The Internal Relationships of Formosan Languages

Paul Li

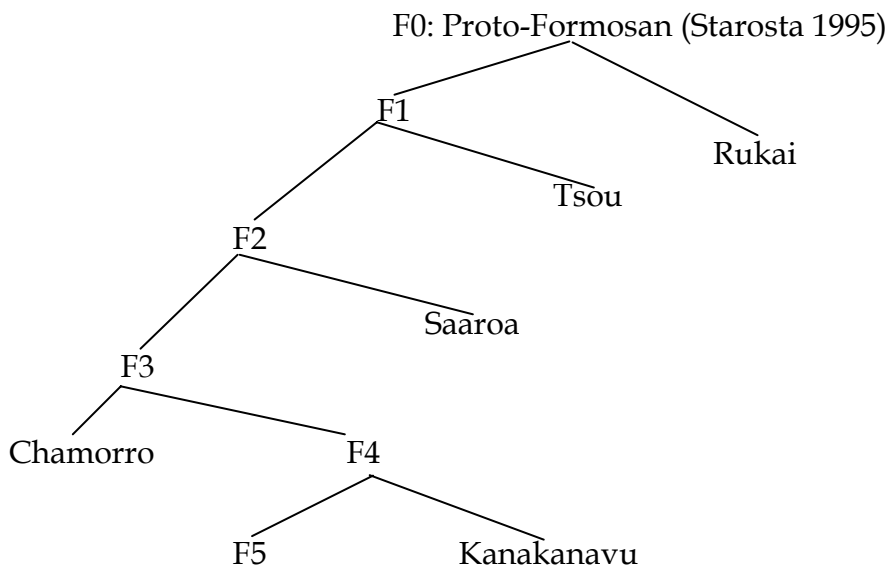Academia Sinica

1. Introduction

Different subgrouping hypotheses of the Formosan languages have been proposed, as based on different types of linguistic evidence, including (1) three main subgroups: Atayalic, Tsouic and East Formosan by Dyen (1963), as based on lexicostatistic evidence, (2) two main subgroups: Rukai and the rest by Starosta (1995), as based on morphological evidence, and (3) nine main subgroups by Blust (1999): Atayalic, East Formosan, Puyuma, Paiwan, Rukai, Tsouic, Bunun, Western Plains, and Northwest Formosan, as based on phonological evidence. Which one of the above is the most acceptable? I shall discuss problems, supporting or counter-evidence for each of the above subgrouping hypotheses, and then suggest a revised subgrouping hypothesis of my own.

An advantage of lexicostatistics is that it may give us a quick result and some clue as to what we can look for. Nevertheless, the problem of lexicostatistic classification is obvious: Mutual influence among the Formosan languages is almost inevitable, and it is not easy to distinguish between early loanwords and inherited words. Hence the percentage of cognate sets shared by each pair of languages may not be a very reliable criterion.

I have found further supporting evidence for Blust's subgroups of East Formosan (Li 2004) and Western Plains (Li 2001, 2003b). Nevertheless, the main problem with Blust's subgrouping is that there are too many subgroups. It is extremely unlikely that Proto-Austronesian would split into ten subgroups (including Malayo-Polynesian) all at once at the earliest stage. Notice that each of his four main subgroups consists of only a single language: Puyuma, Paiwan, Rukai and Bunun. We may be skeptical, especially when we consider the fact that the Japanese anthropologists could not distinguish between Puyuma, Paiwan and Rukai in an early stage of their work on the Formosan aborigines. If no phonological evidence can be found to establish a closer genetic relationship between some of the subgroups, we had better look for other types of evidence, such as morpho-syntactic. An aim of this study is to see if we can find any linguistic evidence for a closer relationship among the four languages and/or their relationship with the other Formosan languages.

Starosta's binary classification seems to be feasible and looks fine at least as a working hypothesis: The first split is Rukai, the second split is Tsou, and so forth. However, scholars who are familiar with the Southern Tsouic languages will be skeptical about the strange relationship among Saaroa, Chamorro and Kanakanavu in that Chamorro is more closely related to Kanakanavu, as shown

in his family tree. In fact, Saaroa and Kanakanavu are so closely related that mutual intelligibility is very high, as I have learned from the native speakers of both languages.

