

## Origins of the East Formosans: Basay, Kavalan, Amis, and Siraya<sup>\*</sup>

Paul Jen-kuei Li  
*Academia Sinica*

Both phonological and lexical evidence is examined to show the internal relationships of four ethnic groups in Formosa: Basay, Kavalan, Amis, and Siraya. The merger of \*j and \*n, occurring nowhere else, defines the entire East Formosan group. Two phonological innovations, (1) merger of \*j, \*n, and \*N as *n* and (2) split of \*k into *k* and *q* (before \*a), are exclusively shared by Basay and Kavalan. Shared innovations are also found in sporadic changes in a few lexical items in these two languages. However, linguistic evidence for a close relationship between Amis and Siraya is much weaker. This study includes a list of cognates shared by the East Formosan group.

Key words: Austronesian subgrouping, East Formosan languages, phonological & lexical comparison, migration history

### 1. Introduction

The term *East Formosan* can refer to various ethnic groups in Taiwan: to the Amis of Hualien and Taitung Counties, which do indeed lie along the east coast; and also to the Kavalan of I-Lan County in the northeast; but also to the Basay in the far north. The term does not, however, refer to the Puyuma or Paiwan occupying the southeast coast. Also note that the term *East Formosan* also includes the people once using Siraya to the southwest of the island. At any rate, rather than using *East Formosan* in an inconsistent, geographic sense, we shall be using it linguistically. It was thus first used by Blust (1999:45), who proposes that Basay-Trobiawan, Kavalan, Amis, and Siraya belong to a single subgroup as indicated by the following phonological evidence: (1) the merger of

---

<sup>\*</sup> This study was supported in part by grants from the National Science Council (NSC91-2411-H-001-051), Chiang Ching-kuo Foundation for International Scholarly Exchange (RG002-D-'00), and Academia Sinica. I wish to thank K. Alexander Adelaar, Robert Blust, James Fox, and Stanley Starosta for their comments on an earlier version of this paper when it was presented at the *International Symposium on Austronesian Cultures: Issues Relating to Taiwan*, Taipei, December 8-11, 2001. I would also like to thank the two anonymous reviewers for their very helpful suggestions, which have led to great improvements in this paper. I remain responsible for any errors that still remain.

\*t/C; (2) the merger of \*j/n; and (3) the shift of \*q > ʔ.<sup>1</sup>

The oral traditions of the Basay, Kavalan, and Amis share a homeland motif that has all of them coming originally from an island called “Sinasay” or “Sanasay” (Mabuchi 1953-4). My purpose here is to find linguistic support for the idea that the four members of East Formosan could have had a shared ancestry. However, in the case of Basay and Kavalan we must bear in mind that geographical proximity is conducive to cultural and linguistic diffusion.

Now we shall attempt to show the internal relationships of the four East Formosan languages by examining their phonological and lexical correspondences.

## 2. Phonological evidence

In Blust (1999:45), the Formosan languages are classified by shared phonological innovations, resulting in nine groups. East Formosan is further subdivided into three branches.

East Formosan

1. Northern branch: (1) merger of \*q/ø, (2) merger of \*n/N; (3) irregular change in \**susu* > /sisu/ ‘breast’
  - 1.1 Basay-Trobiawan: (1) merger of \*s/l
  - 1.2 Kavalan
2. Central branch (Amis)
3. Southwest branch (Siraya)

Blust is essentially correct in considering Basay and Kavalan to be most closely related to one another; and a lot more solid phonological and lexical evidence can be adduced to substantiate this view. Li 2001 shows that Basay and Kavalan share the following phonological innovations: (1) the merger of \*t and \*C as *t*; (2) the merger of \*D and \*Z; (3) the merger of \*q, \*H, \*ʔh and ø; (4) the merger of \*j, \*n, and \*N as *n*; and (5) the split of \*k into *k* and *q* (before \*a). The last of these further underwent the change *q* > *h* in Basay, though not in the closely related Trobiawan dialect.<sup>2</sup> Of the five changes, (4) and (5) are phonological innovations found only in Basay and Kavalan. For example,

---

<sup>1</sup> As a matter of fact, the Amis reflex for \*q is a pharyngealized glottal stop, represented by *Q* in our transcription, which is phonemically distinct from the glottal stop /ʔ/.

<sup>2</sup> The change *q* > *h* is evidenced in dialectal differences, namely *h* in Basay corresponds to *q* in Trobiawan, as in the following lexical forms: *harona* ~ *qaLuna* ‘ant,’ *habateng* ~ *mia-qabateng* ‘sick,’ *habitte* ~ *q<um>avit* ‘to harvest rice,’ *haed*’ ~ *qael* ‘abalone.’

