

# The Great Diversity of Formosan Languages\*

Paul Jen-kuei Li

*Academia Sinica*

Formosan languages are extremely diverse at all linguistic levels, from phonology to morphology to syntax. In fact, the Formosan languages are the most diverse within the entire Austronesian language family. This great linguistic diversity indicates a great time depth for an early settlement on the island, so as to make Taiwan the prime candidate for Austronesian homeland and the center of Austronesian dispersal. In this paper, we shall briefly survey the diversity of the Formosan languages, highlighting eight linguistic features: word order, focus systems, auxiliaries, numerals, personal pronouns, compounding, affixation, and phonology.

Key words: Austronesian homeland, Formosan languages, linguistic structure

## 1. Introduction

The Austronesian language family is geographically the most wide-spread, covering the Indo-China and Malay peninsulas, and the Pacific and Indian Oceans. The number of languages is estimated to be from 800 to 1,200, depending on the criteria of what we consider to be separate languages or different dialects.

Formosan languages are generally believed to be the most diverse in the entire Austronesian language family. The more diverse a group of genetically-related languages are found in a geographical area, the more time depth it implies for its early settlement in that area (Sapir 1916). Hence, Taiwan is the most likely Austronesian homeland and center of Austronesian dispersal.

Although there are only a small number of Formosan languages, some twenty in total, yet there is great diversity in their linguistic structures. The great diversity of

---

\* An earlier version of this paper was presented as a keynote speech at the 14<sup>th</sup> Annual Conference of the International Association of Chinese Linguistics and 10<sup>th</sup> International Symposium on Chinese Languages and Linguistics Joint Meeting, Taipei, 25-28 May 2006. I have greatly benefited from comments by Edith Aldridge, Henry Yungli Chang, Dah-an Ho, Stacy Fang-ching Teng, Shigeru Tsuchida, and Elizabeth Zeitoun. However, I remain responsible for any errors still contained in this paper.

Formosan languages at all levels of linguistic structures makes them most interesting for linguistic investigation, yet it presents a great problem for classification, for instance, Blust's (1999) nine major subgroups of Formosan languages. An aim of this paper is to find out the diversity of various Formosan languages, including three that are extinct, Siraya, Favorlang, and Basay. The linguistic features covered in this paper include word order, focus systems, auxiliaries, numerals, personal pronouns, compounding, affixation, and phonology. In this paper I shall illustrate some of their diversity with examples from the less well known Formosan languages whenever possible.

## 2. Diversity of Formosan languages

### 2.1 Word order

Word order varies to a great extent in having VSO, VOS, or SVO in Austronesian languages: VSO or VOS in most Formosan and western Austronesian languages, and SVO in east Indonesia and many other languages elsewhere in the Pacific. All these variations of word order are found in Formosan languages. There is a preferred word order in each language. Roughly speaking, in languages with explicit case markers before nouns, such as AtaMx<sup>1</sup> (1) (Huang 1995c:23), Amis (2)-(3), Kavalan (4), and Paiwan (5), the order of S and O after V is relatively free, while languages without or with few case markers, such as Saisiyat (6)-(9) and AtaSq, the order of S and O is more rigid. Sentences with personal pronouns or common nouns may manifest different word order constraints, depending on the individual languages, such as (10) in Seediq below. Similarly, a language may prefer a different order of S and O for Agent-focused constructions or non-Agent-focused constructions, as in Amis (2)-(3). In fact, slight variations in word order may indicate subtle semantic differences in these languages. See Aldridge (2004) for further discussion of word order variations in Austronesian languages.

Most Formosan languages are verb-initial, VOS or VSO. But Saisiyat (6)-(8), Pazih (11a) and Thao (12c) also have SVO order, due to strong Chinese influence. The languages with predominantly VOS order include Tsou (13a), Paiwan (5b), AtaMx (1a), and Seediq (10a); the languages with predominantly VSO order include Bunun (14),

---

<sup>1</sup> Abbreviations used in this paper are as follows: AF: Agent-focus; ASP: aspect; AtaMx: Mayrinax dialect of Atayal; AtaSq: Squliq dialect of Atayal; AUX: auxiliary; BF: Beneficiary-focus; CAU: cause; CM: case marker; DEF: definite; EXCL: exclusive; GEN: genitive; HUM: human; IF: Instrumental-focus; INCL: inclusive; LF: Locative-focus; LIG: ligature; LOC: locative; NEU: neutral; NOM: nominative; OBL: oblique; PASS: passive; PF: Patient-focus; PL: plural; PRF: perfective; PROG: progressive; Q: question; REC: reciprocal; RED: reduplication; RF: Referential-focus; SG: singular; TOP: topic.

Favorlang (15), and Basay (16); and the languages with either VSO or VOS order include Puyuma (17), Rukai, and Kavalan (4). In addition to verb-initial constructions in spontaneous speech, as in free running texts, there are also subject-initial constructions in elicited sentences in Saisiyat, Thao, and Pazih.

Word order may be determined by semantic roles (Agent and Patient) rather than by grammatical functions (nominative, genitive, oblique, etc.) in a few Formosan languages, such as Amis, in which the Agent always precedes the Patient, as in (2) and (3) below.

## Atayal Mayrinax

- (1) a. ma-nubuag cku' hani ku' kanayril.<sup>2</sup> (VOS)  
 AF-drink OBL this NOM woman  
 'The woman drank this.'  
 b. nanuan ku' taal-an nku<sup>3</sup> 'ulaqi'? (VSO)  
 what NOM see-LF GEN child  
 'What did the child see?'

## Amis

- (2) a. mi-patay ko cahiw to wawa. (VSO)  
 AF-kill NOM hunger OBL child  
 'Hunger killed a child. = A child was starved to death.'  
 b. cahiw-en no ina ko fafoy. (VOS)  
 Starve-PF GEN Mom NOM pig  
 'The pig was starved by Mother.'  
 (3) a. mi-palo ko mama to wawa. (VSO)  
 AF-hit NOM Dad OBL child  
 'Father hit a child.'  
 b. palo-en no mama ko wawa. (VOS)  
 hit-PF GEN Dad NOM child  
 'The child was hit by Father.'

