9

SINO-TIBETAN–AUSTRONESIAN

An updated and improved argument

Laurent Sagart

In several articles (Sagart 1990, 1993, 1994) I have argued from sound correspondences, shared vocabulary and shared morphology that Chinese and AN are genetically related within a macrophylum which I called 'Sino-Austronesian'. The accuracy of the Chinese and AN material used in the comparisons has not been faulted, and neither have the sound correspondences underlying the comparisons. Criticism has concentrated on three points: first, paucity of comparisons involving basic vocabulary; second, the position of TB outside Sino-Austronesian; third, sound correspondences that leave out non-final syllables of AN words. I present here an updated and improved argument which answers these issues.

In this chapter, Old Chinese (OC, *c*.2,500 BP) is reconstructed according to the system presented in Sagart (1999), a modification of Baxter (1992). PAN reconstructions are drawn from the literature, a few are mine. I adhere to the view that PAN was spoken in Taiwan from around 5,500 BP on, on archaeological grounds. The first diversification of PAN took place on the West coast of Taiwan. Soon one group of West coast speakers moved to the East coast where a second diversification occurred, resulting in a dialect linkage (ECL). Later on, perhaps around 4,500 BP, a group of ECL speakers left Taiwan to settle the northern Philippines. Their language, PMP, is ancestral to all conventionally recognised AN languages outside of Taiwan. Another group of early AN speakers left Taiwan to settle coastal areas in Guangdong or Guangxi, where their language, which I call AAK was to a great extent relexified by a local language, later to become Proto-Kadai (more on this in Chapter 10, this volume). The subgrouping of AN is therefore as in Figure 9.1 (based on Ho 1998 with modifications).

For PMP innovations, see Blust (1977). The following innovations are shared uniquely by PMP and ECL languages:

- PAN *C ⇒ *t (Siraya, Bunun, Amis, Kavalan, Basay-Trobiawan, PMP: Ferrell 1969)
- PAN *N ⇒ *n (Bunun, Kavalan, Basay-Trobiawan, Kanakanabu,[1] PMP: Ho 1998)
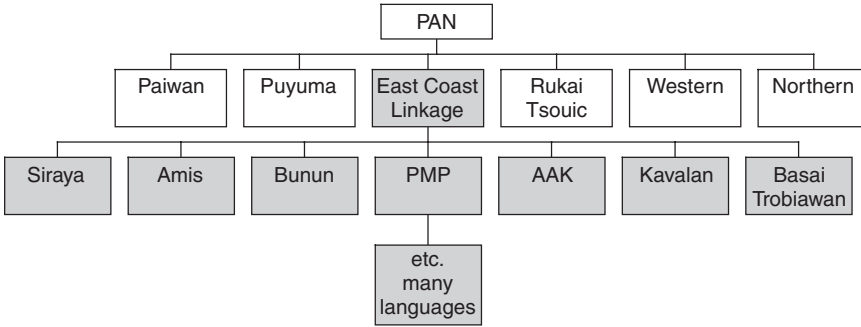
Figure 9.1 Higher AN subgrouping.

- PAN *qayam replaced by *manuk in the meaning 'bird' (Basay-Trobiawan, PMP)
- *pang-V* instrumental construction (Amis, PMP: Starosta 2001)

From Figure 9.1 one can see that MP material is not essential in reconstructing PAN forms; these can be based on Formosan exclusively. In my earlier work, all AN reconstructions were drawn from the works of Blust or Dempwolff and necessarily included MP material. The practice of reconstructing PAN forms based on evidence from Formosa only was initiated by Blust himself (1999). Reconstructing PAN in this way adds a significant number of basic vocabulary comparisons between PAN and Chinese. Most of these have TB comparanda, as shown in Table 9.1. Remarkably, TB and PAN agree against Chinese in certain matters of phonology (Table 9.5). TB morphology, better preserved than Chinese morphology, also has many points of agreement with PAN. Since TB and Chinese do have more basic vocabulary in common than either does with PAN and since some features shared by TB and Chinese against PAN appear to be innovations (Sagart forthcoming), I recognise here (*contra* Sagart 1990) that ST is a valid construct and claim that it, as a whole, is genetically related to PAN. I refer to the resulting macrophylum as STAN. Available reconstructions of TB (Benedict 1972) and ST (Coblin 1986; Gong 1995; Peiros and Starostin 1995) differ widely, due to continuing uncertainty on subgrouping, sound correspondences and the amount of contact between Chinese and the rest of ST. For this reason, Old Chinese will serve here as the main representative of ST.

Linguistic evidence

In the following three sections I present evidence of basic and cultural vocabulary shared with sound correspondences, and of shared morphological processes.

