

# Classification of Formosan Languages

*Paul Jen-kuei Li and Elizabeth Zeitoun*

30.1	Introduction	2
30.2	The Formosan Languages: Not a Single Subgroup	2
30.3	Classification of Formosan Languages	3
30.3.1	<i>Early Classification (Asai 1936)</i>	4
30.3.2	<i>Lexicostatistical Classifications</i>	4
30.3.2.1	Dyen (1963, 1965)	4
30.3.2.2	Ferrell (1969)	6
30.3.2.3	Tsuchida (1976)	8
30.3.2.4	Discussion of Lexicostatistic Classifications	10
30.3.3	<i>Phonological Classifications</i>	11
30.3.3.1	Ho (1998)	11
30.3.3.2	Blust (1999b)	12
30.3.3.3	Discussion of Phonological Classifications	14
30.3.4	<i>Morphosyntactic Classifications</i>	14
30.3.4.1	Starosta (1995)	14
30.3.4.2	Ross (2009)	16
30.3.4.3	Zeitoun & Teng (2016)	20
30.3.4.4	Aldridge (2015, 2016, 2021)	20
30.3.5	<i>Numeral-Based Classification</i>	21
30.3.6	<i>Summary</i>	22
30.4	Discussion of Controversial Subgroups	23
30.4.1	<i>Tsouic</i>	23
30.4.2	<i>Atayal-Seediq</i>	24
30.4.3	<i>Amis-Sakizaya</i>	25
30.4.4	<i>Sirayaic</i>	25
30.5	Internal Relationships of Formosan Languages	27
30.5.1	<i>Atayal</i>	27
30.5.2	<i>Bunun</i>	28
30.5.3	<i>Rukai</i>	28
30.5.4	<i>Paiwan</i>	29
30.5.5	<i>Puyuma</i>	30
30.5.6	<i>Summary</i>	30
30.6	Conclusion	31
	References	31

### 30.1 Introduction

The geographical delimitation of the Austronesian language family is well understood: it is overwhelmingly insular, covering about two-thirds of the globe from west to east. However, the distant genetic relationship between Austronesian and other language families (e.g., Indo-European, Austroasiatic, Semitic, Japanese, Tai-Kadai, Chinese) is still a matter of controversy. Several hypotheses have been advanced, and the reader is referred to Blust (2013, pp. 702–721) for a detailed assessment, as such an overview is beyond the scope of this handbook.

The purpose of this chapter is threefold: it intends to provide (i) an assessment of subgrouping hypotheses, (ii) a discussion on controversial subgroups, and (iii) a short overview of internal subgroups. More specifically, after having presented evidence that the Formosan languages do not form a single subgroup (§30.2), we review various subgrouping hypotheses that have been proposed regarding the Formosan languages (§30.3). We then discuss four subgroups that are controversial either linguistically or politically (§30.4) before examining internal subgroups (§30.5). In his 1936 monograph, Erin Asai admitted that “[o]ur present knowledge of the Formosan languages does not permit us to construct a full table showing the affinity of each language *definitively* [...]”. Nearly a century later, the situation has not changed: the classification of the Formosan languages and their internal relationships are still a matter of debate. We have very few early language records, and most only date back to the Japanese period (1895–1945). It is thus difficult to retrace with exactitude the changes over time in phonological and grammatical structures that might have resulted, among other things, from language contact, language change, language loss, and/or language shift. We are also limited by the fact that the Formosan languages are at the highest level of the Austronesian phylogeny, and subgrouping and reconstruction are subject to circularity. Identifying subgroups depends on shared innovations relevant to a reconstructed proto-language, but reconstructing that proto-language depends on these reconstructions (Malcolm Ross, pers. comm.).

### 30.2 The Formosan Languages: Not a Single Subgroup

Before starting this overview, it is crucial to first examine the evidence against the claim that the Formosan languages form a single subgroup.

Ogawa & Asai’s (1935) milestone work showed that the Formosan languages<sup>1</sup> retain many archaic phonological and grammatical features, and this may have

led to the hypothesis that they form “one common subgroup of the [Austronesian] family” (Dahl 1973, p. 125).<sup>2</sup> On the one hand, such a claim was short-lived because there are no exclusively shared innovations that support it. On the other hand, phonological, lexical, and morphological innovations are found in all Austronesian languages spoken outside Taiwan and have been taken as evidence to propose a single non-Formosan subgroup called “Malayo-Polynesian” (Blust 1977, 1995b). Phonologically, PAN \*S and \*h, \*C and \*t, and \*n and \*N have merged as PMP \*h, \*t, and \*n, respectively (Ogawa & Asai 1935, p. 6, Blust 2013, p. 749). Lexically, PAN \*Siwa ‘nine’ is reflected as PMP \*siwa (for the expected form \*\*hiwa) (Ross 2012, p. 1256). Morphologically, there has been a “politeness shift”, in which the PAN pronoun \*-mu ‘2PL.GEN’ shifted to the singular PMP genitive pronoun \*-mu ‘2SG.GEN’ (Blust 1977, p. 11), and at least one innovation in verbal morphology, with the AV prefix \*maN- found on dynamic verbs in PMP but not in the Formosan languages (Ross 2012, p. 1256).

### 30.3 Classification of Formosan Languages

Blust (2013, p. 721) mentions that “[t]he problem of linguistic subgrouping can be subdivided into three areas: 1. models of subgrouping, 2. methods of subgrouping, and 3. results of subgrouping.”

Taking into account only the hypotheses made by linguists, at least 25 proposals have been advanced, all very diverse in methodology and outcome (Asai 1936, Dyen 1965, Ferrell 1969, Dahl 1973, Tsuchida 1976, Blust 1977, Marsh 1977, Dahl 1981, Harvey 1982, Reid 1982, Ho 1983, Li 1985, 1990, Dyen 1990, Starosta 1995, Ho 1998, Blust 1996, 1999a, 1999b, Ho & Yang 2000, Li 2003, Sagart 2004, Ross 2009, Zeitoun & Teng 2016, Aldridge 2021). Lack of space prevents us from presenting all these classifications,<sup>3</sup> and we will only examine the first linguistic subgrouping proposed by Asai (1936) (§ 30.3.1), followed by lexicostatistical (§ 30.3.2), phonological (§ 30.3.3), morphosyntactic (§ 30.3.4), and numeral-based classifications (§ 30.3.5). This presentation, while necessarily selective,

1 Ground work was made by Ogawa, who published a number of papers between 1930 and 1935 (see P. Li, this handbook, Chapter 3, Liao, this handbook, Chapter 31).

2 Ten years before O. Dahl, Dyen (1963, p. 268) also concluded that lexicostatistic results “indicate a single Formosan subfamily of Malayo[-P]olynesian”, an assumption also held by Tsuchida (1976). Another type of evidence that they presented was the dozens of “Formosan-only” cognates, e.g., \*Cumay ‘bear’, \*LikuNaw ‘leopard’, \*DakeS ‘camphor laurel’, \*Cabu ‘wrap’, \*NuqeS ‘marrow’, \*imah ‘drink’ (see Blust & Trussel 2020).

