

NOTES ON CHINESE WORDS IN SHORTO'S PROTO-AUSTROASIATIC RECONSTRUCTIONS¹

Mark J. Alves
Montgomery College
mark.alves@montgomerycollege.edu

Abstract

This paper evaluates Chinese lexical data in Shorto's 2006 Proto-Mon-Khmer reconstructions to prevent misapplication of his reconstructions, which in a few dozen instances are based on problematic data that affect or even refute his reconstructions. First, Shorto notes about 20 Chinese items to consider for their comparable semantic and phonological properties. While several are probable Chinese loanwords spread throughout the region, a majority of these are unlikely to be Chinese as they are either Wanderwörter seen in multiple language families with undetermined origins or, in most cases, simply partial chance similarities, and these latter items can thus be removed from consideration in Proto-Austroasiatic reconstructions. Second, Shorto also listed about 50 Vietnamese words as supporting data for proto-Austroasiatic etyma which are either (a) clearly Sino-Vietnamese readings of Chinese characters (about 20 instances) or (b) Early Sino-Vietnamese colloquial borrowings (about 30 instances). Many of those proposed proto-Austroasiatic reconstructions must be reconsidered due to the exclusion of these Sino-Vietnamese items. While excluding such Sino-Vietnamese or Early Sino-Vietnamese items in some cases has no impact on those reconstructions, other exclusions result in slight changes in the reconstructed forms, and in several cases, proposed reconstructions must be entirely excluded as only Vietnamese and one other branch of Austroasiatic are available as comparative evidence. Finally, both the exclusions of proposed attestations (and the clarification of their actual origin) and the hypotheses of regional spread of Chinese words must be considered not only for Proto-Austroasiatic but also in comparative historical linguistic studies in the region.

Keywords: Proto-Austroasiatic, Sino-Vietnamese, etymology, historical linguistics
ISO 639-3 codes: bdq, bru, khm, kjg, kuf, ltc, mnw, mtq, och, pac, vie, zh

1 Introduction

Since Shorto's reconstruction of Proto-Austroasiatic² consonants, vowels, and 2,000 entries of lexical comparisons were published in 2006, a tremendous quantity of additional linguistic data and digital tools have been made available on Austroasiatic as well as neighboring language groups with which Austroasiatic has been in contact. These digital resources and tools to search through them have facilitated research tremendously, giving a broader picture than ever before. It is, therefore, important to review Shorto's work with the goal of improving it based on new data. In this paper, the focus is on Chinese words in Shorto's

¹ On such a complex interregional topic which involves multiple language families, it is always beneficial to elicit the views of experts in related areas. I must thank Paul Sidwell for his insights on aspects of Austroasiatic and Marc Miyake for his views on matters of Sinology, both Chinese and Sino-Vietnamese phonology. Fangzhi Jia gave additional suggestions which have resulted in modifications of the claims in some places in the paper. Both offered ideas that led to refinement of details in the paper. For any remaining flaws or incorrect claims—as undoubtedly future investigation will inevitably reveal—I alone readily bear that responsibility.

² Shorto used the term Proto-Mon-Khmer, which is now a questionable taxonym. See Sidwell and Blench 2011.

Proto-Austroasiatic (Proto-AA hereafter), some of which he explicitly noted and a few dozen others which he did not know of, but which still have an impact on his reconstructions. This article is structured as follows: (1) the background of the inquiry and the method of investigation, (2) relevant historical phonological matters in the inquiry, (3) key findings and the role of Chinese loanwords in both Shorto's reconstructions and in Southeast Asian etymological study, (5) notes and evaluations of all entries containing Sino-Vietnamese (SV hereafter) items among the proposed attestations, and (6) a table listing those four dozen items with brief notes to allow anyone using Shorto's work to check whether the items are still valid or ought to be modified in any way.

1.1 Background

Shorto's Proto-AA lexical reconstructions, with numerous additional notes and other data, were organized and published in 2006 published by Pacific Linguistics, edited by Sidwell, Cooper, and Bauer. Each entry with an identifying number (e.g. #502) contains a proposed reconstruction (sometimes with multiple forms, as discussed below), various proposed attestations in various Austroasiatic languages, and sometimes comparative data from other language groups, including Tibeto-Burman (based on Benedict 1972 and other sources),³ Old Chinese (based on Benedict 1972), Proto-Tai (based on Benedict 1975), and Austronesian (from various sources), along with instances of comparable forms in individual modern languages.

In these entries, Shorto mentions SV vocabulary in only two places in the work (2006:237 in #790 and 560 in item B8), with the implicit understanding that these items cannot contribute to Proto-AA reconstructions. However, I gradually found that Shorto provided examples of other SV words, that is, officially recognized Vietnamese readings of Chinese characters listed in Sino-Vietnamese dictionaries, which he apparently did not consult. Moreover, in other cases, a number of Early Sino-Vietnamese (ESV hereafter) items, which are colloquial borrowings that predate character readings by some centuries, were also included in the data. ESV lexical items are less well known, and so it is not surprising that he did not identify these originally Chinese items. As the number of various SV items that I identified grew, I realized excluding these items had effects on his reconstructions.

I then began to pay attention to the Chinese items that Shorto listed, including about twenty Chinese etyma for general comparative purposes, with no suggestion of any shared genetic affiliation of Austroasiatic and Sino-Tibetan. At the time Shorto was doing his work over four decades ago (Shorto had a complete draft by 1977 (Shorto 2006:ix)), digital databases, such as Proto-Tai-o-Matic, STEDT, Blust's Austronesian online comparative dictionary, the Munda Etymological Dictionary, ancient Chinese text corpora such as ctext.org, and the Mon-Khmer Etymological Dictionary⁴ (MKED hereafter) were not yet available. And while some language reconstructions were available at the time he did his work, several major reconstructions have been generated, such as Ratliff's Proto-Hmong-Mien (2010), Pittatayaporn's Proto-Tai (2009), Zhengzhang's Old Chinese (2003), and Baxter and Sagart's Old Chinese (2014), in addition to several branches of Austroasiatic included in the MKED.⁵

Thus, I decided to fully review Shorto's reconstructions to identify all likely instances of Chinese loanwords in Vietnamese included in his offered attestations and to review all of the Chinese words he listed for comparison. To do this, with a digital version of the text, I searched for text strings of all mention of both Vietnamese and Muong. Of Shorto's nearly 2,100 Proto-AA reconstructions, Vietnamese words were provided as comparable data in approximately 600 entries. Of those, about 50 are Chinese loanwords in Vietnamese, from both the SV and ESV loanword strata. Having identified as many SV items as I could (and it is certainly possible for more to be uncovered later), I began to evaluate the impact on the reconstructions in terms of the phonological shape or the likelihood of items as Proto-AA etyma.

³ Tibeto-Burman here is used as a shorthand to refer to all non-Sinitic groups within Sino-Tibetan/Trans-Himalayan and not to adhere to any hypothesis of the division of the the branches.

⁴ Shorto's reconstructions are a central part of the MKED, but over the past decade, 1/4 million items of both synchronic data and Austroasiatic branch reconstructions have since been added, thereby allowing for productive re-evaluation of Shorto's hypothesized reconstructions.

⁵ This is not to exclude Ostapirat's 2000 Proto-Kra and Norquest's 2007 Proto-Hlai, but studies of historical contact between these groups and Austroasiatic is largely lacking, so no clear statements can be made until such matters are explored and contact is either clarified (e.g. contact in an early period in Kradai history) or excluded (e.g. contact is limited to the Tai part of Kradai).

The following sections first review key phonological issues, focusing on final consonants in Austroasiatic with respect to Vietnamese tones—as Shorto appears to have used Vietnamese in his reconstructions—and then summarize the types of Chinese words in Shorto's work, intentionally and unintentionally included, and their impact on each reconstructed entry. A table in the appendix provides a summary list of the SV items and notes the level of impact on each entry.

1.2 Phonological Issues

In this section, I briefly summarize what Shorto considered in his phonological reconstructions, particularly with respect to consonants, vowels, and their relationship with Vietnamese tones. I then present the current understanding of the status of the diachronic relationship between Vietnamese tones and both Austroasiatic and Sinitic.

1.2.1 Shorto's Assumptions

Shorto's 70-page introduction to reconstructing Proto-Mon-Khmer is divided into roughly half on the vowel system and half on the consonant system. His reconstruction consists of a basic nine-vowel system expanded by a length distinction and three diphthongs, all prior to the registrogenesis that later spread through much of the language family. Shorto's reconstructed system is roughly similar to the system presented by Sidwell and Rau (2015:253), with the exception of some modest modifications, namely, Shorto's reconstruction of * ϵ , not Sidwell and Rau's *ai, and no rhymes reconstructed with the pattern *-VCh or *-VC?. A special section covers back vowels before *-h and *-s (Shorto 2006:29-35). Shorto's proposed Proto-Mon-Khmer consonant inventory is the same as that of Sidwell and Rau (2015:238). Thus, there has been a general consistent perception of the proto-AA phonological system.

More relevant to what the Chinese and SV data show is the issue of tones and how Shorto viewed these matters in his reconstructions. Shorto was well aware of tone height. He recognized that (a) Vietnamese upper-register tones (the *ngang*, *sắc*, *hỏi* tones in Vietnamese) correspond to voiceless initials such as *k, *c, *t, and *p (2006:36), fricatives *s and *h and glottal stop (ibid.:48), as well as glottalized *d' and *b' (ibid.:42), realized as /n/ and /m/ respectively in modern Vietnamese and that (b) Vietnamese lower-register tones (called *huyền*, *nặng*, *ngã*) correspond to voiced stops *g, *dʒ~j, *d, and *b (ibid.:39) and sonorants *j, *r, *l, and *w (ibid.:45). He also recognized the correspondence between *hỏi* and *ngã* tones for words with final *-s (ibid.:32, 33) and *-h (ibid.:35). Where tone height is unexpected, he considered possible affixes as a factor (e.g. Vietnamese *nhấc* 'to lift) for Proto-Mon-Khmer *juk #385 (ibid.:45).⁶ The data provided will show, in some cases, no need to posit affixes as the items in Vietnamese were borrowed from Chinese.

As for the relationship between Proto-AA finals and Vietnamese tones, Shorto's exact view is not entirely clear. Yet, the data bears evidence of the application of Vietnamese tones in the reconstructions. Sidwell (Shorto 2006:x) notes how Shorto's reconstructions incorporated understanding of Vietnamese tonogenesis in reconstructing Proto-AA final *-ʔ and *-h. However, Sidwell also writes (ibid.:ix) that the original text of Shorto's that he used to produce the final work was an incomplete presentation of rhymes with final glottal stop and *-h. Thus, there is no explicit statement showing that Shorto reconstructed final glottal stop in words when Viet-Muong languages had the *sắc* or *nặng* tones and final fricatives in those with *hỏi* and *ngã* tones, but his reconstructions are consistent enough throughout the work to suggest that this was the case (e.g. #1590 *[l]jaarh 'the right way up, supine' with Vietnamese *ngửa* 'to lie on back' with the *ngã* tone as an attestation). However, oddly, Shorto suggests at one point (ibid.:51) that Vietnamese and Muong non-stop-final words with the *sắc* tone be reconstructed with *CC, rather than final glottal stop (which is considered a primary source of that tone), but this method seems not to have been applied in the process of reconstruction.

Chinese words in the data have no impact on the reconstructed phoneme inventory of Austroasiatic, but they do impact his views of reconstructed words, sometimes minimally and sometimes more significantly. For instance, Shorto hypothesizes that the tones of Vietnamese *bái* 'to pray' and *ánh* 'to shine' (ibid.:12, 13) are associated with the initial sequences in those words. However, in both cases, these are Chinese loanwords

⁶ Interestingly, these suggest knowledge of Haudricourt's (1952, 1954) hypothesis of Vietnamese tonogenesis, and he knew Haudricourt personally (Sidwell p.c.), but there is no mention of Haudricourt in this introductory section and noted only twice elsewhere in the rest of the book (Shorto 2006:469 and 490)

(see the respective entries #1644 and #489 below). The result is not a change to the reconstruction of the Proto-AA phonological system, but there is a small reduction in the amount of supporting data for his claims.