Table 9.1 Sixty-one basic vocabulary comparisons between AN, Chinese and TB

	PAN or PECL	OC	TB
1 Body hair	gumuN	眉 ᵇmu[r] (eyebrow)	B. mul (Moshang kemul)
2 Bone	kukut	骨 ᵃkut	
3 Brain	punuq	腦 ᵃnuʔ	B. (s-)nuk
4 Elbow	siku(H₂)	肘 ᵇt-r-kuʔ	Gyarong tkru
5 Female breast	nunuH₁	乳 ᵇnoʔ	B. nuw
6 Foot	kakay		B. kriy
7 Head	quluH₁	首 ᵇhluʔ	Lushai lu
8 Palm of hand	dapa	扶 ᵇpa	B. pa
9 Pus	nanaq		Tib. rnag
10 Mother	ina(-q)	女 ᵇnraʔ (woman)	B. m-na
11 Egg	qiCeluR	卵 ᵃCə-lo[r]ʔ	B. twiy < t-l-ʔ
12 Horn, antler	(q)uRung	角 ᵃk-rok	B. rung = rwang
13 Leech	Limatek	蛭 ᵇtik	
14 Snake	bulay	蛇 ᵇm-la[r]	P-Loloish lay₁/₂ 'python'
15 Worm	[ ]ulej	蚓 ᵇlinʔF?	
16 Cloud, cloudy	-qem	陰 ᵇʔïm	Bur. um'
17 Earth	-taq	土 ᵃthaʔ	Tib. ⁿdag pa 'mud'
18 Moon	qiNaS		B. s-la
19 Salt	siRaH₁	鹵 ᵃraʔ S!	B. la I!
20 Sunlight	siNaŋ	陽 ᵇlang	Bur. lang 'to be light'
21 Water	daNum	潘 ᵇt-hlïmʔ (liquid,	
22 Wind	bali	juice)	B. g-liy
23 Cave, hole	b[e]lung	洞 ᵃlong S!	Kachin kin₃₁ luŋ₃₃
24 Year	kawaS	歲 ᵇs-hwat-s S!	
25 Carry	baba		B. ba
26 Chew	paqpaq	哺 ᵃm-paʔ-s	
27 Close, shut	kupit	閉 ᵃpit	
28 Come, go	duwa	于 ᵇwa	B. s-wa
29 Cut off, short	[p,b]utul	斷 ᵃto[r,n]ʔ;	Lepcha tultul
		ᵃto [r,n]	
30 Dig	-kut	堀 ᵇkhut,	Kachin kot
		掘 ᵇm-kut	
31 Drown, disappear	Nemes	滅 ᵇmet	B. mit 'extinguish' (fire)
32 Fall	-luR	墮 ᵃlo[r]ʔ	
33 Flow > water, river	qaluR 'to flow'	水 ᵇhlu[r]ʔ (water, river)	B. twiy < t-l-, lwiy 'to flow'
		冰 ᵇt-lu[r]ʔ(water)	
34 Follow	duNuR	隨 ᵇs-lo[r]	
35 Grasp, embrace	-kep	夾 ᵃm-kep, ᵃs-kep, ᵃk-r-ep	
36 Hold sth in fist/mouth	gemgem (in fist)	含 ᵃgïm (in mouth)	B. gam 'put into mouth'
37 Lick	dilaq	舐 ᵇm-leʔ	B. m-lyak
38 Meet	Cebung	逢 ᵇbung S!	PS pung 'assemble'?

Table 9.1 Continued

	PAN or PECL	OC	TB
39 Open	-kaq	啟 ᵃkheʔ	Kachin khaʔ < -k 'parted, open'
40 Put together	pulung	同 ᵃlong	
41 Ruin, damage	r[i]bas	敝 ᵇbet-s	
42 Scrape I	kuSkuS	括 ᵃk-r-ot	Tib. r-ko, Gyarong ka rkos Kuki-Naga d-kew
43 Scrape II	ku[Ct]ku[Ct]	括 ᵃk-r-ot	B. kut
44 Sink	-neb		B. nup
45 Sleep	-zem	寢 ᵇtshimʔ	Tib. gzim, Dhimal dʒim
46 Speak, say	kawaS	話 ᵃm-kw-r-at-s; 曰 ᵇwat S!	Tib. s-go
47 Think	nemnem	念 ᵃnim-s	Tib. s-nyam-pa 'to think'
48 Vomit, spit	utaq	吐 ᵃthaʔ	B. (m-)tuk V!
49 Wash	basuq	溲 ᵇs(r)uʔ	Lushai shuk, Luoba ɕuk
50 Wrap around (belt)	-kes	繫 ᵃket	
51 Bent, crooked	-kuk	局 ᵇN-k(r)ok	B. kuk
52 Broad	-bang	旁 ᵃbang	Boro go2 bang1 'wide, many'
53 Curled, bent	-kul	卷 ᵇN-k(h)ro[r,n]	PS kuar
54 Dark	-lem	黯 ᵃlïmʔ, ᵃhlïmʔ	
55 Far	ma-dawiN	遠 ᵇwa[r,n]ʔ V!	B. wiy
56 High, tall	-kaw	高 ᵃkaw	Bur. kaw: (heavy tone) 'rise up, swell, bulge'
57 Hot	qa(i)nget	熱 ᵇnget	
58 Old, grownup	-daŋ	丈 ᵇdrangʔ	
59 Sharp	Cazem	[GSR 660a] ᵇtsïm	
60 Thick	-tul	敦 ᵃtu[r,n]	PS tu:r
61 This	di	時 ᵇdï Iʔ	Tib. ⁿdi 'this'

Note

I! V! F! T! S! irregular *I*nitial, *V*owel, *F*inal, *T*one, *S*yllable type.

