*uka*
Tsou	Tsouic	*uk'a*
Saaroa	Tsouic	*'uka'a*
Rukai	Rukai	*okaodho* (near comparison)
Atayal	Atayalic	*ungat* (near comparison)

(30) *Replacement innovation of PAn *uka 'existential negator' in EF and MP languages (ACD)*[27]

Amis	East Formosan	*awa*
Ivatan	Batanic/Philippines, Malayo-Polynesian	*ava*
Ibatan	Batanic/Philippines, Malayo-Polynesian	*aba*
Ilokano	Philippines, Malayo-Polynesian	*awan*
Isneg	Philippines, Malayo-Polynesian	*awan*
Casiguran Dumagat	Philippines, Malayo-Polynesian	*ewan*

There also exist a number of lexical items shared between East Formosan and Malayo-Polynesian languages while unattested in any other Austronesian primary branches. Consider the sample list in (31)-(35) based on the Austronesian Comparative Dictionary (ACD) and our survey of Formosan literature. As seen below, all these lexical items are attested in both Amis and Kavalan (two better-described East Formosan languages), and across various MP primary branches (hence reconstructable to Proto-Malayo-Polynesian), but not in any other Formosan language. Note in particular that the regular sound correspondences shown between Amis and Kavalan suggests these lexical items existed prior to the split of Proto-East-Formosan. [28]

(31) *Cognates of *bubu 'conical bamboo basket trap for fish'*[29]

Amis	East Formosan	*fofo*	a fish trap for catching fish and crabs in rivers
Kavalan	East Formosan	*bubu*	conical bamboo basket trap for fish
Ilokano	Philippines, Malayo-Polynesian	*bobo*	kind of large bow net used to catch fresh-water shrimps
Tagalog	Philippines, Malayo-Polynesian	*bubo*	fish trap
Aklanon	Philippines, Malayo-Polynesian	*bobo*	fish trap
Kelabit	Western Indonesian, Malayo-Polynesian	*bubuh*	bamboo basket trap for fish
Bintulu	Western Indonesian, Malayo-Polynesian	*buvew*	conical bamboo fish trap
Malagasy	Western Indonesian, Malayo-Polynesian	*vovo*	kind of basket used for fishing
Moken	Moken, Malayo-Polynesian	*bubey*	fish trap
Palauan	Palauan, Malayo-Polynesian	*bub*	trap (usually for fish)
Hawu	CMP, CEMP, Malayo-Polynesian	*wuwu*	fish trap
Rotinese	EMP, CEMP, Malayo-Polynesian	*bufu*	fish trap
Buli	SHWNG, CEMP, Malayo-Polynesian	*pup*	fish trap
Kowiai	CMP, CEMP, Malayo-Polynesian	*fuf*	kind of fish trap
Proto-Oceanic	CEMP, Malayo-Polynesian	**pupu*	conical bamboo basket trap for fish

[27] This set of correspondences comes from the ACD, where they were listed as near comparisons (referring to 'comparisons in which the observed similarity appears too great to attribute to chance, but because of imprecise agreement the reconstruction of a well-defined form is not yet possible.') Note, importantly, that reflexes of PAn *uka are not attested in any East Formosan and Malayo-Polynesian languages, to the best of our knowledge. According to available descriptions, Kavalan (Li & Tsuchida 2006), Siraya (Adelaar 2012), and Basay-Trobiawan (Li 2014) all employ an innovative existential negator not etymologically linked to PAn *uka.

[28] Under the consensus subgrouping, Amis is a single-member primary branch of East Formosan; Kavalan belongs to a separate branch, as seen earlier in (5) (Blust 1999).

[29] Note importantly that the same form is attested in multiple Formosan languages and reconstructable to Proto-Austronesian with a distinct meaning 'grandparent/grandchild (reciprocal term of address)': Favorlang/Babuza (Western Plains) *bubu*, Paiwan (Paiwan) *vuvu*, Seediq (Atayalic) *bubu* 'mother'. The case of **bubu* 'fish trap' thus constitutes another potential case of lexical innovation between EF and MP.