- (1) Merger of \*t and \*C as *t*:  
 \*t-inaH > Bas<sup>3</sup> tina, Kav tina ‘mother’  
 \*maCa > Bas mata, Kav mata ‘eye’
- (2) Merger of \*D and \*Z as *r* or *l* in Basay, as *z* in Kavalan:  
 \*DaNum > Bas lanum, Kav z anum ‘water’  
 \*DaDaN > Bas raran, Kav ra-razan (<D)<sup>4</sup> ‘thousand’  
 \*Zalan > Bas cacan (<A), Kav razan (<M) ‘road’  
 \*ZaRum > Kav razum (<M) ‘needle’
- (3) Merger of \*q, \*H, \*ʔh and ø:  
 \*panaq > Bas pana ‘to throw,’ Kav p<m>ani ‘to shoot’  
 \*t-amaH > Bas tama, Kav tama ‘father’  
 \*kaʔen > Bas k<um>an, Kav q<m>an ‘to eat’
- (4) Merger of \*j, \*n, and \*N as *n*:  
 \*Suaji > Bas suani, Kav suani ‘younger sibling’  
 \*pajey > Kav panay ‘rice plant’  
 \*-ajem > Bas anem, Kav anem ‘heart’  
 \*ngajan > Bas nganan, Kav nangan (<M) ‘name’  
 \*bulaN > Bas bucan, Kav buran ‘moon’  
 \*tiaN > Bas tian ‘belly,’ Kav m-tian ‘pregnant’
- (5) Split of \*k into *k* and *q* > *h* (in Basay only) before \*a  
 \*kaRat > Bas harate, Kav qaRat ‘to bite’  
 \*kaRaw > Bas halaw, Kav qaRaw ‘to scratch’  
 Exceptions: \*kaka > Bas kaka, Kav qaqa ‘elder sibling’  
 \*sakay > Bas c<um>aka-cakay, Kav saqay ‘walk’

However, the phonetic environment for the change \*k > *q* was subsequently extended to all the back vowels in Kavalan (although there are exceptions),<sup>5</sup> but not in Basay, as in:

		<u>Basay</u>	<u>Kavalan</u>	
(6)	*baki	> baki	baqi	‘grandfather’
	*ku(S)kuS	> k<an>ukus	q<an>uqus	‘fingernail’
	*bukeS	> bukes	buqes	‘hair’
	*kuCu	> kutu	qutu	‘head louse’

<sup>3</sup> Abbreviations as used in this paper are as follows: Ami = Amis; AmiFa = Fata’an dialect of Amis; AmiFr = Farangaw dialect of Amis; Bas = Basay; Kav = Kavalan; Sir = Siraya; Paz = Pazih; Tha = Thao.

<sup>4</sup> The symbol <A stands for assimilation, <D for dissimilation, and <M for metathesis.

<sup>5</sup> There are exceptions in Kavalan; e.g., \*siku > siku ‘elbow’ (\*k did not become *q* before \*u), \*kita > m-qita ‘see’ (\*k became *q* before \*i).

Shared innovations are also found in the sporadic changes of a few lexical forms: *\*susu* > Bas *cicu*, Kav *sisu* ‘breast’ (anticipatory dissimilation of the first vowel), *\*piliq* > Bas *p<am>ici*, Kav *p<am>il*<sup>6</sup> ‘to choose’ (insertion of <am> after the stem-initial consonant), or alternatively nasal substitution of the initial consonant: Bas *pa-mici* and Kav *pa-mil*.<sup>7</sup> These sporadic changes shared exclusively by Basay and Kavalan provide strong evidence for their close relationship.

However, linguistic evidence for a close relationship between Amis and Siraya is much weaker. They share the following phonological innovation: merger of \*j and \*n, but not with \*N. In addition, they share the following sporadic sound change: *\*CalingaH* > Ami *tangila*, Sir *tangira* ‘ear,’ where both forms show metathesis of \*l and \*ng.<sup>8</sup> However, Kavalan and Amis (but not Siraya) share other sporadic metatheses, as in *\*taSiq* > Kav *tais*, Ami *mi-taQis* (Sir *t<m>ahy*) ‘to sew’; i.e., metathesis of \*S and \*q in both the Kavalan and Amis forms (but not in Siraya); *\*mula*<sup>9</sup> > Kav *pa-ruma*, *m-ruma* (<M>), Ami *pa-ruma* (<M>) ‘to plant’ (a Siraya cognate not attested).

