

# HISTORICAL ETHNOLINGUISTIC NOTES ON PROTO-AUSTROASIATIC AND PROTO-VIETIC VOCABULARY IN VIETNAMESE<sup>1</sup>

Mark Alves

*Montgomery College*

[mark.alves@montgomerycollege.edu](mailto:mark.alves@montgomerycollege.edu)

**Abstract:** This study provides updated numbers of and historical ethno-linguistic observations on Austroasiatic and Vietic etyma in Vietnamese. Lexical data from two dozen Vietic lects were assembled, half from the Mon-Khmer Etymological Database (MKED hereafter) and half from various other published and unpublished sources. Based on Ferlus's preliminary reconstructions of Proto-Vietic (by Ferlus 2007 in the MKED), and data from Austroasiatic, Proto-Tai, and Old and Middle Chinese, approximately 800 items have been evaluated as viable reconstructions. However, of these, nearly 100 are Chinese loanwords of differing periods, and several are early Tai loanwords. The remaining nearly 700 items are native, including about 200 Proto-Austroasiatic etyma, with a few dozen local Austroasiatic words, and over 460 items specific to Vietic. Statistics have been gathered for cultural domains of the reconstructed vocabulary. A combination of etymological sources, semantic domains, and ethnohistorical data (i.e. archaeology, historical texts, and ethnographic information) allow for hypotheses about the ethno-linguistic circumstances of the early Vietic speech community and language contact situations. Many of the cultural domains are readily identified as part of a Neolithic lifestyle (i.e. words related to the natural environment, generic actions, etc.). Some, on the other hand, demonstrate social stratification (e.g. words related to economic practices) and developed agricultural practices (e.g. a large set of terms related to rice production). Others shed light on regional spread of cultural practices (e.g. betel-nut chewing and tooth-blackening) and intergroup contact (e.g. with Sinitic and Tai). Questions related to the spread of metallurgy and metal implements strongly support the influence of Chinese in metal terms and implements.

**Keywords:** Vietnamese, Vietic, Austroasiatic, historical ethno-linguistics

**ISO 639-3 codes:** vie, mtq, scb, thm, pkt, aem, hnu, tou

## 1. 160 years of research on the Austroasiatic origins of Vietnamese

Progress in identifying Austroasiatic vocabulary in Vietnamese has been extremely slow. In the 1850s—over 160 years ago—Logan (1852-1855) first suggested the term “Mon-Annam”, thereby connecting Vietnamese to what Schmidt would call “Austroasiatic” in 1906 (though Schmidt was not persuaded to associate Vietnamese with that group). The related term “Mon-Khmer” occurred as early as 1904 in Grierson's volume, and it has generally been used to refer to a major branch of Austroasiatic not including Munda languages of India or of Nicobarese of the Nicobar Islands. However, evidence is increasingly showing that there is no distinct Mon-Khmer group. Instead, the branches of Austroasiatic are more equally distributed (Sidwell and Blench 2011). Also, Munda languages, which have long been placed in opposition to the hypothesized Mon-Khmer branch, constitute just one among the various branches and are in fact

---

<sup>1</sup> Thanks goes to Paul Sidwell, Robert Blust, and Nam Kim for reviewing this paper and offering thoughts and comments. I take full responsibility for inevitable errors. I can only hope the errors are small in size and number.

typologically closer to the rest of Austroasiatic than has been previously considered (Anderson 2020). Thus, only the term Austroasiatic will be used in this paper, which is ultimately an uncontroversial position.

Data connecting Vietnamese to Austroasiatic continued to be gathered from the beginning of the 20<sup>th</sup> century. For example, while not convinced of the status of Austroasiatic, Blagden (1894:27-40) published a study with over ten pages of comparative Austroasiatic etyma, including a small number of relevant cognates in Vietnamese for kinship (e.g. *con* ‘child’), body parts (e.g. *chân* ‘leg’, etc.), and animals (e.g. *chim* ‘bird’, etc.). While positing a Tai source of Vietnamese, Maspero (1912) nevertheless provided numerous comparative word lists of basic Vietnamese vocabulary attested in several Muong dialects and what are now known to be Pong and Chut languages, as well as in Austroasiatic languages, some several dozen word meanings altogether (e.g. Maspero 1912:22, 33, 36, 44, 63, 64, etc.).

A century after Logan’s initial proposal, Haudricourt (1953, 1954b), in response to Maspero’s claims of Tai affiliation of Vietnamese, used comparative Vietnamese and Austroasiatic evidence to posit a theory of the diachronic development of tones in Vietnamese. This hypothesis of language change (i.e. from segmental to suprasegmental phonological features) in effect severed the proposed affiliation with Tai and solidly connected Vietnamese with Austroasiatic. In an ethnographic treatment of minority groups in Vietnam, Vuong (1963:133-155) presented comparative linguistic data shared by Vietnamese and other Austroasiatic languages. Thomas and Headley (1970) published statistical analysis of lexical relationships among Austroasiatic, showing most branches to have about 20 to 25 percent shared cognates, including what they called Viet-Muong (the term “Vietic” was first introduced by Hayes in 1982). Huffman (1977) summarized the history of research connecting Vietnamese to Austroasiatic and included a list of 100 basic vocabulary in Vietnamese with Austroasiatic roots.

Another twenty years after that, Hồ Lê (1992) contributed comparative evidence of possible Austroasiatic cognates for some 200-plus Vietnamese words. However, he did so without detailing phonological patterns to establish cognate pairs, nor did he fully discuss other possible etymological origins of words. This is problematic as the Southeast Asian linguistic area is well known for experiencing multiple stages of long-term language contact. Considering this, Hồ Lê’s publication must be usefully considered notes that still need sifting through historical linguistic methods and emerging comparative data (many of the items in that study are in the current study). Nguyễn Tài Căn’s 1995 treatise on Proto-Vietic phonology offers important historical linguistic insights and reconstructions of the phoneme system. Unfortunately, his work does not list the original lexical data nor provide detailed information about Austroasiatic etyma in Vietnamese.

In the last decade, I have published two studies addressing the issue of Austroasiatic in Vietnamese. In one study on loanwords in Vietnamese (Alves 2009), 1,200 items selected for their distribution in semantic domains were considered. While about 25 percent were Chinese loanwords, Austroasiatic vocabulary constituted only 10 percent of the total, only about 110 items (and Tai loans constituted only 0.5 percent), with many others of undetermined origin. One weakness is that the paper did not list original Proto-Austroasiatic reconstructions or describe phonological patterns used to verify their Austroasiatic status. Moreover, the high number of words of uncertain etymological origin demonstrates the limitations of Vietnamese etymological investigation at that time. The second paper (Alves 2017) focused on 119 basic vocabulary items but with the incorporation of reconstructions of Proto-Austroasiatic and Proto-Vietic from the MKED (see the list of acronyms in the Appendix) and data on other neighboring language groups. The percentage of Austroasiatic items identified in this study focused on core basic vocabulary was substantially higher, nearly half, with one-third Vietic or Vietnamese innovations, and about 10 percent Chinese loanwords (Alves 2017:189).

Thus, the quantity of Austroasiatic cognates identified in Vietnamese is notable, and the perception that Vietnamese has been relexified by Chinese (e.g. the often repeated claim of the Vietnamese lexicon being a massive portion of Chinese origin, but with no clear criteria for that percentage) is at least somewhat excessive and not an accurate characterization of the Vietnamese lexicon. The question is what portions of the lexicon are considered, such as basic vocabulary, daily spoken vocabulary, formal and/or written vocabulary, and so on. The semantic domains of Chinese vocabulary and those of native Vietic vocabulary are, to a good extent, complementary. This is an expected contrast between core elements of human lifestyle and culturally specific implements, practices and concepts, though with plenty of intrusion of Chinese vocabulary, even in more basic parts of the Vietnamese lexicon.

Most Chinese loanwords in Vietnamese can be readily identified, making such vocabulary prominent, even to the point of masking other etymological sources. Considering the highly visible Chinese portion of Vietnamese, much more data on Austroasiatic and Vietic etyma must be identified to present an overall realistic view. The current study substantially increases the number of proto-language items to present an updated picture of native etyma in Vietnamese. Hundreds of Proto-Vietic etyma are identified, supported with phonological and comparative evidence. This altogether portrays not a relexified language but one with a strong native core, albeit one with a surrounding but transformative quantity of Chinese loanwords.

Overall, in the last 170 years, etymological research on Vietnamese has been stunted by a lack of data and challenges in sorting out the complex mess of lexical material in the Southeast Asian linguistic area. The data on other Vietic languages has only been made available since the early 1990s, and even then, the largest quantity of lexical data was, until the early 2000s, available only for the Vietic languages Rục and Thavung. As for language contact, many of the languages in the region have tremendous typological linguistic similarities and numerous shared loanwords which spread in multiple directions due to multiple periods of language contact over many centuries. It is now possible to begin to make more substantial progress in understanding native Vietnamese etymology as well as to correct past assumptions of both historical linguistics and questions of language contact and sociocultural exchange that were based on limited data and resulting misunderstandings. The assumptions herein must be considered tentative, but I believe they offer a functional foundation and clearer direction for progress.

The rest of this paper covers (a) the intended goals of this paper, (b) a brief summary of the lexical data sources, (c) issues of historical phonology that help establish Vietnamese words as Vietic and/or Austroasiatic etyma, (d) an overview of the etymological sources, and (e) preliminary observations of the semantic domains of the several hundred identified native etyma. Finally, an Appendix provides a table of contents of the article, comments regarding the Vieto-Katuic hypothesis, and several tables of data referred to in various sections of the paper.

## **2. Reevaluating Vietic and Austroasiatic etyma in Vietnamese**

By the end of the first decade of the 21<sup>st</sup> century, massive digital lexical databases and substantive reconstructions of all major language families in southern China and Mainland Southeast Asia—Austroasiatic, Tai, Sino-Tibetan, Austronesian, Hmong-Mien—had become available. It is now possible to more effectively and efficiently assess origins of Vietnamese vocabulary and to connect these to aspects of the Vietic ancestral speech community. Crucially, in 2006, Shorto's 2,100-plus potential Proto-Austroasiatic etyma by Shorto were made available, though many of those reconstructions must be treated as tentative as Shorto himself lacked much of the data we have today. Also, in 2007, the MKED digital database with all available Austroasiatic data—including Shorto's reconstructions, numerous proto-language reconstructions of branches, and several dozen languages—was made freely available online.

The MKED also contains some 1,200 similarly tentative Proto-Vietic reconstructions by Ferlus based on data for a dozen lects, including several little-studied highly conservative varieties, many of which are posted at <https://hal.archives-ouvertes.fr/>. I have assembled a database based on Ferlus's Vietic data and have added data from nearly a dozen more Vietic languages, listed in Section 3. Also, I have evaluated the existing lexical reconstructions to exclude those lacking support for proto-language reconstruction, as well as to add probable etyma of Proto-Vietic or early Vietic and Proto-Austroasiatic or locally attested Austroasiatic. A couple hundred of these items have insufficient support for the Proto-Vietic level and must be excluded. Probable loanwords from both earlier and later periods of Chinese and Tai are noted, and occasionally, instances of regional words of uncertain etymological origin.

Finally, I have noted how many Vietic sub-groups (i.e. Việt-Mường, Pong-Cuôi, and the remaining archaic languages) have lexical reflexes for any posited Proto-Vietic word. Reconstructions with attestations in three sub-groups are considered more probable Proto-Vietic etyma, while those with reflexes in only two sub-groups can be considered ancient but with less certainty of dating to the Proto-Vietic period, but still useable to consider the ethnolinguistic circumstances of early Vietic. Those with Proto-Austroasiatic etyma, as retentions in the pre-Proto-Vietic period, can be considered reconstructable to the Proto-Vietic level, even with reflexes in only Việt-Mường. Those with reflexes in only one branch and which have no Austroasiatic source are not used in this study.

### 3. Sources of Vietic data and other comparative data

While assembling data for 24 lects of Vietic, as listed in Table 1, I focused on identifying likely reflexes among the Vietic languages. To seek additional possible etyma, I explored the la Vaughn Hayes Digital Library, which has extensive, but rather crudely and incompletely assembled, comparative data focused on Vietnamese with Vietic and Southeast Asian data. I also checked for comparable Proto-Austroasiatic etyma of Shorto and of comparable forms of words in reconstructions of Austroasiatic branches and individual languages in the MKED and periodically in the Munda Etymological Dictionary. I noted likely Chinese sources, including those from Late Middle Chinese (i.e. what is considered standard Sino-Vietnamese) and those of Late Middle Chinese or Early Middle Chinese (i.e. what is considered Early Sino-Vietnamese). The Early Sino-Vietnamese items are from my own database of several hundred likely items. The Chinese reconstructions are primarily those of Baxter and Sagart (2014). Finally, in many instances, I consulted Proto-Tai reconstructions of Li (1997) in the Proto-Tai-o-Matic and of Pittayaporn's dissertation on Proto-Tai (2009) and, only as deemed necessary, I consulted the Sino-Tibetan Etymological Dictionary and Thesaurus and the Austronesian Comparative Dictionary (Blust and Trussel 2010). Undoubtedly, in the future, additional items will be noted and added to the current working Vietic database (which I will freely share in response to requests by email), making the current number of items approximate and the observations preliminary. Nevertheless, they will hopefully offer a reasonable overview of statistics of semantic domains and ethnolinguistic inferences.

Table 1 is a list of the varieties of Vietic in the database. It is not meant to represent a phylogenetic tree nor even to make claims about the status of languages versus dialects, hence the need to use the neutral term "lects". Over the past few decades, several phylogenetic trees of Vietic have been proposed based primarily on typological features and developments in final consonants (e.g. Ferlus 1979, Sidwell 2015, Trần Trí Dõi 2018, etc.). These proposals differ in significant ways, and that topic is beyond the scope of this paper and must be dealt with in another study. For current purposes, we will simply consider the three general subgroups of Việt-Mường (including Nguồn), Pong-Cuối, and the remaining archaic varieties of Vietic.

In the process of assessing the data, I indicated the number of Vietic subgroups with reflexes of each item, that is, one, two, or three subgroups. Of the over 700 native items in Table 2, about 500 items have reflexes in two or three lects but not in Austroasiatic (at least not yet identified as such). Another 200-plus items are Austroasiatic etyma, and even with reflexes in only one Vietic subgroup (nearly 10 items), these must be considered native lexical retentions. Beyond the core 700 native items, about 120 items which Ferlus reconstructed are loanwords, most (over 100) from Chinese but also some from Tai languages. Again, these numbers will increase, and the relative percentages will change as data is further processed.

**Table 1:** *Vietic lects used in this study*

- Vietnamese (Hanoi)
- Mường lects
  - Mường Hòa Bình (Ferlus 2007)
  - Mường Sơn La (Ferlus 2007)
  - Mường Thanh Hóa (Ferlus 2007)
  - Mường Bi (Ferlus 2007)
  - Mường Bi (Nguyễn Văn Khang, et al. 2002)
- Nguồn lects
  - Nguồn Cổ Liêm (Nguyễn Phú Phong 1997)
  - Nguồn Yên Thọ (Nguyễn Phú Phong 1997)
- Cuối lects
  - Thổ [Cuối Chăm] (Ferlus 2007)
  - Thổ [Làng Lỡ] (Ferlus 2007)
  - Cuối Thái Hòa (Nguyễn Hữu Hoàn, unpublished)
  - Cuối Tân Hợp (Nguyễn Hữu Hoàn, unpublished)
- Phong lects
  - Phong (Ferlus 2007)
  - Tourn (Ferlus 2007)
  - Liha (Ferlus 2007)
- Archaic Vietic Languages (i.e. polysyllabicity)
  - Chứt lects
    - Sách (Ferlus 2007)
    - Mây (Babaev and Samarina 2019)
    - Rục (Ferlus 2007; Nguyen, Tran & Ferlus 1988; Nguyen Van Loi 1993)
    - Arem (Kasuga 2008)
  - Other archaic lects
    - Thavung lects
      - Thavung (Ferlus 2007)
      - Thavung Phon Soung (Ferlus 2007)
      - Số (Thavung) (Suwilai 1998, 1999)
    - Malieng lects
      - Malieng (QB) (Ferlus 2007)
      - Malieng (HT) (Ferlus 2007)
    - Kri (HT) (Enfield and Diffloth 2009)

**Table 2:** *Vietic Etyma according to etymological sources and number of subgroups with reflexes*

<b>Etymological Sources</b>	<b>Tentative Numbers of Items</b>
Austroasiatic in 3 Vietic subgroups	About 155 of probable Proto-Austroasiatic origin (about 20 in local AA languages)
Austroasiatic in 2 Vietic subgroups	Nearly 50 of probable Proto-Austroasiatic origin (plus several in local AA languages)
Austroasiatic in 1 Vietic subgroup	9 of probable Proto-Austroasiatic origin
Vietic in 3 subgroups	Nearly 300
Vietic in 2 subgroups	Over 210
Chinese and Tai loans	About 120 (all are attested in 2 or 3 Vietic sub-groups)

#### 4. Issues of Vietic and Austroasiatic historical phonology

We next consider aspects of Vietic and Vietnamese historical phonology. The goal here is not to make a complete statement of the historical phonological changes from Proto-Vietic to Vietnamese. Rather, some general phonological patterns must be considered to support claims that certain Vietnamese words belong to Vietic and/or Austroasiatic.

