THE VOCABULARY OF CEREAL CULTIVATION AND THE PHYLOGENY OF EAST ASIAN LANGUAGES

Laurent Sagart

CNRS, CRLAO, 54 Bd Raspail, 75270 Paris cedex 06, France

ABSTRACT
This study investigates some important terms of cereal agriculture in the five language families of East Asia, in an attempt to gain some insights into processes of cereal domestication, demographic expansions and the formation of language families. The principal findings are:

There is some evidence that the proto-Austronesians cultivated wet rice, albeit without elaborate irrigation techniques;

Two words for rice are uniquely shared by Sino-Tibetan and Austronesian, with sound correspondences;

A word for ‘Setaria italica’, a millet, is possibly shared by Chinese and Austronesian.

There is no evidence for agricultural words uniquely shared by Austroasiatic and Austronesian (Austric).

1. INTRODUCTION
In recent years, the formation and diversification of language families has been linked to population expansions occasioned by the first domestication of cereals (Bellwood 1984-85). In East Asia, where at least two cereals – rice and Setaria italica, a millet –, and possibly a third – Panicum miliaceum, another millet – were domesticated in the period 10,000-8000 BP, an examination of the vocabulary of cereal cultivation may yield clues to the population history of the entire region. In this paper I build on earlier scholarship (Haudricourt 1970; Zide and Zide 1976; Revel 1988; Ferlus 1996; Bradley 1997; Pejros and Shirelman 1998; Vovin 1998; Blench forthcoming) to examine some East Asian cereal-related terms (names for rice, millets, grain, fields etc.), in an attempt to draw some inferences on the history of East Asian populations and languages.

The higher-order phylogeny of East Asian languages is controversial. There are five basically uncontroversial building blocks:

Austroasiatic (Eastern: Khmer, Mon, Vietnamese, Nicobar, Aslian, Khasi; Western: Munda etc.). Proto-language spoken maybe 6000 or 7000 BP in south-western China;

Austronesian (Atayal, Paiwan, Puyuma, Bunun, Amis, Rukai, Tagalog, Malay, Malagasy, Maori etc.); proto-language spoken c.5500-4500 BP in Taiwan;

Sino-Tibetan (Chinese, Tibetan, Burmese, Lushai, Kachin, Bodo, Garo etc.); proto-language spoken maybe 6000-7000 BP in the mid- and upper Huang He Valley.

Hmong-Mien (a.k.a. Miao-Yao: Hmong, Ho Nte, Bunu, Mien etc.); proto-language spoken c.2500 BP in the mid Yangzi Valley;

Tai-Kadai (Tai, Li, Kam, Sui, Gelao etc.); proto-language spoken c.2500 BP in southeastern China; note that Tai-Kadai is much more compact and recent than previously believed, cf. Ostapirat 2000 and pers. comm. 2001.

There have been suggestions that the five taxa listed above are branches of larger genetic entities. The main proposals are as follows.

The Austroasiatic theory (Schmidt 1906; Reid 1994; Blust 1996) states that Austroasiatic and Austronesian share an exclusive common ancestor. Blust (1996:130sq) links rice domestication with the Austric expansion, and claims that the hypothetical Austric homeland was in north-western Yunnan, c.9000 BP. Higham (1996) supports this, but appears to favor a homeland further downstream in the Yangzi Valley: from Pengtoushan to Hemudu, which agrees better with the archaeological data on early domesticated rice. Under Blust’s and Higham’s versions of Austric, one expects to find some elements of the vocabulary of rice exclusively shared by Austroasiatic and Austronesian.
The **Austro-Tai** theory (Benedict 1942) states that Austroasian and Tai-Kadai share an exclusive common ancestor, spoken in south-Eastern China by rice-growing populations. Pejros and Shnirolman (1998:380) say the disintegration of Austro-Tai had begun by the sixth millennium BCE. According to Benedict, one group moved out to the Pacific, there to form the Austronesian family. Tai-Kadai is the name used by Benedict for those Austro-Tais who, in his theory, remained on the mainland. He claimed that Tai-Kadai gave out many loanwords to Chinese (Benedict 1975). Benedict later expanded his Austro-Tai to include Hmong-Mien (1975) and Japanese (1990). Under the Austro-Tai theory, one should find some elements of the vocabulary of rice exclusively shared by Tai-Kadai and Austronesian.

The **Sino-Tibetan-Austronesian** theory (Sagart 2001, 2002 extending the results of Sagart 1993) states that Austroasian and Sino-Tibetan share an exclusive common ancestor. The proto-language is seen as having been spoken in the mid- and lower Huang He Valley c.8500 - 7500 BP. The first split occurred between a western and an eastern group. The eastern group, in the lower Huang He and Huai River Valleys, expanded southward along the coast, eventually reaching Taiwan by c.5500 BP where their language began to diversify into the Austronesian language family. Stay-at-homes in northeastern China were later submerged by Chinese expansions. The western group developed into Sino-Tibetan, with Chinese remaining in situ and Tibeto-Burman expanding in a south-westerly direction. Under this theory, Tai-Kadai is an aberrant part of Austronesian: a daughter language of proto-Austronesian, sharing some innovations with Malayo-Polynesian, and having moved from Taiwan to coastal south China, where it underwent extensive relexification by a mainland Southeast Asian language (Sagart forthcoming [a]). The Sino-Tibetan-Austronesian expansion is linked to the domestication of the millets by a population who also cultivated rice as a second cereal. An early site with both rice and Setaria is Longquzhuan in Jiangsu (Wang and Zhang 1998), c.7000-5500 BP. Under that theory, one should find some words relating to rice and millet exclusively shared by Sino-Tibetan and Austronesian (including Tai-Kadai).

