



**HAL**  
open science

## LANGUAGE FAMILIES OF SOUTHEAST ASIA

Laurent Sagart

► **To cite this version:**

Laurent Sagart. LANGUAGE FAMILIES OF SOUTHEAST ASIA. Charles F.W. Higham and Nam C. Kim. The Oxford Handbook of Southeast Asian Archaeology, pp.321-338, 2022, The Oxford Handbook of Early South-east Asia, 978-0-19-935535-8. <10.1093/oxfordhb/9780199355358.001.0001>. <hal-03099922>

**HAL Id: hal-03099922**

**<https://hal.science/hal-03099922v1>**

Submitted on 6 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# LANGUAGE FAMILIES OF SOUTHEAST ASIA

Laurent Sagart, CNRS and INaLCO, Paris

## <1> Introduction

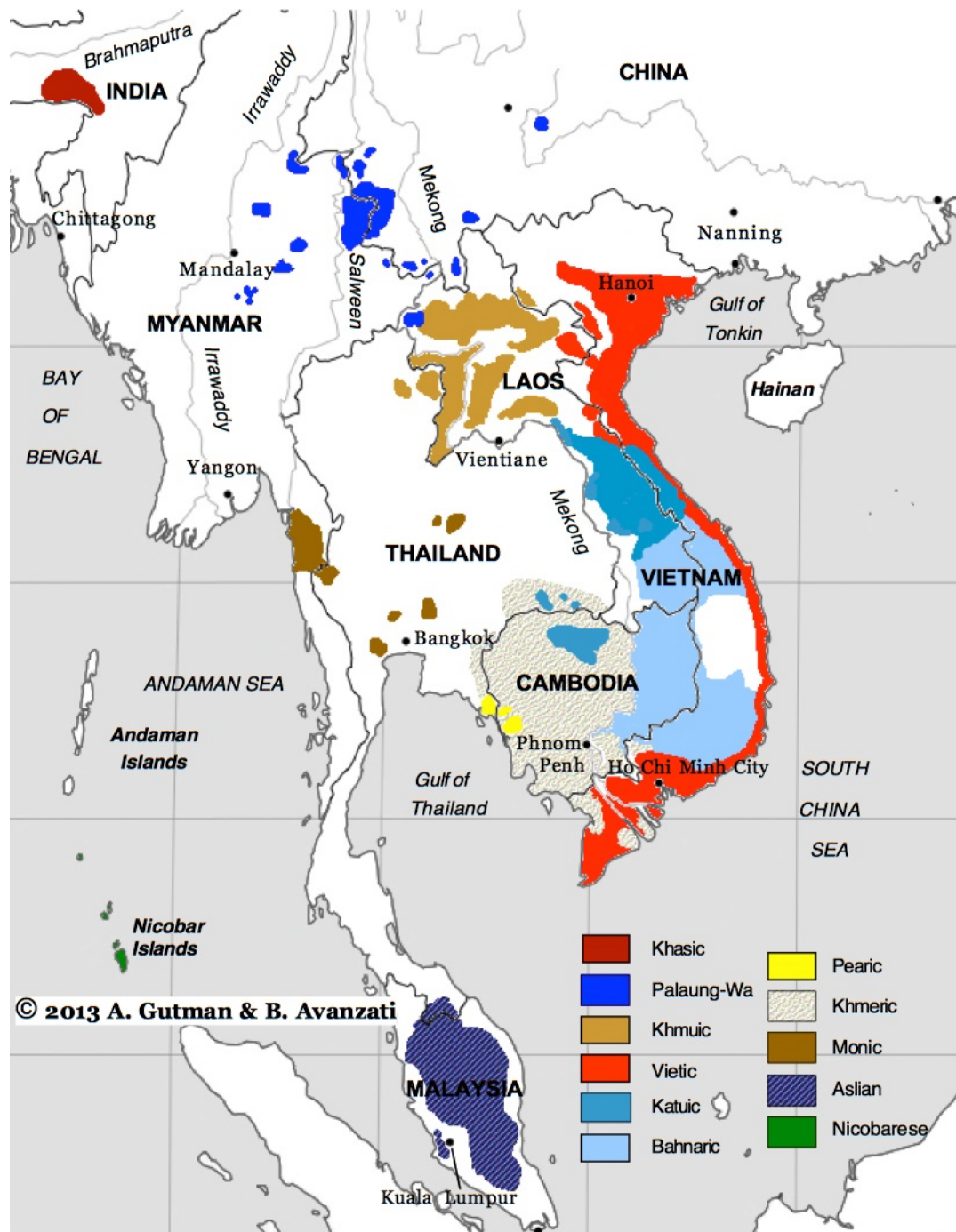
Five language families are represented today in the region and in south China:

Austroasiatic, Austronesian, Tai-Kadai (also ‘Kra-Dai’), Hmong-Mien (formerly ‘Miao-Yao’) and Sino-Tibetan (sometimes now ‘Trans-Himalayan’). With the exception of Austronesian, which is only monophyletic with the addition of Tai-Kadai (below), there is broad agreement among linguists that each of these groups goes back to a private ancestral language. There is broad agreement also as to the affiliation of most of the individual languages in the region. The Andamanese languages and Kusunda, a moribund language of Nepal, are isolates.

## <2> Austroasiatic

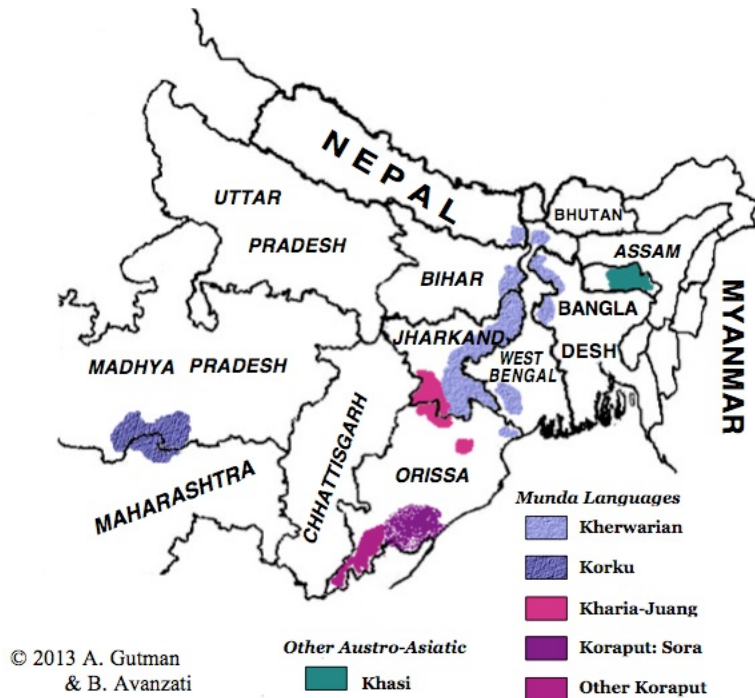
The Austroasiatic family extends from southeast Asia (‘Mon-Khmer’, map 1) to north India (‘Munda’, map 2). The family’s distribution is characterized by extreme territorial fragmentation: Austroasiatic has become dislocated under the penetration of Tai-Kadai, Hmong-Mien, Austronesian and Sino-Tibetan. Lepcha or Rong, a Tibeto-Burman language of eastern Nepal, and Acehnese, an Austronesian language of NW Sumatra, include an Austroasiatic substratum. Witzel’s claims about an Austroasiatic presence in pre-Indo-Aryan northeastern India (Witzel 1999), based on apparently prefixing loanwords of unknown origin in the Rg-Veda, seem quite speculative. Several of the Austroasiatic etymologies of Chinese words which Norman and Mei (1976) say indicate

an old Austroasiatic presence in south China and the Yangtze Valley are errors (Sagart 2008), yet the Chinese name of the Yangtze River: jiāng 江, Old Chinese \*kʰroŋ, has been regarded as an Austroasiatic toponym (compare Old Mon *kruŋ* ‘river’ and similar forms). There are, however, probable Sino-Tibetan cognates (STEDT database, set #2322).



