Reevaluating two schwa-initial reconstructions in Proto- and Pre-Proto-Austronesian numerals with some help from Kra-Dai

Alexander D. Smith
Fudan University

Proto-Austronesian numeral reconstruction typically includes the reconstructions *əsa 'one' and *ənəm 'six'. These lexemes are noteworthy because they contain the only examples of schwa in word-initial position in a Proto-Austronesian reconstruction as presented in the *Austronesian Comparative Dictionary*. In this study, both *əsa and *ənəm are critically evaluated with the hypothesis that Proto-Austronesian descended from an ancestor which did not, in fact, contain schwa-initial words and that these examples arose through historical processes that involve the deletion of a word-initial *h which was present in pre-Proto-Austronesian. Evidence from Kra-Dai suggests that this is true for *ənəm 'six', where Kra-Dai evidence suggests Proto-Kra-Dai *xənəm 'six'. In the development of Proto-Austronesian, *x reduced to *h followed by deletion in Proto-Austronesian. Furthermore, it is shown that the putative reconstruction *əsa 'one' relies on evidence from an entry in Ferrell's (1982) Paiwan dictionary, which, under closer inspection, turns out to not be a valid Paiwan word. This, in turn, means that *əsa cannot be reconstructed to Proto-Austronesian due to a lack of Formosan evidence but must instead be a Proto-Malayo-Polynesian innovation. The only well-attested Proto-Austronesian reconstruction for 'one' is the doublet pair *isa/*asa.

Keywords: Proto-Austronesian, Kra-Dai, comparative linguistics, phonology, numerals

1. Introduction

With some exceptions (for example, Sagart 2004; 2013), there has long been a general consensus in Austronesian (AN) scholarship on the shape of Proto-

https://doi.org/10.1075/lali.00245.smi | Published online: 5 August 2025
Language and Linguistics 26:4 (2025), pp. 673–694. ISSN 1606-822X | E-ISSN 2309-5067

Austronesian (PAn) numerals, which are listed in the first table. There is little in the shape of these lexemes which may suggest that they contain any special or unusual combinations of phonemes. They all adhere to the expectation that PAn content words were, by a large majority, disyllabic (Chrétien 1965). They also all follow the PAn syllable-shape constraints, which disallow consonant clusters except those in reduplicated monosyllables (Dempwolff 1937; Ross 1992; Blust 2013). Closer inspection, however, reveals a conspicuous pattern: in virtually all other semantic categories the vowel schwa is not allowed to appear in word-initial position, but here we have two words with such a shape: PAn *əsa 'one' and *ənəm 'six'. Indeed, these are the *only* two words with a schwa-initial reconstruction to PAn according to the *Austronesian Comparative Dictionary* (henceforth *ACD*, Blust et al. 2023).[1] These, and other, restrictions on PAn schwa are organized in more detail by Smith (2023), who proposes a weight-based (or mora-based) explanation for schwa's peculiar behavior in Austronesian languages. The numerals are shown below in Table 1 as they appear in the *ACD*.

Table 1. PAn Numerals 1–10 (Blust et al. 2023)

PAn 1–5	PAn 6–10
*əsa, *isa, *asa 'one'	*ənəm 'six'
*duSa 'two'	*pitu 'seven'
*təlu 'three'	*walu 'eight'
*Səpat 'four'	*Siwa 'nine'
*lima 'five'	*puluq 'ten'

1. A reviewer notes that some apparently vowel-initial words may have a glottal stop or that PAn itself may have had an initial glottal stop in some apparently vowel-initial words. To quote the reviewer, "Glottal stops are not written word-initially in many Formosan and Philippine languages, which may give the illusion of vowel-initial words" and also "A number of languages…distinguish words with an initial glottal stop from true vowel-initial words (e.g. Hla'alua…and Paiwan)". In Hla'alua (Saaroa in the *ACD*) a distinction between glottal-initial and vowel-initial words follows from a historical distinction between *q-initial and vowel-initial words, since PAn *q became Hla'alua glottal stop (see further examples from Li (2004) such as *ʔavuʔu* 'ash' from PAn *qabu and *ʔaciʔi* 'liver' from PAn *qaCay). The same is true for Paiwan. Although not all dialects have a glottal stop from *q, Chang (2006) does indicate a glottal stop in Northern Paiwan, which is also from *q (*ʔuma* 'field' from PAn *qumah). No consensus has been reached on the presence or absence of a (phonemic) glottal stop in word-initial position at the PAn level. Furthermore, most discussion of the possibility of a PAn *ʔ is restricted to non-initial positions (see Zorc (1982; 1996) for work in support of a glottal stop reconstruction). The present study therefore does not treat reconstructed vowel-initial words as possibly having a hidden glottal stop. All cases of glottal-initial/vowel-initial distinctions in Austronesian languages are the result of sound change that has reduced a historically non-glottal consonant to glottal stop. None are retentions.

Given the rarity of Schwa-initial reconstructions, their presence in these two numeral terms begs the question of their ultimate origins. The evidence for both *əsa and *ənəm is robust, at least at first glance. In Table 2, reflexes of *əsa and *ənəm are shown.[2] Note that *əsa is accompanied by two doublets, *isa and *asa, as reconstructions for 'one'. Those are discussed later in § 4.2. Note also that there is only a single Formosan language, Paiwan, providing non-Malayo-Polynesian evidence for PAn *əsa.[3]

Table 2. PAn reconstructions *əsa 'one' and *ənəm 'six' with selected supporting evidence

Language	*əsa reflex	Language	*ənəm reflex
Paiwan	*əta*	Puyuma	*ənəm*
Cebuano	*usa*	Paiwan	*ənəm*
Malay	*əsa*	Cebuano	*unum*
Sangir	*əsa*	Malay	*ənam*

The examples in Table 2 all point to *əsa and *ənəm, but some cases of the word 'six' in Malayo-Polynesian may appear at first glance to reflect an initial *h-*. These are Itbayaten *haʔnəm* 'six, of animals and humans', Manobo *heʔenem* 'six', and Dampelas *honoŋ* 'six'. Itbayaten *haʔnəm* is morphologically complex. The cardinal numeral 'six' is Itbayaten *aʔnəm*, and the initial *h-* in Itbayaten *haʔnəm* follows from a Proto-Malayo-Polynesian (PMP) prefix for counting humans and

2. The *ACD* does not adequately distinguish between dialects in much of its Formosan data. Language names listed here match those listed in the *ACD*, including Saaroa, which may also be written as Hla'alua. However, in the discussion that follows much additional data is taken from sources that do adequately record dialects, and in those cases the proper dialects are listed. All data without direct citation are from the *ACD*.