(32) *Cognates of *Rabiqi 'late afternoon, evening, evening meal'*

Amis	East Formosan	*lafi*	'evening'
Kavalan	East Formosan	*Rabi*	'evening, dinner; the evening meal'
Isneg	Philippines, Malayo-Polynesian	*xabi*	'evening, night'
Ifugaw	Philippines, Malayo-Polynesian	*labi*	'night'
Cebuano	Philippines, Malayo-Polynesian	*gabi'i*	'night'
Supan	Western Indonesian, Malayo-Polynesian	*gabpi-n*	'night'
Mongondow	Celebic, Malayo-Polynesian	*gobii*	'night'
Bimanese	CMP, CEMP, Malayo-Polynesian	*awi(na)*	'yesterday'
Wandamen	SHWNG, CEMP, Malayo-Polybnesian	*ravi-nena*	'afternoon'
Ron	SHWNG, CEMP, Malayo-Polynesian	*rob*	'night'
Numfor	SHWNG, CEMP, Malayo-Polynesian	*rob*	'night'
Proto-Oceanic	CEMP, Malayo-Polynesian	**Rapi*	'afternoon, evening, yesterday'

(33) *Cognates of *buSaw 'cold, of lefrover food; leftover from a meal'*

Amis	East Formosan	*fasaw*	'cooled off'
Kavalan	East Formosan	*basaw*	'fever went down, abated; to become cold'
Itbayaten	Batanic/Philippines, Malayo-Polynesian	*vahaw*	'idea of being cold (of food esp.)'
Ilokano	Philippines, Malayo-Polynesian	*baaw*	'left over, cold rice; tepidity, coolness'
Pangasinan	Philippines, Malayo-Polynesian	*baaw*	'cooked rice'
Tagalog	Philippines, Malayo-Polynesian	*bahaw*	'left-over food, especially boiled or steamed rice'
Bikol	Philippines, Malayo-Polynesian	*bahaw*	'cold, referring only to food once served hot'
Aklanon	Philippines, Malayo-Polynesian	*bahaw*	'cold rice; cool off, get cool (said of food)'
Mansaka	Philippines, Malayo-Polynesian	*baaw*	'food prepared for a trip'
Bintulu	Western Indonesian, Malayo-Polynesian	*pa-vaw*	'cold, of hot food that has gotten cold'
Agutaynen	CEMP, Malayo-Polynesian	*baw*	'breakfast, morning snack'

(34) *Cognates of *laRiw 'run, run away, flee, escape'*

Amis	East Formosan	*laliw*	'escape'
Kavalan	East Formosan	*m-RaRiw*	'run, run away'
Itbayaten	Batanic/Philippines, Malayo-Polynesian	*yayo-h*	'race'
Hanunóo	Philippines, Malayo-Polynesian	*lagiw*	'running'
Abaknon	Philippines, Malayo-Polynesian	*lahi*	'to run, run away'
Cebuano	Philippines, Malayo-Polynesian	*lagiw*	'run away'
Maranao	Philippines, Malayo-Polynesian	*lagoy*	'rush, hurry'
Samal	Philippines, Malayo-Polynesian	lahi-lahi	'to run'
Iban	Western Indonesian, Malayo-Polynesian	*lari*	'run away'
Malay	Western Indonesian, Malayo-Polynesian	*lari*	'escape'
Kambera	CMP, CEMP, Malayo-Polynesian	*lao*	'to run'
Hawu	CMP, CEMP, Malayo-Polynesian	*rai*	'run, run away'
Soboyo	CMP, CEMP, Malayo-Polynesian	*lahi*	'to run, run away'
Ron	SHWNG, CEMP, Malayo-Polynesian	*farar*	'to run'