Shared vocabulary

I present 61 lexical comparisons involving basic vocabulary items (Table 9.1), and 14 comparisons involving cultural items (Table 9.2). The Chinese and AN members of these comparisons conform to the sound correspondences presented in the next section.

Basic vocabulary comparisons

Ten among these comparisons between Chinese and AN: bone, breast, head, egg, horn, earth, salt, speak, hot, this, are on Swadesh's 100-word list, and six: bone, egg, horn, salt, year, this, on Yakhontov's highly basic 35-word list. It is significant that

Table 9.2 Fourteen cultural vocabulary comparisons between AN, Chinese and TB

1 Setaria	beCeng	稷 ᵇtsïk	
2 Panicum sp.	Numay	麻＋黍 ᵃmaj	
3 Husked rice	beRas	糲 ᵇmə-rat-s	Tib. ᵐbras 'rice' < m-ras
4 Paddy	Sumay 'rice as food'	米 ᵃmijʔ 'grain of cereal'	B-G may 'rice; paddy'
5 Chicken	kuka	雞 ᵃke	B. ka 'kind of fowl'
6 Cage, enclosure	kurung	籠 ᵃkə-rong	B. kru:ŋ
7 Net	aray	羅 ᵃraj	
8 Broom	CapuH₁	帚 ᵇt-puʔ	
9 Stopper, plug	seŋseŋ	塞 ᵃsïk	
10 To bury, tomb	-buN 'to bury'	墳 ᵃbu[r] 'tomb, tumulus'	
11 Loincloth, robe	sabuk	服 ᵇbuk	Tib. ᵐ bog 'k. o. garment'
12 To plait, braid	-pid	編 ᵃpin(ʔ) F?	B. byar~pyar
13 To shoot	panaq	弩 ᵃnaʔ (crossbow)	
14 To hunt	qaNup	獵 ᵇCə-lap	Chepang krup

the percentage of hits on the more basic list (Yakhontov's) is higher than on the less basic list (Swadesh's): 17 per cent against 10 per cent. I do not consider these figures to be final. Missing are the personal pronouns and numerals, which have undergone far-reaching paradigmatic changes (analogy, politeness shifts involving deictics). They will be discussed elsewhere.

Cultural vocabulary comparisons

One notes the presence of terms for agriculture, animal husbandry, hunting, house utensils and the absence of terms for metal. This points to a Neolithic, pre-metal, ancestral culture.

Sound correspondences

Due to canonical reduction of the initial syllable(s) of ancestral polysyllables, sound correspondences relate the last syllable of PAN words with Chinese and TB monosyllabic word stems. In addition, Old Chinese syllable type (A or B) correlates with the nature of the initial of AN penultimate syllables, as detailed in Table 9.7. Tables 9.3 and 9.4 present the correspondences of syllable-initial and final consonants, and Table 9.6 presents the vowel correspondences.

One can see from Table 9.4 that OC -ʔ has two corresponding sounds among the final consonants of PAN: -q and -H. This distinction, lost by Chinese, is actually maintained by TB, which has -k and zero corresponding to PAN -q and -H respectively, as shown in Table 9.5.

Table 9.3 Correspondences of syllable-initial consonants (PAN final syllable initial: Chinese root initial: TB)

PAN	OC	TB	Examples
p-	p(h)-	p-	Palm of hand, chew, plait, close, broom
t-	t(h)-	t-	Leech, earth, vomit, thick, short
k-	k(h)-	k-	Elbow, bone, chicken, dog, high, curled, crooked, dig, grasp, wrap around, scrape I, scrape II, open
q-	ʔ-	0-	Cloud(y)
b-	b-	(p-)	Carry, broad, loincloth, meet, tomb, ruin
d-	d-	d-	Old, this
g-	g-	g-	Hold in fist or mouth
m-	(h)m-	m-	Body hair, drown
n-	n-	n-	Brain, breast, pus, mother, think, shoot, sink
ŋ-	ŋ-	ŋ-	Hot
N-	(h)l-	l-	Hunt, water, follow, sunlight, moon
l-	(h)l-	l-	Head, snake, head, flow, lick, put together, fall, wind, cave, worm
R-	r-	r-	Horn, salt, husked rice
w-	(h)w-	w- (Tib. g-)	Year, far, say, come/go
s-	s-	?	Wash, stopper
z-	ts-	?	Sharp, sleep, wink

Table 9.4 Correspondences of syllable-final consonants

PAN	OC	TB	Examples
-0	-0	-0	Palm of hand, chicken, carry, this, come/go
-k	-k	-k	Leech, crooked, loincloth
-t	-t	-t	Bone, hot, dig, close
-p	-p	?	Hunt, grasp
-ng	-ng	-ng	Put together, broad, cage, meet, sunlight, cave, old
-ng	-k	-ng	Horn, stopper, Setaria
-m	-m	-m/-p	Water, think, hold in fist or mouth, dark, cloud
$-H_{1,2}$	-ʔ	-0	Head, female breast, elbow, salt, broom
-q	-ʔ	-k	Brain, pus, earth, lick, vomit, chew, shoot, wash, open
-l	[-r]	-r	Curled, thick
-R	[-r]	-y	Dog, egg, flame, flow, fall, follow
-y	-j	-y	Snake, net, Panicum sp.
-S	-t	-0	Say, year, scrape I, moon
-s	-t	-s (/a_) -t (else)	Husked rice, drown, wrap around, ruin
-N	[-r]	-y~-l	Body hair, far, tomb

