The Sama-Bajaw Languages

Daniel Kaufman

To appear in: Antoinette Schapper and Sander Adelaar (eds.)
Oxford Handbook of Western Austronesian Languages. Oxford: Oxford University Press

1. Introduction

The Sama-Bajaw (henceforth SB) languages comprise one of the most remarkable subgroups of the Austronesian family. Despite being scattered throughout the Sulu archipelago, coastal areas of Borneo (including Pulau Laut in southern Borneo), Kangean Island (facing Madura), Sulawesi, the Timor area, and the Moluccas, the SB languages clearly display a phylogenetic unity suggesting descent from a single language. Equally remarkable is the fact that speakers of SB languages, to a large extent, maintain a common way of life as sea-nomads, with many present-day sedentary communities having settled down within historical memory.[1] It is very difficult to know the full geographical range of SB communities, even those that have been long established. For instance, Pallesen (1985) mentions several communities of Luzon in the Philippines but there is scant information confirming their existence and no information on their language. The more prominent SB communities are mapped out by Pallesen (1985) and Mead and Lee (2007) provide a comprehensive mapping of SB communities of Sulawesi, but those further east are not well documented.

There exist other sea-nomads speaking non-SB Austronesian languages closer to the southeast Asian mainland, most notably the Moken of Thailand (Larish 1999, Pittayaporn 2005) the Urak Lawoi' (Hogan 1988, 1999), and Malayic-speaking *orang laut* (sea people) populations around the Malaysian peninsula (Anderbeck 2012), but the majority of sea-nomads in the MPSEA region speak SB languages. This makes for a clear parallel between the Sama-Bajaw and the Roma "Gypsy" people of Europe who share both a historically nomadic lifestyle and a widely dispersed language family that, despite many layers of contact-induced changes, can be traced to a single ancestral tongue. Given the obvious similarity in nomadic lifestyle, Sama-Bajaw peoples are often referred to as the "Sea Gypsies" in both the scholarly and popular literature. Endonyms often employ a reflex of Proto-Sama-Bajaw (PSB) *saməh. In areas of Sulu, the term Bajaw has a stronger association with sea nomadism while Sama implies a degree of sedentarism, but Bajaw has become an endonym of more general use among the Indonesian communities regardless of whether they are sedentary or nomadic.

As elsewhere in the Austronesian world, it is not easy to distinguish independent SB languages from dialect. Pallesen's (1985) careful study of SB subgrouping, suggests the family tree in Figure 1, which has been adopted by the Ethnologue (Eberhard et al. 2020) and Glottolog (Hammarström et al. 2020).

[1] Notwithstanding the surprising correspondence between language and lifestyle, it should be stressed that modern SB groups represent "a wide range of economic and cultural types," in the words of Sopher (1965:54). Groups like the Yakan, Jama Mapun and Abaknon are thoroughly sedentary and have engaged in farming for many generations.

Sama-Bajaw
- **Inabaknon** [abx]
- **Sulu-Borneo** (7)
  - Borneo Coast Bajaw (3)
    - *Indonesian Bajau* [bdl] (Jampea, Jaya Bakti, Kajoa, Matalaang, Poso, Roti, Same', Sulamu, Togian 1, Togian 2, Wallace)
    - *Mapun* [sjm]
    - *West Coast Bajau* [bdr] **(**Banggi**,** Kawang**,** Kota Belud**,** Papar (West Coast Bajau)**,** Pitas Bajau**,** Putatan**,** Sandakan Bajau)
  - Inner Sulu Sama (3)
    - *Balangingi* [sse] (Daongdung, Kabinga'an, Lutangan, Nuclear Balangingi, Sibuco-Vitali, Sibuguey)
    - *Central Sama* [sml] (Dilaut-Badjao)
    - *Southern Sama* [ssb] (Bajau Banaran, Bajau Darat, Bajau Laut, Bajau Semporna, Balimbing, Bongao, Laminusa, Languyan, Obian, Sama, Sapa-Sapa, Sibutu', Sikubung, Simunul, Sitangkai, Tandubas, Ubian)
  - Pangutaran Sama [slm]
- **Yakan** [yka]

Figure 1. A Sama-Bajaw family tree (Pallesen 1985)

The Sama languages of the Philippines are relatively well described, with comprehensive dictionaries existing for Mapun (Hashim, Collins and Collins 2001), Yakan (Behrens 2002) and smaller dictionaries for Sama-Pangutaran (Walton and Walton 1992), Sama Bangingi' (Diment and Gault 1980, Diment 1995), as well as word lists for Abaknon (Jacobson 1999). Descriptive grammars and grammar sketches exist for Southern Sinama (Akamine 1996, 2005), West Coast Bajau (Miller 2007), Sama Bangigui' (Gault 1999), Central Sinama (James 2017), Pangutaran Sama (Walton 1986), Yakan (Brainard and Behrens 2002) and eastern Indonesian varieties (Verheijen 1986), although thorough descriptions of the SB languages of eastern Indonesia are completely lacking. Abaknon (also known as Inabaknon) is an SB outlier within the Philippines. It is the northernmost documented variety, whose speakers have a long history of settlement in the eastern Visayas region. Unlike most other SB communities, the Abaknon have assimilated in lifestyle to their non-SB (Visayan) neighbors and appear to have separated from other SB groups before Islamization (and consequently lack the Arabic loans found in other SB varieties).

The phylogeny and contact-induced developments of the SB languages of Sulu are dissected in exquisite detail by Pallesen (1985), a work which seems underappreciated outside of Philippine linguistics but which has greater import to the broader field of language contact. Pallesen demonstrates how contact led to bidirectional influence between Tausug, a Central Philippine language, and the SB languages of the Sulu archipelago. He furthermore locates the center of greatest phylogenetic diversity in the Sulu archipelago (specifically, the Sulu-Sibuguey Bay area). As seen in Figure 1, he identifies Abaknon as a first order branch of the family followed by Yakan. The place of Abaknon in the family tree may, however, be obscured by its heavy contact with Central Philippine languages and its relative isolation from other SB languages for far longer than any other variety. Blust (2007:78) in fact notes several diagnostic sound changes suggesting Yakan as the first SB language to branch off, rather than Abaknon. On

the basis of lexicostatistics, Pallesen proposes that PSB began to diversify in the area of Sulu circa 800 CE. He notes, however, on the basis of typology, that SB shows closer connections to the south rather than the Philippines:

> "A number of distinctive characteristics (e.g. the *h* reflex of PAN *R, the semantic features of the phrase marking particles or prepositions, the lack of verbal inflection to mark the action-begun vs action-not-begun contrast, a 7-vowel system, a uniquely marked agentive phrase) suggest an Indonesian origin rather than any close relationship to the Central Philippine languages with which many SB daughter languages are currently in geographical proximity." (Pallesen 1985:245)

In the next major work on the history of the subgroup, Blust (2007) demonstrates a more specific link between PSB and the Barito languages of southeast Borneo and claims that the speakers of PSB emerged from the Barito river basin at around the same time that the Proto-Malagasy left Borneo for the east African coast. He infers from borrowed maritime vocabulary that the PSB communities were "an originally landbound population drawn out of southeast Borneo by trade contacts with a more maritime people" (Blust 2007:103), similar to the Malagasy (Adelaar 1989, 1995). Blust (2007:91-95) sums up all the correspondences between PMP and PSB and presents novel evidence for his argument involving both sound changes and lexical replacements.[2]

In contrast to the high diversity of Sulu, the area with the least internal diversity seems to be Sulawesi and eastern Indonesia. Verheijen (1986), based on informal comprehension tests with the eastern varieties, suggests that they are in a dialect relation to each other. Mead and Lee (2007) tentatively confirm this with a lexical similarity score of around 90% across the varieties of Sulawesi, Moluccas and the Lesser Sunda Islands.