(35) *Cognates of *bangeSiS 'fragrant/fragrance'*

Amis	East Formosan	*fangsis*	'sweet'
Kavalan	East Formosan	*bangsis*	'fragrant'
Ilokano	Philippines, Malayo-Polynesian	*bang'i*	'to smell of toast'
Cham	Western Indonesian, Malayo-Polynesian	*bangi*	'used of all agreeable sensations: good, tasty, redolant, etc'
Simalur	Western Indonesian, Malayo-Polynesian	*fangi*	'odor, fragrance'
Old Balinese	Western Indonesian, Malayo-Polynesian	*wangi*	'fragrance'
Ngadha	CMP, CEMP, Malayo-Polynesian	*vangi*	'to smell'

6.2 Support from sister fields

Not only does the current subgrouping point to a scenario in line with that expected from a socio-historical perspective where the out-of-Taiwan population belongs to a certain Formosan community, but it also aligns with a major finding that point to other similarities between Malayo-Polynesian and East Formosan. Capelli et al. (2005) report that Y chromosomes from the Amis but not other non-East Formosan aboriginal communities of Taiwan are distributed throughout selected Austronesian communities outside Taiwan. Chen et al. (2011:44) revisit Capelli et al.'s data, noting that the genetic

distance between the Amis and Filipinos was shorter than that between the Amis and other Formosan tribes, quoted in (36).[30]

(36) *Genetic distance between Formosan and Philippines populations*

population compared	genetic distance
Amis - Philippines	0.025
other Formosan aboriginals - Amis	0.068
other Formosan aboriginals - Philippines	0.073

Trejaut et al. (2005) reports two consistent findings: (i) among all Formosan populations tested, the Amis in particular are more closely related to island Southeast Asian populations than to populations from mainland East Asia, and (ii) Y-chromosome haplogroup B4a1a occurs frequently among the Amis and the Yami (Batanic, Malayo-Polynesian) but not in any other Formosan communities (Paiwan, Puyuma, Rukai) examined in the study. Contra the conventional view in Linguistics that Malayo-Polynesian does not have a closer relationship with any Formosan group, this suggested link between East Formosan and Malayo-Polynesian finds a *home* for the out-of-Taiwan population in Taiwan, offering a scenario that is potentially more compatible with the perspective of human expansion.

7 Conclusion

A remaining question in Austronesian higher-order subgrouping concerns the linguistic position of Malayo-Polynesian, the linguistic subgroup comprising all Austronesian languages spoken outside the homeland, Taiwan. While previous subgrouping proposals recognize it as an independent primary branch due to a lack of evidence tracing its precise origin, this treatment is inconsistent with the consensus in archaeology that there was a (relatively) long pause in Taiwan before the settlement of Luzon – which would in principle yield a subgrouping scenario whereby the out-of-Taiwan population represents a linguistic subgroup embedded under a certain Austronesian primary branch. In this paper we put forward a subgrouping proposal in line with this view drawing on previously overlooked evidence for potential connections between Malayo-Polynesian and East Formosan, a linguistic subgroup comprising four languages distributed around the coastline of Taiwan. Support for this claim comes from an understudied morphosyntactic innovation shared exclusively between East Formosan and Malayo-Polynesian, PAn stative intransitive *ma- to a transitive affix featuring a Patient Voice-like case frame. Given one other innovation reflected in both groups (merger of PAn *C and *t), we argue that the two groups share a common origin. Not only does this proposal point to a migration scenario more consistent with a socio-historical picture whereby the ancestor of Malayo-Polynesian moved out Taiwan once speech community had developed a seafaring tradition, it also aligns with a major finding of recent genetic research that East Formosan speakers are genetically closer to Austronesian communities located outside Taiwan.

References

Adelaar, Alexander. 2011. *Siraya: Retrieving the phonology, grammar, and lexicon of a dormant Formosan language*. Berlin & Boston: De Gruyter Mouton.

Aldridge, Edith. 2016. Ergativity from subjunctive in Austronesian languages. *Language and Linguistics* 17(1):27–62.