1.2.2 Vietnamese and Chinese Tone Systems

Other aspects show more significant impact. Another consequence of Shorto's uncertainty of Proto-AA *-ʔ and *-h is multiple reconstructions of entries, often with a mixture of syllables with and without final glottal stop, or with either glottal stop or final *-h. In various instances, the Vietnamese items provided as support appear to have been considered by Shorto as to whether such finals were reconstructed. And it is precisely in the words which are Chinese loanwords in Vietnamese that the reconstructions must then be modified. These most often have an impact on the reconstruction of certain finals in light of the hypothesis of tonogenesis in Vietnamese and Chinese.

As mentioned in the first section, while Shorto apparently sometimes noted that Vietnamese has Chinese loanwords/SV vocabulary, he missed a number of them. In a few places, Shorto notes that certain words in Vietnamese occurred only in bisyllabic compounds, but somehow, he did not recognize that these were part of the spread of large amounts of standard Sino-vocabulary in the early 20th century that spread as well among Japanese and Korean (cf. Sinh 1994, Alves 2015). Moreover, he appears unaware of what Wang Li (1949, 1958) termed Old Sino-Vietnamese (Chinese 古漢越語 *Gǔ Hàn Yuè yǔ*), and which is now sometimes called (as previously noted) Early Sino-Vietnamese in English or *tiền Hán-Việt* "Pre-Sino-Vietnamese" in Vietnamese linguistic publications. The two main layers of SV words have different phonological characteristics in terms of both segmental and suprasegmental features (see Phan 2013 and Alves 2016 for additional information).

The segmental model for tonogenesis in Vietnamese has remained largely unchanged since Haudricourt's often cited 1954 article (preceded by a short article in 1952). The current view is that both Sinitic and Vietic had final fricatives and final glottal stops, in addition to open syllables, syllables with final nasals, and those with final voiceless p/t/k stops, and these are the segmental environments in which the four main tone categories emerged.⁷ In historical phonological literature, these tone categories are often labeled A/B/C/D in Sinitic, Vietic, Kra-dai, and Hmong-Mien. In the Chinese historical phonological tradition going back to the 600s CE, the names of these categories—*píng* (level), *shǎng* (rising), *qù* (going), and *rù* (entering)—were used in rhyming dictionaries to identify the tone categories of Chinese characters.

Table 2: Tone categories and corresponding finals in Sinitic and Vietic

| Type | A | B | C | D |
|------------------------------|-----------------|-------------------|-------------------|------------------|
| <i>Chinese</i> | Ping (level) | Shang (rising) | Qu (departing) | Ru (entering) |
| <i>Sino-Vietnamese</i> | ngang | hỏi/ngã | sắc/nặng | sắc/nặng |
| <i>Early Sino-Vietnamese</i> | ngang | sắc/nặng | hỏi/ngã | sắc/nặng |
| <i>Finals of syllables</i> | open/nasal | glottal stop | fricatives | p/t/k |

The crucial distinction between the ESV items and the later standard SV items is primarily in the B and C categories. When Sinitic words were borrowed initially into Vietic, the pretonal category B words still had glottal stops, while the pretonal category C words still had final fricatives. By the time of SV vocabulary, the speech communities had lost the final fricatives and final glottal stops, while tonal contour remained. This is very important in identifying the early borrowings. There is another category, category A tones for category C tones: the level A tones are the result of borrowing category C tone words before early Vietic had developed tones but after Sinitic had lost final fricatives (Alves 2016).

The four tone categories were eventually divided into eight tones by height (i.e. high and low tones) corresponding to voicing of the initials, with voiceless initials leading to high tones and voiced initials to low tones. When devoicing of all initials occurred in the second millennium CE, the tones became phonemically distinct. While the voicing-to-height tone patterns are very consistent in SV vocabulary, disparities in height

⁷ To present a more precise method of the mechanism of tonogenesis in Vietnamese, and more broadly among languages in that region, Thurgood (2002) has built on the model by applying a phonetically-based laryngeal approach, rather than the more segmental-based approach of Haudricourt.

differences do appear occur in the ESV vocabulary (e.g. #158 'to see', #895 'to dispute', #1444 'to bow, prostrate oneself action custom', etc.). Possible reasons for these inconsistencies are (a) they were borrowed from Sinitic at a time of voicing changes in the variety of Chinese in that region or (b) there was a pre-initial material that was added by Vietic speakers.⁸

In any case, there is sufficient consistency in Shorto's reconstructions to consider that he kept in mind the likely final glottal stops and fricatives in Vietnamese when reconstructing Proto-AA. Therefore, excluding the Chinese loanwords in Vietnamese has the most significant impact on the phonological reconstructions of those entries in which Vietnamese appears to have influenced his decision to reconstruct those final sounds.

1.3 Overview of Chinese Words in Shorto's Reconstructions

In exploring Chinese and Tibeto-Burman vocabulary, Shorto cited the works of Benedict (1972 on Sino-Tibetan) and Pou and Jenner (1973 on Chinese loanwords in Khmer, with numerous notes on Tai, Vietnamese, and even Austronesian comparanda and 1975 on Mon-Khmer and Austronesian comparisons). As mentioned, the words of Chinese origins in Shorto's work included those he explicitly gathered and were entered in his notes as comparanda⁹ and those he did not seem to recognize as SV words, whether SV readings or ESV words. We first deal with those Chinese words that he did note before those he did not.

1.3.1 Chinese Words Shorto Noted

Of those entries in which his notes did allow for the inclusion of Chinese items, there are two categories: trade items and relatively more basic (or at least not culturally specific) vocabulary. The trade item terms—hat, horse, citrus, dish, ring, and goose (see Table 2 in which comparable Old Chinese (OC) and Middle Chinese (MC) forms (Baxter and Sagart 2014)¹⁰ are provided)—have consistent enough phonological patterns and intuitively reasonable scenarios for borrowing; however, the direction and source of borrowing among the main language groups in the region is not always clear. Also, in some cases, textual evidence appears relatively late in the written Chinese record, such as the word for 'hat'.¹¹ This suggests the possibility that such words were borrowed after the Southward expansion into originally Tai-Kadai and even Austroasiatic territory, but this has yet to be determined. In general, these items are either clearly borrowed from other languages or of uncertain origins, making any of these questionable Proto-AA etyma.

Two items Shorto listed with Chinese forms to compare are for 'orange' and 'ring', shown in Table 3. Botanical archaeological studies put the origins of citrus in Mainland Southeast Asia, though details are sketchy. However, considering the wide geography range of *kruuc, including Nicobaric, it seems unlikely to be a borrowing from and more likely a borrowing into Sinitic. For information on Chinese *jú* 'orange', see Fuller et al (2017). As for 'ring,' it appears more likely related to Western Malayo-Polynesian languages (e.g., Malay *cincin*, Javanese *cincin*, Tagalog *siŋsiŋ*, Maranao *sisin*, etc.),¹² and the region around the South China Sea is well known for its early jade trade, including jade earrings and bracelets (though finger rings

⁸ Another possibility is that the items in question are not Chinese in origin. However, in all cases in this study, pending additional data, the items are strong candidates for status as ESV items considering the overall solid phonological (including tone categories), semantic, and sociocultural supporting evidence. Identification of the ESV stratum is admittedly more tenuous than standard Sino-Vietnamese, though there is a good foundation for this area based on previous research by Wang Li (1948, 1958), Haudricourt (1954, etc.), Pulleyblank (1981, 1984, etc.), Starostin (in the Starling Chinese Character Database 1998-2003), Baxter and Sagart (2014, etc.), John Phan (2013), and my own growing database of items with currently over 360 items of high certainty and over 100 items of medium certainty (with another 250 plus items of low certainty or complete exclusions).

⁹ Shorto listed these items in his paper presented at the First International Conference of Austroasiatic linguistics in 1973, in which he provided thoughts in reference to Benedict's work on Sino-Tibetan (1972). I must thank Paul Sidwell for providing me with a scanned copy of that unpublished work.

¹⁰ Baxter and Sagart (2014) reconstructed Old Chinese with a reasonably precise phonetic system, albeit with symbols as stand-in symbols where uncertainty remains due to the limits of data and reconstructive methods. In the Middle Chinese (MC) reconstructions, tone category B is indicated by a final X, while tone category C is indicated with a final H.

¹¹ For 'hat', an early sample is from the 論衡 *Lún Héng*, a book of the Eastern Han in the late first century CE.

¹² This possibility was suggested to me by Paul Sidwell.

are not generally part of this archaeological tradition) stretching back 3,000 years (e.g. Hung et al. 2007, Hung 2014). Overall, these items appear not to be Chinese loanwords; instead, this word for 'orange' is perhaps Austroasiatic in origin, and 'ring' appears a likely Western Austronesian word.

Table 2: Probable Chinese loanwords for trade items¹³

| ID | Proto-MK | Chinese |
|-------|-----------------------------|--|
| #381 | *muuk; *muək 'hat' | 帽 mào 'headpiece', OC *m ^h uk-s, MC mawH |
| #707 | *mraŋ 'horse' ¹⁴ | 馬 mǎ 'horse', OC *m ^h raʔ, MC maeX (a more likely source is Tibeto-Burman *k-m-raŋ ɤ s-raŋ 'horse' (Matisoff 2003)) |
| #1170 | *baan 'pedestal dish' | 盤 bān 'vessel; tub, tray, dish, plate...', OC *[b] ^h an, MC ban |
| #1216 | *haan 'goose' ¹⁵ | 雁 yàn 'wild goose', OC *C.[ŋ] ^h rar-s, MC ngaenH |

Table 3: Items not likely of Chinese origin

| ID | Proto-MK | Chinese |
|-------|--|--|
| #846 | *kruuc; *kruəc 'citrus' | 橘 jú 'orange', OC *[s.k] ^{wi} [t], MC kjwit |
| #1145 | *nciin[]; *nciən[s]; *nciən[] 'ring' | 瑗 yuàn 'large jade ring', OC *w(h)anh, MC wən (Starling 1998-2003) |

Table 4: Relatively basic vocabulary of uncertain origin

| ID | Proto-MK | Chinese |
|--------------|--|---|
| Verbs | | |
| #8 | Pre-Proto-Mon-Khmer *ciʔaʔ > *caʔ, Pre-Khmer ʔ[c]iʔ 'to eat' | 食 shí 'to eat', OC *mə-lək, MC zyik |
| #11 | *cʔ[au]ʔ 'to retch, vomit' | 嘔 ǒu 'to vomit', OC *q ^h (r)oʔ, MC 'uwX |
| #1221 | *khən 'thick, to congeal' | 乾 gān 'dry', OC *[k] ^h ar, MC kan |
| #1239 | *giəp; *gap; *gaap 'to grip, squeeze' | 挾 xié, OC *m-k ^h ep, MC hep 'to grasp' |
| #1242 | *[t]gap; *[t]gaap 'fork, clamp, to clamp') | 挾 xié, OC *m-k ^h ep, MC hep 'to grasp'; 夾 jiā 'to press between', OC *k ^h <r>ep, MC keap; 挟 jiā 'chopsticks', OC *C.k ^h <r>ep, MC kaep |
| #1409 | *[c]limʔ; *[c]liəmʔ; *[c]laim[] 'to lick' | 舔 tiǎn 'to lick', OC *l ^h [i]mʔ, MC themX |
| Nouns | | |
| #20 | *jkooʔ 'body, self' | 軀 qū 'human body', OC NA, MC NA |
| #48 | *kmciʔ 'collateral relation' | 姐 jiě 'elder sister', OC NA, MC NA |
| #132 | *maʔ 'mother' | 媽 mā 'mother', OC NA, MC NA 母 mǔ 'mother', OC *məʔ, MC muwX |
| #199 | *kl[o]ʔ 'shellfish, snail' | 蝸 wō 'snail', OC *k.r ^h oj, MC kwae |
| #265 | *ʔiək '(part of) arm' | 翼 yì 'wing', OC *g ^w rəp, MC yik |
| #269 | *kʔik; *kʔiək; *kʔaik 'armpit' | 腋 yè 'armpit', OC *[c](r)Ak, MC yek |
| #358 | *bo[]k 'belly' | 腹 fù 'belly', OC *p(r)uk, MC pjuwk |

¹³ Schuessler (2007) provides many additional comparative notes for these and other items, including comparanda from Tibeto-Burman, Tai-Kadai, and Austroasiatic.