(7) Sound Correspondences in the East Formosan Group

PAN	Basay	Kavalan	Amis <sup>10</sup>	Siraya
*p	p	p	p	p
*b	b	b	f	v, -f
*t	t	t	t	t
*C	t	t	t	t

<sup>6</sup> An anonymous reviewer offers a quite plausible explanation for the missing \*-i in the Kavalan form *pamil*: *\*paN-piliq* > *pamili* (nasal substitution and loss of \*q) > *pamill* (assimilation of \*-i to the preceding \*l) > *pamil* (simplification of the final geminate consonant). The geminate *ll* is retained when the following suffix begins with a vowel, as in *pamill-i* ‘Choose it!’ These sound changes parallel the ones in *\*Caqi* > *\*tavi* > *\*tali* > *tal* (final) ~ *tall* (when suffixed) ‘faeces’ in Kavalan.

<sup>7</sup> Tsuchida (p.c.) pointed out to me that since it is not found in any Austronesian language having an infix <am>, it is more likely to have the nasal substitution for the stem-initial stop; and yet, a nasal substitution is also rather rare in Formosan languages. Remnants of a similar infix <ar> is occasionally attested in a few forms, such as Tha *b<ar>amban* ‘plant sp., Hyptis rhomboides Mart & Gal,’ Paz *b<ar>an(a)ban* ‘urn,’ *d<ar>ung(u)dung* ‘gong, drum’ (Li and Tsuchida 2001: 18-19).

<sup>8</sup> As pointed out by one of the reviewers, Pazih also shows a metathesis of \*l and \*ng in its form *sangira*, and there is a tendency of metathesis involving a nasal and lateral in various different languages, such as *\*langaw* > Ami *ta-ngalaw* ‘fly.’

<sup>9</sup> Three Formosan languages have undergone the metathesis of \*m and \*l, while only one has not: *\*mula* > Saisiyat *ma-moLa*, Pazih *paxu-ruma*, Kavalan *m-ruma*, Amis *pa-ruma* ‘to plant.’

<sup>10</sup> These sound correspondences are based on the Sakizaya dialect of Amis. All Amis forms are cited from the same dialect unless specified otherwise.

*D	r, l	z	d	s
*Z	r, l	z	d	d
*R	r, l	r, l, R	l	g [x], ø /i i
*k	k, h / _a	k, q	k	k
*q	ø	ø	Q	ø
*H	ø	ø	h, ʔ	ø
*ʔ	ø	ø	ʔ	ø
*l	c	r, l, R	l	r, l, g
*s	c	s	c	s, -ø
*S	s	s	s	h, ø
*j	n	n	n	n
*n	n	n	n	n
*N	n	n	d	l
*m	m	m	m	m
*ng	ng	ng	ng	ng
*w	w	w	v	w
*y	y	y	y	y
*a	a	a, i /*q	a	a
*i	i	i	i	i
*u	u	u	u	u
*e	e, u, i	e, i	e, a /*q	i, ø

## (8) Examples of cognates in the East Formosan group:

- \*t-inaH > Bas tina, Kav tina, Ami ina, Sir ina ‘mother’  
 \*t-amaH > Bas tama, Kav tama, Ami ama, Sir s-ama ‘father’  
 \*baki > Bas baki, Kav baqi, Ami baki ‘grandfather’  
 \*kaka > Bas kaka, Kav qaqa, Ami kaka, Sir s-acka, j-akka ‘elder sibling’  
 \*Suaji > Bas suani, Kav suani ‘younger sibling’  
 \*baHi > Bas b<in>ay ‘woman,’ Kav bai ‘grandma,’ AmiFr fa-fahi’-an ‘woman’  
 \*RuqaNay > Kav Runanay (<A) ‘man, male’  
 \*Cau > Bas tau ‘person,’ Ami Tau ‘other person,’ Sir j-tou ‘person’  
 \*aNak > Bas wanake, Sir allack ‘child’  
 \*Duma > Kav zuma ‘other person,’ Ami duma ‘other’  
 \*qasawa > Bas cawaa (<M) ‘husband’  
 \*i-ku, \*i-aku > Bas yaku, Kav iku, Ami aku, Sir jau, -au ‘I’  
 \*i-ta, \*i-kita > Bas -ita, kita, Kav ita, Ami kita, Sir -eta -kytta ‘we (incl)’  
 \*i-mi, \*i-kami > Bas yami, Kav imi, Ami kami, Sir ymi-an, -kame ‘we (excl)’  
 \*i-Su, \*i-kaSu > Bas isu, Kav isu, Ami k-isu, Sir -hou ‘you (sg)’