#### 4.1 Historical periods from Austroasiatic to Vietnamese

To provide context of the historical linguistic connection between Vietnamese and Austroasiatic, hypothesized stages of developments are offered in Table 3 (comparable to those presented in Trần 2016:3-6). Maspero's 1912 monograph is sometimes cited for historical stages of Vietnamese (e.g. protoannamite (i.e. Proto-Vietnamese), Annamite moyen (i.e. Middle Vietnamese), etc.). The new data and concepts herein require a new approach, one not centered on Vietnamese in the stages of Proto-Việt-Mường and before. These include proposed periods from Austroasiatic to modern Vietnamese and very approximate dates based primarily on archeological data and dates of archaeological cultures and broad periods in the historical era.

At each stage, certain linguistic typological features are hypothesized, such as polysyllabicity versus monosyllabicity, and the presence of tones (putting aside the more complex issues of phonation systems). Other phonological and lexical developments must have occurred that further differentiated the various speech communities. For example, Vietnamese, Mường, and Nguồn share a great deal of vocabulary not seen in the other Vietic languages (not statistically verified, but evident from the massive quantity of shared vocabulary seen in the hefty Mường Bi dictionary), and at least some of this represents lexical innovations in Việt-Mường. However, we cannot yet demonstrate what kinds of phonological distinctions may have developed differentiating Vietic in the northern region around the Red River Delta and groups farther south in north-central Vietnam. Details await further accumulation and assessment of data.

**Table 3:** Tentative stages from Austroasiatic to modern Vietnamese

Period	Timing	Hypothesized typological traits
<i>Austroasiatic dispersal</i>	c. 2,000 BCE	polysyllabic, nontonal
<i>Proto-Vietic as distinct speech group</i>	c. 1,000 BCE	polysyllabic, nontonal, emerging phonation (?)
<i>Heavy Sinitic-Vietic contact</i>	c. 1 <sup>st</sup> mill. CE	polysyllabic, nontonal, emerging north-south distinction (?)
<i>Việt-Mường</i>	c. 1,000 CE	reduced polysyllabicity (?), tonal (pre-register)
<i>Vietnamese</i>		
<i>Archaic</i>	early 1 <sup>st</sup> millennium CE	reduced polysyllabicity, tonal (post-register ??)
<i>Middle</i>	mid 1 <sup>st</sup> millennium CE to the 1800s	monosyllabic but with initial clusters, tonal
<i>Modern</i>	late 1 <sup>st</sup> millennium CE	monosyllabic, mostly lost clusters, tonal

The hypothesized periods in Table 3 can be accounted for as follows. First, no claims of the earliest period of Austroasiatic are made. Instead, the timing of the dispersal of Austroasiatic groups is here based on the predominant view in numerous archaeological studies of the “Two-Layer” hypothesis (e.g. Koenigswald 1952, Matsumura et al. 2019, inter alia). Crucially, this involves the arrival of Neolithic agriculturalists from Southern China encountering hunter-gatherer groups in the region of northern Vietnam (Higham 2017a, etc.).

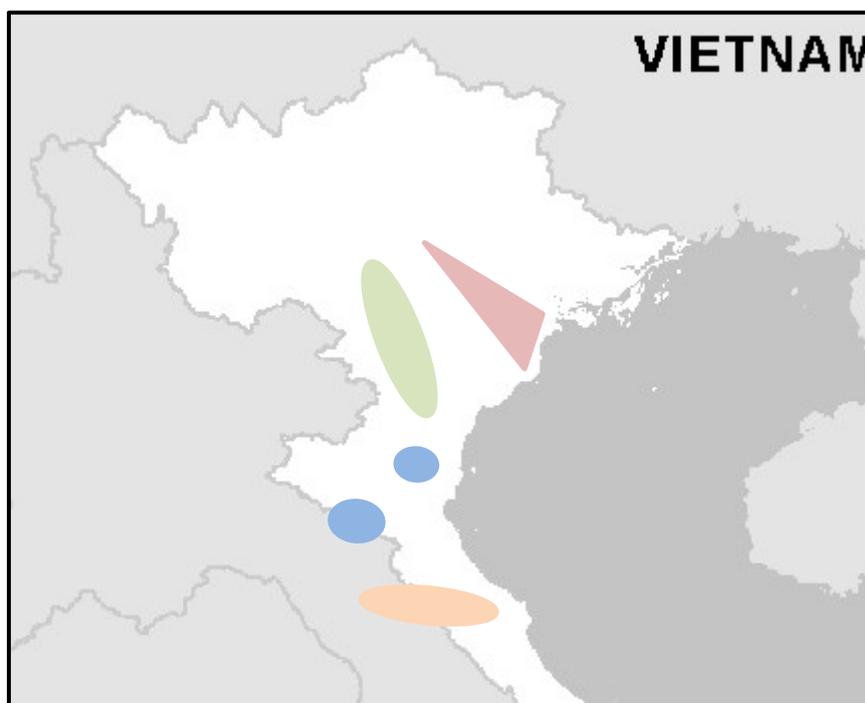
The topic of the locus of the Austroasiatic homeland is of considerable debate as seen in the variety of hypotheses. For now, we accept as reasonable Bellwood's (20004:22) hypothesis of an early Austroasiatic in part of Southern China (Yunnan and Guangxi) and bordering northern parts of Mainland Southeast Asia (Northern Vietnam and Laos). It is in line with the “Two-Layer” hypothesis, noted above. Also, recently, an international team of three dozen researchers (including some from the Institute of Archaeology in Hanoi) of a recent archaeogenetic study claim, “Our results provide genetic support for the hypothesis that agriculture was first practiced in mainland Southeast Asia by (proto-)Austroasiatic-speaking migrants from southern China” (Lipson et al. 2018:94). This conclusion is similarly supported by the scenario of Aslian groups in Malaysia: the original hunter-gatherer Negrito groups shifted to Austroasiatic, the agriculturalists' language. However, determining the origins of Austroasiatic is not the goal of this paper, and these ideas are only presented to provide context for the discussion.

Next, as with the dispersal of Proto-Austroasiatic, the differentiation of Vietic as a branch of Austroasiatic cannot yet be precisely timed based on linguistic data. Instead, we can consider archaeological

evidence of the progression of several stages from the Phùng Nguyên (2000-1500 BCE) to the Đông Sơn (500 BCE-200 CE) archaeological cultures in northern Vietnam (Kim 2015, Hà Văn Tấn 2018, etc.). It has been claimed that the Bronze Age begins in the Đông Đậu culture (1500-1000 BCE). If so, this marks sociocultural change and allows for a possible distinct Vietic speech community around the end of the 2<sup>nd</sup> millennium BCE.

The debate regarding the homeland of Vietic is not addressed here, but sufficient archaeological and archaeogenetic data makes connections between Vietic and the Red River Delta region (e.g. Alves 2019b, Trần 2019). Vietnamese has a lengthy presence in the Red River Delta and is today spoken throughout Vietnam. The modern distribution of Vietic languages other than Vietnamese extends to Thanh Hóa, Nghệ Tĩnh, and bordering areas of Laos, as in Figure 1 (roughly adapted from Ferlus 2014). We assert that the northern region today represents the early geographic extent of Vietic, though archaeological data offers no details of the history of Vietic outside of the Red River Delta. The significant amount of linguistic diversity in the southern region has been posited to be an indicator of the Vietic homeland in this southern region (Chamberlain 1998). But of course, such a scenario is also the case for Sinitic languages, with tremendous diversity in southern China, even though Sinitic has a solidly established northern homeland. Thus, at this point, we follow the archaeological evidence and assume that at the very least, Vietic was spoken.

**Figure 1:** Vietic subgroups in the northern portions of Vietnam and bordering parts of Laos



**Key:** Red = Red River Delta; Green = Muong; Blue = Phong & Cui; Orange = Archaic languages

Whether sub-branching in Vietic occurred prior to the arrival of northerners from China cannot yet be proven with linguistic data. Nevertheless, we can assume (a) linguistic diversity was gradually increasing from the earliest period of Proto-Vietic and (b) two periods of early language contact intensified linguistic differentiation among Vietic groups. First, archaeogenetic data, including studies of genetic material from ancient human remains, suggests likely cultural contact between Tai and Vietic groups in the Red River Delta during the Đông Sơn period (McColl et al. 2018). Lexical data in this study (Section 5.2 and various places in Section 6) further supports early Tai-Vietic contact, but earliest in the more northern part. Second, there were large settlements of Sinitic speakers and intense Sinitic-Vietic language contact during the Han Dynasty (e.g. Taylor 1983, etc.), the East Jin Dynasty (e.g. Phan 2013, etc.), and the Tang Dynasty after the Huang Chao / 黃巢 / Hoàng Sào Rebellion in the early 900s. The incoming Chinese-speaking populations were probably a significant factor in the linguistic differences between Vietic speakers of the Red River Delta and those to the south (cf. Ferlus 1975:22).

As noted, available linguistic evidence cannot yet help to posit a specific time in which Việt-Mường became a distinct dialect group and speech community. Some potential periods include (a) the end of

Chinese administrative rule in the 10<sup>th</sup> century CE, (b) the early centuries CE of East Han and Jin dynasty migrations, or (c) centuries earlier during BCE-era Tai-Vietic contact. All the events in these periods of change in language contact situations may have contributed to gradual differentiation of southern and northern Vietic. The generally accepted view of the linguistic differentiation of the Việt-Mường from the rest of Vietic was by around the 1<sup>st</sup> millennium CE. This is assumed to be the time Chinese-speaking communities were absorbed into Việt-Mường via language shift (cf. Phan 2013, Taylor 2013:29). It seems the shift towards monosyllabicity spread out over a period of centuries, with some remnants of presyllables in the early 1<sup>st</sup> millennium CE (Shimizu 2000, Shimizu, Lê, and Shiro 2005) and the last of initial clusters into the 1800s (Shimizu 2015, Vu 2019). After that, Vietnamese itself emerged as a distinct variety, eventually splitting into numerous dialectal varieties as Vietnamese speakers migrated southward.

Vietnamese is the most typologically transformed language of Vietic, and yet, it has retained a sizable native lexical core as well as a large quantity of proto-Vietic phonological segments, as discussed in Section 4.2.

#### 4.2 Phonological changes from Austroasiatic and Vietic to Vietnamese

We now consider how Vietnamese consonants, vowels, and tones can be traced to Proto-Vietic sounds in Ferlus's Proto-Vietic reconstructions. Regarding phonological changes from Vietic, the main typological factors are (a) loss of presyllables, (b) loss of initial clusters, (c) loss of final fricatives and glottal stop (or phonation features of the syllable), (d) the development of tones, and (e) the development of diphthongs. Though Ferlus's Proto-Vietic reconstructions and Shorto's Proto-Austroasiatic reconstructions were done separately, there is substantial overlap in segmental features. This allows us to focus on the relationships between Vietnamese and Proto-Vietic, rather than with Proto-Austroasiatic, while still showing Austroasiatic data with which Vietnamese phonology shows patterns of correspondences.

The process of the loss and development of features spread over many centuries, and some lingered until relatively recently. Vietnamese retained some archaic features long after it became differentiated from other Việt-Mường languages. Initial consonant clusters were present in Vietnamese well into the 1700s (e.g. Vu 2019), and thus only relatively recently were they completely lost (except for clusters with the /-w-/ medial glide, as in /lw/, /tw/, and so on). The timing of all these changes is complex and must be dealt with in other studies.

Nguyễn Tài Cẩn (1995) has tracked modern Vietnamese consonants and vowels back to reconstructed sounds of Middle Chinese and his own reconstructed version of the Proto-Vietic phoneme inventory. However, rather than rely on or attempt to replicate his work, we here summarize the patterns of phonological retentions and changes from Vietic to Vietnamese based on Ferlus's reconstructed Proto-Vietic etyma in the MKED, with samples in Tables A, B, and C of the Appendix. An explanation of these complex sets of changes, which vary tremendously throughout the typologically diverse Vietic languages, will require a large-scale comparative study.

**Table 4:** Common changes in initials from Proto-Vietic to modern Vietnamese

• Retentions (e.g. *k > /k/, *h > /h/, *m > /m/ etc.)
• Changes in voicing (e.g. *p > b, etc.)
• Reductions of clusters/complex initials with stops leading to affrication and lenition (e.g. *k <sup>h</sup> > /x/, *Ck > /ç/, etc.)
• Losses of presyllabic material before sonorants (e.g. *Cl > /l/, *Cm > /m/, etc.)
• Nasalization of earlier implosives (e.g. *b > /m/, etc.)
• Several mergers of Proto-Vietic initials into single consonants in Vietnamese (e.g. the four Proto-Vietic initials *ŋ, *C-ŋ, *ml, *f > /ɲ/, etc.)
• Splits and diphthongization of vowels (e.g. *a > /uaə/, *ɛ > /ɛ/ & /iə/, etc.)

The overall tendency is that (a) many proto-language phonemes have been retained, but (b) the Vietic system of initials has been restructured and decreased through many mergers in Vietnamese, and (c) Vietic vowels have increased in number, partly due to diphthongization. The most complex type of changes involved alveolar initials such as Vietic \*s as well as a number of Sino-Vietnamese alveolars that merged with /t/. The general types of changes from Vietic to Vietnamese among initials include those in Table 4. A detailed list of these changes with examples is provided in Table A of the Appendix.

Word-final consonants variously show retentions (e.g. \*-k > /-k/, \*-p > /-p/, \*-m > /-m/, etc.), changes (e.g. \*-c > /-t/, \*-l > /-j/, etc.), and loss of segments with rephonologization of suprasegmental phonemes/tonemes (e.g. \*-ʔ > -Ø plus *sắc/nặng* tones and -h > -Ø plus *hỏi/ngã* tones), as per Haudricourt's (1954b) hypothesis of tonogenesis in Vietnamese. Whether these were full glottal stop codas or instead glottalization as a syllabic feature is uncertain, and other related matters are still debated. There are instances glottalization which are not reconstructed with glottal features in Austroasiatic. These appear to represent developments of syllabic phonation in Vietic, such as those considered by Diffloth (1989).

As for vowels, the original Proto-Vietic monophthongs developed in Vietnamese as a system of monophthongs, diphthongs, and monophthongs plus final off-glides. Monophthongs in Vietnamese are mostly retained from Vietic, with some minor changes. As Vietic has been reconstructed without diphthongs, all Vietnamese diphthongs necessarily come from monophthongs. Diphthongization is a widespread but extremely complex diachronic phenomenon among Austroasiatic languages in relation to registrogenesis (e.g. Gehrman 2015 regarding the phenomenon in Katuic).

The patterns of changes of the consonants noted above are generally shared by both Ferlus's Proto-Vietic and Shorto's Proto-Austroasiatic, which supports in general the connection between Vietnamese vocabulary to their proposed Vietic and Austroasiatic etymological sources. Nevertheless, all of these listed phonological changes must be considered tentative pending more data and additional analysis of the accumulated comparative data.

Having presented phonological correspondences between Austroasiatic, Vietic and Vietnamese, we explore etymological sources, including both native etyma and loanwords from Chinese and Tai.

## 5. Etymological sources of Proto-Vietic and early Vietic

Nearly 60 percent of the Vietnamese words considered in this study are here evaluated to be Proto-Vietic items; that is, they are not in Proto-Austroasiatic nor seen in neighboring Austroasiatic languages. In some cases, Vietic words are found in neighboring Austroasiatic branches, such as Katuic, Bahnaric, Khmeric, or Khmuic. However, these local Austroasiatic items are not widespread enough to reconstruct to Proto-Austroasiatic and may be the result of regional innovation and spread. Vietnamese loanwords into those languages generally represent recent language contact and are excluded from this study.