3. Blust (1999) proposes nine Formosan primary branches, with Paiwan occupying one of those branches. Others, for example Ross (2009), reconstruct less first-order diversity in Taiwan. Ross's subgrouping assumes that PAn diversified into four primary divisions, Rukai, Puyuma, Tsou, and Nuclear Austronesian with Paiwan falling under Nuclear Austronesian (see also Sagart 2004 and Aldridge 2016 for other subgrouping proposals). Under Ross's proposal, *əsa is not reconstructable to PAn on subgrouping grounds, since it is not found in multiple primary branches. This paper assumes Blust's proposal, so it is necessary to show precisely why *əsa* in Paiwan is not an actual reflex of a putative PAn *əsa under the assumptions made here. Those who assume Ross's proposal or proposals like it will likely find the argument against Paiwan *əsa* beside the point. Note, however, that this paper still demonstrates the absence of this word from any Formosan language, restricting it to Malayo-Polynesian. This is true no matter which subgrouping one assumes.

living creatures. The same morpheme is found in the other Itbayaten numerals, for example, *atlo* 'three, in counting and of inanimate objects' vs *hatlo* 'three, of animals and humans' and *aʔpat* 'four, in counting and of inanimate objects' vs *haʔpat* 'four, of animals and humans' (Yamada 2002). It is also clearly reflected in on other Philippine languages, for example Binukid *ha-epat* and Dampelas *h-apat*. Manobo *heʔenem* and Dampelas *honoŋ* are analyzed as containing the same prefix. These words are therefore not to be used to reconstruct an initial *h- on 'six', especially considering the shape of Itbayaten and Manobo which clearly involve morphology that attaches to a typical disyllabic root. The glottal stop in these words is part of an automatic process in nearly all Philippine languages which separates bases from affixes with a glottal stop if no consonant is present and in Itbayaten all schwa-initial words, not just 'six', are reflected with *aʔ-* at the beginning (Blust 2017).[4]

In the remainder of this paper, I look to evidence from outside the Austronesian family to explain how schwa-initial lexemes may have arisen in a language that otherwise banned such word-shapes. The evidence is from Kra-Dai (KD), which, by some hypotheses, is a sister to Austronesian (the Austro-Tai Hypothesis, see Schmidt 1906; Benedict 1942; Ostapirat 2005; Smith 2022), a subgroup within Austronesian (Sagart 2004), or an unrelated family with an ancient period of contact with PAn or pre-PAn speakers (Thurgood 1994).[5] Although I have written in support of the Austro-Tai hypothesis, the exact nature of the Austronesian-Kra-Dai relationship is not central to the claims made in this paper. Regardless of whether Kra-Dai and Austronesian are sisters or unrelated groups that underwent a period of extensive contact, the evidence suggests that at its earliest reconstructable state the word for 'six' was *xənəm and that the initial *x- underwent reduction to *h- and eventually deleted as the language developed in PAn. This, in turn, reinforces the view that an ancestor to Austronesian languages banned schwa from word-initial position.

A second consequence of the present research is the removal of *əsa as a valid PAn reconstruction, although it remains valid as a PMP reconstruction. PAn *əsa has only one Formosan witness, Paiwan *eta* [əta], which is listed in the *ACD* and sourced from Ferrell's (1982) *Paiwan Dictionary*. It is shown that the dictionary entry in Ferrell (1982) is a hypothetical, not actual, root listed in accordance with Ferrell's standards for cross-listing entries which includes hypothetical and unat-

4. This also supports the interpretation that *h* in this word is not from *hənəm but rather from an affix. Words with a schwa penult in PMP underwent metathesis in Itbayaten, whereby Cə- become aC-. *hənəm would therefore be reflected as *ahnəm* if this were considered an inherited *h*.

5. The hypothesis that Austronesian and Kra-Dai are sisters is referred to as the Austro-Tai Hypothesis. The ancestor to both families under this hypothesis is referred to as Proto-Austro-Tai (PAT).

tested roots. Because Ferrell's dictionary was the main reference used in the *ACD*, the inclusion of Paiwan *eta* resulted in *əsa being mistakenly reconstructed to PAn. In reality, only *isa and *asa stand as valid PAn reconstructions with the meaning 'one'.

2. Justifying the use of Kra-Dai

As mentioned earlier, there are three main positions regarding the shared history of Austronesian and Kra-Dai: (1) They are sister groups; (2) Kra-Dai is an Austronesian subgroup; and (3) they two are unrelated. One may argue that, because the status of the relationship between these two families is unresolved, using Kra-Dai evidence for an argument about Austronesian historical phonology is a poor methodological decision. It is not the case, however, that the status of the Austro-Tai hypothesis has bearing on the arguments made in this paper. If, for example, the Austro-Tai hypothesis were widely accepted, then Kra-Dai evidence may be used to reconstruct putative Proto-Austro-Tai (PAT) vocabulary. In the case of 'six', Kra-Dai witnesses may form critical evidence for reconstructing a PAT word for 'six' with an initial consonant. If, however, we reject the basic premise of the Austro-Tai hypothesis and insist that all shared vocabulary between the two families is the product of language contact which took place in South China before Austronesian speaking people migrated to Taiwan, then the Kra-Dai evidence may still be useful.

It is not uncommon for borrowed vocabulary to contain evidence for linguistic structures which have since been lost in the donor language. Campbell (2020) demonstrates that Finnish borrowings from Germanic contain evidence for a contrast between *e and *i before *n, a contrast that has been neutralized in all Germanic languages via the change e > i / _n. Finnish, for example, borrowed the word *rengas* 'ring' from a Proto-Germanic source. The borrowing maintains a mid-vowel before *n*, an environment that cannot be reconstructed with Germanic evidence alone (English, Dutch, German, all have *ring*, Icelandic has *hringur*, and Norwegian *ringe*). In much the same way that Finnish maintains evidence for a distinction that is now lost in Germanic, Kra-Dai may contain evidence for distinctions now lost in Austronesian, even if the evidence is the product of borrowing.

The status of the Austro-Tai hypothesis therefore has no bearing on the usefulness of Kra-Dai vocabulary in this study. The two families have a shared history which originates from modern-day southern China and this alone is enough for researchers to utilize Kra-Dai evidence in arguments of Austronesian linguistic prehistory.

3. PAn *ənəm and the initial *h hypothesis

It has been established that schwa was restricted in PAn such that it appears in word-initial position in only two words, *əsa 'one' and *ənəm 'six'. The exceptionality of the shape of *əsa and *ənəm is inexplicable since these are the only two well-supported reconstructions with a word-initial schwa in PAn in the *ACD*. One may posit that initial schwa was more common in PAn but deleted irregularly in many words thereby giving a false sense of rarity. This scenario is unlikely, however, since most languages have *expanded* the allowable position of schwa in lexemes to include additional schwa-initial words, for example, PMP *əpat < PAn *Səpat. What appears to have happened, rather, is that a total ban on schwa-initial words was gradually relaxed over time as Austronesian languages developed. One of the main pathways that schwa-initial words expanded in early Austronesian was through the irregular deletion of *h where it appeared in word-initial position, especially during the development of PMP. In this section, a review of the behavior of *h- is given, wherein it is shown that *h- irregularly deleted both in PMP and in Formosan languages in a number of lexemes. Then, evidence from Kra-Dai is analyzed which indicates that there was an initial consonant, *x-, in word initial position in Kra-Dai reflexes of 'six'. From this it is concluded *xənəm was reduced to *hənəm and eventually to *ənəm via deletion.