- \*i-mu, \*i-kamu > Bas imu, Kav imu, Ami kamu, Sir -kamou ‘you (pl)’  
 \*quluH > Bas ucu, Kav uRu ‘head’  
 \*bunguH > Sir vongo ‘head,’ cf. Tsou fnguu, Kan na-vungu, Sar vunguʔu ‘head’  
 \*lima > Bas cima, Kav rima, Ami lima, Sir rima ‘hand’  
 \*maCa > Bas mata, Kav mata, Ami mata, Sir matta ‘eye’  
 \*CalingaH > Ami tangira (<M>), Sir tangira (<M>) ‘ear’  
 \*Dilaq > Sir da-dila ‘tongue’  
 \*siku > Bas ciku, Kav siku, Ami ciku ‘elbow’  
 \*likuD > Kav k<m>u-rikuz ‘follow,’ pata-rikuz ‘reply,’ Sir ricos ‘back’  
 \*tiaN > Bas tian ‘belly,’ Kav m-tian ‘pregnant,’ AmiFr tiahL ‘belly’  
 \*DapaN > Kav zapan ‘leg and foot,’ Sir sapal ‘foot’  
 \*CuqelaN > Bas tulan, Kav tiRRan ‘bone’  
 \*Caqi > Bas tai, Kav tal, Ami taiQ (<M>), Sir tai-ng ‘faeces’<sup>11</sup>  
 \*C<in>aqi > Bas t<in>ai, Kav t<n>al, Ami t<in>aiQ (<M>) ‘intestines’  
 \*qetut > Bas mu-tutte, Kav ʔtut, AmiFr Qtut ‘fart,’ Sir ma-tout ‘break wind’  
 \*ikuR > Ami k-ikul (cf. Sai k-ikoL), Sir ikog ‘tail’  
 \*-ajem > Bas anem, Kav anem ‘mind’  
 \*quCeluR > Bas (telud’),<sup>12</sup> Kav tiRuR ‘egg’  
 \*Sesi > Bas ci, Kav ʔsi, Ami heci ‘meat’  
 \*SimaR > Kav simaR, Ami simal, Sir gimach (<A>) ‘fat, grease’  
 \*ku(S)kuS > Bas k<an>u:kus (Abe 1930:432), Kav q<n>uqus, AmiFa (k<an>uʔus)  
 ‘fingernail’  
 \*bukeS > Bas bukes, Kav buqes, Ami bukes, Sir vukig, but-sek (<M>) ‘hair’  
 \*kuCu > Bas kutu, Kav qutu, Ami kutu, Sir kutu ‘head louse’  
 \*susu > Bas cicu (<D>), Kav sisu (<D>), Ami cucu ‘breast’  
 \*luSeq > Kav Rusi, Ami lusaQ ‘tears’  
 \*sibu > Bas cubu (<A>) ‘urine’  
 \*iSiq > Ami isiQ ‘urine’  
 \*babuy > Bas babuy, Kan babuy, Ami pabuy (<D>), Sir vavoy ‘pig’  
 \*wasu > Bas wacu, Kav wasu, Ami wacu, Sir assou ‘dog’  
 \*luCung > Bas (lutung), Kav Rutung, AmiFa lutung, Sir routong ‘monkey’  
 \*Cumey > Bas tomai, Kav tumay, Ami tumay, Sir tumey ‘bear’  
 \*qaNuang > Sir lowang ‘carabao’

<sup>11</sup> An anonymous reviewer pointed out the possible cognates of Kavalan *tal* < \*Caqi ‘faeces’ and *t<n>al* < \*C<in>aqi ‘intestines.’ Cf. the irregular Kavalan reflex of *l* for \*y adjacent to \*a in the form *alam* < \*qayam ‘bird’ (if it is inherited). But further work needs to be done before the whole picture becomes clear.

<sup>12</sup> A suspect cognate with one or more unanticipated segments is given in parentheses.