Some Chinese and Tai words have also been widespread enough to reconstruct in early Vietic, though they are obviously not Proto-Vietic items. Instead, they demonstrate that these loanwords spread widely among early Vietic speech communities. Of these widespread early loanwords in Vietic, a solid majority are from Chinese (about 12 percent of the 800 items), while another smaller percentage appear to be likely early Tai loanwords (1 percent). These are discussed more in Sections 5.1 and 5.2.

*Table 5: Etymological sources of Vietnamese words in the reconstructed data*

Language group	Quantity (% of total)
Vietic	520 (60.1%)
Proto-Austroasiatic	186 (22%)
Neighboring Austroasiatic	28 (3.2%)
Late Old Chinese/Early Middle Chinese	87 (10.2%)
Late Middle Chinese	18 (2.1%)
Tai	13 (1.5%)
TOTAL	c. 853

The numbers in Table 5 give an indication of the etymological distribution in Vietnamese. Vietic vocabulary constitutes nearly six out of ten of the items. About 23 percent are of Austroasiatic origin, with another five percent attested in neighboring Austroasiatic languages, altogether making up over one quarter of the data. Several Proto-Austroasiatic (PAA hereafter) etyma have possible comparable forms in multiple language families (e.g. *này* (PAA \*niʔ, \*nih) 'this', *vác* (PAA \*ʔɔʔ) 'carry on back/shoulder', *bóc* (PAA \*pɔk) 'to peel', etc.) and may either represent deeper linguistic origins or are the result of regional spread. Such words are treated here as PAA etyma, though they hypothetically may indicate a larger shared language family. The various multi-language family hypotheses of Austro-Tai (Benedict 1942), Austric (cf. Reid 2005), Sino-Austronesian (cf. Sagart 2005), and Proto-East-Asian (the latter encompassing all five of the

modern language families (Starosta 2005)) are all of considerable debate, and as such cannot be considered useable points of reference for this paper.

### 5.1 Chinese loanwords in early Vietic

The word “Chinese” is a highly ambiguous term. For this study, it refers broadly to languages belonging to the Sinitic branch of Sino-Tibetan at various historical stages. Vietic words of likely Chinese origin include those from Late Middle Chinese (LMC) and from Late Old Chinese (LOC) or Early Middle Chinese (EMC), which are most of the Chinese loanwords considered in this study. In Vietnamese, the former loanwords from the later period are standard Sino-Vietnamese (i.e. those listed in Sino-Vietnamese dictionaries as Chinese character readings), while the latter belong to the stratum of Early Sino-Vietnamese (ESV). The ESV words were borrowed throughout the first millennium CE, with some possibly borrowed in the BCE period. Though borrowed several centuries to a millennium after the stage of Proto-Vietic, some Chinese loanwords in conservative Vietic languages, such as Rục, appear to have spread early enough to retain evidence of presyllabic material in OC. As a result, they have been influential in OC reconstructions (cf. Baxter and Sagart 2014a).

Loanwords from LMC are evidence of the continued language contact among Vietic speech communities and possibly with Annamese Chinese, Phan’s (2013) hypothesized Chinese speech community in northern Vietnam in that period, after linguistic differentiation of Vietic subgroups. Careful study of the phonology of these words among Vietic languages is needed to determine which words were genuinely spread in the early period of Sinitic-Vietic language contact or in later periods as loanwords from early Việt-Mường or modern Vietnamese. However, as the number of these is relatively small, at this point, there is little that can be gleaned from them in terms of understanding language contact and the history of their spread throughout Vietic.

Of the 80-plus items from the LOC to EMC period in the first several centuries of the 1<sup>st</sup> millennium CE, about 70 percent are nouns. Only about a dozen are verbs, several are adjectives, and only two are function words. The latter two are quantity expressions: *nhieu* ‘much/many’ (LSV *nhieu*) and *đôi* ‘a pair’ (LSV *đôi*) both of which have standard literary Sino-Vietnamese (LSV) pronunciations with different tones following patterns of other ESV items. However, as Vietnamese itself has a much larger number of ESV grammatical loanwords (e.g. Lê 2002), we can assume the amount of Sinitic-Vietnamese bilingualism was much more intense and qualitatively different from that of other Vietic groups in the region leading up to the formation of Proto-Việt-Mường. Still, no studies yet exist that have identified Early Sino-Mường loanwords to clarify this matter.

LOC and EMC loanwords throughout Vietic include several main semantic domains: agriculture (a dozen items), general trade items (several), animals for trade (several), clothing (several), kinship terms (several), metals and related color terms (several), and internal organs (a few). In the compound *buôn bán* ‘commerce / trade’, both *buôn* and *bán* are both possible ESV forms of Chinese 販 *fàn* ‘to sell’ (LSV *phiến*, *phán*), clearly highlighting trade as a part of the impact. Overall, while loanwords related to trade and food production are dominant, the domains of household culture and marriage were also part of the impact in Vietic communities of early language contact with Sinitic speakers.

Some of the Chinese loanwords (e.g. *đực* ‘male (of animals), *bạc* and *ngân* ‘silver’, etc.) have also been borrowed into Tai languages (e.g. Proto-Tai \**thuk*<sup>D</sup> ‘silver’, Proto-Southwestern Tai \**ŋəʔn*<sup>A</sup> ‘silver’ (Proto-Tai-o-Matic)). This shared language contact with Chinese has resulted in confusion about the historical connection between Tai and Vietnamese. In response to Maspero, Haudricourt (1954a) pointed out many shared Chinese loanwords in Vietnamese and Thai.

Unless evidence can demonstrate otherwise, we must assume that Chinese (or rather the ancestral Sinitic language group in the early period) was the direct donor language separately to Vietic and Tai, not that these were Chinese loanwords borrowed from Tai into Việt-Mường or Vietnamese. This point is supported by the hundreds of early Chinese loanwords in Vietnamese that are not in Tai languages, which is true in a large majority of cases. The only explanation is direct borrowing from Chinese into Việt-Mường languages, not through Tai.

Chinese influence goes beyond Tai and Vietnamese. Tai has borrowed the Chinese numerals from 2 to 99, and the decimal numerals (e.g., 20, 30, etc.) subsequently spread into Khmer (while Vietic languages have retained native numeral terms, as noted in Section 6.4). A few Chinese words in Vietic have also made their way throughout Austroasiatic languages in Mainland Southeast Asia, and Tai as well, including ‘hat’,

‘shirt/upper garment’, and ‘bean’ (cf. Alves 2018). The details of the historical spread of such words are not yet well studied.

Table D in the Appendix presents examples of the Chinese loanwords in Vietic using the OC and MC reconstructions of Baxter and Sagart (2014a). However, as the Baxter and Sagart reconstructions lack substages (e.g. LOC, etc.), they are only general points of reference and admittedly require additional consideration and some speculation of smaller stages of phonological developments. But overall, the phonological patterns of initials, vowels, codas, and tones demonstrate these are highly likely to be ESV items, as described in various sources (e.g. Wang Li 1948, Phan 2013, Alves 2016, etc.).

As for widespread LMC loanwords, a majority of those reconstructed for Proto-Vietic in Ferlus 2007 are items of trade (e.g. *ách* ‘yoke’ from 軛 *è* ‘yoke’ (MC *eak*); *bao* envelope/bag from 包 *bāo* ‘a wrap/bundle’ (MC *paew*); *câu* ‘hook/fishhook’ from 鈎 *gōu* ‘hook’ (MC *kuw*); *đồng* ‘copper/bronze’ from 銅 *tóng* ‘copper’; *đường* sugar/cane sugar from 餉 *táng* ‘sugar’ (MC *dang*); *ngân* ‘silver’ from 銀 *yín* ‘silver’ (MC *ngin*), etc.). Others demonstrate the spread of Chinese cultural practices (e.g. *khách* ‘guest’ from 客 *kè* ‘guest’ (MC *khaek*); *ma* ‘ghost/spirit’ from 魔 *mó* ‘devil/magic’ (MC *NA*); *ông* ‘grandfather/old man’ (it often means ‘husband’ in Vietic languages) from 翁 *wēng* ‘elderly man/father-in-law’ (MC *qung* (Karlgren))). As these are later loanwords, likely after the merge of Annamese Chinese with Vietnamese, it seems more likely that they were borrowed from Vietnamese (or Việt-Mường more broadly) rather than a variety of Chinese.

### 5.2 *Tai loanwords in early Vietic*

The list of possible Tai loanwords in Proto-Vietic in Table E in the Appendix contains only nine items. As noted in 5.1, previously, many Chinese loanwords in both Tai languages and Vietnamese were mistakenly considered Tai loanwords. Moreover, various past publications positing Tai loanwords in Vietnamese have often not followed strict historical linguistic methods of identification, and the words noted often have problematic phonological correspondences, or the studies give insufficient attention to regional comparative data (cf. discussion in Alves 2006). Also, Vietic languages other than Vietnamese have additional loanwords from various Tai languages, but these are not at the Proto-Vietic level and came in later periods. I assume more valid early Tai loanwords will be identified, but many previous publications’ claims cannot be accepted until their data are given proper application of the historical-comparative method and compared with all available modern linguistic data in the region.

Of the words in Table E, it is not always certain that the direction of borrowing was from Tai into Vietic, or from or into Austroasiatic. In some cases, when the geographic distribution in Austroasiatic is more scattered with apparent different periods of borrowing, such as for *rũa* ‘bush-knife’, the direction of Tai-to-Vietic appears more likely. As for timing, it is reasonably possible that waterway terms (*bè* ‘raft’, *muong* ‘ditch/canal’) and food production terms (*muống* ‘water spinach’, *vịt* ‘duck (domestic)’) were borrowed in an early period, if not during the Đông Sơn BCE-era, though the phonological aspects require more data to identify patterns that clarify the matter.

A few other Tai loanwords have an ancient history in Vietic (e.g. Proto-Vietic *\*bə:ʔ* ‘daughter-in-law’ from Tai *\*bəu<sup>c</sup>* (Li), Proto-Vietic *\*pa:ʔ* ‘father’s elder brother’s wife / father or mother’s elder sister (bá (dialect))’ from Tai *\*pa<sup>c</sup>* (Li)) but are not in mainstream Vietnamese. Such data highlights differing language contact situations—and periods of contact—of Tai with Vietic to the north versus with Vietic to the south.

### 5.3 *Native etyma*

Despite the sizeable number of early Chinese loanwords in Vietnamese, Vietnamese vocabulary still retains a significant number of native Proto-Austroasiatic and Vietic etyma. This becomes even more evident in looking at the wide range of semantic domains of Proto-Vietic vocabulary in Section 6. Yet the balance of native versus borrowed vocabulary raises several points to consider.

One question is how to view the relationship between native and borrowed vocabulary. Vietnamese can hardly be considered a relexified Chinese dialect if several hundred words (and undoubtedly hundreds more yet to be identified) are native Austroasiatic and Vietic words. However, these tend to be more basic in nature or part of rural settings, while Chinese loanwords are statistically dominant in more urbanized settings with literacy as a dominant cultural feature. This situation has contributed to the view of a Chinese-dominant language, but the substantial native portion cannot be simply ignored: a new characterization is needed.

A related matter is the concentration of proto-language vocabulary that is primarily related to groups living in rural areas. The tremendous range of lifestyles among the various Vietic groups—from cities to small hunter-gatherer groups (cf. Chamberlain 1998)—necessarily results in a situation in which the shared proto-level vocabulary consists of relatively more culturally generic vocabulary, the cultural common denominator. But is this the only type of vocabulary in early Vietic? Proto-Vietic vocabulary naturally includes that of the Neolithic period, but for later periods of Vietic in the pre-Qin Bronze Age period, other lexical developments are expected. Recent archaeological, historical, and linguistic evidence supports the hypothesis that Vietic was a speech community in the Đông Sơn culture (cf. Trần 2019, Alves 2019). Also, archaeological evidence suggests a state-level society at the Cỗ Loa archaeological site (Kim 2015). Many of the words retained in Vietnamese (and likely Mường lects as well) from the Đông Sơn era of the proto-urban Cỗ Loa site would generally not be part of the lexicons of small Vietic groups living in forests. However, it may not be possible to certify that some Việt-Mường words not seen in other Vietic languages date to the Đông Sơn era. Recall that in Table 3, the Việt-Mường stage is listed as c. 1000 CE, a millennium later than the period just discussed

Also, how can we distinguish the introduction of lexical replacements from the introduction of entirely new words? A clear instance of the likely retention of a pre-Qin word is *rìu* ‘axe’ (PV \*m-ri:w), an object which is well represented in archaeological sites in the Đông Sơn and earlier Bronze Age periods. Archaeological evidence makes it possible to show when Chinese loanwords may have been additions to previous words. Early Sino-Vietnamese such as *đồng* ‘bronze’ and *tên* (LSV *tiên*) ‘arrow’ have no native counterpart in Vietnamese, and yet, neither represent concepts introduced after the establishment of Chinese communities in the Red River Delta. The complete replacement of a native word for ‘bronze’ by the Chinese etymon in a central locus of the pre-Qin Đông Sơn bronze drums and bronze arrowheads is a significant lexical loss. In other cases, there are no archaeological items corresponding directly to words for actions or concepts, and thus no way to know what lexical items have been replaced by Chinese loanwords.

Archaeological data of the Đông Sơn era clearly suggests social stratification and related sociopolitical developments, but only in a general way. As for the linguistic data, the semantic domains of Vietic vocabulary support some of the expected sociocultural elements of a more complex sociopolitical system. Some evidence of economic differentiation is attested in the Proto-Vietic lexicon (see Section 6.3). Support for a large community is supported by terms of rice production: Proto-Vietic has a substantial core of reconstructed items, including many lexical innovations beyond those in Proto-Austroasiatic (see Section 6.9). With this large rice-production lexicon of Proto-Vietic, it is reasonable to expect large-scale agricultural production necessary for the building of the Cỗ Loa site (e.g. Kim 2015:214, 222). Other thoughts on the matter of timing words in Vietnamese but not attested in other Vietic languages are provided in the conclusion.

Section 6 covers several semantic domains in more detail. As mentioned, in the Appendix, there is a table of contents listing the sections and subsections.

## 6. Overview of semantic domains of native etyma in Vietnamese

The statistical distribution of parts of speech of native etyma in Vietnamese in Table 6 appears reasonable. Nouns constitute over half, with verbs, adjectives, and function words showing decreasing percentages. These percentages contrast with Chinese loanwords, of which, as noted in Section 5.1, 70 percent are nouns, with verbs and adjectives a total of just over 25 percent.

**Table 6:** Parts of speech of Vietic and Austroasiatic etyma

Part of Speech	Numbers and Percentages
Nouns	c. 360 (51.7%)
Verbs	c. 220 (31.6%)
Adjectives	c. 87 (12.5%)
Function words	c. 27 (3.8%)
Total	c. 695

In Table 7, the semantic domains are grouped by areas of semantic domain overlap, such as kinship with social and political relations or quantity, spatial relations, and time. The semantic domains here are based on the WOLD (World Loanword Database at <https://wold.clld.org/>). No domain lacks lexical items, from the

natural environment to society, to conceptual/perceptual notions. However, these categories provide only very broad sociocultural context and lack additional semantic field subgroups, which are instead considered in more detail in subsections of Section 6.

**Table 7:** *Semantic domains of Vietic and Austroasiatic etyma*

Semantic Domains	Numbers
Agriculture and vegetation	82
Animals	103
Food and drink	37
Basic actions and technology	129
Motion	31
Physical world	37
Body	97
House	9
Clothing and grooming	14
Kinship	25
Social and political relations	4
Warfare and hunting	5
Possession	9
Emotions and values	9
Cognition	6
Speech and language	4
Sense perception	28
Quantity	18
Spatial relations	28
Time	18

While this makes up only a small portion of the total original lexicon of Proto-Vietic, these several hundred items can still offer a glimpse into the material culture and lifestyle. There are many gaps in the lexical data for various reasons: limits of available data, the impact of differing lifestyles and hence differing lexicons among the Vietic groups, changes from earlier cultural practices and earlier natural environments, and so on. Evidence of early Vietic culture (with likely differing cultural subgroups depending on socioeconomic status and location) can only be more fully characterized through more ample lexical data, archaeological studies, and modern ethnographic evidence. Still, the current lexical data can and should be utilized.