Map 1. The Mon-Khmer languages. Source: The Language Gulper

(<http://www.languagesgulper.com/eng/Austromap.html>).



Map 2. The Munda languages. Source: The Language Gulper (<http://www.languagesgulper.com/eng/Mundamap.html>).

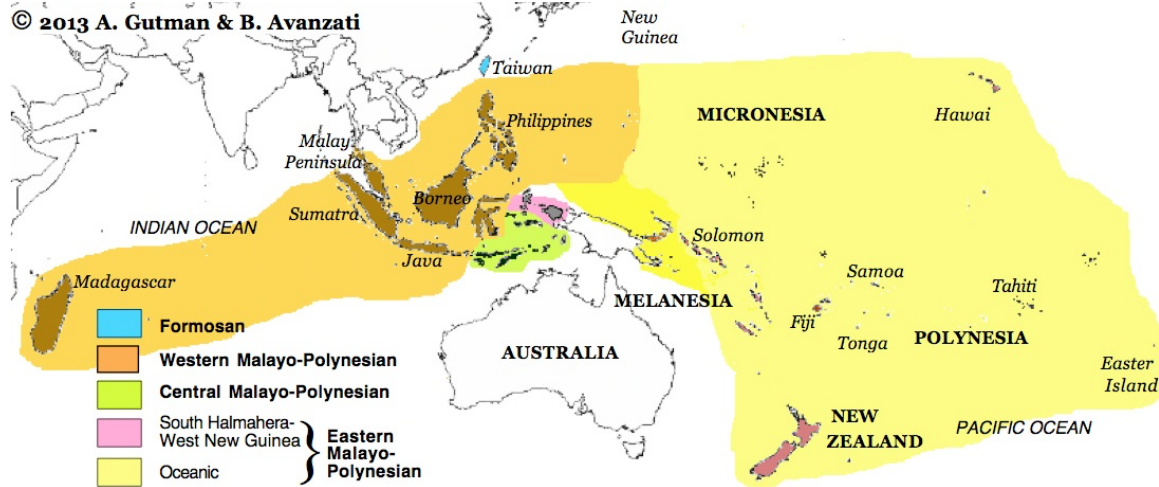
Austroasiatic linguistic typology is of an East Asian type: monosyllables, or disyllabic words with final stress and reduced first syllables; mostly derivational morphology using prefixes and infixes; object-final word order. The Munda languages show evidence of having adapted to the south Asian type after the breakup of proto-Munda: this implies an East Asian origin in a region not too far removed from the Sino-Tibetan and Austronesian homelands. A small number of basic vocabulary items shared by Austroasiatic and Hmong-Mien have suggested to Pejros and Shnirelman (1998:155) and others that the two are genetically related. There is also evidence of a shared Y-chromosome haplotype between Hmong-Mien and Austroasiatic peoples, and the same genetic marker has been found in skeletons from Daxi culture (Li et al. 2007). However, an episode of intimate

contact between the two groups in the middle Yangtze is an alternative explanation. Until about 2000 the dominant view was that the family had two branches: Munda vs. all the rest ('Mon-Khmer'). Diffloth (2005) carved a third, northern branch ('Khasi-Khmuic') out of Mon-Khmer; he identified elements of a diversified vocabulary of rice and argued from plant and animal names for a tropical homeland, in the Burma/Bengal region at an impressionistic 5000 BCE. Sagart (2011a) notes that the Austroasiatic vocabulary of rice is independent from the other East Asian groups, implying an independent Austroasiatic domestication of rice. Arguing from negative evidence, Sidwell (2008) presents the family's phylogeny as a rake of a dozen or so equidistant branches; yet failure to detect tree-like structure can be due to problems with cognate-word encoding and should not be regarded as final. Taking consideration of the Austroasiatic subsistence vocabulary reconstructed by Shorto (2006)—words for *taro*, *yam*, *rice* as well as for the *boat*—, Sidwell and Blench (2011) propose that the Austroasiatic family broke up after group(s) of hunter-gatherers practicing tuber-culture in the Mekong valley acquired rice agriculture. They argue that the arrival of rice, c. 4100 BP in the northern part of region, precedes and partly triggers the Austroasiatic dispersal, which they place at c. 3800 BP. However Sidwell and Blench's assignment to Proto-Austroasiatic of individual vocabulary items is intimately dependent upon Austroasiatic phylogeny, which is disputed, as we have seen. The circumstances surrounding the formation of Austroasiatic await clarification.

## <2> Austronesian

Most Austronesian languages are spoken in insular southeast Asia and the Pacific (map

3).



Map 3. The Austronesian languages. Source: The Language Gulper

(<http://www.languagesgulper.com/eng/Austronesmap.html>).

There is broad agreement among linguists that the homeland was in Taiwan, at the northern end of the family's territory, and where the highest diversity in the family is found. The sound correspondences across Austronesian languages are relatively well understood, allowing for several good-quality Proto-Austronesian reconstructions by Tsuchida, Blust, Wolff. All the Austronesian languages outside of Taiwan ('Malayo-Polynesian') share linguistic innovations, e.g. the change of \*S to an h-type sound and the replacement of \*S by \*s in 'nine'. This shows that speakers of an early Austronesian language of Taiwan, in which these and other linguistic changes were already completed, established settlements, presumably in the northern Philippines, out of which the rest of the Austronesian world was eventually settled. Sagart (2008) argues from shared innovations in the system of numerals that Proto-Malayo-Polynesian was part of a southern Formosan group. Hung (2008) similarly notes that southern Taiwan is the

precursor of the earliest neolithic sites in the northern Philippines; Chang et al. likewise (2015) demonstrate a south Taiwan origin for the domesticated paper mulberry carried by the expanding Austronesians. Remarkably consistent dates for the initial Austronesian settlement of Taiwan and for the out-of-Taiwan event come from archaeology (Hung 2008), linguistic phylogenetics (Gray et al 2009) and population genetics (Ko et al. 2014): c. 3500-3000 BCE and c. 2000 BCE. Judging from the reconstructable Proto-Austronesian vocabulary, the first Austronesians built *houses* (\*Rumaq), *wove* cloth and baskets (\*tenun ‘to weave’), had *boats* (\*qaCu, \*qabaŋ), fished with *fishnets* (\*aray), practiced hunting/warfare with *bows* (\*buSuR), kept *dogs* (\*asu), and raised *pigs* (\*beRek); they cleared *swiddens* (\*qumah) to grow *Setaria italica* (\*beCeŋ), *Panicum miliaceum* (\*baCaR) and *japonica rice* (\*pan<sup>y</sup>ay), as confirmed by finds of charred grains at Nan Kuan Li and Nan Kuan Li East, two 3rd mill. BCE sites on the southwest coast of Taiwan (Tsang et al. 2017). Proto-Austronesian is usually equated with the Ta-Pen-K’eng culture in Taiwan (Hung 2008). The absence of archaeological antecedents and the similarity with contemporary archaeological sites on the mainland side of the Taiwan straits argue that the first Austronesians reached Taiwan from the mainland by boat in the late 4th millennium BCE. Analysis of the evolution of the numeral systems shows that the first Austronesian languages to branch off were those in northwest Taiwan (Sagart 2004), where the straits are narrowest and Taiwan is visible from the mainland: this was probably where the first Austronesians set foot. A northern point of entry is confirmed by human and plant genetics (Ko et al. 2014; Chang et al. 2015).