Table 9.5 Preservation by PAN and TB of a contrast in consonant endings lost by Chinese

	PAN	OC	TB
Brain	punuq	腦 ᵃnuʔ	(s-)nuk
Pus	nanaq		Tib. rnag
Lick	dilaq	舐 ᵇm-leʔ	m-lyak
Open	-kaq	啟 ᵃkheʔ	Jingpo khaʔ < -k 'parted, open'
Wash	basuq	溲 ᵇs(r)uʔ	Lushai shuk, Luoba çuk
Female breast	nunuH₁	乳 ᵇnoʔ	nuw
Head	quluH₁	首 ᵇhluʔ	Lushai lu
Salt	siRaH₁	鹵 ᵃraʔ S!	la I!

Table 9.6 Vowel correspondences (PAN last vowel : Chinese root vowel)

STAN	PAN : Chinese	Examples
u (before labials)	-u- : -ï-	Water
u (elsewhere)	-u- : -u-	Head, brain, elbow, bone, body hair, dog, flow, thick, dig, meet, tomb
o (before labials)	-u- : -a-	Hunt
o (elsewhere)	-u- : -o-	Breast, egg, horn, fall, put together, curl, crooked, cut off, cage, cave
a (before y)	-a- : -i-	Grain
a (elsewhere)	-a- : -a-	Palm, mother, snake, year, salt, earth, vomit, shoot, speak, broad
æ	-a- : -e-	Chicken, lick, ruin, open
e (after grave cons.)	-e- : -e-	Grasp, wrap around, drown, hot
e (elsewhere)	-e- : -i-	Think, leech, worm, sleep
i (open syll.)	-i- : -ï-	This
i (closed syll.)	-i- : -i-	Plait, close
ə	-e- : -ï-	Dark, sink, hold in fist, stopper, sharp

Table 9.7 Correspondences of Chinese syllable type and manner of articulation in PAN penultimate syllable initial consonant

PAN penultimate syllable initial	Chinese syllable type	Examples
Voiceless stop (except q), or zero	ᵃ(non-division 3)	Bone, brain, horn, close, put together, spit
Other initials (including q)	ᵇ(division 3)	Elbow, head, palm, leech, snake, water, drown

Old Chinese had two contrastive syllable types: A and B, of uncertain phonetic interpretation. In my notation these are marked by superscript 'a' and 'b' preceding the reconstruction. These syllable types exhibit a statistical correlation with the nature of the penultimate syllable initial of PAN words: if the penultimate

syllable of a PAN word begins with a voiceless stop (excluding q) or zero, then type A is predicted in Chinese. If the penultimate syllable began in another sound (including q), type B is predicted. With PAN monosyllables and roots (always monosyllabic), including reduplicated monosyllables/roots, no prediction can be made. With PAN penultimate initial *C, no prediction can be made either (perhaps PAN *C results from the merger of two PSTAN sounds, a voiceless stop/affricate, and another kind of sound).

Shared morphology

Several morphological processes are shared by AN and ST, including three of the main verbal 'focus' constructions which form the backbone of AN verbal morphology:

The Proto-Austronesian nominaliser and Goal Focus marker -ən and the TB nominalising suffix -n

A process deriving nouns from verbs by means of a suffix -n or -ən exists in AN and in ST:

AN	Atayal	niq 'to eat' : niq-un 'eaten thing'
	Paiwan	alap 'to take' : alap-en 'object being taken'
	Amis	aʃik 'to sweep' : aaʃik-en 'place to sweep'

ST	Tibetan	za-ba 'to eat' : za-n 'food, fodder, pap, porridge'
		skyi-ba 'to borrow' : skyi-n-pa 'borrowed thing, loan'
		rdzu-ba 'to delude, to falsify' : rdzu-n-pa 'falsehood, fiction, lie'
	Lepcha	hru 'to be hot' : ă-hru-n 'heat'
		bu 'to carry' : ă-bu-n 'vehicle'

In AN, according to the theory known as SPQR, this nominalising process is the source of the GF construction, where -ən is the GF marker. Consider the following verb-initial Atayal sentence in GF (from Egerod 1980), where -un is the GF marker:

baq-un	makuʔ	tuqii
know-GF	my$_{GEN}$	way
I know the way		

Under SPQR, the verb-initial, GF parsing of this sentence is a reinterpretation of an earlier cleft sentence meaning 'my known thing (baq-un makuʔ) [is] the way (tuqii)'. Comparison with TB provides the STAN source of this AN nominalising suffix: that is precisely the -n nominaliser found in TB languages. The reinterpretation of the NP 'my known thing' as a GF verb meaning 'I know' occurred after verb-initial word order became generalised in pre-PAN.