This chapter is organized as follows. I review the salient phonological features of the SB languages (§2), examine questions of lexical category in the better described SB varieties (§3). Section 4, to which the majority of the chapter is devoted, examines SB syntax through a comparative lens. Finally, I sum up in section 5 with notes on the role of language contact and suggestions for further research.

2. Phonology

Pallesen reconstructs PSB with a seven vowel system: /a, e, o, i, u, ə, ʉ/ although *ʉ and *ə are not contrastive with each other in all positions and their independence is still open to question.[3] Typical vowel systems of Sulu either show a six-vowel system (/a, e, o, i, u, ə/) system or a five-vowel system (/a, e, o, i, u/), where *ə has merged with one or more other vowels. Abaknon has reduced the system even further to a three vowel system (/a, i, u/), as found in many Central Philippine languages.

---

[2] The following changes, among others, distinguish the SB languages from all their Philippine neighbors: PMP *R>h, gemination of onsets following PMP *e, vowel lowering before *$ʔ$, and final devoicing. Blust (2007) shows that lexical replacements, such as PMP *$qulu$ 'head' > PSB*$takuluk$, PMP *$qudip$ 'living, alive' > PSB *$belum$, betray SB's Bornean origins.

[3] The only attested SB languages to show a full 7-way vowel distinction are Sama Batuan and Sama Pangutaran. Pallesen (1979, 1985:74) argues that PSB *$ə$ and *$ʉ$ result from "a phonemic split of PAN *$ə$, probably conditioned by stress."

In the native stratum, SB languages show final devoicing, merger of historic **l* and **r*, as well as post-schwa gemination. Additionally, pre-tonic vowel neutralization and intervocalic /l/ deletion are commonly found in the SB languages of Sulu. Both of these processes are seen in Yakan (1), where vowels are reduced to [ɛ] in pre-stress (pre-penultimate) positions and a productive rule of intervocalic /l/ deletion operates on the root onset.

	Yakan, vowel reduction		(Pallesen 1985:76, Brainard and Behrens 2002:7)
(1)	/mag-pa-ˈlaboʔ-an/	→	[mɛg-pɛːˈboʔan]
	AV-CAUS-drop-LOC		‘repeatedly drop something’

The native stratum also shows a simpler phonotactic template, only allowing a limited range of consonant sequences (nasal-stop clusters and geminates). Contact with Central Philippine languages has led to more complex phonotactics while contact with languages of Sulawesi has, in some cases, led to further simplification. For instance, the varieties of eastern Indonesia only allow /l, r, s, ŋ, ʔ, h/ as word-final codas, with historical oral stops (both voiced and voiceless) having gone to ʔ in word-final position and word-final nasal stops merging with the velar nasal, as found commonly in Sulawesi.

In almost every SB language, there exists a minor degree of vowel harmony, typically occurring with suffixes that harmonize with their stem. This can be seen in West Coast Bajau /pogos-an/ force-APPL → [pəgoson] where the suffix has assimilated to the final root vowel and the first vowel of the root has been neutralized to [ə] (in pre-tonic position).

No SB language shows the phonemic vowel length distinction in penultimate syllables characteristic of Philippine languages. Rather, most SB languages have been described as having a right-aligned trochaic, stress pattern. Some varieties, such as Pangutaran Sama, have been further described as having iterative secondary stress preceding the primary stress on the penultimate syllable (Pallesen 1979:192). The stress window includes suffixes and genitive pronominal enclitics, as shown in (2).[4]

	Central Sinama		(Pallesen 1985:94)		
(2) a.	*bónoʔ*	b.	*bonóʔ-un*	c.	*pag-bonoʔ-án-bi*
	kill		kill-UV.IMP		GER-kill-NMLZ-2PL.GEN
	‘to kill’		‘Kill it!’		‘the cause of your act of fighting together’

All SB languages possess a reflex of the PMP prefixes *maŋ- and *paŋ-, which are used for various derivational functions and which trigger a range of different morphophonological behaviors depending on the type of segment the stem begins with (Blust 2004). The most common pattern in SB languages, found both in Sulu and eastern Indonesian varieties, involves assimilation to and deletion of stem-initial /p, b, t, s, k, ʔ/. However, with nasals, liquids and the voiced obstruents /d, dʒ, g/, we find the allomorph /ŋaŋ-/ (followed by assimilation or deletion of the affixal coda), e.g. *ŋan-doleʔ* ‘to anger’ (Akamine 1996:40). Blust’s (2004) survey shows that the pattern of vowel epenthesis with stem-initial voiced stops (e.g. /ŋ-t.../ vs. /ŋa-d.../) is strongly centered in Borneo (with the exception of Sundanese), and is found across different subgroups (e.g. Tombonuwo *moŋod-*, Kadazan *momod-*, Timugon Murut, *mamad-*, Kayan *ŋed-*). SB languages are unusual in that, despite the vowel epenthesis, a nasal stop cluster still occurs. In

[4] To facilitate comparison, the glossing and presentation of examples from various sources has made to conform with a relatively neutral analysis (e.g. with reference to voices instead of their purported morphosyntactic functions).

the Bornean languages, epenthesis serves to separate the prefix-final nasal from the stem-initial voiced stop.

3. Lexical categories and their basic ordering relations

On the level of full words, descriptive grammars of SB languages have defined lexical categories on the basis of semantics (e.g. Miller 2007:95) or by voice and person morphology, although morphological criteria have not been applied rigorously. I thus adopt a standard, uncritical view of lexical categories in the following, with a notional categorization of nominal, verbal and adjectival roots.

Nouns are most often underived roots but can also be derived from verbal roots with a reflex of PMP *-*an* in combination with PMP **ka-* and **paŋ-*. Brainard and Behrens (2002:11) treat the Yakan suffixal determiner *-in* as deriving nouns in forms like Yakan *ma-hāp-in* (ADJ-good-DET) 'the good one' and *mag-belli-hin* (AV-buy-DET) 'the one who buys', but such data also easily support an analysis where the general determiner *-in* simply does not discriminate with regard to the category of its host. Because there exist event-denoting roots that function as predicates without voice, aspect or agreement morphology, there is an ambiguity with roots such as *uran* in (3), just as we find in Malay.

West Coast Bajau

(3) a. *Ai* *uran.*
PFV rain
'It's begun to rain.'

b. *uran* *pan duwai...*
rain also fell
'... the rain fell...' (Miller 2007:178)

James (2017:59) notes a more general difficulty in distinguishing between verbal and non-verbal predicates in Central Sinama, as putative nouns function as predicates without a copula and putative verbs can function as arguments to a certain extent. He notes that evidence for a noun-verb distinction may, however, also be found in TAM marking; while the tense/aspect markers very commonly precede verbal predicates, they only rarely precede nominal ones, although such combinations may not be completely ungrammatical.

As in nearly all Austronesian languages, entity-denoting roots can take voice morphology and thus become event-denoting words, as in (4).