## Kavalan

- (4) a. sim-subut tu baqian a sunis. (VOS)  
 REC-bow OBL old.man NOM child  
 'The child bowed to an old man. = The child and old man bowed to each other.'

<sup>2</sup> All examples for the extant Formosan languages are based on my field notes unless indicated otherwise. Examples for the extinct languages are based on written documents.

<sup>3</sup> The label O is interpreted in a broad sense: It refers not only to oblique, but also to genitive and locative, namely any case other than nominative.

- b. si-qnabil=ti ya kbaran tu maytumal. (VSO)  
become-enemies=ASP NOM Kavalan OBL Taroko  
'The Kavalan people became enemies with the Taroko.'

Paiwan

- (5) a. k<em>a-kan a kaLang tua velvel. (VSO)  
RED-AF-eat NOM monkey OBL banana  
'The monkey is eating a banana.'  
b. si-kesa tua vurasi na kina a kakDian. (VOS)  
BF-cook OBL potato GEN Mom NOM child  
'Mother cooked sweet potatoes for the child.'

Saisiyat Ta'ay

- (6) tatini' masay=ila. (SV)  
old die=ASP  
'The old man has died.'  
(7) 'oya' ma-yakai' ka korkoring Sa'ila pakakSia'. (SVO)  
Mom AF-tell OBL child go play  
'The mother told the child to go and play (by himself).'  
(8) ka korkoring Sibil-in ray taLoe'aen. (SVO)  
DEF Child leave-PF LOC home  
'The child was left at home.'  
(9) m-waLi'=ila ka Loko'. (VS)  
AF-come=ASP NOM loom  
'The loom has arrived.' (1-9)

Notice that Saisiyat tends to have the subject before the verb, due to the heavy influence from its close contact with the Chinese. However, many verb-initial sentences also appear in the free-running texts.

Seediq Paran

- (10) a. pusu na wada s<um>ipaq tanah tunux mona rudo. (VOS)  
begin AUX kill<AF> red head name name  
'At the outset Mona Rudo killed the Japanese.'  
b. m-ekan miyan ido, cukau miyan aga. (VSO)  
AF-eat IPL.NOM rice use IPL.NOM hand  
'When we ate rice, we used our hands.'

Pazih

- (11) a. yaku kasibat pazih a rahan. (SVO)  
 ISG.NOM teach pazih LIG language  
 ‘I teach the Pazih language.’
- b. ma-laleng xani-xanisay a kawas, liaka partisan  
 AF-live RED-several LIG year then give.birth  
 adang a rakihan ki kalayu. (VOS)  
 one LIG child NOM name  
 ‘After living for several years, Kalayu had a baby.’

Thao

- (12) a. shkuda nak a shnaw maqa uka-wan sa taun. (VS)  
 ache my LIG heart because not.have-yet CM house  
 ‘I feel bad because I don’t have a house yet.’
- b. maka-rishkish yamin malh-kakrikriw. (VS)  
 MAKA-diligent IPL.EXCL AF-work  
 ‘We work hard.’
- c. yaku inshiraq-in ama. (SVO)  
 ISG.NOM scold-PF Dad  
 ‘I was scolded by Father.’

Tsou

- (13) a. pei'i ta ucey 'e ino. (VOS)  
 AF.cook OBL taro NOM mother  
 ‘The mother cooked taros.’
- b. i-si p-‘an-i si oko ta ino-si. (VSO)  
 AUX-3SG.GEN CAU-eat-LF NOM child OBL mother-his  
 ‘The child was fed by his mother.’

Bunun Isbukun

- (14) ma-saiv tama zaku sui tu mas'an. (VSO)  
 AF-give Dad 1SG.OBL money LIG ten  
 ‘Father gave me \$10.’

Favorlang

- (15) a. in-i-pa-ita ja aukat o Deos pana torro  
 PRF-1SG-CAU-see NOM love OBL God to IPL.INCL.OBL  
 boa. (VSO)  
 person  
 ‘God’s love was shown to our person.’

- b. alli mamchap ijonoe. (VS)  
not shall steal 2PL.NEU  
'Thou shalt not steal.'

Basay

- (16) a. ma-kawas ta lamuan-ia,  
AF-say NOM villager-2SG.GEN  
"kuakua=ti=ita." (Text 4) (VSO)  
nothing.to.do=ASP=1PL.EXCL.NOM  
'His villagers said, "There is nothing we can do about it."'  
b. mai-mu vlai kuanan tuLnijan, mataai takna.  
not-2PL.NOM give 1SG.OBL flute die shall  
'If you don't give me the nose-flute, I shall die.' (Text 4) (VSO)

Puyuma

- (17) a. sagar=yu Da bunga? (VSO)  
like=2SG.NOM OBL potato  
'Do you like sweet potatoes?'  
b. m-ekan Da kuLabaw na ngiaw. (VOS)  
AF-eat OBL mouse NOM cat  
'The cat ate a mouse'

Formosan languages are predominantly head-initial in both verbal and nominal constructions.

## 2.2 Focus systems

Focus systems are found in all Formosan and western Austronesian languages except Rukai, which has no focus system at all. The most common type of focus system is the so-called "Philippine type": AF *\*-um-*, PF *\*-en*, LF *\*-an*, and RF *\*Si-* (or *\*Sa-*) in the indicative voice, which is found in Atayal, Seediq, Saisiyat, Paiwan, and Amis. Yet divergent types of focus systems are found in Formosan languages, as shown below:

**Table 1:** Focus systems in Formosan languages

	AF	PF	LF	RF
Philippine type	-um-	-en	-an	Si-
Tsou	-m-	-a	-i	-(n)eni
Puyuma	-em-	-aw	-ay	-anay
Bunun	m(a)-	-un	-an	'is-
Pazih	mu-	-en	-an	sa~saa-
Amis	-um-	-en, ma- mi-...-an	-an	sa-
Kavalan/Basay	-m-	-an, ma-	Ø	ti-
Siraya	-m-, m-	-en, -an	Ø	

Notice that non-Agent focus markers of verbs look quite different, especially in Tsou and Puyuma. They are, in fact, the non-indicative voice markers in the other languages, but they function as indicative in these two languages. Such drastic morphosyntactic changes must have taken place in these two languages if we assume the common types of focus systems as the proto-type, as reconstructed by Ross (1995:739), and as generally found in many other Formosan and Philippine languages.