3 See Blust (1999b, pp. 39–40) for a review of the hypotheses proposed prior to 1996.

attempts to show in detail the evidence presented by these authors to support their subgrouping hypotheses.

### 30.3.1 *Early Classification (Asai 1936)*

One of the earliest linguistic classifications, reproduced in Figure 30.1, dates back to Asai (1936). It is based on “the reciprocal relation of the phonology, morphology and vocabulary” (p. 5) of the Formosan languages, which are divided into four main groups—northern, Bunun, Tsou-Paiwan, and Amis—with Yami recognized as part of the Batan group. While Bunun and Amis are shown to each include three main dialects (northern, central, and southern), the two other groups (northern and Tsou-Paiwan) consist of much more complex language clusters. Interestingly, they subsume geographically close languages. The “northern group” includes “Original Atayal” (further divided into Atayal proper and Seediq) and Saisiyat. The “Tsou-Paiwan group” is composed of four main languages: (1) Tsou, (2) “Original Saaroa-Kanakanavu” (with Saaroa and Kanakanavu treated as two distinct dialects), (3) “Original Paiwan” (with Paiwan proper and Puyuma as two distinct languages), and (4) Rukai, among which the status of Mantau[r]an is questioned. None of the sinicized languages (including Thao, Pazeh, Taokas, Babuza, Hoanya, Papora, Siraya, Kavalan, and Basay) is included in this classification.

### 30.3.2 *Lexicostatistical Classifications*

No other classification of the Formosan languages was proposed for almost 30 years after Asai (1936), until a new subgrouping hypothesis, based on lexicostatistics, was advanced by Dyen (1963, 1965) (§ 30.3.2.1), followed by Raleigh Ferrell (1969) (§ 30.3.2.2) and Shigeru Tsuchida (1976) (§ 30.3.2.3).

#### 30.3.2.1 Dyen (1963, 1965)

Dyen (1963, 1965) proposes the “existence of at least three subgroups among the languages [...] reported by Ogawa & Asai (1935)”: Atayalic (Atayal and Seediq) in the north, Tsouic (Tsou, Kanakanavu and Saaroa)<sup>4</sup> in the center, and East Formosan, which includes all the remaining languages (Kavalan, Pazeh, Saisiyat, Thao, Bunun, Amis, Puyuma, Paiwan, and Rukai). He mentions that Thao, Bunun, Ami[s], Puyuma, and Paiwan cluster together, while Pazeh is closer to Saisiyat; Kavalan and Rukai do not subgroup with any of the other languages,

4 Dyen (1963, p. 263) mentions that Tsou, Kanakanavu, and Saaroa might be distinct languages rather than dialects as assumed by Ogawa & Asai (1935) and Asai (1936), with Kanakanavu and Saaroa closer to each other.