In the post-Taiwan period, the pace of Austronesian settlement quickened, thanks

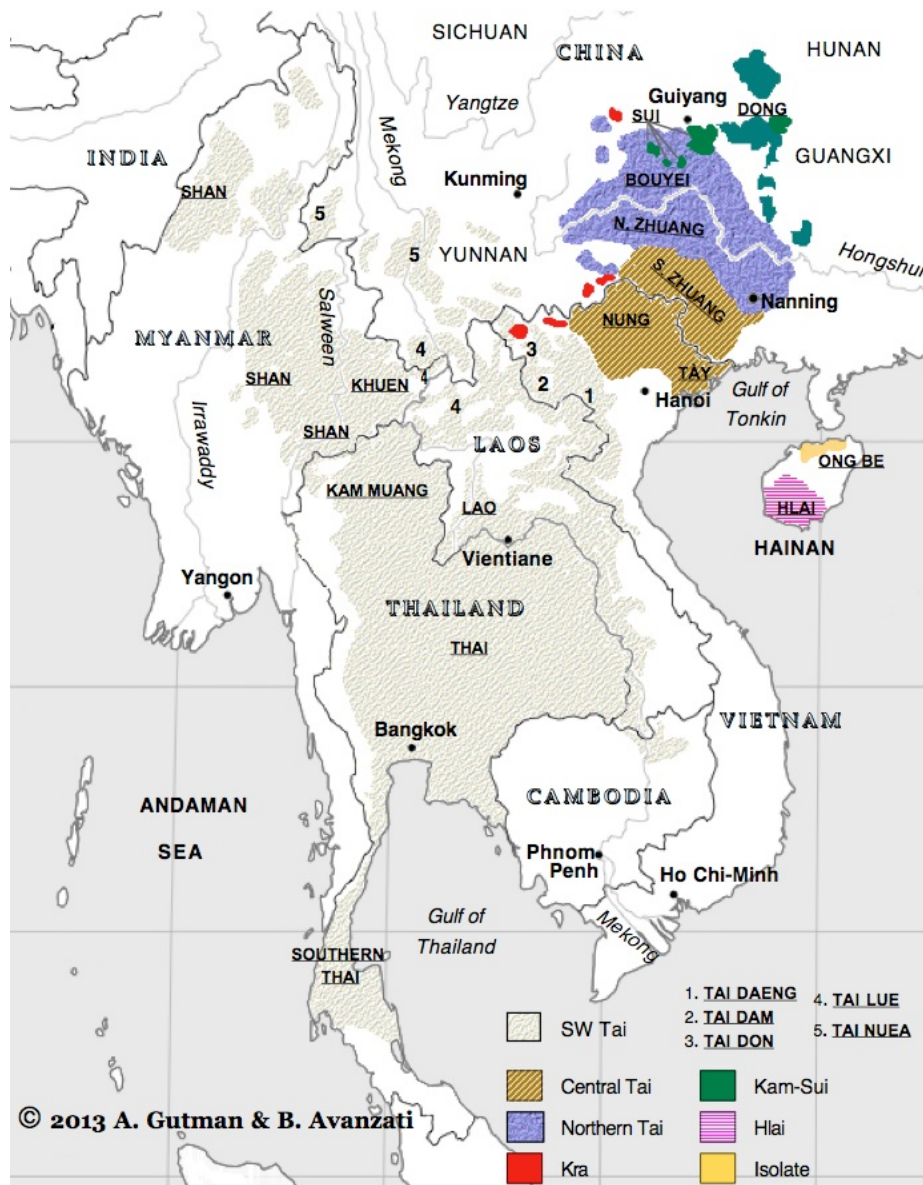
apparently to improvements in nautical technology. The Philippines were settled, followed by Borneo, Sumatra, the Celebes, the Sunda islands, Maluku, Timor, New Guinea and from there the rest of the Oceanic world: Micronesia, Melanesia and Polynesia. Speakers of proto-Chamic, a language related to Malay possibly from Borneo, may be behind the Sa Huynh culture on the coast of Vietnam. A later migration by boat out of Borneo brought Austronesian speakers in contact with the Bantu languages on the east coast of Africa and Madagascar c. 400 CE (Adelaar 2009).

Outside of Taiwan, the Austronesians received gene flow from preexisting populations who also at times shifted to Austronesian languages: this is sometimes interpreted as evidence for an Austronesian presence in the Sahul plate region in Paleolithic times: that cannot be true in a linguistic sense, at least. Cultural contact introduced the expanding Austronesians to new food resources such as the banana, sago and yam: the availability of these resources allied to changes in the natural environment led to rice and millet being abandoned as the Austronesians reached New Guinea. The constant factor throughout the Austronesian expansion, at least since the Formosan stage, was fishing with gathering of marine resources. A study of dental health among the earliest Austronesians (Pietruszewsky et al. 2014) finds that marine products, rather than grain, formed the basis of the diet at Nan Kuan Li. Shell gathering clearly had economic importance among genetic pre-Austronesians just north of Taiwan c. 6000 BCE (Ko et al. 2014) .

## <2> Tai-Kadai

The Tai-Kadai languages are spoken in south China and mainland southeast Asia (map 4).

The area of highest diversity, and probable Tai-Kadai homeland, is in south China, especially in Guangxi province, Hainan island, and adjacent areas in Guangdong, Guizhou and more marginally Yunnan and north Vietnam. Tai-Kadai toponyms in Cantonese-speaking areas of western Guangdong province show that there was once continuity between the Tai-Kadai speaking areas in Guangxi and in Hainan island.



Map 4. The Tai-Kadai languages. Source: The Language Gulper (<http://languagesgulper.com/eng/Taikadaimap.html>).

The Cantonese dialect of Chinese has certain unique features which clearly indicate a Tai-Kadai substratum, such as a distinction between long and short vowels, otherwise unknown in Chinese. All the languages outside of that region: in Vietnam, Laos, Cambodia, Thailand, Burma, eastern Yunnan and Assam, including standard Thai, belong to the low-level subgroup southwestern Tai. The family's most recent common ancestor may have been spoken as recently as 3000-2500 years ago. The historically documented southward and westward spread of southwestern Tai occurred much later, in the late first millennium CE and the early second. Four subdivisions of the family are recognized: the Tai group, which includes the southwestern languages just mentioned, as well as other languages in north Vietnam (Nung, Tho) and the Chinese provinces of Guangxi and Guizhou, (Zhuang, Buyei); the more northerly Kam-Sui group in eastern Guizhou, western Hunan and northern Guangxi (Dong, Sui etc.); the Hlai languages of Hainan island in the southeast; and the small Kra languages, scattered principally at the family's northwestern periphery. The Tai and Kam-Sui groups form the Kam-Tai branch of Tai-Kadai, on the same taxonomic level as Hlai and Kra. Reconstructions are available for each branch but none as yet exists for Proto-Tai-Kadai.

The Chinese conquest of the regions south of the Yangtze and the establishment c. 204 BCE of the Chinese-led Nanyue state with its capital at Pan-yu in the Pearl River delta for the first time brought the Chinese language in contact with Tai-Kadai on a significant scale. Since then cultural domination by Chinese on Tai-Kadai in south China has not ceased. The effects on the Tai-Kadai languages have been far-reaching but also not

uniform, an indication that Tai-Kadai was already diversified at Chinese contact. Like Vietnamese, also subjected to long-term Chinese influence, the Tai-Kadai languages have become structurally similar to Chinese in word structure: monosyllabism, few affixes, a three-tone system; and in basic sentence structure. The Tai-Kadai languages have naturally absorbed a large number of Chinese loanwords, principally relating to cultural notions. The permeability of Tai and Kam-Sui has been greatest: in contrast to Hlai and Kra, they have replaced their indigenous numerals with the Chinese ones. Possibly the 2nd-century BCE precursor of the (Kam-)Tai group was spoken at or near the Nanyue capital, while the precursors of the Hlai and Kra groups were southern and western provincial dialects.