¹⁴ k-m-raŋ ɤ s-raŋ horse *Tibeto-Burman Matisoff 03 HPTB

¹⁵ This item has a complex situation, with the potential for onomatopoeia and borrowing via Tai as sources for this form. However, even if borrowed from Tai, which would account for the initial fricative versus the initial velar nasal, that makes an indirect Sinitic loan. See Alves 2015 for an overview.

As for the more basic, non-tradeable terms including both verbs and nouns (see Table 4), comparable words are similarly seen in various language groups around the region, including Chinese. However, there are confounding factors for most of them. Sound-symbolism (e.g. ‘to vomit’ and ‘to lick’) and extremely simple syllable shapes (cf. ‘(part of) arm’, ‘armpit’, ‘belly’, ‘mother’) must clearly be excluded as borrowings. The evidently related items meaning ‘to grip’ and ‘to clamp’ appear to be regionally seen in Sinitic, Tai, and Austroasiatic. Its occurrence in Proto-Kuki-Chin (*tsep-I, tse?-II) and Loloish (as opposed to a higher-level reconstruction in Proto-Sino-Tibetan/Trans-Himalayan) further supports this as a regional item. Other than this word form, the items do not clearly represent either contact or shared origins.¹⁶

Ultimately, most of the trade item words may be Wanderwörter with no clear explanations as to their etymological origins, but regardless, they cannot be reconstructed to Proto-AA. It is nearly impossible to support the notion that words such as ‘ring’ (appearing in the archaeological record only from the Iron Age) and ‘horse’ (not indigenous to Mainland Southeast Asia)¹⁷ could have been introduced into Proto-AA culture from groups other than Sinitic speech communities. In contrast, among the more basic terms, most appear to be chance similarities, and thus these are mostly not borrowed words and are at least potentially reconstructable as Proto-AA. However, while these may be reconstructable within Austroasiatic, some appear to have even more complex histories that remain obscure for now.

1.3.2 Words Shorto Did Not Identify as Chinese in Origin

I have identified 50 items which are Chinese in origin, but which Shorto did not appear to recognize as non-native. Of these 50, 29 are ESV words, 20 are standard SV, and one is a Sino-Muong form.

These 50 items are also categorized in terms of their impact on the reconstructions, ranging from no impact, some modification, or complete exclusion. While 17 items have no impact, 33 items are impacted, with about 16 items either excluded or considered regional Chinese loanwords. In 9 instances, some modification of the phonological reconstructions are required (primarily regarding finals as related to tones). Several other items, while not modified or completely excluded, must be considered less likely reconstructable to the Proto-AA level and must either be categorized as sub-regional items in Austroasiatic, or the items used for the reconstruction are smaller in number, which, combined with phonological regularities, make them seem less certain as related forms.

No impact: 17 items

Modified reconstruction: 9 items

Weakened status as Proto-AA etymon: 7 items

Excluded: 8 items

Regional Chinese loanword: 7 items (6 total as the two items for ‘cage’ are doublets)

Uncertain: 1 item

As for the words of likely Chinese origin with regional distribution, it is uncertain exactly when or whether they were borrowed via Tai, Vietnamese, Tibeto-Burman, or Chinese directly. Nevertheless, these are cultural items that seem likely—but not certainly—to have Chinese cultural origins. As was the case with the Chinese words Shorto did note, these are all trade items, as shown in Table 5. Such words clearly cannot be considered part of the Proto-AA of 4000 BP as the primary period of contact with Sinitic began only during the Han Dynasty (200s BCE to 200s CE) expansion into modern day Southern China. These are ESV vocabulary, likely borrowed in the first half of the first millennium CE, and so these are unlikely to have been borrowed earlier elsewhere in Austroasiatic. Further details are provided in the individual entries in Section 2.

¹⁶ Deeper shared origins of the language families (e.g. Austric, Austro-Tai, Sino-Austronesian, etc.) could account for these, though such matters are unclear at best and cannot play a role in the current discussion.

¹⁷ In Higham (2015), early instances of metal rings are either seen made of gold (late in the Iron Age) (ibid.:220) or bronze (ibid.:235). As for horses, Higham writes of a ring with an image of a horse on it discovered in the Prohear archaeological site in the Lower Mekong, “Until these discoveries, horses were unknown in any context in Southeast Asia other than that on a bowl from Ban Don Ta Phet” (ibid.:221).

Table 5: Regional Chinese Loanwords in Austroasiatic

| ID | Proto-AA | Gloss | SV | Chinese |
|-------|--|----------------------|--|---|
| #697 | *[t]ruŋ; *[t]ruuŋ; *[t]ruəŋ | stable, sty, cage | <i>chuông</i> 'cage, shed, coop, stable, sty' | 籠 <i>lóng</i> 'cage/coop', SV <i>lông/lung</i> , OC *k.r ^o ŋ, MC luwng; 籠 <i>lǒng</i> 'basket/cage', SV <i>lông</i> , OC *k.r ^o ŋʔ, MC luwngX |
| #725 | *luŋ; *luuŋ; *luəŋ | stable, sty, cage | <i>lông</i> 'cage' | 籠 <i>lóng</i> 'cage/coop', SV <i>lông/lung</i> , OC *k.r ^o ŋ, MC luwng; 籠 <i>lǒng</i> 'basket/cage', SV <i>lông</i> , OC *k.r ^o ŋʔ, MC luwngX |
| #791 | *d ^h huŋ (& *d ^h huuŋʔ) | tub | <i>thùng</i> 'tub' | 桶 <i>tǒng</i> 'bucket', SV <i>dùng</i> , OC *l ^o ŋʔ, MC thuwngX |
| #1337 | *n ^ɰ juŋ; *n ^ɰ juuŋ; *n ^ɰ juəŋ; *n ^ɰ juəŋ | dark in colour | <i>nhuộm</i> 'to dye' | 染 <i>rǎn</i> 'to dye', SV <i>nhiểm</i> , OC *C.n[a]mʔ, MC nyemX |
| #1813 | *ʔaawʔ | upper garment | <i>áo</i> 'clothing' | 襖 <i>ǎo</i> 'coat/jacket', OC quuʔ (Zhengzhang 2003:266) |

2 Previously Unidentified Chinese Loanwords in Shorto's Data

This section lists the Vietnamese words in Shorto's entries that he considered either attestations of his proposed Proto-AA etyma or simply notes for general comparative purposes. Researchers who have suggested ESV items are noted. Where there is no citation, such items are my proposals as they have not, to the best of my knowledge, been cited elsewhere. For each ESV item, supporting comparative Chinese data includes a Chinese character, pinyin pronunciation for reference, a gloss, the standard SV reading when the sample is an ESV item, the OC reconstruction and the MC transcription. However, when the items are SV vocabulary, the type listed in Sino-Vietnamese character dictionaries, no reconstructions are needed as there is complete certainty of their status as Chinese loanwords.

'mother' *ʔboʔ (#113) – NO IMPACT

Shorto's data in this entry includes Mon, Aslian, and Viet-Muong. However, Vietnamese *vợ* 'wife' is most likely an ESV item (Wang Li 1958:375), from Chinese 婦 *fū* 'woman; wife', SV *phu*, OC *mə.bəʔ, MC bjuwX. Within Vietic, this loanword is seen primarily in Vietnamese and Muong languages, but the southern Chut/Minor Vietic languages have a distinct etymon (*p-ke:ʔ 'wife' (Fer2xx7:R:35)). As for Shorto's proposed etymon, the MKED has additional terms for 'female' in Aslian and terms for female relatives in Pearic and Palaungic, often with final glottal stops. Overall, this seems to be a reasonable reconstruction, and excluding the Vietnamese item does not impact the phonological reconstruction.

'to see' *[s]jəəʔ; by metathesis *[s]ʔəəj (#158) – NO IMPACT

For this entry, Shorto provides data from several branches, including Palaungic, Bahnaric, Nicobaric, Khmuic, and Khasic. As for Vietnamese, *thấy* 'to see' is indicated as secondary data. It is a probable ESV item from Chinese 視 *shì* 'to see', SV *thị*, OC *gijʔ-s, MC dzyijH. The initial is expected (i.e. the tendency in SV for MC initial dental fricatives and affricates to merge with stops), diphthongization is common in ESV vocabulary, and the category B tone matches the OC final glottal stop, though this word's high-register *sắc* tone is unexpected (i.e. the corresponding SV *thị* has a low-register *nặng* tone). Other AA attestations in Shorto's data have final glottal stops, so the exclusion of the SV loanword from PAA comparative data does not impact Shorto's reconstruction.

'staff, cudgel' *[c]duuk (#328) – NO IMPACT

Shorto provides data for Mon, Khasi, Aslian, and tentative data for Khmer and Vietnamese. Vietnamese *thước* 'meter/ruler' is most likely an ESV form (Wang 1958:370) from Chinese 尺 'ruler' *chǐ/chě*, SV *xích*, OC *^hʔAk, MC tsyhek. The semantics (the Vietnamese gloss is lacking in the entry) of this Vietnamese word clearly match the Chinese source and are rather different from the AA item. Moreover, borrowing this type of implement is reasonable considering the types of culturally specific items borrowed into the linguistic ancestor of Vietnamese. Nevertheless, due to the sufficient supporting data from other Austroasiatic languages, exclusion of this item has no immediate impact on Shorto's proposed reconstruction.

'to bind (round)' *bək; *buuk; *buək (#357) – NO IMPACT

The data Shorto provides comes from Mon, Aslian, Palaungic, Nicobaric, and Viet-Muong. A possible source for Vietnamese *buộc* (cf. Muong *puộc*)¹⁸ is Chinese 縛 *fù* 'to bind', SV *phước, phọc, phạc*, OC *bak, MC bjak. A complication comes from Tai, including Proto-Southwest Tai *ph?uuk and Proto-Tai *fruəkD (reconstructions of Li in Proto-Tai-o-Matic). One hypothesis is that this is a Chinese loanword in AA, but that hypothesis is weakened by the presence of cognates in both Nicobarese, as noted, as well as Munda (e.g. [tupak'-] 'to bind, to bundle' (Bondo [Plains]) (bhattacharya1968bonda:C:c1.p71.r12.il427 .s1422)¹⁹).²⁰ That being the case, this item is actually a reasonably strong, widespread candidate for a proto-AA etymon. Moreover, the item in Vietnamese is more likely an AA retention. It is not surprising that a form with similar shape and meaning spread in this geographic region (though I have not found this form in Hmong-Mien, Austronesian, or Tibeto-Burman), though this makes it extremely difficult, if not impossible, to determine etymological origins, direction of borrowing, or even whether items were borrowed.

'to shine, be light' *cʔaiŋ (#489) – NO IMPACT

In this entry, Shorto provides data from Bahnaric, Palaungic, Nicobaric (which is marked less certain), and Vietnamese.²¹ The Vietnamese item *ánh* 'beam, ray' is an SV word, from Chinese 映 *yìng* 'project, reflect light'. Shorto's note in the entry suggesting a hypothesis (ibid.:175) for the unexpected tone is thus moot. As for the AA data, the MKED provides instances in Mon, reasonable entries in Nicobaric, and Khmer.²² Sidwell (p.c.) suggests this entry is related to #487 *cʔaŋ; *cʔaaŋ; *chaŋ; *chajŋ; *cʔiəŋ; *chiəŋ 'to expose to heat', which, if correct, would further strengthen this as a Proto-AA form.