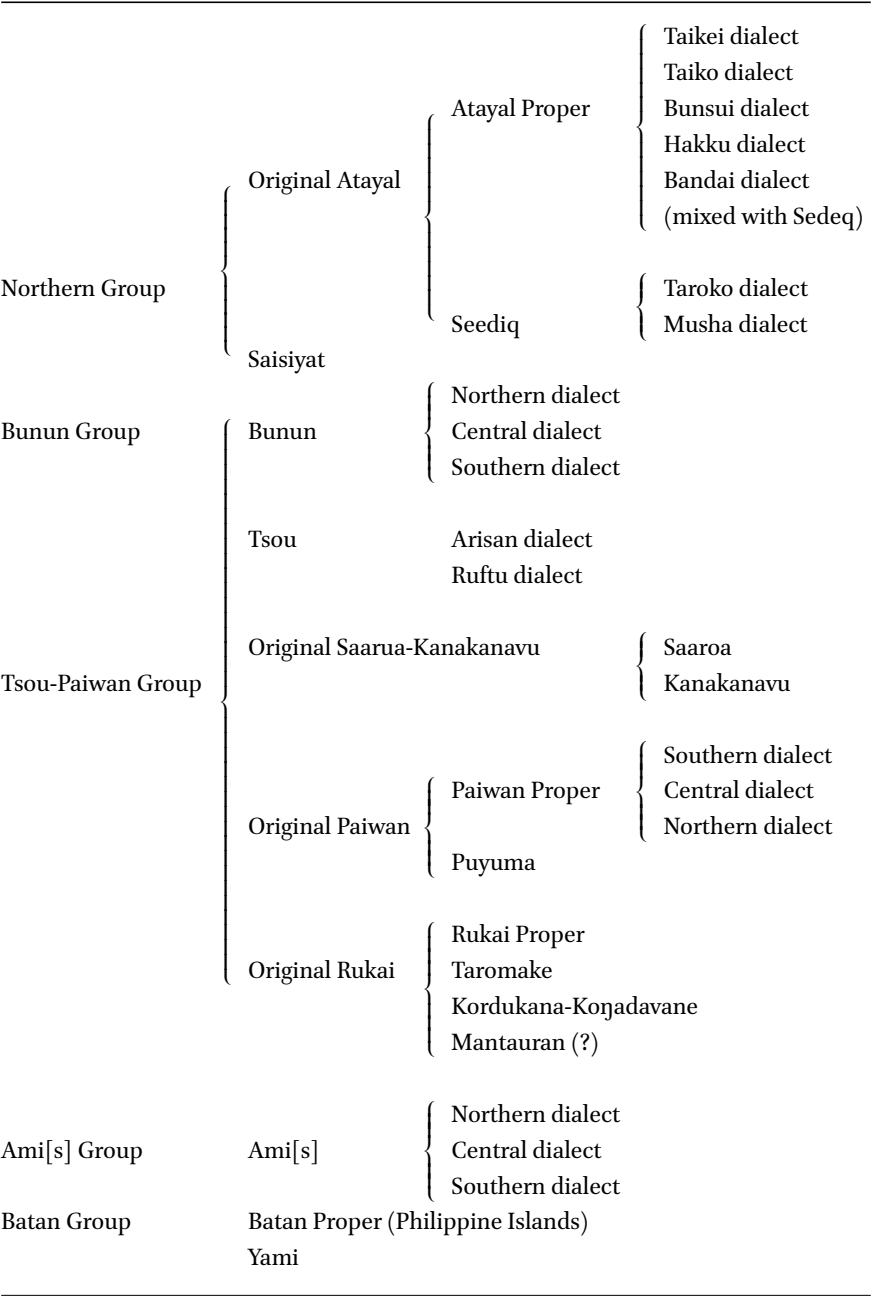


FIGURE 30.1 Asai's (1936, p. 6) classification of Formosan languages

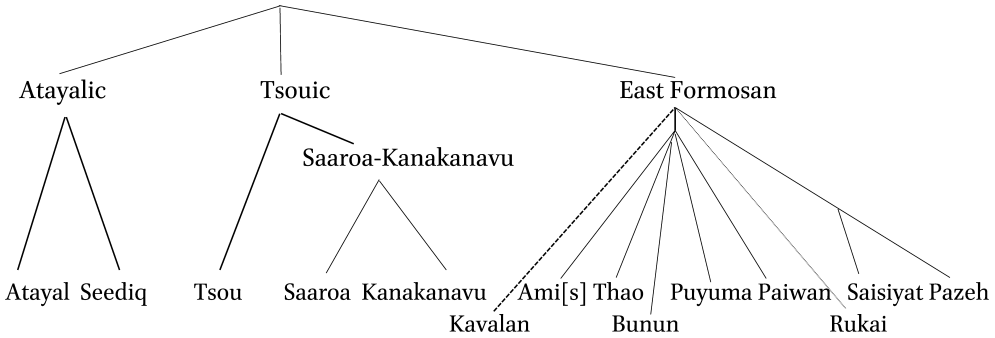


FIGURE 30.2 Dyen's (1965) subgrouping hypothesis

but “the former seems to be closest to Ami[s], and the latter to Puyuma and Paiwan” (Dyen 1965, p. 287). His subgrouping hypothesis is represented schematically in Figure 30.2. Dyen (1965) makes two other interesting remarks: first, the high lexical divergence among Atayalic, Tsouic, and East Formosan shows that Taiwan might “be the homeland for the languages to the South and perhaps for all of the Austronesian languages” (ibid.); second, “East Formosan is perhaps more closely related to the languages of the South, in the Philippines and Indonesia, than to Atayal” (ibid.).

Dyen's (1963, 1965) subgrouping hypothesis served as the basis for Ferrell's (1969) and Tsuchida's (1976) proposals, though it is now well-known that languages do not change at a constant rate of lexical replacement and that we thus need to be careful when relying on lexicostatistics. Also, plausible subgroupings imply the discrimination of retention from innovations, and this is not an easy task. For instance, linguists such as Dyen often failed to identify true cognates in languages with which they were not familiar.

### 30.3.2.2 Ferrell (1969)

Ferrell's (1969) study consists of a brief introduction to the cultural and linguistic traits of the Formosan languages with a list of classified vocabulary. On the basis of Dyen's (1963) lexicostatistical findings, the higher percentage of cognates and the (non)distinction of the PAN reflexes \*t and \*C, he hypothesizes a tripartite division of the Formosan languages into Atayalic, Tsouic, and Paiwanic, as schematized in Figure 30.3. He proposes that (1) Atayalic is composed of Atayal and Seediq; (2) Tsouic<sup>5</sup> includes Tsou, Kanakanavu, and Saaroa (see § 30.4.1); and (3) Paiwanic is split in two, Paiwanic I (Rukai, Pazeh, Saisiyat,

5 Ferrell (1969) challenges the hypothesis that Kanakanavu and Saaroa subgroup with Tsouic. One the one hand, he mentions that these three languages share “many features [that] set

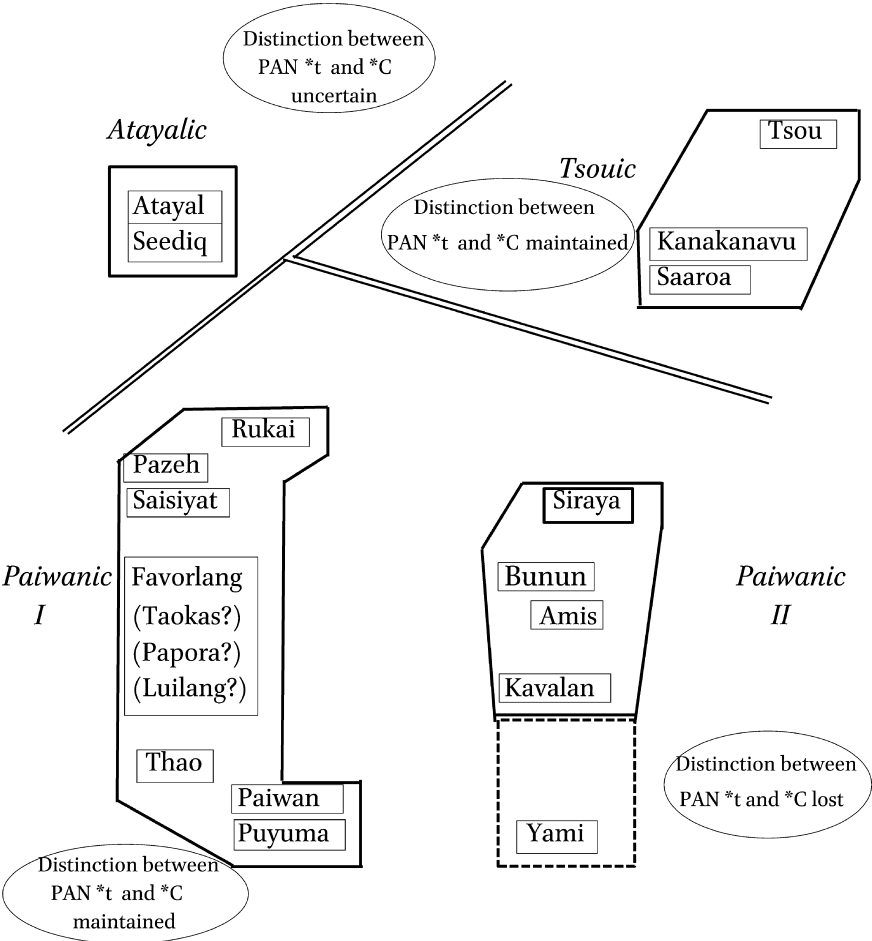


FIGURE 30.3 Ferrell's (1969, p. 69) subgrouping hypothesis

Thao, Puyuma, and Paiwan) and Paiwanic II (Siraya, Bunun, Amis, Kavalan, and Yami). His proposal prevailed until the 1990s and was adopted as a working hypothesis by Dyen (1971), Tsuchida (1976), Blust (1977), and Dahl (1981), even though it was generally acknowledged that there were problems with Ferrell's so-called Paiwanic group, which Blust (2013, p. 744) summarizes as follows: "With regard to Paiwanic II Yami is a Philippine language, and Bunun shares

them apart from all other Taiwan groups" (p. 36), and "superficial phonological examination" lead him to continue to consider Tsouic a "discrete grouping". On the other, he also argues that it cannot be excluded that Saaroa is actually a "Paiwanic language" with "heavy Tsouic overlay" (p. 68).