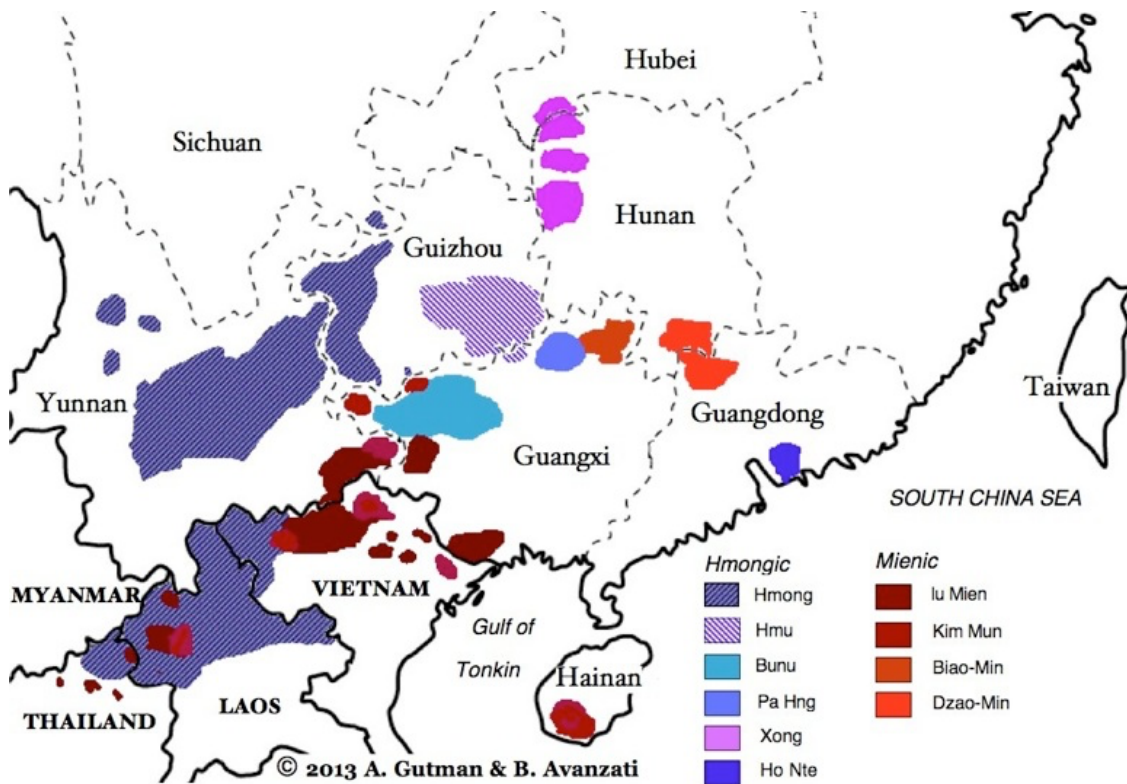
The Tai-Kadai languages have a history of contact with the Austroasiatic languages. Few loanwords can be found in all the Tai-Kadai branches, like ‘ant’: Proto-Tai \*mɯc D, Proto-Hlai \*amuc, Proto-Kra \*mot D, Proto-Kam-Sui \*mwit 8. This word’s source is evidently Shorto’s Proto-Mon-Khmer \*s<m>uuc ‘stinging insect; ant’, an infixed derivative of \*suuc ‘to sting’. Another insect name borrowed from Austroasiatic is ‘termite’, Proto-Tai \*mo:t ‘termite’ < Proto-Mon-Khmer \*kmuət ‘woodworm, weevil’. Relatively basic items were also borrowed: ‘brain’, Proto-Tai \*ʔe:k < Proto-Mon-Khmer \*ʔuək; ‘egg’, Proto-Kra \*təmA ‘egg’, Proto-Tai \*tram (‘testicles’) < Proto-Mon-Khmer \*ktɨəm; ‘elder sibling’ Proto-Tai \*bi: B < Proto-Mon-Khmer \*mbiiʔ, suggesting a period of intimate contact with intermarriage. Particularly interesting are agriculture-related terms: ‘rice’, Proto-Tai \*Ç.qaw C < Proto-Mon-Khmer \*rk[aw]ʔ/\*rkaawʔ ‘husked rice’; ‘swidden’, Proto-Tai \*rɯj B < Proto-Mon-Khmer \*sreʔ ‘field’ (note, however, that the

Tai-Kadai name of the wet rice field: \*na A, is likely from Proto-Austronesian \*bena ‘lowland field’, like Proto-Tai \*sal A ‘husked rice’ < Proto-Austronesian \*qasaN and Proto-Tai \*C.wal B ‘to sow broadcast’ < Proto-Austronesian \*sabeR; see below on the Tai-Kadai-Austronesian connections). The name of the water buffalo: Proto-Tai \*ɣwa:jA, Proto-Kra \*kwai A, is from Proto-Mon-Khmer \*[k].[b]ay ‘gaur, water buffalo’. Likewise for the Tai name of the banana: Proto-Tai \*kluəj C, from Proto-Mon-Khmer \*t<sub>1</sub>luəyʔ ‘banana’. The Proto-Tai artefact names such as string \*sa:i A and the hunting crossbow \*<sup>h</sup>nwɿ:C are also of Austroasiatic provenience: Proto-Mon-Khmer \*ks[i]ʔ; \*ksih ‘string, cord, rope, thread’; \*snaʔ ‘crossbow’. That loans from Austroasiatic usually do not occur in all the Tai-Kadai branches suggests high-intensity contact between the two stocks following the breakup of Proto-Tai-Kadai. The cultural vocabulary of Tai-Kadai indicates knowledge of agriculture: a word for *Setaria italica* (Proto-Tai \*<sup>h</sup>wuəŋ C; Proto-Hlai \*apa:ŋ C) is present. How it relates to Proto-Austronesian \*beCeŋ ‘*Setaria italica*’ is uncertain. Although the name of the rice plant is not reconstructable—possibly because the Kam-Tai word has been lost to a loan from Austroasiatic (above)—there are words for the paddy field and for dehusked rice (above); taro (Proto-Tai \*pruək; Proto-Kra \*p-ɣak) had economic importance. A word for ‘canoe’ (Proto-Tai \*C.rwuə A, Proto-Kra \*da A, Proto Hlai \*ura A) is present. Both are likely Austronesian words: Proto-Austronesian \*biRaŋ ‘broad leaf’ and \*aluja ‘paddle’ respectively. If consideration is taken of Tai-Kadai cultural terms of Austronesian origin, which were in all likelihood part of Proto-Tai-Kadai even when they are preserved in a single Tai-Kadai branch (below), one has verbs for ‘to bury seeds in the ground’, ‘to sow broadcast’, ‘to winnow’, ‘pestle’; and a diversified vocabulary of fishing: line fishing (Proto-Tai \*bet ‘fish hook’ < Proto-

Austronesian \*kabit), fishing with poison (Proto-Tai \*C.buə A ‘to poison fish’ < Proto-Austronesian \*tuba), and net fishing (Proto-Hlai \*aRəi C ‘fish net’ < Proto-Austronesian \*aray).