The Benefactive-focus marker is *si-* or *sa-* in many Formosan languages, *-(n)eni* in Tsou, *-anay* in Puyuma, *ti-* in Kavalan, but *'is-* in Bunun, as in (25).

Tsou

- (18) a. mo bon~~u~~ ta cnum~~u~~ si amo.  
 AUX AF-eat OBL banana NOM Dad  
 'The father is eating a banana.'  
 b. cuma na i-si an-a ta amo?  
 what NOM AUX-3SG.GEN eat-PF OBL Dad  
 'What is Father eating?'
- (19) a. sia na mo baito ta oko?  
 who NOM AUX AF-see OBL child  
 'Who saw a child?'  
 b. os-'o ait-i si oko.  
 AUX-1SG.GEN see-LF NOM child  
 'I saw the child.'
- (20) t'es-**neni** ta ino 'e amo.  
 sew-BF OBL Mom NOM Dad  
 'Mother sewed clothes for Father.'

Puyuma

- (21) a. m-ekan Da kuLabaw na ngiaw.  
 AF-eat OBL mouse NOM cat  
 ‘The cat ate a mouse.’  
 b. iDiyu na kuLabaw i, ki-kan Da ngiaw.  
 that NOM mouse TOP PASS-eat OBL cat  
 ‘That mouse was eaten by a cat.’
- (22) ku=pukpuk-aw iDu na walak.  
 ISG.GEN=beat-PF that NOM child  
 ‘That child was beaten by me.’
- (23) a. iDi na walak i, aTebung Da paysu.  
 this NOM child TOP AF.pick.up OBL money  
 ‘As for this child, he picked up some money.’  
 b. iDi na paysu i, tu=aTebung-ay Da walak.  
 this NOM money TOP 3SG.GEN=pick.up-PF OBL child  
 ‘As for this money, it was picked up by a child.’
- (24) ku=kibuLas-anay Da kabung. (cf. Tan 1997:53, (24b))  
 ISG.GEN=borrow-BF OBL hat  
 ‘I borrowed a hat for him.’

In addition to a focus system similar to most other Formosan languages and western Austronesian languages, Puyuma has also a passive marker *ki-*, as in (21b) above, just like Rukai (Li 1973:193).<sup>4</sup> However, it has a focus system of its own peculiarity, with quite different markers for non-Agent- focus, *-aw*, *-ay*, and *-anay* on the verbs, as in (22)-(24), from all the other Formosan languages.

Bunun Isbukun

- (25) a. m-aun 'uvaz-a' mas tai'.  
 AF-eat child-NOM OBL taro  
 ‘That child ate a taro.’  
 b. kaun-un mas 'uvaz-cia' tai'.  
 eat-PF OBL child-OBL taro  
 ‘The taro was eaten by that child.’
- (26) 'is-'anat cina-cia' 'uvaz-a' mas tai'.  
 BF-cook Mom-OBL child-NOM OBL taro  
 ‘Mother cooked taros for the child.’

<sup>4</sup> The prefix *ki-* ‘passive’ has restricted usage in Puyuma, and so does the personal marker *-li* ‘my’ in Puyuma. These two instances indicate that there is some linguistic influence from Rukai.



Unlike the majority of Formosan and western Austronesian languages, Bunun has no infix *-um-* for its Agent-focus verbs. Moreover, the Referential-focus marker in Bunun is *'is-*, rather than *si-* or *sa-*, as commonly found in the other Formosan languages.

## Pazih

- (27) a. **mu**-baket rakihan ki aba.  
 AF-beat child NOM Dad  
 'The father beat a child.'
- b. baked-en ni aba ki rakihan.  
 beat-PF GEN Dad NOM child  
 'The child was beaten by Father.'
- (28) pa-batu'-an lia ki babaw daran.  
 pave-stone-LF ASP NOM above road  
 'The surface of the road was paved with stones.'
- (29) a. **saa**-xe'et nuang ki saris.  
 IF-tie cow NOM string  
 'The string was used to tie a cow.'
- b. saa-talek alaw ki bulayan  
 IF-cook fish NOM pan  
 'The pan was used to cook fish.'

Unlike the majority of Formosan languages and western Austronesian languages, but similar to Bunun, Pazih has only the prefix *mu-* or its variants for Agent-focus marker on the verb, but no infix *-um-*.

## Kavalan

- (30) a. s<m>aqut=ti=iku tu paRin.  
 AF-move=ASP=1SG.NOM OBL wood  
 'I have transported firewood.'
- b. saqut-**an**-ku=ti ya paRin.  
 move-PF-1SG.GEN=ASP NOM wood  
 'The firewood has been transported by me.'
- c. **ti**-saqut-ku tu paRin ya qRitun.  
 IF-move-1SG.GEN OBL wood NOM car  
 'I transported wood with the car.'

Patient-focus and Locative-focus have merged as *-an* in Kavalan. Moreover, the Referential focus marker is *ti-* rather than the anticipated *si-*.

### 2.3 Auxiliaries

An auxiliary is generally required for each clause in Tsou, whereas there is no auxiliary at all in a few languages, such as Pazih or Kavalan.<sup>5</sup> There are also languages that have only a few auxiliaries, such as Atayal or Seediq. Moreover, the auxiliary and main verb of each clause must agree in focus in Tsou, whereas there is no such requirement in the other Formosan languages.