F0: Proto-Formosan (Starosta 1995)

```
                    F0: Proto-Formosan (Starosta 1995)
                   /                              \
                 F1                                \
                /   \                               Rukai
              /       Tsou
            F2
           /   \
         /       Saaroa
       F3
      /   \
  Chamorro   F4
            /    \
          F5      Kanakanavu
```

The internal relationship of the languages in the southwestern plains, Siraya, Taivoan and Makatao, will also be discussed in this paper.

2. The Great Diversity of Formosan Languages

The great diversity of Formosan languages at all levels of linguistic structures presents a great problem for classification. Focus systems are found in all Formosan and western Austronesian languages except Rukai. The most common type of focus system is the so-called "Philippine type": AF *–um-*, PF *-en*, LF *-an*, RF *Si-* (or *Sa-*), which is found in Atayal, Seediq, Saisiyat, Paiwan and Amis. Yet divergent types of focus systems are found in Formosan languages, as shown below:

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

Morphologically, in addition to the two infixes, *-um-* and *–in-*, commonly found in many Formosan and western Austronesian languages, there are at least two more infixes *-ar-* and *-al-*, as attested in a a few fossilized forms in some Formosan languages, 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>issuzuk* 'is hiding' in Pazih.   Infixes of *<ar>*, *<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 and Saisiyat, while the reverse order is found only in a few languages, such as *-in-um-* in Favorlang (Li 2003a).   Both types of order are found in Formosan languages.

3. The Relationships of Some Formosan Languages

3.1 Rukai and Puyuma

A striking syntactic feature of Rukai is that it has no focus system, as found in all other Formosan and western Austronesian languages. Focus system is generally believed to be an important and typical syntactic feature in all these Austronesian languages. However, there is no evidence that Rukai has had any focus system throughout its history. Rukai is thus believed to be the very first offshoot from Proto-Austronesian, while the other subgroup has evolved a focus system, as evidenced in all the other Formosan and western Austronesian languages (Starosta, Pawley and Reid 1982).

Nevertheless, there is some morpho-syntactic evidence that Rukai shares some features with Puyuma.   One is the passive construction indicated by the prefix *ki-*, which is found in all sorts of constructions in both Rukai (Li 1973:193-97) and Puyuma (Tan 1997:64-65, Stacy Teng, p.c.), as illustrated in (1) and (2) respectively below.   Another is the first person genitive *–li* 'my', which is also found in both languages, as illustrated in (3) and (4) below.

Rukai (1)a. wa-kane   sa   umas   kuani Likulaw.
            eat   Acc person that   leopard
              'That leopard ate a person'

        b. ki-a-kane kuani umas   sa   Likulaw.
            was-eaten that   person Acc leopard
              'That person was eaten by a leopard'

Puyuma (2)a. Dua   b<en>ekas-a   I   tugi. (Stacy Teng)
              come   <AF>interrogate-Proj. Nom   name
                'Tugi came to interrogate someone.'

          b. m-uka i   Tau-Tau,   m-uka ki-bekas-a.
              AF-go Loc Red-person AF-go Pass-interrogate-Proj
                'He went to others; he went to get interrogated.'

Rukai (3). kuani   ababay swa-swaswa' inia   daan-li.
            that/Nom woman   Red-sweep   that/Acc house-my
              'That woman is sweeping my house.'

Puyuma (4). mu-ruma la   i   nana-li.
              AF-return Asp Nom   mother-my
                'My mother has returned'

We cannot rule out the possibility of borrowing from one language to another, especially when the languages are geographically adjacent to each other. A case in point is the first person genitive *-li* 'my', which is an innovation in all dialects of Rukai and that it may have been borrowed into Puyuma. The usage of *-li* 'my' is restricted to the inalienable constructions, which "occur only in nouns of kinship terms and of body parts used in a figurative sense" in the Tamalakaw dialect of Puyuma (Tsuchida 1995). However, it is much less likely that Puyuma would borrow the passive construction indicated by the prefix *ki-* from Rukai or vice versa. If both languages have inherited this unique passive construction among Formsoan languages, then it is a good piece of morpho-syntactic evidence for their closer genetic relationship.

3.2 Rukai and Paiwan

Lexicostatistic evidence indicates that Paiwan is more closely related to Puyuma (Dyen 1963, 1971). However, Puyuma has a focus system different from all the other Formosan languages, including Paiwan, which has a typical Philippine type of focus system.