3.1 *h-initial reconstructions and their reflexes

Blust (2018) has pointed out the tendency for laryngeal consonants to have irregular reflexes in daughter languages throughout the Austronesian family.[6] In this earlier study, it was shown that some irregular patterns suggest the existence of an additional phoneme, *x, which he reconstructed to the final-syllable of some PAn words.[7] Blust focused his attention on reflexes of *S and *h in word-final posi-

---

6. Other studies have similarly tried to incorporate irregularities in laryngeal in reconstructions of PAn phonology. Examples include Tsuchida's (1976) *s, *θ, *S_1, *S_2, and *H and Ross's (2015) *s, *θ, *x, *S, and *h. The overall picture of PAn voiceless fricatives is mixed and no clear consensus exists on the quality and quantity of voiceless fricative contrasts. Following Blust (2018), this is thus interpreted here as a consequence of h-instability, not of multiple laryngeal contrasts.

7. Note that although Blust found evidence for a PAn phoneme *x from his analysis of *h in word-final position in numerous Austronesian languages, it is not immediately clear that Blust's *-x and my *x- were the same phoneme. Although it is likely that Blust's *-x and the *x- which reduced to *h- and eventually deleted are related, there is no Austronesian evidence that *x- persisted in word initial position at the PAn level. In addition, Blust's *x is not necessarily a velar fricative. Although he does reconstruct separate fricatives *h and *x in word-final position, there

tion, but in word-initial position, the situation is by no means regular. Numerous examples can be found of an *h irregularly deleting in word-initial position, many of which have been known for some time. A well-known example is the irregular deletion of h- from *S- in Malayo-Polynesian reflexes of PAn *Səpat 'four' (Blust 1995). Not only do *h-initial lexemes often undergo irregular deletion of *h-, there are few *h-initial reconstructions which date back to PAn in the *ACD*. The best-supported such reconstruction is of PAn *huRaC 'artery; blood vessel; nerve; tendon' which is reconstructed with an initial *h- with evidence from Pazeh *huhas* 'vein, blood vessel'. Other reconstructions are less widely attested than *huRaC, but include PAn *həsək 'to drive in stakes or posts' with evidence from Amis *həcek* 'post, pillar', PAn *hisəp 'to suck; inhale' with evidence from Amis *hicep* 'to slurp water', and PAn *hukaq 'to loosen; open' again with evidence from Amis *hoka?* 'loosen something that was tight, as a belt'.

These four reconstructions, *huRaC, *həsək, *hukaq, and *hisəp represent the totality of PAn *h-initial reconstruction in the *ACD*. What is most relevant for the present argument is that all but *hukaq appear to have lost the initial *h- in PMP. For example, the PMP reconstructions *uRat, *əsək, *isəp show *h- deletion, but PMP *hukaq does not. The irregular deletion of *h- in PMP is likely related to the irregular deletion of *S from PAn *Səpat, as already mentioned. Since *S typically merged with *h in PMP as PMP *h, and it is known that *h- irregularly deleted in PMP, then it is reasonable to assume that *S- first became PMP *h-, and then some of the *h-initial words underwent further deletion of *h-. *Səpat, therefore, is assumed to have gone through intermediate changes as follows: *Səpat > *həpat > *əpat. In Table 3, all cases of where PAn *h- or *S- irregularly delete in PMP are listed, as well as two examples of the regular development where *h and *S merge as PMP *h.

It is not only PMP, however, where word-initial *h was irregularly deleted. As mentioned earlier, PAn *huRaC is reconstructed with an initial *h based on Pazeh *huhas* and *həsək, *hukaq, and *hisəp are all reconstructed with evidence from Amis. This indicates that the typical Amis reflex of PAn *h is *h*, at least in initial position. Amis does have a reflex of *huRaC, but this reflex does *not* contain an initial /h/: Amis *olat*, which may indicate irregular *h- deletion in this word in Amis. In other cases, we find that *S- sometimes underwent irregular reduction in Formosan languages to /h/, which sometimes led to deletion. The initial *S in reflexes of *Sapuy, for example, was deleted in Thao *apuy* even though *S-

is no phonetic evidence to suggest that *x was strictly velar. This is an issue that will need to be addressed in more detail in future research since this paper is principally concerned with the fact that schwa was not allowed in word-initial position in pre-PAn and not with the possible addition of a new PAn phoneme *x.

Table 3. Reflexes of *h and *s in PMP

Change	PAn	PMP	Gloss
	*huRaC	*uRat	artery; blood vessel; nerve; tendon
	*həsək	*əsək	to drive in stakes or posts
	*hisəp	*isəp	to suck; inhale
*S-, *h- > *h > Ø	*Səpat	*əpat	four
	*Si-	*i-	instrumental voice prefix
	*Sika-	*ika-	ordinal number prefix
	*Sipəs	*ipəs	cockroach
*S-, *h- > *h-	*hukaq	*hukaq	to loosen; open
	*Sapuy	*hapuy	fire

is typically reflected as Thao *sh*. A similar irregular development also occurs in Saisiyat *hapoy* where *S- became *h-* but is normally retained as Saisiyat *s-* and in Proto-Seediq *hapuy (Ochiai 2022).[8] For comparison, *Sanaq 'otter' became Thao *shanaq* and *Sajək 'smell' became Saisiyat *s-om-azək*.

Austronesianists trying to make sense of these data find themselves in a difficult situation, since PMP *h from both PAn *h and *S may irregularly delete in word-initial position, making accurate reconstruction difficult. This is not much of a problem for words that reflect *S because Formosan languages will typically provide evidence for accurate reconstruction to PAn. However, with so few Formosan languages retaining *h in initial position it is particularly difficult to make any sort of inference on PMP words that may reflect a PAn word-initial *h-, since corresponding Formosan evidence will rarely provide insights. Referring back to the problem of schwa-initial reconstructions for the numerals *əsa and *ənəm, there appears to be no family-internal evidence which suggests a similar history; that is, a word-initial *h- which deleted in PMP as follows: PAn *hənəm > PMP *ənəm. Where Austronesian-internal evidence fails, however, evidence from Kra-Dai may be utilized to fill in gaps such as that concerning the history of 'six'. Such Kra-Dai evidence is discussed in more detail next.

8. This Seediq example was brought to my attention by an anonymous reviewer, who also notes that *hapuy may be further reconstructed to Proto-Atayalic, providing yet another case where Atayal irregularly reduced *S- to *h-. The same reviewer also points out that the personal case marker PAn *Si is reflected in Proto-Atayalic as *i-, presumedly having gone through a stage of reduction from *S- to *h-, followed by further irregular deletion (Ochiai 2021). This further demonstrates how common such irregular developments are.

