Roger Blench

The origins of nominal classification markers in MSEA languages

Convergence, contact and some African parallels

1 Introduction

The languages of the world characteristically have morphological strategies both to classify nouns and to signify to speakers and hearers alike aspects of the semantics of those nouns. These strategies can be broadly divided into two categories: noun classes and numeral classifiers. Noun classes are here treated as devices for categorizing nouns semantically. Noun classes can be expressed overtly through bound affixes or other direct marking on the noun, or indirectly, on verbs (as in Navajo) or other parts of speech such as dependent verbs, adjectives, pronouns and prepositions, as in North Caucasian. Where the class is marked with a bound affix, noun class marking can be concordial: that is, other parts of speech, typically adjectives and demonstratives, agree with the class marker. This agreement may be alliterative, where the segmental material is the same or closely related, or non-concordial, where the relationship between class and marker is regular but the segments are dissimilar[1]. A classifier is a word which accompanies a noun in certain grammatical contexts, and reflects a semantic classification of nouns. Typically, where the noun is counted or measured, the classifier is an obligatory accompaniment to the noun.

Classifiers are typical of most South and East Asian language phyla, as well as being scattered across the New World. Compared with noun class markers, classifiers in individual languages tend to be very numerous, and some languages may have several hundred. The lists given in Adams (1989) or Post (2007) are extremely long, and the classes are strongly concerned with the

1 This is typical of many Bantoid languages, where an originally alliterative system has been restructured, producing a disjunction between agreement markers.

shape and appearance of the noun. Although this type of semantic association is also characteristic of Niger-Congo languages, noun class markers are always restricted to a small number of bound affixes.

Although noun classes are often thought to be absent in SE Asia, there is increasing evidence that they exist as a parallel system in both Daic and Hmong-Mien. Ratliff (2010: 267) treats the nominal prefixes of Hmong-Mien as 'weakly classifying'. Enfield (2007: 146) calls one system of marking nominal semantics in Lao [and also Thai] 'class-terms[2]. These consist of obligatory accompaniments to nouns, such as Thai marking all fish with a preceding *pla*, corresponding to Lao *pa* (reduced from *paa*[3]). Lao has quite a number of these terms, which are almost always etymologically transparent, although Lao *ka-* is somewhat opaque, applied to small creatures and objects (Enfield 2007: 150). They typically define taxonomic essences, colours, roles and functions. Exactly how widespread they are in SE Asia is unclear since their description is often conflated with numeral classifiers. De Lancey (1986) argues that class-terms can be reconstructed back to proto-Tai. English has a fragmentary system of this type in that the names of birds and fish are sometimes accompanied by the term itself ('blackbird', 'mutton-bird', 'dogfish', 'catfish'). In contrast to numeral classifiers, they are unrelated to number and quantity. Reduced noun class systems, such as the four-term systems in Australian languages like Dyirbal (Dixon 1972) are quite common in the Papuan and Australian language areas (Harvey and Reid 1997).

A third system, characteristic of phyla such as Indo-European and Afroasiatic, is the sex-gender system, which classifies nouns through notional male/female oppositions. Although animates with biological gender are marked with the appropriate sex marker, their extension to non-animates rarely conveys further semantic information. In French, for example, extremely similar body parts, such as 'arm' and 'hand', are assigned different genders. Noun classes in concordial languages are often referred to as 'genders' but this is a confusing terminology as it conflates a genuine semantic categorization with sex-gender systems. Sex-gender systems are usually concordial, unless they are part of a broader system of noun classes. Dyirbal, for example, marks male/ female distinctions, but includes water, fire and violent acts with the female class (Dixon 1972).

Languages are, broadly speaking, conservative and phyla can be characterized by particular strategies. So the great majority of Niger-Congo languages

2 A term said to have been introduced by Mary Haas (1964). See also Beckwith (1993) for further discussion of this terminology.

have noun classes or nothing; numeral classifiers or sex-gender systems rarely develop. Afroasiatic languages exhibit sex-gender throughout the phylum and indeed the morphology used to express this is highly conservative. Austroasiatic, Austronesian, Hmong-Mien and Sino-Tibetan languages all have numeral classifier systems which do not show agreement. However, at least in two phyla of SE Asia, semantically associated affixes show formal and functional similarities. This paper considers various models to account for these formal and functional similarities.

Concordial noun classes are not found in SE Asia, but they do occur sporadically in Papua and Australia, notably in the Ngarnic language Yanyuwa, which has sixteen classes (Kirton 1988). Hammarström (2013) has reviewed the occurrence of these in some detail. Astonishingly, Yanyuwa has a *ma-* prefix for fruits, which although identical to a common SE Asian class-term, is presumably just coincidence. However, concordial systems are common in Niger-Congo languages, as well as in some New World language phyla, such as Arawan and Kiowan. The evolution of these systems is not well understood, but the recent description of predicate classifiers in Nilo-Saharan (Ahland 2010) may provide a clue as to how these have developed in Niger-Congo. A secondary argument of this paper is that some of the morphological processes at work in SE Asia also help shed light on nominal classification in African languages.

2 Noun class affixes

Typical noun class morphology consists of a root and an affix. The affix can be prefixed, suffixed, infixed or appear as a circumfix. In rare cases, languages exhibit double-affixing: two separate affixes which alternate according to distinct rules. Examples of such languages in Africa are Bassari on the Togo-Ghana borderland and the Tivoid languages of SE Nigeria (Greenberg 1977). As part of the erosion of such systems, various types of fusion can occur, but the original morphemes are usually reconstructible. The affix in principle has a semantic assignation, which may or may not be opaque. Affixes frequently alternate; thus singulars can have one or two marked plurals. In Nilo-Saharan it is often considered that the 'middle' is the unmarked term and a singulative and a plurative can be formed from the root (Dimmendaal 2000). But this is not a necessary requirement of a noun class language; in Niger-Congo the *m-* class for mass nouns is always an unpaired class (Greenberg 1963; Blench 1995).

In SE Asian language phyla, word structure is often described as 'sesquisyllabic' (Matisoff 1973). That is, words have major and minor syllables, i.e. an

iambic structure. The major syllables are in the stem and the minor syllable a prefix, generally C or CV. Since the –V is often represented orthographically with a mid-central vowel, it may be that it is not realized phonetically. Minor syllable prefixes are either lost or optional in many languages, and they seem to change in ways that do not suggest phonological shift but affix substitution. Austroasiatic, Sino-Tibetan and Hmong-Mien all show this behaviour. In Austronesian languages, the tendency is for the prefix to have a (C)V form and to be conserved. In other words, once a prefix has been fused to a stem, it is retained, from Taiwan to New Zealand, as it were. Daic languages are typically CV(C) and except in rare cases no longer retain the minor syllable. The loss of prefixes in Daic may well explain the adoption of class-terms as a substitute strategy for semantic marking of nouns.

Minor syllables thus have the appearance of optional prefixes in many languages. Anderson (2004) observes that in Munda the final syllable is the 'stable, meaning-associated element' while the prefixed syllables are unstable and cannot be assigned a meaning. Intriguingly, Blust (1988) also identifies an apparently similar system in Austronesian, where roots seem to retain a cross-language basic meaning, but are preceded by a variety of CV prefixes which transform the meaning in individual languages. Extended examples can be seen in the 'roots' section of the online Austronesian Comparative Dictionary[3]. Blust considers this as an example of phonosemantic association, similar to phonaesthemes identified elsewhere in the world (e.g. sl/gl in English). However, as Sagart (2011) observes, this system has striking similarities to the MSEA structures identified here.

It is certainly the case that the minor or prefix syllables have no obvious semantic assignations, and in no SE Asian language do they show concord. But to assume that they have 'no meaning' suggests a curious model of language. A general postulate of morphology is surely that the elements of words either do have or formerly have had meanings. These can be obscured over time, but one task of linguistics is surely to tease them out. In the case of prefixes in MSEA languages, the fact that they vary dynamically from one language to another is surely a reflection of their significance for speakers. They cannot be simply euphonious noise. Outside SE Asia, Nilo-Saharan languages show a wide range of affixes which suggest a former nominal marking system (Bender 1996; Storch 2005), but synchronically, no Southeast Asian language shows a productive

3 http://www.trussel2.com/acd/. It is striking that the majority of Blust's examples focus on Western Malayo-Polynesian, especially the Philippines. It is as if the system is completely dropped in Oceanic.

system comparable to those in Niger-Congo. Affixes certainly change to mark number, but alliterative concord is unknown. This system is stable, and Nilo-Saharan shows no sign of eliminating this unproductive morphological baggage and developing in the direction of a SE Asian type system.