Rukai and Paiwan share some affixes with the same or similar functions, not or rarely found in any other Formosan languages. For example,

Rukai	Paiwan	
-nga	*-anga*	'completive', e.g. Ruk *wakane-nga* 'to have eaten', Pai *vaik-anga* 'already going' (Ferrell 1982:58)
ka-…-ane	*ka-…-an*	'genuine, native, traditional', e.g. Ruk *ka-bava-ane* 'native wine', Pai *ka-paysu-an* 'real money'
ki-	*ki-*	'to get, pick', e.g. Ruk *ki-tai* 'to pick taroes', Pai *ki-vasa* 'to pick taroes'
ki-	*ki-*	'by oneself', e.g. Ruk *ki-vaevang* 'to play by oneself', Pai *ki-vangvang* 'to play by oneself', *ki-elaela* 'to ask oneself', Pai *ki-sia'* 'to feel embarassed'

It is possible that Rukai and Paiwan may be more closely related. The relationship among Rukai, Puyuma and Paiwan remain to be worked out.

3.3 Tsou

Tsou is unique in the following aspects of syntax: (1) Every clause requires an auxiliary verb, (2) The auxiliary and main verbs of a clause must agree in focus, (3) The focus system in Tsou is quite different from all other Formosan and western Austronesian languages, as shown before in Section 2; see examples below. There are other syntactic idiosyncrasies of Tsou, including the following: (1) Compound verbs are quite common, and (2) The causative verb indicated by the prefix *poa-* (< **pa-*) is used only in non-Agent-focus constructions, and these syntactic features are Tsou innovations (Yungli Chang, p.c.). Tsou, rather than Rukai, could be the very first offshoot from Proto-Austronesian.

Tsou (5) m-oso   m-imo   to   emi   'o   ic'o.
          AF-Aux AF-drink Obl   wine   Nom   that
             'That person has drunk wine'
Tsou (6)a. m-oh   ta   m-imo   to   emi.
          AF-Aux he/Nom AF-drink Obl   wine
             'He has drunk wine'
       b. oh   ta   im-a   (na) 'o   emi.
          PF-Aux him/Obl drink-PF   Nom   wine
             'The wine has been drunk by him'

In short, Tsou is unique with a number of morpho-syntactic features of its own. Thus Tsou, rather than Rukai, could be the very first offshoot from PAN. If so, how could we account for the fact that Rukai has no focus system while Tsou has one?

3.4 Bunun

No phonological evidence indicates that Bunun is genetically closer to any other Formsoan languages.   As in Kavalan, Basay and Amis, *t and *C merged as *t* in Bunun, and as in Kavalan and Basay, *n and *N merged as *n* in Bunun, and these are shared phonological innovations of all extra-Formosan languages. However, unlike Kavalan, Basay and Amis, *j did not merge with *n or *N in Bunun or extra-Formosan languages.

Like many other Formosan languages, Bunun has a four focus system: AF *m(a)-*, PF *–un* (< *-en), LF *–an*, and RF *'is-*. It is noticeable that its RF is *'is-* rather than *si-* or *sa-* as in most other Formosan languages. Like Pazih, Bunun has no infix *–um-* indicating Agent-focus.

Starosta, Pawley and Reid (1982) reconstruct *iSi- to account for the Bunun prefix *'is-*. In fact, we can account for the form *'is-* as derived from *s-* < *Si- from the synchronic point of view. The segments *'i* is inserted to avoid consonant clustering, e.g. *ma-snava* (AF), *'isnava-n* < *snava-an (LF), *'isnava* < *s-snava (RF) 'to teach'. If the root form is *snav*a, there is no need to delete *'i* to derive the AF form *ma-snava* and a perfective form with the inserted infix *<in>* correctly (Hsiu-hsu Lin, p.c.), e.g. *s<in>ava-an* 'tuition' (Nihira 1988:231).