In the subsections of Section 6, I make observations about the types of semantic domains of the current data. These include ethnolinguistic notes combined with information gleaned from archaeological information and historical records. These ideas must be considered starting points for discussion as more data is gathered. The lexical data gathered from Vietic groups is far from complete, and without historical records of those groups, there is no way to even speculate about those groups' histories of movement or changes in sociocultural and sociopolitical practices. We can, therefore, only consider a range of general hypotheses, and over time, as new information is accumulated, these hypotheses can be supported, re-evaluated, modified, and/or refuted.

### 6.1 *The physical world, body parts, and colors*

Core types of basic vocabulary in Proto-Vietic include those of the physical and natural world, body parts, and color terms. While the first two categories are solidly native Vietic or Austroasiatic, Proto-Vietic color terms consist of only three items and show the influence of language contact.

A few dozen Proto-Vietic etyma, of which nearly ten have Austroasiatic roots, are of **the physical world**. These primarily include parts of the natural environment (e.g. *hang* (PV \*ha:ŋ) 'cave', *sao* (PV \*k-ra:w) 'star', etc.), natural materials and substances (e.g. *vỏ* (PV \*-pəh, PAA \*pus) 'bark/shell', *mun* (PV \*p-lə:) 'ashes', *sỏi* (PV \*k-rə:s) 'gravel', etc.), and the weather (e.g. *sét* (PV \*p-rə:t) 'lightning', *gió* (PV \*k-jə:ʔ > kʰjə:ʔ, PAA \*kjaal) 'wind', *mây* (PV \*k-məl, PAA #Cmil) 'cloud', etc.). Note that an asterisk \*

means that a reconstruction is that of Shorto, while the hashtag # indicates a tentative reconstruction I have made based on available data.

While several Early ESV **body part terms** have been borrowed, several dozen native items have been reconstructed (including conditions related to the body, such as *điếc* ‘deaf’ (PV \*dɛ:k, tɛ:k)). These total about two dozen Proto-Austroasiatic etyma, while the remaining dozens are Proto-Vietic etyma. They include some 45 terms for external body parts and over fifteen terms for internal organs. This solid native lexical core is robust data connecting Vietnamese to its linguistic origins.

In contrast, only three **color terms** are reconstructable to the Proto-Vietic level: *trắng* (PV \*k-laŋʔ) ‘white’ (likely from ‘white of the eyes’), *đen* (PV #tɛ:n) ‘black’, and *đỏ* (PV \*tɔh) ‘red’. The color system of Vietnamese has been significantly impacted by early lexical borrowing from Chinese of both color terms and terms for metals (Alves 2019a), such as *vàng* (黃 huáng ‘yellow’, OC \*N-k<sup>w</sup>ʰaŋ, MC hwang, LSV hoàng) ‘yellow/gold’ and *bạc* (白 báí ‘white’, OC \*b<sup>h</sup>rak, MC baek, LSV bạch) ‘white/silver’ (Alves 2016:265-266). Terms for gold and silver likely spread in the Han Dynasty (a period of the spread of gold and silver in Mainland Southeast Asia (Reinecke 2015)). Over subsequent centuries, additional Chinese color words were incorporated, which ultimately impacted the semantic structural system of Vietnamese color terms.

## 6.2 General actions and conditions

Words for **general actions and descriptions** are well represented, with about 200 reconstructed items. As these are basic and/or culturally general words, they further highlight the linguistic affiliation between Vietnamese and both Vietic and Austroasiatic. However, as vocabulary of a more basic nature, they are not particularly telling of early Vietic culture and instead represent the general human experience. General actions include semantic subdomains such as actions of manipulation of objects, motion, grooming, cognition and emotion, among others, as exemplified in Table 8.

*Table 8: Numbers and examples of Vietic words for general actions*

Semantic Domains	Quantities	Examples
general actions	several dozen	lụt (PV *p-lu:t) ‘to flood’ nhỏ (PV *k-fɔh) ‘to drip’ trượt (PV *b-la:t) ‘to slip’
manipulation of objects	60+	vả (PV *t-pah) ‘to hit with hand/slap’ vặn (PV *vaŋʔ, PAA *wɨŋ) ‘to twist/wring’ bóc (PV #pɔ:k, PAA *pɔɔk) ‘to peel (something)’
motion	30+	chạy (PV *jalʔ, PAA *jarʔ) ‘to run’ trở (PV *p-ləh) ‘to return/give back’ sá (PV *k-ra:ʔ > k <sup>h</sup> la:ʔ, PAA #kra:ʔ) ‘way/path’
sounds and actions of animals	several	sủa (PV *k-rəh, PAA #-roh/rəh) ‘to bark’ gáy (PV *t-karʔ) ‘to crow (cock)’ ve vẫy (PV *vasvas, PAA *was ‘to wag (as of a tail)’)’
actions of consuming or speaking	dozen+	đói (PV #CV.to:jʔ) ‘lie/tell an untruth’ ăn (PV *ʔan) ‘to eat’ nhậu (PV *ju:ʔ, PAA *[ ]juuc) ‘to drink’ (in Vietnamese, drinking alcohol)
grooming	several	chải (PV *ca:s) ‘to comb’ búi (PV *pu:lʔ) ‘to plait (hair)’ vá (PV *k-pa:ʔ) ‘to sew/repair’
cognition and emotions	several	ngẫm (PV *-ŋamʔ) ‘to think about’ nhớ (PV *k-ŋə:ʔ) ‘to remember’ ghét (PV *t-kɛ:t) ‘to hate’

Of the smaller set of Proto-Vietic adjectives, some semantic subcategories include **conditions of the body** (e.g. *đui* (PV \*du:l/tu:l) ‘blind’, *sói* (PV \*k-rɔ:lʔ ~ k-lɔ:lʔ) ‘bald’, etc.); **emotions** (e.g. *vui* (PV \*t-pu:j) ‘happy’, *nhác* (PV \*k-ŋa:k) ‘lazy’, etc.); **tastes and smells** (e.g. *đắng* (PV \*taŋʔ, PAA \*ktaŋ) ‘bitter’, *nhạt/lạt* (PV \*m-la:c) ‘tasteless/bland’, etc.); **sensory conditions** (e.g. *rắn* (PV \*k-sanʔ) ‘hard/firm’, *mát* (PV \*t-ma:c) ‘cool’, etc.); **size and space** (e.g. *ngái* (=xa) (PV #s-ŋa:jʔ, PAA #cŋaaj) ‘far’, *dày* (PV \*k-daj)

‘thick’, etc.); and other basic categories. Again, such etyma of basic vocabulary only serve to further demonstrate the Vietic and Austroasiatic origins of Vietnamese but are not culturally specific and offer little insight into early Vietic culture.

### 6.3 Possessions and exchange

The Đông Sơn period is one of substantial quantities of bronze production, and the archaeological data at the Cổ Loa site suggests a proto-urban dwelling with a large population (Kim 2015). Despite this, reconstructed Vietic words for **metals** largely stem from Chinese of the Han Dynasty and early centuries of the first millennium CE, as is the case in Proto-Tai and Proto-Hmong-Mien (Alves 2016). As noted in Section 5.1, the words *bán* ‘to buy’ and *buôn bán* ‘commerce’ appear to be likely Early Sino-Vietnamese words. The borrowing of Chinese words for economic practices (e.g. the LMC Sino-Vietnamese words *tiền* ‘money’, *giá* ‘price’, and *thuế* ‘taxes’, etc.) is clear evidence of the impact of sociocultural contact.

Proto-Vietic words in the semantic domain of **possessions** nevertheless provide some evidence of socioeconomic differentiation and practices (e.g. ‘debt’, ‘rich’, etc.), as shown in Table 9. Native etyma in this domain indicate that at least some economic practices were in place in the initial period of Sinitic-Vietic contact. Vietnamese has other relevant words which have no apparent Chinese origin, but also no comparable words in other non-Việt-Mường languages (e.g. *cuộc* ‘cost’, *nghèo* ‘poor’, etc.), so it is not impossible that other early Vietic terms for economic practices may yet be found.

*Table 9: Proto-Vietic words for possessions and exchange*

Vietnamese	Gloss	Proto-Vietic
nợ	debt	*b-n-ə:ʔ > nə:ʔ
giàu	rich	*k-ɟaw
mua	buy	#muə
vay	borrow	#βal (PAA *pəl/*pul)
muợn	borrow	*ma:nʔ (PAA *smaap ‘to ask’)
mất	lose	*bət (PAA #bit)
lỡ	lose	#lo:h
có	have/there is	*kə:ʔ

### 6.4 Numerals and measure words

Proto-Vietic has a fully reconstructed system of **numerals** from 1 to 10, the numeral 100, and a couple of general quantity words (Table 10). Some unit nouns can also be reconstructed (Table 11). The Vietnamese numerals 1 to 4 appear related to Proto-Austroasiatic etyma, though both 1 and 2 have unexplainable phonological issues (i.e. the final \*c versus \*j of ‘one’ and the initial \*h versus \*b of ‘two’). In contrast, numerals 5 and to 10 can be reconstructed to the Proto-Vietic stage, but they have a complex, uneven distribution throughout the Austroasiatic language family (cf. Sidwell (2012)), making the reconstructions of a Proto-Austroasiatic system of numerals uncertain. Available data (Sidwell Ibid.) clearly shows proto-language numeral terms from 5 to 9 shared by Vietic and Bahnaric, Monic, and Aslian, but not other branches. Surprisingly, the neighboring Katuic branch has a very different set of numerals above 4, as discussed in the Appendix note about problems with the Vieto-Katuic hypothesis. In any case, these etyma in Vietic are Austroasiatic words and represent an early stage of speech community, but they are of uncertain status in relation to the Proto-Austroasiatic stage.

Of particular significance is that, despite the intense language contact with Chinese, Vietnamese retained its complete native system of numerals. This is in contrast with Tai, whose numeral system from 2 to 99 is entirely borrowed from Chinese, a contrast noted in Section 5.1. This is yet another example of the highly different sociocultural circumstances of Tai and Vietic in the period of Chinese southward expansion. I posit that retention of a native numeral system is also evidence of the social status of Vietic speech communities even under apparent Chinese linguistic and cultural influence.

**Table 10:** Numerals in Proto-Vietic

Vietnamese	Gloss	Proto-Vietic
một	one	*mo:c (PAA *muuj)
hai (cf. vài ‘a few’)	two	*ha:r (PAA *baar)
ba	three	*pa: (PAA *pi? (#pee)
bốn	four	*po:n? (PAA *pun?)
năm	five	*đam
sáu	six	*p-ru:ʔ > p <sup>h</sup> ru:ʔ / k <sup>h</sup> lu:ʔ
bảy	seven	*pəs
tám	eight	*sa:mʔ
chín	nine	*ci:nʔ
chục	ten	*ɟu:k
mười	ten	*ma:l
trăm	hundred	*k-lam
mấy	how many	*bəlʔ
ít	few	*ʔi:t
nữa	more	#đə:h
nửa	half	#CVđə:h

Correspondingly, Proto-Vietic has a number of **quantifiable nouns and measure words**. These include units of time and of physical measurements. The etyma for ‘day’ and ‘year’ are the only items with possible Proto-Austroasiatic roots. There is no evidence of classifiers, which likely developed in Vietic (and Southeast Asia in general) much later, primarily under the influence of Chinese (Alves 2020). These are conceptual terms and thus the kind of data that cannot otherwise be revealed in archaeological studies. However, it is an open question as to whether such data can be useful in making hypotheses about early Vietic culture.

**Table 11:** Unit nouns in Proto-Vietic

Vietnamese	Gloss	Proto-Vietic
ngày	day	*-ŋi: (PAA *tjii?)
tháng	month	*k-ra:ŋʔ
năm	year	*c-n-əm < cəm (PAA *cn lam)
gang	span	*c-ka:ŋ
vóc	handful	*po:k
miếng	mouthful/piece of	*-mɛ:ŋʔ
sải	fathom (n)	*p-la:s

### 6.5 Locational terms

Vietnamese locational words (in Table 12) are largely native, with a few stemming from Austroasiatic and the rest either solidly Proto-Vietic or at least at the stage of Viet-Muong and Pong-Cuoi. None of these constitute cultural vocabulary and instead represent what can be considered basic vocabulary. While spoken Vietnamese has borrowed some locational terms from Chinese, such as all cardinal directions *đông/tây/nam/bắc* ‘east/west/south/north’ and *ngoài* ‘outside’, it appears that the overall system of locational terms is native.

**Table 12:** Locational terms in Proto-Vietic

Vietnamese	Gloss	Proto-Vietic
ở	stay/be at	*ʔəh
trái	left side	*k-la:jʔ
đăm (archaic)	right side	*dam   tam (PAA #Ctam)
ngái (dialectal)	far	#s-ŋa:jʔ (PAA #cŋaaj)
gần	near	*t-kəŋ
dưới	under	*-ta:lʔ (PAA *ktlaal)
trên	top/upstream	*k-le:jŋ
trước ‘before’	before/forehead	*k-la:k
trong	inside	*k-lɔ:ŋ (PAA *kluuŋ ‘belly/middle’)
giữa	middle	#-Cah
cuối	end/extremity > nipple	*gɔ:jʔ / kɔ:jʔ

The one word of possibly uncertain status is *gần* ‘near’, for which Old Chinese \*N-kərʔ (MC gj+nH, LSV cậ) is a potential source. Trần (2016) has argued that this is an Austroasiatic etymon, but while it is found widely in Vietic, there is no comparable evidence in Austroasiatic. Furthermore, Ferlus also reconstructed \*s-də: ‘near’, which has a possible Austroasiatic origin, \*t2dəh (but attested only in Aslian and Palaungic). This item could, therefore, allow for a slot in which the Chinese word could have been borrowed. Nevertheless, it is reasonable to first consider *gần* a native item in light of the large overall system of native locational terms, but a possible Chinese origin cannot be completely excluded.

### 6.6 Pronouns and kinship terms

The modern Vietnamese system of **interpersonal reference terms** is heavily based on kinship and social positions, as is the case in modern Khmer, Thai, Lao, and Indonesian. These languages have a tendency to avoid pronouns for politeness, a typological feature restricted largely to majority/national languages in Southeast Asia as well as Japan and Korea (cf. Helmbrecht 2013). Modern Vietnamese has a blend of Vietic and Chinese elements in the Vietnamese system of pronouns and terms of reference (cf. Alves 2017). Original pronouns have peripheral usage in Vietnamese, primarily in intimate situations, resulting in the restructuring of the referential system. In contrast, most Austroasiatic languages are pronoun-dominant, as are highly conservative Vietic languages (though some degree of use of kinship terms for address are present, but not as much as in Vietnamese). Correspondingly, they have more typologically typical systems of pronouns than in Vietnamese. This suggests that Proto-Austroasiatic and Proto-Vietic were also pronoun-dominant, with kinship terms used primarily with kin.

While a full set of **pronouns** cannot be reconstructed for Proto-Vietic at this point, there is a solid core of 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> person pronouns, as shown in Table 13. Most are Proto-Austroasiatic etyma, with the exception of the first-person singular pronoun. This pronoun is unrelated to pronouns in other Austroasiatic languages and appears to be a true Vietic lexical innovation. The Proto-Austroasiatic reconstructions in Table 13 are those of Sidwell (2015b:104).

**Table 13:** Proto-Vietic pronouns

Vietnamese	Gloss	Proto-Vietic	Proto-Austroasiatic
tao	I/me	*so:	NA
bay	you (plural)/2p	#baj	*pej
nó	he/she/3s	#na:ʔ	*nVʔ
mày	you/2s	*mi:	*mi[i]ʔ; *miih
hắn, ( nó )	he/she/3s	*hanʔ	*[ʔ]anʔ

Most native Proto-Vietic **kinship** terms are not Proto-Austroasiatic etyma, with the exception of *cháu* (PV \*cu:ʔ, PAA \*cuuʔ) ‘grandchild’ and *chắt* (PV \*cat, PAA \*ceʔ) ‘great grandchild’. The Vietnamese system of kinship has been dramatically transformed by the borrowing of Chinese kinship terms (cf. Alves 2017). Between the restructured Vietnamese system and the insufficient comparative data of Vietic kinship terms, no ethnolinguistic claims can yet be made. However, native terms for marriage are reconstructable (e.g. *cưới* (PV \*-ga:lʔ / ka:lʔ) ‘to marry’, *chồng* (PV \*p-ʃo:ŋ / \*jo:ŋ) ‘husband’, (PV \*-ju:), etc.) *dâu* ‘daughter-in-law/bride’, etc.). Such native words are evidence of marriage practices prior to the arrival of the Chinese Han administration and historical descriptions of mandates of Chinese-style marriages in Giáo Chi province. As noted in Section 5.2, some Tai loanwords indicating intermarriage have affected Vietic, but not Vietnamese.