A little-publicized but serious theory says that Austroasian and Hmong-Mien have a common ancestor (Davies 1909; Haudricourt 1954, 1966, Pejros and Shnirolman 1998:155sq). The homeland of this hypothetical phylum would have been in the mid-Yangzi Valley, precisely where rice appears to have been domesticated c.9000 BP, with Hmong-Mien and Austroasiatic later diversifying as an Eastern versus a Western group. I will use Starosta’s term ‘Yangzian’ for this construct (Starosta, forthcoming).

A larger construct, defended in e.g., Ruhlen (1991) and Peiros (1998), consolidates Schmidt’s Austric and Benedict’s Austro-Tai under the same name ‘Austric’. Pejros and Shnirolman (1998) date the disintegration of this macrofamily to the ninth to eighth millennium BCE: I will refer to this hypothesis as ‘Greater Austric’, to distinguish it from the classical Austric theory. Pejros and Shnirolman (1998) do not see the Macro-Austric speakers necessarily as rice farmers. Indeed they claim that the proto-Austronesian vocabulary of rice was borrowed from Sino-Tibetan.

Then come proposals which aim at unifying all of the five language families of East Asia: Schiller’s **Macro-Austric** (Schiller 1987) and Zhengzhang’s **Pan-Sino-Austronesian** (Zhengzhang 1993) consolidate Sino-Tibetan, Austro-Tai and Austroasiatic into a macrophylum without an explicit subgrouping. Starosta’s **proto-East Asiatic** (Starosta forthcoming) is a conjecture consolidating Sino-Tibetan and Yangzian, and Sino-Tibetan-Yangzian further with Austroasian. My own very tentative conjecture is of an East-Asian superfamily having Sino-Tibetan-Austronesian as its northern branch and Yangzian as its southern branch. Under that tentative view, all five East Asiatic taxa would have diversified out of the language of the first domesticators of rice, spoken around 9000 BP in the mid-Yangzi Valley. Millet was added to rice by speakers of the northern languages the northern branch.

2. THE WORDS

The terms to be investigated here are the names of the principal domesticates: rice and the millets (Setaria italica and Panicum miliaceum); the names of the unhusked, husked and cooked grains of rice, and the terms for fields: the swidden and wet ricefield.

2.1 Rice plant, paddy, unhusked grain

Few languages distinguish between the terms for growing rice plant and the unhusked grain, the same term often also functioning as a general term for “rice”.

2.1.1 Sino-Tibetan

The least unlikely candidate for the meaning ‘paddy’ in Sino-Tibetan is a form something like *mey*, including Benedict’s (1972:149) *mey* or *may* ‘paddy’ for Bodo-Garo (a SW group within Sino-Tibetan), plus Gyarong smaj, presumably a general term for grain (in smaj khri ‘millet’), Karen me < may ‘boiled rice’, Ch. miš * *mij ‘husked grain’ (general term). This Chinese word has unexpected high-tone reflexes in some modern dialects (O’Connor 1976 for Hakka), which indicates a lost voiceless consonant before initial m- (thus really *C_voiceless-mij*). Voiceless hm- in the Chinese loan to Mien *hmọja ‘husked rice’ (see below) also points to a lost voiceless element before m-.
2.1.2 Hmong-Mien

2.1.3 Austroasiatic
The proto-Austroasiatic word for ‘paddy’ was certainly *pajay (Blust 1976), with widespread reflexes both inside and outside of Formosa. Proposals to connect this term with proto-Hmong-Mien *mblauŋ 'id.' (Pejros and Sniegelman 1998:383), or with proto-Tai (Li 1977) *klaː, 'rice seedlings' (Benedict 1975:364) are extremely unlikely. No convincing comparisons for PAN *pajay are known so far outside of Austroasiatic.

2.1.4 Austroasiatic
There are two proto-Austroasiatic candidates for ‘paddy’. One is a set including North Munda *babə, Khasi kba, Bahmaric and Danaw ba~ka and Aslian baa?, all ‘paddy’ (Zide and Zide 1976; Ferlus 1988:87). Another includes Sora soro and Old Mon soro?, both ‘paddy’. Ferlus (1996) argued that the second set originally referred to taro (Riang soro?, Khmu soro? etc.) and later applied to paddy, because of the similarity in cultivation techniques of the two plants (wet field).