## <2> Hmong-Mien

The small Hmong-Mien family has two branches: Hmongic and Mienic, distributed principally in south China and accessorially in the northern parts of Vietnam, Laos and Thailand (map 5). Hmong-Mien presence outside of the boundaries of China is but a few hundred years old. Mienic is smaller and tighter, and its geographical distribution is more southerly and easterly, than Hmongic. In China the Hmong-Mien languages are spoken in Hunan, Yunnan, Guizhou, Guangxi, Guangdong and in Hainan island. The center of diversity lies north of the tropic of Cancer, perhaps around northern Guangxi. As with Austroasiatic, Hmong-Mien language geography is marked by strong discontinuity. Here the dislocating factor has been the influx of Chinese speakers following the conquest of south China by Emperor Qin Shi Huang in the late 3rd century BCE. As with Tai-Kadai, Chinese political dominion over the Hmong-Mien languages has not ceased until now: the result, again, has been an evolution towards Chinese linguistic type: a tendency towards invariable monosyllabic words with three tones, loss of affixation patterns, and object-final sentences.



Map 5. The Hmong-Mien languages. Source: The Language Gulper (<http://languagesgulper.com/eng/Hmongmap.html>).

Ratliff's reconstruction of Proto-Hmong-Mien (2010) includes numerous Chinese loanwords relating to metallurgy, agriculture, food and cooking, crafts, economy, the calendar, as well as names of domesticated plants and animals and terms for various artefacts. These point to intense contact between the two groups in the later part of the Old Chinese period, not before c. 300 BCE judging from the phonetic shapes of the words (tones already present; Old Chinese nasal preinitials still pronounced, without having induced voicing in following voiceless stops). The Proto-Hmong-Mien numerals present several historical layers: it is understood that '2' and '3' are indigenous, that '3' to

'9' are borrowed from Tibeto-Burman and that '10', '100' and '1000' are from Chinese. This shows that pre-Proto-Hmong-Mien was exposed to Tibeto-Burman influence before entering into contact with Chinese. The case of '1', Proto-Hmong-Mien \*ʔi, is interesting: it resembles both the Chinese form, Old Chinese \*ʔi[t], and the Proto-Tibeto-Burman form: \*it. For that reason it is often treated as a loanword (loss of final -t in Chinese loanwords has parallels). However numerals are normally borrowed 'from the top down': it is implausible that Hmong-Mien should retain '2' and '3' but borrow '1': more probably '1' is inherited (Table 2 below). The Hmong-Mien homeland must have been both in contact with Chinese and westerly enough for significant interaction with Tibeto-Burman. Niederer's proposal (1998:22-23) of a homeland area corresponding approximately to the ancient state of Chu, north of the Yangtze, at a time-depth of at least 2000 years B.P. satisfies these requirements. The Qujialing-Shijiahe culture (5400-4000 B.P.) could be an early precursor. The natural environment included *ice*, bodies of water (*boat, cross a river*), the *bear, pangolin, porcupine, tortoise, eagle/hawk* and *bamboo*. The vocabulary of subsistence is for the most part unremarkable in a late first-millennium BCE central/western China context. The Proto-Hmong-Miens grew a variety of *crops: rice* and '*millet*', including *glutinous* varieties; there was a general word for *field* (resembling the Sino-Tibetan word), but remarkably no specific term for the wet rice field. *Slash-and-burn* was practiced. *Sickles* were used in harvesting grain, which was kept in *granaries*. Processing included *pounding* and *winnowing*. *Rice steamers* were used in cooking; grain could also be *ground* into *flour* to make *rice cakes*. Other plant domesticates were *buckwheat, taro, beans, peaches, plums, eggplants, cucumbers* and *ramie/hemp*. *Wheat, sorghum* and *Perilla sp.* (a source of oil) reconstruct only at Proto-

Hmongic level. Domesticated animals included *dogs, cattle, sheep/goats, chickens* and *ducks*; there were presumably pigs but the original term cannot be determined as the two branches have different words for the animal. Bee-keeping is indicated by a Proto-Hmong-Mien term for *beeswax* and Proto-Hmongic terms for *honey* and the *bee*. Hunters shot game with the *crossbow*. Terms for the *trap* or *noose* are found in Proto-Hmongic. People lived in *villages* with *houses* covered with *cogon grass* or *tiles*. *Boats* were used. *Fish* and *bamboo shoots* were among the staple complements and, at Proto-Hmongic level, *Houttuynia cordata* was consumed. A taste for *spicy* foods was recognized. *Bowls* and *chopsticks* were known. *Iron* was in use, as well as *copper/bronze* and *gold*. Common artefacts included *cord*, the *bag, pillow, stove* and *bucket*. *Cloth* was *woven*. Objects could be *bought* and *sold* for *money*. Time was reckoned in lunar *months*. *Scales* were used in weighing objects; weight was reckoned in units such as the *catty*. *Writing* was known. The Proto-Hmong-Mien verb for ‘to write’ is a Chinese loanword \*xjaB but there is an intriguing indigenous synonym \*hrueiC in Proto-Hmongic, whose relation to the undeciphered Ba script of eastern Sichuan and Hunan is worth investigating. Many of the cultural terms listed above are borrowed from Chinese. A few: the *soybean, buckwheat* and *pig* are more plausibly related to Tibeto-Burman and may have been borrowed at the same time as the numerals from ‘4’ to ‘9’—unless, like the word for ‘1’, they are inherited and point to an old genetic relationship with Sino-Tibetan. Authors (Haudricourt, van Driem), who believe that the Chinese language spread to north China from areas more to the west regard Hmong-Mien as the language of the original north Chinese populations, from whom the intruding Chinese acquired agriculture and commerce. There is however no strong evidence that Chinese is intrusive in north China;

and there is convergent evidence that much of the borrowing of agricultural and commercial vocabulary went from Chinese into Hmong-Mien (Sagart 1995, 2011a). For instance, *pace* van Driem, the Chinese word for ‘rice steamer’: zèng 甑, Old Chinese \*s-təŋ-s, cannot be borrowed from Hmong-Mien \*tsjɛŋH ‘rice steamer’ because it is an Old Chinese deverbal instrumental with prefix \*s- of a verb ‘to steam’: zhēng 蒸, Old Chinese \*təŋ. Clearly the Hmong-Mien word was borrowed from Chinese after the change \*s-t- > ts-. Truly indigenous Hmong-Mien vocabulary of rice cultivation is very limited. In particular the absence of a term for the wet rice field is noteworthy. An association of the Hmong-Mien family with the domestication of rice in the mid-Yangtze Valley, tepidly suggested by the family’s present-day extension, is not particularly affirmed by the reconstructed vocabulary.

<2> Sino-Tibetan.

The Sino-Tibetan family originates in the late Císhān or early Yǎngsháo cultures, on the eastern edge of the northern Chinese loess plateau, c. 7000 years ago and has spread over most of China, the Tibetan plateau, parts of southeast Asia and the southern edge of the Himalayas (map 6).



Map 6. The Sino-Tibetan languages. Source: The Language Gulper (<http://languagesgulper.com/eng/Tibetomap.html>).

The received view, supported by two recent studies (Zhang et al. 2019 ; Sagart et al. 2019) opposes Sinitic in the east to non-Sinitic (also known as ‘Tibeto-Burman’) in the west. There is a recent reconstruction of Old Chinese by Baxter and Sagart and a pre-reconstruction of Proto-Tibeto-Burman by Matisoff (pre-reconstructions are cited here preceded by the symbol ‘#’). The Sinitic vs. non-Sinitic view of Sino-Tibetan subgrouping is disputed, yet shared non-Sinitic innovations exist, supporting the non-

Sinitic subgroup's validity: loss of /r/ in cognates of 三 \*s.rum '3', 沙 \*s<sup>h</sup>raj 'sand', 彡 s.rom 'mustache' etc.; merger of final glottal stop with -k in Tibeto-Burman cognates of 腦 \*n<sup>h</sup>[u]ʔ 'brain', 武 \*m(r)aʔ 'military' etc. (Sagart 2017).