Aldridge, Edith. To appear. Syntactic conditions on accusative to ergative alignment change in Austronesian languages. *Journal of Historical Linguistics*.

Andersen, Henning. 1990. The structure of drift. In Henning Andersen & Konrad Koerner (eds.), *Historical Linguistics 1987*. Papers from the 8th International Conference on Historical Linguistics, 1–20. Amsterdam: Benjamins.

[30] Capilli et al. (2005) did not specify the exact Filipino community examined in their study, and the Amis was the only East Formosan population examined in this study.

Baxter, Gareth, Richard Blythe, William Croft & Alan McKane. 2006. Utterance selection model of language change. *Physical Review E* 73(4):46–118.

Bellwood, Peter. 1979. *Man's conquest of the Pacific*. New York: Oxford University Press.

Bellwood, Peter. 1980. *The peopling of the Pacific.* Scientific American 243(5):174–85.

Bellwood, Peter. 1983. New perspectives on Indo-Malaysian prehistory. *Bulletin of Indo-Pacific Prehistory Association* 4:71-83.

Bellwood, Peter. 1984–85. A hypothesis for Austronesian origins. *Asian Perspectives* 26:107–17.

Bellwood, Peter. 1988. Archaeological research in south-eastern Sabah. Kota Kinabalu: Sabah Museum and State Archives.

Bellwood, Peter. 1989. Archaeological investigations at Bukit Tengkorak and Segarong, south-eastern Sabah. *Bulletin of the Indo-Pacific Prehistory Association* 9, 122–62.

Bellwood, Peter. 1995. Austronesian prehistory in Southeast Asia: Homeland, expansion and transformation. In Peter Bellwood et al. (eds.), *The Austronesians*, 96–106. Canberra: The Department of Anthropology, Research School of Pacific and Asian Studies, The Australian National University.

Bellwood, Peter. 2005a. *First Farmers: The origins of Agricultural Societies.* Blackwell, Oxford.

Bellwood, Peter. 2005b. Coastal south China, Taiwan, and the prehistory of the Austronesians. In Chung-Yu Chen & Jian-Guo, (eds.), *The archaeology of south-east coastal islands of China Conference*, 1–22. Taipei: Executive Yuan, Council for Cultural Affairs

Bellwood, Peter. 2007. *Prehistory of the Indo-Malaysian Archipelago*, revised ed. Canberra: Australian National University E Press.

Bellwood, Peter. 2017. *First islanders: prehistory and human migration in Island Southeast Asia*. John Wiley & Sons.

Beratha, Ni Luh Sutjiati. 1992. Evolution of verbal morphology in Balinese. PhD dissertation, Australian National University.

Blust, Robert. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *Working Papers in Linguistics* 9.2:1–15. Honolulu: Department of Linguistics, University of Hawai'i.

Blust, Robert. 1984. The Austronesian Homeland: A linguistic Perspective. *Asian Perspectives* 26:45–67.

Blust, Robert. 1991. The Greater Central Philippines hypothesis. *Oceanic Linguistics* 30:73–129.

Blust, Robert. 1999. Subgrouping, circularity, and extinction: some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (eds.), *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Taipei: Institute of Linguistics, Academia Sinica.

Blust, Robert. 2001. Malayo-Polynesian: New stones in the wall. *Oceanic Linguistics* 40:151–55.

Blust, Robert. 2013. *The Austronesian languages* (2nd edition). Canberra: Pacific Linguistics.

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

Blust, Robert. 2020. Response to Comments on "The Resurrection of Proto-Philippines". *Oceanic Linguistics 59*:450–79.

Blust, Robert and Victoria Chen. 2017. The pitfalls of negative evidence: 'Ergative Austronesian', 'Nuclear Austronesian' and their progeny. *Language and Linguistics* 18(4):577–621.

Blust, Robert, and Stephen Trussel. Ongoing. Austronesian comparative dictionary. Online: http://www.trussel2.com/ Austronesian Comparative Dictionary.