3.2 Kra-Dai reflexes of 'six'

Although Austronesian evidence suggests a word-initial schwa for 'six', evidence from Kra-Dai suggests that the initial syllable had an onset at some point. The evidence comes from the Kra branch, but before the evidence is presented, it is necessary to discuss two points about Kra-Dai numerals and the shared Kra-Dai-Austronesian vocabulary.

First, many Kra-Dai languages have replaced their numerals with Chinese loans. For example, 'six' was replaced with a Chinese loan in Proto-Tai *krok^{D}, Lakkja *lok*^{24}, and Southern Kam *ljok*^{21} (Long & Zheng 1998; Pittayaporn 2009; Fan 2019).[9] Because of this, the overall number of comparisons between Kra-Dai and Austronesian numerals are rather low — they are restricted to Kra and Hlai. Despite this restriction, however, it is still possible to make reconstructions to PKD with the evidence that remains from only Hlai and Kra.

Second, the correspondences between PAn *ə in the final syllable and corresponding vowels in Kra-Dai are irregular. At some point in the shared history of Austronesian and Kra-Dai there was variability in the pronunciation of what is typically reconstructed as *ə. Sometimes, Kra-Dai and Austronesian languages agree on a schwa reconstruction. Other times there is conflicting evidence where some languages reflect a schwa, and others reflect *u (Smith 2022). Table 4 below lists some of the irregular correspondences. Kra-Dai cognates of PAn *dəmdəm 'dark'/'black' and *ŋipən 'tooth', show regular correspondences that reflect PAn and PKD *ə. Kra-Dai cognates of PAn *ənəm 'six' and *datəŋ 'arrive', show cases where Kra-Dai language have irregular correspondences with a PAn *ə. Finally, Kra-Dai cognates of PAn *daNum 'water' and *Rabun 'cloud'/'sky' are irregular plus the Austronesian vowel is *u rather than *ə. The irregularities typically take the shape of a mid-to-high back vowel where a schwa or /a/ is expected. Unexpected correspondences are bolded in the table. Proto-Tai (PT) data are from Pittayaporn (2009), Proto-Hlai (PH) data are from Ostapirat (2004), Proto-Ong-Be (POB) data are from Chen (2018), Proto-Kra (PK) data are from Ostapirat (2000), Lakkja (Lk) data are from Fan (2019), and Southern Kam (S.K) data are from Long & Zheng (1998).

These irregularities in correspondences that involve *ə in the final syllable are also present in Kra-Dai cognates of PAn *ənəm: some reflect *ənəm, others *ənum. Nevertheless, words for 'six' in Kra-Dai still hold important evidence for reconstructing the word-initial phoneme. In the following Example (1), both Hlai and Kra reflect a numeral that may have descended from PAT or was borrowed into PKD at an early stage (note that *hn* = n̥. Data from Ostapirat 2000; 2004).

---

9. Kra-Dai tones are typically organized into A, B, C, and D tones. I keep whichever tone markings are used in the source materials.

Table 4. KD and AN words showing unexpected *u : *ə correspondences

Gloss	Dark	Tooth	Six	Arrive	Water	Cloud
PAN	*dəmdəm	*ŋipən	*ənəm	*datəŋ	***daNum**	***Rabun**
PT	*Ç.dam^{A}	*wan^{A}	–	***C.tɤŋ^{A}**	*C.nam^{C}	***ɓɯn^{A}**
PH	*(ʔ)dam^{C}	*ipan^{A}	***ənum^{A}**	–	*nam^{C}	–
POB	*zam^{A1}	–	–	***ɗəŋ^{A1}**	*nam^{BC2}	–
PK	*dəm^{A}	*l-pən^{A}	*x-nəm^{A}	***m-duŋ^{A}**	–	–
LK	lam^{51}	wan^{2}	–	taŋ^{231}	**num^{11}**	**bən^{51}**
SK	nɐm^{55}	pjɐn^{55}	–	tɐŋ^{55}	nɐm^{31}	**mən^{55}**

(1) a. Proto-Hlai *ənum^{A} 'six': Hlai *tom^{1}* 'six'
b. Proto-Kra *x-nəm^{A} 'six': Gelao *nam^{A1}* 'six', Paha *nam^{A2}* 'six', Pubiao *hnam^{A1}* 'six', Buyang *nam^{A1}* 'six'

In Hlai, all medial nasals are reflected as voiceless oral stops. The reflex *tom^{1}* in Hlai is therefore interpreted as a regular development and cannot be said to have been influenced by the presence or absence of an initial consonant in the Proto-Hlai penult. Kra evidence, on the other hand, points to a Proto-Kra word that began with a consonant, reconstructed as *x- by Ostapirat, not a vowel. The reconstruction is inferred from two observations relating to: (1) the tone of Kra reflexes of 'six'; and (2) the voicing of the non-initial nasal. Inferences are from Ostapirat (2000) and are summarized below.

3.2.1 Tone and initial consonants in Kra

Tones in Kra-Dai are typically split into four classes, A, B, C, and D, which usually correspond to some quality of the reconstructed final-syllable rhyme. These tone classes may be further split to different degrees in different languages and subgroups. In Tai, the A, B, C, and D tones may be themselves split into 4 on conditions relating to the shape of the onset, driven by a loss of voicing and glottalization of onsets (Ostapirat 2000: 53). Tone splits driven by a loss of voicing in onsets are common. Such splits occur, for example, in Chamic languages where voicing distinctions have been lost and replaced by tonal distinctions (Thurgood 1999).

In Kra, onset-driven tonal splits are also present and are triggered by the voice quality of the initial consonant referred to as the "1–2 voicing series" (Ostapirat 2000: 50). In Pubiao, for example, series 1 tones correspond to voiceless and glottalized initials and series 2 tones correspond to voiced and breathy initials. Each series is present in the four base tones, providing a historical analysis of A^{1}, A^{2},

B[1], B[2], and so forth. In all Kra languages, series 1 tones are triggered at least in part by voiceless onsets, and series 2 tones by voiced onsets. In reflexes of 'six' the realization of tone in Gelao, Buyang, and Pubiao have an A[1] tone, conditioned by a voiceless initial /n̥/, even in cases where the voicelessness has been lost due to subsequent sound change, as in Gelao and Buyang. Paha, on the other hand, has an A[2] tone.

The tonal evidence therefore indicates that after the breakup of Proto-Kra, some aspect of the proto-word for 'six' triggered voiceless reflexes of the medial *-n- in many languages. In the next section, the specific conditions for voice-loss in nasals in Kra languages are discussed.