- \*langaw > Kav rangaw, Ami ta-ngalaw (<M), Sir a-ga-gangaw<sup>13</sup> ‘fly’  
 \*waNuH > Kav wanu, Ami wadu, Sir hou-walou ‘honeybee’  
 \*qayam > Kav alam, Ami Qayam, Sir aiam ‘bird’  
 \*tuNa > Ami tuda, Sir toula ‘eel’  
 \*kaRang > Ami kalang, Sir kagan (<D) ‘crab’  
 \*kuRita > Kav qlita ‘octopus’  
 \*Sanaq > Kav sani, Ami sanaQ ‘otter,’ Sir hanna ‘fox’  
 \*paRiS > Sir pagig ‘rayfish’  
 \*paNij > Sir pa-palis ‘wing’  
 \*waNiS > Ami wadis, Sir walig ‘tooth (of animal)’  
 \*kaʔen > Bas k<um>an, Kav q<m>an, AmiFr k<um>aʔen ‘eat,’ Sir k<m>ang (<A)  
 ‘eat some fruit’  
 \*sepsep > Bas cepcep, Kav s<m>epsep ‘suck,’ Ami mi-cepcep, Sir s<m>ippsip  
 ‘suck blood’  
 \*utaq > Bas uta, Kav m-uti, Ami m-utaQ ‘to vomit’  
 \*panaq > Bas pana ‘throw,’ Kav p<m>ani, Ami mu-panaQ ‘shoot’  
 \*taktak > Bas taktak, Kav taktak ‘chop’  
 \*sakay > Bas c<um>aka-cakay, Kav saqay ‘walk’  
 \*Nanguy > Bas nanuy (<A), Kav nanguy, Ami mi-danguy, Sir l<m>angoy ‘swim’  
 \*Cawa > Kav tawa, Ami ma-tawa, Sir ma-ttawa ‘laugh’  
 \*Cangis > Bas t<um>angice, Ami t-om-angic (Fey 1986:301), Sir t<m>angi  
 ‘weep’  
 \*paCey > Bas patay ‘hit,’ Kav patay, Ami patay, Sir ma-patey ‘die’  
 \*alap > Bas acap ‘take’  
 \*ala > Kav ara, Ami mu-ala, Sir m-ara ‘take’  
 \*beRay > Bas balai (<A), vulai (Tr), Kav bura, Ami pa-beli, Sir (picha) ‘give’<sup>14</sup>  
 \*Sajek > Kav sanek, Ami mi-sanek ‘smell’  
 \*kaRaC > Bas harate, Kav qaRat, Ami mu-kalat, Sir k<m>agat ‘bite’  
 \*kita > Kav m-qita, Sir k<m>itta ‘see’  
 \*kiRim > Kav k<m>irim, Ami mi-kilim, Sir k'iim-au, k<m>i'ym ‘seek’  
 \*Suab > Kav kar-suab, Ami ma-suab ‘yawn’  
 \*qiNep > Kav qa-inep ‘sleep’  
 \*taSiq > Kav tais (<M), Ami mitaQis (<M), Sir t<m>ahy ‘sew’  
 \*kaRaw > Bas halaw, Kav qaRaw ‘scratch’

<sup>13</sup> As pointed out by one of the reviewers, the Siraya form shows an unanticipated reflex of *g* for the initial, suggesting a proto-form of \*Rangaw.

<sup>14</sup> The loss of \*-y in Kavalan is unexplained, and irregular reflex of Amis *i* for \*-ay is also a problem.