The Proto-Austronesian Actor Focus prefix and infix m-/-m- and the ST intransitive prefix m-

The AN AF marker is a nasal affix m- (prefix) or -m- (infix) depending on language and root shape. In Starosta's ergative interpretation of AN grammar (Starosta 1991, 1994), assumed here, all verbs in AF are intransitive, with m-/-m- deriving intransitive verbs from transitive ones. Contrast the following Tagalog sentences:

s-in-agot	ng-istudyante	ang-propesor
OF-answer	GEN-student	NOM-professor

s-um-agot	sa-propesor	ang-istudyante
AF-answer	LOC-professor	NOM-student

Both sentences mean 'The student answered the professor'. The first sentence is in OF, marked on the verb with infixed -in-. It is a typical ergative construction, with the patient marked as nominative, and the agent marked as genitive. Starosta regards verbs in OF as transitive. The second sentence is in AF, marked with infixed -um-. In Starosta's analysis this is really an antipassive construction, with the patient marked in an oblique case form (locative). Infixed -um- marks the verb as intransitive, even though it occurs with two arguments.

   PST had a prefix m- which turned transitive verbs into intransitives. Wolfenden (1929: 25–26, 76) characterised it as 'inactive' and 'intransitive'. Examples (Wolfenden 1929: 30 for Tibetan and Kachin; Bhattacharya 1977: 184, 328–330 for Boro):

Tibetan	m-nam-ba	'to smell (intr.), stink'
Kachin	ma-nam	'to smell' (intr.)
	ma-ni	'to laugh'
Boro	$m\bar{o}_2$-nam_1	'to spread smell'
	mi_2-ni_2 'to laugh'	

This prefix, illustrated before nasals in the preceding examples, reduced to prenasalisation preceding voiceless stops. In Gyarong, a TB language from Sichuan, prenasalisation has secondarily voiced the following stop. Examples (Lin Xiangrong 1993: 193):

Gyarong	ka-tʃop 'to set fire to' : kə-ndʒop 'to catch fire'[2]
	kɐ-p'ɐk 'to split open' : kə-mbɐk 'to be rent'
	kɐ-tɕ'op 'to break' : kə-ndʑop 'broken'
	kɐ-klɐk 'to wipe off' : kə-ŋglɐk 'to fall'

In Tibetan, Kiranti, Bahing, Vayu, Bodo-Garo, prenasalisation has further been lost and only secondary voicing of the root initial marks the intransitive member (Benedict 1972: 124 for examples and discussion[3]). MC (mid-first millennium CE)

likewise had contrasting pairs of transitive verbs with voiceless stop initials vs intransitive verbs with voiced stop initials:

別 pjet (III) 'to separate, distinguish': 別 bjet (III) 'to take leave'
箸 trjak 'to put something in a certain place': 箸 drjak 'to occupy a fixed position'
斷 twanH 'to cut, sever': 斷 dwanH 'broken off, cut off from; to cease'
折 tsyet 'to break, to bend' (trans.): 折 dzyet 'to bend' (intrans.)

I have shown (Sagart 1994, 1999, 2003) that intransitive voicing in MC verbs reflects OC prenasalisation, as shown in particular by early loans to Miao-Yao.

The Proto-Austronesian Instrumental/Beneficiary Focus prefix Si- and the valency-increasing s- in Sino-Tibetan

A prefix PAN Si-: OC s-: TB s- allows a verb to take a NP with real-world roles such as causer, beneficiary, instrument, etc. and treat it formally as its patient (that is, as its grammatical object in Chinese, an accusative language, and as its subject in ergative AN). The AN Si-V construction is known as 'Instrument focus' (also 'Beneficiary Focus') but its semantics are complex. Huang (1991: 45) characterises the Si-construction in Atayal as 'circumstantial voice' and states that one characteristic of circumstantial voice is 'increased transitivity'. As an illustration, I cite here examples with a transitive/causative character, because the semantic difference between prefixed and non-prefixed forms can be apprehended directly through simple lexical glosses, even though this is an oversimplification of the functions of this prefix.

Atayal   m-ŋuŋuʔ 'to be afraid': s-ŋuŋuʔ 'to frighten'
Paiwan   k/m/avuL 'to beg': si-kavuL 'cause someone to beg'
Bunun   daŋadx 'to stop' (intr.): is-daŋadx 'to stop' (trans.)
Old Chinese   順 *^{b}m-lun-s 'to be pliant, obedient': 馴 *^{b}s-lun 'to tame'
Tibetan   Nbar 'to burn, catch fire, be ignited': s-bar-pa 'to light, to kindle, to inflame'
                  m-nam-pa 'to smell, stink' (intransitive): s-nam-pa 'to smell' (transitive)
Gyarong   rong 'to see': s-rong 'to show'
Boro   gi 'to be afraid of, fear': si-gi 'to frighten'
Proto-Loloish[4] (C)-no$_2$ 'to awake': sə-no$_2$ 'to awaken' (tr.)