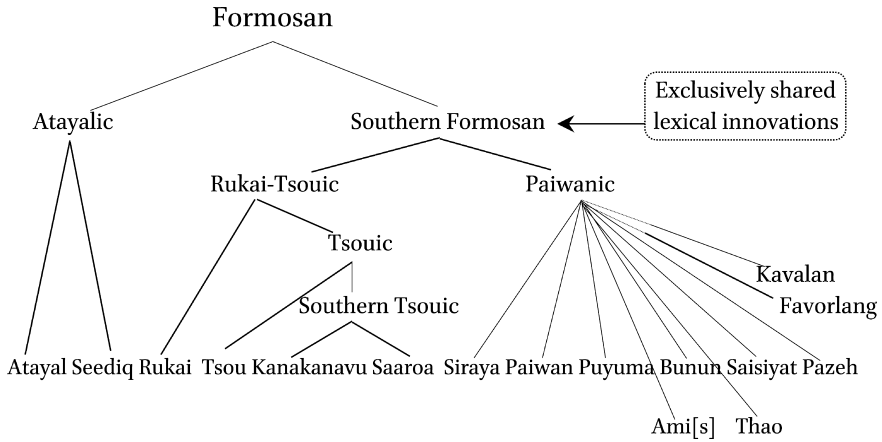


FIGURE 30.4 Tsuchida's (1976, p. 15) subgrouping hypothesis

no known innovations exclusively with Siraya, Amis and Kavalan. Paiwanic I is likewise a diverse collection with no exclusively shared innovations, apart from members of the Western Plains group.”

### 30.3.2.3 Tsuchida (1976)

Roughly following Dyen (1963, 1965) and Ferrell (1969), Tsuchida (1976, pp. 1–10, 14–15) assumes that Proto-Formosan is divided in two main groups, Atayalic and Southern Formosan. He accepts, *de facto*, Atayalic (Atayal and Seediq) and Paiwanic (one of the two subgroups under Southern Formosan), i.e., all remaining languages with the exclusion of Kavalan due to scarcity of data, and Yami because it is a Batanic language, but suggests that Tsouic and Rukai form a subgroup that he labels Rukai-Tsouic, which further clusters with Paiwanic as the Southern Formosan subgroup, as shown in Figure 30.4.

The phonological innovations that characterize the Tsouic subgroup are summarized in (1).

#### (1) a. Sibilant dissimilation

PAN \*C<sub>1</sub> > PT \*t / \_\_ \*S/\*s

PAN C<sub>1</sub>aS<sub>6</sub>i<sub>q</sub><sub>3</sub> > Kan t<um>a-taʔisi, Sar t<um>a-talis-ua, Tso t<m>eʔsi  
‘to sew’ (Tsuchida 1976, p. 160)

PAN C<sub>1</sub>aŋis > Kan t<um>a-taŋi, Sar t<um>a-taŋii, Tso m-oŋsi (tŋis-i) ‘to cry, weep’ (Tsuchida 1976, p. 149)

#### b. Echo vowels altering the PAN predominant CVC canonical syllable structure to CVCV, where the last vowel of a word echoes the penul-



timate vowel as in CV<sub>1</sub>CV<sub>1</sub>#; if the penultimate vowel is /a/, then V<sub>1</sub># is schwa, e.g.,

PAN \*w<sub>1</sub>iRiH<sub>2</sub> > PT \*wíríhi > Kan *íri*, Sar *iri*, Tso *vri-na* 'left' (ibid., p. 145)

PAN \*busuR > PT \*vusúru > Kan *vuúru*, Sar *vuuru*, Tso *fsuru* 'bow' (ibid., p. 128)

PAN \*ʔənəm > PT \*ənəmə > Kan *u-num*, Sar *u-num*, Tso *nom* 'six' (ibid., p. 251)

PAN \*D<sub>2</sub>a(ŋ)D<sub>2</sub>aŋ > PT \*č<um>a~čaŋəčaŋə > Kan *c<um>a~caŋucaŋu*, Sar *s<um>a-saasaŋu*, Tso *t<m>a-cŋucŋu* 'to dry by fire' (ibid., p. 154)

c. PAN \*-an > PT \*-ã

PAN \*(q<sub>a</sub>a-)lipan<sup>6</sup> > PT \*ʔ<sub>1</sub>alálipã > Sar *ʔ~al~alipa*, Tso *r-erpa* 'centipede' (ibid., p. 216)

d. Loss of PAN \*S<sub>2</sub> (Ross' \*x)

PAN \*kaS<sub>2</sub>iw > PT \*káiwu > Kan *kaálu*, Tso *evi* 'tree, firewood' (ibid., p. 247)

e. Mergers

PAN \*C, \*d > PT \*d

PAN \*D<sub>1</sub>, (\*D<sub>4</sub>), \*D<sub>2</sub>, \*Z > PT, PAN (\*z<sub>1</sub>) > PT \*č

PAN \*k, \*k<sub>2</sub>, \*g > PT \*k

PAN \*R, \*r > r

PAN \*θ, \*θ<sub>1</sub>, \*S<sub>6</sub> > PT \*θ

To support the Rukai-Tsouic subgroup, Tsuchida (1976, pp. 11–12) proposes two phonological innovations, as in (2), and twelve cognate sets, which are exclusively found in the Tsouic languages and in the Rukai dialects, as shown in (3):

(2) Phonological innovations

1. loss of PAN final \*n after an unstressed /a/ in pre-Proto-Tsouic, Maga, and Mantauran
2. echo vowels (see (1b))

6 Tsuchida (1976) indicates provisional reconstructions with sub-numerals. He states that Seediq is the only language that exhibits different reflexes for \*q<sub>1</sub> and \*Q<sub>2</sub>. If no Seediq cognate is found, reconstruction is indeterminably \*[q<sub>1</sub>Q<sub>2</sub>] and thus represented as \*q<sub>a</sub> (ibid., pp. 165–167).

## (3) Exclusively shared cognate sets

1. Tso *ho*, Sar, Mt Rukai *la* 'and' (< PRT \**la*)
2. Kan, Mg/To Rukai *si* 'because' (< PRT \**S*<sub>16i</sub>)<sup>7</sup>
3. Kan *tukúnu*, Tso *tɔuhu*, Mg Rukai *a-túklú* 'heart' (< PRT \**tukulu*)
4. Tso *cɔrha*, Kan *cakuranu*, Sar *sakuralu*, Bd Rukai *ɖakəralə*, Mg Rukai *ɖkərli*, To Rukai *ɖakəalə*, Mt Rukai *ɖakəralə* 'river' (< PRT \**Dakəralə*)
5. Tso *coŋroh-a*, Sar *ʔa-caŋərat-a*, Mg Rukai *ca-cŋálə* 'star' (< PRT \**Caŋə-Ralə*)
6. Tso *ma-free* 'tasty', Sar *um-a-a-vali*, To Rukai *ua-sa-bái*, Mt Rukai *o-ʔa-vali* 'to smell, sniff' (< PRT \*-*bali*)
7. Kan *kúuku*, Mg Rukai *kúku*, Mt Rukai *u-kuku* 'leg' (< PRT \**kəku*)
8. Sar *ɲuu~ɲuru*, Mt Rukai *ɲu~ɲuɔu* 'nose', Kan *ɲurúɔu* 'nasal mucus' (< PRT \**ɲuRuq<sub>2</sub>u*)
9. Kan *ramúcu*, Sar *ramucu* 'hand', Tso *mucu* 'hand, arm', Mt Rukai *ramucu* 'finger' (< PRT \**ramuCu*)
10. Kan *siʔípi*, Mt Rukai *ʔipi* 'arm', Mg Rukai *i-sípi*, To Rukai *a-sípi* 'shoulder' (< PRT \**S*<sub>16i</sub>*iqipi*)
11. Tso *nutnu* 'lungs', To Rukai *tətənə* 'liver', Mt Rukai *tətənə* 'heart' (< PRT \**nətənə* or \**tətənə*)
12. Kan *ʔapútu*, Tso *pútu* 'hammer', Mt Rukai *aputu* 'stone' (< PRT \**qaputu*)

Tsuchida's (1976) Rukai-Tsouic subgroup was first rejected by Ho (1983), who concluded that Rukai shared more grammatical traits with Paiwanic, but was favored again in later studies (see Ho 1998, Ho & Yang 2000). Li's (1990) lexical study shows quite clearly that depending on the Rukai dialect being compared, different subgroupings (Rukai-Tsouic or Rukai-Paiwan) obtain, but one or both of these may simply be the result of contact rather than close genetic relationship.