Yakan

(4) *Sinna-ku* *mag-luma?* *dem puweblo*
like-1SG.GEN AV-house in town
'I like to live in town.' (Brainard and Behrens 2002:236)

Adjectives are argued by Miller (2007: 101-105) to not constitute an independent morphosyntactic class in West Coast Bajau but rather to be a subtype of intransitive verb. Verbs and putative adjectives can be modified by the same aspectual markers, intensifying adverbs, and form predicates in their bare form. They also share the same morphological potential. Property-denoting words in most SB languages are also bare roots, again as in Malay. Exceptions to this include Yakan and Sama Bangingi', where a stative prefix *ma-*/*a-* is commonly found on adjectives (e.g. *a-ha:p* STA-good 'good', *a-bottoŋ* STA-stomach 'pregnant') and Abaknon, where we find a *ma-* prefix in the same function, a likely borrowing from a Central Philippine language

rather than a direct retention of PMP **ma-*. Similarly, in Bajau Mola (southeast Sulawesi), we find occasional use of an adjectival *ma-* but these are apparent loans from Bugis (e.g. *marannu* 'happy'). James (2017:33) notes that not all "adjectives" require the *a-* prefix in Central Sinama and that event denoting predicates like *lahi* 'flee', *həlliŋ* 'say' also take this prefix.

Manner adverbs are formed with adjectival bases using the *pa-* prefix in a number of SB languages, such as Mapun (5), even if the roots are unaffixed when used as adjectival modifiers.

Mapun

(5) *Lay ya* *sonse* *pa-taŋkas*
PFV 3SG.NOM run ADV-quick
'She ran quickly.' (Hashim et al. 2001:32)

Negation offers evidence for a two split across lexical categories in most SB languages. Typically, one negator is used for verbal and adjectival predicates (e.g. Abaknon *gaʔi*, Yakan *gaʔ*, Manuk Mangkaw *maha*, Sama Dilaut *mbal*) while another is used for nominal and prepositional phrase predicates (e.g. Abaknon *maʔin*, Yakan *dumaʔin* [both derived from PWMP **laqin* 'different, another', Blust & Trussel ongoing], Manuk Mangkaw *sikeyya*, Sama Dilaut *halam*).

Unlike most languages of the Philippines, there exists a class of bound roots in many SB languages, which are claimed to not occur without stem-forming voice or valency affixes. Miller (2007:97) argues that bound roots in West Coast Bajau (e.g. **puleʔ* 'return') are inherently verbal, based on their morphological behavior. In Philippine-type languages, valency is typically determined by voice in combination with valency changing morphology with great flexibility on the part of the roots. Transitivity in SB languages may not be as flexible. For instance, monovalent roots, e.g. *teko* 'arrive', generally resist taking the passive/undergoer voice (Miller 2007:98, Donohue 1996:785).

SB languages have a richer inventory of true prepositions when compared to the languages of the Philippines. Akamine (2005:385) enumerates five for Simunul Island Sama: *leʔ* AGENT/REASON, *ma* LOCATION, *ni* GOAL, *min* SOURCE and *maka* INSTRUMENT, COMITATIVE. Unlike the prepositions of other CSP languages, these take noun phrase complements directly rather than as oblique case phrases. The origin of some but not all of these prepositions can be traced to earlier sources. The agent marker *leʔ* and its many SB cognates descend transparently from PMP **uliq* 'get, obtain' (see §4.4 below). A common goal marker *pa*, as seen in (6), most likely represents a degrammaticalization of PMP **pa-* in one of its common non-causative functions (cf. Cebuano *pa-iŋun sa gawaŋ* DIR-toward OBL door).

Pangutaran Sama

(6) *t<um>uju* *aʔa* *pa* *lumaʔ* *saupak*
<AV>toward person OBL house Saupak
'The man is headed for Saupak's house.' (Walton 1986:87)

On the other hand, the instrumental preposition *maka*, appears to be a grammaticalization of Malay *makai* < *memakai* (/məŋ-pakai/) 'use', as a fuller form can still be seen in the same function in (7).

West Coast Bajau

(7) *Boi_jo* *Ø-tataʔ=ni* *anak makay* *gayuŋ…*
after UV-pour.water=3SG.I child AV.use bucket
'Just after he poured water on the child using a bucket…' (Miller 2007:235)

Pallesen (1985) reconstructs PSB pronouns as in Table 1. Note that the nominative pronouns, which are second position clitics in some SB languages, only differ minimally from the independent pronouns. SB pronouns differ from most surrounding languages in having lost the number distinction in the third person nominative and independent sets. The distinction is reasserted through a number of independent innovations, e.g. Abaknon *maŋa iya* (a 3sg pronoun preceded by a common Central Philippine plural marker), Central Sinama *sigala* (via Malay *segala* 'all' ultimately from Sanskrit *sakala* 'complete, entire, all', see Pallesen 1985:201), Bajau Mola *disiʔ iru* (via a demonstrative meaning 'those over there', see Donohue 1996:784).

Table 1. PSB Pronouns (Pallesen 1985:103)

		Set I NOMINATIVE	Set II GENITIVE	Set III INDEPENDENT
Minimal	1	**aku, *ku*	**-ku*	**aku*
	2	**kaa, *kaw*	**-nu*	**kaʔa, *kaʔaw*
	1+2	**kitəh*	**-təh*	**kitəh*
	3	**iəh*	**-nəh*	**iəh*
Augmented	1	**kami*	**-kami*	**kami*
	2	**kaam*	**-bi(i)*	**kaʔam*
	1+2	**kitəh*	**-təh*	**kitəh*
	3	**iəh*	**-dəh*	**iəh*

Pronominal objects of actor voice verbs are avoided in Philippine varieties but when there is no choice, SB languages vary in which set they employ for this function with the independent set (Set III) enjoying preferred status. Note also that, in accordance with a widespread Philippine pattern, a first person singular agent acting on a second person patient is often expressed with a reflex of **-təh* rather than the expected **-ku* (James 2017:25). In contrast, Bajau Mola has adopted an areal pattern of south Sulawesi in using the first person plural inclusive as a polite second person pronoun.

Besides the above major categories, there are a number of TAM markers and adverbial clitics which cannot be reviewed fully here. The independent aspect markers supplant PMP syllable reduplication marking progressive (not to be mistaken with full reduplication marking continued action, cf. Miller 2007:78) and the PMP perfective marker **<in>* on verbs. As discussed below, the reflex of **<in>* survives as a marker of passive voice or a resultative.

The basic word order relations in the SB languages appear similar to Malay. These can be exemplified with West Coast Bajau (Miller 2007):

(i) Possessors and adjectives must follow the phrases they modify, e.g. *moto Deli* (eye Deli) 'Deli's eye', *beriu daras* (wind strong) 'strong wind'.
(ii) Complementizers (e.g. *engko'*) and adpositions uniformly precede their complements.
(iii) Relative clauses tend strongly to follow the head noun they modify (see §5.1).
(iv) Negation and auxiliaries precede verbs, e.g. *nya' buli s<in>egir* (NEG can <PASS>touch) 'most not be touched'.

(v) The comparative marker precedes the adjective which precedes the standard of comparison, e.g. *lagi langa man kam* (more tall than 2PL.II) 'taller than you'.

The ordering relations that are more difficult to generalize over involve adverbs and the relative order of subject and predicate. The northern SB languages are more thoroughly predicate initial while a subject initial order is claimed to be unmarked for actor voice clauses in West Coast Bajau and certain SB languages of Indonesia. The positioning of adverbs depends on adverb type and discourse factors but has not been described in any detail.

4. Grammatical relations

4.1 Voice and valency

An unexpected feature of Philippine SB languages is their apparent maintenance and even elaboration of the basic four-way PMP voice system, despite the simplification of the case system and total loss of verbal aspect morphology. The five-way voice distinction of Central Sinama is shown in (8).