'male animal' *kuəŋ; *kwaəŋ (#502) – NO IMPACT

Shorto here identifies data in Bahnaric, Palaungic, Khasi, and Viet-Muong. In this case, Vietnamese does not have a comparable item, and this author does not know of a comparable form, nor can one be found in various Vietnamese dictionaries. Instead, Shorto provided the Muong word *kông* in the noted compound for 'rooster/cock' *kà kông* (chicken-male). This seems possibly related to Chinese 公 *gōng* 'male (of animals)', which occurs even in Cantonese similarly with the reversed head-modifier pattern, 雞公 *gai' gung'* (chicken-male) 'rooster', and the tone category also matches. The comparable Proto-Vietic form *p-ku:ŋʔ 'mâle, homme vir, male, man' (Fer2xx7:R:964) has an unexplained final glottal stop, which does not match the Muong tone category. It is reasonable to consider this a Sinitic borrowing, but this cannot be proven with absolute certainty. More recent Austroasiatic branch reconstructions provide further support for this as a viable item in Austroasiatic (e.g. proto-Bahnaric *kwəŋ 'male (animal)' (Sid2011:R:479); proto-Katuic *koop 'male (often referring to animals)' (Sid2005:R:1121); proto-Khasic *kwaŋ male (animal) Sid2012:R:502.B). Regardless, even if the Muong item is excluded, there would be no significant impact on this proposed Proto-AA reconstruction.

'to dig' *k[ua]ŋ (#503) – WEAKENED STATUS AS PROTO-AA ETYMON

The samples come from Khmuic, Palaungic, and Viet-Muong. The Vietnamese item *cuốc* 'a hoe; to hoe' (cf. Muong *cuốc*) is a likely ESV word (Starostin in Starling 1998-2003) from Chinese 鋤/鏟 *júé* 'hoe', SV *quốc*, OC *k^wag (Zhengzhang 2003:388), MC kwak. The final voiceless stop in Vietnamese versus the final nasal in the AA reconstruction and the semantic contrast are also factors against affiliation with the proposed

¹⁸ Muong is not a single language but rather a group of related speech varieties (cf. Phan 2012). For Muong data, Shorto used the work of Barker, whose data consisted of the Muong Khen variety.

¹⁹ The citation form in this and items with similar forms are the identifiers of sources in the MKED website.

²⁰ A related form is *bat; *buət 'to tie, bind' (#1032), which is seen in Mon, Khmer, Bahnaric, and possibly Palaungic. In Vietnamese, *bắt* 'to force/compel' (cf. Muong *pắt*) occurs in a compound *bắt buộc* with the same meaning, and *buộc* also has the second meaning 'to force/compel.' Whether this represents semantic shift of *buộc* or chance similarity is not possible to ascertain. Regardless, this further complicates the matter of etymological source.

²¹ Also in this entry, Shorto writes "Connect Kuy *je:ŋ*, Bru *je:ŋ*, Chrau *yà:ŋ* gold", but this is most likely from the Vietnamese word *vàng* 'gold', pronounced with initial /j/ in Central and Southern Vietnamese, where Bahnaric is spoken. Vietnamese *vàng* is an ESV item from Chinese 黃 *huáng* 'yellow', SV *hoàng*, OC *N-k^waj, MC hwang, and thus yet another instance of the spread of a word of Chinese origin in the region.

²² Vietnamese *sáng* 'to shine' is unrelated. It comes from Proto-Vietic *p-la:ŋʔ 'to shine' (Fer2xx7:R:1153).

etymon. This ESV item has also been borrowed into some AA languages in Vietnam (e.g. Pacoh [kuək] 'hoe' (Wat2009:C:1553); Khmu (in Nghe Am, Vietnam) [kuək] 'hoe' (Suw2002:C:930)). Sidwell (2015:113) reconstructs proto-forms for this only in Palaungic *kəəŋ and Khmuic *ku:ŋ, thus restricting this to the northern part of Austroasiatic, not the entire language family.

'stable, sty, cage' *[t]ruŋ; *[t]ruuŋ; *[t]ruəŋ (#697) & 'stable, sty, cage' *luŋ; *luuŋ; *luəŋ' (#725) – REGIONAL CHINESE LOANWORD

Shorto presents two somewhat similar reconstructions with the same gloss. These two posited etyma appear to have been borrowed from Sinitic in both Thai and Vietnamese but at different stages of the phonological shape of the word. Shorto lists two words in Vietnamese: *lồng* 'cage' and *chuồng* 'cage, shed, coop, stable, sty.' These are both likely ESV items (WL 1958:372 noted *lồng*) from two related Chinese items: first, Chinese 籠 *lóng* 'cage', SV *lung*, OC *k.r⁶oŋ, MC *luwng*, and second, the same character a different reading, 籠 *lóng* 'to cage', SV *lông*, OC *k.r⁶oŋ?, MC *luwngX*. This item is also seen in Sino-Tibetan *kru:ŋ (STEDT), Tai-Kadai *kruoŋA (Li 1977), and Western Malayo-Polynesian *kuruŋ (Blust and Trussel).²³ Moreover, this implement is associated with domesticated birds and bird husbandry, of which there is also Chinese lexical evidence in the region in this period (Alves 2015:51-52). It thus seems likely that Shorto's item #697 with the initial cluster, seen in Mon, Khmer, Bahnaric, and Katuic branches of AA (and Mangic: Mang [zəŋ⁶] 'bird cage' (Loi2008:C:1378)), could be related to the form in any of those language groups (i.e., Kradai, Tibeto-Burman, or Sinitic). The connection between Vietnamese *chuồng* and the OC item is of less certainty: it may be from that earlier period with a cluster, as early as the Han Dynasty, though admittedly, OC *kr has been shown to be realized as retroflex 's' /ʃ/ in ESV vocabulary, rather than the palatal stop 'ch' /ç/, which more often results from other presyllabic segments in OC.²⁴ In contrast, item #725, seen only in Vietnamese and Bahnaric, appears to be the result of borrowing Vietnamese *lồng*, which is probably from the later Early MC period and thus a relatively more recent borrowing. Item #670 *ruuŋ; *r[ə]ŋ 'unpartitioned building', which Shorto made note of, appears different enough in meaning and form (the vowel) that it is not likely related.²⁵

'hole; hollow, empty' *dhoonŋ; *dhe[e]ŋ [] (#790) – MODIFIED RECONSTRUCTION (no final *-s)

Shorto presents data for this item from several branches of Austroasiatic as far west as Khasi. For Vietnamese, as secondary material, he mentions the SV reading *không* (one of the two instances in which he explicitly recognizes a word in Vietnamese as SV) from Chinese 空 *kōng* 'empty/void', and also reference to Vietnamese *hông* 'to be hollow, vacant' and what he notes as Cantonese *hōŋ* (presumably /hoŋ⁵⁵/). There is no reason to assume the Chinese item is anything more than chance partial similarity to *hông*, which has an entirely different tone category (level versus non-level), clearly marking it as unrelated to the Chinese form.²⁶ Excluding *không* for comparison does not impact the reconstruction, though exclusion of *hông* would require a reconstruction without final *-s (assuming Shorto reconstructed it based on the Vietnamese *hỏi* tone, as no other Austroasiatic items he noted have a final fricative).

'tub' *d1huŋ (& *d1huuŋ?) (#791) – REGIONAL CHINESE LOANWORD

Shorto lists comparable forms in Mon, Khmer, Katuic, Bahnaric, Khmuic, and Viet-Muong, though comparanda can also be found in Aslian (e.g., (Tonga [Tean Ean]) *thuŋ* 'bucket' (Pha2006:C:76-4)) and Nicobaric (e.g. Nancowry *toŋ* 'wooden bucket' (Man1889:C:5595)). The Vietnamese form *thùng* 'tub/bucket' is a possible ESV item related to Chinese 桶 *tǒng* 'bucket', SV *dũng*, OC *ʃ⁶oŋ?, MC *thuwnŋX*, though the tone category is unexpected (Category A instead of Category B). Shorto notes comparable items

²³ See Schuessler (2007:363) for additional comparative notes on this item.

²⁴ I appreciate Fangzhi Jia for pointing out the phonological concerns with this proposed ESV item.

²⁵ In Muong, in addition to these items, there are distinct etyma: *thòng* '(bird)cage', *cùm* 'cage,' *hàn* 'pen/cage,' and *cũl* 'cage' (related to Vietnamese *cũl*). In Nicobarese, Car has *kinloŋ* 'cage (wooden, built on ground)' (Das1977:C:707), which, if it is related, suggests the depth of the age of this item as well as gives hints about the timing of Nicobarese migration.

²⁶ However, also of partial similarity is 孔 *hǒng* 'opening, hole', SV *khỏng*, though the semantics are only vaguely connected, and it would require the assumption of a change from /x/ to /h/, which is tendency in southern, not northern, Vietnamese.

in Chamic languages as well, and Proto-Hmong-Mien also has *thəŋ(X), also likely borrowed from Chinese (Ratliff 2010). Proto-Southwestern Tai has *thaŋ (Jonsson in Proto-Tai-o-Matic), though the vowel and tone are unexpected, but Central Tai has clearer instances of this, as shown in Gedney's comparative Tai data (Hudak 2008:180). While not all the data is consistently, it seems reasonable to posit that this is a regional Chinese loanword until new data can present a stronger case to separate these comparable forms.

'work' *k[i]ŋ; *kuŋ (#890) – WEAKENED STATUS AS PROTO-AA ETYMON MODIFIED RECONSTRUCTION (No back round vowel)

In this instance, Shorto notes forms in Mon, Palaungic, and Vietnamese, and Khmuic has comparable forms (e.g. [ká:n] 'work (n.)' (T'in [Mal]) Fil2009:C:2361). The proposed Vietnamese item *công* is a standard SV reading of Chinese 工 *gōng* 'work'. Excluding this item, the only one with a back round vowel, suggests there is no need for the reconstruction of *kuŋ, and the remaining instances in Mon, Palaung, and Khmuic form a geographic region, suggesting the possibility of inter-branch borrowing rather than a retention from Proto-AA.

'quarrelsome' *tjaap (#895) – NO IMPACT

Shorto notes instances in Mon, Bahnaric, Nicobaric, and, tentatively, Palaungic. The noted Vietnamese items include *tranh* (with an alternate pronunciation *chanh*, see footnote 23) and *giành*. *Tranh* is a standard SV reading of Chinese 爭 *zhēng* 'to dispute'. The form *giành* has been hypothesized to be a doublet of the Chinese etymon (Pulleyblank 1984:283), though its low tone for an expected high tone word cannot be easily explained. Vietnamese *tranh* must be excluded as a clear SV item, but *giành* is of uncertain origin unless the tone height can be explained (e.g. a presyllable affecting voicing). Excluding the Vietnamese forms does not affect the reconstruction.

'to distribute, disburse' *phat; *phaat (#1112) – WEAKENED STATUS AS PROTO-AA ETYMON

For this item, Shorto lists relevant words in Khmer, Bahnaric (only Sre), and Viet-Muong.²⁷ There is a semantic division between 'repay' and 'distribute,' with 'repay' in Khmer and Bahnaric (and note Palaungic: Palaung [wiət] 'to repay (borrowed money)' (Palaung) Mil1931:C:2584), but not in Vietnamese. Vietnamese *phát* 'to distribute' (with the same item in Muong) is a standard SV reading from Chinese 發 *fā* 'to emit'. Sidwell has reconstructed *paak 'distribute/share' in both Katuic and West Bahnaric²⁸ (Sidwell 2005 and 2003), and the pronunciation of *phát* in Central Vietnamese [fak⁴⁵], similarly with final /k/ (also the form in Southern Vietnamese), suggests this is probably a Vietnamese loanword in those languages. The Vietnamese item must thus be excluded from Shorto's entry, the items in Katuic and Bahnaric ultimately have SV origins, and at best, Khmer and Palaung have a shared form meaning 'to repay,' perhaps an inter-branch borrowing.