Capelli, C., J. F. Wilson, M. Richards, M. P. Stumpf, F. Gratrix, S. Oppenheimer, P. Underhill, V. L. Pascali, T. M. Ko, and D. B. Goldstein. 2001. A Predominantly Indigenous Paternal Heritage for the Austronesian-speaking Peoples of Insular Southeast Asia and Oceania. *Am J Hum Genet* 68:432–43.

Cauquelin, Josiane. 2015. *Nanwang Puyuma-English dictionary*. Taipei: Academia Sinica.

Chang, Anna, Hsiou-chuan. 2000. *A reference grammar of Paiwan* (in Chinese). Taipei: Yuanliu.

Chang, Anna, Hsiou-chuan. 2006. A reference grammar of Paiwan. PhD dissertation, Australian National University.

Chen, Yaofeng, Shuzhuo Chen, and Muzhu Xu. 2011. Homeland or transfer station? Genetic studies' contribution to our understanding of Austronesian diaspora (in Chinese). *Humanities and social sciences newsletter quarterly* 12(3):41–49.

Chen, Victoria. 2017. A re-examination of the Philippine-type voice system and its implications for Austronesian primary-level subgrouping. PhD dissertation, University of Hawai'i.

Chang, Henry Yung-li, and Chia-jung Pan. 2018. *A sketch grammar of Tsou* (in Chinese). *Series on Formosan Languages* 7. New Taipei: Council of Indigenous Peoples.

Chung, Sandra. 2000. *Chamorro grammar*. California Digital Library University of California.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Longman.

Croft, William. 2006. The relevance of an evolutionary model to historical linguistics. In O. Nedergaard Thomsen (ed.), *Competing Models of Linguistic Change: Evolution and beyond*, 91–132. Amsterdam: John Benjamins.

Dahl, 1973. Proto-Austronesian. *Scandinavian Institute of Asian Studies Monograph Series*, No. 15. London: Curzon Press.

De Busser, Rik. 2009. Towards a grammar of Takivatan Bunun: Selected topics. PhD dissertation, La Trobe University.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Amsterdam: Mouton de Gruyter.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: problems in cultural and linguistic classification.* Institute of Ethnology, Academia Sinica, Monograph no. 17. Taipei: Academia Sinica.

Gibson, Jeanne. 1980. Clause union in Chamorro and in Universal Grammar. PhD dissertation, University of California dissertation, San Diego, CA.

Gibson, Robert. 1993. Palauan causatives and passives: An incorporation analysis. PhD dissertation, University of Hawai'i.

Himmelmann, Nikolaus. 2004. On statives and potentives in Western Austronesian (mostly Tagalog). In Paul Law (ed.), *Proceedings of AFLA 11. ZAS Papers in Linguistics* 34, 103–19. Berlin: Zentrum fuer Allgemeine Sprachwissenschaft, Typologie und Universalienforschung (ZAS).

Himmelmann, Nikolaus. 2006. How to miss a paradigm or two: Multifunctional *ma-* in Tagalog. In Felix K. Ameka, Alan Dench, & Nicholas Evans (eds.), Catching language: The standing challenge of grammar writing, 487–526. Berlin: Mouton de Gruyter.

Ho, Dah-an. 1998. Genetic relationships among the Formosan languages [in Chinese]. *Chinese Studies* 16(2):141–71.

Ho, Dah-an, and Hsiu-fang Yang. 2000. *Prologue: Austronesian and Formosan languages*. In Lilian Huang (ed.), *Formosan languages grammar series* (in Chinese), 1–25. Taipei: Yuanliu. .

Huang, Shu-ping, and Li-may Sung. 2008. The undergoer focus *ma-* in Kavalan. *Oceanic Linguistics* 47: 159–84.

Huang, Lilian. 2000. *A reference grammar of Atayal* (in Chinese). Taipei: Yuanliu.