2.1.5 Tai-Kadai
There is no specific word for ‘rice plant, paddy’ in the Tai languages proper and in Kam-Sui (Haudricourt 1970). Words for the plant are general terms which also refer to unhusked, husked and cooked rice, with the requisite modifiers. These forms (*khaŋ,*gawə or the like) are moreover generally recognized as borrowings from the Austroasiatic term for ‘husked rice’, compare for instance Vietnamese gau < rakaw (Haudricourt 1970), Lévy 1988:64, Mahdi 1994, Blench forthcoming). The situation is more complex outside of Kam-Tai. For proto-Kra, Ostapirat (2000:227) reconstructs a term for ‘paddy’ (=unhusked grain) *caŋ, and a general term for rice *khaŋ. Yet no single item qualifies for proto-Tai-Kadai ‘rice plant/unhusked grain’.

2.2 Husked rice
2.2.1 Sino-Tibetan
The main Sino-Tibetan cognate set for this meaning is probably represented by Chinese li 筷 *mo-rat-s 'coarse (unpolished) husked rice' and Written Tibetan རbras (orthographic 'bras') 'rice' (general term), earlier *m-ras, with parasitic /b/ arising between /m/ and /r/ (this idea will be further developed elsewhere), and Lushai ra? 'fruit, to bear fruit'. In view of the Lushai word, Benedict (1972:17; 1975:104) reconstructed a proto-Tibeto-Burman root *ras, something like 'fruit/bear fruit'. The Sino-Tibetan meaning appears to have been 'fruit' in a broad sense, including 'kernel, grain'. If so, proto-Sino-Tibetan had no specific term for 'husked rice' and referred to husked rice as 'fruit/kernel of the rice-plant'.

2.2.2 Hmong-Mien
Wang and Mao (1995) reconstructed three distinct proto-Hmong-Mien forms for the meaning ‘husked rice’: *hm̄ †tʒ occurs only in Menic, and is most likely a loan from Chinese miŋ (*miŋ) 'husked grain', as mentioned earlier. A second form: *tshaŋ, is found only in Hmongic. Only *tšətə, is found in both Menic and Hmongic, but no external comparison is available.

2.2.3 Austroasiatic
The proto-Austroasiatic word for ‘husked rice’ was *beRas, with many reflexes both inside and outside of Taiwan, and with the same husked-rice ~ fruit/kernel polysemy as in Sino-Tibetan. In Taiwan, the Paiwan reflex of *beRas is vɔt 'seed, kernel, grain', and in northern Bunun, a language of East Formosa (Duris 1987) the word for ‘fruit’ is las, which reflects proto-Austroasiatic *Ras, identical with the second syllable in *beRas ‘husked rice’. Reid (1994:374) describes similar semantics for the reflex of *beRas in the Cordilleran languages of the Philippines.

2.2.4 Austroasiatic
Zide and Zide (1976), citing Pinnow (1959), observe that Western and Eastern Austroasiatic have reflexes of an etymon which “is presumably a biformic form composed of *raj and *kug”’, for instance Khmu rako?.. The view that a form something like *rajko is proto-Austroasiatic for ‘husked rice’ has generally been accepted by later scholarship (Ferlus 1988, 1996; Mahdi 1994).

2.2.5 Tai-Kadai
Again, Tai-Kadai has no specific form for ‘husked rice’. The Kam-Tai languages (a branch of Tai-Kadai) use the generic term plus a modifier sān (earlier saak, as indicated by conservative Sayek). For proto-Kra, Ostapirat reconstructs *sal, identical with the Kam-Tai modifier. Michel Ferlus (pers. comm. August 2002) argues that this word has its origin in an Austroasiatic verb meaning 'to peel, remove the skin'; Khmu hae:l < s- 'to peel', supported by Kisingmal sal 'strips of bamboo bark used for tying' and Mon s-n-aaa 'mat' (made of bamboo strips).

2.3 Cooked rice
2.3.1 Sino-Tibetan
The most likely candidate for proto-Sino-Tibetan ‘cooked rice’ includes proto-Tamang *kən, *kan ‘cooked rice’ (Mazaudon 1993-1994) and Chinese gan, 畲 'thick gruel of rice'. This word is probably to be analyzed as *ka-n, from Sino-Tibetan *ka ‘to eat’ (Tuja ka ‘to eat’, proto-Tamang *ka: ‘to eat’, Bogaer Luoba ka: ‘to eat’, Chinese ‘ka-s ‘rice gruel’ plus nominalizing -n suffix).
2.3.2 Hmong-Mien
The proto-Hmong-Mien term for 'cooked rice' is *hnaapc (Wang and Mao 1995), usually regarded as indigenous (e.g., Haudricourt and Strecker 1991), however a connection with the well-attested Chinese word xiàng 餐 *s-hnaap-s 'bring food to people, especially labourers in the field, soldiers etc.' appears extremely probable. It is primarily a verb of giving, cognate with Tibetan g-nang 'give, grant, concede'; in Chinese it also occurs as a noun 'provisions of food'. The Hmong-Mien term is thus very likely a loan from Chinese. The tones conform to the habitual loan correspondence pattern (Chinese qusheng: Hmong-Mien tone C).