Austroasiatic and many branches of Sino-Tibetan[4] have a common word structure where the root is preceded by a C- prefix. Although C- prefixes may have semantic correlates, this is inconsistent between languages. The prefix may disappear or be substituted, while the root remains static. The C- prefix can sometimes be incorporated into the stem, and a new prefix added, leading to complex initial sequences (cf. examples in Matisoff 2003). Additional evidence for this is drawn from the typical pattern of pronouncing initial consonant sequences as individual segments; thus 'spr' in Austroasiatic and Sino-Tibetan is pronounced s.p.r, rather than a cluster as in Indo-European. These similarities between the two phyla are rather perplexing, as few historical linguists consider them to be related. Globally, such systems are extremely rare, and for them to have arisen independently given their direct geographical proximity is unlikely. However, clear examples of common lexemes of any time-depth are few[5], and these are often shared with other regional phyla such as Daic and Hmong-Mien. A historical scenario to account for this structural convergence is not obvious; the likely Urheimats of these two phyla are far apart.

This paper describes the features of word structure in Austroasiatic and Sino-Tibetan that appear to be convergent and suggests how they might have arisen. It will argue that such features are transitional towards the evolution of true noun classes and introduces a typological parallel from West Africa. The hypothesis is that the SE Asian affix system originates from frozen numeral classifiers (and noun class terms) and that as the system is renewed, these prefixes co-exist together with productive classifiers. It will examine possible borrowing scenarios and suggest that while these can be detected, they are inadequate to explain the diachronic morphology.

4 I am aware of the controversy between this term and Tibeto-Burman. See below for further discussion. Whatever the case, Sinitic languages are in consideration, as the evidence for the type of canonic form described here is well attested in Old Chinese.

5 Benedict (1990: 4) says, 'there is little evidence of any borrowing of lexical items of 'core' type by TB/ST from AA/MK'. However, he does give some striking examples from kin terms, as also the animal names 'hawk' and 'tiger' (cf. Table 7).

3 Word structure in Austroasiatic and Sino-Tibetan

Word structure in Austroasiatic (at least for nouns) seems to consist of one or more optional C(V) prefixes, a CV(CV) or CVA stem [where A is an approximant] and a C suffix, often weakened to a glottal stop or deleted. The optional C(V) prefixes are sometimes referred to as a 'pre-syllable' in the literature. Many nouns may have had a labial or palatal approximant in final position and this has a strong tendency to be incorporated into the stem. If it represents a different prosody from the vowel of the stem, then the synchronic output may be either a diphthong or a long vowel.

There is a background murmur in the literature suggesting the existence of old affixes with semantic content which derive from frozen classifiers (e.g. Costello 1996, 1998). For example, Thomas (1969: 105) gives evidence for a *sa-* prefix in the Bahnaric language Chrau which denotes animals. She says "For the most part the first syllable is never dropped, except in direct address" which of course does mark its optionality in the minds of speakers. Although she was unable to find comparative evidence, in fact this prefix appears to be quite widespread, as Shorto (2006: 469) notes a number of cognates. Table 1 cites cognates for the Chrau term for 'bear' [the animal] which shows that the *s-* prefix occurs in Bahnaric, Katuic and Vietic, with further possible cognates for the root itself in Aslian and Pearic.

Table 1. An Austroasiatic root for 'bear' with variable prefixes

Language	Subgroup	Attestation
Jahai	Aslian	kaw.ip
proto Bahnaric	Bahnaric	*c.kaw ~ *gaw
Laven [Jru']	Bahnaric	h.kaw
Sedang	Bahnaric	rə.kǫw
Chrau	Bahnaric	si.kaw
Ngeq	Katuic	haŋ.kaw
Bru	Katuic	sa.kaw
Chong [of Kanchanaburi])	Pearic	kəw.x̯aj suːʔt
Vietnamese [Hanoi])	Vietic	gấu
Chứt [Rục])	Vietic	cə.kuː

Smith (1975) points to the widespread presence of a velar prefix for animal names both in Sedang, and more broadly in the Vietnamese languages he sampled. But this prefix is found across Austroasiatic and also, strikingly, widely in Sino-Tibetan. Matisoff (1973) draws attention to its presence in Lolo-Burmese, but as Benedict (1990) notes, it is present on the words for ‘tiger’ and ‘hawk’ throughout the phylum.

The Sino-Tibetan language phylum has a disputed internal structure and thus debatable reconstructed forms. Van Driem (2008), Handel (2008) and Blench and Post (2013) give an overview of some of the key issues. These swirl around the position of Sinitic, formerly considered a primary branching, but now often treated as simply another branch within Sino-Tibetan, hence the rechristening by some authors of Sino-Tibetan as Tibeto-Burman. Proposals to rename the phylum in a more neutral fashion (e.g. Tibeto-Burman or Trans-Himalayan) certainly have merit. Sinitic shares far more lexically with Tibeto-Burman than some of the isolated groups of Arunachal Pradesh (Blench and Post 2013). As with Austroasiatic, Sino-Tibetan words tend to have a core segmental structure, very often CV(N), and then one or more affixes, both prefixed and suffixed. Affixes can shift right or left into root medial position, inducing consonant and vowel changes, and prosodies affecting C_1. Change in the segmental character of C_1, such as n→ɲ or ŋ, is deemed to be driven by a shift of palatalization or velarization into the core. The perceived incorporation of a consonant within the stem leads to affix renewal, and thus stacking of unproductive morphemes.

In some languages of the region, the use of the fricative to mark animal names is notable. Table 2 gives an example from Western Miji, an only doubt-

fully Sino-Tibetan language spoken around Nafra in Arunachal Pradesh[6]. The palatal fricative /ʃ/ is the most common prefix, but I am assuming *s~ʦ* are probably its allomorphs.

Table 2. Animal names in Nafra Miji

Gloss	ʃ	s/ʦ
'animal'		stɕõ
'horse'	ʃgrɔ	
'stallion'	ʃgrɔ mbŭ	
'mare'	ʃgrɔ mnĕʔ	
'colt'	ʃgrɔ i	
'sheep'	ʃgθɔʔ	
'goat'	ʃprn	
'dog'	ʃazi	
'barking deer'		tstshũ
'deer'		tstsə
'flying squirrel'	ʃbiã	
'leopard'	ʃnmu	
'monkey'	ʃbŏ	
'musk deer'		tstsɲăw
'pangolin, anteater'	ʃgʤɔ	
'wild cat'	ʃgrɛ̆	
'wild dog'	ʃkʃə	
'sparrow'		slĭʔ
'ant'	ʃɲi	
'fish sp. I'		sθɯ̆
'fish sp. II'		sviaʔ
'fish sp. III'		sgiɔʔ

The neighbouring Hruso language also shows an S- prefix for animals, although it shares almost no lexical cognates with Miji, except probably 'ant' *ʃn* (Table 3).

Table 3. Hruso animal names with S- prefix

Gloss	Hruso
'wild animal'	sm ʧi
'dog'	ʃʌuɔ
'bear'	stɔ̄ɔ

[6] All data from NE Indian languages is based on my own fieldwork in 2010 and 2011, and I would like to take the opportunity to thank the many people who helped me, as well as Jummar Koyu and Jiken Bomjen, who arranged my field trips.

Gloss	Hruso
‘otter’	sžɛ
‘rat’	ʒmɔ
‘ant’	ʃn
‘caterpillar’	ʃblu
‘flea’	sgzə
‘bloodsucking fly’	sdžm
‘cobra’	ʒ̄tɔ̌
‘python’	ʒ̄ʃaba
‘snake sp. I’	ʒmə
‘frog I’	ʃʤa
‘snail’	svankɔ̌

These dissimilarities suggest strongly that what has been transferred is the idea of the semantics of a prefix rather than actual lexical items. By contrast, the neighbouring Koro language, which is structurally very similar to Hruso and Miji, shows no trace of S- prefixes.

Miji and the related Bangru also have a very marked *m-* prefix related to body parts, both for humans and animals. Table 4 shows a comparative list of Western and Eastern Miji as well as Bangru. The Bangru citations are orthographic, and on comparative grounds, it is assumed the vowel following the *m-* prefix is epenthetic. Forms in square brackets are cited for completeness, where one branch has an *m-* prefix and the other lacks it.