The personal marker is *ti* in Kavalan, e.g. *ti api*, *ti abas*, *ti ula*w. The RF marker is also *ti-* in Kavalan, e.g. *ti-kiras* 'to cut with something', *ti-Ramaz* 'to cook for someone'. Similarly, the RF marker is *'is-* in Bunun, and Bunun personal names all have the prefix *'is-*, e.g. *'is-bukun*, *'is-lituan* (Hsiu-hsu Lin, p.c.).   This seems to indicate that Bunun may be more closely related to Kavalan and the other languages in the East Formosan group, which includes Siraya, Amis, Kavalan and Basay (Blust 1999, Li 2004). Notice that the personal marker is *ci* in Amis, and *ti* in Siraya.

Similar to Tsou (Tsuchida 1990) and Thao (Blust 2003:91-186), Bunun has a rich repertoire of verb-deriving prefixes, called "lexical prefixes" by Nojima (1996).   I am not suggesting that Bunun is closer to either of these two

languages. But a careful comparison of their prefixes may shed some light on their relationship.

4. Languages of the Southwestern Plains

Aside from Siraya proper, there are only short wordlists originally collected by scholars, laymen and policemen for the extinct languages and dialects in the southwestern plains. Ogawa (see Tsuchida 1991) prepared a comparative vocabulary of 75 villages or sources and arranged the data in three separate groups: Siraya, Taivoan and Makatao. Baed on Ogawa's manuscript, Tsuchida (1991) presented some lexical evidence to prove that they belonged to three separate languages: (1) Like all the other Formsoan languages, their indigenous self-appellation terms are *siraya, taivoan* and *makatao* respectively, and (2) their terms for 'wine', again like the other Formosan languages, are all different, namely *it, tau* and *lihu* respectively. In addition, Tsuchida presented some phonological evidence for their differences:

PAN	Siraya	Taivuan	Makatao
(1) *l	r	ø~h	r
(2) *N	l	l	n

A second type of language data for the languages in this area is available in the so-called "Sinkang manuscripts", 168 contracts written in the Romanized indigenous languages. Based on this language material, I have found two more phonological differences between Siraya and Taivoan:

	PAN	Siraya	Taivoan	
(3)	*D, *d	s	r~d	
		sa	ra, da	'and'
		hiso	haijro	'if, as'
	*Daya	saija	raija	'east'
	*laHud	raos	raor	'west'
		posoh	poroh	'land'
(4)	*k	ako-saij	au-saij	'not have'
	*g(?)	dagogh	daoh	'price'
		ligig	liih	'sand'

The velar obstruents *k* and *g* in the intervocalic position are retained in Siraya, but lost in Taivoan, as shown in (4) above.

In addition to the lexical and phonological differences, there is some morphological evidence for their difference: The suffix *-ali* or *-ili* 'future' in Siraya corresponds to *-ah* in Taivoan.

In short, all the linguistic evidence indicates that there are three separate languages in the southwestern plains: Siraya, Taivoan and Makatao. Yet they appear to more closely related to each other than any other Formosan language elsewhere. Their reflex for PAN *j is *n* in all the three languages, e.g. *bukij >

Siraya *vaukyn*, Taivoan *bukin*, Makatao *bukin* 'mountain'.   That PAN *j merged with *n is a typical phonological innovation in the East Formosan group.

5. Conclusion

No satisfactory classification of Formosan language is available today.   A lot more careful work on the comparative study of Formosan language is required to get a more satisfactory solution.   My tentative classification is given below (see next page).

The northern group, comprised of the Atayalic and Northwestern languages, was postulated in Li (1985).   Recently Meili Yeh (2006) has found that two grammatical properties are exclusively shared by Atayal and Saisiyat: (1) the distinction of the future tense between Agent-focus and non-Agent-focus constructions, and (2) nominalization patterns. Phonological and lexical evidence for a close relationship between Saisiyat and Pazih is not strong.   It is conceivable that morpho-syntactic evidence might show that Saisiyat is closer to the Atayalic, rather than Pazih.