Each clause has an auxiliary before the main verb in Tsou:

- (31) a. m-i-ta                      eobako    ta    oko    'e    taini.  
 AF-AUX-3SG.NOM    AF.beat    OBL    child    NOM    3SG  
 'He is beating a child.'
- b. i-ta                              eobak-a    'e    taini.  
 PF.AUX-3SG.GEN    beat-PF    NOM    3SG  
 'He was beaten by him.'

There are two common auxiliaries, *nyux* and *cyux*, in the Sqliq dialect of Atayal, but they are not required for each clause:

- (32) a. nyux    saku'              m-anig    qulih.  
 AUX    1SG.NOM    AF-eat    fish  
 'I am eating fish.'
- b. cyux    'inu'    laqi'    =su'?  
 Aux    where    child    2SG.GEN  
 'Where is your child?'
- c. q<n>at-an    hozil    qu'    laqi'    qani.  
 bite-PRF-LF    dog    NOM    child    this  
 'This child was bitten by a dog.'

Similarly there are two common auxiliaries, *gaga* and *wada*, in the Paran dialect of Seediq, but they are not required for each clause:

- (33) a. gaga              t<m>aus    dangi-na              tama-mu.  
 AUX.PROG    beckon-AF    friend-3SG.GEN    Dad-1SG.GEN  
 'My father is beckoning to his friend.'

<sup>5</sup> Zeitoun (p.c.) also does not believe that there are any auxiliaries in Paiwan. Auxiliaries attract pronouns, as in Tsou (31), Atayal (32a), and Seediq, whereas *uri* in Paiwan, which appears sentence-initially before the main verb, does not.

- b. wada            s-bege    pila        laqi-na.  
 AUX.past    BF-give    money    child-3SG.GEN  
 ‘The money has been given to his child.’
- c. b<n>ege        na        bubu-na            pila        ni.  
 give-PRF    GEN    Mom-3SG.GEN    money    this  
 ‘This money was given by his mother.’

In short, there is a great deal of variation in the auxiliaries of Formosan languages: From languages without any auxiliary, as in Pazih and Kavalan, to a language that requires an auxiliary in each clause, as in Tsou.

## 2.4 Numeral systems

All Formosan languages, except Saisiyat, Pazih, and Favorlang, distinguish human vs. nonhuman in their numeral systems.<sup>6</sup> Many of them indicate the distinction by *Ca*-reduplication, while the others indicate it by a prefix (*ta-* in Rukai, *ma-* in Paiwan, *mi(a)-* in Nanwang dialect of Puyuma,<sup>7</sup> *kin-* in Kavalan) or by alternation of vowels and/or consonants, as in the Tsouic languages (Li 2006). Western Austronesian languages either indicate the distinction by *Ca*-reduplication, such as Yami, or have lost it, such as Ivatan (Tsuchida et al. 1987:161-163). It is clear that Formosan languages are much more diverse than western Austronesian languages. However, different sets of numerals and classifiers must be chosen according to the type of object being counted in Oceanic languages, such as Mokilese (Harrison 1976:93-108). This type of complexity is not found in Formosan languages.

Siraya

- (34) a. **ra**-rouha    ki        voual. (St. Matthew 24:40)  
 HUM-two    OBL    field  
 ‘Two (men) will be in the field.’
- b. ro-rouha    ka        pani. (St. Matthew 25:17)  
 RED-two    LIG    another  
 ‘(He gained) another two (talents).’

<sup>6</sup> Both Saisiyat and Pazih belong to the “Northwest” subgroup (Blust 1999), while all these two languages plus Favorlang belong to the “Northwestern” subgroup (Li 1985). If so, the loss of distinction must be an innovation in this subgroup.

<sup>7</sup> In the other dialects of Puyuma, such as Tamalakaw (Tsuchida 1980:287), human numerals are indicated by *Ca*-reduplication. In fact, there are still a few residual forms of *Ca*-reduplication indicating human numerals, e.g., *mi-Da-Dua* ~ *mi-a-Dua* ‘two people,’ *mi-ta-teLu* ~ *mi-a-teLu* ‘three people’ even in Nanwang. Thus *mi(a)-* is an innovation in Nanwang.

Basay (*nia-* + number ‘human’)

- (35) a. sulaw-an-ia          apun-ia          zusa. (Text 1.1)  
 hunt-LF-3PL.GEN    deer-3PL.GEN    two  
 ‘They hunted two deer.’
- b. masang    sipalu    **nia**-zusa. (Text 5)  
 formerly    friend    HUM-two  
 ‘In the past there were two friends.’
- (36) a. qumnipa    ngazi          **nia**-zusa-na    vakvakia    tsuwai. (Text 1.1)  
 formerly    stone.god    HUM-two-ASP    old.man    this  
 ‘In the past there were two people, a stone god and this old man.’
- b. aui-na    zusa    ta    tiquL-na    mulin. (Text 7)  
 go-ASP    two    NOM    stuck-ASP    iron  
 ‘(He) went (to the field) but was stuck by iron for two days.’

Favorlang

- (37) a. ja    paga    narroa    da    boa    inni    icho-şar,    so-o    bo'o  
 Q    be    two    LIG    person    in    3SG-OBL    and    person  
 Deos-şar? (Text 13)  
 God-OBL  
 ‘Are there two persons in Him, the person of a man and the Person of God?’
- b. narroa    da    tупpo:    aukat    o    Deos,    so-o    aukat    o  
 two    LIG    thing    love    OBL    God    and    love    OBL  
 ma-achaddik    o    cho. (1<sup>st</sup> sermon)  
 neighbor    OBL    people  
 ‘Two things: to love God and to love our neighbour.’
- (38) a. numm'ija    maunis    a    tataap    o    natorroa    da    bo'o  
 what    peculiar    work    OBL    three    LIG    person  
 mini? (Text13)  
 these  
 ‘What is the peculiar special work of these three people?’
- b. tупpa    natorroa    da    tупpono    mini. (Text 10)  
 remember    three    LIG    thing    these  
 ‘Remember these three things.’