'to hide' *[]ʔuun[]; *[]ʔən[s] (#1121) – EXCLUDED

In this entry, Shorto lists items in Bahnaric, Palaungic, and Vietnamese. The listed Vietnamese form *ẩn* is a SV reading of Chinese 隱 *yǐn* 'to hide', which is not a common free morpheme for this sense (Vietnamese *trón* 'to hide' from Proto-Vietic *k-lo:nʔ 'to flee, hide' (Fer2xx7:R:866)). The reconstructed Proto-AA form with final *-s, presumably based on the Vietnamese *hỏi* tone, must be excluded. The other comparative data for Shorto's reconstruction is extremely minimal. Such data is in only a few languages in Palaungic and Bahnaric, and it seems not to be reconstructable to proto-language stages in either Austroasiatic branch. Shorto suggests a possible relationship with item #1168 *[d]puun; *[d]puən 'to hide'. In Aslian, one language, Semelai, has a possible comparable item, ʔəŋ 'hide' (Kru2004:C:563), but no other such items appear in available Aslian data, and considering this extremely simple syllable, it could well be chance

²⁷ The Munda Etymological Dictionary contains a few comparable forms (Bondo (Plains) [baʔa-] 'to be distributed' (bhattacharya1968bonda:C:c1.p91.r14.i1828.s1823)), though as with the items in this entry, the sense is 'distribute' rather than 'repay'. If these are not chance similarities, this proposed etymon could be strengthened, but more data and evaluation is needed.

²⁸ One problem is that North Bahnaric, which is spoken closest to Central Vietnamese, has a different reconstruction, while West Bahnaric, which is spoken in Laos in regions bordering central Vietnam has the *paak etymon. This also borders Katuic, such that that language group could have helped spread this item. Details remain to be clarified.

similarity. It appears Shorto considered the Vietnamese tone in reconstructing final *-s, but this is no longer feasible as none of the items in the other Austroasiatic languages have final fricatives. Altogether, with just a few seeming random samples of a minimal phonological form with unmarked segments, it is best to exclude this item.

‘to gnaw’ *kiən (#1125) – WEAKENED STATUS AS PROTO-AA ETYMON

The data Shorto provides for this proposed etymon includes South Bahnaric, Khmuic, and Vietnamese. Vietnamese *ghen* ‘to be jealous’ has been posited (Starostin in Starling 1998-2003; Baxter and Sagart 2014:191) to be an ESV item (though they do not use ESV to refer to these early loanwords) from Chinese 羨 ‘covet, desire’ *xiàn*, SV *tiện*, OC *s-N-qa[r]-, MC *zjenH*. This is feasible considering the fricative initial ‘gh’ /ɣ/ and the type A tone for a type C tone word, as noted in Section 2. Moreover, ‘gh’ may stem from the earlier cluster in Chinese, and as the Proto-AA form lacks the prefix, the OC form accounts for Vietnamese ‘gh’, while Proto-AA does not (Miyake p.c.). If one posits this as related to the Austroasiatic item, there is no explanation for the fricative, and the semantic shift, while not unreasonable, does not appear elsewhere in available data. Excluding it might not change the reconstruction, though it reduces the comparative evidence to two branches. A review of data in the MKED shows proto-Bahnaric *kiəl, with only some varieties having final /-n/, and some varieties of Khmu final /-r/. At best, this is a word shared by two branches in somewhat close proximity, though even that could be questioned.

‘to receive, to suffer, endure’ *pənʔ; *pən[] (#1151) – MODIFIED RECONSTRUCTION (No final *-ʔ); WEAKENED STATUS AS PROTO-AA ETYMON

Shorto provides data from Vietnamese and Khmuic with the sense of ‘to receive’, Old Mon as a passive auxiliary, and Palaungic with the sense of ‘to endure’. Of the items Shorto notes, Vietnamese *nhận* ‘to receive’ is a standard SV reading of Chinese 認 *rěn* ‘to admit’. Vietnamese has also borrowed *nhịn* ‘to endure’ (Haudricourt 1954:77), which is an ESV item, from Chinese 忍 *rěn* ‘to endure’, OC *nə[n]ʔ, MC *nyinX*. Other data from the MKED includes Katuic (Katu (An Diem) [pən] ‘receive’ (Cos1971:C:2499-1)), but this is feasibly considered a Vietnamese loanword. The other proposed attestations have no final glottal stop, so this exclusion of a Vietnamese *nặng* tone word refutes the reconstruction with a final glottal stop. Also, without the Vietnamese form, there are distinct semantic senses among only three branches in a geographic region. Altogether, the status of this item as a Proto-AA etymon is quite weak at best.²⁹

‘time (quantifier)’ *lən; *l[a]n (#1199) – EXCLUDED

Shorto lists comparanda in only Mon and Vietnamese for this item. Vietnamese *lần* ‘time/turn (unit)’ is most likely an ESV version of SV *luân* from Chinese 輪 *lún* ‘turn/round’, OC *[r]u[n], MC *lwin*. Retention of a monophthong and the low tone corresponding to an initial sonorant (standard SV have high tones for initial sonorants, as is the case for *luân*) supports this. With only data from Mon, this proposed proto-form must be rejected.

‘coiled, to wind, bend’ *win; *wiin; *wiən; *wən[ʔ]; *wan; *waan (#1208) – NO IMPACT

The data Shorto identifies for this proposed etymon includes most branches of AA, making this a very widespread item in the language family. He provides six items from Vietnamese, but two of these are Chinese in origin. First, Vietnamese *khuyên* ‘circle, ring’ is a standard SV reading of Chinese 圈 *quān* ‘circle, ring’. Next, *quấn* ‘to be rolled round, to roll round’ is a probable ESV word from Chinese 卷 *juǎn* ‘to roll,’ SV *quyển*, OC *[k](r)o[n]ʔ, MC *kjwenX*, which has overall comparable form and the expected tone for the final glottal stop of OC.³⁰ Nevertheless, excluding these two items seems not to impact the overall reconstructions.

²⁹ Also seen in AA is a hypothesized *rap, as attested in Khmer, Khmuic, Monic, Palaungic, and Pearic. However, the Proto-Southwest Thai *rap and Proto-Tai *rɔp ‘to accept; receive’ (as is)” (cf. Thai รับ *ráp*) is a likely source.

³⁰ Two other SV readings of this character are *quyển*, from the corresponding noun version of Chinese 卷 *juàn* ‘(a) scroll’, and *quyến*, related to 卷 *quán* ‘handsome’.

'to suffer constraint' *ʔap; *ʔaap (#1224) – WEAKENED STATUS AS PROTO-AA ETYMON

In this instance, Shorto provides evidence in Bahnaric, Mon, and Vietnamese. Vietnamese *áp* 'to press' is a SV reading for Chinese 壓 *yā* 'to press', and thus it cannot be considered an attestation for this proposed etymon. As for the remaining data among the other AA languages, various items have initials other than Shorto's reconstructed initial glottal stop, and moreover, the semantics of the provided items seem to have tenuous semantic correspondences. Excluding the Vietnamese data and considering the remaining data for this minimally distinctive word form results in substantial weakening of this posited reconstruction with data in only two AA branches.

'molar tooth, jaw' *dgam; *dgaam; *dgəm (#1318) – NO IMPACT

Shorto provides ample evidence from several Austroasiatic branches for this item, including Vietnamese *cằm* 'chin', which has suitably matching phonology,³¹ if a slight semantic difference. In his notes, he mentions Vietnamese *hàm* 'jaw', apparently in connection with Chinese, and we indeed consider this an ESV item from Chinese 颌 *hàn* 'jaw, chin', SV *həm*, OC *[g]ʰ[ə]mʔ, MC homX. This is part of the group of ESV items with tone category A for words with C in the later borrowings. Excluding this item to some extent highlights *cằm* as a native form, but for the reconstruction, it otherwise has no impact.³²

'dark in colour' *nɯm; *nɯum; *nɯəm; *nɯəm (#1337) – MODIFIED RECONSTRUCTION (no initial nasal, excluding items meaning "to dye"); REGIONAL CHINESE LOANWORD (items with initial palatal nasal meaning "to dye")

The data for this etymon is complex, with a range of phonological forms and semantic features in several branches of AA. As for Vietnamese, *nhuộm* 'to dye' is a probable ESV item (Haudricourt 1954:77), related to SV *nhiễm* from Chinese 染 *rǎn* 'to dye', OC *C.n[a]mʔ, MC nyemX. It is a solid candidate for ESV status in light of its tone category (the ESV C tone category for a SV B tone category) and the lower vowel. This Chinese word has also been borrowed in Proto-Mien *ɲumC (Ratliff 2010), and Tai languages have also borrowed the Chinese etymon (i.e. Proto-Southwest Tai *ñɔmC (Jonsson in Proto-Tai-o-Matic) and Proto-Tai *ñuəmC (Li in Proto-Tai-o-Matic)). Tai and Vietnamese are likely source languages for the words with the palatal nasal initial meaning 'to dye' in AA languages, such as Bahnar and Palaungic (cf. proto-Palaungic *ɲəm 'to dye' (Sid2010:R:738)), as Shorto noted, but also Khmuic (e.g. Phong *nhɔ:m* 'dye' (Bui2000:C:1511) and Katuic (e.g. Bru *ɲum* 'to dye' (The1980:C:1974)). A few languages in different branches (e.g. Mon, Muong, and Palaung) have a form with the general shape of *jom and mean 'dark', which is the more likely reconstructable item. But in light of the data above, an item with a palatal nasal initial meaning 'to dye' cannot be considered a proto-AA etymon but rather a loanword originally from Chinese that likely entered Austroasiatic through both Tai and Vietnamese.³³

'(to catch) cold' *ksaam; by metathesis *kaams (#1420) - EXCLUDED

Shorto provides data only for Mon and Vietnamese in this entry. Vietnamese *cám* is a standard SV reading of Chinese 感 *gǎn* 'to feel', for which Vietnamese has a similar sense. In Chinese, it occurs in the compound 感冒 *gǎn mào* 'to catch a cold', and in Vietnamese, it also occurs in a number of compounds, often including other SV words, to refer to catching a cold. It appears that the final *-s in the proposed reconstruction *kaams reflected by the Vietnamese *hỏi* tone cannot be retained, but this point is moot as the exclusion of this word leaves only Mon, making this an entirely untenable Proto-AA etymon.

'to bend, nod, drowse' *kuujʔ (& *kuəjʔ?) (#1444) - MODIFIED RECONSTRUCTION (No final *-ʔ)

³¹ There are additional complications. Proto-Vietic data suggests a reconstruction of *-ha:m (< hɲ-ʔa:m < tɲ-ʔa:m ?) "molar" (Fer2xx7:R:728), but *-ha:m is more likely the result of the ESV form, while forms with presyllables may stem from the Proto-AA etymon. However, the speculation that it is an ESV item is still reasonable considering the (a) Vietnamese 'c' /k/ from the OC initial, (b) the short vowel 'ă' /ǎ/ from OC schwa, and (c) the expectation of lenited 'g' /ɣ/ from initial *dg rather than the actual Vietnamese 'c' /k/ (Miyake p.c.).

³² As for the seeming similarity shared by the Proto-AA and OC form, it is a basic CVC syllable with unmarked phonemes, so chance similarity is the best explanation for now.

³³ Shorto also reconstructs *lɔk; *lɔk 'to dye' (#428), with only Mon, Khmer, and Katuic items.

For this entry, Shorto provides attestations from several AA branches with a wide range of semantic meanings. Vietnamese *cúi* ‘to bow, prostrate oneself action custom’ (which is semantically quite different from the proposed AA etymon) is most likely an ESV form, from Chinese 跪 *guì* ‘to kneel; foot’, SV *quì*, OC *[g](r)ojʔ, MC *gjeX*. The tone height is unexpected, but the semantics and overall phonetic correspondence to the Chinese form make this a likely Chinese loanword. The tone type matches (category C for B), and it seems likely this is the reason for Shorto's reconstruction with final glottal stop as none of the other AA languages have final glottal stop for this item. Excluding this Vietnamese item thus suggests a reconstruction of *kuuy.