All the suggestions given in this paper are tentative in nature and require further investigation and careful comparison to confirm or disconfirm them.   It is obvious that this paper is more of a progress report than a completed research project.
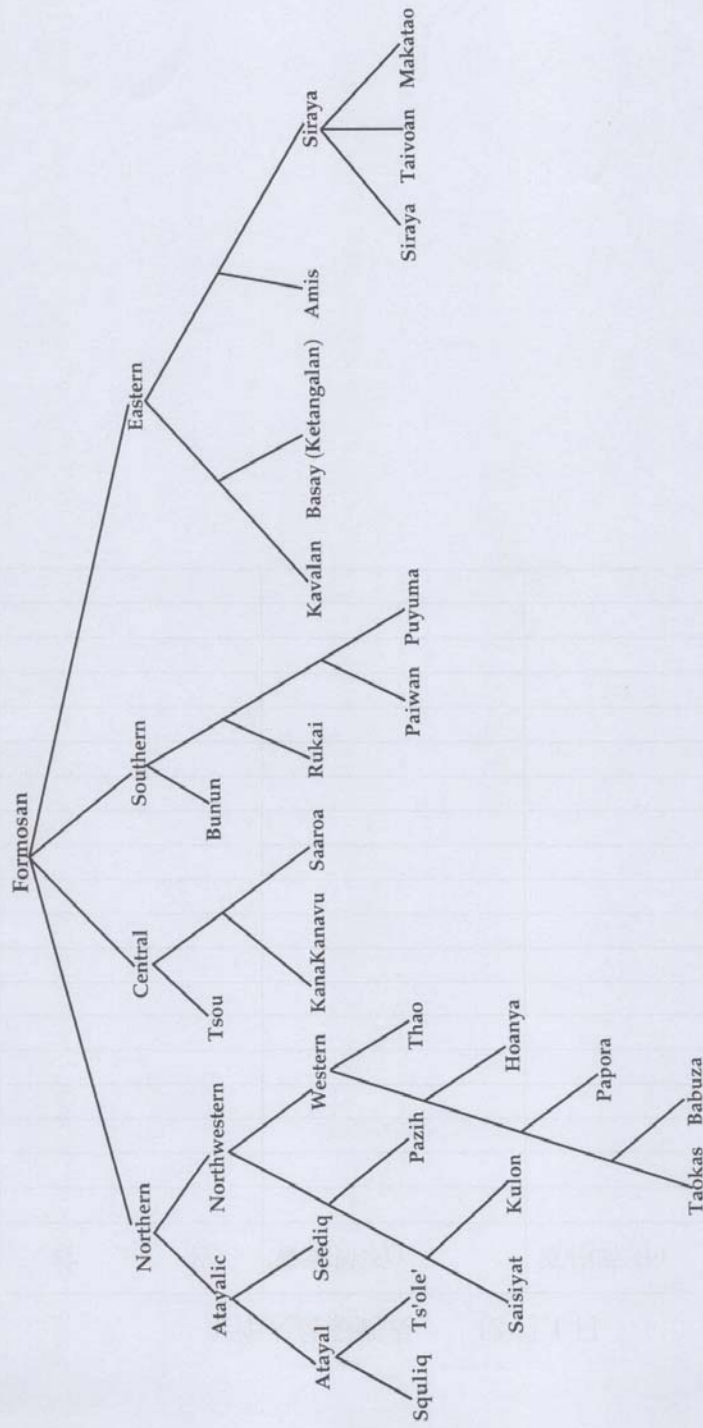
Classification of Formosan Languages

Figure 3. Paul Li

8

References

Blust, Robert

    1996   Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35.2:272-294.

    1999   Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun and Paul Jen-kuei Li, eds. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, 31-94. Symposium Series of the Institute of Linguistics (Preparatory Office), No.1. Taipei: Academia Sinica.

    2003   *Thao Dictionary*. Institute of Linguistics, Academia Sinica, Language and Linguistics Monograph Series No.A-5.

Chang, Yung-li（張永利）

    2002   Distributivity, plurality, and reduplication in Tsou. *Tsing Hua Journal of Chinese Studies*.

Dyen, Isidore

    1963   The position of the Malayopolynesian languages of Formosa. *Asian Perspectives* 7.1/2:261-271.

    1971   The Austronesian languages of Formosa. *Current Trends in Linguistics* 8:168-199.

    1995   Borrowing and inheritance in Austronesianistics. In Li, Paul *et al.*, eds. *Austronesian Studies Relating to Taiwan*, 455-519. Taipei: Academia Sinica.

Ferrell, Raleigh

    1982   *Paiwan Dictionary. Pacific Linguistics* C-73. Canberra: Australian National University.