- \*mula > Bas (pa-luma)<sup>15</sup> (<M), Kav pa-ruma (<M), Ami pa-luma (<M) ‘plant’
- \*piliq > Bas pam-ici, Bas pam-ili, Ami mi-piliQ, Sir peri, mou-piri ‘choose’
- \*Siup > Bas siupe, Kav s<m>iup, Ami mi-ʔiup ‘blow’
- \*laRiw > Bas rareu, Kav RaRiw, Ami mu-laliw ‘escape’
- \*qeNeb > Kav m-ineb, Ami mi-Qedeb, Sir illif ‘close’
- \*esa > Bas ca, Kav issa ‘one’
- \*DuSa > Bas lusa, Kav zusa, Ami tusa (<A),<sup>16</sup> Sir sa-soa ‘two’
- \*teluH > Bas cuu, Kav turu, Ami tulu, Sir tou-touro ‘three’
- \*Sepat > Bas sepat, Kav spat, Ami spat, Sir pa-gpat ‘four’
- \*lima > Bas cima, Kav rima, Ami lima, Sir ri-rima ‘five’
- \*enem > Bas anem, Kav nnem, Ami enem, Sir (ninnam) ‘six’
- \*pitu > Bas pitu, Kav pitu, Ami pitu, Sir pi-pitu ‘seven’
- \*walu > Bas wacu, Kav waru, Ami walu ‘eight’
- \*Siwa > Bas siwa, Kav siwa, Ami siwa ‘nine’
- \*DaDaN > Bas raran, Kav ra-razan (<D) ‘thousand’
- \*ngajan > Bas nganan, Kav nangan (<M), Ami ngangan (<A), Sir nanang (<M) ‘name’
- \*layaR > Bas rayar (<A), Kav RayaR (<A) ‘sail’
- \*busuR > Bas (vucal) ‘bow,’ Kav busuR ‘bow string,’ Ami bucul ‘bow’
- \*qaSeluH > Bas (li-cu), Kav saRu, AmiFa Qasulu, Sir haiero (<M) ‘pestle’
- \*ZaRum > Kav razum (<M) ‘needle’
- \*Zalan > Bas cacan (<A), Kav razan (<M), Ami dadan (<A), Sir darang ‘road’
- \*Damar > Bas ramar, Kav zamaR, Ami lamal (<A) ‘fire’
- \*Sapuy > Sir apoy ‘fire’
- \*baRaH > AmiFr falah ‘charcoal,’ Sir vaga ‘burning charcoal’
- \*qabu > Bas abu, Kav ibu, Ami Qabu, Sir avo ‘ash’
- \*batuH > Bas batu, Kav btu, Ami baʔtu, Sir vahto, vatto ‘stone’
- \*siNaR ‘light, ray’ > Bas cenal, Ami cilal (<A) ‘sun’
- \*waRiH > Kav waRi, Ami wali ‘east,’ Sir wagi ‘sun, day’
- \*bulaN > Bas bucan, Kav buran, Ami bulad, Sir vourel ‘moon’
- \*bali > Bas baci, Kav bari, Ami bali, Sir vare ‘wind’
- \*quDaN > Bas uran, Kav uzan, Ami Qudad (<A), Sir oudal ‘rain’
- \*SuReNa > Kav surna, Ami sulda, Sir ougla ‘snow’
- \*DaNum > Bas lanum, Kav zanum, Ami nanum (<A), Sir salom ‘water’

<sup>15</sup> Since the regular Basay reflex of \*l is c, the anticipated form for the verb ‘to plant’ would be \*muca. In fact, the actual form we get is *paluma*, which is probably a loan from Kavalan.

<sup>16</sup> As pointed out by a reviewer, the change of the initial *t*- in Amis could be due to analogy to *tulu* ‘three,’ or more likely both assimilation and analogy.



- \*pajey > Kav panay, Ami panay ‘rice plant’
- \*beRas > Bas bulace, Kav bRas, Ami belac ‘husked rice,’ Sir pchag (<A) ‘stamped rice’
- \*Semay > Bas sumay, Kav ?may, Ami hemay ‘cooked rice,’ Sir mey ‘smoked rice’
- \*quay > Bas wa-wai, Kav uway, Ami Quway, Sir uweg ‘rattan’
- \*panguDaN > Kav pangzan ‘pandanus,’ Sir pandal (<A) ‘pineapple’
- \*biRaq > Bas bila, Kav biRi ‘leaf’
- \*RameC > Bas lamit, Ami lamit ‘root’
- \*buaq > Sir boa ‘fruit’
- \*Daya > Bas laya, Kav zaya ‘west,’ Sir taga-seia ‘east’
- \*laHud > ‘seawards’ > Kav rauz ‘tip’ (?), Sir r<m>aos ‘west’
- \*qamiS > Bas amis, Kav imis, Ami Qamis-an, Sir tag-amig ‘north’
- \*babaw > Kav babaw, Ami pabaw (<D), Sir vavau ‘above’
- \*wiRiH > BasTr ka-wiLi, Kav ka-wiri-an, Ami ka-wili, Sir (m-ougy) ‘left’
- \*wanaN > Kav ka-wanan, Ami ka-wanan, Sir m-onal ‘right’
- \*Rabi?iH > Bas rabi, Kav sa-Rabi-an, Ami labi ‘evening’
- \*ma(nge)taq > Bas mata, Kav mti, Ami mantaQ ‘raw’
- \*baqeRuH > AmiFr faQeluh-ay, Sir vahu ‘new’
- \*cengeN > Kav tngen ‘black’
- \*Raya > Kav Raya ‘big’
- \*NiSepis > Kav inpis (<M), Ami (duhpic) ‘thin’
- \*buSuk > Kav m-busuq, Ami ma-busuk ‘drunk’
- \*i-nu > Bas a-inu, Sir h-ino ‘where’
- \*pijaH > Bas pina, Ami pina, Sir pi-pina ‘how many’

Now we are still left with the following questions about this classification: Are the three branches of East Formosan really coordinate with one another, as suggested by Blust (1999)? Or can any two given branches be even more closely related? As of yet, we have no solid linguistic evidence either way. As mentioned above, the phonological evidence indicates that \*j has merged with \*n, but \*N has not merged with \*j/n in either Amis or Siraya, as it has in Basay and Kavalan. But the distinction between \*N and \*j/n is one of phonological retention, rather than innovation. As a working hypothesis, we should only assume that Amis and Siraya may be more closely related to one another than either is to Basay/Kavalan.