‘late’ *[l]juuj; *[l]jəj (#1462) - EXCLUDED

The data Shorto provides includes Mon, North Bahnaric, and Vietnamese. Vietnamese *chây* ‘late’ is most likely an ESV item (Wang Li 1958:379) from Chinese 遲 *chí* ‘slow’, SV *tri*, OC *l<r>ə[j], MC *drij*. The diphthong and the palatal initial are expected for ESV. Excluding this from the data leaves just a couple of languages in two branches, which themselves have questionable phonological correspondences. It is thus reasonable to exclude this entirely as a PAA reconstruction.

‘to emit smoke or steam’ *chuj; *chuj (#1546) - NO IMPACT

The data in Shorto's entry includes Khmer, Katuic, Bahnaric, Khmuic, and Nicobaric, with tentative inclusion of Vietnamese. Vietnamese *hơi* ‘steam/vapour’ (a noun, not a verb like Shorto's proposed etymon) is a possible ESV word from Chinese 氣 ‘vapor’ *qì*, SV *khi*, OC *C.qʰəp-s, MC *khj+jH*. The level tone A for a contour tone C word is reasonable in an ESV item borrowed during the transitional period from OC to MC. Excluding this from the data has no impact on Shorto's reconstruction. Vietnamese *hơi* is also noted in item #1807 ‘air, gas’.

‘to kneel’ *ʔbaar (#1644) - WEAKENED STATUS AS PROTO-AA ETYMON

Shorto's attestations for this entry include words from Mon, Khasi, and Vietnamese. As for Vietnamese *bái* ‘to bow, pay homage to’, this is a standard SV reading of Chinese 拜 *bài* ‘to pay respect, worship’. Shorto even hypothesized (2006:12) that the tone in the Vietnamese word is associated with the initial sequence, now no longer an issue to consider. With attestations in only two branches, the status of this proposed PAA etymon is questionable at best.

‘to go round, to turn round’ *wir; *wiir; *wiər; *wər; *war; *waar; *wuur; *wuər (#1669) - NO IMPACT

Shorto provides data for this entry from most branches of Austroasiatic, making it a strong Proto-AA candidate. As for Vietnamese, he noted two items: *về* ‘to return’, which is evidently related to this item via Muong *vêl*, and *vi* ‘to surround’ ‘circumference’, which, as Shorto notes, occurs in compounds. Not surprisingly, it is a standard SV morph, from Chinese 圍 *wéi* ‘to surround’, occurring only in SV bisyllabic compounds. This Vietnamese item must be excluded from the data, but this has no impact on the reconstruction. Instead, one sees that Vietnamese *về* ‘to return,’ from proto-Vietic *ve:r ‘rentrer, to return’ (Fer2xx7:R:473), while similar to *vi*, is a good instance of a chance resemblance.

‘skin’ *huur (#1687) - NO IMPACT

Shorto provides reasonable comparanda for Khmuic, Palaungic, and Munda. However, he also includes Vietnamese *phủ* ‘skin’, noting that it occurs in compounds. This is a formal SV reading 肤 *fū* ‘skin’. Excluding this Vietnamese item has no impact on the reconstructed form.

‘to support, help’ *kəl[]; *kəlʔ (#1705) - MODIFIED RECONSTRUCTION (No final *-ʔ)

For this entry, Shorto provides attestations from Khmer, Katuic, Bahnaric, Khasi, and Viet-Muong. Vietnamese *cứu* is a formal SV reading of Chinese 救 *jiù* ‘to help’. The number of AA languages with the general semantics and word shape support this as a valid Proto-AA reconstruction. However, none of the other AA items have final glottal stops, so Shorto's reconstruction with a final glottal stop appears based solely on the Vietnamese *sác* tone. Thus, while the basic word shape may be valid, the reconstructed final glottal stop must be removed from the reconstruction.

‘to call, to bark’ **[c]kuul* (?); **[c]kuəl*; **[c]kəl*; **[c]kiil*; **[c]kiəl* (#1709) - MODIFIED RECONSTRUCTION (Primarily back vowels)

The data Shorto provides comes from Aslian, Bahnaric, Khmuic, Mon, and Viet-Muong. The Vietnamese word *kêu* ‘to call’ is an ESV item from Chinese 叫 *jiào* ‘to shout’, SV *khiêu*, OC **k^hewk-s*, MC *kewH*. The segments are expected (e.g. no medial palatal glide), and the category A tone for this C tone word is also supporting evidence. Shorto also notes Vietnamese *gâu* ‘to bark (of a dog)’, though it is of course possible that this is the result of onomatopoeia, particularly as final *-l in PAA tends to become a palatal off-glide /-j/ in Vietnamese (Alves 2017), not /-w/. There are enough comparable items beyond Vietnamese to support the basic word shape, though removing the Vietnamese data, with front /e/ and mid /ə/ vowels, should result in reconstructions primarily with back vowels in the several items Shorto proposed.

‘to give, make over’ **[b]kəl* []; **[b]kəls* (#1712) - EXCLUDED

Shorto notes forms in Mon, Khmer, and Vietnamese. Vietnamese *gửi/gởi* ‘to send’ is a likely ESV item (Haudricourt 1954:363, Wang 1958:359) from Chinese 寄 ‘to entrust, to send’ *jì*, SV *ký*, OC **C.[k](r)aj-s*, MC *kjeH*. Without the Vietnamese item, the reconstruction with final *-s is no longer justified. This then limits comparative data to only two AA branches, and while the Khmer form has final /l/, Mon does not. Moreover, Sidwell (p.c.) claims that Old Mon *k and Old Khmer *g is not a typical correspondence. Altogether, this entry should be excluded.

‘to carry on head’ **du*l?; **duul* []; **duəl* [] (#1742) - MODIFIED RECONSTRUCTION (No final *-?)

Shorto’s supporting data here consists of items from Khmer, Katuic, Bahnaric, Nicobaric, and Viet-Muong. Vietnamese *đội* ‘wear/don/carry on head’ is a possible ESV word from Chinese 戴 *dài* ‘carry on the head’, SV *đái*, OC **Cə.ɬək-s*, MC *tojH*. The tone category is correct, though the low tone can only be explained by a voiced initial at the time of borrowing. The proto-Vietic reconstruction of **do:j?* (Fer2xx7:R:314) shows a lack of final *-l, which Vietic languages have preserved in other etyma, thereby providing more support for this item as a loanword. The impact of excluding this item is that **du*l? with a glottal stop is not valid, but the other comparative data leaves this an otherwise reasonable Proto-AA etymon.

‘steam, breath, vapour, gas, air, vapour’ **khuul* (#1807) - NO IMPACT

Shorto posits cognates in Mon, Bahnaric, Katuic, and Vietnamese. He considers Vietnamese *hơi* ‘steam, air, gas’ to be a possible connection (also noted in item #1546). It is feasibly an ESV item from Chinese 氣 *qì* ‘air, gas’, SV *khi*, OC **C.q^həp-s*, MC *khj+jH*. The semantic and phonological features are reasonable (again, see #1546 for details). Excluding this item has no impact on Shorto’s reconstruction.

‘upper garment’ **?aaw?* (#1813) - REGIONAL CHINESE LOANWORD

For this entry, Shorto’s selected data comes from Khmer, Bahnaric, and Viet-Muong. MKED data shows comparable items also in Katuic, but not elsewhere. The Vietnamese item *áo* ‘clothing/shirt’ is listed as an SV word in SV dictionaries, coming from Chinese 襖 *ǎo* ‘coat/jacket’. However, based on the tone category, it should be **áo* with a *hỏi* tone, but with the Vietnamese *sắc* tone, it must have still had a final glottal stop at the time of borrowing (cf. OC *quu?* (Zhengzhang 2003:266) and thus was borrowed in Vietnamese, perhaps in the late OC to early MC period in the first several centuries of the first millennium CE. Notably, none of the noted languages have final glottal stops in this word, suggesting this was borrowed by those languages sometime later after the loss of final glottal stop in Chinese and Vietnamese. It is reasonable to consider this word a Chinese loanword in Khmer, Bahnaric, and Katuic whether indirectly from Vietnamese or directly from Chinese (I have not found comparable forms in Tai languages), the result of trade throughout this eastern part of Austroasiatic.

‘to sit, stay’ **tkaw*[?] (#1818) - NO IMPACT

For this item, Shorto includes data in Katuic, Khmuic, Palaungic, and, tentatively, Nicobaric. He also mentions two Vietnamese items: (a) *cư* ‘to dwell’, which is a SV word from Chinese 居 *jū* ‘to reside’ and (b) *cứ* ‘to continue to’, which is also a likely SV form from Chinese 據 *jù* ‘according to’ (Fangzhi Jia, p.c.). Overall, it appears unlikely to be related to the AA item and must be excluded, though with no impact on the reconstruction as some of the AA languages have final glottal stops.

'to bargain' *.caaw (#1821a) - EXCLUDED

In this entry, Shorto notes items in Khmuic, Khasi, and Viet-Muong. Vietnamese *giao* 'to hand over' is a SV item from Chinese 交 *jiāo* 'to submit/hand over'. Excluding the Vietnamese item makes this entire posited etymon extremely weak, as the items in Khasi and Khmuic seem phonologically too dissimilar, and this posited etymon should be excluded from consideration for PAA status.

'to hand over' *Jaw? (?); *Jaaw[] (#1822) - WEAKENED STATUS AS PROTO-AA ETYMON

The primary data Shorto provides comes from Palaungic, Khasi (tentative), Khmer, and Bahnaric. In addition, he lists with uncertainty Vietnamese *chợ* 'market-place', which is a probable ESV word (Haudricourt 1954:76), from Chinese 市 *shì* 'market (n.)', SV *thị*, OC *C.[d]əʔ, MC *dzyiX*. Excluding this item does not affect the reconstruction. However, some of the Bahnaric forms appear similar in form and meaning 'to hand over' to SV *jiao* 'to hand over' (see notes in #1821a), with possible borrowing through Vietnamese. At the very least, such data requires an alternative reconstruction, but in general, this is significantly weakened as a potential Proto-AA item.

'cat' *miəw (#1838) – REGIONAL CHINESE LOANWORD (PLUS ONOMATOPOEIA)

Shorto notes data for this item in Bahnaric, Palaungic, Viet-Muong, Aslian, and Nicobaric, in addition to Thai and Shan. Moreover, comparable forms are widespread in Tai, as seen in Gedney's notes (Hudak 2008:163), while in Tibeto-Burman, there are numerous instances of general items with the shape [mi] (see STEDT), which does lack the diphthong of the other languages (As for Hmong-Mien, Kra, or Hlai, I have not found comparable forms). As for Vietnamese, *mèo* is a possible ESV item (Wang 1958:365) from Chinese 貓 *māo* 'cat', SV *miêu*, OC *C.m'raw, MC *maew*. Archaeogenetic studies of cats in East Asia suggest domestication of cats in Southwest China around 3000 BCE (Vigne et al. 2016), not Mainland Southeast Asia. It is also important to note that cat bones are not generally seen in archaeological studies in Mainland Southeast Asia in the pre-Qin era, while bones of dogs and pigs, for example, can be easily located. This is clearly not a Proto-AA item but rather a possible Chinese loanword through Vietnamese and/or Tai languages. Still, the strongly onomatopoeic word seen in multiple language groups in the region, combined with hypothesized domestication of cats in Southwest China rather than the more northern homeland of Sinitic, does further obscure the origins, so the status as a Chinese loanword is still somewhat tentative.

'to untie, unfasten' * []kah (#1968) - NO IMPACT

Shorto's primary data for this item is from Bahnaric, Khmuic, and Palaungic. As an extra note, he provides Vietnamese *cởi/cởi* 'to unfasten', which is a probable ESV item (Wang 1958:360) from Chinese 解 *jiě* 'cut up, unloose', SV *giải/giải/giới*, OC *k'reʔ, MC *keaX*. Excluding this has no impact on Shorto's reconstruction as some of the other AA languages also have final [-h], so the Vietnamese *hỏi* tone is not the sole factor.