3.2.2 Nasal consonant voicing and initial consonants

Proto-Kra is reconstructed with both a voiced and voiceless nasal series. Ostapirat (2000: 211) hypothesizes that the voiceless nasals "…resulted from preceding onsets…" and singled out *s- as a possible trigger for Proto-Kra voiceless initial nasals but did not make any firm claim about the shape of the onset. In Kra, then, voicelessness arose in stages: pre-Kra *CN-, where C may have been /s/, first became *hN- (a cluster), which then triggered nasal devoicing yielding Proto-Kra *N̥-. Words that followed this path tend to have regular correspondences across Kra languages in both tone and voicing. However, reduction of a putative *s- is not the only path to voicelessness in Kra nasals. Some words in Kra, as is the case with 'six' and others, tone and voicing evidence indicate loss of voicing in Gelao, Pubiao, and Buyang, but a retention of voicing in Paha. In these cases, Ostapirat reconstructs Proto-Kra *x- and notes that *x- causes a split in voicing realization of the nasals, since *x- in some languages was altered to *ɣ- before voicelessness was transferred from the initial to medial consonant. In Kra, reflexes of *x-nam show this split both in voicing realization as well as in tone-split realization.

3.2.3 Comparison between Kra voiceless nasals and other Kra-Dai branches

At least in reflexes of 'six', Kra evidence unambiguously points to the presence of an initial consonant, Proto-Kra *x-. An important question to ask is how 'six', as well as other voiceless nasal-initial words, correspond to cognates in other Kra-Dai branches. Such comparisons might shed light on the reconstruction of the initial consonant. In this section, Proto-Kra reconstructions are compared with Proto-Tai (Pittayaporn 2009), Lakkja (Fan 2019), and Southern Kam (Long & Zheng 1998). In some cases, words that do not have a voiceless nasal in Kra nevertheless correspond to voiceless nasals in other branches. This is due to the influence of an initial consonant, reconstructed as *C- by Ostapirat (2000). Such

words are also included here if they have correspondences in both Tai and Lakkja. Table 5 below contains the comparisons:

Table 5. Comparisons of voiceless initial nasals in KD

PK	PT	Lakkja	Southern Kam	Gloss
*x-mu^{A}	*^{h}mu:^{A}	khũ^{51}	ŋu^{453}	pig
*x-ma^{A}	*^{h}ma:^{A}	khuə̃^{51}	ŋwa^{35}	dog
*x-mət^{D}	*^{h}mat^{D}	khwat^{55}	ŋwɐt^{35}	flea
*hŋwu^{B}	–	khjãu^{51}	–	pus
*C-me^{A}	*^{h}mwɯj^{A}	kũi^{51}	me^{55}	bear
*C-na^{A}	*^{h}na:^{A}	tsã^{51}	na^{55}	thick

Wherever *x- is reconstructed in Kra, it corresponds to a voiceless nasal in Tai, a voiceless non-nasal velar in Lakkja, and a velar nasal in Southern Kam. Reflexes of 'six' are not included since relevant languages have replaced it with a Chinese borrowing. Inherited voiceless velar nasals (without an initial x-), also correspond to velars in Lakkja as evidenced by reflexes of 'pus'. Relevant cognates with 'pus' are unfortunately absent from Tai and Southern Kam. Where voicelessness was triggered by a consonant other than *x, as seen in 'bear' and 'thick', different correspondence sets are seen. With reflexes of 'bear', the labial-velar features of the reconstructed penultimate vowel *u (cognate with PAn *Cumay), caused velarization in Lakkja but not in Southern Kam (Ostapirat 2018; Smith 2022). This shows that the velarization in Lakkja kũi^{51} was not triggered by the initial consonant. Finally, cognates of 'thick' again show that in words without *x-, velarization is not triggered in Lakkja or Southern Kam.

Several observations can be made with these data. First, it is not the case that *any* initial consonant triggered voicelessness after the breakup of Proto-Kra. Only *x- caused such a change, although Proto-Kra itself also had a set of initial voiceless nasals derived from an earlier stage. Second, voiceless nasals in Kra always correspond to a voiceless non-nasal stop in Lakkja. If the nasal was velar, or velarized from *x-, then the corresponding Lakkja consonant is *kh*, and this corresponds to a velar nasal in Southern Kam. Otherwise, the non-nasal stop in Lakkja typically has the same place of articulation as the corresponding voiceless nasal. This is true even in words without Kra cognates, for example, Proto-Tai *^{h}nak^{D} 'heavy': Lakkja tsak^{55}.

From this analysis it is possible to hypothesize that Kra reflexes of 'six' developed under the influence of an initial consonant. This consonant can be reconstructed to PK *x-nam and may ultimately reflect a PKD word of similar shape:

PKD *xənəm. With this information from Kra-Dai it is now possible to reconstruct the history of 'six'. In doing so, it is important to keep in mind the key points made so far: (1) In addition to multiple distributional restrictions in other positions, schwa was highly restricted in PAn word-initial position. According to the *ACD*, only two words may be reconstructed with a schwa-initial: *əsa and *ənəm; (2) There are very few reconstructions of *h in word-initial position, and the reconstructions that we do have suggest irregular deletion of *h from this position in Formosan and Malayo-Polynesian languages; (3) Kra-Dai evidence suggests an *x-initial word, which reduced to *h- and eventually deleted altogether in Austronesian.

Given what is known about both the development of *h- and *ə-initial words in Austronesian languages, it is reasonable to take the Kra-Dai evidence as support for a hypothesis that the pre-PAn word for 'six' was *xənəm. The phonetic value of *x was likely [x], which lead to the velarization of consonants in Lakkja and Southern Kam as shown in Table 5. This word later underwent a process of lenition whereby *x- became *h-, yielding an intermediate numeral *hənəm. This fed the familiar process of irregular *h-deletion as discussed earlier in § 3.1. The reduction of *x- to *h- is a cross-linguistically common sound change so its appearance in pre-PAn is not surprising. Regarding the shape of PAn 'six', one may hypothesize that *h- from earlier *x- was retained as *h- in PAn *hənəm. Under such a hypothesis Amis must have irregularly deleted *h from *hənəm, yielding Central Amis *nnəm* (or *ənəm* when pronounced with emphasis as noted in Amis fieldnotes stored in Kaipuleohone's *Blust collection* (Blust 1994)). This process parallels irregular *h-deletion in Amis' reflex of *huRaC. This must then be followed by an additional irregular deletion of *h in PMP, which parallels the irregular deletion of *h- in so many other words. A competing hypothesis is that *h- from earlier *x- was deleted before PAn, resulting in the familiar reconstruction PAn *ənəm. In this second scenario, a strict ban on schwa-initial words first began to fall apart at the PAn level, as *h- started irregularly deleting, a process which continued for some time thereafter (Blust 2018).

It is tempting to support the first scenario, whereby *hənəm is reconstructed to PAn, because this reduces the number of exceptional schwa-initial reconstructions. However, without any sort of Austronesian-internal evidence for an initial *h it is difficult to justify such a reconstruction. Scenario 2, however, can be readily defended with the evidence considered thus far. If the Austro-Tai hypothesis is accepted, then the development path is from PAT *xənəm to *hənəm, finally yielding PAn *ənəm. If one does not accept the Austro-Tai hypothesis, then one may simply state that at some point in early pre-PAn the word *xənəm was passed on to PKD speaking communities via contact before it was reduced to *hənəm in late pre-PAn and subsequently reduced to *ənəm in PAn.