As illustrated above, Siraya indicates human by *Ca-* reduplication, and Basay by the prefix *nia-*, similar to Kavalan (by *kin-*), its closely related language, whereas there is no such distinction in Favorlang, a dialect of Babuza, an extinct western plains language (Li 2003).

Most Formosan languages employ the decimal system or modified decimal systems, while Pazih nearly has a quinary system (see the tables below). The larger numerals in Formosan languages are derived from the smaller base numerals by addition and/or multiplication, occasionally by subtraction, but never by division, just as in extra-Formosan languages. The smaller base numerals include 3, 4, 5, and 6.

**Table 2:** Numeral systems in Formosan languages

Language	Numerals
Pazih	1, 2, 3, 4, 5, 5+1, 5+2, 5+3, 5+4, 10
Saisiyat	1, 2, 3, 4, 5, 6, 6+1, 2x4, 10-1, 10
Atayalic	1, 2, 3, 4, 5, 2x3, 7, 2x4, 9, 10
Thao	1, 2, 3, 4, 5, 2x3, 7, 2x4, 10-1, 10
Taokas/Babuza	1, 2, 3, 4, 5, 6, 7, 2x4, 10-1, 10

**Table 3:** Numeral forms in Formosan languages

	1	2	3	4	5	6
Pazih	ida	dusa	turu	supat	xasep	xaseb-uza
Saisiyat	'aehae'	roSa'	toLo'	Sepat	Laseb	SayboSiL
Atayal	qutux	'usa-ying	tu-gal	sapaat	ima-gal	ma-tuu'
Thao	taha	tusha	turu	shpat	rima	ka-turu
Taokas	taanu	dua	turu	lupat	hasap	takap
Babuza	na-ta	na-roa	na-torro-a	na-spat	na-hup	na-taap
	7	8	9	10		
Pazih	xaseb-i-dusa	xaseb-i-turu	xaseb-i-supat	isit		
Saisiyat	SayboSiL o 'aehae'	ka-Spat	Lae-'hae'	langpez		
Atayal	pitu	ma-spat	maqisu'	magalpug		
Thao	pitu	ka-shpat	ta-na-thu	makthin		
Taokas	pitu	maha-lpat	ta-na-so	ta-isit		
Babuza	na-ito	maa-spat	na-ta-xa-xo-an	tsixit		

All sorts of numeral systems in the entire Austronesian language family are found in Formosan languages, even though there are only a small number of languages on this island.

The Tanan dialect of Rukai, however, differs from most other Formosan languages in that it distinguishes animacy together with humanness in its numerals, marking both human beings and the other animals with the same prefix *ta-*, while ordinary objects are marked by zero. See Li (2006) for further details.

## 2.5 Personal pronouns

Each Formosan language usually has three or four sets of personal pronouns: Nominative, genitive, locative and oblique or accusative (Li 1997). A few languages have even more, by distinguishing between genitive and possessive, as in Saisiyat<sup>8</sup> and Amis. A set of pronouns may be called “neutral” (Huang 1989, 1995c:128), “free” (Zeitoun 1992) or “topic” (Li 1978, 1997, Zeitoun 2007), which occurs freely in the nominative, oblique, or topic position. Personal pronouns may be free (long) or bound (short) forms, both of which may appear in nominative and/or genitive sets. A language may have only free personal pronouns, as in Saisiyat; mostly free, as in Thao, which has only one bound pronominal form *-k ~ -ik ~ -ak* ‘1SG.GEN;’ or mostly bound, as in the Mantauran dialect of Rukai, which has only a set of free forms that appear in the topic position (Zeitoun 2007:284). There is no bound nominative third person pronoun in any Formosan language except Tsou, which does have *-ta* ‘3rd, Nom,’ as in (31a) above and (39) below:

- (39) m-i-ta                      mo-fti'i    ta    pangka.  
 AF-AUX-3SG.NOM    AF-jump    OBL    table  
 ‘He jumped to a table.’

When two personal pronouns co-occur in sequence, the order varies to a great extent. The order is determined by the functions of the pronouns, as in Kavalan (40), in which the genitive always precedes the nominative. However, it is a little more complex in Seediq: The nominative precedes the genitive, as in (41a), while the nominative precedes the oblique, as in (41b).

Kavalan

- (40) tayta-an-ku=isu.  
 see-PF-1SG.GEN=2SG.NOM  
 ‘I saw you.’

Seediq

- (41) a. kuxun=su=mu.  
 love.PF=2SG.NOM=1SG.GEN  
 ‘You are my love. = I love you.’  
 b. s<un>kuxun=ku    isu.  
 AF-love=1SG.NOM    2SG.OBL  
 ‘I love you.’

<sup>8</sup> Saisiyat is one of the few Formosan languages that manifest the same set of pronouns as that of case markers (Zeitoun, p.c.).

As in Kavalan, the genitive usually precedes the nominative in the Mayrinax dialect of Atayal. However, the genitive pronoun *niam* ‘we (exc)/Gen’ usually follows a nominative:

Atayal Mayrinax

- (42) a. *baiq-ay=mi=su' cu' pila'.*  
 give-LF=1SG.GEN=2SG.NOM OBL money  
 ‘I shall give you some money.’
- b. *ba-bahiy-un=si' cami.*  
 RED-hit-PF=2SG.GEN 1PL.EXCL.NOM  
 ‘You will hit us.’
- (43) a. *baiq-ay=su'=niam cu' pila'.*  
 give-LF=2SG.NOM=1PL.EXCL.GEN OBL money  
 ‘We (excl) shall give you (sg) some money.’
- b. *baiq-ay=cimu=niam cu' pila'.*  
 give-LF=2SG.NOM=1PL.EXCL.GEN OBL money  
 ‘We (excl) shall give you (pl) some money.’

Apparently the order is determined by more than one single factor, as discussed in Huang (1989, 1995c), Chang (1999), and Liao (2005).