### 3. Lexical evidence

Languages of the East Formosan group have the following exclusively shared cognates:

- (9) \*lasibu > Bas RACIBU, Kav m-Rasibu, Ami a-lasubu (<A) ‘hundred’  
 \*sabak > Bas sabak, Kav sabaq, Ami sabak ‘grains’

Basay and Kavalan have quite a few uniquely shared cognates:

(10)	<u>Basay</u>	<u>Kavalan</u>	
a.	bancaw	bangRaw	‘tooth’
	pakaw	paqaw	‘spider’
	kulupu	qrupu	‘eyelash’
	ma-sikat’	m-sikaz	‘shy, shameful’
b.	tasaw	tasaw	‘year’
	baute	baut	‘fish’
	tnayan	tnayan	‘bamboo’
	mutun	mutun	‘rat’

Note that Amis contains a few lexical forms found only in the western plains languages; e.g., Ami *nu-tipan*, Taokas *tipan*, Babuza *tsipan* ‘west.’ This may suggest that Amis has had a closer relationship or contact with the western plains languages; or, this could simply be a case of retention in Amis, with loss of the etymon in the other East Formosan languages.

### 4. Other types of evidence

Some **syntactic** evidence seems to indicate that Siraya (Adelaar 1997) has greater affinity with Basay (Li 1999) and Kavalan (Li 1996, Chang 1997) in the merger of Patient-focus and Locative-focus forms, while Amis has retained that distinction. Note, however, that the merger was not complete in Siraya, since there are still both *-en* and *-an* verb forms (although no functional difference in use). And there is always the possibility that the languages first split and then developed in parallel fashion.

Formosan **ethnic** groups have lived in settlements of different sizes and patterns reflecting variation in social organization (Mabuchi 1960). Among Taiwan mountain groups, “The closest parallel to the Siraya is the matrilineal Amis”<sup>17</sup> (Shepherd 1993:

<sup>17</sup> As far as we can determine, Kavalan, Amis, and Siraya are all matrilineal.

458, note 88).

In conclusion, the linguistic evidence for a close relationship between Basay and Kavalan is solid; but the evidence for a closer relationship of Siraya to either Amis or Basay-Kavalan is still inadequate, as is also the case for proving a closer affinity of Amis with either Siraya or Basay-Kavalan.

The linguistic evidence for an East Formosan group—Basay, Kavalan, Amis, and Siraya—is not in doubt. If this subgrouping hypothesis is correct, then Basay, Kavalan, and Amis may have originated in the southwestern plains of Taiwan, where the Siraya people have resided for centuries. The relative chronology of an eastward migration would then largely depend on the linguistic closeness of Amis to Siraya.

## References

- Abe, Akiyoshi (安倍明義). 1930. 《蕃語研究》 [*Study of the Languages of Native Tribes*]. 台北：蕃語研究會。
- Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36.2:362-397.
- Asai, Erin (淺井惠倫). (ed). 1939. *Gravius's Formulary of Christianity in the Siraya Language of Formosa*. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University [台北帝國大學文政學部紀要] 4.1. Formosa: Taihoku Imperial University.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31-94. Symposium Series of the Institute of Linguistics (Preparatory Office), No.1. Taipei: Academia Sinica.
- Campbell, Rev. William. 1888. *The Gospel of St. Matthew in Sinkang-Formosan, with corresponding versions in Dutch and English edited from Gravius's edition of 1661*. London: Trubner and Co.
- Chang, Yung-li (張永利). 1997. *Voice, Case, and Agreement in Seediq and Kavalan*. Hsinchu: National Tsing Hua University dissertation. Abstract appeared in *Glott International* 4.6:10, Holland and Academic Graphics (2000).
- Fey, Virginia (方敏英). 1986. 《阿美語字典》 [*Amis Dictionary*]. 台北 [Taipei]: 財團法人中華民國聖經公會出版 [The Bible Society].
- Li, Paul Jen-kuei (李壬癸). 1982. Kavalan phonology: Synchronic and diachronic. *GAVA: Studies in Austronesian Languages and Cultures Dedicated to Hans Kähler* 17:479-495. Berlin: Dietrich Reimer Verlag.
- Li, Paul Jen-kuei (李壬癸). 1996. 《宜蘭縣南島民族與語言》 [*The Formosan Tribes and Languages in I-Lan*]. 宜蘭 [I-Lan]: 宜蘭縣政府 [The Government of I-Lan].
- Li, Paul Jen-kuei (李壬癸). 1999. Some problems in the Basay language. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 635-664. Symposium Series of the Institute of Linguistics (Preparatory Office), No.1. Taipei: Academia Sinica.
- Li, Paul Jen-kuei (李壬癸). 2001. 〈巴賽語的地位〉 [The linguistic position of Basay], 《語言暨語言學》 [*Language and Linguistics*] 2.2:155-171。
- Li, Paul Jen-kuei (李壬癸), Shigeru Tsuchida (土田滋). 2001. *Pazih Dictionary* [巴宰語詞典]. Language and Linguistics Monograph Series, No.A2. Taipei: Academia Sinica.
- Mabuchi, Tôichi (馬淵東一). 1953-4. 〈高砂族の移動及び分布〉, 《民族學研究》