Central Sinama

(8)a. *amonoʔ* *aku* *edoʔ* *maka* *lahut* *itu*
AV:kill 1SG.NOM dog with knife this
'I will kill the dog with this knife.'

b. *bonoʔ-ku* *edoʔ* *maka* *lahut* *itu*
kill-1SG.GEN dog with knife this
'I will kill the dog with this knife.'

c. *pamonoʔ-ku* *edoʔ* *lahut* *itu*
IV:kill-1SG.GEN dog knife this
'I will kill the dog with this knife.'

d. *bonoʔ-an-ku* *kaʔa* *edoʔ*
kill-BV-1SG.GEN 2SG.NOM dog
'I will kill the dog for you.'

e. *pamonoʔ-an-ku* *edoʔ* *lantay* *itu*
LV:kill-LV-1SG.GEN dog floor this
'I will kill the dog on the floor.' (Pallesen 1985:96-7)

The actor voice is expressed with a reflex of PMP **maŋ-* in (8a) (usually *ŋ-* with nasal substitution), originally a pluractional/distributive actor voice prefix. In SB languages, it has been analyzed as an antipassive/intransitive marker (Gault 1999, Trick 2008, Brainard & Behrens 2002) as well as an inchoative (Walton 1986) but is most commonly glossed as actor voice/focus. The patient voice in (8b), on the other hand, is expressed with the bare verb stem. The morphology of both voices thus resembles Malay and languages of Indonesia that have replaced PMP **-en* PATIENT VOICE with a bare verb stem and widened the function of PMP **maŋ-* to become a default actor voice, subsuming PMP **<um>*.

Whereas the PAn voice marker **Si-* was polysemous in selecting an instrumental, beneficiary and conveyed object as the pivot, these functions are cleaved apart in SB. The instrumental voice is marked uniquely by *paŋ-* < PMP **paŋ-* INSTRUMENTAL, as seen in (8c), while beneficiaries, recipients and kindred roles are selected by *-an* < PMP **-an* LOCATIVE, as in

(8d).[5] The use of *paŋ-* to mark instrumental voice is also found in Central Philippine languages and elsewhere, which may either indicate parallel innovation or a function that was already present in PMP (perhaps to derive instrumental nouns), but not fully incorporated into the voice system. The locative (which corresponds to true location, rather than directional or oblique roles) is expressed with a combination of PMP **paŋ-* and **-an*, as seen in (8e).

Pallesen reconstructs the PSB voice markers as shown in Table 2, with two moods and two aspects in the indicative. What Pallesen terms "perfective" really derives from the PMP potentive paradigm, which is often used to mark accomplishment, ability and accidental action, functions that are also continued in SB languages. Note also that while the PMP **-en* patient voice has been eliminated in the indicative voice, it survives in the imperative as **-un* (as in modern Javanese). Where we might expect a reflex of PAn **Si-* in the instrumental (Pallesen's "accessory"), we again find a reflex of **-en*. Finally, PSB shows a unique innovation in the imperatives of the benefactive/referent and locative voice, which appears to be a blend of the PMP dependent mood locative **-i* (used commonly for locative imperatives in Philippine languages) and the final *-n* found in all the other suffixes in the voice paradigm.

Table 2. Transitive verbal affixes (Pallesen 1985:99)

VOICE	MOOD		
	INDICATIVE		IMPERATIVE
	IMPERFECTIVE	PERFECTIVE	
Actor	**N-*	**maka-*	**N-*
Patient	**∅*	**ta-*	**-un*
Accessory	**paN-*	*(*)tapaN-*	*(*)paN- -un*
Referent	**-an*	**ka- -an*	**-in*
Locative	**paN- -an*	*(*)kapaN- -an*	*(*)paN- -in*

The presentation of Pallesen's paradigm in (8) and Table 2 somewhat deceptively makes SB appear as a Philippine-type language. In reality, SB languages may be more amenable to a two-voice (actor vs. undergoer) analysis with the other voices in the paradigm being applicatives. All SB languages allow "referent voice" to combine with either undergoer voice (unmarked) or actor voice. For instance, in West Coast Bajau, we find *beli-an* 'to buy for X' but also with the actor voice stem, yielding *meli-an* 'X buys for Y', as well as the passive stem, yielding *b<in>eli-an* 'X is bought something (by Y)' (Miller 2007:274). The promotion of a prepositional object to a direct object in this manner is shown in (9).

West Coast Bajau

(9) a. *Endo=ku* *muan* *peranggi' e* *m-aku*
wife=1SG.I AV:give pineapple DEM PREP-1SG.II
'My wife gave the pineapple to me.'

b. *Endo=ku* *muan-an* *aku* *peranggi' e*
wife=1SG.I AV:give-APPL 1SG.II pineapple DEM
'My wife gave me the pineapple.' (Miller 2007:282)

[5] The benefactive use of the locative voice is also found in Philippine languages as an option under certain conditions.

Walton (1986:87-94) argues explicitly that Pangutaran Sama *-an* should be treated as a valency increaser (i.e. applicative) rather than its own voice. Note that Pangutaran Sama, like all other SB languages, has an applicative use of *-an*, similar to Malay, but at the same time resembles Philippine languages in its verb-initial syntax and strong tendency for an indefinite interpretation of actor voice objects and definite interpretation of the pivot (see §4.3). In (10a), we see a plain undergoer voice clause with a benefactive adjunct expressed as an oblique phrase (*ma si Andi*). In (10b), we see the promotion of the benefactive to the pivot of the undergoer clause with the help of *-an*. In (10c), we see an actor voice clause where the actor is the pivot and the benefactive is again an oblique phrase. So far, the facts abide by typical the Philippine pattern. But in (10c) we see an actor voice clause co-occurring with *-an* and the former oblique phrase promoted to object. This divergence from the Philippine pattern is another typological feature which points to a southern origin for SB languages.

Pangutaran Sama (Walton 1986:88-89)

(10)a. *Ø-bəlli-ku taumpaʔ ma si Andi*
UV-buy-1SG.GEN shoes OBL PM Andy
'I bought the shoes for Andy.'

b. *Ø-bəlli-an-ku si Andi taumpaʔ*
UV-buy-LOC-1SG.GEN PM Andy shoes
'I bought Andy some shoes.'

c. *məlli=aku taumpaʔ ma si Andi.*
AV:buy=1SG.NOM shoes OBL PM Andy
'I bought some shoes for Andy.'

d. *məlli-an=aku si Andi taumpaʔ*
AV:buy-LOC=1SG.NOM PM Andy shoes
'I bought Andy some shoes.'

The status of the instrumental and the locative is even less clear. There is little evidence from any SB variety that actor voice ever co-occurs with the instrumental derived from **paŋ-* and the locative voice circumfix **paŋ- -an* (cf. James 2017:66). The possibility of recent convergence with the cognate Malay nominalizations, *pəŋ-* AGENT NOMINALIZER and *pəŋ- -an* GERUND, is unlikely, as the semantics of these in Malay/Indonesian and SB have drifted apart considerably (cf. Miller 2007:296-8). If Pallesen is correct in the reconstruction of an imperative mood for all voices, as in Table 2, then Indonesian varieties must have simplified this part of the paradigm. This, however, remains to be worked out, because imperatives for the instrumental and locative voices are also lacking in Philippine varieties (Walton 1986:10, Akamine 2005:389). More likely is a scenario in which nominalizations derived with **paŋ-* have been incorporated into the voice system to a greater extent in SB languages of Sulu through contact with Philippine languages.