Table 4. Miji and Bangru body parts with an m- prefix

Gloss	W. Miji	E. Miji	Bangru	Comment
‘arm, hand’	(m)gĭ	(m)gĭ	m(e)gey	No Tibeto-Burman cognates
‘beard’	mɔmyuʔ	mmɯʔ	m(a)maŋ	Widespread Tibeto-Burman root, though not with m- prefix
‘bone’	mriaŋ	mriaŋ	mnii	Possibly cf. Northern Naga **raŋ*
‘brain’	mɲɔʔ	mɲɔʔ		No certain external cognates, though cf. Bodic, e.g. Tshangla *ɲ̥ok taŋ*
‘breast’	mɲu	mnɯʔ		*m-* ‘body part prefix’ plus widespread Tibeto-Burman etymon **nu(w)*
‘chest’	mθm kʸu	mlọŋ kə̣ʔ		The kV- element has widespread Tibeto-Burman cognates. Note Puroik *tə kɯ*
‘chin’	mugudza	mgutɕă		No Tibeto-Burman cognates
‘ear’	mʒɔʔ	mzɔʔ	m(i)bwa	No certain external cognates, but cf. Memba *namʤo*
‘eye’	mmreʔ	mreʔ		No Tibeto-Burman cognates
‘face’	mgmiaʔ	mkmiaʔ	m(e)kwii/mekuyi	Matisoff (2003) proposes *#s.myal* for PTB. The best cognates are in Maraic, e.g. Lakher *h.mia*,

Gloss	W. Miji	E. Miji	Bangru	Comment
				but the velar preceding the Miji stem is of unknown origin.
'finger'	mgi tso		m(e)gey ʧowa	cf. 'arm'
'flesh'	mzaʔ	mʒaʔ		#sa is widespread in Tibeto-Burman, but this may be coincidence
'heart'	luŋ, [θɔm vʸu]		mloŋ	#luŋ is widespread in Naga complex languages
'kidney'	mkbɔ̆		mpega	Neither root has a Tibeto-Burman cognate
'liver'	mtn		m(a)tayiŋ	cf. Chin roots such as Thado *tʰin*, and possibly proto-Tani **zin*.
'lungs'			mloŋ wasayi	cf. 'heart'
'mouth'	mugɔ̆		m(i)niŋ	STEDT relates the *gɔ* element to proposed PTB *#ku(w)*. Some Tani languages have apparently similar forms, e.g. Apatani *a.gũ* but this is not apparently proto-Tani. The Bangru form has no obvious cognates.
'navel'	mʃmay			No Tibeto-Burman cognates. The ʃ- appears to be an earlier prefix.
'neck'	[dmuzɔ̄]		m(i)niŋri	Scattered attestations in Kuki and Chin, e.g. Lushai *#riŋ*. Possibly related to much more widespread *#luŋ*
'nose'	[ɲubyuŋ]		m(i)niiko	Miji has ɲi 'blow nose'. Chin languages have common *niit* for 'blow nose'
'rib'			mpelowa	No Tibeto-Burman cognates.
'shoulder	'mfa		mpoʧ	Miji has very scattered Tibeto-Burman cognates, e.g. Chinbon *pá*, though forms with a back high vowel are widespread. No obvious cognates for Bangru.
'stomach'	mrɔ̆		mulgu	No Tibeto-Burman cognates.
'thigh'	mləʔ		murʰ	No clear Tibeto-Burman cognates. Isolated Thado *mʌ́l*, also possibly metathesis of Bodic *lum* (e.g. Tshangla).
'throat'	mryɔnza			No Tibeto-Burman cognates.
'tooth'	mtɾ̄		m(e)tʰu	No clear Tibeto-Burman cognates, except possible Puroik *kətuŋ*
'vein'	mdtʔ			No clear Tibeto-Burman cognates.
'wrist'	gi mvθɛ			No clear Tibeto-Burman cognates.
Animals				
'horn'	mʃʒɔ̃		m(e)ws	No clear Tibeto-Burman cognates.
'tail'	mdmray		m(u)lwe	Tibeto-Burman has widespread **may* or similar. If this is cognate then it is an example of multiple re-affixing
'hump'	mkbʸu			Isolated possible cognate Bokar (Tani) *gur bɯŋ*
'tusk'	mʈŭ			No Tibeto-Burman cognates. In many Sino-Tibetan languages, the same word as 'tooth'

Gloss	W. Miji	E. Miji	Bangru	Comment
				but not here.
'udder'	mɲŭʔ			Possibly cf. Tangkhulic Huishu ʔa-*nə-nuk*
'fur, feather'	mɔmyŭʔ			Widespread Tibeto-Burman *mu(l)* but no other language shows palatalization
'wing'	mktɕi			No clear Tibeto-Burman cognates.

Table 4 shows that Mijiic has a strong preference for an *m-* prefix for human and animal body parts, even where this is not attested in external cognates. There is limited comparative evidence for a Tibeto-Burman *m-* prefix, see for example Matisoff (2008: 183) on **m-ley*~**m-li* for 'penis'.

A language spoken nearby, Mey [=Sherdukpen] of Rupa, also has the *m-* prefix but marking fruits (Table 5), which parallels the widespread *ma(k)* class-term, found in Tai languages.

Table 5. m- prefix for fruits in Mey of Rupa

Gloss	Rupa
'fruit'	m.laŋ
'banana'	m.suŋ
'lemon'	m.kẽ
'sugar-cane'	m.ʧi
'walnut'	m.ku

Forrest (1962), in an article not often cited, points out that Lepcha (Rong) has the same *kV-* prefix for animals noted for Palaungic and Khmer. Rong also shares other prefixes with Austroasiatic, for example, the *sV-* prefix mentioned above, which is also attested in Khasian and Palaungic. Rong uses a *ma-* prefix for trees and fruits, similar to the examples above. The nominalizer which forms abstracts in Rong, *nun/num-*, is also widely attested in Austroasiatic.

As an example of how the *kV-* animal affix is realized synchronically, Table 6 shows a widespread root for 'buffalo' attested in most branches of Austroasiatic[7]. The term is borrowed into Austronesian and gives us the common English name carabao. The attestations in different languages provide an example of the complex build-up of prefixes that characterizes this type of morphology.

7 Munda has *bɔŋtel* throughout, which may be the same root with the *–tel* an old compound. Mangic languages have *vɔ*, which again could well be cognate but a lack of morphology makes this speculative.

Shorto (2006) reconstructs **krpiʔ* for PMK[8], but the evidence seems to better support either a back or central vowel and a final palatal, thus the suggestion **k.r.pu*.y. I have analysed the synchronic forms as a combination of a root, plus segmental affixes, each separated by a full stop. The proposal for the leftwards movement of the final palatal to the interior of the root is shown with a raised ^{y}, thus $p^{y}u$. The front vowels arise from the final –y being incorporated into the stem. Sometimes this is merely lost and the back vowel is retained or lengthened. Whether the earliest form had a three consonant cluster in initial position is debatable. The original could have been **r.pu* as in Khmuic, which subsequently gained a *k-* animal prefix. Proto-Khmuic must have had something like **g.r.pu* to explain the synchronic forms. In Vietic, the b/p of the root was lost and r→l, generating *k.l.Vw* structures. The final nasal in Mon is mysterious unless it arose under the influence of the *k-* prefix.

Table 6. 'Buffalo' *#k.r.pu.y* in Austroasiatic

Phylum	Branch	Language	Attestation	Formula	Comment
Austroasiatic		PMK (Shorto)	*krpiʔ	k.r.pi.ʔ	
Austroasiatic		PAAS (RMB)		k.r.*pu*.y	
Austroasiatic	Monic	Mon	preaŋ	r.$p^{y}u$.ŋ	
Austroasiatic	Monic	Nyah Kur	chəlo̱w	k.r.(p)*u*	? < Vietic
Austroasiatic	Vietic	proto Vietic	*c-lu	k.r.(p)*u*	
Austroasiatic	Vietic	Thavung	khuay[1]	k.r.(p)*u*.y	
Austroasiatic	Vietic	Pong	klow	k.r.(p)*u*	
Austroasiatic	Khmeric	Khmer	krəbɤy	k.r.*pu*.y	? < Stieng
Austroasiatic	Pearic	Pear	krəpaːw	k.r.*pu*.y	
Austroasiatic	Pearic	Chong	kapaːw^{A}	k.$p^{y}u$	
Austroasiatic	Bahnaric	PNB	*kapɔː	k.*pu*	
Austroasiatic	Bahnaric	Sedang	kopôu	k.*puu*	
Austroasiatic	Bahnaric	Tampuon	kəpəu	k.$p^{y}u$	
Austroasiatic	Bahnaric	Bahnar	kəpoː	k.$p^{y}u$	
Austroasiatic	Bahnaric	PSB	*g~rəpuː	k.r.*pu*	
Austroasiatic	Bahnaric	Mnong	rpu	r.*pu*	
Austroasiatic	Katuic	Proto-Katuic	*krpiiw	k.r.$p^{y}u$	
Austroasiatic	Katuic	Pacoh	kərbɤː	k.r.*pu*.w	alligator; dragon [!]
Austroasiatic	Katuic	Chatong	karpiiw	k.r.$p^{y}u$	
Austroasiatic	Khmuic	Sre	rəpu	r.*pu*	
Austroasiatic	Khmuic	Chrau	gəpuː	k.*pu*	
Austroasiatic	Khmuic	Biat	rpuː	r.*pu*	
Austroasiatic	Aslian	Kensiw	kɛˈpaw	k.$p^{y}u$	unless < Malay

8 I use proto-Mon-Khmer when citing previous literature, but in general this terminology should be discouraged, as perpetuating an outmoded classification (Sidwell and Blench 2011).