'bean' *t1uh; *t1uuh; *t1uəh; *t1əh (#2002) – REGIONAL CHINESE LOANWORD

In this instance, Shorto provides data only for Bahnaric (treating North Bahnaric and South Bahnaric separately) and Vietnamese. He provides Vietnamese *đậu* and *đỗ* 'bean' as posited attestations. Vietnamese *đỗ* is a probable ESV item from Chinese 豆 *dòu* 'bean', OC *[N.t]ʰo-s, MC *duwH*, while *đậu* is the standard SV counterpart. The ESV tone B for SV tone C is entirely reasonable. Proto-Bahnaric *tu:h 'bean' (Sid2011:R:882) has final *h, and it is possible that this product was traded and the word borrowed into Bahnaric when Vietic still retained a final /h/.³⁴ This word is also a probable Chinese loanword in Tai (*thue B) and with evidence as well in Hmong-Mien. Overall, the data suggests that this is a regional Chinese loanword.

³⁴ Admittedly, for this scenario to have occurred, it would have been in the first millennium CE, when Vietic likely still had final *-h. However, there is no clear Vietic-Bahnaric contact for that rather early part of history. Such time depth further obscures the situation.

'to pour out' *tuh; *tuəh; *təh (#2003) - WEAKENED STATUS AS PROTO-AA ETYMON

Shorto provides data for this entry from Bahnaric, Palaungic, Aslian, and Viet-Muong. Shorto (2006:509) mistakenly shows Vietnamese *đổ*, which means 'to pass (of a test)', but most likely this should be *đổ* 'to spill/pour', which is a possible ESV item from Chinese 倒 *dǎo* 'to pour', SV *đào*, OC *tʰawʔ-s, MC tawH.³⁵ Excluding the Vietnamese data does not affect the phonological reconstruction, though without the Vietnamese and related Muong item *tổ*, the semantics of the original becomes less certain: the varied meanings include 'to pour', 'to fall/let fall/put in', 'sprinkle', 'decant', 'rain', 'to put (cooked rice) into bags,' among others. Thus, at the very least, the meaning 'to pour' is no longer primary, and of the remaining items with mixed semantic correspondences, these are more likely chance phonological similarities.

'to use a bow' *pooh (#2024) - UNCERTAIN

Shorto's data here includes Mon, Khmer, Palaungic, and Viet-Muong. The data he provides from Vietnamese is *nỏ* 'crossbow'. This form, as well as *ná* (Haudricourt 1954:361), are possible ESV items from Chinese 弩 *nǚ* 'crossbow', SV *nỗ*, OC *C.nʰaʔ, MC nuX. The circumstances surrounding these words is quite complicated:³⁶ (a) there are three phonological forms in Vietnamese, (b) it is possible for a native noun to have been derived with a verb via a nasal infix, with only the nasal remaining, (c) Shorto posits Proto-AA *tʰaʔ #32 'bow' and *snaʔ #97 'crossbow',³⁷ and comparable forms are scattered throughout AA languages, including Mon, Khmer, Katuic, Pearic, Bahnaric, Vietic, and Munda, (d) it is also reconstructed in Proto-Tai *hnaC, and (e) the archaeological history of crossbows is complicated (e.g. bronze triggers were developed in central China during the mid-first millennium BCE, but the history of wooden crossbows in MSEA is less clear). One possibility is that, considering correspondences of earlier OC finals and MC tone categories, *ná* is an ESV item (though also hypothetically a retention of the AA etymon), while *nỗ* is a standard SV reading. Whether *nỏ* is an ESV item as well, borrowed chronologically between the other two, or is actually a retention of the posited Proto-AA item, cannot be determined with certainty based on current data.

'to flow out' *bah; *baas (#2032) - NO IMPACT

For this entry, Shorto provides data from Bahnaric, Mon, Khmer, Katuic, and Viet-Muong, and also notes Proto-Austronesian data (the latter of which appears rather unlikely). Vietnamese *vãi* 'to spill, to be spilled, to strew', possibly in relation to scattering seed. This has been posited (Nguyễn 1995:210) to be an ESV word from Chinese 播 *bō/bò/bǒ* 'to sow', SV *bá/bả*, OC *pʰar-s, MC paH. It may be impossible to determine whether this is a retention from Proto-AA or an Old Chinese loanword. The overall shape and meaning are comparable in both cases. If it is related to scattering seed, it could be a Sinitic loan. There is ultimately no impact of retaining or excluding this item, but its origin not certain.

'to appear (through)' *luh; *luuh; *luəh; *luʔ (#2071) - MODIFIED RECONSTRUCTION (no final *-ʔ)

Shorto provides data from Bahnaric, Mon, Khmer (marked as less certain), and Viet-Muong. Vietnamese *lộ* 'to appear' is a standard SV reading of Chinese 露 *lù* 'to reveal/show'. No other attestations in the AA language data have final glottal stops, so Shorto may have included a reconstructed form with final glottal stop in light of the Vietnamese and Muong data. If so, by removing the Viet-Muong data, only final *h need be reconstructed.

References

- Alves, Mark J. 2015. Chinese Loanwords in Vietnamese. in *Encyclopedia of Chinese Language and Linguistics*, General Editor Rint Sybesma. http://dx.doi.org/10.1163/2210-7363_ecll_COM_000170.
- Alves, Mark J. 2015. Etyma for 'Chicken', 'Duck', & 'Goose' among Language Phyla in China & Southeast Asia. *Journal of the Southeast Asian Linguistics Society* 9:39-55.

³⁵ SV items tend to retain MC –aw and not develop –ow, which does reduce the certainty of this as a Chinese loanword.

³⁶ See also Schuessler (2007:404-405) for additional comparative notes with other language families in the region.

³⁷ Shorto's two reconstructions would seem to be one etymon. Additional in the MKED, including reconstructions in various branches, are difficult to reconcile, but still, a form approximating *snaʔ is likely.

- Alves, Mark. 2015. Historical notes on words for knives, swords, and other metal implements in early Southern China and Mainland Southeast Asia. *Mon-Khmer Studies* 44:39-56.
- Alves, Mark J. 2016. Identifying Early Sino-Vietnamese Vocabulary via Linguistic, Historical, Archaeological, and Ethnological Data. *Bulletin of Chinese Linguistics* 9:264-295.
- Alves, Mark J. 2017. Does Vietnamese have evidence for OC *-r? *Cahiers de Linguistique Asie Orientale* 46 (2017):151–173.
- Baxter, William H. and Laurent Sagart. 2014. *Baxter-Sagart Old Chinese reconstruction, version 1.1*. Web. <http://ocbaxtersagart.lsa.umich.edu/BaxterSagartOCbyMandarinMC2014-09-20.pdf>.
- Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus*. New York, Cambridge University Press.
- Benedict, Paul K. 1975. *Austro-Thai language and culture, with a glossary of roots*. New Haven, Human Relations Area Files Press.
- Blust, Robert & Stephen Trussel. 2010. *Austronesian comparative dictionary*. Web. http://www.trussel2.com/ACD/acd-ak_a.htm. Updated December 21, 2014.
- Fuller, dorian q., cristina castillo, eleanor kingwell-banham, ling qin and alison weisskopf. 2017. Charred pummelo peel, historical linguistics and other tree crops: approaches to framing the historical context of early Citrus cultivation in East, South and Southeast Asia. in *AGRUMED: Archaeology and History of Citrus Fruit in the Mediterranean (Acclimitization, Diversifications, Uses)*, ed. by Véronique Zech-Matterne and Girolamo Fiorentino. Publications de Centre Jean Bérard.
- Haudricourt, André-Georges. 1952. L'origine môn-khmère des tons en vietnamien. *Journal Asiatique* 240:264-265.
- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242:69-82.
- Hudak, T. J. 2008. *William J. Gedney's comparative Tai source book*. Honolulu: University of Hawai'i Press.
- Hung, Hsiao-chun. 2014. Jade in Southeast Asia. in *Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures*. DOI 10.1007/978-94-007-3934-5_9927-1.
- Hung, Hsiao-chun, et al. 2007. Ancient jades map 3,000 years of prehistoric exchange in Southeast Asia. *PNAS* 104.50. 19745–19750. www.pnas.org/cgi/doi/10.1073/pnas.0707304104.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*, Berkeley: University of California Press.
- Nguyễn, Tài Cẩn. 1995. *Giáo trình lịch sử ngữ âm tiếng Việt (sơ thảo)* [A history of Vietnamese phonology (a draft)]. Hanoi: Nhà Xuất Bản Giáo Dục.
- Phan, John D. 2012. Mường is not a subgroup: Phonological evidence for a paraphyletic taxon in the Viet-Muong sub-family. *Mon-Khmer Studies* 40:1–18.
- Phan, John Duong. 2013. Lacquered words: the evolution of Vietnamese under Sinitic influences from the 1st century BCE through the 17th century CE. Ithaca: Cornell University. Doctoral dissertation.
- Pou, Saveros and Philip N. Jenner. 1973. Some Chinese loanwords in Khmer. Hong Kong University Press. *Journal of Oriental Studies* 6.1:1-90.
- Pou, Saveros, and Philip N. Jenner. 1974. Proto-Indonesian and Mon-Khmer. in *Applied psycholinguistics: psychological studies of language processes*, 112-24.
- Proto-Tai-o-Matic*. Thai Lexicography Resources. <http://sealang.net/crcl/proto/>.
- Pulleyblank, Edwin G. 1981. Some notes on Chinese historical phonology. *Bulletin de l'École Française d'Extrême-Orient* 69.1:277–288.
- Pulleyblank, Edwin G. 1984. *Middle Chinese: A study in historical phonology*. Vancouver: University of British Columbia Press.
- Pulleyblank, Edwin G. 1984. *Middle Chinese: a study in historical phonology*. University of British Columbia.
- Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.
- Schuessler, Axel. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawai'i Press.

- SEAlang Mon-Khmer Etymological Dictionary. 2007-2011. SEAlang Mon-Khmer Languages Project. <http://sealang.net/monkhmer/dictionary/>.
- SEAlang Munda Etymological Dictionary. 2007-2011. SEAlang Munda Languages Project. <http://www.sealang.net/munda/dictionary/>.
- Shorto, H. L. 1973. Mon-Khmer contact words in Sino-Tibetan: commentary on Paul K. Benedict, Sino-Tibetan (Cambridge 1972). Paper given at the First International Conference of Austroasiatic Linguistics, Honolulu.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra Pacific Linguistics 579.
- Sidwell, Paul and Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: volume 1 — West Bahnaric*. Canberra: Pacific Linguistics 551.
- Sidwell, Paul. 2005. *The Katuic Languages: Classification, reconstruction and comparative lexicon*. Munich: Lincom Europa.
- Sidwell, Paul. 2015. *The Palaungic languages: Classification, reconstruction and comparative lexicon*. Munich: Lincom Europa.
- Sidwell, Paul and Felix Rau. 2015. Austroasiatic comparative-historical reconstruction: an overview. in *The Handbook of Austroasiatic languages*, Vol. 1, ed. by Mathias Jenny and Paul Sidwell. 221-363.
- Sidwell, Paul and Roger Blench. 2011. The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In *Dynamics of Human Diversity*, ed. by Nick Enfield. Pacific Linguistics. 317-345.
- Sinh, Vinh. 1993. Chinese characters as the medium for transmitting the vocabulary of modernization from Japan to Vietnam in Early 20th century. *Asian Pacific Quarterly* 25.1:1-16.
- Starling Chinese Character Database. 1998-2003. in *The Tower of Babel: An International Etymological Database Project*. <http://starling.rinet.ru/cgi-bin/query.cgi?basename=%5Cdata%5Cchina%5Cbigchina>.
- Thurgood, G.: Vietnamese and tonogenesis: revising the model and the analysis. *Diachronica* 19.2: 333-363.
- Vigne, Jean-Denis, Allowen Evin, Thomas Cucchi, Lingling Dai, Chong Yu, Songmei Hu, Nicolas Soulages, Weilin Wang, Zhouyong Sun, Jiangtao Gao, Keith Dobney, Jing Yuan. 2016. Earliest “domestic” cats in China identified as leopard cat (*Prionailurus bengalensis*). *PLOS One*, January 22, 2016, <https://doi.org/10.1371/journal.pone.0147295>.
- Wáng Li (王力). 1948. Hànyuèyǔ yánjiū (漢越語研究) [Research on Chinese and Vietnamese]. *Língnán Xuébào* (嶺南學報) 9.1:1-96.
- Wang, Li (王力). 1958. Hanyueyu Yanjiu 汉越语研究 (A study on Sino-Vietnamese). *Hanyu Shi Lunwen Ji* (汉语史论文集). Beijing: Kexue Chubanshe. 290-406.
- Zhengzhang, Shangfang (郑张尚芳). 2003. Shàngǔ yīn xì (上古音系) [Old Chinese Phonology], 上海教育出版社 (Educational Press of Shanghai).