4. What about *əsa, the other schwa-initial?

PAn *əsa is a more problematic reconstruction, not because of the initial schwa, but because of competing reconstructions, referred to as "doublets" (Blust 2011), without the initial schwa. According to the *ACD*, Paiwan is the only Formosan language listed as containing a reflex of *əsa: Paiwan *eta* (e = [ə]). This is not the only word for 'one' in Paiwan, however. A reflex of the competing form *isa is also reflected by Paiwan *ita*, which has corresponding forms in Kavalan *issa*, and Budai Rukai *iθa*. An additional competing form, *asa, is reflected by at least Taokas *aha* and Saisiyat *ʔæhæ*. Both *əsa and *isa are widely reflected in Malayo-Polynesian. Reflexes of *asa are less common but still attested in Malayo-Polynesian.

4.1 Kra-Dai reflexes of 'one'

Kra-Dai does not provide much assistance in determining the shape of PAn or pre-PAn 'one'. There are even fewer comparisons available for 'one' than there are for 'six', but those that are available suggest that the Kra-Dai languages reflect *isa, not *əsa. In Table 6 below, with data from Ostapirat 2000; 2004), a high-vowel in Hlai and Gelao, and palatalization in Pubiao and Qabiao, suggest earlier *isa, since the presence of a high vowel in the penultimate syllable of PKD reconstructions causes palatalization and high-vowel reflexes in numerous daughter languages (Ostapirat 2018).

Table 6. Evidence for a high-front penultimate vowel in Kra-Dai

Language	Reflex of *isa 'one'
PH	*ciː
Gelao	*tsi*^{55}
Pubiao	*tɕja*^{C1}
Qabiao	*tcia*^{33}

4.2 Paiwan and formosan evidence for PAn *əsa

Because the Kra-Dai evidence points to *isa but not *əsa, and because the only Formosan evidence for *əsa is a form in a language which also contains a reflex of the more common *isa variant, one may question the validity of PAn *əsa altogether. In fact, grammatical and comparative descriptions of Paiwan call into question the listing of Paiwan *eta* on the *ACD*. For example, Chen (2006: 70) states

that schwa is banned from word-initial position in Paiwan, and explicitly rules out [əta] as a possible Paiwan word. Similarly, Chang's (2006) thesis, *A Reference Grammar of Paiwan*, lists *ita* as Paiwan 'one' and *eta* appears nowhere in the document. It would be reasonable to hypothesize that these mismatches are the product of dialectical differences, but this does not seem likely. Chen (2006) utilizes data from all three major Paiwan divisions, Northern, Central, and Southern, and Chang (2006) focuses on Saichia and Santi, both Northern Dialects. The major distinctions in Paiwan dialects are in the consonants, not the vowels (Chen 2006), and both authors indicate where dialectical differences are attested, and neither list dialectical differences for the word 'one'.

To get to the bottom of *əsa, it is necessary to inspect the source for the *ACD* entry for Paiwan *eta* (e = ə), Ferrell (1982). In doing so, it is discovered that an error in PAn reconstruction has occurred due to a misunderstanding of the cross-listing procedures used in the dictionary. On page ix of the dictionary, Ferrell goes into some detail on the cross-listing conventions used in the dictionary, and for good reason — Ferrell regularly lists entries based on hypothetical roots which are unattested in the language, but which may be hypothesized to exist depending on one's morphological analysis. Specifically, Ferrell states that "Certain putative roots are attested only in frozen, complex forms…". He gives the example of *kałay*, which is listed as a separate root in the dictionary, despite only appearing in a complex form *parakałay* 'priest'. The hypothetical root entry is then cross-listed with the attested word, so one may look up *kałay* which will direct the reader to the *parakałay* listing. To reiterate, *kałay* is *not* a root in Paiwan, but is listed as such because one *could* argue that it is a root in a hypothetical complex form *parakałay*.

This is not the only case where Ferrell (1982) lists words that do not exist as head entries. Ferrell also employed this cross-listing method with words where he saw a possible morpheme boundary within a word, even in the absence of actual morphological complexity. Again, on page ix, he states that *valanga* 'mortar' "…may be found listed under three separate entries: (1) *langa* (as if the form were derived from a root **langa* with prefixed *va-*); (2) *vanga* (as if the form were derived from a root **vanga* with an infixed /al/); (3) *valanga* (as a frozen complex form or a trisyllabic root)" (Underlined words from Ferrell are replaced with italics). Here, too, Ferrell's intention was to direct readers to the proper definition of a word, regardless of their analysis or misanalysis of the true shape of the root.

Finally, Ferrell (1982) goes on to state that "Scholars of comparative Austronesian linguistics will recognize that such cross-listing is not meant to imply that each of these multiple entries represents a viable productive root in Paiwan." Ferrell was not being unreasonable with these hypothetical root entries. One can imagine that they might be useful for someone without a good grasp of Paiwan.

However, the decision to list words like *eta* despite no actual evidence that the word exists in Paiwan has inadvertently caused errors to be made in PAn reconstruction.

The discussion may now turn to the issue of PAn *əsa and Paiwan with a review of all numeral entries in Ferrell (1982). On pages 41–43, Ferrell lists all numeral data in a detailed section that includes cardinal, general counting, person counting, living plant counting, frequency, duration, distributional, and ordinal forms of the numerals. *ita* is listed throughout, but *eta* is not listed. On page 90, one finds the entry for *eta*, which appears in its entirety as follows: *eta : ta* (one). According to Ferrell's own description, this form could be a hypothetical root with a cross-listing directing the reader to the full entry for 'one' which is listed under *ta*. Similarly, the entry for *ita* on page 102 also directs the reader to *ta*. Under the entry for *ta*, like in the earlier *numerals* section on pages 41–43, there are numerous entries containing the root *ita*, but none with the putative root *eta*. The hypothetical word *eta* is also not listed in the reverse-entry for 'one' on page 446. Here, the reader is given *ita* and *ta-* as the root forms for 'one'.

All of this, along with the statements from both Chen (2006) and Chang (2006) that schwa is not a permissible word-initial vowel, suggests that the entry *eta* in Ferrell (1982) is a hypothetical root, not an attested root. With that, the only piece of Formosan evidence for PAn *əsa falls apart, leaving just *isa and *asa as well-attested reconstructions for PAn 'one'. Although there is much evidence for PMP *əsa, it is important to not conflate PMP's size with its usefulness in reconstructing PAn vocabulary. No matter how well-attested a PMP reconstruction may be, it requires at least one Formosan cognate for reconstruction to PAn. Only *isa and *asa satisfy this requirement. In Table 7, the Formosan evidence for PAn *isa and *asa are listed, including evidence not listed in the *ACD* (Puyuma (Katipul) and Hoanya from Li (2004) and Tsuchida (1982) respectively). Malayo-Polynesian evidence for *asa is also listed from Motonene, Popalia, and Buli, since *asa has fewer Formosan cognates.[10]

10. A reviewer asks why Bunun *tasʔa* is not included here (see Li (2004) for entries from Takituduh, Takbanuaz, and Iskubun Bunun). These examples are not exhaustive (relevant data sets include hundreds of entries, not appropriate for an article of this length). Some forms, like Bunun *tasʔa* contain irregularities, in this case, an initial *t-* and a medial *-ʔ-*, and are therefore not included in this list which only focuses on a few forms.