## 30.3.2.4 Discussion of Lexicostatistic Classifications

Dyen (1963, 1965), Ferrell (1969), and Tsuchida (1976) all assume that Atayalic, which subsumes Atayal and Seediq, is the most distantly related group among all Formosan languages. That's because they failed to identify many cognates in Atayal and Seediq, which exhibit irregular male speech forms, as shown in (4) and (10). This register was discovered by Li (1980, 1982), who explains in detail the different phonological and morphological changes that have taken place in

<sup>7</sup> \**S*<sub>16</sub> refers indeterminably to \**S*<sub>1</sub> or \**S*<sub>6</sub>.

Squliq, the prestige dialect of Atayal, which was the most documented in the 1960s and 1970s, and thus served as the basis for reconstruction. He showed that in Squliq, many male forms were adopted, with the loss of many consonants and vowels before stressed syllables, making it extremely difficult to identify cognates.

(4)	PAN	Squliq Atayal male speech forms	Mayrinax Atayal male speech forms	Gloss
a.	*bulaN	<i>bya-cing</i>	<i>bua-ting</i>	‘moon’
b.	*batu	<i>tu-nux</i>	<i>batu-nux</i>	‘stone’
c.	*kuCu	<i>ku-hing</i>	<i>ku-hing, kucu’</i>	‘head louse’

30.3.3 *Phonological Classifications*

Different phonological classifications have been made (e.g., Ferrell 1980, Blust 1999a, 1999b, Ho 1998, Ho & Yang 2000). In this section, we focus on two recent classifications, Ho (1998) and Blust (1999b), which are quite similar but were developed independently.

30.3.3.1 Ho (1998)

Ho (1998) divides all Formosan languages into two main types, “Formosan type” and “non-Formosan type”. The former distinguishes \*t and \*C, as in Paiwan, Rukai, Saisiyat, Pazeh-Kaxabu, Tsou, Kanakanavu, Saaroa, Atayal, and Seediq, while the latter does not. This latter group includes Bunun, Kavalan, Amis, and Siraya.

In his classification of “Formosan type” languages, there are five major subgroups: Paiwan, Puyuma, and three others, which are unlabeled, but which may be referred to as “Rukai-Tsouic”, “Western Plains” and “North Formosan”.

Ho (1998) shows the shared phonological innovations in each major subgroup, as summarized in (5) and further shown in Figure 30.5:

- (5)
- Phonological innovations in major subgroups according to Ho (1998)

1.

Merger of \*ŋ and \*n as *n*, \*s and \*t as *t* in the Western Plains subgroup

2.

Merger of \*d, \*D, and \*z in Atayal, Saisiyat, and Pazeh

3.

Merger of \*R and \*r as *r* in Rukai-Tsouic; merger of \*k and \*g as *k* in Tsouic, but not in Rukai

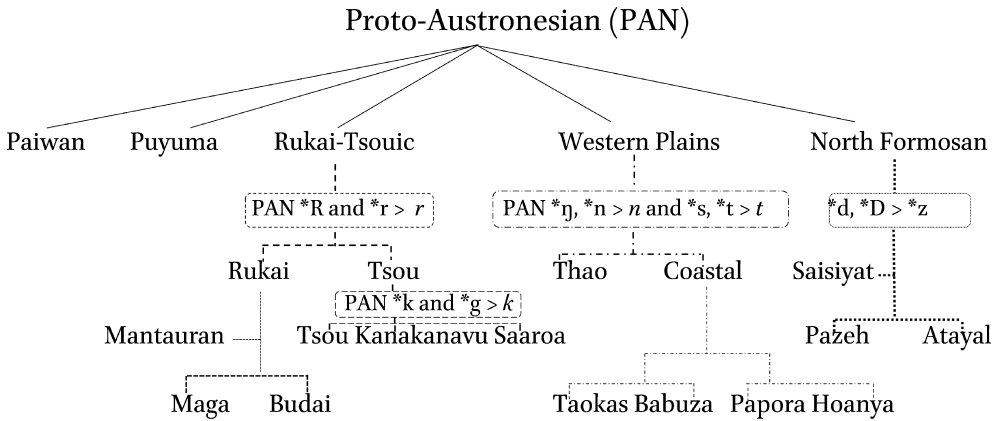


FIGURE 30.5 Ho's (1998) phonological subgrouping

Unfortunately, Ho (1998) did not consider the retention or the loss of *\*S* as a subgrouping criterion. As was first shown by Tsuchida (1976, p. 13) and later reassessed by Blust (1993), *\*S* metathesis is a clear demarcation between Formosan and Malayo-Polynesian, e.g., PAN *\*bukeS* vs. PMP *\*buSek* 'hair (head)', PAN *\*CaqiS* vs. PMP *\*CaSiq* 'sew'.

### 30.3.3.2 Blust (1999b)

Blust (1999b), who rejects Starosta's (1995) subgrouping hypothesis (see §30.3.4), proposes, a rake-like family tree, which consists of 10 primary offshoots including (1) Atayalic, (2) East Formosan, (3) Puyuma, (4) Paiwan, (5) Rukai, (6) Tsouic, (7) Bunun, (8) Western Plains, (9) Northwest Formosan, and (10) Malayo-Polynesian based on significant mergers (1999b, pp. 44–45).

Blust (1996) was the first to identify Thao as a Western Plains language and to demonstrate that similarities shared with Bunun were the result of language contact. Blust (1999b) was also instrumental in recognizing "East Formosan", a previously unrecognized group including Basay-Trobiawan, Kavalan, Amis, and Siraya and exclusively sharing the merger of PAN *\*j* and *\*n*. He states, "All languages which share the merger of PAN *\*j* and *\*n* [as *n*] also share the merger of PAN *\*t* and *\*C* [as *t*]. Although the later change has also taken place in Bunun and in PMP, it is otherwise unknown in Taiwan. The simplest hypothesis is therefore to posit three convergent mergers of PAN *\*t* and *\*C*: one in Proto-East Formosan, another in Bunun, and a third in PMP. Together these two distinctive changes strongly suggest that the languages of Taiwan's eastern coast from Basay in the north to Amis in the south, together with Siraya, constituted a single prehistoric language community at some time after the break-up of PAN" (p. 46).