Huang, Lilian, and Hsin-sheng Wu. 2018. Atayal. *A sketch grammar of Tsou*. (In Chinese). Series on Formosan Languages 2. New Taipei: Council of Indigenous Peoples.

Kimoto, Yukinori. 2017. A grammar of Arta: A Philippine Negrito Language. PhD dissertation, Kyoto University.

Kirch, Patrick. 2002. *On the road of the winds: An archaeological history of the Pacific islands*. Berkeley: University of California Press.

Kroeger, Paul. 1991. Phrase structure and grammar relations in Tagalog. PhD dissertation, Stanford University.

Kulikov, Leonid. 2003. The labile syntactic type in a diachronic perspective: The case of Vedic. SKY Journal of Linguistics 16: 93-112.

Lavidas, Nikolaos. 2007. Transitivity alternations in diachrony: Changes in argument structure and voice morphology. Newcastle: Cambridge Scholar Publishing.

Li, Lilian Li-ying. 2018. A grammar of Isbukun Bunun. PhD dissertation, National Tsing Hua University.

Li, Paul Jen-kuei. 2004. Origins of the East Formosans: Basay, Kavalan, Amis, and Siraya. *Language and linguistics* 5. 363-376.

Li, Paul Jen-kuei. 2010. *Origins and dispersal of the Formosan natives in the east coast*. Taiwan Journal of Indigenous Studies 3(4):19.

Li, Paul Jen-kuei. 2014. *Texts of the Trobiawan dialect of Basay. Asian and African Lexicon Series* 56. Research Institute for Languages and Cultures of Asia and Pacific, Tokyo University of Foreign Studies.

Li, Paul Jen-kuei, and Shigeru Tsuchida. 2006. *Kavalan dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Liao, Hsiu-chuan. 2011. Some morphosyntactic differences between Formosan and Philippine languages. *Language and Linguistics* 12(4):845–76.

Liao, Hsiu-chuan. 2020. A reply to Blust (2019) "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:426–49.

Lin, Shu-yi. 2011. Reconstructing negative morphemes in Proto-Austronesian: Evidence from Formosan languages. MA thesis, National Taiwan Normal University.

Lin, Ying-chin. 2000. *A reference grammar of Pazeh* (in Chinese). Taipei: Yuanliu.

Liu, Dorinda Tsai-Hsiu. 2007. Basay nominal constructions. *Working papers in Linguistics* 38(5):1–29. Department of Linguistics, University of Hawai'i.

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

Jespersen, Otto. 1922. *Language, its nature, development, and origin*. H. Holt.

Mabuchi, Toichi. 1960. The aboriginal peoples of Formosa. In G. Murdock (ed.), *Social Structure in Southeast Asia*, 127–40. Chicago: Quadrangle.

McMahon, April. 1994. *Understanding language change*. Cambridge University Press.

Mills, Roger. 1975. Proto South Sulawesi and Proto Austronesian phonology. PhD dissertation, University of Michigan at Ann Arbor.

Pan, Jia-jung. 2012. A grammar of Lha'alua, an Austronesian language of Taiwan. PhD dissertation, James Cook University.

Rau Victoria and Maa-neu Dong. 2006. *Yami Texts with reference grammar and dictionary*.

Reid, Lawrence. 1966. An Ivatan Syntax. *Oceanic Linguistics Special Publication*. Honolulu: Pacific and Asian Linguistics Institute, University of Hawai'i.

Reid, Lawrence. 2016. Accounting for variability in Malayo-Polynesian pronouns. *Journal of Historical Linguistics* 6(2):130–64.

Ross, Malcolm, 1995. Some current issues in Austronesian linguistics. In D. Tryon (ed.), *Comparative Austronesian Dictionary* 1, Fascicle 1, 45–120. Berlin: Mouton de Gruyter.

Ross, Malcolm. 2002. The history and transitivity of Western Austronesian voice and voice marking. In Faye Wouk & Malcolm Ross (eds.), *The history and typology of Western Austronesian voice systems*, 63–78. Canberra: Pacific Linguistics.