Phylum	Branch	Language	Attestation	Formula	Comment
Austroasiatic	Aslian	Temiar	kəɹbau	k.r.$p^{y}u$	< Malay

'Buffalo' indicates clearly the morphological path these nouns characteristically take: prefixes seem originally to have been numeral classifiers with semantic coherence. As they became semantically bleached, a new prefix was added, and the initial prefix incorporated into the stem. This creates a consonant string in initial position (i.e. k.r.p) and any one of these consonants can be deleted. This can lead to highly diverse synchronic outcomes. An Austronesian language such as the Chamic Rhade, which today has *kbao*, probably originally had a longer, more characteristically Austronesian form, and has restructured it under the influence of its Austroasiatic neighbours. A Daic language such as Nung has a synchronic form *tú vai*, i.e. classifier plus stem, which may have been borrowed from a Vietnamese Austroasiatic language. This could also explain deviant Katuic forms such as Katu *tariiq*, which would originally have resembled Chatong *karpiiw*. The *k-* prefix became a suffix, the stem consonant *p-* was deleted and a now unproductive *t-* prefix was added, perhaps on the model of the Daic nominal classifiers.

Table 7 is a second illustration of the *k-* prefix for animals in SE Asian languages, showing a common root for 'tiger', attested across phylic boundaries.

Table 7. The *#kVla* root for 'tiger' in SE Asian languages

Phylum	Branch	Language	Attestation
Sino-Tibetan	Sinitic	OCM	*hlâʔ
Sino-Tibetan	Burmic	Old Burmese	klya
Sino-Tibetan	Bodish	Monpa	khai-la
Austroasiatic	Khmeric	Angkorian Khmer	khlaa
Austroasiatic	Pearic	Samre	kanɔh^{A}
Austroasiatic	Bahnaric	PB	*kəlaa
Austroasiatic	Bahnaric	Sedang	klá
Austroasiatic	Katuic	Pacoh	kulaa
Austroasiatic	Katuic	Ir	kalaʔ
Austroasiatic	Katuic	So	kula
Austroasiatic	Palaungic	Shinman	kaʔ^{4} vai^{3}
Austroasiatic	Monic	Proto-Monic	*klaaʔ
Austroasiatic	Aslian	Sakai	kla
Austroasiatic	Khasian	War Jaintia	k^{h}la
Austroasiatic	Muṇḍā	Muṇḍā	kula
Daic	Tai	Thai	kla

The root for ‘tiger’ illustrates how semantically assigned affixes are borrowed. The tiger is an animal of great symbolic importance across the region and the word has probably been borrowed extensively, including fossil morphology. Once a semantic association of a *k-* prefix for ‘animal’ is set up (cf. Table 6) it is easily generalized to other animals within a particular speech community and thence to other languages in the same geographic region.

Table 8 is intended to demonstrate how this works in Tibeto-Burman with the root for ‘two’, not a noun, but subject to analogous processes. The starred forms are drawn from standard sources, and are not necessarily endorsed, merely cited for a convenient comparison. It is assumed there was a core *ni*, with a velar prefix and two suffixes, a fricative and a high front vowel or an approximant. The velar prefix was regularly suffixed and weakened to *ʔ*. The fricative suffix was either affricated or weakened to *–h* and switched to a prefix. Forms like Cho *hŋgih* may represent copying, so that the affix appears at both ends of the word. Other more sporadic affixes are added, such as *p-*, *t-*, *r-* and possibly *a-*.

Table 8. The root C.ni.C(C) for 'two' in Tibeto-Burman

Language	Group	Form	Formula
*Sino-Tibetan	Sino-Tibetan	gnyis	g.nʸi.s
*Tibeto-Burman	Tibeto-Burman	g-ni-s	g.ni.s
*Karen	Karenic	hni	h.ni
*Lolo-Burmese	Lolo-Burmese	ʔnit	ʔ.ni.t
*Loloish	Loloish	s-ni(k)[2]	s.ni.k
*Northern Naga	Northern Naga	ʔ-ni	ʔ.ni
Bugun	Bugun	ɲeŋ	nʸi.ŋ
Taraon	Mishmic	kaiŋ	k.ni
Idu	Mishmic	kaɲi	k.nʸi
Puroik	Puroik	ɲi	nʸi
Kamengic	Mey of Shergaon	ɲit	nʸi.t
Miji	Mijiic	gni	g.ni
Miju	Mijuish	knîn	k.ni.n
Koro	Siangic	ki-ne	k.ni
Milang	Siangic	nə	ni
Karbi	Mikir	hiní	h.ni
Meithei	Meithei	ə-nì	ə.ni
Newar (Dolakhali)	Newar	nis	ni.s
Atong	Bodo-Garo	ni	ni
Garo	Bodo-Garo	gəni	g.ni
Kokborok	Bodo-Garo	nəy	ni.y
Ao (Mongsen)	Naga	anət	a.ni.t
Rongmei	Naga	kənəi	k.ni.y
Tangkhul	Naga	³khə ³ni	k.ni

Language	Group	Form	Formula
Phom	Northern Naga	ñi³¹	nʸi
Cho (Mindat)	Chin	hngih	h.n(g)i.h
Daai	Chin	ŋ̥ɲ̊iʔ	ŋ.ni.ʔ
Khumi	Chin	nue(ng)	ni.ŋ
Lai (Hakha)	Chin	pa-hniʔ	p.h.ni.ʔ
Lakher [Mara]	Chin	³sa ²nɒ	s.ni
Lakher [Mara]	Chin	pā-nō	p.nʷi
Lushai [Mizo]	Chin	hnih	h.ni.h
Matu	Chin	pa̱ʌniʔ	p.ni
Nyhmoye	Chin	ŋ̥ɲ̊iʔ	ŋ.ni
Bhramu	Himalayish, Western	nis	ni.s
Kanauri	Himalayish, Western	nis	ni.s
Motuo Menba	Monpa	ŋa̱ik tsiŋ	nʸi.k
Kaike	Bodic	nghyi	g.h.nʸi
Tshona (Mama)	Bodic	nʌi¹³	a.ni
Tibetan (Alike)	Tibetic	ɣŋa̱i	g.nʸi
Tibetan (Amdo: Bla-brang)	Tibetic	hŋa̱i	h.nʸi
Tibetan (Balti)	Tibetic	ŋis	n(g)i.s
Tibetan (Sherpa)	Tibetic	ngyi	g.nʸi
Dirang	Tibetic	nitsiŋ	ni.ts.ŋ
Tawang	Tibetic	neʲ	ni
Memba	Tibetic	ɲi	nʸi
Meyor	Tibetic	ni	ni
Burmese (Written)	Burmish	hnaʦ	h.ni.ʦ
Marma	Burmish	hnɔiʔ	h.nʷi.ʔ
PNL	Loloish	ʔnitᴸ	ʔ.ni.t
Ahi	Loloish, Central	ni²¹	ni
Lalo	Loloish, Northern	ni²¹	ni
Nasu	Loloish, Northern	ŋa̱i⁵⁵	nʸi
Nusu (Southern)	Loloish, Northern	ɦĩ³⁵	h.ni
Akha	Loloish, Southern	nyì	nʸi
Mpi	Loloish, Southern	ɲiʔ²	nʸi.ʔ
Naxi	Naxi	ŋa̱i²¹	nʸi
Chinese (Old)	Sinitic	njijs	nʸi.(y)s
Sak	Luish	níŋ-hvú	ni.ŋ
Anong	Nungic	əni	ə.ni
Dulong	Nungic	ɑ³¹ ni⁵⁵	a.ni
Nung	Nungic	ɑ³¹ ŋ̊⁵⁵	a.nʸi
Ersu	Qiangic	nɛ⁵⁵	ni
Guiqiong	Qiangic	ŋa̱i³³	nʸi
Namuyi	Qiangic	ŋa̱i⁵³	nʸi
Qiang (Mawo)	Qiangic	ɣnə	g.ni
Tangut [Xixia]	Qiangic	nji	nʸi
Caodeng	rGyalrongic	ʁnes	r.ni.s
Daofu	rGyalrongic	ɣnə	g.ni
rGyalrong	rGyalrongic	kěněs	k.ni.s