## 2.6 Compounding

Formosan languages have from little compounding to rich compounding. The languages with little compounding include Pazih, Thao, and Atayal, while the languages with rich compounding include Tsou, Bunun, and Siraya.

Tsou has unusually rich compounding, with hundreds of compounds, a very high percentage of the vocabulary, as compared with other Formosan and western Austronesian languages (Huang & Tanangkingsing 2005). Tung (1964:196-213) devotes a whole chapter to an analysis of it. Each Tsou compound is composed of two free forms, either or both of which may be shortened, e.g., *meobango* ‘to chase’ + *uafeihni* ‘to follow’ > *meofeihni* ‘to run after’ (Tung 1964:198). There are three main types of compounds, as based on the elements of their formation: (1) N+N, e.g., *'ua* ‘deer’ + *chumu* ‘water’ > *'uachumu* ‘buffalo,’ (2) V+N, e.g., *ea* ‘to search’ + *fou* ‘meat’ > *eafo* ‘to hunt,’ (3) V+V, e.g., *ma'mi* ‘to think’ + *asonə* ‘probably’ > *ma'sonə* ‘to guess’ (Tung 1964:200-203).

Compound words are italicized in the following examples:

Tsou

(44) *te-'o pasu-mameoi* (< pasunaeno-mameoi).

AUX-1SG.NOM sing-old

'I shall sing like an old man.'

(45) *i-si sivie-i ta ino-si si mo tma'-congo*

AUX-3SG.GEN carry-PF OBL Mom-3SG.GEN NOM AUX fall-pain

*ci oko.* (tma'- < tmadə 'to comply with')

LIG child

'The sick child is carried by his mother on her back.'

Bunun

(46) *mis-utmang mis-busuk.* (Nojima 1996)

burn-careless burn-drunk

'(he) carelessly became drunk'

Siraya

(47) *ka ara-auu ta t'hing ka mou-ouro mou-vavau.*

and take-PF NOM fish LIG come.first.AF come.up.AF

'And take up the fish that comes up first.' (17-27)

Notice that there may be prefix harmony of the compounds that appear in the same clause, as in Bunun and Siraya.

## 2.7 Affixes

A language may have a rich repertoire of prefixes, such as Tsou (Tsuchida 1990) and Thao (Blust 2003:91), each of which has hundreds of prefixes. On the other hand, another language may have only a few dozen prefixes, such as Pazih (Li & Tsuchida 2001:10-19) and Kavalan (Li & Tsuchida 2006:14-24).

Morphologically, in addition to the two infixes, *-um-* and *-in-*, commonly found in Formosan and western Austronesian languages, there are at least two more infixes *-ar-* and *-al-*, as attested in a few fossilized forms in some Formosan languages,<sup>9</sup> e.g. *q<ar>afqaf* 'house (arch.)', *b<ar>imbin* 'vehicle', *b<ar>umbun* 'thunder', *k<ar>ungkun* 'to wind', *sh<ar>inshin* 'bell', and *b<al>umbun* 'bell' in Thao; *k<ar>makmaz* 'to blink', *t<r>aqitaq* 'talkative', *p<r>inipin* 'to walk unsteadily', *k<r>awkaway* 'to work', *b<aR>qian* 'old people' (<*baqi* 'grandpa'), *t<R>abtab* 'to eat and make noise like a pig' and *q<R>ezqez* 'stable' in Kavalan; *s<a>ungusung* 'is counting' and *p<a>isuzuk* 'is hiding' in Pazih; *k<ar>otkot* in Siraya. Infixes of <*ar>*,

<sup>9</sup> These two infixes also appear in fossilized forms in Philippine languages (Tsuchida, p.c.).



<al> and <alj> occur in quite a few lexical forms in Paiwan; see Ferrell (1982:16) for examples.

When *-um-* and *-in-* co-occur, they appear in that order, *-um-in-* in most languages, such as Atayal (48) and Saisiyat (49), while the reverse order is found only in a few languages throughout the entire Austronesian language family, such as *-in-um-* in Bontok and Ilokano (Philippines, Reid 1992), Rungus Dusun and Kimaragang Dusun (in Sabah, Beatrice Clayre, p.c.), Kananavu (51), and Favorlang (50), an extinct western plain language in Taiwan (Li 2003). That is to say, both types of order are found in Formosan languages:

Atayal Mayrinax

- (48) g<um-in>aliq cu' ruas ku' 'ulaqi'.  
 tear-AF-PRF OBL paper NOM child  
 'The child has torn some paper.'

Saisiyat

- (49) s<om-in>'iael=ila pa-ki-kita' si-sa'i'.  
 eat-AF-PRF=ASP CAU-RED-see SI-defecate  
 'When he had eaten, he made them watch him defecate.'

Favorlang

- (50) ts<in-um>agach i boesum, airo-ossen i kallamas  
 ascend-PRF-AF in heaven sit on right  
 o choa tamau, kamabarr'ija tapos o ai.  
 OBL 3SG.GEN father Almighty all OBL that  
 'He ascended to Heaven and sat on the right hand of His Father, the Almighty.'

Kanakanavu

- (51) c<in-m>e'era=ku vunai.  
 see-PRF-AF=1SG.NOM snake  
 'I saw a snake.'

## 2.8 Phonology

Phonologically, Formosan languages manifest many different points of articulation, including dental vs. retroflex stops and lateral (*t, d, l* vs. *T, D, L*), as in Puyuma, Paiwan and Rukai; palatal vs. nonpalatal stops and liquid, as in Paiwan (*t, d, l* vs. *tj, dj, lj*); uvular stop vs. velar stop (*q* vs. *k*), as in Atayal, Seediq, Thao, Bunun, and Paiwan; velar vs. pharyngeal fricative (*x* vs. *h*), as in Atayal, Seediq, Amis and Pazih; pharyngeal stop

and fricative, as in Amis (*q, h*); and uvular fricative, as in Kavalan (*R*). There are three phonemically distinct liquids in some languages, as in Rukai (*l, L, r*), Puyuma (*l, L, r*), Paiwan (*l, r, lj*), Thao (*l, r, lh*), and Saaroa (*l, r, lh*). In manner of articulation, there is a contrast between voiced and voiceless lateral, as in Thao and Saaroa. In many Formosan languages, all or most of their consonants may occur word-finally. There are consonant clusters in word-initial and medial positions in several languages, namely Thao, Tsou and the Maga dialect of Rukai. There are also geminate consonants in Maga and Kavalan (Li & Tsuchida 2006:5-6).