The Sino-Tibetan homeland was in the northern Chinese area of foxtail millet domestication. Today, foxtail is cultivated by speakers of the main Sino-Tibetan branches: Tani, Sal, Kiranti, Bodic, Karen, Burmo-Qiangic, Tujia and Sinitic, at least. Its names in Old Chinese, Lhokpu (Bhutan) and Trung : 稷 \*[ts]ək, cək and tɛja<sup>255</sup> respectively, can be derived from a Proto-Sino-Tibetan #tsək. This makes *Setaria* a cultivated plant at the earliest Sino-Tibetan level, in full agreement with a Císhān-Yǎngsháo origin. Scholars (Haudricourt, Peiros, Starostin) who think that the family originates in the southern Himalayan region, where the modern Sino-Tibetan center of language diversity is located, have the lack of early archaeological evidence for *Setaria italica* there to explain. Linguistic diversity is not a reliable indicator of the Sino-Tibetan homeland, as all the diversity in northern China has been erased by a succession of Chinese standard languages over the past 3000 years. Other reconstructable agricultural vocabulary includes the name of the swidden: Old Chinese 田 \*l<sup>h</sup>iŋ '(dry) field', Written Tibetan zying < \*lying 'field', Lepcha *lyǎŋ*, Cuona *leŋ*<sup>13</sup>, etc. In the Nungish and Sal branches, the word has shifted its meaning to 'forest', pointing to forest swiddens fields opened by slash-and-burn. *Panicum miliaceum* is archaeologically prominent in Císhān-Yǎngsháo and is still cultivated by the Chinese and some non-Sinitic peoples, but the plant's names in western Sino-Tibetan are severely under-recorded. Rice is absent in Císhān-Yǎngsháo: moreover the two Sino-Tibetan branches have different names for the plant. Rice

cultivation appears to have entered the Sinitic and non-Sinitic branches independently from early rice cultivation areas in Henan—Jiahu, Baligang—south of the Yellow river (Sagart et al. 2019).

Other plants of economic importance included beans (genus and species uncertain):

Tibeto-Burman #nuk, Old Chinese 菘 \*nuŋ; Brassica spp.: 芥 \*kʰr[e][t]-s ‘mustard plant’ (probably \*kʰr[e]p-s), Proto-Kuki-Chin \*kram ‘cabbage’. Unsurprisingly there are no reconstructable terms for the western domesticates wheat and barley, introduced after the breakup of Proto-Sino-Tibetan; buckwheat was probably domesticated by non-Sinitic speakers and then transmitted to the Hmong-Miens and Chinese. Proto-Sino-Tibetan animal domesticates included the *dog* and *pig*: the Tibeto-Burman word #C.pak ‘pig’ corresponds to 富 \*pək-s ‘wealth’. A term for ‘cattle’: 牛 \*[ŋ]wə, Tibeto-Burman #ŋwa, is reconstructable at the earliest Sino-Tibetan level, despite the late introduction of western *Bos taurus* in East Asia: the term appears to have designated morphologically wild cattle managed by humans in early Holocene northeastern China (Zhang et al. 2013). A term for domesticated ovines can be reconstructed: 羊 \*gaŋ, Written Tibetan g.yang. The horse was first introduced into western Sino-Tibetan from the steppes and subsequently transmitted to China: the Chinese name 馬 \*mʰraʔ was borrowed from a Sino-Tibetan language where Proto-Tibeto-Burman #m-raŋ ‘horse’ (STEDT) had evolved to [mrã], explaining the lack of -ŋ in the Chinese term. Contrary to a widespread misconception, its source is not Mongolic *morin* nor is it relatable to English *mare*. Reconstructable terms for the *fishnet* (Sagart 2011b), *weaving* and *arrow* (several kinds) attest to the importance

of fishing, weaving, hunting and warfare. While the Sinitic branch largely evolved *in situ*, Majiayao-culture sites like Haxiu and Yingpanshan in western Sichuan, with *Setaria i.* and *Panicum m.* beginning 3300 BCE (d'Alpoim Guedes 2011) as well as similar and only slightly later sites on the northeastern Tibetan plateau (Chen et al. 2015) probably represent a phase in the non-Sinitic Sino-Tibetan expansion. Some genetic evidence links this expansion to modern Sino-Tibetan peoples (Su et al. 2000).

<1> higher-level connections

The strong likelihood that the first Austronesians reached Taiwan from the continent has given rise to several hypotheses of genetic relatedness between Austronesian and one or the other family of mainland East Asia. Sagart (2005) details the linguistic evidence for a genetic relationship between Sino-Tibetan and Austronesian. Tsang et al. (2017) identify abundant grains of *Panicum miliaceum* in Nan Kuan Li, the earliest neolithic site in Taiwan, alongside previously identified *Setaria italica* and rice: this increases the similarities between early Formosan agriculture and several early northeastern Chinese agriculture, such as Yuezhuang in northern Shandong. Sagart et al. (2018) argue from these similarities that the Austronesians originate in a Shandong population of the 6th millennium BCE, cumulating farming (foxtail, broomcorn, *japonica* rice) and nautical expertise, and practicing ritual tooth ablation; after 5000 BCE, these groups expanded south along the China coast, introducing these traits to Taiwan by 3500-3000 BCE.

Genetic support was provided by Ko et al. 2014 (mtDNA) and by Wei et al. 2017 (Y chromosome). Fuller (2011) supports the Shandong-to-Taiwan expansion model. At an earlier level, the Sino-Tibetan-Austronesian homeland can be traced back to precursors of

Císhān and Peiligang cultures, on the eastern edge of the loess plateau, where millets and pigs were domesticated beginning c. 10.000 BP. While the later Císhān and Yǎngsháo cultures north of the Yellow River have been equated with Proto-Sino-Tibetan (Sagart et al. 2019), Císhān’s southern sister Peiligang culture may be the source of an eastward expansion leading to Shandong and eventually to the first Austronesians in Taiwan. The sudden appearance, shortly after 8000 BP, of sites with foxtail, broomcorn and rice, immediately south of the Yellow River from Henan (Tanghu) to north Shandong (Yuezhuang, Hexi) materializes the continuity between Peiligang culture and the pre-Austronesians in Shandong.

Table 1 presents items of the subsistence vocabulary shared by Proto-Austronesian and Sino-Tibetan.

	Proto-Austronesian	Old Chinese	Tibeto-Burman
<i>Setaria italica</i>	*beCeŋ	稷 *[ts]ək	Lhokpu <i>cək</i> , Trung <i>teja</i> <sup>25</sup>
<i>Panicum miliaceum</i>	*baCaR	糜 *[ts][a][t]-s	—
seed	*RaH <sub>1</sub> ap	粒 *p.rəp	W. Himalayish <i>yeb-mo</i> ‘to sow’
dehusked grain	*beRas	糲 *[r]at-s	PTB #b-ras rice/fruit/bear fruit (STEDT)
polished grain/grain as food	*Semay	米 *(C).m <sup>ʰ</sup> [e]j?	PTB #mey rice/paddy (STEDT)
to toss/sow broadcast	*sabeR~*sabuR	播 *p <sup>ʰ</sup> ar-s	PTB #bwar throw away/cast/sow/toss (STEDT)
to cut with tool/reap	*ritrit	利 *C.ri[t]-s ‘profit’ (what is reaped)	PTB #ri:t reap/cut/sever (STEDT)

pig	*beRek	(富 *pək-s 'wealth')	PTB #C.pak 'pig'
fishnet	*aray	羅 *r <sup>h</sup> aj 'kind of net'	Wr. Tib. rgya < rya 'net'

*Table 1 Sino-Tibetan-Austronesian subsistence vocabulary*

The Old Chinese and Tibeto-Burman forms in Table 1 should be compared with the last syllable of the corresponding Austronesian forms, except when the medial consonant disappears, which happens regularly with -H<sub>1</sub>- ('seed') and with the kind of -R- in 'pig'. The resemblances among the forms in Table 1 are of a non-accidental kind: for the most part they fit the sound correspondences in Sagart (2005).