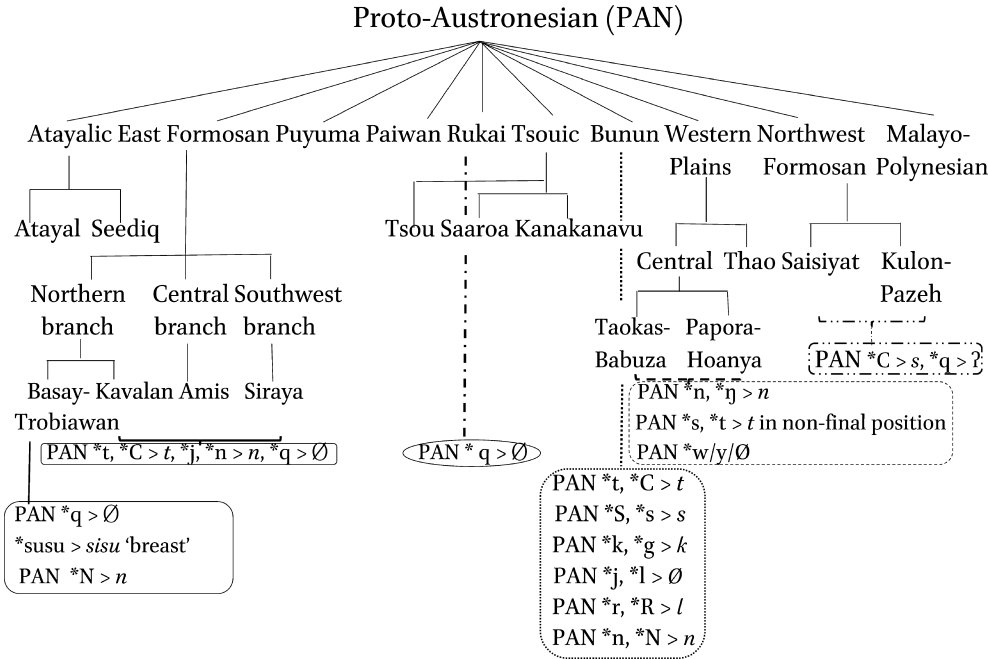


FIGURE 30.6 Blust's (1999b) phonological subgrouping

Other significant mergers are listed in (6) and further indicated in Figure 30.6 above.

- (6) Phonological innovations in major subgroups (Blust 1999b)
1. Mergers in East Formosan of PAN \*t, \*C > t; \*j, \*n > n; \*q > Ø
  2. Mergers in the Western Plains of \*n and \*ŋ as n; \*s, \*t as t; and \*w/y/Ø through truncation of the diphthongs \*-ay and \*-aw
  3. PAN \*q > Ø in Rukai
  4. Mergers in Bunun of PAN \*S, \*s > s; PAN \*k, \*g > k; PAN \*j, \*l > Ø; PAN \*r, \*R > l; and PAN \*n, \*N > n
  5. PAN \*C > s and \*q > ? in Saisiyat and Kulon-Pazeh

Blust (1999b) does not provide any phonological evidence for the Atayalic subgroup and simply considers it “self-evident”. There is, in fact, no phonological evidence per se, so we need to look for other types of evidence, as shown in §30.4.1. Blust also takes Tsouic for granted, so does not provide any phonological evidence that Tsou, Kanakanavu, and Saaroa subgroup together. As shown in §30.3.2.3, they share the merger of \*C and \*d as c, and the merger of \*R and \*r as r, the first of which is exclusively shared, whereas the second is not.

### 30.3.3.3 Discussion of Phonological Classifications

Although Ho (1998) and Blust (1999b) worked independently on the subgrouping of the Formosan languages, they arrived at some similar conclusions, for instance, that Paiwan, Puyuma, and the Western Plains constitute major subgroups. They seem to have missed the fact that only three Formosan languages (Rukai, Paiwan, and Puyuma), all geographically located in the south, retain the distinction between \*k and \*g, which merged as *k* in all Formosan languages. Also, many subgroups (up to ten) are posited, and there is no center of the greatest linguistic diversity in Blust's subgrouping hypothesis. Chen et al. (2022) argue that both Malayo-Polynesian and East Formosan languages display the same innovative use of the prefix *ma-* (from an intransitive stative verb to a transitive verb) and the merger of \*t and \*C as *t*. These two groups should be viewed as sharing a common origin rather than as two separate primary subgroups.

### 30.3.4 Morphosyntactic Classifications

Morphosyntactic classifications have been proposed by Starosta (1995) (§30.3.4.1), Ross (2009) (§30.3.4.2), Zeitoun & Teng (2016) (§30.3.4.3), and Aldridge (2015, 2016, 2021) (§30.3.4.4).

#### 30.3.4.1 Starosta (1995)

Starosta (1995) proposes a subgrouping with binary branching based on top-down morphosyntactic innovations, in which Rukai constitutes the first offshoot from a hypothetical "Proto-Formosan" linguistic group that is ancestral to all Austronesian languages. Starosta's analysis is complex, and only the major points are summarized and shown in Figure 30.7. Starosta (1995, p. 698) assumes that Proto-Formosan was an ergative language, which had a developed set of auxiliary verbs and bound pronouns. In Rukai, a *kV*-relator noun fused with the determiners \*i and \*a in the formation of nominative determiners and demonstratives (ibid., p. 701). The UV voice system developed in F<sup>8</sup> through the fusion of the determiners \*a and \*i onto the verbs, Tsou having elaborated a complex system of auxiliary verbs. In Saaroa, the primary innovation was the development of the prefix *saa-* marking instrument, whose origin is unclear. By analogy, \*-en was innovated in the Chamorro language of Guam (F<sub>3</sub>) and replaced the earlier UVP suffix \*-a, which, in turn, was downgraded to subordinate clauses. In Kananavu, there were three innovations: first, the lexicaliza-