- [*Journal of Ethnology*] 18.1-2:123-154, 18.4:23-72。
- Mabuchi, Tôichi (馬淵東一). 1960. The aboriginal peoples of Formosa. *Social Structure in Southeast Asia*, ed. by George P. Murdock, 127-140. Chicago: Quadrangle.
- Murakami, Naojirô (村上直次郎). 1933. *Sinkan Manuscripts* [新港文書]. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University [台北帝國大學文政學部紀要] 2.1. Formosa: Taihoku Imperial University.
- Shepherd, John Robert. 1993. *Statecraft and Political Economy on the Taiwan Frontier, 1600-1800*. Stanford: Stanford University Press.
- Tsuchida, Shigeru (土田滋). 1976. *Reconstruction of Proto-Tsouic Phonology*. Study of Languages & Cultures of Asia & Africa, Monograph Series, No.5. Tokyo: Tokyo University of Foreign Studies.
- Tsuchida, Shigeru (土田滋). 1982. *A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan I: West Taiwan*. Memoirs of the Faculty of Letters, University of Tokyo, No.7 [東京大學文學部研究報告7・語學・文學論文集]. Tokyo: Faculty of Letters, University of Tokyo.
- Tsuchida, Shigeru (土田滋). 1993. 〈カバラン語〉 [Kavalan language]・《言語學大辭典》 [*Encyclopedia of Linguistics*] 5:89-99。東京 [Tokyo]: 三省堂 [Sanseido]。
- Tsuchida, Shigeru (土田滋), Yukihiro Yamada (山田幸宏), Tsunekazu Moriguchi (森口恒一). 1991. 《台灣・平埔族の言語資料の整理と分析》 [*Linguistic Materials of the Formosan Sinicized Populations I: Siraya and Basai*]。東京 [Tokyo]: 東京大學 [University of Tokyo]。

[Received 18 June 2003; revised 10 November 2003; accepted 19 December 2003]

Institute of Linguistics  
Academia Sinica  
130, Sec. 2, Academia Road  
Nankang, Taipei 115, Taiwan  
hspaulli@ccvax.sinica.edu.tw

## 東部台灣南島語群的來源： 巴賽、噶瑪蘭、阿美、西拉雅

李壬癸

中央研究院

「東部台灣南島語群」包括過去在大台北地區定居的巴賽，蘭陽平原的噶瑪蘭，現在仍在花東縱谷平原的阿美，以及在西南部平原的西拉雅。這四個族群的語言都有一個共同的音變：古南島語的  $*j$  跟  $*n$  合併，這種音變並不見於其他任何南島語言。巴賽語跟噶瑪蘭語又有兩種共同的音變：(一) 古語的  $*j$ ,  $*n$ ,  $*N$  的合併，(二) 古語的  $*k$  在  $*a$  元音之前分化為  $k$  和  $q$ ，而這兩種音變並不見於阿美語或西拉雅語。此外，前兩種語言又都含有一些相似的詞彙形式的特殊變化。因此，它們的密切關係，證據相當確鑿。可是阿美跟西拉雅是否關係較密切，語言學的證據仍嫌薄弱。本文列有東部台灣南島語群的同源詞。

關鍵詞：東部台灣南島語群，歷史語言學，音韻演變，詞彙演變，語言分支