2.3.3 Proto-Austronesian
Proto-Austronesian *Semay or *Sumay 'cooked rice' is widely attested both in Taiwan and outside of Taiwan, with irregularities in the reflexes of the initial consonant. Also of interest is the term kaan 'cooked rice' in Taokas, an extinct language of the Formosan west coast (cited by Ferrell 1969:219). This word appears somehow related to the Proto-Austronesian word *ka- in 'to eat', with root ka 'eat' and suffixed -en 'goal focus', an analysis originally proposed by Dempwolff (1938). While it is true that in modern languages reflexes of *ka-en require an additional -en suffix to be converted to goal focus forms, the verb+goal focus suffix analysis for *ka-en is indicated by the fact that proto-Austronesian phonotactics do not allow sequences of two nonhigh vowels in the same morpheme. It would appear that the -en suffix was incorporated into the root shortly after PAN times. Since, moreover, a proto-Austronesian *ka-en would give Taokas kaen or kaam, it is likely that Taokas kaan 'cooked rice' is analyzable as *ka'eat' plus -an, a nominalizing suffix. This form resembles the Sino-Tibetan set discussed in section 2.3.1 above.

2.3.4 Austroasiatic

2.3.5 Tai-Kadai
Again, no specific term for 'cooked rice' exists in Kam-Tai, but Ostapirat (2000) reconstructs proto-Kra *mla(w)C, phonetically similar to forms meaning 'evening meal' in Tai and Kam-Sui (Lakkia blau, Maonan mbja:uo, Longzhou pja:uc). The semantic evolution 'cooked rice' > 'meal' is common. The correspondence between the Kam-Tai and Kra forms, however, is not wholly regular (the tones, in particular, do not fit Ostapirat's sound correspondences).

2.4 Setaria italica
*Setaria italica*, one of the millets, was domesticated by 8500 BP in the mid-Huang He region (Lu forthcoming). The history of terms for millets is difficult to reconstruct because investigators often do not distinguish between *Setaria* and *Panicum*, or confuse them.

2.4.1 Sino-Tibetan
The Old Chinese term for *Setaria* was ji 成 *tsik (Sagart 1999). There are no known cognates for this form in the rest of Sino-Tibetan.

2.4.2 Hmong-Mien
I know of no specific etymology for *Setaria* in Hmong-Mien languages.

2.4.3 Austronesian
Proto-Austronesian *beCen 'Setaria italica' has reflexes in Rukai-Tsouic, a primary branch of proto-Austronesian in Formosa, and in Malay-Polynesian languages.

2.4.4 Austroasiatic
There have been few reports of *Setaria*, and of millets in general, being cultivated by Eastern Austroasiatic peoples, although this may in part be due to gaps in our information on the lexicon of these languages. For Western Austroasiatic, Zide and Zide (1976) reconstructed proto-Munda *(h)oxy 'Setaria' (Sora bur-oy, Remo wi-dar, Gta? ü hwe, Mundari oe), where 'X' is "a vowel feature which must be reconstructed for proto-Munda" (Zide and Zide 1976:1330 note 4).

2.5 Panicum miliaceum and other unidentified millets
2.5.1 Sino-Tibetan
One Sino-Tibetan term (Luce 1981:36; Peiros and Starostin 1996) is reconstructible on the ground of Chinese ji 成 *tsap-s > Middle Chinese tsjeβ 'Panicum miliaceum', non glutinous' and Written Burmese hsap < tshap, an indeterminate kind of millet (authors vary: Judson: "Panicum"; Bernot: "Setaria italica"; Luce 1981:36 'Panicum millet (Setaria)', etc.). We may reconstruct *tsap 'kind of millet' at proto-Sino-Tibetan or some intermediate level.

Another Sino-Tibetan cognate set for a millet includes Written Tibetan khr 'millet', Dzongkha khe 'Panicum miliaceum', Jingpo ?a kji 'millet', perhaps also Chinese qis 成 *kh(r)i 'a cereal with white stem, much appreciated' (Shi Jing), said by Karlgen (1964) to be a kind of millet. In some languages, the term has changed its meaning to 'rice': Gyarong khti 'cooked rice' (G. Jacques, pers. comm. 2000), Lushai taj 'a species of early rice', considered to be cognate by Peiros and Starostin (1996). The Dzongkha term, reliably determined to be *Panicum* m. by Driem (2001:385), suggests the original referent may have been *Panicum*, but this will have to be confirmed from other languages. In Benedict's system, the cognate set seems best represented by a reconstruction Tibeto-Burman *krey. The vowel correspondence with Old Chinese *kh(r)i* is uncertain, however.
2.5.2 Hmong-Mien
Wang and Mao (1995) reconstructed *taiI ‘millet’. The form occurs only in Mien. It is possibly to be compared with the Chinese word for *Panicum* millet *tsap-s*. The Hmong-Mien form shows loss of final -s and palatalization of the initial, close to the Middle Chinese form of the word: tsjejH. Tone A in the Mienic word is unexpected, however (tone C is expected). An alternative comparison is with one of the Austronesian words discussed in 2.5.3 below, either *baCanR *Panicum miliaceum* or *baCaJ *sp. of millet*.