-ar- distributed action; distributed object

This infix was inserted between the root initial and the first vowel of a stem. Attached to verbs of action it indicated that the action was distributed in time (occurring over several discrete occasions), or in space (involving several agents/patients/locations); attached to nouns it indicated a referent distributed in

space, that is having double or multiple structure. The reflex of this infix in the AN languages is -ar-, marking verbs of distributed action and nouns of distributed object, including names of paired or multiple body parts. Infixation is often, but not always, in the first of two reduplicated syllables:

Paiwan	k-ar-akim 'to search everywhere' (kim 'search')
	k-ar-apkap-an 'sole of foot'
Puyuma	D-ar-ukap 'palm of hand'
Bunun	d-al-apa 'sole of foot' (PAN *dapa 'palm of hand')
Amis	p-ar-okpok 'to gallop'
	t-ar-odo' 'fingers, toes'
	k-ar-ot 'harrow'
Tagalog	d-al-akdak 'sowing of rice seeds or seedlings for transplanting'
	(dakdak 'driving in of sharp end of stakes into soil')
	k-al-aykay 'rake'
Malay	ketap 'to bite teeth' : k-er-etap 'to bite teeth repeatedly'

Other AN languages show an infix -aR- with similar functions (not illustrated here). According to the sound correspondences presented above, both -r- and -R- correspond to OC -r-. Although no living TB language has -r- infixation as a living process, paired nouns and verbs with what appears to be an infix -r- show up here and there, with similar semantics as in Chinese:

Burm.	pok 'a drop (of liquid)' : prok 'speckled, spotted'
	pwak 'to boil up and break, as boiling liquid' : prwak 'ibid.'
	khwe$_2$ 'curve, coil' : khrwe$_2$- 'to surround, attend'
Kachin	hpun 'of pimples, to appear on the body' : hprun 'pimples, on the body; to appear on the body, of pimples'
Chepang -r-	pop, prop 'the lungs'
	brok 'be partly white, grey, streaked' (of hair); compare TB bok 'white'.

I first identified the Chinese -r- distributed action/object infix from minimal pairs in Old Chinese (Sagart 1993). Later on, I described some infixed pairs in modern dialects where the infix showed up as the regular modern reflex -l-, preceded either with a schwa or with a full or partial copy of the syllable's rime (Sagart 1994, 2001). Here are some examples of infixed nouns and verbs from Yimeng, a Jin dialect of Inner Mongolia, where the infixed string is -ə^{ʔ}l-[5] (Li 1991):

p-ə^{ʔ}l-ai$_3$ 'to swing, oscillate'
p-ə^{ʔ}l-ən$_1$ 'to run on all sides'
xu-ə^{ʔ}l-a$_4$ 'to scribble'
t-ə^{ʔ}l-əu$_1$ 'cluster(s) of fruit hanging from branches'
khu-ə^{ʔ}l-u$_3$ 'wheel(s) of a car'

Reduction to monosyllables and maintenance of prefixation and infixation

How did PSTAN prefixes and infixes survive the loss of non-final syllables, to which they were attached, in the evolution to Chinese? The answer was provided by Starosta (1995). Starosta argued that PSTAN had both monosyllables and polysyllables: only polysyllabic words were affected by the loss of initial syllables and attached affixes: monosyllables could then act as a refuge for prefixes and infixes. PSTAN monosyllables survive in PAN as roots and reduplicative disyllables. Judging from the high number of verbs among PAN roots, and from the high number of PAN roots in the lexical comparisons for verbs presented above (Table 9.1), it appears likely that many PSTAN verbs were monosyllabic. PSTAN verbal morphology, then, could easily continue in ST languages after canonical reduction had started operating.

Archaeology and agricultural origins

What historical reality lies behind the proposed linguistic relationship? Both in the modern cultures and archaeologically, evidence of a substantial cultural unity between the AN peoples of Taiwan and the ST peoples can be discerned. The principal is an agriculture based on two millets: *Setaria italica* and *Panicum miliaceum*, with rice as a third cereal. In northern China, the millets appear archaeologically in different sites of the Cishan-Peiligang culture between 8,500 and 7,500 BP (Lu, Chapter 3, this volume), and continue to be present down to historical times. The earliest Chinese inscriptions and texts (late second to first millennium BCE) show millets to be the main crops of the Shang and Zhou states. The Zhou rulers thought themselves descended from a mythical ancestor, Hou Ji 侯稷 ('Lord *Setaria*'). Millets played a major role in religious rituals. Domesticated *Setaria* also occurs in the Karuo culture of Eastern Tibet, *c*.5,555–4,750 BP (Fu Daxiong 2001: 66) and in Changguogou in the mid-Yalu Tsangpo River Valley, *c*.3,370 BP (Fu Daxiong 2001). Many TB peoples cultivate millets to this day. In the lower Huang He Valley, downriver from the Peiligang culture, the Beixin and Dawenkou cultures of Henan, south Shandong and northern Jiangsu (from *c*.7,000 BP) were also millet-based (Chang 1986). Chang regards them as a probable eastward expansion from the mid-Huang He Valley communities of millet farmers. Millets, regarded by the AN peoples of Taiwan as sacred, had long been missing from the archaeological record in Taiwan, generating speculations that these cereals could have been acquired at a relatively recent date, even though one millet-related term: *beCeng '*Setaria*' can be securely reconstructed to PAN. The recent discovery in southwestern Taiwan of thousands of carbonised grains of millet (Tsang, Chapter 4, this volume), in conjunction with rice grains, in a TPK cultural context dated to 4,500 BP, has laid these speculations to rest. TPK, the oldest ceramic culture in Taiwan, is generally identified with the PAN speech community. The antiquity of millets in AN culture cannot

now be doubted. The PAN speakers were farmers, and their main crops were rice and millet. In contrast, the scarcity of the millets, not just archaeologically, but ethnologically, in South China, is striking. It is not clear how the early Austronesians could have possessed millet if their immediate ancestors were a southern Chinese people.