### 6.7 The home and clothing

Proto-Vietic terms related to **homes** include *nhà* (PV \*ja:) ‘house’ and words for parts of homes (e.g. *cửa* (PV \*kiah, only Việt-Mường and Cuối) ‘door’, *cột* (PV \*go:t) ‘stilt/post of a house’, *mái* (PV \*6a:lʔ) ‘roof’, etc.). Such words are expected, and yet, it contributes evidence, however basic, to the archaeological record. Lexical evidence connects to that of studies of homes in prehistoric Southeast Asia, or gaps in the archaeological record due to the deterioration of materials in traditional homes (cf. Higham 2017b).

Reconstructed Proto-Vietic terms for **clothing** include a core set of traditional clothing: *khố* (PV \*kʰo:ʔ) ‘loincloth’ (with possible cognates in Katuic, Bahnaric, and Khmer), *nón* (PV \*d̥o:nʔ) ‘conical hat’, *dép* (PV #cɛ:p) ‘sandals’, *váy* (PV \*6ə:lʔ) ‘skirt’, *trâm* (PV \*p-lam (only Việt-Mường and Pong-Cuối)) ‘earring’ and *búi* (PV \*c-pu:lʔ) ‘hairbun’. Additional actions related to grooming are listed in Table 8. The Chinese loanwords in this cultural domain appear to be either clearly introduced (e.g. *lụa* ‘silk’, *kim* and *ghim* ‘needle’, *nhuộm* ‘to dye’(spread also into Tai and Hmong-Mien), etc.) or in somewhat complementary distribution to the existing types (*áo* ‘shirt/upper garment’, versus ‘loincloth’ and ‘skirt’).

Archaeological evidence is available for bracelets and even the loincloth (on images of human figures) seen in Đông-Son-era bronzes. Other Proto-Vietic words for clothing lacking in the archaeological record can be considered viable evidence of items in the pre-Qin era. While not reconstructable in Proto-Austroasiatic, the Proto-Vietic \*d̥o:nʔ ‘conical hat’ appears related to words in neighboring Austroasiatic branches: Khmer *duən*; Katuic \*đuən, and Bahnaric (comparable forms in Bahnar, Sre, Mnong, and Chrau). This suggests a regional spread of the word possibly in the early Vietic period, a type of headwear before the borrowing of the Chinese word ‘hat’, Vietnamese *mũ* (帽 mào, OC \*mʰuk-s, MC mawH, LSV mào), which also spread into other Austroasiatic languages and Tai.

Some additional words of note are those for **jars and pots**, items which are well represented in lists of archaeological objects back to 4000 BP, but there are complications in the data. Two terms for container lids (*nấp* (PV #hnap) ‘lid/cover of jar’ and *vung* (PV #CV.puəŋ) ‘lid/cover of pot’) are both widespread and one is reconstructable with a presyllable. The reconstruction of \*vɔ: ‘jar’, Vietnamese *vò*, is attested only in Việt-Mường and Pong languages, making it less likely to be Proto-Vietic level and thus less likely to stem to Proto-Vietic but can be considered possibly early Vietic. Vietnamese *ché* ‘small jar’ has widespread counterparts in (Old Khmer (*ceh*), Mon (*ceh*), Katuic (Proto-Katuic \*ceh, \*ceeʔ), and Bahnaric (*ceh* in multiple languages)), and it is possible to tentatively reconstruct in Vietic the form #ce:ʔ, despite the lack of further attestations in Vietic. Beyond these, Vietnamese has at least two dozen items referring to jars and pots (e.g. *chum* ‘water jar’, *chĩnh* ‘jar (to store rice, salt, etc.)’, *khạp* ‘jar’, *kiệu* ‘big jar’, *liễn* ‘pot’, *lon* ‘earthenware pot’, *lọ* ‘vase/jar’, etc.). Overall, these words are thus suggestive of early pottery of the kind seen frequently in archaeological data in the region.

It is significant that a check of these items reveals very little evidence of Chinese loanwords. Some possible comparanda are seen in neighboring Austroasiatic languages (cf. Bahnaric \*drap; Katuic (Ngeq tɬ:p hɬ:p); Khmer daap ‘bottle/jar/pitcher/flask’). Vietnamese *lu* ‘jar’ and *khạp* ‘jar’ have comparable items in Khmer, *lù*: and *khap* respectively. Some words related to porcelain are Chinese in origin (e.g. later-era Middle-Chinese *bình* ‘vase’, *sứ* ‘porcelain’, etc.), but overall, this semantic domain is indigenous, but also often without clear etymological origins in Vietic. One item of note in Vietnamese *thạp* ‘metal jar’, a term specifically used to refer to situlae, the ornate bronze pots seen in many Dong Son archaeological sites. Unfortunately, in available data, this is only attested only in Việt-Mường. The question of whether this word can be connected to the Dong Son era situlae cannot be supported or refuted at this time.

## 6.8 Food production and materials for manufacturing

Not surprisingly, a good deal of the vocabulary is related to food production. Such words can be divided into several semantic subdomains.

- **Terms for food and drink** include descriptors (e.g. *chín* (PV \*ci:nʔ, PAA \*ciinʔ) ‘ripe/cooked’, *nép* (PV \*dê:p) ‘glutinous’, *mắm* (PV \*bámʔ) ‘salted’, etc.), ingredients (e.g. *muối* (PV \*bɔ:jʔ, PAA \*bɔɔh) ‘salt’, *ớt* (PV \*ʔə:t) ‘chili’, *nghệ* (PV \*ŋɛ:lʔ) ‘turmeric’, etc.), and cooking (e.g. *nung* (PV \*dũŋ) ‘to bake’, *nướng* (PV \*dã:ŋʔ, PAA \*t̪aŋ) ‘to roast’, *rán* (PV \*-ra:nʔ) ‘to fry’, etc.). Indeed, the traditional *ngũ vị* ‘five flavors’ (from Chinese 五味 wǔ wèi), including *cay* ‘spicy’, *ngọt* ‘sweet’, *chua* ‘sour’, *đắng* ‘bitter’, *mặn* ‘salty’, are all proto-language etyma. As for implements of food production, the words *cối* (PV \*t-ko:lʔ, PAA \*guul) ‘mortar’ and *chày* (PV \*tʃ-re:) ‘pestle’ (see discussion in Section 6.9) for food processing have Austroasiatic reconstructions, which is strong evidence for these as part of early Vietic and possibly Proto-Vietic practices.
- Of the approximately 100 reconstructed **words for animals**, several dozen are wild animals, including several terms for birds, 20 for mammals, and nearly two dozen for insects. The number of terms for domestic animals is naturally much smaller: *chó* (PV \*ʔa-cɔ:ʔ, PAA \*cɔʔ) ‘dog’, (*heo*) *cúi* (PV \*gu:rʔ, ku:rʔ) ‘pig’, *gà* (PV \*r-ka:) ‘chicken’, *voi* (PV \*-vɔ:j (only Việt-Mường and Pong-Cuối)) ‘elephant’, *dê* (PV \*-te: (only Việt-Mường and Cuối)) ‘goat’, and *vịt* (PV \*vi:t) ‘duck’ (probably from Tai \*pet<sup>D</sup>). Of the several early Chinese loanwords in the semantic field of animals, a few appear to be likely introduced practices: *ngựa* ‘horse’, *mèo* ‘cat’, and *kén* ‘cocoon’ (for silk production, as ‘silk’ is also borrowed).
- Of the 80-plus terms for **agriculture and vegetation**, over half are terms for produce, a variety of fruits, vegetables, and grains. Several are agricultural activities, such as *rây* (PV \*-re:) ‘to sieve/sift’, *cấy* (PV \*kəlʔ) ‘to transplant rice seedlings’, *hái* (PV \*ha:rʔ) ‘to harvest/gather’, *mới* (PV \*mɔ:lʔ) ‘to sow rice (in a hole)’, *trồng* (PV \*m-lo:ŋ) ‘to plant (a tree)’, and *xay* (PV \*tʃe:) ‘to grind/husk (rice)’. These relatively basic terms are expected for any culture in the region back to periods of horticulture, not necessarily developed agricultural practices. However, as will be discussed in Section 6.9, the practice of rice agriculture is even more fully represented by the lexical evidence.
- As for **agricultural materials used for production**, there are multiple terms for (a) various types of bamboo (e.g. *dang/giang* (PV \*k-ta:ŋ), *nứa* (PV \*-na:ʔ), *pheo* (PV \*p-hɛ:w (Việt-Mường and Cuối)), *tre* (PV \*k-lɛ: (Việt-Mường and Cuối)), (b) thatch-grass (*tranh/gianh* (PV \*p-lɛŋ, PAA #plang), *bái* (PV \*pa:jʔ)), and (c) sap/resin (*nhựa* (PV \*ɲa:ʔ (Việt-Mường and Cuối))). As for production of items, the verb *đan* (PV \*ta:ŋ, PAA \*taaŋ) ‘to weave’ in particular is in every branch of Austroasiatic and clearly attests to the deep time of that practice. While weaving is a well-known Neolithic practice, the widespread retention of this word is not trivial. Woven materials usually deteriorate completely in the humid climate and acidic soil of Southeast Asian, but evidence of weaving can be seen on pots on which patterns of woven material were used for decoration (e.g. in the Đông Đậu culture (c. 1500-1000 BCE) (Hoàng 2003)). This handful of lexical items is certainly limited compared to what the original vocabulary must have been. Nevertheless, they highlight essential materials and activities to manufacture implements, decorations, and homes, all core aspects of daily life and culture. Those of Austroasiatic etyma in particular can be hypothesized to be part of the Neolithic/Pre-Metal Age. They attest to the means by which cultural diversity may have developed in the region, notably via practices involving bamboo (White 2011).

## 6.9 Rice agriculture

Diffloth (2005:77) has noted a core set of 11 terms related to **rice and rice production** in Austroasiatic. Proto-Vietic vocabulary in the field of rice and rice production and processing has most of these and about a dozen other related terms; thus, there are both lexical retentions and innovations in this cultural domain. Table 13 lists these items, including both nouns and verbs, though all the verbs are restricted to Vietic. Several nouns are also Proto-Austroasiatic etyma, but Proto-Vietic terms predominate.

The etymon for ‘**pestle**’ is of note. Ferlus (2009) has claimed that *chày* ‘pestle’ is a Đông-Son era lexical innovation via derivational morphology that spread into other Austroasiatic languages. Considering the reflexes of this etymon in some Munda languages (e.g. Sora [ɔn-rɪj] ‘pestle’), which Ferlus notes, it is reasonable to posit that the word predates the Đông Sơn period. Indeed, Rau and Sidwell provide evidence that show a possible migration of the Munda portion of Austroasiatic to India around 2000 to 1500 BCE (2019:50). Moreover, a list of archaeological objects at the Xóm Rền site of the Phùng Nguyên period (c.

2000-1500 BCE) (Hán 2009:222-224) includes a few pestles and mortars. Unfortunately, I cannot locate other instances of mortars and pestles in archeological studies from this early period, as one would hope for further clarification. Still, this implement was certainly present at the time of the arrival of the Han, as was this etymon. How deep the history is of the pestle in Austroasiatic culture will depend on future archaeological investigations.

Overall, while one can posit that not all of these words are necessarily from the earliest Proto-Vietic period, archaeological data from the Phùng-Nguyên-era Mán Bạc site supports settled communities in which hunting was less significant than before (Oxenham et al. 2011:110, noting only seven animal taxa in Phùng-Nguyên-era data versus twenty in the earlier Hoabinhian period). That early Vietic groups were increasingly reliant on settled food production practices is lent support by the linguistic evidence.

**Table 14:** Proto-Vietic terms related to rice agriculture

Vietnamese	Gloss	Proto-Vietic
<i>Verbs</i>		
trấu	to husk (rice)	*k-lu:ʔ
cấy	to transplant rice seedlings	*kəlʔ
mới	to sow rice (in a hole)	*mɔ:lʔ
xay	to husk (rice)	*tʃe:
xôi ‘steamed rice’	to steam (rice)	*so:j
<i>Nouns (Vietic)</i>		
nếp	glutinous (of rice)	*dɛ:p
kê	millet (setaria)	*k-hiɛl > kiɛl
mương	canal/ditch for irrigation	*-miəŋ
ruộng	rice-field	*rɔ:ŋʔ
cháo	gruel/porridge of rice	*ca:wʔ
lòn ‘variety of rice’	nonglutinous (rice)	*lɔ:ŋ
nám	croquette of rice	*-namʔ
vắt	croquette of rice	*k-pat
<i>Nouns (Vietic &amp; Austroasiatic)</i>		
gạo	rice, husked	*r-ko:ʔ (PAA *rk[aw]ʔ)
nong	winnowing basket	*dɔ:ŋʔ (PAA #Cduŋ)
sàng	a winnow	*g-ra:ŋ (PAA *k-ra:ŋʔ)
chày	pestle	*tʃ-re: (PAA *nrəjʔ, *nrəj[ ], *rəjʔ)
cối	mortar (for rice)	*t-ko:lʔ (PAA *guul)
mạ	seedlings, rice (for transplanting)	*s-ma:ʔ (PAA *maʔ)
cám	bran	*t-ka:mʔ (PAA *skaamʔ)
mới 梅 ‘a thing used when transplanting’	stick for digging	*c-mɔ:lʔ (PAA #C-mɔ:l)

Despite the evidence of an existing system of rice production, as noted above, Sinitic may have contributed a significant word: Vietnamese *lúa* ‘paddy rice’ (PV \*ʔa-lɔ:ʔ) is a possible Old Chinese loanword (稻 dào ‘paddy’, OC \*[l]ʰuʔ, MC dawX, LSV dào). As for Tai, Vietnamese *nong* ‘basket for winnowing’ (PV \*dɔ:ŋʔ, PAA #Cduŋ) appears related to Proto-Tai \*dɔŋ<sup>C</sup> ‘winnowing basket’. However, there are sufficient attestations among Austroasiatic languages to tentatively reconstruct a form #Cduŋ (cf. Katuic \*kdɔŋ; Bahnaric \*-dɔ:ŋ; Khasic \*pduŋ; Mangic (Mang ʔa:<sup>1</sup> dɔ:ŋ<sup>6</sup>); Nicobarese (Car) tɔluŋ). Moreover, considering the presyllabic material in the Austroasiatic languages, and the lack of that in the Proto-Tai form, the direction of borrowing would more likely be from Austroasiatic (and hypothetically,

even Vietic specifically) into Tai. This is yet another instance in which more data sifting is necessary to clarify the situation. Regardless, the lexical data suggests *nong* ‘winnowing basket’ of Vietnamese may be connected to a woven object not easily identified in archaeological data.

A final note is regarding the origin of the Vietnamese word *cày* ‘plough (v)/a plough (n)’. According to both archaeological (e.g. Higham 2017) and linguistic evidence (Diffloth 2005), Austroasiatic groups appear to have had agricultural practices that likely supported the dispersal throughout Mainland Southeast Asia. Nevertheless, the appearance of metal ploughshares in the mid-1<sup>st</sup> millennium BCE, such as those at the Cỏ Loa site, are contemporaneous with and likely facilitated rapid socioeconomic developments in the region. Less certain in the archaeological literature is the origin of this crucial metal agricultural implement. Vietnamese *cày* is a reflex of with the reconstructed Vietic form *\*gal/\*lɲgal/\*ŋgal* ‘plough’. This form is widespread throughout Austroasiatic in Mainland Southeast Asia as well as in Malayic, including both varieties of Malay and Chamic languages. It has been posited to be related to Sanskrit *taṅgāla* ‘plough,’ and even an Austroasiatic loanword into the Indic region (Przyluski 1929:8-15). However, I have seen no archaeological studies of metal plough styles of these two regions to indicate whether they are related or whether the sharing went east to west or west to east, so at this point, the linguistic data cannot resolve the matter.