Ross, Malcolm. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2): 1–23.

Ross, Malcolm. 2012. In Defense of Nuclear Austronesian (and Against Tsouic). *Language and Linguistics* 13(6):1253–330.

Ross, Malcolm. 2015. Reconstructing Proto Austronesian verb classes. *Language and Linguistics* 16(3):279–345.

Ross, Malcolm. 2020. Comment on Blust "The Resurrection of Proto-Philppines." *Oceanic Linguistics* 59: 366–73.

Rubino, Carl Ralph Galvez. 1997. A reference grammar of Ilocano. PhD dissertation, University of California, Santa Barbara.

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

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2):411–44.

Sagart, Laurent. 2014. In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic. *Language and Linguistics* 15(6):859–82.

Sapir, Edward. 1921. *Language*. An introduction to the study of speech. New York: Harcourt, Brace and Company.

Shepherd, John Robert. 1993. *Statecraft and Political Economy on the Taiwan Frontier, 1600-1800*. Stanford: Stanford University Press.

Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. In Paul Li et al. (eds.), *Austronesian studies relating to Taiwan*:683–726. Symposium Series of the Institute of History and Philosophy, Academia Sinica, no. 3. Taipei: Academia Sinica.

Steels, Luc and Eors Szathmary. 2018. The evolutionary dynamics of language. *Biosystems* 164:128–37.

Smith, Alexander. 2017. The Western Malayo-Polynesian Problem. *Oceanic Linguistics* 56(2):435–90.

Sung, Li-May. 2018. *A sketch grammar of Kanakanavu* (in Chinese). Series on Formosan Languages 16. New Taipei: Council of Indigenous Peoples.

Tanakingsing, Michael. 2009. A functional reference grammar of Cebuano. PhD dissertation, Nataionl Taiwan University.

Tanakingsing, Michael. 2013. A study on the behavior of Cebuano pronouns in discourse. *Concentric: Studies in Linguistics* 39(1):59–89.

Teng, Stacy Fang-ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics.

Trejaut, J. A., T. Kivisild, J. H. Loo, C. L. Lee, C. L. He, C. J. Hsu, Z. Y. Li, and M. Lin. 2005. Traces of Archaic Mitochondrial Lineages Persist in Austronesian Speaking Formosan Populations. *PLoS Biol* 3:e247.

Ward, J. V., J. S. Athens, and C. Hotton. 1998. Holocene pollen records from Babeldaob island, Palau, Western Caroline islands. Paper presented at the Annual Meeting of the Society for American Archaeology, Seattle, March 29.

Wild, Ilka. 2018. Voice and transitivity in Kanakanavu. PhD dissertation, Universitat Erfurt.

Wu, Joy Jing-lan. 2006. Verb classification, case marking, and grammatical relations in Amis. PhD dissertation, State University of New York at Buffalo.

Yamada, Yukihiro. 2014. *A grammar of the Itbayat language of the Philippines*. Himeji, Japan.

Zeitoun, Elizabeth. 2000. A reference grammar of Rukai (in Chinese). Taipei: Yuanliu.

Zeitoun, Elizabeth. 2000. *A reference grammar of Tsou* (in Chinese). Taipei: Yuanliu.

Zeitoun, Elizabeth. 2007. *A grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth. 2018. *A sketch grammar of Saisiyat* (in Chinese). Series on Formosan Languages 8. New Taipei: Council of Indigenous Peoples.

Zeitoun, Elizabeth. 2018. *A sketch grammar of Rukai* (in Chinese). Series on Formosan Languages 8. New Taipei: Council of Indigenous Peoples.

Zeitoun Elizabeth, Tai-hwa Chu, and Lalo a Tahesh Kaybaybaw. 2015. A study of Saisiyat morphology. *Oceanic Linguistics Special Publication* no. 40. Honolulu: University of Hawaii Press.

Zorc, David. 2020. Reactions to Blusts "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59:394–425.