An alternative view (Blust 1996), places the pre-Austronesians in the lower Yangtze rice neolithic. That location provides a link, *via* the upper Yangtze Valley, to the great rivers which irrigate southeast Asia and to the presumed Austroasiatic homeland. That proposal accounts for no shared vocabulary at all, and leaves Formosan *Setaria italica*, *Panicum miliaceum* and tooth ablation in limbo. Lower Yangtze rice was moreover considerably more advanced on the path to domestication than early Taiwan rice: tiny Formosan rice grains are better compared to early northern Chinese rice (Fuller 2011) and unlike in the lower Yangtze, the first Austronesian farmers did not irrigate or drain their fields. The importance of the rice vocabulary in Proto-Austronesian is often overstated: recent fieldwork in Taiwan shows that \*beRas and \*Semay were general terms for dehusked and cooked grains, not limited to rice (Sagart et al. 2018). Blust (1996) thinks a single domestication of rice, centered on the Yunnan/Burma border area, caused both the Austroasiatic and Austronesian expansions. The two families are treated by him as two branches of an 'Austric' superfamily with a time depth of c. 9000 BP. However the reality

of Austric is doubtful. The limited linguistic evidence published in its support is entirely morphological: that is in part disputed and in part also found in Sino-Tibetan (infixes <r>; pa- causative). There is little shared vocabulary, none of it agricultural.

Archaeological evidence for rice cultivation in Blust's proposed homeland is too recent: hardly older than 4500 BP. Aside from Austroasiatic, the lower Yangtze origin theory of Austronesian also aims at accounting for a set of obviously genetic resemblances—shared basic vocabulary with sound correspondences—between the Tai-Kadai and Austronesian vocabularies (Benedict 1942). Benedict, Blust (1996) and Ostapirat (2005) explain these by supposing an 'Austro-Tai' language family, originating on the south China mainland: the Tai-Kadai and Austronesian languages would be Austro-Tai's two branches. In Blust's version of Tai-Kadai, the lower Yangtze neolithic language would be Proto-Austro-Tai, putative Austric's eastern branch. The difficulties pointed out earlier with the lower Yangtze hypothesis are not improved by the addition of Tai-Kadai: foxtail millet and tooth ablation, absent in the lower Yangtze neolithic, are part of Tai-Kadai culture. While the relationship of Austronesian and Tai-Kadai is clearly genetic, exclusive sharing by Tai-Kadai and Malayo-Polynesian of major innovations in the numerals shows instead that Tai-Kadai, like Malayo-Polynesian, is a branch, rather than a sister group, of Austronesian (Sagart 2004). Proto-Tai-Kadai goes back to the back-migration toward the mainland, in the second millennium BCE, of an Austronesian language, probably out of the Philippines.

Finally Ratliff (2013) presented an argument that the Hmong-Mien and Austronesian families are genetically related. The most striking piece of evidence compares the Proto-

Hmong-Mien pair \*dəjH ‘die’ vs. \*təjH ‘kill’ to the Proto-Austronesian pair \*ma-Cay ‘die, be dead’ vs. \*pa-Cay ‘kill’, where \*pa- and \*ma- are causative (‘kill’ = ‘cause to die’) and stative prefixes respectively. In both languages one has a single root for ‘kill/die’: \*Cay in Proto-Austronesian, \*təjH in Hmong-Mien, with the addition of voicing in the stative member (‘be dead’): in Proto-Austronesian voicing takes the form of a nasal prefix; in Hmong-Mien of a change from \*t to \*d—potentially due to a lost nasal prefix. This example combines a very basic lexical comparison and a shared grammatical feature. Ratliff also cited resemblances in the personal pronouns, and the comparison Proto-Hmong-Mien \*m-nək ‘bird’ vs. Proto-Austronesian \*manuk ‘bird’; one may add the resemblance between Proto-Hmong-Mien \*NKəjX ‘excrement’ and Proto-Austronesian \*Caqi ‘excrement’. The numeral ‘1’ (above) potentially matches Proto-Austronesian \*isa ‘one’—assuming the -a at the end is another morpheme. Finally we seem to have a sound correspondence between Mienic initial \*hm- in \*hmei B ‘dehusked rice’ and \*hmej A ‘animal fat/oil’ and Proto-Austronesian \*S\_m- in \*Semay ‘grain as food’ and \*SimaR ‘grease, oil, fat’. Overall the evidence is slight, as Ratliff recognizes, but varied and of fair quality. Ratliff observes that the material Hmong-Mien shares with Austronesian is also often shared with Tai-Kadai: this is expected if Tai-Kadai is part of Austronesian. Part of this material, which includes agricultural vocabulary, is also shared with Sino-Tibetan (Table 2). This argues that Hmong-Mien might be a sister of Austronesian within the eastern branch of Sino-Tibetan-Austronesian.

Austroasiatic is a more southerly and westerly group, with no shared agricultural vocabulary with any other group: any genetic relationship between Austroasiatic and the

other groups can only be at a very early, pre-agricultural stage.

	Proto-Austronesian	Proto-Hmong-Mien (R.)	Proto-Tai (P.) Proto-Kra (O.)S)	Old Chinese (B-)	Proto-Tibeto-Burman	Proto-Mon-Khmer (Sh.)
die/kill	maCay/paCay	dəjH/təjH	*p.ta:j A (< m.t-)	死 *sij?/—	#siy/—	*kc[ə]t/*gət ~ *gut
1sg	-ku	ku(N) (Ratliff 2013)	*ku: A	—	#ka	—
2pl	-mu	m̥əu	—	—	—	*pi?
one	isa (< is-a ?)	ʔi (< ʔit ?)	*ci C	— ʔi[t]	#it	*muuy ~ *muəy ~ *muup
excrement	Caqi	N-KəjX	*Ç.quj C	屎 *[qʰ]ij?	#qiy	*ʔic ~ *ʔiə[c] ~ *ʔ[ə]c ~
bird	manuk	m-nək	*Ç.nok D	—	—	*ci(i)m ~ *ciəm ~ *caim ~ *cum
grain as food	Semay	hmei B	—	米 (C.)m̥[ɛ]j?	#may	—
fat, grease/oil	SimaR	hmej A	*man A	麻 *C.m̥'raj 'hemp' (source of lamp oil)	#mar	*priiŋ ~ *prieŋ ~ *pru[ə]ŋ

*Table 2: shared elements between Hmong-Mien and Austronesian, and corresponding forms in other groups.*

## References

Adelaar, Karl A. “Towards an integrated theory about the Indonesian migrations to Madagascar.” In *Ancient human migrations: a multidisciplinary approach*, edited by P.