Reviewed: Received 1 August 2018, revised text accepted 15 November 2018, published 10 December 2018

Editors: Editor-In-Chief Dr Mark Alves | Managing Eds. Dr Paul Sidwell, Dr Nathan Hill, Dr Sigrid Lew

Table 6: SV and ESV Words in Shorto's Data and Summary of the Impact on Reconstructions

| ID# | Reconstr. | Gloss | Type | Viet. | Chinese | Evaluation |
|-----|-------------------------------------|-------------|------|---------------|--|------------|
| 113 | *[ʔ]boʔ | mother | ESV | vợ ‘wife’ | 婦 fù ‘woman; wife’, SV phụ, OC *mə.bəʔ, MC bjuwX | No impact |
| 158 | *[s]jəəʔ; by metathesis *[s]ʔəəj | to see | ESV | thấy ‘to see’ | 視 shì ‘to see’, SV thị, OC *gijʔ-s, MC dzyijH | No impact |
| 328 | *[c]duuk | meter/ruler | ESV | thước ‘ruler’ | 尺 ‘ruler’ chǐ/chě, OC *tʰAk, MC tsyhek, SV xích | No impact |

| | | | | | | |
|------|--|-------------------------------------|-----|--|--|---|
| 357 | *bək; *buuk; *buək | to bind (round) | ESV | buộc 'to bind' | 縛 fù 'to bind', SV phước, phọc, phạc, OC *bak, MC phjuwH | No impact |
| 489 | *cʔaiŋ | to shine, be light | SV | ánh 'beam, ray' | 映 yìng 'project, shine light' | No impact |
| 502 | *kuəŋ; *kwaəŋ | male animal | NA | NA | 公 gōng 'male (of animals)' | No impact |
| 503 | *k[uaŋ] | to dig | ESV | cuốc 'pickaxe' | Chinese 鍬/鑿 jué 'hoe', SV quǎc, OC *k ^w ag (Zhengzhang 2003:388), MC kwak (Starostin in Starling 1998-2003) | Weakened status as proto-aa etymon |
| 697 | *[t]ruŋ; *[t]ruuŋ; *[t]ruəŋ | stable, sty, cage | ESV | chuồng 'cage, shed, coop, stable, sty' | 籠 lóng 'cage/coop', SV lộng, OC *k.r ^s oŋ, MC luwng; 籠 lǒng ³ 'basket/cage', SV lộng, OC *k.r ^s oŋʔ, MC luwngX | Regional Chinese loanword |
| 725 | *luŋ; *luuŋ; *luəŋ | stable, sty, cage | ESV | lồng 'cage' | 籠 lóng 'cage/coop', SV lộng, OC *k.r ^s oŋ, MC luwng; 籠 lǒng ³ 'basket/cage', SV lộng, OC *k.r ^s oŋʔ, MC luwngX | Regional Chinese loanword |
| 790 | *dhoŋs; *dhe[e]ŋ[] | hole; hollow, empty | SV | không 'empty' | 空 kōng 'empty/void' | Modified reconstruction (no final *-s) |
| 791 | *d ^h huŋ (& *d ^h huuŋʔ?) | tub | ESV | thùng 'tub' | 桶 'bucket' tǒng, SV dǔng, OC *ʔ ^s oŋʔ, MC thuwnɡX | Regional Chinese loanword |
| 890 | *k[i]n; *kuŋ | work | SV | công 'work' | 公 gōng 'work' | Modified reconstruction (no back round vowel) |
| 895 | *tjaaŋ | quarrelsome | SV | tranh/chanh 'to quarrel' | 争 zhēng 'to dispute' | No impact |
| 895 | *tjaaŋ | quarrelsome | ESV | giành 'to quarrel' | 争 zhēng 'to dispute' | No impact |
| 1112 | *phat; *phaat | to distribute, disburse | SV | phát 'to distribute' | 發 fā 'to emit' | Weakened status as Proto-AA etymon |
| 1121 | *[]ʔuun[]; *[]ʔəŋ[s] | to hide | SV | ẩn 'to hide' | 隐 yǐn 'to hide' | Excluded |
| 1125 | *kiəŋ | to gnaw | ESV | ghen 'to be jealous, envious' | 羨 'covet, desire' xiàn, SV tiệŋ, OC *s-N- qa[r]-, MC szjenH | Weakened status as Proto-AA etymon |
| 1151 | *nəŋʔ; *nəəŋ[] | to receive, to suffer, endure | SV | nhận 'to receive' | 認 rěn 'to admit' | Modified reconstruction (no final *-ʔ); Weakened status as Proto-AA etymon |
| 1199 | *ləŋ; *l[a]n | time (quantifier) | ESV | lần 'turn' | Chinese 輪 lún 'turn/round', SV luán, OC *[r]u[n], MC lwin | Excluded |
| 1208 | *win; *wiin; *wiəŋ; *wəŋ[ʔ]; *wan; *waan | coiled, to wind, bend | ESV | quán 'to be rolled round, to roll round' | 圈 quān 'circle, ring' | No impact |

| | | | | | | |
|------|---|-------------------------------|-----|---|---|---|
| 1208 | *win; *wiin; *wiən; *wən[ʔ]; *wan; *waan | coiled, to wind, bend | SV | khuyên 'circle, ring' | 卷 juǎn 'to roll,' SV quyển, OC *[k](r)o[n]ʔ, MC kjwenX | No impact |
| 1224 | *ʔap; *ʔaap | to suffer constraint | SV | áp 'to oppress' | 壓 yā 'to press' | Weakened status as Proto-AA etymon |
| 1318 | *dgam; *dgaam; *dgəm <i>molar tooth, jaw</i> | jaw | ESV | hàm 'jaw' | 颌 hàn 'jaw, chin', SV hạm, OC *[g]ʰ[ə]mʔ, MC homX | No impact |
| 1337 | *n̄jum; *n̄juum; *n̄juəm; *n̄jəm | dark in colour | ESV | nhuộm 'to dye' | 染 rǎn 'to dye', SV <i>nhiễm</i> , OC *C.n[a]mʔ, MC nyemX | Modified reconstruction (no initial nasal, excluding items meaning "to dye"); Regional Chinese loanword (items with initial palatal nasal meaning "to dye") |
| 1420 | *ksaam; by metathesis *kaams | (to catch) cold; | SV | cảm 'to catch cold' | 感 gǎn 'to feel' | Excluded |
| 1444 | *kuujʔ (& *kuəjʔʔ) | to bend, nod, drowse | ESV | cúi 'to bend, nod' | 跪 guì 'to kneel; foot', OC *[g](r)ojʔ, MC gjweX, SV <i>qui</i> | Modified reconstruction (no final *-ʔ) |
| 1462 | *[l]juuj; *[l]jəj | late | ESV | chày | 遲 chí 'slow', SV tri, OC *l<r>ə[j], MC drij | Excluded |
| 1546 | *chuj; *chuuj | to emit smoke or steam | ESV | hơi 'steam, vapour' | 氣 'vapor' qi4, SV <i>khi</i> , OC *C.qʰəp-s, MC khj+jH | No impact |
| 1644 | *ʔbaar | to kneel | SV | bái 'to bow, pay homage to' | 拜 bài 'to pay respect, worship' | Weakened status as Proto-AA etymon |
| 1669 | *wir; *wiir; *wiər; *wər; *war; *waar; *wuur; *wuər | to go round, to turn round | SV | Vietnamese in compounds vi 'to surround; circumference' | 圍 wéi 'to surround' | No impact |
| 1687 | *huur | skin | SV | Vietnamese in compounds phu 'skin' | 肤 fū 'skin' | No impact |
| 1705 | *kəl[]; *kəəlʔ | to support, help | SV | cứu 'to help' | 救 jiù 'to help' | Modified Reconstruction (no final *-ʔ) |
| 1709 | *[c]kuul (?); *[c]kuəl; *[c]kəl; *[c]kiil; *[c]kiəl | to call, to bark | ESV | kêu 'to call' | 叫 jiào 'to shout', SV khiếu, OC *kʰewk-s, MC kewH | Modified reconstruction (primarily back vowels) |
| 1712 | *[b]kəl[]; *[b]kəəls | to give, make over | ESV | gửi, gởi 'to send, despatch, to leave in someone's care' | 寄 'entrust' jì, SV ký, OC *C.[k](r)aj-s, MC kjeH | Excluded |
| 1742 | *dulʔ; *duul[]; *duəl[] | to carry on head | ESV | đội 'to wear or carry on head' | 戴 dài 'carry on the head', SV <i>đái</i> , OC *Cə.tʰək-s, MC tojH | Modified reconstruction (no final *-ʔ) |

| | | | | | | |
|-------|---------------------------------------|----------------------|------|---|---|--|
| 1807 | *khuul | vapour | ESV | hơi 'steam, breath, vapour, gas, air' | 氣 qì 'air, gas', SV khí, OC *C.q ^h əp-s, MC khj+jH | No impact |
| 1813 | *ʔaawʔ | upper garment | ESV | áo 'clothing' | 襖 áo 'coat/jacket', OC quuʔ (Zhengzhang 2003:266) | Regional Chinese loanword |
| 1818 | *tkaw[ʔ] | to sit, stay | SV | cu 'to dwell' | 居 jū 'to reside' | No impact |
| 1818 | *tkaw[ʔ] | to sit, stay | SV | cứ 'to continue to' | Native Vietnamese, chance partial similarity | No impact |
| 1821a | *.caaw | to bargain | SV | giao 'to hand over' | 交 jiāo 'to submit/hand over' | Excluded |
| 1822 | *jjawʔ (?); *jaaw[] | to hand over | ESV | chợ 'market-place' | 市 shì 'market (n.)', SV thị, OC *C.[d]əʔ, MC dzyiX | Weakened status as Proto-AA etymon |
| 1838 | *miəw | cat | ESV? | mèo 'cat' | 貓 māo 'cat', SV miêu, OC *C.m ^h raw, MC maew | Regional Chinese loanword (?) |
| 1968 | *[]kah | to unfasten | ESV | cởi, cởi 'to untie, unfasten' | 解 jiě 'cut up, unloose', SV giải/giái/giới, OC *N-k ^h reʔ, MC keaX | No impact |
| 2002 | *tluh; *tluuh; *tluəh; *tləh | bean | ESV | đỗ 'bean' | 豆 dòu 'bean', SV đậu, OC *[N.t] ^h o-s, MC duwH | Regional Chinese loanword |
| 2002 | *tluh; *tluuh; *tluəh; *tləh | bean | SV | đậu 'bean' | 豆 dòu 'bean' | Regional Chinese loanword |
| 2003 | *tuh; *tuəh; *təh | to pour out | ESV | đổ 'to spill' | 倒 dào 'to pour', SV đảo, OC *t ^h awʔ-s, MC tawH | Weakened status as Proto-AA etymon |
| 2024 | *pooh | to use a bow | ESV | nỏ 'crossbow' | 弩 nǚ 'crossbow', SV nỏ, OC *C.n ^h aʔ, MC nuX | Uncertain |
| 2032 | *bah; *baas | to flow out | ESV | vãi 'to spill, to be spilled, to strew' | 播 bō 'to sow', SV bá, OC *p ^h ar-s, MC paH | No impact |
| 2071 | *luh; *luuh; *luəh; *luʔ | to appear (through). | SV | lộ 'to appear' | 露 lù 'to reveal/show', OC *p.r ^h ak-s, MC luH | Modified reconstruction (no final *-ʔ) |