Not only were rice and millets grown by the early TBs, Chinese and Austronesians, the very names of these cereals are shared, with the same sound correspondences as the rest of the shared vocabulary (Table 9.2). My current interpretation of the facts is as follows. Between 8,500 and 7,500 BP, farming communities with domesticated *Setaria*, *Panicum* and rice began to appear in the mid-Huang He Valley, whether as a northern extension of the Yangzi rice Neolithic (Bellwood, Chapter 1, this volume), or as an independent transition to the Neolithic (Lu, Chapter 3, this volume) is still uncertain. I call PSTAN the language spoken by these early farmers. Subsequent population growth resulted in geographical expansion, both up- and down-river, of PSTAN speakers. A western and an eastern dialect individualised. The western dialect, in the mid- and upper Huang He Valley, later evolved into PST, whose speakers eventually expanded southward and westward. The eastern dialect was spoken in the lower Huang He and Huai He Valleys. There its speakers adapted to a wetter environment (marine, riverine, lacustrine). The site of Longqiuzhuang, dated to *c*.7,000–5,500 BP in the lower Huai Valley, has both rice and millet (Lu, Chapter 3, this volume, Figure 3.1). A migration brought some of the speakers of this eastern dialect speakers to Taiwan,[6] reached by 5,500 BP. There their language began to diversify into the modern AN languages. Southern elements (cord-marked pottery, bark beaters, etc.) probably entered early AN culture through contact with peoples of southern China. These southern elements do not, however, indicate a south mainland origin of the Austronesians. As to the Tai-Kadai languages, which show strong evidence of relatedness with the AN languages, I have hypothesised that they are not a sister group of AN having remained on the mainland when the pre-Austronesian migrated to Taiwan, but a daughter group of AN, sharing some innovations with the MP languages (see my other Chapter, this volume).

Conclusion

In this chapter I have answered criticisms levelled at earlier versions of my theory. I have significantly increased the number of basic-vocabulary comparisons with sound correspondences between OC and PAN. I have shown that these comparisons, for the most part, have comparanda among the TB languages, and that in some cases TB preserves phonological distinctions reflected in AN but lost in Chinese. I have shown that the OC syllable-type distinction correlates with the nature of the penultimate syllable's initial consonant in AN and that important sections of AN and ST morphology are shared, as well as how PSTAN prefixes and infixes survived the loss of initial syllables. Finally, I have argued that, better than any other theory, a STAN unity explains the spread of a millet-based agriculture

to Taiwan. I therefore maintain, with increasing confidence, my original verdict, voiced in 1990: Chinese and AN *are* genetically related. *Contra* my original assessment, however, I am claiming here that the relationship with AN includes not just Chinese but the whole of ST.

Acknowledgement

Abbreviations

AAK	Austronesian Ancestor of Kadai
AF	Actor Focus
AN	Austronesian
B.	Benedict (1972)
Bur.	Written Burmese
ECL	(Formosan) East Coast Linkage
GEN	Genitive
GF	Goal Focus
LOC	Locative
MC	Middle Chinese
MP	Malayo-Polynesian
OC	Old Chinese
OF	Object Focus
PAN	Proto-Austronesian
PECL	Proto-East Coast Linkage
PMP	Proto-Malayo-Polynesian
PS	Peiros and Starostin (1995)
PST	Proto-Sino-Tibetan
PSTAN	Proto-Sino-Tibetan-Austronesian
SPQR	Starosta, Pawley and Reid (1982)
ST	Sino-Tibetan
STAN	Sino-Tibetan-Austronesian
TB	Tibeto-Burman
Tib.	Written Tibetan
TPK	Tapenkeng

Notes

1 It is assumed that this change spread secondarily to Kanakanabu.
2 Gyarong ka-, kɐ- and kə- are verb prefixes for controllable (ka-, kɐ-) and non-controllable (kə-) actions.

3 Facts from Gyarong were not available to Benedict: he did not realise that voiceless-transitive vs voiced-intransitive alternations in TB verb roots have their origin in intransitive prenasalisation. Neither did he realise that TB intransitive prenasalisation/voicing and Wolfenden's intransitive m- prefix are in complementary distribution with respect to initials: he therefore treated them as two distinct processes.

4 Bradley (1979).

5 The glottal stop was probably artefactually introduced by the transcriber, who assigned the first syllables to the 'entering tone', a glottal-stop-carrying tone, on account of their shortness.

6 A cultural trait found in essentially identical form in the Dawenkou culture of coastal north Jiangsu and south Shandong (in the region of the mouths of the Huang He and Huai River Valleys) and among the modern Formosans, is ritual extraction of upper lateral incisors in both boys and girls, in puberty. Although this feature is widespread among modern southern Chinese populations, it first appears archaeologically in south Shandong *c.*6,500 BP, and is found nowhere else in China at that date (Han and Nakahashi 1996).

Bibliography

Baxter, W.H.III (1992) *A Handbook of Old Chinese Phonology*, Berlin: Mouton de Gruyter.