### 6.10 Betel-nut chewing and tooth-blackening

The deep history of the practice of betel-nut chewing among early Vietic speakers is supported in the linguistic data. Words for ‘betel’, ‘areca’, and ‘lime (mineral)’ used in betel-nut chewing preparations (see Table 14) are all attested widely among the three Vietic subgroups and readily reconstructed at the Proto-Vietic level. The overlapping practices of betel-nut chewing and tooth-blackening are spread throughout both Mainland and Insular Southeast Asia. The origins of the practices of betel-nut chewing and overlapping practice of tooth-blackening are of debate. Such stains on teeth have been excavated among a group of individuals in the Philippines dating to 2660 BCE, and the practice may have spread in the region via Austronesian expansion (e.g. Zumbroich 2007).

As for northern Vietnam, archaeological data of dental remains show that betel-nut chewing and tooth-blackening (*nhuôm răng*) were practiced in the Núi Nấp archaeological site considerably later, as early as 400 BCE (Oxenham, Locher, Nguyen, and Nguyen 2002) in the Đông Sơn era. Przyluski (1929:15-24) noted some of the comparable Austroasiatic terms for ‘betel’ (suggesting possible spread of this to Aryan languages). Moreover, two of the Proto-Vietic words have also been reconstructed for Proto-Austroasiatic, though the geographic distributions of the words appear restricted to the central part of Mainland Southeast Asia. Corresponding to the archaeological data noted above, Blust and Trussel (2010) reconstructs Proto-Malayo-Polynesian ‘betel pepper’ as *\*bu-bulu*, a promising comparable form. An additional complicating factor is the Proto-Tai reconstruction *\*blu<sup>A</sup>* ‘betel nut’ (Li), a likely match that suggests a regionally spread word. It is thus impossible to determine the precise route of transmission, and it is possible that it spread in multiple directions at multiple times. More archaeological data is required to be able to provide more details.

Regardless of various complicating factors, the terms are widespread in Vietic, with some degree of complex phonological structure (e.g. presyllabic material). This provides support that they date at least to the pre-Qin part of the Đông Sơn in northern Vietnam.

*Table 15: Proto-Vietic and chewing betel*

Vietnamese	Gloss	Proto-Vietic	Proto-Austroasiatic
vôi	lime, mineral	*k-pu.r	*knpur
trầu / giầu	betel leaf	*b-lu:	*ml[əw] (or #blu:)
cau	areca nut	*kaw	NA

## 7. Insights, challenges, and future directions

Thus far, Vietic lexical data offers insights in three main areas: (a) further clarification of the linguistic affiliation of Vietnamese in Austroasiatic, (b) evaluation of early Sinitic and some Tai loanwords in early Vietic, and (c) preliminary insights into ethnolinguistic traits of early and Proto-Vietic. In this final section, these issues are summarized and concluding thoughts are given.

### **7.1 Linguistic affiliation and ethnolinguistic inferences**

Some of the lexical data, namely, basic vocabulary, primarily serves to highlight the linguistic affiliation between Vietnamese and its Vietic and Austroasiatic origins. The 200-plus words of Austroasiatic origin cover the domains of the natural world, general actions, basic implements, body parts, numbers, and so on, and Proto-Vietic etyma cover similar terms with numerous lexical innovations in virtually all semantic fields. Some semantic domains, such as kinship and marriage terms, pronouns (a pronoun-dominant system, in contrast with the modern Vietnamese kinship/title-dominated one), rice cultivation and production, economic practices, and aspects of material culture, such as clothing and grooming, altogether paint a general image of early Vietic culture. Words in the domain of rice and rice production are particularly ample, with the implication that Vietic culture was one in which rice agriculture (i.e. not just horticulture) was part of a broader cultural package. Nevertheless, these are very general ideas and can only be used in coordination with archaeological, historical, and ethnographic data to make more specific inferences of the ethnohistory of Vietic-speaking groups.

### **7.2 Language contact and loanwords**

Also, regarding the issue of ethnohistory, early loanwords highlight the intensity of early language contact between Sinitic and Vietic. The notable number of loanwords suggests unequal sociocultural status, with Sinitic as a donor language having perceived social status (but see the comments about the retention of native numerals in Section 6.4). In contrast, the much smaller number of Tai loanwords in this period suggests a very different situation, and it is possible that Tai and Vietic speakers had more equal sociocultural status.

This latter point is in contrast with some previous hypotheses by both archaeologists and historical linguists of ancestors of Vietnamese being recipients of rice production practices from Tai groups, but it is in line with more recent views. Questions of sociopolitical organization of these latter two groups are largely unanswered (e.g. Was there a near state-level society at the Cỗ Loa site? Was there a confederation of chiefdoms including different language groups, including both Vietic and Tai speakers?), so we cannot make precise claims about the language contact scenario. Regardless, the archaeological data noted in Section 4 points to multiple groups in southern China with rice-production capacity in the period of Austroasiatic dispersal, and Proto-Austroasiatic has a reasonably solid core of rice-production vocabulary.

As for technology, how metallurgical practices spread in the earliest periods into southern China and Southeast Asia have not been articulated. But at the very least, archaeological evidence shows that the Metal Age in the region long predates large-scale migration from northern China, during the Đông Đậu period (1500-1000 BCE), when archaeologists note evidence of the Bronze Age in this region (e.g. Sinh 2003). Thus, the lack of any native, pre-Qin terms for metals in Proto-Vietic is surprising. Indeed, metal terms in Proto-Vietic, Proto-Tai, and proto-Hmong-Mien are all early Chinese loanwords (Alves 2016), even though all three groups were in regions where the Metal Age began centuries prior to Chinese southward expansion. Again, this suggests significant sociocultural status of the Sinitic, whether due to economic capacity, types of technology, cultural status, or likely a combination of factors. But the lexical evidence does not indicate that Chinese speakers introduced metallurgy practices, but apparently influenced long-established practices among all the language groups in the region. The apparent Sanskrit or Pali term for ‘ploughshare’ is another apparent instance of shared metal-based technology, as many bronze ploughshares have been discovered at the Cỗ Loa site.

As for the practice of betel-nut chewing and tooth-blackening, both lexical and archaeological data demonstrate such practices are deep in Vietic history, but where the tradition comes from is still an unanswered question. As a regional practice and some regionally shared etyma, it highlights early sociocultural contact. Hopefully, an interdisciplinary approach will allow for hypotheses to be furthered about the origins and spread of this practice in Vietnam and Greater Southeast Asia as it will undoubtedly provide insight into early cultural contact in the region.

### **7.3 Challenges in evaluating the data and final thoughts**

Many challenges and uncertainties remain in assessing Vietic lexical data and reconstructions, with implications for piecing together aspects of Proto-Vietic and early Vietic ethnolinguistic traits. Words such as *vùi* ‘to bury/put in the ground’ (PV#bu:1) and *quê* ‘village’ (PV \*k-ver, cf. Proto-Katuic \*wiil, \*weel) are suggestive of cultural practices. However, ‘to bury’ is somewhat general and cannot be linked to early

practices with any certainty, while ‘village’ is related to communities, but with no clear indication of size approaching that of Cồ Loa. As much as we want to tie proto-language words to ethnoarchaeological data and inferred cultural features, we can only note this kind of lexical data of intangible culture but make no strong claims of association or clarification of sociocultural circumstances.

Another point of uncertainty is when words may have spread later than suggested by widespread occurrence in a language group. Words which appear throughout a language group can be lexical retentions from an earlier shared linguistic ancestor, but words can also spread both within and among speech communities at any point after the proto-language stage. Though the number is small, the instances of widespread Late Middle Chinese loanwords, thousands of years later than the period of Proto-Austroasiatic or even the earliest period of Proto-Vietic, demonstrate the latter situation. Some words in minority Vietic languages are clearly recent loans from Vietnamese or from Lao, as indicated by their phonological patterns. This matter has not yet been studied in detail and is important for increasing understanding and checking hypotheses.

As for dating of etyma, Austroasiatic etyma with more archaic features (e.g. complex initials, presyllables, or finals such as /-h/ or /-l/) could hypothetically be among the oldest in Vietic. Similarly, Vietic etyma that exhibit archaic features could be associated with Vietic dating back a few thousand years. As comparative linguistic, archaeological, and ethnographic data is accumulated and analyzed, growing amounts of evidence do strongly support the deep history of many of the words of the current database. But in most cases, it is not possible to tie words with any precision or even to general archaeological periods.

Another challenge is filling in the lexical gaps in Proto-Vietic for items which are in the archaeological or historical record, but which are not reconstructable to the Proto-Vietic level. For instance, one well-known type of Đông Sơn object is the situla, a kind of bucket often made of bronze, called *thạp* in Vietnamese. This word has not been noted in data from other Vietic languages. It is thus not possible to say whether the Vietnamese word stems to the original object or is a more recent lexical innovation, and no Proto-Vietic or even early Vietic word can yet be offered. We can go back further and consider the Mán Bạc Phung-Nguyen-era site, which contains bracelets, adzes, bone hooks, shell ornaments, and pottery (Higham 2017:14). Vietnamese words for such items include a mixture of etymological origins, or there are no precise modern equivalents, and modern lexical compounds must be created. Thus, in many cases, there are no continuous connections between real-world objects of the past and modern words. Lexical developments and innovations have left no remnants in many cases.

As for early loanwords that have been reconstructed in early Vietic, while many Late Old Chinese and Early Middle Chinese words were borrowed and spread in Vietic in an early period and are retained in Vietnamese, many of these words are not seen elsewhere in Vietic data. It may simply be they were unlikely to have become part of the lifestyles of Vietic groups in rural areas. Moreover, abstract concepts such as *ngĩa* ‘righteousness’ (LSV *ngĩa*) and *vốn* ‘capital’ (LSV *bổn*) are unlikely to be gathered during fieldwork. But they are also less likely to have become part of languages spoken in circumstances where such concepts may not have been socioculturally significant. The Vietnamese word *ngói* ‘roof tile’ (LSV *ngõa*) is a very early loanword, potentially dating to the Western Han in the BCE era, as supported by the archaeological evidence of tens of thousands of Chinese-style tiles in the Cồ Loa site. And yet, it is not seen in data outside of the Việt-Mường branch in current data, meaning it cannot be reconstructed at the Proto-Vietic level. Like *thạp* ‘situla’, *ngói* ‘tile’ appears limited to Vietnamese (until more data shows otherwise). The items appear frequently in archaeological studies in the Red River Delta, and so they may represent lifestyles of the Vietic groups but not those to the south, but they cannot yet be reconstructed to early Vietic.

Throughout this paper, I have noted many caveats about the limitations of the data as well as many areas in need of attention, such as Sino-Mường data and identification of phonological evidence for a north-south split in Vietic. As the data is further developed and analyzed for various historical linguistic purposes, additional challenging questions can be considered.

- What can be done to compensate for gaps in the data in seeking to answer questions of ethnolinguistic history?
- What can be done with Vietnamese vocabulary associated with the Đông Sơn archaeological tradition, but which is not corroborated by other Vietic languages? These include words which were either not part of the groups in the proto-urban dwellings or which were lost as part of subsequent cultural developments. Yet it is possible such words could date to that period.

- How much more data can be gathered among minority Vietic before some of them are further transformed by language contact or completely lost to language shift to other languages?
- How can we certify which items among Vietic languages are not later loanwords from Vietnamese or Việt-Mường languages? Various criteria must be considered to evaluate relevant data.
- The influence of Chinese through complex long-term language contact is profound, but the evidence provided herein shows clear Austroasiatic and Vietic affiliation of Vietnamese. What is a suitable way to describe the combination of native and non-native elements of Vietnamese?

Thus, while the data herein is more than previously gathered since Logan's publication in the 1850s, in many ways, it is still preliminary. Other researchers must continue to judiciously use all available analytical tools and continue to seek relevant data, whether from Vietic languages, neighboring languages, or from various ethnohistorical data sources.

## References

- Alves, Mark J. 2005. The Vieto-Katuic hypothesis: lexical evidence. *SEALS XV: Papers from the 15th meeting of the Southeast Asian Linguistics Society*, ed. by Paul Sidwell. Pacific Linguistics: Research School of Pacific and Asian Studies, Electronic Publication E-1, the Australian National University, 169-176.
- Alves, Mark. 2006. Linguistic research on the origins of the Vietnamese language: an overview. *Journal of Vietnamese Studies* 1.1-2:104-130.
- Alves, Mark J. 2009. Loanwords in Vietnamese, in *Loanwords in the World's Languages: A Comparative Handbook*. De Gruyter Mouton, ed. Martin Haspelmath and Uri Tadmor, 617-637.
- Alves, Mark J. 2016. Identifying Early Sino-Vietnamese Vocabulary via Linguistic, Historical, Archaeological, and Ethnological Data, in *Bulletin of Chinese Linguistics* 9 (2016):264-295.
- Alves, Mark J. 2017. Etymological research on Vietnamese with databases and other resources. *Ngôn Ngữ Học Việt Nam, 30 Năm Đổi Mới và Phát Triển (Kỷ Yếu Hội Thảo Khoa Học Quốc Tế)*, 183-211. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Alves, Mark J. 2018. Chinese Loanwords in Vietnamese Pronouns and Terms of Address and Reference, in *Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29)*, 2017, Volume 1, ed. by Marjorie K.M. Chan and Hana Kang, 286-303. Columbus, Ohio: The Ohio State University.
- Alves, Mark J. 2019a. Chinese Loanwords in the Vietnamese System of Color Terms, in *Proceedings of the 30th North American Conference on Chinese Linguistics (NACCL-30)*, Volume 2:341-355. Columbus, Ohio: The Ohio State University.
- Alves, Mark J. 2019b. Dữ liệu liên ngành: chi Vietic kết nối với văn hóa Đông Sơn. Presentation given to the University of Hanoi, Department of Social Science and Humanities, 23 May, 2019.
- Alves, Mark J. 2020. Initial Steps in Reconstructing Proto-Vietic Syntax. In *Austroasiatic Syntax in Areal and Diachronic Perspective*, ed. by Paul Sidwell, Mathias Jenny, and Mark Alves, 46-81. Brill Publishers.
- Anderson, Gregory. 2020. Proto-Munda Prosody, Morphotactics and Morphosyntax in South Asian and Austroasiatic Contexts. in *Austroasiatic Syntax in Areal and Diachronic Perspective*, ed. by Mathias Jenny, Paul Sidwell, and Mark Alves. Brill Publishers. 157-197.
- Babaev, Kirill V. and Irina V. Samarina. 2018. *May language: Materials of the Russian-Vietnamese linguistic expedition* (in Russian). YASK Publishing House.
- Baxter, William H. and Laurent Sagart. 2014. Baxter-Sagart Old Chinese reconstruction, version 1.1 (20 September 2014). <http://ocbaxtersagart.lsa.it.lsa.umich.edu/BaxterSagartOCbyMandarinMC2014-09-20.pdf>.
- Bellwood, Peter. 2004. The origins and dispersals of agricultural communities in Southeast Asia. in *Southeast Asia: From Prehistory to History*, ed. by Ian Glover & Peter Bellwood. 21-40.
- Benedict, Paul K. 1942. Thai, Kadai, and Indonesian: A new alignment in South-Eastern Asia. *American Anthropologist* 44:576-601.