- N. Peregrine, I. Peiros and M. Feldman, pp. 149-172. Salt Lake City: University of Utah Press, 2009.
- d'Alpoim Guedes, Jade. "Millets, Rice, Social Complexity, and the Spread of Agriculture to the Chengdu Plain and Southwest China." *Rice* 4 (2011), 104-113.
- Benedict, Paul K. "Thai, Kadai and Indonesian: a new alignment in Southeastern Asia." *American Anthropologist*, n.s. 44 (1942), 576-601.
- Blust, Robert A. "The Prehistory of the Austronesian-speaking peoples: A View from Language." *Journal of World Prehistory* 9, 4 (1996): 453-510.
- Chang, C.S., Liu, H.L., Moncada, X., Seelenfreund, A., Seelenfreund, D., Chung, K.F. "A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry." *PNAS* online, October 5, 2015.
- Chen F. H., Dong, G. H., Zhang, D. J., Liu, X. Y., Jia, X., An, C. B., Ma, M. M., Xie, Y. W., Barton, L., Ren, X. Y., Zhao, Z. J., Wu, X. H., Jones M. K. "Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P." *Science* 347, 6219 (2015): 248-250.
- Diffloth, Gérard. "The contribution of linguistic paleontology to the homeland of Austroasiatic." In *The peopling of East Asia: Putting together Archaeology, Linguistics and Genetics*, edited by Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, 77-89. London: RoutledgeCurzon, 2005.
- Fuller, Dorian. "Pathways to civilizations: Tracing the origins and spread of rice and rice cultures." *Rice* 4 (2011): 78-92.
- Gray, Russell D., A.J. Drummond, and S. Greenhill. "Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement." *Nature*, 323 (2009): 479-483.
- Hung, Hsiao-Chun. "Migration and Cultural Interaction in Southern Coastal China, Taiwan and the Northern Philippines, 3000BC to AD 100: The Early History of the Austronesian speaking Populations." PhD diss., Australian National University, 2008.
- Ko, Albert M.-S., Chen, C.Y., Fu, Q.M., Delfin, F. Li, M.K., Chiu, H.L., Stoneking, M. and Ko, Y.-C.. "Early Austronesians: Into and Out of Taiwan." *American Journal of Human Genetics*, 94 (2014): 426-436.
- Li, Hui, Huang, Y., Mustavich, L.F., Zhang, F., Tan, J.Z., Wang, L.E., Qian, J., Gao, M.H., Jin, L. "Y chromosomes of prehistoric people along the Yangtze River". *Human Genetics* 122 (2007): 383-388.
- Niederer, Barbara. *Les langues hmong-mjen (miao-yao). Phonologie historique*. München, Newcastle, Lincom Europa, 1998.
- Norman, Jerry, and T.-L.Mei. "The Austroasiatics in ancient south China: some lexical evidence." *Monumenta Serica* Vol. XXXII (1976), 274-301.
- Ostapirat, Weera "Kra-Dai and Austronesian: notes on phonological correspondences and vocabulary distribution." In *The peopling of East Asia: Putting together Archaeology, Linguistics and Genetics*, edited by Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, 109-133. London: RoutledgeCurzon, 2005.
- Peiros, Ilya. and Shnirelman, V. "Rice in Southeast Asia: a regional interdisciplinary approach". in: *Archaeology and Language, archaeological data and Linguistic hypotheses*, edited by Roger Blench and Matthew Spriggs, Vol. II: 379-389. Routledge 1998.
- Pietruszewsky, Michael, A. Lauer, C.H. Tsang, K.T. Li, M. T. Douglas. "Dental Indicators of Health in Early Neolithic and Iron Age Skeletons from Taiwan." *Journal of Austronesian studies* 4 (2014), 2:1-34.
- Ratliff, Martha. *Hmong-Mien Language History*. Australian National University, 2010.
- Ratliff, Martha. "Grammatical features shared by Hmong-Mien and Austronesian". Powerpoint presentation for the 23rd Annual conference of Southeast Asian Linguistics Society, Bangkok, May 21-23, 2013. Online at <http://www.jseals.org/seals23/ratliff2013grammaticalp.pdf>.
- Sagart, Laurent. "Some remarks on the Ancestry of Chinese." In *The Ancestry of the Chinese Language*, edited by William S.-Y. Wang, 195-223. Journal of Chinese

- Linguistics monograph series n. 8, 1995.
- Sagart, Laurent. "The higher phylogeny of Austronesian and the position of Tai-Kadai." *Oceanic Linguistics* 43,2 (2004): 411-444.
- Sagart, Laurent. "Sino-Tibetan-Austronesian: an updated and improved argument." In *The peopling of East Asia: Putting together Archaeology, Linguistics and Genetics*, edited by Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas, 161-176. London: RoutledgeCurzon, 2005.
- Sagart, Laurent "The expansion of *Setaria* farmers in east Asia: a linguistic and archaeological model." In *Past human migrations in East Asia: matching archaeology, linguistics and genetics*, edited by Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross, Ilya Peiros, Marie Lin, 133-157. London: Routledge, 2008.
- Sagart, Laurent. "How many independent rice vocabularies in East Asia ?" *Rice* 4, 3 (2011a): 121-133.
- Sagart, Laurent. "华澳语系发源于何时何地" [The Sino-Tibetan-Austronesian homeland: where and when ?]. *Communication on Contemporary Anthropology* 5 (2011b):143-147/e21.
- Sagart, Laurent. "A candidate for a Tibeto-Burman innovation". *Cahiers de Linguistique Asie Orientale* 46 (2017): 101-119.
- Sagart, Laurent., T.-F. Hsu, C.T. Yuan, C.C. Wu, L.T. Huang, Y.C. Chen, Y.F. Chen, Y.C. Tseng, H.Y. Lin, Y.I. Hsing, "A northern Chinese origin of Austronesian agriculture: new evidence on traditional Formosan cereals". *Rice* 11, 57 (2018): 1-16.
- Sagart, Laurent, G. Jacques, Y.F. Lai, R.J. Ryder, V. Thouzeau, S.J. Greenhill, J.M. List "Dated language phylogenies shed light on the ancestry of Sino-Tibetan". *Proceedings of the National Academy of Sciences* 116, 21 (May 2019): 10317-10322; DOI: 10.1073/pnas.1817972116.
- Shorto, Harry. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics, 2006.
- Sidwell, Paul. "Is Mon-Khmer dead? Long live Austroasiatic!" Paper presented at the 18th Annual Meeting of the Southeast Asian Linguistics Society (2008), Bangi: Universiti Kebangsaan Malaysia.
- Sidwell, Paul, and R. Blench. "The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis." In *Dynamics of Human Diversity*, edited by Nick J. Enfield, 317-345. Pacific Linguistics, 2011.
- Su, B Xiao, C.J., Deka, R., Seielstad, M.T., Kangwanpong, D., Xiao, J.H., Lu, D.R., Underhill, P., Cavalli-Sforza, L., Chakraborty, R., Jin, L. "Y chromosome haplotypes reveal prehistorical migrations to the Himalayas." *Human Genetics* 107 (2000):582–590.
- Tsang Chenrg-hwa, Li, K.T., Hsu, T.F., Tsai, Y.C., Fang, P.H, Hsing, Y.I. "Broomcorn and foxtail millet were cultivated in Taiwan about 5000 years ago". *Bot Stud* 58,1(2017):3
- Wei L.H., Yan S., Teo Y.Y., Huang Y.Z., Wang L.X., Yu G., et al. "Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia". *PLoS ONE* 12, 4 (2017): e0175080.
- Witzel, M. "Early sources for south Asian substrate languages." *Mother Tongue* special issue (October 1999), 1-76.
- Zhang, H.C. et al. "Morphological and genetic evidence for early Holocene cattle management in northeastern China". *Nature Communications* 4.1 (2013).
- Zhang, M., Yan, S., Pan, W. et al. "Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic". *Nature* 569 (2019), 112–115, doi:10.1038/s41586-019-1153-z

web site:

STEDT [The Sino-Tibetan Etymological Dictionary and Thesaurus]. Director: Matisoff, James A. Online at <http://http://stedt.berkeley.edu/>

**Abstract.** This paper briefly describes the state of the art of linguistic research on the main language families represented in southeast Asia: Austroasiatic, Austronesian, Tai-Kadai, Hmong-Mien and Sino-Tibetan. It reviews the vocabulary of agriculture, and more generally of subsistence, that can be reconstructed to each family's proto-language. It attempts to clarify the dates and locations of the respective proto-languages and to throw light on the circumstances surrounding the population expansions that gave rise to them; it also discusses issues of long-distance relationships among them.

**Keywords:** east Asia, language families, agriculture, reconstruction, cultural vocabulary, genetic relationships, population expansions.