Huang, Lillian M.（黃美金）

    2000   《卑南語參考語法》[Puyuma reference grammar]。台北 [Taipei]：遠流出版公司 [Yuan-liu Co.]。

Li, Chao-lin（李剣麟）

    2004   *The Spatial Representations in Paiwan*. Unpublished MA thesis, National Chungcheng University.

Li, Paul Jen-kuei（李壬癸）

    1973   *Rukai Structure*. Taipei: Institute of History and Philology, Academia Sinica, Special Publications No.64.

    2001   On the linguistic position of Thao—Some remarks on Blust's (1996) 'Some remarks on the linguistic position of Thao'. Su-chuan Jan, Ying-hai Pan eds., *Papers Presented at the Symposium on the Plains Aborigines and Taiwan History*, 165-184. Institute of Taiwan History Preparatory Office, Academia Sinica.

    2003a Notes on Favorlang, an extinct Formosan language. In Paul Li, ed., *English-Favorlang Vocabulary*, 1-13, by Naoyoshi Ogawa. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Asia-African Lexicon

Series No.43.

2003b The internal relationships of six western plains languages. *Bulletin of the Department of Anthropology* 61:39-51.

2004 Origins of the East Formosan peoples: Basay, Kavalan, Amis and Siraya. *Language and Linguistics* 5.2:363-376.

Nihira, Yoshiro（仁平芳郎）

1983 *A Bunun Vocabulary: A Language of Formosa*, 3rd edition, enlarged.

Nojima, Motoyasu（野島本泰）

1996 Lexical prefixes of Bunun verbs. *Gengo Kenkyu* [Journal of the Linguistics Society of Japan] 110:1-27.

Ross, Malcolm D.

1995 Reconstructing Proto-Austronesian verbal morphology: Evidence from Taiwan. In Li, Paul *et al.*, eds., *Austronesian Studies Relating to Taiwan*, 727-791.

Ross, Malcolm D. and Stacy Fang-ching Teng

2005 Clause constructions in Nanwang Puyuma. *Concentric* 31.1:119-158.

Starosta, Stanley（帥德樂）

1995 A grammatical subgrouping of Formosan languages. In Paul Li, *et al.*, eds., *Austronesian Studies Relating to Taiwan*, 683-726. Institute of History and Philology, academia Sinica.

Starosta, Stanley, Andrew K. Pawley & Laurence A. Reid

1982 The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A. Wurm, eds., *Papers from the Third International Conference on Austronesian Linguistics, Vol.2, Tracking the Travellers* 145-170. *Pacific Linguistics* C-75.

Tan, Cindy Ro-lan（譚若蘭）

1997 A Study of Puyuma Simple Sentences, 140 pp. M.A. thesis, National Taiwan Normal University. Taipei.

Teng, Stacy F.（鄧芳青）

1997 *Complex Sentences in Puyuma*. MA thesis, National Taiwan Normal University.

Tsuchida, Shigeru（土田滋）

1980 Puyuma (Tamalakaw dialect) vocabulary—with grammatical notes and texts. *The Peoples, Cultures and Languages of the Black Currents*, 183-307.

1990 Classificatory prefixes of Tsou verbs. *Tokyo University Linguistics Papers '89*, 17-52. Tokyo: University of Tokyo.

1995 Alienable and inalienable distinctions in Puyuma. In Li, Paul *et al.*, eds., *Austronesian Studies Relating to Taiwan*, 793-804. Taipei: Academia Sinica.

Tsuchida, Shigeru, Yukihiro Yamada and Tsunekazu Moriguchi

1991 *Linguistic Materials of the Formosan Sinicized Populations I: Siraya and Basai*. University of Tokyo.

Tung, T'ung-ho (1911-1963) *et al*.

1964 *A Descriptive Study of the Tsou Language, Formosa*. Taipei: Institute of History and Philology, Academia Sinica, Special Publications No.48, Reprinted in

1991 with Notes and Errata prepared by Paul J.K. Li.

Wu, Chun-ming（吳俊明）

2004 *A Study of Lexical Categories in Paiwan*. Unpublished MA thesis, National Chungcheng University.

Zeitoun, Elizabeth

1992 *A Syntactic and Semantic Account of Tsou Focus System.* Unpublished MA thesis. Hsinchu, Taiwan: National Tsing Hua University.

Yeh, Marie Meili

2006 The linguistic position of Saisiyat. Manuscript.