2.5.3 Austronesian
Tsuchida (1977:90) reconstructed “proto-Formosan” *baCaR *Panicum miliaceum* based on Formosan evidence. In north-western Formosan some languages reflect proto-Austronesian *baSaR, with medial -s-, as reconstructed by Li (1992:274) based on Atayal basag, Sediq basaw, Saisiats basal ‘small grain millet’, Taokas basu ‘millet’. Extra-Formosan forms often show divergent final consonants, however: Blust (Austronesian Comparative Dictionary) reconstructs proto-Austronesian *baCaJ ‘sp. of millet’, with final -j accounting for some Formosan and Extra-Formosan forms (but not for all). In the same work, Tsuchida reconstructed a ‘proto-south-Formosan’ *hl-(al)umay ‘Panicum crus-galli’, with reflexes in Kanakanavu, Saroqa, Tsou, Rukai and Paiwan. This set of languages indicates a PAN form by my subgrouping assumptions, where Malayo-Polynesian is part of a Formosan East Coast Linkage (Sagart 2001, forthcoming [a]).

2.5.4 Austroasiatic
Zide and Zide (1976) argued that proto-Munda (Western Austroasiatic) possessed a term for *Panicum miliaceum* which in one of the Munda branches reconstructs to *takrig*. This has no known cognates in Eastern Austroasiatic. In a short description of Khasi, Daladier (in press) assembles a larger cognate set for a term for a millet, perhaps *Panicum miliaceum*, with reflexes in both Western and Eastern Austroasiatic: Munda, Sora kro ‘large millet’ (=either sorghum, bajra or maize, cf. Zide and Zide 1976:1312), Korku kya, Mundari kai ‘kind of millet’, Khasi krai ‘millet’ (Panicum), Amwai kaj ‘millet’ (Panicum), Black and White Riang (Palaungic) kaff (a larger set, with full reconstructions, will appear in Daladier, in press). Note however that as far as Eastern Austroasiatic is concerned, the term is represented only in the westernmost and northernmost branches Khasian and Palaungic, being completely absent in the core Eastern languages. Khasian and Palaungic moreover have a history of contacts with Tibeto-Burman. A borrowing from Tibeto-Burman *krey ‘k.o. millet’ (2.5.1 above) is a possibility.

2.5.5 Tai-Kadai
Liang and Zhang (1996:133) reconstructed *plwuaac ‘millet’. Reflexes include Tai faj, Dong pjaac, Li (Hainan) faj. It is not clear from the evidence at hand what species of millet this set referred to. It bears a certain resemblance to PAN *beCeng ‘millet’. Phonetic evolution through bCeng > pCeng > pleng > *plwuaac is conceivable, and there are even partial parallels in Lakkia, a Tai language, for the evolution to pl-: ‘eye’ PAN *maCa > mCa > pCa > Lakkia pla; ‘die’ PAN *maCay > mCay > pCay > Lakkia plei. Evolution to initial f- in Sianese and Li is a problem, however.

2.6 Swidden
2.6.1 Sino-Tibetan
Haudricourt and Streeker (1991) observed that, while Hmong-Mien has two terms for the swidden and the wet rice field, Chinese has only one: *tian ‘field’ and *ting. This word is cognate with Written Tibetan zing ‘field’ and Lepcha lyang ‘field’. We may reconstruct Sino-Tibetan *tian ‘wet or dry field’.

2.6.2 Hmong-Mien
Chang Kun and Herbert Purnell assembled a Hmong-Mien cognate set for ‘dry field, swidden’, discussed by Haudricourt and Streeker (1991:338). The relevant forms include a Hmongic set reconstructing to proto-Hmong *tlen or *tla, plus forms in Mienic. In terms of Wang and Mao’s reconstruction of proto-Hmong-Mien, the entire cognate set reconstructs to proto-Hmong-Mien *tlaac.

2.6.3. Austroasiatic
The proto-Austronesian etymology for ‘dry field’ is *qmaH, a form without any proposed outside comparisons.

2.6.4 Austroasiatic
See ‘wet rice field’ below.

2.6.5 Tai-Kadai
Michel Feuris (pers. comm., Aug. 2002) reconstructs proto-Tai-Yai (L Fangkuai’s proto-Tai, more or less) *tlaa ‘swidden’, which plausibly regards as an Austroasiatic loanword. Ostapirat (2000:229) reconstructs proto-Kra *tac ‘dry field’. It is unclear whether the two forms are identical.

2.7 Wet rice field
2.7.1 Hmong-Mien
Wang and Mao (1995) provide the proto-Hmong-Mien reconstruction *jiaa, distinct from *jaac ‘dry field’ discussed above.