- Blagden, Charles Otto. 1894. Early Indo-Chinese influence in the Malay Peninsula, as illustrated by some of the dialects of the aboriginal tribes. *Journal of the Straits Branch of the Royal Asiatic Society* 27:21-56.
- Blust, Robert and Stephen Trussel. 2010. *Austronesian Comparative Dictionary*. Web. [http://www.trussel2.com/ACD/acd-ak\\_a.htm](http://www.trussel2.com/ACD/acd-ak_a.htm). Last accessed: 15 August 2020.
- Chamberlain, James R. 1998. The origin of Sek: Implications for Tai and Vietnamese history. In *The International Conference on Tai Studies*, ed. by Somsonge Burusphat. Bangkok, Thailand, 97-128. Institute of Language and Culture for Rural Development, Mahidol University.
- Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *The Mon-Khmer Studies Journal* 15:139-154.
- Diffloth, Gérard. 1992. Vietnamese As a Mon-Khmer Language. *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*, ed. by Martha Ratliff and Eric Schiller, 125-139. Arizona State University, Program for Southeast Asian Studies.
- Enfield Nick J. and Gérard Diffloth. 2009. Phonology and sketch grammar of Kri, a Vietic language of Laos. *Cahiers de Linguistique Asie Orientale* 38.1:3-69.
- Ferlus, Michael. 1979. Lexique Thavùng-Français. *Cahiers de Linguistique, Asie Orientale* 5: 71-94.
- Ferlus, Michel. 1975. Vietnamien et Proto-Viet-Muong. *Asie du sud-est et monde insulindien* 6.4:21-55.
- Ferlus, Michel. 2014. Arem, a Vietic language. *Mon-Khmer Studies* 43.1:1-15.
- Gehrmann, Ryan. 2015. Vowel height and register assignment in Katuic. *The Journal of the Southeast Asian Linguistics Society* 7:56-70.
- Grierson, George Abraham. 1904. Mon-Khmer and Siamese-Chinese Families. Vol. II, *Linguistic Survey of India*: Delhi. Banarashidaas.
- Hà, Văn Tấn. 2018. *The Making of Việt Nam*. Thế Giới Publishers.
- Haudricourt, André-Georges. 1953. La Place Du Viêtnameien Dans Les Langues Austroasiatiques. *Bulletin de la Société de Linguistique de Paris* 49.1:122-8.
- Haudricourt, André-Georges. 1954a. Comment Reconstruire Le Chinois Archaïque. *Word: Journal of the International Linguistic Association* 10:351-64.
- Haudricourt, André-Georges. 1954b. De L'origine Des Tones En Viêtnameien. *Journal Asiatique* 242:69-82.
- Hayes, La Vaughn. 1982. The Mutation of \*R in Pre-Thavung. *The Mon-Khmer Studies Journal* 11: 83-100.
- Hán, Văn Khấn. *Xóm Rền: Một di tích khảo cổ đặc biệt quan trọng của thời đại đồ đồng Việt Nam* [Village: An important archaeological relic of the Vietnamese Bronze Age]. Nhà Xuất Bản Đại Học Quốc Gia Hà Nội.
- Helmbrecht, Johannes. 2013. Politeness distinctions in pronouns. in *The World Atlas of Language Structures Online*, ed. by Dryer, Matthew S. & Haspelmath, Martin. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/45>, Accessed on 2020-11-18.
- Higham, Charles F. W. 2017a. First farmers in Mainland Southeast Asia. *Journal of Indo-Pacific Archaeology* 41:13-21.
- Higham, Charles. 2017b. The prehistoric house: a missing factor in Southeast Asia. in *New Perspectives in Southeast Asian and Pacific History*, ed. by Philip J. Piper, Hirofumi Matsumura and David Bulbeck, ANU Press, the Australian National University. 369-384.
- Hồ, Lê. 1992. Từ Nam Á trong tiếng Việt [Austroasiatic words in Vietnamese]. in *Tiếng Việt và Các Ngôn Ngữ Dân Tộc Phía Nam*, 65-110. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội.
- Hoàng, Văn Khoán. 2003. Nghề đan của người Đông Đậu (qua các dấu đan in trên đồ gốm) [The craft of the Dong Dau people (through imprints of weavings on pottery)]. *Văn Hóa Đông Đậu: 40 Năm Phát Hiện và Nghiên Cứu* (1962-2002). Nhà Xuất Bản Khoa Học Xã Hội. 172-180.
- Huffman, Franklin E. 1977. An Examination of Lexical Correspondences Between Vietnamese and Some Other Austroasiatic Languages. *Lingua* 43.2:171-98.
- Kasuga, Atsushi. 1994. A Study on Arem Phonology. Master's thesis. Tokyo University.
- Kim, Nam. 2015. *The Origins of Ancient Vietnam*. Oxford Studies in the Archaeology of Ancient States. Oxford University Press.

- Koenigswald, G.H.R. 1952. Evidence of a prehistoric Australomelanesoid population in Malaya and Indonesia. *Southwestern Journal of Anthropology* 8.1:92-96.
- La Vaughn Hayes Digital Archives. <http://sealang.net/archives/hayes/>. Last Accessed: 20 August 2020.
- Lê, Đình Khản. 2002. *Từ Vựng Gốc Hán trong Tiếng Việt* (Vocabulary of Chinese Origin in Vietnamese). Hồ Chí Minh City: Nhà Xuất Bản, Đại Học Quốc Gia Thành Phố Hồ Chí Minh.
- Lipson, Mark et al. 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* 361:92-95.
- Logan, J. R. 1852-1855. Ethnology of the Indo-Pacific islands. *The Journal of the Indian Archipelago and Eastern Asia* VI (1852), VII (1853), VIII (1854), IX (1855).
- Maspero, Henri. 1912. Etude sur la phonétique historique de la langue Annamite: Les Initiales. *Bulletin de l'École Française d'Extrême-Orient* 1-127.
- Matsumura, Hirofumi et. al. 2019. Craniometrics Reveal “Two Layers” of Prehistoric Human Dispersal in Eastern Eurasia. *Scientific Reports* 9:1451. doi:10.1038/s41598-018-35426-z.
- McColl, Hugh, et al. 2018. Ancient Genomics Reveals Four Prehistoric Migration Waves into Southeast Asia. Preprint, *Science*, doi:10.1126/science.aat3628.
- Mon-Khmer Etymological Dictionary*. SEALang Mon-Khmer languages project. <http://www.sealang.net/monkhmer/dictionary/>. Last accessed 20 August 2020.
- Munda Etymological Dictionary*. SEALang Mon-Khmer languages project. <http://www.sealang.net/munda/dictionary/>. Last accessed 20 August 2020.
- Nguyễn, Phú Phong. 1997. *Le parler Nguôn: Language d'un minorité ethnique des hautes vallées du Sông Gianh Quảng Bình*, Việt Nam. Université Paris 7, Denis Diderot.
- Nguyễn, Phú Phong, Trần Trí Dõi, and Michel Ferlus. 1988. *Lexique Vietnamien-Ruç-Français*. Université de Paris.
- Nguyễn, Tài Cẩn. 1995. *Giáo trình lịch sử ngữ âm tiếng Việt (sơ thảo)* [Textbook of Vietnamese historical phonology (a draft)]. Hà Nội: Nhà Xuất Bản Giáo Dục.
- Nguyễn, Văn Khang, Bùi Chi, and Hoàng Văn Hành. 2002. *Từ điển Mường-Việt* (A Mường-Vietnamese dictionary). Hà Nội: Nhà Xuất Bản Văn Hoá Dân Tộc.
- Nguyen, Viet. 2005. The Da But culture: evidence for cultural developments in Vietnam during the Middle Holocene. *Indo-Pacific Prehistory Association Bulletin* 25, 2005 (Taipei Papers, Volume 5). 89-93.
- Oxenham, Marc F., Cornelia Locher, Nguyen Lan Cuong & Nguyen Kim Thuy. 2002. Identification of areca catechu (betel nut) residues on the dentitions of Bronze Age inhabitants of Nui Nap, Northern Vietnam. *Journal of Archaeological Science* 29:909-915.
- Przyluski, Jean. 1929. Non-Aryan Loans in Indo-Aryan. Pre-Aryan and Pre-Dravidian in India by S. Lévi, J. Przyluski and J. Bloch, translated from French by Prabodh Chandra Bagchi. University of Calcutta. 3-24. <https://indianculture.gov.in/rarebooks/pre-aryan-and-pre-dravidian-india>.
- Rau, Felix & Paul Sidwell. 2019. The Munda maritime hypothesis. *Journal of the Southeast Asian Linguistics Society* 12.2:35-57.
- Reid, Lawrence A. 2005. The current status of Austric: A review and evaluation of the lexical and morphosyntactic evidence. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Sagart, Laurent; Blench, Roger; Sanchez-Mazas, Alicia. London: Routledge Curzon. 132-160. <http://hdl.handle.net/10125/33009>.
- Reinecke, Andreas. 2015. Ancient gold and silver jewelry and the beginnings of gold working in Mainland Southeast Asia. *Gold in Early Southeast Asia (Selected Papers from the Symposium Gold in Southeast Asia)*, ed. by Ruth Barnes, Emma Natalya Stein, & Benjamin Diebold. Yale University Council on Southeast Asia Studies. 125-166.
- Sagart, Laurent. 2005. Sino-Tibetan–Austronesian: an updated and improved argument. In *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Sagart, Laurent; Blench, Roger; Sanchez-Mazas, Alicia. London: Routledge Curzon. 161–176.
- Schmidt, Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie* 5:59-109.

- Shimizu, Masaaki. 2000. Khảo sát sơ lược về cấu trúc âm tiết tiếng Việt vào thế kỷ XIV-XV qua hai cứ liệu chữ Nôm [A brief survey of Vietnamese syllable structure in the XIV-XVth centuries in two Nom scripts]. In *Việt Nam Học - Kỷ Yếu Hội Thảo Quốc Tế Lần Thứ Nhất*. Hà Nội: Thế giới. 252-265.
- Shimizu, Masaaki. 2015. A reconstruction of ancient Vietnamese initials using Chữ Nôm materials. *国立国語研究所論集 (NINJAL Research Papers) 9*:135-158.
- Shimizu, Masaaki, Lê Thị Liên, and Shiro Momoki. 2005. A trace of disyllabicity of Vietnamese in the 14th century: Chữ Nôm characters contained in the inscription of Hộ Thành Mountain. *Kobe City University of Foreign Studies 64*:17-49.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Paul Sidwell, Doug Cooper, and Christian Bauer eds. Canberra: Pacific Linguistics.
- Sidwell, Paul and Roger Blench. 2011. The Austroasiatic urheimat: the southeastern riverine hypothesis. in *Dynamics of Human Diversity: The Case of Mainland Southeast Asia* (Pacific Linguistics, 627), 317-345, ed. by N. J. Enfield.
- Sidwell, Paul. 2012. Austroasiatic numerals. Data presented at the 22nd Annual Meeting of the Southeast Asian Linguistics Society, June 1st, 2012, Argay, France. Last accessed: 27 August 2020. <https://sites.google.com/view/paulsidwell/austroasiatic-numerals>.
- Sidwell, Paul. 2015a. Austroasiatic classification. in *The Handbook of Austroasiatic Languages*, ed. by Paul Sidwell and Mathias Jenny, 144-220. Boston: Brill.
- Sidwell, Paul. 2015b. *The Palaungic languages: Classification, Reconstruction and Comparative Lexicon*. Munich: Lincom Europa.
- Sinh, Trịnh. 2003. Đồng Đậu một bước nhảy vọt của nghề đúc đồng [Dong Dau: a leap in copper casting]. in *Văn Hóa Đồng Đậu: 40 Năm Phát Hiện và Nghiên Cứu (1962-2002) (Kỷ Yếu Hội Thảo Khoa Học)*, 162-171. Nhà Xuất Bản Khoa Học Xã Hội.
- The Sino-Tibetan Etymological Dictionary and Thesaurus*. The University of California, Berkeley. <http://stedt.berkeley.edu/~stedt-cgi/rootcanal.pl>. Last accessed: 15 August 2020.
- Starosta, Stanley. 2005. Proto-East Asian and the origin and dispersal of languages of east and southeast Asia and the Pacific, in *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Sagart, Laurent; Blench, Roger; Sanchez-Mazas, Alicia. London: Routledge Curzon. 182-197.
- Suwilai, Preamsrit. 1998. So (Thavung)-English-Thai Glossary, Part I. *The Mon-Khmer Studies Journal 28*:189-218.
- Suwilai, Preamsrit. 1999. So (Thavung)-English-Thai Glossary, Part II. *The Mon-Khmer Studies Journal 29*:107-132.
- Taylor, Keith. 2013. *A History of the Vietnamese*. Cambridge University Press.
- Thomas, David D. and Robert K., Jr. Headley. 1970. More on Mon-Khmer Subgroupings. *Lingua 25*:398-418.
- Trần, Trí Dõi. 2016. Từ “gần” trong tiếng Việt không phải là từ gốc Hán “近 cận” [The word “gần” ‘near’ in Vietnamese is not from Chinese “近 cận”]. *Ngôn ngữ 9* (2016):3-15.
- Trần, Trí Dõi. 2018. Một tham chiếu về nguồn gốc của tiếng Việt [A reference regarding the origin of Vietnamese]. in *Tiếng Việt Lịch Sử: Một Tham Chiếu Hồi Quan [Vietnamese Language History: A Reference]*, ed. by Đinh Văn Đức. 10-86. Hanoi: Nhà Xuất Bản Văn Học.
- Trần, Trí Dõi. 2019. Ngôn ngữ của cư dân văn hóa Đông Sơn ở Việt Nam. *Ngôn Ngữ* (2019:8-9):66-82.
- Vu, Duc Nghieu. 2019. Vietnamese Initial Consonant Clusters in Quốc Ngữ Documents from the 17th to Early 19th Centuries. *The Journal of the Southeast Asian Linguistics Society 12.1*:143-162.
- Vương, Hoàng Tuyên. 1963. *Các dân tộc nguồn gốc Nam Á ở miền bắc Việt Nam*. Hà Nội: Nhà Xuất Bản Giáo Dục.
- White, Joyce. 2011. Emergence of cultural diversity in Mainland Southeast Asia: A view from prehistory. in *Dynamics of Human Diversity*, ed. by N. J. Enfield. Pacific Linguistics. 9-46.
- Zumbroich, Thomas J. 2007. The origin and diffusion of betel chewing: a synthesis of evidence from South Asia, Southeast Asia, and beyond. *eJournal of Indian Medicine Vol. 1* (2007-2008):87-140.

## Appendix

### *Contents of the Article*

1. 160 years of research on the Austroasiatic origins of Vietnamese
2. Reevaluating Vietic and Austroasiatic etyma in Vietnamese
3. Sources of Vietic data and other comparative data
4. Issues of Vietic and Austroasiatic historical phonology
  - 4.1 Historical periods from Austroasiatic to Vietnamese
  - 4.2 Phonological changes from Austroasiatic and Vietic to Vietnamese
5. Etymological sources of Proto-Vietic and Early Vietic
  - 5.1 Chinese loanwords in early Vietic
  - 5.2 Tai loanwords in early Vietic
  - 5.3 Native etyma
6. Overview of semantic domains of native etyma in Vietnamese
  - 6.1 The physical world, body parts, and colors
  - 6.2 General actions and conditions
  - 6.3 Possessions and exchange
  - 6.4 Numerals and measure words
  - 6.5 Locational terms
  - 6.6 Pronouns and kinship terms
  - 6.7 The home and clothing
  - 6.8 Food production and materials for manufacturing
  - 6.9 Rice agriculture
  - 6.10 Betel-nut chewing and tooth-blackening
7. Insights, challenges, and future directions
  - 7.1 Linguistic affiliation and ethnolinguistic inferences
  - 7.2 Language contact and loanwords
  - 7.3 Challenges in evaluating the data and final thoughts

### References

### Appendix

Contents of the article

Abbreviations

A note on the limitations of the Vieto-Katuic hypothesis

On Proto-Language Sources in the Appendix Tables

Table A: Vietnamese versus Vietic initials

Table B: Vietic and Vietnamese syllable codas

Table C: Vietic monophthongs and Vietnamese monophthongs, diphthongs and vowels plus off-glides

Table D: Late Old Chinese and Early Middle Chinese loanwords in Vietic etyma

Table E: Possible early Tai loanwords in Vietic Vietnamese words

### *Abbreviations*

EMC = Early Middle Chinese

ESV = Early Sino-Vietnamese

LMC = Late Middle Chinese

LOC = Late Old Chinese

LSV = Later Sino-Vietnamese

MC = Middle Chinese

MKED = Mon-Khmer Etymological Dictionary

OC = Old Chinese

PAA = Proto-Austroasiatic

PV = Proto-Vietic

**A note on the limitations of the Vieto-Katuic hypothesis**

In support of a Vieto-Katuic branch, Diffloth (1989, 1992) noted shared changes from AA \*ʔ- to Katuic \*h- and Vietnamese \*s-, etyma seen specifically in those two branches, and shared phonation types. Alves (2005) offered additional lexical data and commented on prefixes in certain words. This hypothesis has been incorporated in discussion of Vietic and Austroasiatic (e.g. Nguyễn Tài Căn 1995, Sidwell and Blench 2011, Chamberlain 2018, etc.). In contrast, Sidwell (2015a) has suggested that the situation is more likely one of contact.