Benedict, P.K. (1972) *Sino-Tibetan: A Conspectus*, Cambridge: University Printing House.

Bhattacharya, P.C. (1977) *A Descriptive Analysis of the Boro Language*, Gauhati: Gauhati University Department of publications.

Blust, R.A. (1977) 'The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report', *University of Hawaii Working Papers in Linguistics* 9, 2: 1–15.

—— (1999) 'Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics', in E. Zeitoun and P.J.-K. Li (eds) Selected Papers from the *Eighth International Conference on Austronesian Linguistics*, Taipei: Institute of Linguistics (preparatory office).

Bradley, D. (1979) *Proto-Loloish*, London and Malmø: Curzon.

Chang K.C. (1986) *The Archaeology of Ancient China*, 4th edn, New Haven and London: Yale University Press.

Coblin, S. (1986) *A Sinologist's Handlist of Sino-Tibetan Lexical Comparisons*, Nettetal: Steyler Verlag.

van Driem, G. (1997) 'Sino-Bodic', *Bulletin of the School of Oriental and African Studies* 60, 3: 455–88.

Egerod, S. (1980) *Atayal-English Dictionary*, London and Malmø: Curzon Press.

Feng, A.-Z. (1998) *Fuzhou fangyan cidian*, Nanjing: Jiangsu Jiaoyu.

Ferrell, R. (1969) *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*, Nankang: Academia Sinica.

Fu, D.-X. (2001) Xizang Changguogou yizhi xin shiqi shidai nongzuowu yicun de faxian, jianding he yanjiu, *Kaogu* 3: 66–74.

Gong, H.-C. (1995) 'The system of finals in Sino-Tibetan', in W.S.-Y. Wang (ed.) *The Ancestry of the Chinese Language*, Journal of Chinese Linguistics monograph series no. 8.

Han, K.-X. and Nakahashi, T. (1996) 'A comparative study of ritual tooth ablation in ancient China and Japan' *Anthropol. Sci.* 104, 1: 43–64.

Ho, D.-A. (1998) 'Taiwan Nandaoyu de Yuyan Guanxi', *Chinese Studies* 16, 2: 141–71.

Huang, L.M.-J. (1991) 'The semantics of s- in Atayal', *Studies in English Literature and Linguistics* 17: 37–50, Taipei: Department of English, National Taiwan Normal University.

Li, Z.-G. (1991) 'Yi Meng fangyan de "fenyinci" ', *Fangyan* 3: 206–10.

Lin, X.-R. (1993) *Jiarong yu yanjiu*, Chengdu: Sichuan Minzu.

Peiros, I. and Starostin, S. (1995) *A Comparative Vocabulary of Five Sino-Tibetan Languages*, Melbourne: Department of Linguistics and applied Linguistics, University of Melbourne.

Sagart, L. (1990) 'Chinese and Austronesian are genetically related', Paper presented at the *23rd International Conference on Sino-Tibetan Languages and Linguistics*, Arlington, October 1990.

—— (1993) 'Chinese and Austronesian: evidence for a genetic relationship', *Journal of Chinese Linguistics* 21, 1: 1–62.

—— (1994) 'Old Chinese and Proto-Austronesian evidence for Sino-Austronesian', *Oceanic Linguistics* 33, 2: 271–308.

—— (1999) *The Roots of Old Chinese*, Amsterdam: John Benjamins.

—— (2001) 'Vestiges of Archaic Chinese Derivational Affixes in Modern Chinese Dialects', in H. Chappell (ed.) *Sinitic Grammar – Synchronic and Diachronic Perspectives*, Oxford: University Press.

—— (2003) 'Sources of Middle Chinese manner types: Old Chinese prenasalized initials in Hmong-Mien and Sino-Tibetan perspective', *Language and Linguistics*, 4.4: 757–68.

Starosta, S. (1991) 'Ergativity, transitivity, and clitic coreference in four Western Austronesian languages', Paper presented at *6th International Conference on Austronesian Linguistics*, Honolulu, May 1991.

—— (1994) 'Formosan clause structure: transitivity, ergativity, and case marking', revised version (18 October 1994) of a paper presented at the *4th International Symposium on Chinese languages and Linguistics*, Nankang, July 1994.

—— (1995) 'The Chinese-Austronesian connection: a view from the Austronesian morphology side', in W.S.-Y. Wang (ed.) *The Ancestry of the Chinese Language*, Journal of Chinese Linguistics monograph series no. 8.

—— 'Reduplication and the subgrouping of Formosan languages', Paper presented at *International Symposium on Austronesian cultures: Issues relating to Taiwan*, Taipei, December 2001.

Starosta, S., Pawley, A. and Reid, L. (1982) 'The evolution of Focus in Austronesian', in A. Halim, L. Carrington and S. Wurm (eds) Papers from the *Third International Conference on Austronesian Linguistics, vol. 2*, Canberra: Pacific Linguistics.

Wang, F.-S. and Mao, Z.-W. (1995) *Miao-yao yu guyin gouni*, Beijing: Zhongguo Shehui Kexue.

Wolfenden, S. (1929) *Outlines of Tibeto-Burman Linguistic Morphology*, London: Royal Asiatic Society.