2.7.2 Austronesian
No proto-Austronesian word for ‘wet rice field’ has hitherto been reconstructed. Whether the Austronesians in Formosan ever cultivated wet rice before Chinese contact is a matter of discussion. At least two Formosan languages use a Chinese loanword for ‘wet rice field’: Paiwan can,
Puyuma can, from Minnan Chinese tsan, 'wet rice field'. The early sources on Taiwan do not mention wet rice fields. Chen Chi-lu (1968:322) argues that pre-contact Formosan rice cultivation was entirely of the dry field type. There is, however, some linguistic evidence for early wet rice cultivation by Formosan Austronesians. The conservative Tsouic languages in south central Taiwan have a verb for 'to transplant rice seedlings', reconstructed by Tsuchida (1976:157) as *c:ətđicu, with reflexes in Kanakanabu, Saaroa and Tsou. Transplantation of rice seedlings is typical of wet rice cultivation. Proto-Tsouic very probably predates Chinese contact. In the present work (note 84 p. 197) Tsuchida cites Bunun ma-danuh 'to transplant rice seedlings' which bears a non-accidental relation to the Tsouic word (the correspondences for the first four phonemes are regular, and, at least in the case of the consonants, nontrivial). However, Bunun final -h (<-q) and Tsou /c/ do not correspond. Tsuchida also cites possibly related forms in Philippine languages: Bikol and Samar-Leite dalugi 'rice seedling', although the final syllable -gi is divergent. Disregarding final consonants and vowels, the corresponding proto-Austronesian reconstruction covering the Formosan (Tsouic and Bunun) and Philippine forms just cited would be *daNu-. Even with the phonological irregularities, it is difficult to imagine that the resemblance between the Formosan and Philippine forms can be entirely accidental, and it does not seem likely that it can be explained in terms of recent contact either.

Further evidence for the antiquity of Austronesian wet rice cultivation comes from Formosan words for 'wet rice field': Atayal (as in Li 1981:387) slate (Squiliq), sala? (C?uli?), claqa (Mayrinax), Sediq celaq; Tsou cxana. The Atayal forms mean 'mud' as well as 'wet rice field'. As Benedict (1995) observed, Atayal implies PAN *CaNaq, from which Malayo-Polynesian *tanäq 'earth, land' also derives regularly. I would suggest for proto-Austronesian *CaNaq the meaning 'mud, muddy ground', with extension to 'wet rice field' in Atayalic, and to 'earth, land' in Malayo-Polynesian. Consider now the Tsou word: cxana: 'wet rice field', the first syllable of which: /cxa/ regularly reflects the same proto-Austronesian *CaNaq 'mud, muddy ground'. The Tsou word should therefore be analyzed as cxa-na < *CaNaq-na. The second formative -na is recognizable in Paiwan pana 'river' (includes dry river bed and low land along river: Ferrell 1982), Kavalan zana 'field' (wet and Western Malayo-Polynesian (Dempwolff) *bena 'low-lying country' (Javanese 'flooded'; Malay 'tide bore of a river'). These point to a proto-Austronesian root *na 'low-lying/easily flooded ground'. All this suggests that the proto-Austronesians cultivated dry and wet rice, the latter planted in natural marshes or swamps, or in low, annually flooding river beds, with transplantation of rice seedlings, but without elaborate irrigation techniques.

2.7.3 Austroasiatic

Zide and Zide (1976:1308) assembled a Munda cognate set for the meaning 'wet, low (terraced) paddy field': Gorsum li(y)ø,q, Gta? lajo, Remo leuə, Mundari loeq, etc. To this we can probably add an Eastern Austroasiatic word: proto-Katuic liaq 'field' from Peiros (1996), which makes this etymology proto-Austroasiatic. It is unclear whether Peiros's reconstruction applies to wet fields or swiddens, however. There are other terms for 'field', apparently limited to Eastern Austroasiatic in the current state of our knowledge: M. Flerus (pers. comm., August 2002) gives Khmu he? < sr- Khmer sreə and various Bahnaric forms for 'wet rice field' like srɛ, sray. The term, however, means 'swidden' in northern Austroasiatic languages like Khmu he? < sr-: indeed, proto-Tai *rajə 'swidden' seems borrowed from such a language. Symmetrically some Eastern Austroasiatic languages: De'ang, Blang, Khmu have na 'wet rice field' (Yan and Zhou 1995:570). These are languages in contact with Tai-Kadai and the word was probably borrowed from Tai-Kadai (see next paragraph).

2.7.4 Tai-Kadai

The most general etymon, reflected in Tai proper, in Li of Hainan and in Kra, is *na* (Liang and Zhang 1996:321 for Kam-Tai, Ostapiat 2000:229 for proto-Kra). The northernmost languages: Kam-Sui and Lakkha, show another set: southern Dong ja, northern Dong ja, Shui *ya*, Then ra, Mak ja, etc. Liang and Zhang (1996:353) reconstruct initial *tr- (though the northern Dong tone argues for *hr- but cannot give a reconstruction for the vowel, phonetically [a] almost everywhere. It would seem the second set displaced na, in Kam-Sui and Lakkha. It is either a Kam-Sui innovation, or, more likely (in view of the irregular vowel reflexes), a form borrowed into several already divergent varieties of Kam-Sui from an unknown source (perhaps an Austroasiatic language where the word for 'wet rice field' was something like sre). Our discussion of cereal-related terms is summarised in Table 1.