There are many different types of morphophonemic alternations in Formosan languages, especially the Sqliq dialect of Atayal, Seediq, and Maga dialect of Rukai; see Li (1977) for details.

### 3. Concluding remarks

We have examined several linguistic aspects of Formosan languages, all of which manifest great diversity. (1) All variations of word order in Austronesian languages, VSO, VOS, and SVO, are found in Formosan languages. (2) All western Austronesian languages have the so-called “focus systems” (or voice systems). Formosan languages do have the Philippine type of focus system, *\*-um*, *\*-en*, *\*-an* and *\*Si-*, commonly found in western Austronesian languages. They also have various other types, including non-indicative mood used as indicative mood, as in Tsou and Puyuma. Moreover, Rukai has no focus system at all, and it is a straightforward accusative language rather than ergative. (3) Every clause requires an auxiliary in Tsou, whereas not a single auxiliary has been found in Pazih or Kavalan. Moreover, the auxiliary and the main verb of each clause must agree in focus in Tsou, yet there is no such requirement in other Formosan languages. (4) Most Formosan languages distinguish between human and nonhuman in their numeral systems, as in many western Austronesian languages, yet a few Formosan languages (Saisiyat, Pazih, and Favorlang) have lost such a distinction. Many of them indicate the distinction by *Ca-* reduplication, while the others do so by a prefix or by alternation of vowels and/or consonants, not reported in western Austronesian languages. Furthermore, Kakanavu and Tanan Rukai differ from other Formosan languages by distinguishing between animate and inanimate. Most Formosan languages employ the decimal system or modified decimal systems, while Pazih has nearly a quinary system. The larger numerals in some Formosan languages are derived from the smaller base numerals by addition or multiplication, occasionally by subtraction, just as in extra-Formosan languages. In fact, all sorts of numeral systems in the entire Austronesian language family are found in Formosan languages. (5) All personal pronouns are free in Saisiyat, mostly free in Thao, yet they are mostly bound in Mantauran Rukai. Most Formosan

languages have no nominative third personal pronoun, yet Tsou does. (6) Formosan languages have from little compounding, as in Pazih, Thao, and Atayal, to unusually rich compounding, as in Tsou, Bunun, and Siraya. (7) One language may have a rich repertoire of prefixes, as in Tsou and Thao, while another language may have only a few dozen prefixes, as in Pazih and Kavalan. When the two infixes *-um-* and *-in-* co-occur, they appear in that order, *-um-in-*, in most languages, while the reverse order is found only in a few languages in the entire Austronesian language family. Both types of order are found in Formosan languages. In addition to these two infixes, two more infixes, *-ar-* and *-al-*, are attested in a few fossilized forms in some Formosan languages. (8) Phonologically Formosan languages manifest many different points of articulation, including uvular and pharyngeal, and in manner of articulation there is a contrast between voiced and voiceless lateral.

It is clear that, from the preceding discussion, Formosan languages are very diverse at every linguistic level, from phonology to morphology to syntax. I have discussed only eight types of linguistic features. One can go on and on and then write up a monograph on this topic. We do not find such a great diversity in any area of the extra-Formosan languages, not in the Philippines, Malaysia, or Indonesia.

## References

- Adelaar, Karl Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36.2:362-397.
- Aldridge, Edith Catherine. 2004. *Ergativity and Word Order in Austronesian Languages*. Ithaca: Cornell University dissertation.
- Asai, Erin. 1937. Basai. Field notes.
- 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 & Paul Jen-kuei Li, 31-94. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert. 2003. *Thao Dictionary*. Language and Linguistics Monograph Series A5. Taipei: Institute of Linguistics, Academia Sinica.
- Campbell, Rev. William. 1888. *The Gospel of St. Matthew in Formosan (Sinkang dialect), with Corresponding Versions in Dutch and English Edited from Gravius's Edition of 1661*. London: Trübner.
- Campbell, Rev. William. 1896. *The Articles of Christian Instruction in Favorlang-Formosan Dutch and English: from Vertrecht's Manuscript of 1650*. London: Kegan Paul, Trench, Trübner.
- Chang, Henry Yungli. 1997. *Voice, Case, and Agreement in Seediq and Kavalan*. Hsinchu: National Tsing Hua University dissertation.
- Chang, Henry Yungli. 1999. Seediq bound pronouns: pronominal clitics or agreement affixes? *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 355-370. Taipei: Institute of Linguistics, Academia Sinica.
- Ferrell, Raleigh. 1982. *Paiwan Dictionary*. Pacific Linguistics C-73. Canberra: The Australian National University.
- Harrison, Sheldon P. 1976. *Mokilese Reference Grammar*. Honolulu: University of Hawaii Press.
- Huang, Lillian M. 1989. The pronominal system in Atayal. *Studies in English Literature and Linguistics* 15:115-133.
- Huang, Lillian M. 1995a. The case markers and pronominal system in Amis. *Journal of National Chengchi University* 7:217-257.
- Huang, Lillian M. 1995b. The voice system in Atayal. *Papers from the First International Symposium on Languages in Taiwan*, ed. by Feng-fu Tsao & Mei-hui Tsai, 73-98. Taipei: Crane.
- Huang, Lillian M. 1995c. *A Study of Mayrinax Syntax*. Taipei: Crane.