Language	Group	Form	Formula
Gurung (Ghachok)	Tamangic	ŋĩhq	ŋ.nʸi.h.q
Tamang (Sahu)	Tamangic	'nyi:h	nʸi.h
Thakali	Tamangic	'ngih	n(g)i.h
Kayan (Pekon)	Karenic	θanɬ	t.ni
Magar	Kham-Magar	nis	ni.s
Thulung	Kiranti	nək	ni.k
Limbu	Kiranti, Eastern	nɛccʰi	ni.s
Bahing	Kiranti, Western	nik-si	ni.k.s
Apatani	Tani	tá-ñe	t.nʸi
Nah	Tani	a-ɲi	a.nʸi
Tujia	Tujia	ɳ̩ie⁵⁵	nʸi.V

The rapid switching and replacement of affixes in Tibeto-Burman illustrates the problems inherent in the usual process of reconstruction. The 'method', such as it is, involves choosing a common segmental core and then proposing the most commonly attested affixes to accompany it. But common affixes may well be evidence for lower-level nodes, or indeed diffusion. The similarities between affixes attested in both Austroasiatic and Sino-Tibetan show that these can spread from language to language and indeed across phylum boundaries.

Sinitic historical phonology allows us to see these processes as they occur. Early Zhou Chinese has many more affixes familiar from other Tibeto-Burman languages than its later descendants. Table 9 shows a set of lexemes attested in Zhou which are found either with fewer or no affixes in later forms.

Table 9. Affix movement and loss in the evolution of Chinese

Gloss	Early Zhou	Classical	Character
blood	s.wiːt	wiːt	血
fire	s.məːyʔ	mə̆ːy	火
head	s.luʔ	lŭ	首
black	s.məːk	mə̆ːk	黑
see	keːn.s	keːn.s	見
seed	toŋ.ʔ	tŏŋ	種

To illustrate the semantic convergence of Sino-Tibetan and Austroasiatic affixes, Table 10 shows one of the principal roots for 'bear' in Tibeto-Burman. The original form may have been something like *twŏ.m* , currently attested in rGyalrong. This would account for many synchronic forms with roots such as *vom, wom, hom*, with or without affixes. However, strikingly, the common prefix for 'bear', as for some other animal names in Sino-Tibetan, is *s-*, just as in Austroasiatic.

Table 10. 'Bear' in Tibeto-Burman languages

Group	Language	Form	Formula
Central Loloish	Kucong	sa^{35} mu^{31}	s.-m
Chin	Lakher [Mara]	chā-vỳ	s.vo.m
Chin	Lushai [Mizo]	sà-váwm	s.vo.m
Kham-Magar-Chepang-Sunwar	Chepang	siŋʔ.tyamh.yom	s.t.vo.m
Meithei	Meithei	shaum	s.wom
Naga	Lotha Naga	sēvá̱ŋ	s.vo.m
Naga	Ao (Chungli)	shim	s.hʸom
Naga	Khoirao	chawom	ʧ.wom
Naga	Lotha Naga	seva	s.vo(m)
Naga	Maram	sahom	s.hom
Naga	Rongmei	cagüm	ʧ.g.wom
Naga	Tangkhul	¹si ¹ŋom	s.g.wom
Nungic	Rawang	ʃəwi^{53}	s.wʸo(m)
Sinitic	Chinese (Old/Mid)	g̱ium/g̱iung	g.yom
Loloish, Southern	Akha	xhà-hḿ	g.hom
Loloish, Southern	Hani (Khatu)	sjhí	ʃ.hʸo(m)
Tani	Galo	sotum	s.tom
Tani	Padam-Mising [Abor-Miri]	si-tum	s.tom
Tani	Apatani	si-tĩ	s.tʸo
Tani	Bengni	šu-tum	s.tom
Tani	Bokar	šu-tum	s.tom

This evidence can be taken to demonstrate;

a) that Sino-Tibetan and Austroasiatic have underlyingly similar word structures, without being genetically related
b) that the so-called 'minor syllable' is an optional affix, which can have semantic content, and which can be shifted to a different positions, or incorporated into the stem
c) that unproductive affixes can be subject to renewal, for example reprefixing, without forming consonant clusters
d) that semantic associations of affixes can be borrowed across phylic boundaries, along with the segmental material, and indeed evidently were borrowed at an early stage of the evolution of these phyla

4 Contact, borrowing and metatypy

Describing structural similarities is one thing: accounting for them historically is quite another. A neat explanation would have Sino-Tibetan and Austroasiatic originating in neighbouring areas and these similarities would then be phenomena deriving from early contact. However, this explanation is difficult to support using current hypotheses about geographical origins. There is a long history of varied speculations about the homeland of Austroasiatic (see review in Sidwell and Blench 2011; Blench in press). Diffloth (2005) has generally argued for a southern, tropical locus on the basis of faunal reconstructions. Sidwell and Blench (2011) propose a riverine dispersal from the Central Mekong, based on their parallel array model of Austroasiatic classification. Hypotheses of the homeland of Sino-Tibetan are similarly varied, from the views of Matisoff ('somewhere on the Himalayan plateau'[9]), Van Driem (1998) arguing for Sichuan, and Blench and Post (2013) for Northeast India.

Unless these hypotheses are very misguided, proximate homelands are not the solution. Austroasiatic clearly spread far and fast, probably along the river systems of SE Asia, seeking humid valleys to grow taro while using improved boat technology. Only such a hypothesis would account for the arrival and diversification of the Munda languages in India. If the proposals in Sidwell and Blench (2011) are correct, then this would have been around four thousand years ago, when there is a rapid and sudden expansion of the Neolithic in mainland SE Asia, marked by the spread of 'incised and impressed' pottery (Rispoli 2008). So there may have been intensive contact between Austroasiatic and Sino-Tibetan in the zone between northern Vietnam, Laos and northeast Myanmar, and consequent diffusion of key structural traits. Purely chronologically, these traits are likely to originate in Sino-Tibetan, as they are clearly attested in Sinitic as well as in many of the highly diverse languages of NE India. Probably this question cannot be fully resolved until we have better mapping of the distribution of semantically significant prefixes across multiple language phyla.

Austroasiatic and Sino-Tibetan language phyla are intertwined across much of their geographical range today and we should expect considerable local borrowing. The Munda languages are cut off from the remainder of Austroasiatic by a zone of highly diverse Sino-Tibetan languages. There are evidently two distinct issues: local borrowing, and broader structural similarities between the two phyla. Studies of these issues are sparse; Benedict (1990) discusses Aus-

[9] STEDT Website section: *Homeland and time-depth of Sino-Tibetan*. URL http://stedt.berkeley.edu/about-st (Accessed 09/05/14)

troasiatic loans in Sino-Tibetan and Shafer (1952) is a study of similarities between Khasi and Sino-Tibetan, evaluated in Diffloth (2008). Forrest (1962) and Bodman (1988) both discuss the puzzling issue of apparent Austroasiatic similarities in Lepcha (Rong), a language no longer in direct contact with Austroasiatic.

Although there *are* deep-level lexical borrowings between Austroasiatic and Sino-Tibetan, they appear to be few (Benedict 1990). The similarities of word-structure and affixes are far more striking. What seems to have occurred is extensive metatypy, i.e. long-term bilingualism causing convergence of structures. The infrequency of lexical borrowing must be due to sociolinguistic factors, for example a desire for esoterogeny, marking the separateness of languages. This is probably at its most extreme in Arunachal Pradesh, where neighbouring languages with extremely similar cultural concepts, such as Miji, Hruso and Koro, share no more lexical cognates than could be expected by chance.

The key to these convergent structures is the incorporation and re-analysis of numeral classifiers and class-terms. Both Austroasiatic and Sino-Tibetan may originally have had simple stems, with no affixes marking number, case, semantics or gender. Numeral classifiers, usually CV(C) syllables with semantic assignations, were associated with nouns, usually preceding them, as is still very much the situation in Daic languages. Sino-Tibetan and Austroasiatic numeral classifiers became bound to the root and reduced to C with an epenthetic vowel following, hence their transformation into affixes. Although this occurred to a greater or lesser extent in different languages, consciousness of their separateness was retained. As a consequence, they can be shifted to the end of the root, and even deleted in some languages. The marked template of affix plus stem required a new prefix to be added, either *de novo* or preceding the existing prefix. Meanwhile, distinct numeral classifiers continued to co-exist and continued to be incorporated and renewed. Figure 1 shows a highly schematic visualization of this process of renewal; the examples in the text indicate some of the complexities encountered on the way.