Table 7. Selected evidence for PAn *isa 'one' and *asa 'one'

Language	*isa reflex	Language	*asa reflex
Kavalan	*issa*	Taokas	*aha*
Budai Rukai	*iθa*	Saisiyat	*ʔæhæ*
Paiwan	*ita*	Motonene	*me-ʔasa*
Puyuma (Katipul)	*isa*	Popalia	*asa*
Hoanya	*ittaʔ*	Buli	*asa*

4.3 Reflexes of *əsa in Malayo-Polynesian

Although *əsa may not be a valid PAn reconstruction, there is ample evidence in Malayo-Polynesian for PMP *əsa. Some of this evidence is listed below in Table 8. Note that in the *ACD*, languages that lack an initial vowel and have only *sa* or some related forms are also listed as evidence for PMP *əsa. I exclude such evidence here because it is not clear that the deleted vowel was indeed schwa, or if such forms simply reflect the short-form PAn *sa.[11]

Table 8. Selected reflexes of PMP *əsa

Language	*əsa reflex
Bontok	*ʔəsá*
Talaud	*assa*
Cebuano	*usá*
Malay	*əsa* 'unity; one'
Sangir	*əsa*
Tontemboan	*əsa*
Madurese	*əssaʔ*
Palawan Batak	*ʔəsá*

A natural question to ask then is how a new form for 'one', *əsa, came to be. A possible explanation for *əsa is found in the reduced form *sa 'one', which was a proclitic form that attached to the front of words, for example, Paiwan *ta-iday* 'one hundred' and Malay *sə-ratus* 'one hundred'. Although this is only a specula-

11. Gemination in the Talaud reflex is triggered by a past schwa, which has merged with a. Utsumi (2016) contains many examples showing that historical schwa triggers gemination before merging with a, for example, Lirung and Bulude dialects *mamalli* and Pangeran dialect *mammaddi* 'to buy' from PMP *bəli.

tion, PAn *sa may have been expanded from being a clitic to also functioning as a stand-alone word in PMP. Under such a scenario, *sa operated both as a clitic and as a content word in competition with PMP *isa/*asa. Due to the family-wide pressure to maintain a disyllabic minimum in all content words, a schwa was inserted in word-initial position creating *əsa whenever reflexes of *sa were used as independent words rather than clitics. This change may have occurred in PMP, or in parallel developments after the breakup of PMP, but without any viable Formosan evidence, reflexes of *əsa in Malayo-Polynesian cannot have been inherited from a level higher than PMP.

5. Conclusion: Did PAn have any schwa-initial vocabulary?

The conclusion is mixed regarding the question of schwa-initial vocabulary in PAn. Of the two PAn reconstructions with word-initial schwa in the *ACD*, *əsa has been thrown out completely after it was shown that the Paiwan evidence for *əsa is not actually a valid Paiwan word, but a hypothetical root listed in Ferrell (1982). With the removal of *əsa the best PAn reconstruction for 'one' is *isa plus the doublet *asa. Regarding the reconstruction for 'six', however, even though there is good evidence from Kra-Dai to reconstruct PKD and pre-PAn *xənəm 'six', there is no Austronesian-internal evidence to suggest that the initial consonant was retained in PAn. For now, it is assumed that PAn had a single schwa-initial word, *ənəm, which arose through the irregular deletion of word-initial *h as evident in the many other cases of *h-deletion discussed earlier. The pre-PAn numerals are listed below in Table 9, where 'six' is given as *xənəm. As far as PAn is concerned, the reconstructed numerals will be identical to those in pre-PAn, except that *xənəm became *hənəm then *ənəm, and the doublet *asa may be reconstructed for 'one' was likely not present in pre-PAn. In pre-PAn, only *isa is reconstructed for 'one' because it has Kra-Dai support whereas support for *asa is only found in Austronesian.

Table 9. Pre-PAn numerals 1–10

Pre-PAn 1–5	**Pre-PAn 6–10**
*isa 'one' (> PAn *isa, *asa)	*xənəm 'six' (> *hənəm > PAn *ənəm)
*duSa 'two'	*pitu 'seven'
*təlu 'three'	*walu 'eight'
*Səpat 'four'	*Siwa 'nine'
*lima 'five'	*puluq 'ten'

Regarding schwa, this research suggests that schwa was banned from both word-initial and word-final positions in pre-PAn. Put differently, schwa was banned from word boundaries. This total ban was relaxed over time as deletion began targeting words with an initial *h-. As more evidence becomes widely available it has also become evident that PAn treated schwa differently than the other vowels. These restrictions may provide insights into more abstract properties of PAn phonology.

Funding

This research was supported by the Singapore Ministry of Education grant (MOE-T2EP40121-0003).

Acknowledgements

I wish to thank two anonymous reviewers whose comments helped to improve an earlier version of this paper, as well as audience members at the 31st Annual Meeting of the Southeast Asian Linguistics Society who provided early feedback on this research.

List of abbreviations

(P)An	(Proto-)Austronesian
(P)AT	(Proto-)Austro-Tai
(P)H	(Proto-)Hlai
(P)K	(Proto-)Kra
(P)KD	(Proto-)Kra-Dai
(P)MP	(Proto-)Malayo-Polynesian
(P)OB	(Proto-)Ong-Be
(P)T	(Proto-)Tai
Lk	Lakkja
SK	Southern Kam

References

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

Benedict, Paul K. 1942. Thai, Kadai, and Indonesian: A new alignment in Southeastern Asia. *American Anthropologist* 44(4). 576–601.

Blust, Robert. 1994. Blust collection. *Kaipuleohone, University of Hawai'i Digital Language Archive*. Honolulu: Department of Linguistics, University of Hawai'i at Manoa. (https://hdl.handle.net/10125/7735). (Accessed 2025-04-12.)

Blust, Robert. 1995. The position of the Formosan languages: Method and theory in Austronesian comparative linguistics. In Li, Paul Jen-kuei & Tsang, Cheng-hwa & Huang, Ying-kuei & Ho, Dah-an & Tseng, Chiu-yu (eds.), *Austronesian studies relating to Taiwan*, 585–650. Taipei: Institute of History and Philology, Academia Sinica.

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

Blust, Robert. 2011. The problem of doubleting in Austronesian languages. *Oceanic Linguistics* 50(2). 399–457.