- Huang, Lillian M. 1996. A syntactic and semantic study of *-ay*, *-aw* and *-anay* in Mayrinax Atayal. *Studies in English Literature and Linguistics* 22:15-36.
- Huang, Shuanfan, and Michael Tanangkingsing. 2005. Reference to motion events in six western Austronesian languages: toward a semantic typology. *Oceanic Linguistics* 44.2:307-340.
- Jeng, Heng-syung. 1977. *Topic and Focus in Bunun*. Special Publications No.72. Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 1973. *Rukai Structure*. Special Publications No.64. Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 1977. Morphophonemic alternations in Formosan languages. *Bulletin of the Institute of History and Philology Academia Sinica* 48.3:375-413.
- Li, Paul Jen-kuei. 1978. The case-marking systems of the four less known Formosan languages. *Second International Conference on Austronesian Linguistics: Proceedings*, Fascicle 1, ed. by Stephen A. Wurm & Lois Carrington, 569-615. Pacific Linguistics C-61. Canberra: The Australian National University.
- Li, Paul Jen-kuei. 1985. The position of Atayal in the Austronesian family. *Austronesian Linguistics at the 15<sup>th</sup> Pacific Science Congress*, ed. by Andrew Pawley & Lois Carrington, 257-280. Pacific Linguistics C-88. Canberra: The Australian National University.
- Li, Paul Jen-kuei. 1997. A syntactic typology of Formosan languages—case markers on nouns and pronouns. *Chinese Languages and Linguistics*, Vol. 4: *Typological Studies of Languages in China*, ed. by Chiu-yu Tseng, 343-378. Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 2003. Notes on Favorlang, an extinct Formosan language. *English-Favorlang Vocabulary by Naoyoshi Ogawa*, ed. by Paul Jen-kuei Li, 1-13. Asia-African Lexicon Series No.43. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Li, Paul Jen-kuei. 2004. *Selected Papers on Formosan Languages*, 2 vols. Language and Linguistics Monograph Series C3. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei. 2006. Numerals in Formosan languages. *Oceanic Linguistics* 45.1: 133-152.
- Li, Paul Jen-kuei, and Shigeru Tsuchida. 2001. *Pazih Dictionary*. Language and Linguistics Monograph Series A2. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei, and Shigeru Tsuchida. 2006. *Kavalan Dictionary*. Language and Linguistics Monograph Series A19. Taipei: Institute of Linguistics, Academia Sinica.
- Liao, Hsiu-Chuan. 2005. Another look at the order of clitic pronouns in Wulai Atayal. *Concentric: Studies in Linguistics* 31.1:47-63.

- Nojima, Motoyasu. 1996. Lexical prefixes of Bunun verbs. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]* 110:1-27.
- Reid, Lawrence A. 1992. On the development of the aspect system in some Philippine Languages. *Oceanic Linguistics* 31.1:65-91.
- Ross, Malcolm. 1995. Reconstructing Proto-Austronesian verbal morphology: evidence from Taiwan. *Austronesian Studies Relating to Taiwan*, ed. by Paul Jen-kuei Li et al., 727-791. Taipei: Institute of History and Philology, Academia Sinica.
- Sapir, Edward. 1916. *Time Perspective in Aboriginal American Culture: A Study in Method*. Ottawa: Government Printing Bureau. Reprinted in *Selected Writings of Edward Sapir in Language, Culture and Personality*, ed. by David G. Mandelbaum, 389-462. Berkeley: University of California Press, 1949.
- Starosta, Stanley. 1988. A grammatical typology of Formosan languages. *Bulletin of the Institute of History and Philology Academia Sinica* 59.2:541-576.
- Tan, Cindy Ro-lan. 1997. *A Study of Puyuma Simple Sentences*. Taipei: National Taiwan Normal University MA thesis.
- Teng, Stacy F. 1997. *Complex Sentences in Puyuma*. Taipei: National Taiwan Normal University MA thesis.
- Teng, Stacy F. 2007. *A Reference Grammar of Puyuma*. Canberra: The Australian National University dissertation.
- Tsuchida, Shigeru. 1980. Puyuma (Tamalakaw dialect) vocabulary—with grammatical notes and texts [プユマ語(タマラカオ方言)語彙—附: 語法概説みよびテキスト]. *The Peoples, Cultures and Languages of the Black Currents [黒潮の民族・文化・言語]*, 183-307. Tokyo: Kadokawa Shoten.
- Tsuchida, Shigeru. 1990. Classificatory prefixes of Tsou verbs. *Tokyo University Linguistics Papers* '89:17-52. Tokyo: Department of Linguistics, University of Tokyo.
- Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi. 1987. *Lists of Selected Words of Batanic Languages*. Tokyo: University of Tokyo.
- Tung, T'ung-ho et al. 1964. *A Descriptive Study of the Tsou Language, Formosa*. Special Publications No.48. Taipei: Institute of History and Philology, Academia Sinica. Reprinted in 1991 with Notes and Errata prepared by Paul Jen-kuei Li.
- Zeitoun, Elizabeth. 1992. *A Syntactic and Semantic Account of Tsou Focus System*. Hsinchu: National Tsing Hua University MA thesis.
- Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran (Rukai)*. Language and Linguistics Monograph Series A4-2. Taipei: Institute of Linguistics, Academia Sinica.

[Received 4 May 2008; revised 21 May 2008; accepted 6 June 2008]

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

## 台灣南島語言的多樣性

李壬癸

中央研究院

國際南島語言學界一般都認為台灣南島語言的差異性最大，遠非其他地區（包括菲律賓、馬來西亞、印尼）的南島語言所能比。一個地區的語言歧異性愈大，顯示其時代的縱深愈長，就愈有可能是該語族的起源地和擴散中心。因此，各種台灣南島語言，包括已消失的幾種平埔族語言，都值得做深入的研究。本文首次全面檢驗各種台灣南島語言各層次的多樣性，包括詞序、焦點系統、助動詞、數詞、人稱代詞、複合詞、詞綴、音韻等，並且儘量涵蓋較少人知道的語言，如巴宰、巴賽、法佛朗、西拉雅。

關鍵詞：南島語，祖居地，多樣性，語言結構，台灣南島語