Pallesen also reconstructs what he terms "secondary verbal affixes" related to valency and aspect. These include PSB **si-* RECIPROCAL and **pa-* CAUSATIVE whose forms and functions are inherited directly from PMP. There are also several uses of a *pa-* prefix that are not, strictly speaking, causative but which probably descend from the same PMP morpheme. In many SB varieties, *pa-* forms motion verbs from deictics, location words, and body positions (Walton 1986:75). Several SB languages have the unusual property of combining causative *pa-* with the actor voice prefix to yield an active causative prefix *ma-*. There is little need to avoid homophony with widespread stative *ma-* as the stative prefix is highly marginal in SB and has been reduced to *a-* in several languages.

A prefix cognate to PMP **ka-* is also commonly found and plays a role in the potentive paradigm, typically cancelling out an agentive interpretation (Walton 1986:83). The undergoer counterpart of this prefix is *ta-*, whose distribution is robustly southern and only rarely found in

the Philippines, despite being reconstructable to PAn **taR*- (Blust and Trussel ongoing). Actor voice and undergoer voice potentive clauses are shown in (11).

Pangutaran Sama (Walton 1986:101)

(11) a. *ka-kəlloʔ si mmaʔ daiŋ*
AV.POT-get PM father fish
'Father was able to get some fish.'

b. *ta-kəlloʔ aʔa daiŋ kuhapu*
UV.POT-get PM father fish
'A man was able to catch the grouper fish.'

Most SB languages differ from Malay in making a clearer distinction between undergoer voice, with an unmarked verb, and passive voice, with a reflex of PMP **<in>* (although see van den Berg and Mead, this volume, for parallels in Sulawesi). In the Yakan undergoer voice, neither agent nor patient are case marked but the agent must be adjacent to the verb, as seen in (12a). In (12b), we see what is often considered a passive; the verb is marked with *<in>,* which, unlike Philippine languages, has no association with aspect. With the use of *<in>*, the agent must take the agentive marker *weʔ* and can appear in a wider range of positions in the clause. This appears to be equally true for Pangutaran and other SB languages of Sulu.

Yakan

(12)a. *pogpog [aʔa] sawe-hin [*aʔa]*
hit person snake-DEF person
'A person hit the snake.'

b. *p<in>ogpog [weʔ aʔa] sawe-hin [weʔ aʔa]*
<PASS>hit AGT person snake-DEF AGT person
'A person hit the snake.' (Brainard and Behrens 2002:113)

Prepositionally marked agents are also used for non-agentive causers but in this case, the predicate is typically not marked with a reflex of *<in>, as shown in (13).

Pangutaran Sama

(13) *tutuŋ uk lətteʔ kabbun-kami*
burn AGT lightning plantation-1PL.EX.GEN
'Our plantation was burned by lightning.' (Walton 1986:62)

While (13) shows a passive agent without a passive-marked verb, we also find a passive marked verb with a bare agent in Mola Bajau (14c), where it can also be introduced by the oblique markers *ma* or *aleh*.[6] The actor voice and undergoer voice are shown in (14a) and (b) for comparison. It is not clear how the undergoer and passive voice differ syntactically in Mola Bajau or to what extent (14c) represents a real passive.

[6] A bare agent with a passive marked verb appears to be rare in SB languages and the Mola construction may be a recent calque from Indonesian which allows both bare agents of passive verbs as well as those marked by the agentive preposition *oleh*. Note also that the Mola passive is uniquely (among SB varieties) marked with *di-*, as in Malay.

Mola Bajau (Donohue 1996:784)

(14)a. *ŋ-ita uggoʔ aku*
AV-see pig 1SG
'I saw the pig.'

b. *kita-ku uggoʔ*
see-1SG.GEN pig
'I saw the pig.'

c. *di-kita-ku uggoʔ*
PV-see-1SG.GEN pig
'The pig was seen by me.'

Actor voice clauses show more flexibility in the ordering of arguments in the postverbal domain. This can give rise to ambiguity in some languages, as in (15).

Manuk Mangkaw Sinama

(15) *Bey nipaʔ kambiŋ kuda*
already AV:kick goat horse
'The goat kicked a horse.' OR 'The horse kicked a goat.' (Akamine 1996:73)

Akamine (1996:73) shows that definiteness disambiguates the relations in the Manuk Mangkaw actor voice. If one of the arguments in (15) is marked as definite (with a following demonstrative), it must also be interpreted as the subject, abiding by a common tendency in Austronesian languages for the pivot to be definite and actor voice undergoers to be indefinite. We examine the relation of this constraint to transitivity and alignment below in section 4.3.

4.2 Pronominal arguments

In all documented SB varieties, genitive/ergative pronominal arguments are verb-adjacent enclitics (Set II) while nominative pronouns are either free (Set III) or found in second position (Set I). Unlike most languages of the Philippines, second-position clitics in SB languages of Sulu can be hosted by complementizers like *baŋ* 'if', as in (16).

Sama Bangingi' (Gault 1999:78)

(16) *baŋ=aku iŋgaʔi pa-billi-nu...*
if=1SG.NOM NEG CAUS-buy-2SG.GEN
'If you won't sell to me...'

This type of split in positioning between genitive and nominative clitics recurs in various areas of Indonesia (Haaksma 1933, Billings and Kaufman 2004, Himmelmann 2005, Kikusawa 2003) but the second-position condition on nominative clitics is stronger in SB languages of the Philippines than those of Indonesia. Free pronouns play an expanded role in Indonesian varieties, which tend to place the pivot in clause-initial position, as in (17). Free pronouns are generally used only for emphasis or as predicates in Philippine varieties.

Eastern Indonesian Bajau

(17) *kau korobbaŋ-ku*
2SG slaughter-1SG.GEN
'I will slaughter you.' (Verheijen 1986:21)

Clitic doubling has gone unexamined in SB languages although its existence is clear from published texts and other examples. Verheijen analyzes the ergative clitics in (18) as object clitics, but the interpretation in (18b) would be impossible on his view. Interestingly, the clitics

are doubling an agent introduced by the preposition *alé* rather than an unmarked ergative argument.

Eastern Indonesian Bajau

(18)a. *baraʔ-an-**na** né **alé** **enda-na** ka…*
tell-APPL-3.GEN already AGT wife-3.GEN to
'His wife told it to…'

b. *soho-**na** lagi **alé** **anaʔ-ku** ka kita*
command-3.GEN again AGT child-1SG.GEN to 2.POL
'(I) was again ordered by my son to (go) to you.' (Verheijen 1986:21)

Ergative clitic doubling and agreement are ubiquitous in Sulawesi and other parts of eastern Indonesia but lest this be dismissed as a contact effect, it should be pointed out that Abaknon (19) also shows this pattern, despite being geographically embedded in the Central Philippine subgroup where clitic doubling is absent.

Abaknon

(19) *Ag-laklak-**na** **si** **idoʔ** i luhoʔ*
ASP-drink-3.GEN OBL dog NOM soup
'The dog drinks the soup.' (Jacobson 1999)

4.3 Transitivity, definiteness and alignment

Recent work on SB languages has focused on questions of alignment type and transitivity. In this regard, the SB languages are more conservative than Malay, despite the two subgroups sharing several morphosyntactic innovations enumerated earlier. Definiteness correlates strongly with grammatical relations in Philippine varieties but much less so in Indonesian varieties, which appear more symmetrical in the sense of Himmelmann (2005), that is, having "at least two voice alternations marked on the verb, neither of which is clearly the basic form" (cf. Foley 2007). The resistance of actor voice clauses to definite undergoers in Philippine varieties has led many analysts to treat these languages as ergatively aligned rather than symmetric, with the actor voice as an antipassive (Gault 1999, Akamine 1996, Trick 2008). A concrete example from Yakan is seen in (20).