Blust, Robert. 2013. *The Austronesian languages*. 2nd edn. Canberra: Pacific Linguistics.

Blust, Robert. 2017. Regular metathesis in Batanic (Northern Philippines)?. *Oceanic Linguistics* 56(2). 491–504.

Blust, Robert. 2018. The "mystery aspirates" in Philippine languages. *Oceanic Linguistics* 57(1). 221–247.

Blust, Robert & Trussel, Stephen & Smith, Alexander D. 2023. *CLDF dataset derived from Blust's "Austronesian comparative dictionary"* (Version 1.2). (Data set.)

Campbell, Lyle. 2020. *Historical linguistics: An introduction*. Edinburgh: Edinburgh University Press.

Chang, Anna Hsiou-chuan. 2006. *A reference grammar of Paiwan*. Canberra: Australian National University. (Doctoral dissertation.)

Chen, Chun-Mei. 2006. *A comparative study on Formosan phonology: Paiwan and Budai Rukai*. Austin: University of Texas at Austin. (Doctoral dissertation.)

Chen, Yen-ling. 2018. *Proto-Ong-Be*. Honolulu: University of Hawai'i at Mānoa. (Doctoral dissertation.)

Chrétien, C. Douglas. 1965. The statistical structure of the proto-Austronesian morph. *Lingua* 14. 243–270.

Dempwolff, Otto. 1937. *Vergleichende lautlehre des Austronesischen wortschatzes*. Berlin: Reimer.

Fan, Wenjia. 2019. *A grammar of Lakkja, South China*. Melbourne: University of Melbourne. (Doctoral dissertation.)

Ferrell, Raleigh. 1982. *Paiwan dictionary*. Canberra: Pacific Linguistics.

Li, Paul Jen-kuei. 2004. Basic vocabulary for Formosan languages and dialects. In Li, Paul Jen-kuei (ed.), *Selected papers on Formosan languages*, vol. 2, 1483–1532. Taipei: Institute of Linguistics, Academia Sinica.

Long, Yaohong & Zheng, Guoqiao. 1998. *The Dong language in Guizhou Province, China*. Dallas & Arlington: The Summer Institute of Linguistics & The University of Texas at Arlington. (Translated from Chinese by Geary, D. Norman.)

Ochiai, Izumi. 2021. A fossilized personal article in Atayal: With a reconstruction of the Proto-Atayalic patronymic system. *Journal of Ainu and Indigenous Studies* 1. 99–120.

Ochiai, Izumi. 2022. Fossilized infixes in Seediq: Identification through dialect comparison. *Journal of Ainu and Indigenous Studies* 2. 1–29.

Ostapirat, Weera. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23(1). 1–251.

Ostapirat, Weera. 2004. Proto-Hlai sound system and lexicons. In Lin, Ying-chin & Hsu, Fang-min & Lee, Chun-chih & Sun, Jackson T.-S. & Yang, Hsiu-fang & Ho, Dah-an (eds.), *Studies on Sino-Tibetan languages: Papers in honor of Professor Hwang-cherng Gong on his seventieth birthday*, 121–175. Taipei: Institute of Linguistics, Academia Sinica.

Ostapirat, Weera. 2005. Kra-Dai and Austronesian: Notes on phonological correspondences and vocabulary. In Sagart, Laurent & Blench, Roger & Sanchez-Mazas, Alicia (eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, 107–131. London: Routledge Curzon.

Ostapirat, Weera. 2018. Reconstruction disyllabic Kra-Dai. (Plenary presentation at the 28th meeting of the Southeast Asian Linguistics Society (SEALS 28), Kaohsiung, 17–19 May 2018.)

Pittayaporn, Pittayawat. 2009. *The phonology of Proto-Tai*. Ithaca: Cornell University. (Doctoral dissertation.)

Ross, Malcolm. 1992. The sound of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31(1). 23–64.

Ross, Malcolm. 2009. Proto-Austronesian verbal morphology: A reappraisal. In Adelaar, Alexander & Pawley, Andrew (eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, 295–326. Canberra: Pacific Linguistics.

Ross, Malcolm. 2015. Some Proto Austronesian coronals reexamined. In Zeitoun, Elizabeth & Teng, Stacy F. & Wu, Joy J. (eds.), *New advances in Formosan linguistics*, 1–38. Canberra: Asia-Pacific Linguistics.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444.

Sagart, Laurent. 2013. The higher phylogeny of Austronesian: A response to Winter. *Oceanic Linguistics* 52(1). 249–254.

Schmidt, Wilhelm. 1906. *Die Mon-Khmer völker: Ein bindeglied zwischen völkern Zentralasiens und Austronesiens*. Braunschweig: Friedrich Vieweg und Sohn.

Smith, Alexander D. 2022. New evidence for Austro-Tai and observations on vowel correspondences. In Alves, Mark J. & Sidwell, Paul (eds.), *Papers from the 30th meeting of the Southeast Asian Linguistics Society (2021)*, 73–95. Honolulu: University of Hawai'i Press.

Smith, Alexander D. 2023. Reconstructing non-contrastive stress in Austronesian and the role of the mora in stress shift, gemination and vowel shift. *Diachronica* 40(1). 111–152.

Thurgood, Graham. 1994. Tai-Kadai and Austronesian: The nature of the historical relationship. *Oceanic Linguistics* 33(2). 345–368.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change, with an appendix of Chamic reconstructions and loanwords*. Honolulu: University of Hawai'i Press.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Research Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.

Tsuchida, Shigeru. 1982. *A comparative vocabulary of Austronesian languages of Sinicized ethnic groups in Taiwan, part I: West Taiwan*. Tokyo: University of Tokyo.

Utsumi, Atsuko. 2016. A dialectal study of Talaud: Dialectal classification and a comparative word list. *Bulletin of Meisei University. Department of Japanese and Comparative Culture, School of Humanities* 24. 37–70.

Yamada, Yukihiro. 2002. *Itbayat-English dictionary*. Osaka: ELPR Publications.

Zorc, David R. 1982. Where, o where, have the laryngeals gone?. Austronesian laryngeals re-examined. In Halim, Amram & Carrington, Lois & Wurm, S.A. (eds.), *Papers from the Third International Conference on Austronesian Linguistics, volume 2: Tracking the travellers*, 111–144. Canberra: Pacific Linguistics.

Zorc, David R. 1996. The reconstruction and status of Austronesian glottal stop — chimera or chameleon. In Nothofer, Bernd (ed.), *Reconstruction, classification, description: Festschrift in honor of Isidore Dyen*, 41–72. Hamburg: Abera Verlag Meyer & Co.

Address for correspondence

Alexander D. Smith
Institute of Modern Languages and Linguistics
Fudan University
Room 1213, Guanghua East Main Building
220 Handan Road, Yangpu District
Shanghai 200433
China

alexsmith@fudan.edu.cn
https://orcid.org/0000-0003-2510-8555

Publication history

Date received: 4 August 2022
Date accepted: 10 May 2024
Published online: 5 August 2025