8 The sub-numeral that follows F relates to the position of a particular language in Starosta's subgrouping three.

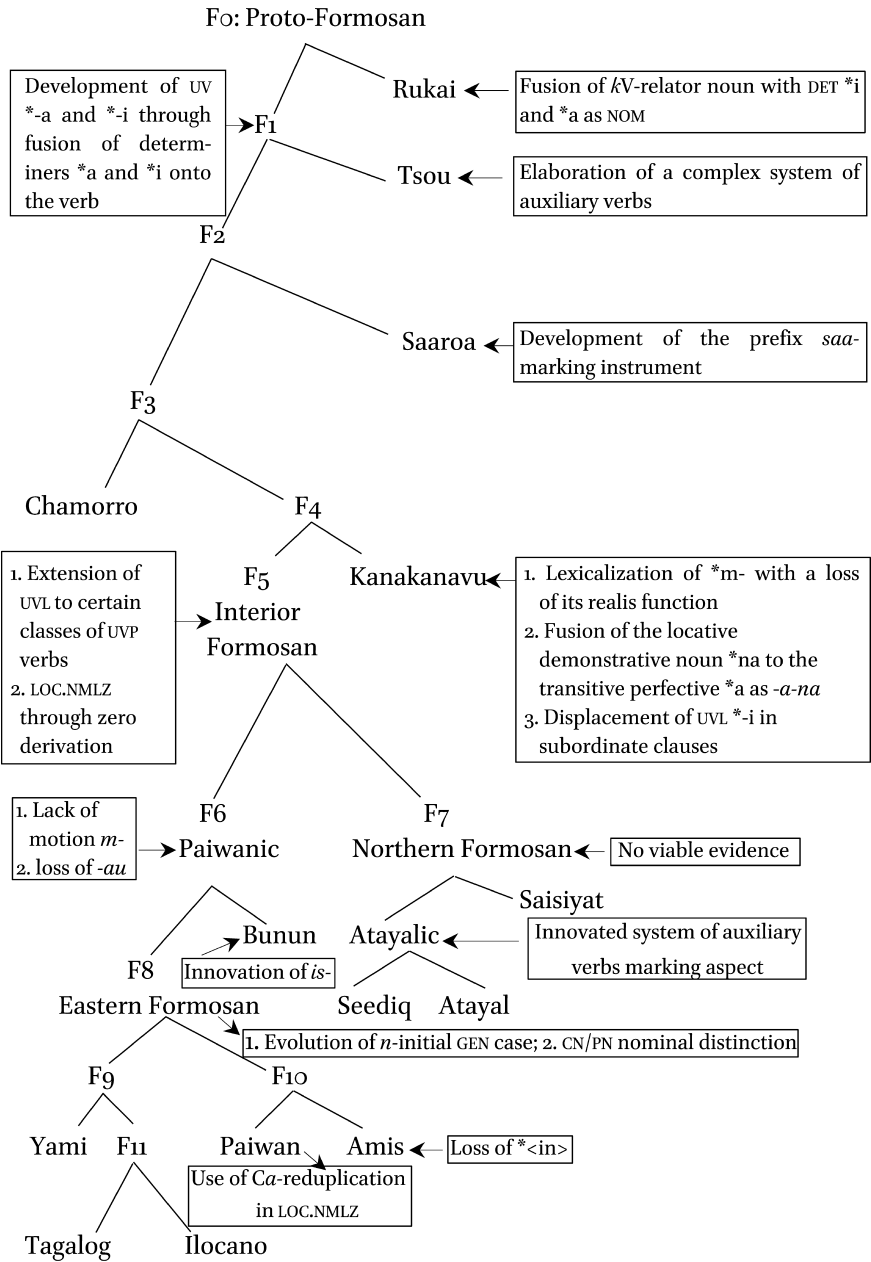


FIGURE 30.7 Starosta's grammar-based subgrouping  
REPRODUCED PARTLY FROM ZEITOUN & TENG 2016, P. 186 AND BASED  
ON STAROSTA 1995, P. 691FF.

tion of the intransitive realis prefix \**m-* with a loss of its realis function; second, the fusion of the locative demonstrative noun \**na* to the transitive perfective \**-a*, as *-a-na* (with the final *a* reanalyzed as an echo vowel); and third, the displacement in subordinate clauses of the suffix \**-i* marking UVL and found in Tsou and Saaroa.

The “Interior Formosan” subgroup (F<sub>5</sub>), “the common ancestor for Paiwanic and Northern Formosan”, is characterized by (i) the extension of UVL to certain classes of UVP verbs and (ii) the innovation of locative nominalization through zero derivation (ibid., p. 711). Starosta (ibid., p. 713) admits that there are no “shared positive grammatical innovations to justify grouping Saisiyat with the Atayalic subgroup.” The Atayalic subgroup (F<sub>7</sub>) (which subsumes Atayal and Seediq) is supported by an innovated system of auxiliary verbs marking aspect. The Paiwanic subgroup (F<sub>6</sub>) lacks the motion verb *m-u-* and the UVP suffix *-au* (ibid., p. 714), and Bunun has innovated *is-*. Eastern Formosan (F<sub>7</sub>) features “the evolution of the *n*-initial case form” and the development of the “common-personal distinction in the nouns” (ibid., p. 715). Paiwan is mostly characterized by use of *Ca*-reduplication in locative nominalization and Amis by the loss of the infix \*⟨in⟩.

Starosta’s (1995) subgrouping is based on rather bold hypotheses. Unfortunately, linguistic data were not as accessible at that time as they are now. This has had two consequences. First, Starosta did not take into account all Formosan languages: Puyuma does not appear in his tree. Second, this led him to make wrong predictions about certain languages: L. Li (2018, p. 11), for instance, mentions that “there is actually a motion verb prefix *m-u-* in [Isbukun] Bunun.”

Adelaar (2009, p. 406) summarizes Starosta’s methodological flaws as follows: “These morphological reconstructions may seem impressive, but the reality behind them is usually less straightforward than appears from Starosta’s presentation, which is based more on general typological considerations than on a painstaking application of the comparative method. Meanwhile, the groundwork of lexical and phonological comparison is apparently left to others.”

#### 30.3.4.2 Ross (2009)

Blust’s (1999b) hypothesis was challenged by Ross (2009), who posits four primary offshoots: Tsou, Rukai, Puyuma, and Nuclear Austronesian (including all other Austronesian languages). One major distinction between these two proposals is that while in Blust’s (1999b) hypothesis, any morpheme occurring in at least two of the ten primary subgroups can be reconstructed in PAN, a morpheme must occur in at least two languages (out of Puyuma, Tsou, and Rukai) and a Nuclear Austronesian language to be reconstructed at the highest (PAN) level; if it only occurs in languages at the NucAN level, then it can only be reconstructed at the PNucAN.



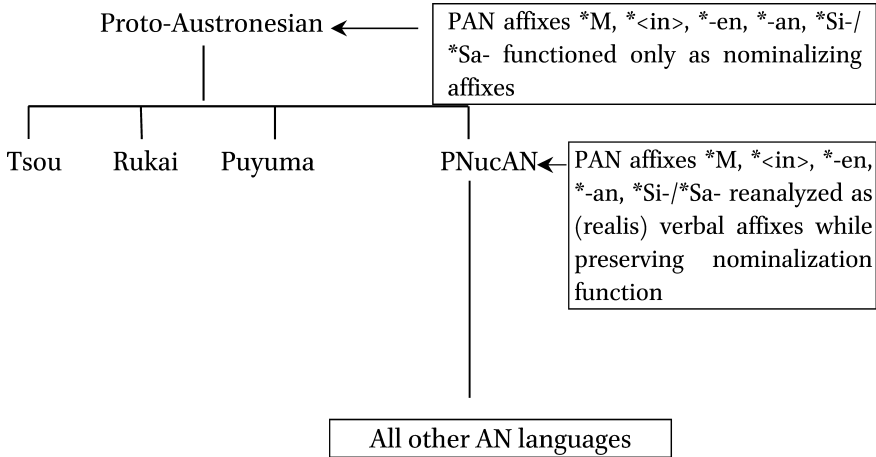


FIGURE 30.8 Ross's (2009) subgrouping hypothesis  
REPRODUCED FROM ZEITOUN & TENG 2016, P. 188