3. DISCUSSION

Some of the matches we have seen are explainable in terms of contact between languages. The Hmong-Mien vocabulary is known to be inundated with Chinese loanwords. We have found that three Hmong-Mien items are probable Chinese loanwords: Mienic *hmlo Internet ‘husked rice’, most likely a loan from Chinese mi 米 *C_viem-mi ‘husked grain’; Mienic *tjai Internet ‘millet’, probably to be compared with 稲 Middle Chinese tsje’hi; and proto-Hmong-Mien *hnpp Internet ‘cooked rice’, in all likelihood from Chinese xiang 香 *s-nang ‘s food, as given to labourers in the field or soldiers’. This does not indicate that rice cultivation was acquired by the Hmong-Mien at the contact of the Chinese: rather, it indicates intense cultural pressure from Chinese. The term for 'cooked rice' apparently began its career as the name of a meal in a
Table 1: Some cereal-related terms in East Asian language families

<table>
<thead>
<tr>
<th></th>
<th>Sino-Tib.</th>
<th>Hmong-Mien</th>
<th>Austro-Asiang</th>
<th>Austro-Asiang (general)</th>
<th>Tai-Kadai</th>
</tr>
</thead>
<tbody>
<tr>
<td>rice plant,</td>
<td>may–mey</td>
<td>mbaru</td>
<td>pajar</td>
<td>6a? (general)</td>
<td></td>
</tr>
<tr>
<td>unhusked grain</td>
<td>(?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>husked rice</td>
<td>m-ras</td>
<td>ntsaiC</td>
<td>beras</td>
<td>räŋko? (general)</td>
<td></td>
</tr>
<tr>
<td>cooked rice</td>
<td>ka-n</td>
<td>hnaŋC</td>
<td>Semay</td>
<td>?uup (Eastern)</td>
<td>khawC–gauC</td>
</tr>
<tr>
<td>Setaria</td>
<td>Ch. hsiāk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panicum</td>
<td>tsap</td>
<td>t'aiA</td>
<td>beCer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>wet ricefield</td>
<td>ling</td>
<td>liŋA</td>
<td>liŋ(ŋ) etc. (W.)</td>
<td>liang (Eastern)</td>
<td>naA (Kra, Kan-Tai)</td>
</tr>
<tr>
<td>swidden</td>
<td>liŋaC</td>
<td></td>
<td>qunaH</td>
<td></td>
<td>raja zan (p-Kra)</td>
</tr>
</tbody>
</table>

Conscripted labour or army context; that for 'husked grain', an item of trade, may have been borrowed in a commercial context, displacing the indigenous term ntsaiC in some Mienic languages.

Proposed Austro-Asiang loans to Tai-Kadai in the vocabulary of rice cultivation include a general word for 'rice' (2.1.5) and a word for 'swidden' (2.6.5). Haudricourt (1970) argued that Tai evolved on an Austro-Asiang substratum and considered the paucity of indigenous Tai words for rice to be a case of loss of complexity due to language shift. Lévy (1988:64) suggested that the Tai relies on Austro-Asiang manpower and for that reason adopted some Austro-Asiang vocabulary terms. Brench (forthcoming) argues that "proto-Daic speakers were not originally rice cultivators (...) they borrowed rice from their Austro-Asiang neighbours during an early period of expansion". However, we will see that at least the Tai-Kadai word for the wet rice field: *naA, stands a good chance of being inherited. Moreover, there are also instances of probable borrowings in the opposite direction at a later date, cf. 'rice seedling' (2.1.3 above), probably also 'wet rice field' (2.7.3). It is interesting that these suspected Tai-Kadai loans to Austro-Asiang relate to wet rice cultivation, while at least one Austro-Asiang loan to Tai-Kadai ('swidden') relates to dry rice cultivation.

Benedict (1995) argued that speakers of his proto-Austro-Tai possessed rice, which they called *(m)beyaw, and a men's language in which this term was given a meaningless suffix -ts, resulting in beya-ts, with loss of -w. The alleged suffixed form according to Benedict was maintained only in Austro-Asiang, where *beyas 'husked rice' (his *beya-ts) is widespread. In Austro-Asiang, alleged reflexes of his *(m)beyaw are also found, but only in Tsouic languages, in the meaning 'cooked rice'. Benedict's interpretation relies crucially on the assumption of a meaningless suffix, which can for this reason be used without any constraints. It must be considered doubtful. It is possible that the Tai-Kadai forms listed by Benedict, such as proto-Tai *(b)rawA 'evening meal' (see also 2.3.5) are related with the Tsouic forms for 'cooked rice' (Rukai *bro etc.), but this relationship, if real, is independent of PAN *beras 'husked rice'.

Benedict (1975:94) further reconstructed an Austro-Tai word for 'wet rice field' based on Tai-Kadai *na and proto-Austro-Asiang *na 'low-lying/easily flooded ground'. This comparison is phonetically reasonable, but if the proto-Austro-Tai meaning was 'wet rice field', it is not clear why this meaning survives only in Tai-Kadai, and why proto-Austro-Asiang apparently had no specialized word meaning 'wet rice field'. It is preferable to suppose that the Tai-Kadai word for 'wet rice field' is inherited from an Austro-Asiang word meaning 'low-lying/easily flooded ground', with shift to 'wet rice field' in Tai-Kadai. Such a shift is straightforward, assuming that seasonal flooding was the main source of irrigation in the kind of rice agriculture practiced by speakers of a language ancestral to proto-Tai-Kadai.