Yakan (Brainard and Behrens 2002:160)

(20)a. *kehet dende-hin kenna-hin* b. *ŋehet kenna dende-hin*
cut woman-DEF fish-DEF AV:cut fish woman-DEF
'The woman cut up the fish.' 'The woman cut up fish.'

Uniquely in Yakan, the pivot argument must be suffixed with the definite marker *-in*. In a transitive clause, such as (20a), the agent can also be suffixed with *-in* and interpreted definitely. However, in a canonical matrix clause such as (20b), the undergoer argument of an actor voice verb cannot be marked with the definite marker. A similar situation holds for Pangutaran Sama, although here the definite interpretation of the pivot comes "for free" without the use of determiners, as shown in (21).

Pangutaran Sama (Walton 1986:7)

(21) a. *Ø-tauʔ-ku kahawa ma siliʔ*
UV-put-1SG.GEN coffee OBL teapot
'I put the coffee in the teapot.'

b. *nauʔ aku kahawa ma siliʔ*
AV:put 1SG.NOM coffee OBL teapot
'I put some coffee in the teapot.'

Even when the undergoer takes a definite possessor, it is interpreted with lower transitivity in the actor voice, as shown in (22).[7]

Pangutaran Sama (Walton 1986:120)

(22) a. *Ø-bonoʔ sultan bantaʔ-na*
UV-kill king enemy-3SG.GEN
'The king killed his enemy.'

b. *monoʔ sultan bantaʔ-na*
AV:kill king enemy-3SG.GEN
'The king kills/fights some of his enemies.'

As suggested by (23), this is not the case for West Coast Bajau, which allows definite actor voice objects more freely, as in Malay.

West Coast Bajau

(23) *Dela e pan nambut* ***iyo*** *taʔ beluang...*
man DEM TOP AV:receive 3SG.II PREP door
'The man welcomed him at the door...' (Miller 2007:163)

In West Coast Bajau, and presumably other varieties of western Indonesia that have been under the constant influence of Malay, we see two concomitant changes; the actor voice allows for definite objects more freely and takes on SVO as the unmarked word order. The asymmetric word order change in the actor voice but not the undergoer voice, which remains largely predicate initial, also reflects the historical development of Malay, as detailed by Cumming (1991).

The status of the "passive" versus the undergoer voice remains an open question for most SB languages. Trick (2008) shows that clefting, relativization and question formation in Southern Sinama are all restricted to the pivot argument, as is true for all SB languages and, more generally, nearly all syntactically conservative Austronesian languages. But there is a clear contrast in control constructions between Philippine-type patient voice clauses and what Trick (2008) treats as a plain transitive clause in Southern Sinama. He presents the data in (24) suggesting that the controllee (the null argument in the lower clause that must corefer with an argument in a higher clause) is restricted to the pivot.

Southern Sinama (Trick 2008)

(24)a. *Ka-bilahi-an-ku ni-liŋan-an leh si Ben*
NMLZ-want-NMLZ-1SG.GEN UV-call-APPL AGT PM Ben
'I want Ben to call [me].'

b. *Ka-bilahi-an si Ben ni-liŋan-an akú.*
NMLZ-want-NMLZ PM Ben UV-call-APPL 1SG.NOM
'Ben wants me to be called [by someone].' NOT: 'Ben wants to call me.'

[7] I maintain Walton's glossing of the null prefix as a marker of undergoer voice although this can just as well be applied to other SB languages described here.

Trick (2008) thus analyzes the Sama control pattern as following an ergative system in that only the absolutive argument can be controlled, whereas the Central Philippine pattern has been analyzed as being sensitive to a number of factors including thematic role and grammatical function (Schachter 1995). The behavior we see in (24) does, however, correspond to how control operates in Malay/Indonesian clauses with *di-* marked verbs and agents introduced by *oleh*. Unfortunately, Trick (2008) does not compare prefixed forms like *ni-liŋan-an* to *liŋan-an*, in the unmarked undergoer voice, and suggests that such a distinction is not at play with full noun phrases in this variety. If the expected opposition between the two types of undergoer voice does exist here, it may be that the *ni-* marked verbs are simply more passive-like than their bare undergoer counterparts, as suggested by other descriptions. This is clearly an area that requires further exploration.

4.4 The *leʔ* actor voice construction

Akamine (1996, 2002, 2003, 2005) describes a highly peculiar perfective construction involving the agentive marker *leʔ* introducing a verb with ACTOR VOICE prefix *ŋ-*, followed by an agent also introduced by *leʔ*, seen in (25).

	Manuk Mangkaw Sinama			
(25)	*leʔ*	*ŋ-ajal*	*leʔ*	*ku*	*manuk*
	?	AV-cook	AGT	1SG.GEN	chicken
	'I have cooked the chicken.' (Akamine 2005:391)				

This is odd on four counts: (i) *leʔ* does not introduce verbs in any other context; (ii) the verb is marked with the actor voice but behaves as an undergoer voice verb in expressing the agent with *leʔ* and giving a definite interpretation to the patient (Akamine 2002:361); (iii) the construction has a perfective meaning whose source is unclear; (iv) the agent can be fronted in this construction, as in (26), but here, only one instance of *leʔ* can appear.

(26)	*leʔ*	*ku*	*(*leʔ)*	*ŋ-ajal*	*manuk*
	by	1SG.GEN	?	AV-cook	chicken
	'I have cooked the chicken.' (Akamine 2005:391)				

I would like to briefly pursue Akamine's (2002:363) suggestion that "The *leq-* prefix and the *leq* preposition are possibly both derived from an earlier verb via different grammaticalisation paths" and attempt an explanation of these puzzling features that departs from previous proposals (Akamine 1996, Ross 2002). It appears all of the above anomalies obtain a natural solution by taking the above pattern to be a remnant of a serial verb construction based on PMP **uliq* 'get, obtain' (Blust & Trussel in progress), serving in two distinct but related functions. The use of reflexes of **uliq* to introduce an agent are widespread in Malayic and even found to some extent in eastern Indonesia (e.g. Bimanese, Manggarai), and are thus rather straightforward. In contrast, the first instance of *leʔ* is unusual but can be understood as an auxiliary based on the verbal meaning of 'get'. It is similar to a 'get' passive of the type commonly found in mainland Southeast Asia in its resultative and perfective semantics, although it appears freely with an agent. I propose that the structure in (25) represents restructuring (or "clause union"), in which the oblique agent is really an argument of the initial verbal *leʔ* but appears after the lexical verb

in the actor voice. We have already seen earlier that oblique agents occur not only with passives in SB languages but also with certain stative predicates, as in (13), above. As seen for several Philippine and Formosan languages in Kaufman (this volume [**CSP**]), actor voice plays a dual role as both a voice marker and a voice-neutral infinitive marker. Actor voice is required on the lexical verb in this construction due to its subordinating function rather than its voice function. Finally, in (26), restructuring has not taken place and the agent is in the expected position of an argument of the verbal auxiliary *leʔ*. On this account, we do not expect an additional *leʔ* to follow the agent in (26) and we predict both the semantics of the construction and the apparent anomalous voice on the lexical verb.[8] Further descriptive work on the details of this construction should support or disconfirm this analysis.