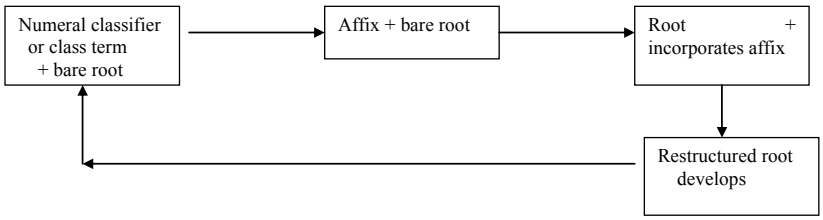

Fig. 1: Cyclical renewal of affixes in SE Asian languages

Most languages in the region also historically had suffixes; it seems likely these were also originally classifiers and indeed, the similarities of segmental material suggest that prefixes became suffixes. Harvey et al. (2006) refer to the notion that such a process is uncommon, while illustrating its operation in Northern Australia (see also Green 1995). The evolution of noun class suffixes in Gur and Adamawa languages within Niger-Congo is similarly an example of prefix-suffix shift, carrying all the segmental material and concomitant alliterative concord. Greenberg (1977) has a perceptive discussion of this issue with respect to the double-affixing languages in Niger-Congo such as Tiv. Within SE Asia, suffixes tend to weaken to glottal stops, nasals or approximants. As the final segments of a word erode, their features are incorporated into the stem, resulting in unstable vowels and vowel length. This is very similar to the sort of word shortening characteristic of languages of the Cameroun Grassfields, where final syllable erosion is responsible for complex tones (e.g. Mambiloid).

5 Parallel processes in African languages

Do these processes in SE Asian languages help model the emergence of noun classes in African languages? Apart from Afroasiatic, African languages are usually considered to characterized by noun classes, and not to have numeral classifiers at any historical time-depth. These categories are not watertight; languages can display aspects of all these, and erosion of one morphosyntactic category can lead to the partial or complete evolution of another. Krongo, one of the Kadu languages, a branch of Nilo-Saharan, appears to have nominal affixes, although these are partly fossilized and unproductive. They have no semantic assignments, but Krongo has instead adopted or developed a sex-gender system (Reh 1985, 1994; Blench 2006). However, there are some cases where numeral

classifiers appear to be developing, for example Kana, a Cross River language in the Niger Delta of Nigeria (Ikoro 1996). The relatives of Kana are classic nominal affix alternation languages, but Kana seems to have pared down this system and compensated by developing numeral classifiers through grammaticalization (Williamson 1985). More surprising is the case of Mambay, an Adamawa language of northern Cameroun, which still has a quite prominent concordial nominal suffix system, but which is developing prefixed numeral classifiers (Anonby 2011). Examples given by Anonby include the 'collectives' which precede nouns, do not show concord, and have broad semantic correlations.

Nilo-Saharan nominal morphology is marked by extensive affix alternation for number, and yet there is no system of alliterative concord and no systematic association of affixes with semantic categories. However, evidence is emerging that we have been quite wrong in our understanding of Nilo-Saharan, and that its underlying morphology is a system of numeral classifiers. Various studies have noted associations between affixes and semantic themes in different branches. For example, both Stevenson (1991) and Gilley (2013) note the semantic associations of affix pairings in Kadu languages, and Storch (2005) analyses these for Western Nilotic. Carlin (1993) observes that some So number markers have broad semantic themes. But the most striking evidence comes Gumuz, a Nilo-Saharan language of the Ethio-Sudan borderland, whose Mayu dialect has been studied by Ahland (2010). Gumuz has a system of nominal incorporation, in which a series of body part nouns is incorporated into verbs and 'classifies' the object, or more rarely the subject or instrument.

Gumuz predicate classifiers mark semantic fields, typically of shape or texture. These are infixed in 'split verbs' and are copied as demonstratives. The major classifiers are *-Vk"^{w}* 'head', *-Vts* 'body', *-Vc* 'eye/seed', *-k"^{w}ós* 'tooth', and *–ts'ê* 'ear'. Ahland (2010), adapting Mithun (1986), describes a verbal classifier whereby "a noun is incorporated into a verb to categorize an extra predicate argument…usually in S or O function." With this type of verbal classifier, there is frequently a generic-specific relationship between the incorporated NP and the external NP which accompanies it. The significance of this system is that classifiers which develop from grammaticalized body parts are governed by the semantics of nouns. For example, Ahland (2010) notes; "entities that are head-like in shape and/or function or closely associated with such objects" govern the following classes of object "fingers, toes, water, sauce, beer, lotion, soap (in a container), ears of corn, pots, pans, cans". In constructions where the classifier refers to the object of the main verb, the classifier is suffixed to the verb and thus abuts the object noun directly. It can thus become attached to the noun rather than the verb. Western Nilotic, as described by Storch (2005), has a system of nominal suffixes which appear to have semantic correlates. There are

both singular and plural suffixes marking length, roundness, part of etc. And there is some evidence for a system of suffix alternation which has been overwritten by the diffusion of 'imperial' number markers *kV-* and *N-* from other branches of Nilotic. As with Gumuz, grammaticalized body parts are a major source of affixes. The T/K and N/K 'substrata' identified by Bryan in the 1950s and Greenberg's (1981) 'moveable *–k* as a Stage III article' are all reflections of this broader phenomenon.

Exactly how the Niger-Congo and Nilo-Saharan systems are related is still under discussion. Some branches of Niger-Congo, such as Mande, Dogon, Ijoid and Kaalak-Domurik, show no clear traces of any affix system. However, the remaining branches, Atlantic, Kwa, Benue-Congo and Gur-Adamawa have elaborate noun classes and alliterative concord, or traces of such systems where they have demonstrably been lost (e.g. in Volta-Niger and Kru). So this system develops *within* Niger-Congo (and is thus probably not to be reconstructed to proto-Niger-Congo, despite an extensive literature to the contrary). Other phonological evidence, such as labial-velars, ±ATR vowels etc. point to extensive contact between Nilo-Saharan and Niger-Congo, and it is not stretching credibility to propose that the noun classes of Niger-Congo represent a regularized metatypy of Nilo-Saharan affix systems. In other words, something that was implicit in the Nilo-Saharan system of numeral classifiers was borrowed as a system into part of Niger-Congo and then analogized as a rich system of nominal classes[10].

From this perspective, the similarities with SE Asian languages become clearer. Affix renewal is very common in Niger-Congo, where noun class affixes become unproductive and a new affix is added (e.g. Childs 1983). Usually, however, the unproductive affix retains its vowel, or else the conjunction of two consonants results in a complex consonant. For example, Hyamic (Plateau) has developed a complex system of alternating initial clusters due to deletion of –V in the prefix. In SE Asia, the weakening of -V does not result in consonant merger but is rather retained as a syllable sequence. Such a word structure is not typical of Niger-Congo in general, but at least one group of languages does take on this appearance synchronically. Nouns in the northwest Kainji languages (cLela, tHun, u̲t-Main, Gwamhi-Wuri), typically have the structure C.CVCV (the prefix is often transcribed with a schwa to ameliorate the otherwise disquieting

10 The origins of alliterative concord can be debated, but a stimulating suggestion is the proposal of Hoffmann (1968) that demonstratives which copy affixes can explain the movement from prefix to suffix. Extending this idea, if affixes became re-analysed as separable, they can easily become demonstratives or articles, and once copied, establish the principle of alliteration.

appearance) (e.g. Hoffmann 1967). Table 11, shortened from Paterson (2012), shows the noun class prefixes of Ut-Ma'in, some of which now only have consonant prefixes, but which retain strong semantic associations. The symbol ə̄ marks the epenthetic vowel for single-consonant prefixes, which are always phonetically mid-tone in relation to the stem-tone melody.