Benedict (1995) also claimed that proto-Austro-Tai had another word for 'rice': *qasal. This is based on the Austro-Asiang side, on a Formosan (allegedly proto-Paiwanic) *qasal (Paiwan qasal/qaqsan 'unhulled rice', Puyuma qašal 'hulled rice') and in Tai-Kadai on the form salA 'husked (rice)' discussed in 2.2.5 above. However, the correspondence Paiwan sal : Puyuma *s indicates a loanword (Blust 1959 for discussion and examples). This reduces the Austro-Asiang side of the comparison to a single witness (here probably Paiwan). Moreover, Feurle's suggestion on the Austro-Asiang origin of Tai-Kadai salA (2.2.5) is worthy of consideration, and Benedict's comparison is probably invalid. Overall, two of Benedict's Austro-Tai comparisons: 'wet rice field' and 'husked rice/rice as food' appear to have a basis in fact, but the former, at least, makes better sense under a theory which considers Tai-Kadai as an offshoot of Austro-Asiang, and Tai-Kadai agriculture as a continuation of early Austro-Asiang agriculture.

Several comparisons are consistent with the hypothesis that the Sino-Tibetan-Austro-Asiang linguistic expansion was based on rice and millet. The comparison Sino-Tibetan *may 'paddy' : AN *Semay 'cooked rice' fits the
sound correspondences in Sagart (2002), except that proto-Austronesian initial S- would predict Chinese type B syllable, thus **mijj in place of actual **mijj. The shift ‘paddy’ > ‘cooked rice’ is attested in Saisiat, Iatan and Karen independently. In the meaning ‘husked rice’, the comparison Written Tibetan bras < m-ras ‘rice’; Chinese 米 mo-rat-s ‘coarse husked rice’; Austronesian beRas ‘husked rice’ fits the sound correspondences in Sagart (2002) exactly. PAN *na ‘low-lying ground, wet ground’ correspond to proto-Sino-Tibetan *na ‘low and wet land’ (consisting of Old Chinese ra-t 諾 *na-s ‘wet ground on the side of a river’ and Written Tibetan na ‘meadow’) according to the system in Sagart (2002); Beaudic (1975:94) overlooked the Chinese cognate and explained Written Tibetan na ‘meadow’ as a loan from Austro-Tai. One wonders why Sino-Tibetan speakers should borrow a word meaning ‘irrigated rice-field’ from Austro-Tai and use it as a term for meadows or wet grounds. On the general implausibility of Benedict’s theory of Austro-Tai loans to Chinese, see my discussion of metal names (Sagart 1999). Finally, Proto-Austronesian *beCen ‘Setaria italica’ and Old Chinese 稲 *išk ‘Setaria italica’ correspond according to the sound correspondences in Sagart (2002, forthcoming [b]). The C : t correspondence was not listed in Sagart (2002), but new comparisons in its support have recently turned up. There are several good parallels for the correspondence of proto-Austronesian final -i to Old Chinese -k. The comparison for ‘Setaria’ is presented here tentatively for the first time.

Table 1 holds an interesting alignment between Hmong-Mien, Austroasiatic and Sino-Tibetan for the meanings ‘swidden’ and ‘wet rice-field’. It is uncertain how this is to be explained. Proto-Hmong-Mien *liŋA ‘wet rice-field’ can be an early loan from Chinese 田 *liŋ ‘field’, with expected tonal correspondence (tone A to Chinese Ping) and rendition of OC l- by *l, as in ‘iron’: proto-Hmong-Mien *liŋK0, this definitely an early loan from Chinese 米 *hek ‘iron’. A loan from Sino-Tibetan *liŋ ‘field’ is also a possibility. On the other hand, similarity with Austroasiatic li(y)on, liang etc. raises the possibility that we are dealing with an old Yangzian word, loaned to Sino-Tibetan at an early date; or even (wow!) with an antique East Asian word for ‘field’, inherited by all of Sino-Tibetan, Hmong-Mien and Austroasiatic!

Finally, there are no agricultural words uniquely shared by Austronesian and Austroasiatic. This is striking in view of claims that rice domestication has been the motor of the Austro dispersal (Blust 1996, Righam 1998).

ACKNOWLEDGEMENTS

I am indebted to Michel Ferlus for valuable information on Austroasiatic and Tai-Kadai words and to Anne Daladier for alerting me to an important Austroasiatic set for a millet, assembled by herself. Wolfgang Behr, Anne Daladier, Alexis Manaster-Ramer, Marc Miyake, Martha Ratliff and Axel Schuessler read a preliminary version of this paper and made valuable observations. Thanks also to David Bradley for useful discussion of a modified version. All errors are mine.

REFERENCES


SAGART: VOCABULARY OF CEREAL CULTIVATION AND PHYLOGENY OF EAST ASIAN LANGUAGES