5. The noun phrase

The typical linear order of elements within an SB noun phrase is shown in (27):

(27) Case | Plural | **Num** | **Class** | Noun | Poss | Adjective | Relative | **Num** | **Class** | Dem

Note that post-nominal adjectives and demonstratives are a southern feature among Malayo-Polynesian languages and are not reflected by most languages of the Philippines (Donohue 2007). The absolute final position of demonstratives (following the relative clause) is typologically unusual but common to languages of Indonesia and possibly an effect of Malay contact.[9] Note also that there are two positions for numeral-classifier constituent, one before the head noun and one after. In Yakan, the choice between these positions depends on the definiteness of the entire NP. As seen in (28), prenominal numerals correspond to indefinite interpretations and postnominal ones with definite interpretations.

Yakan (Brainard & Behrens 2002: 30-31)

(28) a. *ŋite ku lime manuk*
AV:see 1SG.NOM five chicken
'I saw five chickens.'

b. *kite-ku (meʔ) manuk-in lime*
see-1SG.GEN PL chicken-DEF five
'I saw the five chickens.'

Other SB languages also show multiple positions, as in (29) (see also Verheijen 1986:20), but the semantic correlates, if any, have not been described.

West Coast Bajau

(29)a. *duo em-bua' belud oyo*
two CNT-CLF hill large
'two large hills'

b. *enselan di-kau' tin*
gasoline one-CLF can
'one can of gasoline' (Miller 2007:313)

[8] As pointed out in Adelaar (2005), a similar construction exists in Salako. This constellation of properties is so unusual that it is unlikely to have arisen multiple times independently and may be another piece of evidence linking the SB languages to Borneo (Kaufman 2007:629).

[9] Pallesen (1985:180) observes that the strictly post-nominal order of demonstratives and possessors in Sama languages has entered Tausug, a Central Philippine language, via contact, while its closest relatives in the Eastern Mindanaoan subgroup show the same flexibility found in Tagalog.

SB classifier systems are relatively simple. Eastern Indonesia Bajau employs *kau* for both animate and inanimate objects of various types. West Coast Bajau uses *-aŋan* for people, *-kauʔ* for animals and non-round objects, and *-buaʔ* for fruits, round objects, and very large objects (Miller 2007:109). Some SB languages of the Philippines, like Yakan, only have a vestigial use of classifiers.

SB languages only show a remnant of the linker found ubiquitously in Philippine languages which mediates between modifiers and their complements. This is almost always restricted to numeral modifiers in SB languages. The general loss of the linker and presence of classifiers puts the SB languages closer to Austronesian languages outside of the Philippines typologically.

5.1 Relative clauses

Only pivots can be relativized in SB languages without a resumptive pronoun. Relative clauses follow the head noun but are introduced in a variety of ways. West Coast Bajau allows relative clauses to follow the head noun without the mediation of any overt functor, as seen in (30).

West Coast Bajau

(30)a. *uwaʔ nguma e pan beranti.*
dog bark DEM TOP stop
'The dog that was barking (or 'the barking dog') stopped.'

b. *enselan Ø-boo Azam kemuap e*
gasoline UV-bring Azam afternoon DEM
'the gasoline that Azam bought yesterday' (Miller 2007:392-393)

Some SB languages of the Philippines make use of *ya* or *iya*, a relativizer derived from the third singular pronoun. In Southern Sinama, we find relative clauses introduced with a combination of *ya* and the linker *na*, parallel to the etymology of Malay *yaŋ* (Adelaar 1992).[10]

Southern Sinama

(31) *si Ben* ***ya*** ***na*** *bey nengge*
PM Ben NMLZ LNK PFV AV:stand
'Ben is who stood.' (Trick 2008:191)

Intriguingly, Yakan employs a prefix *ma-* specifically for agent oriented relative clauses, as seen in (32), but not for other types of relatives. This use of *ma-* does not exist in surrounding languages (but see Klamer 1998:316-334 for a parallel in Kambera).

Yakan

(32)a. *Iyan sawe* ***ma-****pa-diyalem luma-in*
that snake ACT.RELT-VRB-inside house-DEF
'That is the snake that went into the house'

[10] James (2017:75) notes that the use of *ya* in Sinama Dilaut is obligatory with resumptive pronouns but not typically found with the gap strategy. Several authors note variation in this area among the SB languages of Sulu.

b. *Iyan* *naknak* ***ma**-molong* *buwaʔbuwaʔ-in*
that child ACT.RELT-AV:break toy-DEF
'That is the child who broke a toy.' (Brainard & Behrens 2002:165-166)

6. Conclusion: Internal diversity, contact and convergence

Just like the Roma languages of Europe and Asia, the SB languages provide a unique view into language contact across a family of dispersed communities in various stages of sedentarization. Recent descriptive grammars and dictionaries have shed much light on several SB languages of the Philippines but Pallesen's 1985 study remains the only in depth investigation of language contact in the SB family. As this work focused primarily on the bidirectional contact effects between Tausug and the SB languages of the Sulu area, there still remain large gaps in our understanding of SB languages of Indonesia, Malaysia and the central Philippines. Certain SB languages, in particular, beg for attention. Abaknon has barely been described in published work and, next to Chavacano, is the closest thing to a true mixed language (in the sense of Bakker 1997) in the Philippines. The diverse influences are plain to see in the typical Abaknon utterance shown in (33), where native morphemes are in plain typeface, morphemes with a Central Philippine origin are in bold, English-origin morphemes are underlined and Spanish-origin elements are italicized. (Note that *baligyaʔ* may have its ultimate source in a language of South Asia, although its immediate source is a Central Philippine language.)

Abaknon (Jacobson 1999)

(33) I=**maŋa**=ismaglir *pirmi* hamok **ag-tago** si **maŋa baligyaʔ**-na **kon** niyaʔ *sundalo*
NOM=PL=smuggler always only VRB-hide OBL PL goods-3S.GEN if EXT soldier
'The smugglers always hide their goods for sale whenever there is a soldier.'

The SB languages are clearly a fruitful area for further study, both for Austronesianists and those interested in language contact more generally. As discussed earlier, the SB languages also possess syntactic puzzles that provide interesting evidence of grammaticalization and that may shed additional light on the Bornean origins of the SB subgroup.

References

Adelaar, K. Alexander (1989). 'Malay influence on Malagasy: linguistic and cultural-historical implications'. *Oceanic Linguistics* 28(1): 1-46.

Adelaar, K. Alexander (1992). *Proto-Malayic: the reconstruction of its morphology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

Adelaar, K. Alexander (2005). *Salako or Badamea: sketch grammar, text and lexicon of a Kanayatn dialect in West Borneo*. Wiesbaden: Harrassowitz.

Akamine, Jun (1996). 'A Grammatical analysis of Manuk Mangkaw Sinama.' PhD thesis, University of the Philippines, Diliman.

Akamine, Jun (2002). 'The Sinama derived transitive construction'. In Fay Wouk and Malcolm Ross, *The History and Typology of Western Austronesian Voice Systems*. Canberra: Pacific Linguistics, 355- 366.

Akamine, Jun (2003). *A Basic Grammar of Southern Sinama*. Osaka Gakuin University: Endangered Languages of the Pacific Rim, Faculty of Informatics.