Table 11. Ut-Ma'in noun class prefixes

Class	sg.	pl.	Object Pronoun	Examples
1u	ū-	ú/wá	ū-mákt	'barren woman'
1ø	ø-	wá	ø-hámə̀t	'visitor'
2	ø-	ɛ́	ø-ná	'oxen, bovines'
3u	ū-	ɔ́	ū-bù	'house'
3ø	ø-	ɔ́	ø-bòʔ	'dream'
4	ə̄s-	sɛ́	ə̄s-bòʔ	'dreams'
5	ə̄r-	dɛ́	ə̄r-kɔ́k	'calabash'
6	ə̄t-	tɔ́	ə̄t-kɔ́k	'calabashes'
6m	ə̄m-	mɔ́	ə̄m-nɔ̀ːg	'oil'
7u	ū-	já	ū-ná	'ox, bovine'
7ø	ø-	já	ø-tʃāmpá	'man'
aug	ā-	á	ā-kɔ́k	'huge calabashes
dim	ī-	ɛ́	ī-kɔ́k	'tiny calabash'

These prefixes can be said to bear tone, although as it appears to be always mid, it is no longer functional (also the case in Himalayan Sino-Tibetan, where these prefixes are uniformly low). Similarly, many Kordofanian[11] languages have C.VCV structures, where the initial C is an alternating prefix. Schadeberg (1981a, 1981b) illustrates this for both the Heiban and Talodi groups. For example, Table 12 shows the reconstructed noun class prefixes of Proto-Heiban (Schadeberg 1981a: 133).

[11] 'Kordofanian' is a creation of Greenberg (1963) based on the assumption that the Niger-Congo languages of the Nuba Mountains must form a genetic group, although this now looks like an over-optimistic view (Blench 2013). However, these languages do share common morphological features, perhaps due to contact.

Table 12. Proto-Heiban noun class prefixes

sg.	pl.
gu-	li-
g-	j-
d-	n-
li-	bu-
ŋ-	ɲ-
d̰-	d-

The difference with northwest Kainji is that the typical Kordofanian stem is -CVC. This suggests (perhaps) loss of C_1 of the stem, subsequent loss of –V from the prefix or assimilation of the resultant VV sequence. The overall parallels to be drawn with African languages are as follows;

a) Nilo-Saharan languages have traces of a former numeral classifier system, still realized in Gumuz, which surfaces synchronically as moveable affixes and which has sporadic semantic associations
b) A subset of Niger-Congo languages have noun class affixes with semantic associations, although these are regularly lost and re-evolve
c) These affixes can be shifted, disappear, fossilize or be incorporated into stems, leading to a process of renewal.
d) These affixes typically conserve their co-associated vowel, because it has a strongly associated segmental tone, whereas SE Asian languages weaken the vowel because there is no underlying tone.
e) However, Niger-Congo languages can occasionally lose the -V- of the affix so comprehensively that the result is a segmental affix consisting only of C, with resultant structural similarities to SE Asia

For these reasons, it seems that outcomes in SE Asia, while diverse, are still less exuberant than in West Africa. Excluding Bantoid and Bantu, the principal branches of Benue-Congo are Kainji, Plateau, Jukunoid, Cross River and possibly Ikann. All of these have evidence (and usually synchronic examples) for a Bantu-like system of alternating nominal prefixes exhibiting alliterative concord on adjectives and other parts of speech. These systems are often preserved in a single branch, with other related branches exhibiting very diverse surface morphology. From this it is reasonable to conclude that nouns in the system of the proto-language had a basic (C)V.CVCV morphology, assumed by De Wolf (1971)

in his now outdated study[12]. If so, the surface forms we see today are a development from this. However, those surface forms are astonishingly diverse. As an example of the complexity within Benue-Congo, Table 13 illustrates the possible outcomes from this type of restructuring.

Table 13. Erosion and restructuring of CV nominal prefixes within Benue-Congo

Strategy	Languages
Complete loss of affixes	Shen
Complete loss of prefixes, vowel neutralization in stem, loss of second stem consonant, number marked by contrastive vowel length	Cara
Complete loss of prefixes, addition of generalized prefix, suffix or clitic plural marker	Lower Cross, Central Jukunoid
Complete loss of prefixes, development of nominal classifiers	Kana
Complete loss of original affixes, new affixes added on the basis of reduplication of first syllable of stem	Hasha
Loss of V of affixes leading to C- prefixes	Northwest Kainji (cLela etc.)
Loss of V of affixes leading to long C- stem initials	Kambari, Upper Cross, Jju cluster
Loss of V of affixes leading to stem initial consonant clusters and consonant alternation	Hyamic
Existing affixes become frozen to the stem and are reprefixed	Cibər [Lopa]
Reduction of all CV- prefixes to V-	Ikann, some Plateau
Reduction of all CV- prefixes to u/i- and rightwards shift into stem, leading to contrastive palatalization and labialization	Many Plateau
Reduction of RV- and NV- prefixes to R-, N-, and rightwards shift into stem, leading to sporadic nasalization and rhotacization	Many Plateau
Prefixes become suffixes	Some Mambiloid
Prefixes become suffixes, which are deleted producing complex stem tones	Mambila
Prefixes become suffixes, lose final –V, C is frozen to the stem and number marking is lost	Dakoid
Prefixes partly become suffixes, resulting in systems of double-affixing	Tivoid

Once the descriptive language is changed, many of these processes are also attested in Austroasiatic and Sino-Tibetan.

[12] De Wolf reached this view by working 'backwards' from Bantu rather than actually surveying Benue-Congo as the primary data source.

6 Conclusions

SE Asian languages have competing systems of noun classification, using both numeral classifiers and noun class terms. Intriguingly, not only word structure but some of the segmental morphology appears to be shared between different language phyla. In languages where these systems are residual, they may have reduced C(V) affixes with weak semantic correlations. These may be reduced classifiers or noun class terms which have become bound to the stem. It is unlikely that this is a result of genetic affiliation and thus it appears that both segmental material and the underlying concept of semantically associated affixes is borrowable. Similar classifier systems occur in some Nilo-Saharan languages, and have apparently developed into non-concordial affixes. Niger-Congo languages have taken the next step, developing strong semantic associations and alliterative concord, probably through demonstrative copying. The challenge is to see whether similar pathways can be reconstructed for noun class affix systems in other language phyla, notably Papuan, Australian and some New World languages.

Acknowledgements

This is an extensively revised version of a paper presented at the meeting *Mainland Southeast Asian Languages: The State of the Art in 2012*, 29 November - 1 December 2012 at the Max Planck Institute for Evolutionary Anthropology, Leipzig. I would like to thank the audience for the discussion, and the editors for subsequent comments. An early version of this paper was presented at the 40th Colloquium on African Languages and Linguistics, Leiden, 29th-31st August, 2011.

References

Adams, Karen Lee. 1989. *Systems of numeral classification in the Mon-Khmer, Nicobarese and Aslian subfamilies of Austroasiatic*. Canberra: Pacific Linguistics.

Ahland, Colleen A. 2010. Noun incorporation and predicate classifiers in Gumuz. *Journal of African Languages and Linguistics* 31(2). 159–203.

Aikhenvald, Alexandra Y. 2000. Classifiers: *A typology of noun categorization devices*. Oxford: Oxford University Press.

Anderson, Gregory D.S. 2004. Advances in Proto-Munda reconstruction. *Mon-Khmer Studies* 34. 159–184.

Anonby, Erik. 2011. *A grammar of Mambay*. Köln: Rüdiger Köppe.

Beckwith, C.I. 1993. Class nouns and classifiers in Thai. In Mark Alves (ed.), *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*, 11-26. Phoenix: Arizona State University.

Bender, M.L. 1996. *The Nilo-Saharan languages: a comparative essay*. Munich: Lincom Europa.

Benedict, P. 1990. Austroasiatic loanwords in Sino-Tibetan. *Mon-Khmer Studies* 18-19: 1-13.

Blench, Roger M. 1995. Is Niger-Congo simply a branch of Nilo-Saharan? In R. Nicolai & F. Rottland (eds.), *Proceedings of the Fifth Nilo-Saharan Linguistics Colloquium, Nice, 1992*, 68-118. Köln: Rudiger Köppe.

Blench, Roger M. 2006. The Kadu languages and their affiliation: between Nilo-Saharan, Niger-Congo and Afro-Asiatic. In Al-Amin Abu-Manga, Leoma Gilley & Anne Storch (eds.), *Insights into Nilo-Saharan Language, History and Culture*, 101-127. Köln: Rüdiger Köppe.

Blench, Roger M. 2013. Splitting up Kordofanian. In Thilo Schadeberg & Roger Blench (eds.), *New research in the languages of the Nuba mountains. Nuba mountains language studies,* 571-586. Köln: Rudiger Köppe.

Blench, Roger M. in press. Reconstructing Austroasiatic prehistory. In P. Sidwell & M. Jenny (eds.), *Handbook of Austroasiatic*. Canberra: Pacific Linguistics.