The Vietic data gathered in this study is the most complete to date, and the spreadsheet includes notes on lexical items shared only in Vietic. In contrast with the 40-plus items in Alves’s 2005 paper before the MKED was made available, only about a dozen have been identified through careful checking of the MKED. There also several Vietic words in the spreadsheet shared exclusively with Bahnaric. Also, as noted in Section 6.4, Proto-Vietic shares terms for numbers 5 to 9 with Bahnaric, not Katuic. Additional scattered shared words can be found in neighboring Austroasiatic languages. The best supported position is as Sidwell (Ibid.) suggests: Vietic and Katuic are two distinct branches of Austroasiatic which have been in contact. Questions of some shared features (phonation and some bits of affixation) and a modest number of shared etyma deserve attention.

Regardless, the careful sifting of the lexical data in this study requires that the claim of a combined Vietic and Katuic branch of Austroasiatic to be put aside unless a substantial quantity of data is uncovered to clearly support it.

**On Proto-Language Sources in the Appendix Tables**

Proto-Vietic reconstructions are mainly those of Ferlus 2007 in the MKED, and Proto-Austroasiatic reconstructions are those of Shorto 2006 (also available in the MKED). However, when sufficient evidence allows, I offer additional tentative reconstructed etyma for both Proto-Vietic and Proto-Austroasiatic, marked by #. Old and Middle Chinese reconstructions are those of Baxter and Sagart (2014); Proto-Tai reconstructions are noted as either from Li (1977) or Pittayaporn (2009) (and are indicated as such); proto-Austronesian items are from Blust and Trussel (2010). NA indicates that there is no available or relevant comparable reconstruction. The complete spreadsheet will be sent to readers in response to email requests sent to the author.

**Table A: Vietnamese versus Vietic initials**

Proto-Vietic	Vietnamese (Quoc Ngu)	Vietnamese	Proto-Vietic	Proto-Austroasiatic
*p	b	bay to fly	*pər	*par
*k	g	gãi to scratch	*-ka:s	*kais
*Ck	g	gạo husked rice	*r-ko:ʔ	*rk[aw]ʔ
*C+(pal)	gi	giàu rich há	*k-ʃaw	NA
*h	h	to open (mouth)	*ha:ʔ	NA (cf. Katuic *kahaa, Bahnaric *ha:)
*k	k/c	wing	*kɛ:ŋʔ	*kaiŋʔ
*kw	qu	quay stir, mix	*kwe:	NA
*gw	qu	quạt fan	*gwa:t	NA
*kv	qu	quê village	*k-ve:r	NA
*k <sup>h</sup> (Proto-Việt-Mường/Pong-Cuối)	kh	khế starfruit	*k <sup>h</sup> e:ʔ	NA
*l	l	lá lách spleen	*la:ʔ ~ la:	*slaʔ
*Cl	l	lè to pull out tongue	*t-lɛ:l	#IVVI (cf. Bahnaric, Katuic, Khasic, Aslian)
*m	m	mật gall	*məc	*k.mət
*Cm	m	mọt termite	*k-mɔ:c	*kmuət
*ʃ	m	múc to draw water	*ʃa:lʔ	*ʃək

Proto-Vietic	Vietnamese (Quoc Ngu)	Vietnamese	Proto-Vietic	Proto-Austroasiatic
*d	n	năm five	*dam	NA
*Cd	n	năng to dry on fire	*p-dan	NA
*Cn	n	nanh eye tooth, tusk	*k-nɛ:ŋ	#Cnap
*n	nh	nhà house	*na:	NA
*C-n	nh	nhá / (nhai) to chew, chew tobacco	*s-na:ʔ	NA
*ml	nh	nhặt, lặt to chew, chew tobacco	*m-la:c	NA
*f	nh	nhau placenta	*faw	NA
*ŋ	ng	ngồi to sit	*ŋu:j	*ŋguj
*Cŋ	ng	ngái far	#s-ŋa:jʔ	#cŋaaj (Bahnaric *cŋa:j, Katuic *cŋaaj, Monic *chəŋəj, Palaungic *sŋaaj)
*p <sup>h</sup>	ph	pha to dilute, mix	*p <sup>h</sup> a:	NA
*Cr	r	ruồi fly (insect)	*m-rɔ:j	*ruj
*Cs	r	rắn snake	*p-səŋʔ	*[b]səŋʔ
*r	r	ruột intestines	*rɔ:c	*ruuc
*Cr	s	sấm thunder	*k-rəmʔ	*grəm[ʔ]
*s	t	tóc hair	*-suk	*suk
*t <sup>h</sup>	th	thổi to blow	*t-hu:s > tu:s / t <sup>h</sup> u:s	NA
Cl	tr	trái fruit	*p-le:ʔ > ple:ʔ / tle:ʔ	*pləjʔ
*v	v	vặn to twist/wring	*vapʔ	*wɪŋ
*C+(lab)	v	vôi lime (mineral)	*k-pu:r	*knpur
*p	v	vắt to press (fruit)	*pat	*pit
*c	ch	cháu grandchild	*cu:ʔ	*cuuʔ
*Cʔ	zero	óc brain	*c-ʔɔ:k	*ʔuək
*tʃ	x	xương bone	*tʃ-ʔa:ŋ > ja:ŋ / tʃiəŋ	*cʔaəŋ
*c	d	dứa pineapple	*-ca:ʔ > -ciaʔ	NA
*Ct	d	dựng to build	*pr-təŋʔ > p-dəŋʔ	NA

Table B: Vietic and Vietnamese syllable codas

Proto-Vietic	Vietnamese (Quoc Ngu)	Vietnamese	Proto-Vietic	Proto-Austroasiatic
*-k	-c (/k/)	sóc squirrel	*p-rɔ:k > k <sup>h</sup> ɔ:k	*prɔək
*-n	-n	hắn he, she	*hanʔ	*[ʔ]anʔ
*-ŋ	-n	đan to weave	*ta:ŋ	*taap
*-ŋ	-nh (/p/)	tranh thatch-grass	*p-lɛŋ	NA
*-ŋ	-ng (/ŋ/)	nướng to roast	*da:ŋʔ	*tlaŋ
*-m	-m	chim bird	*-ci:m	*ciim
*-c	-t	rút to pull out (with hands)	*p-ru:c	#p-ruuc (Katuic *pooc pull out; West Bahnaric *tu:c; Khmer rò:c)
*-t	-t	vứt to throw away	*-vət	*wat
*-k	-c (/k/)	bóc to peel	#pɔ:k	*pɔək
*-p	-p	ngáp to yawn	*sŋ-ʔa:p > s-ŋa:p / -ʔa:p	*sʔaap
*-j	i/y (/j/)	cây civet, dog	#gəj	*cgəj
*-l	i/y (/j/)	cối mortar (for rice)	*t-ko:lʔ	*guul
*-r	i/y (/j/)	cúi ( heo cúi )	*gu:rʔ   ku:rʔ	NA
*-s	i/y (/j/)	gãi to scratch (from itch)	*-ka:s	*kais
*-w	-o/u	beo / báo panther	*pe:w	NA

Proto-Vietic	Vietnamese (Quoc Ngu)	Vietnamese	Proto-Vietic	Proto-Austroasiatic
*-h	-Ø	gõ to knock/rap	#gɔ:h	*g[uə]h
*-ʔ	-Ø	lá leaf miệng mouth	*s-la:ʔ *mɛ:ŋʔ	*slaʔ *miəŋʔ

**Table C:** Vietic monophthongs and Vietnamese monophthongs, diphthongs and vowels plus off-glides

Proto-Vietic	Vietnamese	Vietnamese Examples	Proto-Vietic	Proto-Austroasiatic
*a	a	ba three	*pa:	*piʔ (or #pee)
	ã	đắng bitter	*taŋʔ	*kt2aŋ
	ua	muợn to borrow	*ma:ŋʔ	*smaaŋ to ask
*a:w	ao	áo placenta	*ʔa:wʔ	NA
*ɔ:	o	mỏ beak	*k-ɔəh	*jɔbuəh (Katuic *crɔəh; Khmer campuh)
	ua	ruột intestines	*rɔ:c	*ruuc
*ə	â	âm lukewarm	*s-ʔəmʔ	NA
	ư	dựng to build	*pr-təŋʔ > p-dəŋʔ	NA
*ə:	ơ	giơ to lift (ones hand)	*k-jə:	NA
*e:	ê	hết finished	*he:t	*ʔət
	ây	lấy to take	*le:ʔ / -le:ʔ	*liʔ (only Palaungic *leʔ and Vietic)
	ai	ai who	*ʔe:	NA
*ɛ:	e	bẻ to break in two, to pick	*pɛh	NA
	ia	ỉả to defecate	*ʔɛh	NA
	iê	miệng mouth	*mɛ:ŋʔ	NA
*i:	i	chín ripe	*ci:nʔ	*ciinʔ
	ây	vây fin	*c-pi:	NA
*o:	ô	công peacock	*k-vo:ŋ	#k-vo:ŋ (Aslian, Katuic, Palaungic, Vietic)
	ao	gạo rice (uncooked)	*r-ko:ʔ	*rk[aw]ʔ
*u:	u	mũi nose	*mu:s	*muuh
	ô	ngồi to sit	*ŋu:j	*ŋguj
	au	cháu grandchild	*cu:ʔ	*cuuʔ
	âu	gấu bear	*c-gu:ʔ / c-ku:ʔ	*c-gu:ʔ / c-ku:ʔ (Vietic, Katuic, Bahnaric only)

**Table D:** Late Old Chinese and Early Middle Chinese loanwords in Vietic etyma

(OC=Old Chinese, MC=Middle Chinese, LSV=Later Sino-Vietnamese (i.e. standard Sino-Vietnamese))

Viet.	Gloss	Vietic	Late Old Chinese to Early Middle Chinese
giống	seed	*k-co:ŋʔ	種 zhong3 seed, OC *k.toŋʔ, MC tsyowngX, LSV chủng, chủng
lúa	paddy rice	*ʔa-lɔ:ʔ	稻 dào paddy, OC *[l]ʔuʔ, MC dawX, LSV đạo
cải	cabbage mustard	*ka:s	芥 jiè/gài, OC *kʰr[e][t]-s, MC keajH, LSV giới
cà	eggplant	*ga:	茄 qié, jiā, OC NA, MC, NA, LSV gia
mè	sesame	#mV:	麻 má hemp (sesame in compound), OC C.mʰraj, MC mae, LSC ma
vãi	scatter (rice) (v)	*va:s	播 bō to sow, OC *pʰar-s, MC paH, LSV bá
kén	cocoon	*kɛ:nʔ	繭 jiǎn, OC *kʰenʔ, MC kenX, LSV kiến
đực	male (of animal)	*dək	特 de2, OC *dʰək, MC dok, LSV đặc

Viet.	Gloss	Vietic	Late Old Chinese to Early Middle Chinese
ngựa	horse	*m-ŋə:ʔ	午 wǔ year of horse, OC *[m].q <sup>h</sup> aʔ, MC nguX, LSV ngo
dao	knife	*-ta:w	刀 dāo knife, OC *C.t <sup>h</sup> aw, MC taw, LSV đao
bừa	harrow	#ba:	耙 pa2 rake (n.), OC *[b] <sup>h</sup> ra, MC bae, LSV bà
giùi	awl/drill (cf. mallet)	*k-cu:ʔ	椎 chuí hammer, OC *k.druij, MC drwij, LSV trùy, chuy
liềm	sickle	*liem	鎌 lián, OC *[r]em, MC ljem, LSV liềm
gan	liver	*t-ka:n	肝 gān liver, OC *s.k <sup>h</sup> a[r], MC kan, LSV can
phổi	lungs	*p-so:s > p-ho:c > p <sup>h</sup> o:c / po:c (need final *-l)	肺 fèi lungs, OC *p <sup>h</sup> [a][t]-s, MC phjojH, LSV phé
lụa	silk	*luaʔ	縷 lǚ silk, OC *[r]oʔ, MC ljuX, LSV lữ/lâu)
kim (EMC) / ghim (LOC)	needle	*ki:m	鍼 zhēn needle, OC *t.[k]əm, MC tsyim, LSV châm
nhuộm	dye/lacquer (teeth) (v)	*ŋo:mʔ	染 rǎn to dye, OC *C.n[a]mʔ, MC nyemX, LSV nhuộm
giường	bed	*k-ʒə:ŋ	床 chuáng bed, OC *k.dzraŋ, MC dzrjang, LSV sàng
mả (LOC) / mồ (EMC)	grave (n)	*-mah	墓 mù grave (n.), OC *C.m <sup>h</sup> ak-s, MC muH, LSV mộ
giấy	paper	*k-cajʔ	紙 zhǐ paper, OC *k.teʔ, MC tsyeX, LSV chi
góa	widowed	*-kwa:ʔ	寡 guǎ widowed, OC *[C.k] <sup>w</sup> raʔ, MC kwaeX, LSV quả
vợ	wife	*-bə:ʔ	婦 fù woman, wife, OC *mə.bəʔ, MC bjuwX, LSV phụ
mụ old woman	woman, old	*mu:ʔ	LOC/EMC? (姥 mu3 old woman, OC NA, MC NA (shangsheng word), LSV mỗ, mụ
cậu	mother's younger brother	*gu:ʔ	舅 jiù mothers brother, OC *[g](r)uʔ, MC gjuwX, LSV cữu
chị	elder sister	*ji:ʔ	姊 zǐ elder sister; OC *[ts][i]jʔ; MC tsijX, LSV tỉ
sắt	iron (metal)	*k-rac	鐵 tiě iron, OC *ʃik, MC thet, LSV thiết
gang iron	steel	*t-ka:ŋ	鋼 gāng steel, OC *C.k <sup>h</sup> aŋ, MC kang, LSV cương
bạc white / silver	silver	#ba:k	白 bái white, OC *b <sup>h</sup> rak, MC baek, LSV bạch
xanh	blue-green	#eəŋ	青 qīng blue-green, OC *N-s <sup>h</sup> <r>eŋ, MC tsheng, LSV thanh
vàng	yellow/gold	#C-wa:ŋ	黃 huáng yellow, OC *N-k <sup>w</sup> aŋ, MC hwang, LSV hoàng
bạc	white	*ba:k	白 bái white, OC *b <sup>h</sup> rak, MC baek, LSV bạch

**Table E:** Possible early Tai loanwords in Vietic Vietnamese words (some uncertainty of direction of borrowing)

Vietnamese	Gloss	Vietic	Tai	Austroasiatic
ra / rựa	knife / bush-knife	*m-ra:ʔ	*vraC sword (Li)	#mraaʔ (Bahnaric *bra:, Katuic *braa, Khmuic *bra:, Mangic (Mang pja:ʔ), Monic *mraaʔ, Palanguic (Waic *plaʔ))
mương	ditch/canal	*-miəŋ	* <sup>h</sup> muəŋA (Pitt.)	NA
vịt	duck	*vi:t	*petD	NA
bè	raft (n)	#bɛ:	*bɛ (Li) (cf. 排 pái raft)	#bɛ: (tentative: Bahnaric (Jeh, Laven, Tarieng bɛ:); Katuic (Bru pɛ:, Kui phɛ:; Khmeric (Khmer phe:); Khmuic (Mlabri

Vietnamese	Gloss	Vietic	Tai	Austroasiatic
				bεε); Monic (Nyah Kur phèε); Palaungic (Lampet phe:); Pearic (Chong p <sup>h</sup> ε:))
muống	water spinach	*bɔ:ŋ?	*bɔŋC water spinach (Pitt.)	NA
máng	water pipe of bamboo	*ba:ŋ?	*baŋB/C tube, bamboo (Pitt.)	NA
què	crippled / lame	#gwε: (tentative, Tai loan)	*gwεA (Li)	Tai loan? (Bahnar kwεε; Khmer khe:; Mon khe?)
trống	drum (n)	#klo:ŋ?	*klɔŋA (Li)	NA
mùng (dialectal)	net (for mosquitos)	#mu:ŋ	Thai múŋ	#mu:ŋ (Bahnaric (Laven muŋ); Katuic (Katu, Ngeq, Pacoh muŋ); Khmer muŋ; Monic (Nyah Kur muŋ); Pearic (Pear muŋ; Chong mu:ŋ))
đập	thresh/husk rice (v)	*dəp	*tɔpD to slap (Pitt.)	NA
lội	to wade/swim	#lo:j?	*lo:jA 'to swim' (Pitt.)	*luj? (Tai-Austroasiatic contact?)

**Reviewed:** Received 1 October 2020, revised text accepted 15 November 2020, published 10 December 2020

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