Akamine, Jun (2005). 'Sama (Bajau)'. In K. Alexander Adelaar and Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*. London: Routledge, 377-396.
Anderbeck, Karl (2012). 'The Malayic-speaking Orang Laut Dialects and directions for research'. *Wacana* 14(2): 265–312.
Bakker, Peter (1997). *A Language of Our Own: The Genesis of Michif, the Mixed Cree-French Language of the Canadian Métis*. New York and Oxford: Oxford University Press.
Behrens, Dietlinde (2002). *Yakan-English Dictionary*. Manila: Linguistic Society of the Philippines.
Billings, Loren and Daniel Kaufman (2004). 'Towards a typology of Austronesian pronominal clisis'. In Paul Law (ed.), *Proceedings of AFLA 11 (ZAS Papers in Linguistics 34).* Berlin: Zentrum für Allgemeine Sprachwissenschaft, 15–29.
Blust, Robert (2004). 'Austronesian nasal substitution: a survey'. *Oceanic Linguistics* 43(1): 73-148.
Blust, Robert (2007). The linguistic position of Sama-Bajaw. *Studies in Philippine Languages and Cultures* 15: 73-114.
Blust, Robert and Steve Trussel (ongoing). Austronesian Comparative Dictionary. Available online at www.trussel2.com/ACD
Brainard, Sherri and Dietlinde Behrens (2002). *A Grammar of Yakan*. Manila: Linguistic Society of the Philippines.
Cumming, Susanna (1991). *Functional Change: the Case of Malay Constituent Order*. Berlin and New York: Mouton de Gruyter.
Diment, Eunice (1995). Bāngingi Sama: introduction and wordlist. In Darrell T. Tryon (ed.), *Comparative Austronesian dictionary: an introduction to Austronesian studies, part 1: fascicle 1*. Berlin and New York: Mouton de Gruyter, 375-380.
Diment, Eunice and JoAnn Marie Gault (1980). *Manga Bissara: Sama Bangingì-Filipino-English*. Manila: SIL.
Donohue, Mark (1996). 'Bajau: A symmetrical Austronesian language'. *Language* 72: 782–793.
Donohue, Mark (2007). 'Word order in Austronesian from north to south and west to east'. *Linguistic Typology* 11: 349-391.
Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.) (2020). *Ethnologue: Languages of the World.* Twenty-third edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.
Foley, William A. (2007). 'The place of Philippine languages in a typology of voice systems'. In Simon Austin and Peter Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*. Stanford, CA: CSLI Publications, 22-44.
Gault, JoAnn Marie (1999). *An ergative description of Sama Bangingi'.* Manila: Linguistic Society of the Philippines.
Haaksma, Rémy (1933). *Inleiding tot de studie der vervoegde vormen in de Indonesische talen*. Leiden: Brill.
Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank (2020). *Glottolog 4.2.1*. Online resource https://glottolog.org/
Hashim, Sulfilix and Virginia Collins and Millard Collins (2001). *Mapun-English dictionary*. Manila: SIL
Himmelmann, Nikolaus P. (2005). 'The Austronesian languages of Asia and Madagascar: Typological characteristics'. In K. Alexander Adelaar and Nikolaus P. Himmelmann

(eds.), *The Austronesian languages of Asia and Madagascar*. New York: Routledge, 110–181.
Hogan, David W. (1988). *Urak Lawoi': Basic structures and a dictionary*. Canberra: Pacific Linguistics.
Hogan, David (1999). *Urak Lawoi'*. Languages of the World/Materials, 268. München: Lincom.
Jacobson, Marc (1999). Inabaknon wordlist. Available at: https://www.trussel2.com/acd/acd-inab-a.htm
James, Jeremiah Joy. 2017. 'Central Sinama Voice: A Symmetrical Analysis.' MA thesis, Graduate Institute of Applied Linguistics.
Kaufman, Daniel. 2007. 'Salako or Badameà: Sketch Grammar, Texts and Lexicon of a Kanayatn Dialect in West Borneo (review)'. *Oceanic Linguistics* 46(2): 624-633.
Kikusawa, Ritsuko (2003). 'The development of some Indonesian pronominal systems'. In Barry J. Blake, Kate Burridge, and Jo Taylor (eds.), *Historical Linguistics 2001. Selected papers from the 15th International Conference on Historical Linguistics*. Amsterdam and Philadelphia: John Benjamins, 237-269.
Klamer, Marian (1998). *A Grammar of Kambera*. Berlin and New York: Mouton de Gruyter.
Larish, Michael David (1999). 'The Position of Moken and Moklen Within the Austronesian Language Family (Thailand)'. PhD thesis, University of Hawai'i at Mānoa.
Mead, David and Muyung-young Lee (2007). 'Mapping Indonesian Bajau Communities in Sulawesi. Summer Institute of Linguistics (SIL)', Electronic Survey Reports 2007-019, http://www.sil.org/silesr/abstract.asp?ref=2007- 09, accessed 05 July 2020.
Miller, Mark T. (2007). 'A grammar of West Coast Bajau (Malaysia)'. PhD thesis, University of Texas at Arlington.
Pallesen, A. Kemp (1979). 'The pepet in Sama-Bajaw'. In Nguyen Dang Liem (ed.) *South East Asian Linguistic Studies, Vol. 3*. Canberra: Pacific Linguistics, 115-142.
Pallesen, A. Kemp (1985). *Culture Contact and Language Convergence*. Manila: Linguistic Society of the Philippines.
Pittayaporn, Pittayawat (2005). 'Moken as a Mainland Southeast Asian Language'. In Anthony Grant and Paul Sidwell (eds.), *Chamic and Beyond*. Canberra: Pacific Linguistics, 189-209.
Ross, Malcolm (2002). 'Final words: research themes in the history and typology of western Austronesian languages'. In Fay Wouk and Malcolm Ross (eds.), *The History and Typology of Western Austronesian Voice Systems*. Canberra: Pacific Linguistics, 451-474.
Schachter, Paul (1995). The subject in Tagalog : still none of the above. *UCLA Occasional Papers in Linguistics* 15.
Sopher, David Edward (1965). *The sea nomads: a study based on the literature of the maritime boat people of Southeast Asia*. (Memoirs of the National Museum 5). Singapore: Lim Bian Han, Govt. printer.
Trick, Douglas (2008). Ergative control of syntactic processes in Southern Sinama. *Studies in Philippine Languages and Cultures* 19: 184-201.
Verheijen, Jilis A. J. (1986). *The Sama/Bajau language in the Lesser Sunda Islands*. Canberra: Pacific Linguistics.
Walton, Janice R. and Charles Walton (1992). *English-Pangutaran Sama dictionary*. Manila: Summer Institute of Linguistics.
Walton, Charles (1986). *Sama verbal semantics: Classification, derivation and inflection*. Manila: Linguistic Society of the Philippines.

APPENDIX 1. ABBREVIATIONS

ACT.RELT – active relativizer
APPL - applicative
ADV – adverbializer
AGT – agent marker
ASP – aspect marker
AV – actor voice
BV – benefactive voice
CAUS – causative
CNT – counting prefix
DEF – definite
DEM – demonstrative
EXT – existential
GEN – genitive case
IMP - imperative
IV – instrumental voice
LNK - linker
LOC - locative
LV – locative voice
NEG - negative
NOM – nominative case
NMLZ – nominalizer
OBL – oblique case
PASS – passive
PFV – perfective
PL – plural marker
PM – personal marker
POL - polite
POT – potentive,
PREP - preposition
STA - stative
TOP – topic marker
UV – undergoer voice
VRB – verbalizer

APPENDIX 2. LANGUAGES CITED

Abaknon/Inabaknon [abx]
Central Sama/Sinama [sml]
Eastern Indonesian Bajau [bdl]
Manuk Mangkaw Sinama [ssb]
Mapun [sjm]

Mola Bajau [bdl]
Pangutaran Sama [slm]
Sama Balangingi/Bangingi’ [sse]
Southern Sama [ssb]
West Coast Bajau [bdr]
Yakan [yka]