Blench, Roger M. & M. Post. 2013. Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages. In Tom Owen-Smith & Nathan Hill (eds.), *Selected papers from the 16th Himalayan Languages Symposium, September 2010*, 71-104. Berlin & New York: Mouton de Gruyter.

Blust, R. 1988. *Austronesian root theory: an essay on the limits of morphology*. Amsterdam/Philadelphia: John Benjamins.

Bodman, Nicholas C. 1988. On the place of Lepcha in Sino-Tibetan, a lexical comparison. *Linguistics of the Tibeto-Burman Area* 11(1). 1-26.

Carlin, Eithne. 1993. *The So language. AMO2*. Köln: Universität zu Köln.

Childs, Tucker. 1983. Noun class affix renewal in Southern West Atlantic. In J. Kaye, H. Koopman, D. Sportiche & A. Dugas (eds.), *Current Approaches to African Linguistics* (Volume 2), 17-29. Dordrecht: Foris Publications.

Costello, Nancy. 1966. Affixes in Katu. *Mon-Khmer Studies* 2. 63-86.

Costello, Nancy. 1998. Affixes in Katu of the Lao P.D.R. *Mon-Khmer Studies* 28. 31-42.

DeLancey, Scott 1986. Toward a history of Tai classifier systems. In C.G. Craig (ed.), *Noun classes and categorization,* 437-452. Amsterdam: John Benjamins.

De Wolf, P. 1971. *The noun class system of Proto-Benue-Congo*. The Hague: Mouton.

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology and Austroasiatic. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, 77-80. London: Routledge Curzon.

Diffloth, Gérard. 2008. Shafer's parallels between Khasi and Sino-Tibetan. In S. Morey & M. Post (eds.), *Northeast Indian Linguistics*, 93-104. New Delhi: Cambridge University Press.

Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42(2). 214-261.

Dixon, R.M.W. 1972. *The Dyirbal language of north Queensland*. Cambridge: Cambridge University Press.

Driem, George L. van. 1998. Neolithic correlates of ancient Tibeto-Burman migrations. In R.M. Blench & M. Spriggs (eds.), *Archaeology and Language II*, 67-102. London: Routledge.

Driem, George L. van. 2008. To which language family does Chinese belong, or what's in a name, In Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros & Marie Lin

(eds.), *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, 219-253. London & New York: Routledge.

Enfield, N.J. 2007. *A grammar of Lao*. Berlin & New York: Mouton de Gruyter.

Forrest, R.A.D. 1962. The Linguistic Position of Rong (Lepcha). *Journal of the American Oriental Society* 82(3). 331-335.

Gilley, Leoma. 2013. Katcha noun morphology. In Thilo Schadeberg & Roger Blench (eds*.), New research in the languages of the Nuba mountains. Nuba mountains language studies*, 501-521. Köln: Rudiger Köppe.

Green, Ian. 1995. The Death of 'Prefixing': contact induced typological change in Northern Australia. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 21(1). 414-425.

Greenberg, Joseph H. 1963. *Languages of Africa* (Indiana University Research Centre in Anthropology, Folklore, and Linguistics, Publication 25). The Hague: Mouton & Co.

Greenberg, Joseph H. 1977. Niger-Congo noun class markers: prefixes, suffixes, both or neither. *Studies in African Linguistics* (Supplement 7), 97-104.

Greenberg, Joseph H. 1981. Nilo-Saharan Moveable- k as a Stage III Article (with a Penutian Typological Parallel). *Journal of African Languages and Linguistics* 3(2). 105-112.

Haas, M.R. (ed.). 1964. *Thai-English student's dictionary*. Stanford: Stanford University Press.

Hammarström, Harald. 2013. Noun class parallels in Kordofanian and Niger-Congo: evidence of genealogical inheritance? In Thilo Schadeberg & Roger Blench (eds.), New research in the languages of the Nuba mountains. *Nuba mountains language studies*, 549-570. Köln: Rudiger Köppe.

Handel, Zev. 2008. What is Sino-Tibetan? Snapshot of a Field and a Language Family in Flux. *Language and Linguistics Compass* 2/3. 422–441.

Harvey, Mark & Nicholas Reid. 1997. *Nominal Classification in Aboriginal Australia*. Philadelphia: John Benjamins.

Harvey, Mark, Rachel Nordlinger & Ian Green. 2006. From Prefixes to Suffixes: Typological Change in Northern Australia. *Diachronica* 23 (2). 289-311.

Hoffmann, Carl 1967. An outline of the Dakarkari noun class system and the relation between prefix and suffix noun-class systems. In *La Classification nominale dans les langues Négro-africaines*. ed. G. Manessy. 237-259. Paris: CNRS.

Ikoro, S.M. 1996. *The Kana language*. Leiden: CNWS.

Jones, Robert B. 1970. Classifier constructions in Southeast Asia. *Journal of the American Oriental Society* 90(1). 1-12.

Kirton, Jean F. 1988. Yanyuwa, a dying language. In Michael J. Ray (ed.), *Aboriginal language use in the Northern Territory: 5 reports* (Work Papers of the Summer Institute of Linguistics), 1-18. Darwin: Summer Institute of Linguistics.

Matisoff, J.A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman, (ed.), *Consonant Types and Tone* (Southern California Occasional Papers in Linguistics, No. 1.), 71-95. Los Angeles: UCLA.

Matisoff, J.A. 2003. *Handbook of proto-Tibeto-Burman*. Berkeley: University of California Press.

Matisoff, J.A. 2008. *The Tibeto-Burman reproductive system: toward an etymological thesaurus*. Berkeley: University of California Press.

Mithun, Marianne. 1986. The convergence of noun classification systems. In Colette Craig (ed.), *Noun classes and categorization*, 379-397. Philadelphia: John Benjamins.

Paterson, Rebecca. 2012. The semantics of Ut-Ma'in noun classes. In R.M. Blench & S. McGill (eds.), *Advances in minority language research in Nigeria* (Volume I), 239-272. Köln: Rudiger Köppe.

Post, M.W. 2007. *A grammar of Galo.* Bundoora: La Trobe University dissertation.

Ratliff, Martha 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics.

Reh, M. 1985. *Die Krongo-Sprache (nìino mó-dì): Beschreibung, Texte, Wörterverzeichnis*. (Kölner Beiträge zur Afrikanistik, 12). Berlin: Dietrich Reimer.

Reh, M. 1994. A grammatical sketch of Deiga. *Afrika und Übersee* 77. 197-261.

Rispoli, Fiorella 2008. The incised and impressed pottery of Mainland Southeast Asia: following the paths of Neolithization. *East and West*, 57:235-304.

Sagart, Laurent. 2011. The Austroasiatics: East to West or West to East? In N. Enfield (ed.), *Dynamics of Human Diversity in Mainland Southeast Asia*. 345-359. Canberra: Pacific Linguistics.

Schadeberg, Thilo C. 1981a. *A Survey of Kordofanian. Volume 1: The Heiban Group*. Hamburg: Helmut Buske.

Schadeberg, Thilo C. 1981b. *A Survey of Kordofanian. Volume 2: The Talodi Group*. Hamburg: Helmut Buske.

Shafer, Robert. 1952. Etudes sur l'austroasien. *Bulletin de la Société de Linguistique de Paris* 48. 111-158.

Shorto, Harry L. 2006. A Mon-Khmer comparative dictionary. (Edited by Paul Sidwell, Doug Cooper & Christian Bauer), Pacific Linguistics *579*. Canberra: Australian National University.

Sidwell, Paul & Roger M. Blench. 2011. The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In N. Enfield (ed.), *Dynamics of Human Diversity in Mainland Southeast Asia*, 317-345. Canberra: Pacific Linguistics.

Smith, Kenneth D. 1975. The velar animal prefix in Vietnam languages. *Linguistics of the Tibeto-Burman Area* 2. 1-18.

Stevenson, Roland C. 1991. Relationship of Kadugli-Krongo to Nilo-Saharan: Morphology and lexis. In M L. Bender (ed.), *Proceedings of the Fourth Nilo-Saharan Conference. Bayreuth, 1989*, 347-369. Hamburg: Buske.

Storch, Anne. 2005. *The Noun Morphology of Western Nilotic*. Köln: Köppe.

Thomas, Dorothy. 1969. Chrau affixes. *Mon-Khmer Studies* 3. 90-107.

Williamson, K. 1985. How to become a Kwa language. In: Makkai A., Melby AK (eds.), *Linguistics and philosophy: Essays in honor of Rulon S. Wells*. 427–443. Amsterdam: Benjamins.