Ross' (2009) subgrouping is based on the “nominalization-to-verb” hypothesis, an innovation formerly proposed by Starosta et al. (1982), which assumes that the PAN affixes \*-en, \*<in>, \*-an, \*Sa-/\*Si- had only a nominalizing function in PAN—they were “first-generation affixes”—and acquired their verbal usage in PNucAN through reanalysis while preserving their nominalizing functions—as “second-generation affixes”—as depicted in Figures 30.8 and 30.9, respectively. Ross (2009) claims that it is difficult to draw a line between morphology and syntax, and the picture that emerges below relies on the assumption that PAN/PNucAN had an ergatively aligned clause structure, with two voices, AV and UV: AV-clause types were intransitive with only one (nominative) argument, and UV-clause types were transitive with two core arguments (the undergoer being marked as nominative and the actor as genitive).

Ross (2009) assumes that Puyuma verbal morphology reflects PAN morphology, as in (7), while Tsou reflects only the PAN dependent verb forms and lacks reflexes of both the verbal forms and the nominalizing affixes, as in (8).

- (7) Nanwang Puyuma (Teng 2008, p. 109)
- a. *tr<em>akaw dra paisu i Isaw.*  
 <AV>steal OBL.INDF money NOM.SG Isaw  
 ‘Isaw stole money.’ (Realis AV)
- b. *tu=trakaw-aw na paisu kan Isaw.*  
 3.GEN=steal-UVF NOM.DEF money OBL.SG Isaw  
 ‘Isaw stole money.’ (Realis UVP)

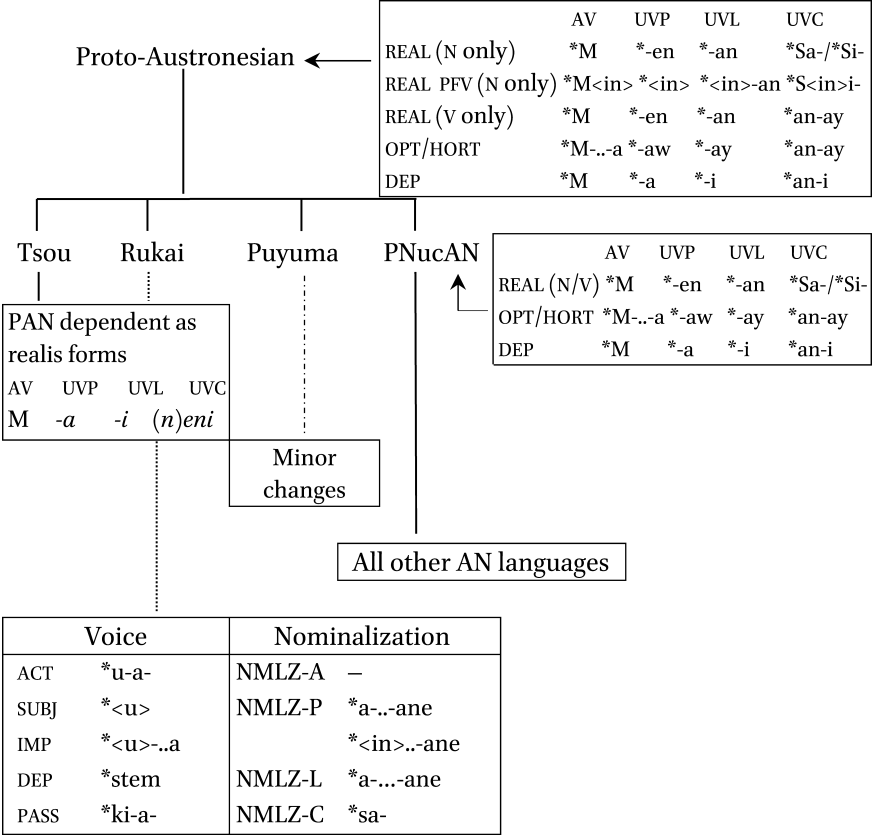


FIGURE 30.9 Ross’s (2009) subgrouping hypothesis  
REPRODUCED FROM ZEITOUN & TENG 2016, P. 188

- c. *tu=trakaw-ay=ku dra paisu kan Isaw.*  
3.GEN=steal-UVL=1SG.NOM OBL.INDF money OBL.SG Isaw  
‘Isaw stole money from me.’ (Realis UVL)
- d. *tu=trakaw-anay i tinataw dra paisu.*  
3.GEN=steal-UVC NOM.SG his.mother OBL.INDF money  
‘Isaw stole money for his mother.’ (Realis UVP)
- (8) Tsou
- a. *m-o m-osi ta pangka to emi ’o amo.*  
AV.REAL.IMM AV-put OBL table OBL wine NOM father  
‘Father put some wine on the table.’ (Zeitoun 1992, p. 3)

- b. *i=si* *si-a* *ta* *pangka to amo 'o*  
 UV.REAL.IMM=3SG.GEN put-UV P OBL table OBL father NOM  
*emi.*  
 wine  
 'The wine was put by father on the table.' (ibid.)
- c. *i=si* *si-i* *ta emi to amo 'e*  
 UV.REAL.IMM=3SG.GEN put-UVL OBL wine OBL father NOM  
*pangka.*  
 table  
 'The table is where father put some wine.' (ibid.)
- d. *i=ta* *si-eni* *to naveu to takubingi 'o*  
 UV.REAL.IMM=3SG.GEN put-UVC OBL rice OBL bowl NOM  
*ba'i.*  
 grandmother  
 'He used a bowl and filled it with rice for grandmother.' (Tsou e-dictionary)

Ross (2009) assumes that Proto-Nuclear Austronesian verbal morphology was largely as in present-day Paiwan, as in (9):

- (9) Puljetji Paiwan (Huang 2012)
- a. *na=k<em>an=aken* *ta demangasan.*  
 PFV=<AV>eat=1SG.NOM OBL goat  
 'I ate goat (meat).' (Realis AV) (p. 97)
- b. *dj<in>adjas=anga* *a ma-drusa a c<em>akav nazua*  
 <PFV.UVP>catch=COS NOM PN-TWO LIG <AV>steal GEN.that  
*a kisac.*  
 LIG police  
 'That police officer caught two thieves.' (Realis UVP) (p. 43)
- c. *p<in>acun-an niamadju azua* *a cemakav.*  
 <PFV>see-UVL 3PL.GEN NOM.that LIG thief  
 'They saw that thief.' (Realis UVL) (p. 43)
- d. *si-vali=anga* *a ku=tapav.*  
 UVC-blow=COS SPEC? 1SG.GEN=hut  
 'My hut was blown away.' (Realis UVC) (p. 113)

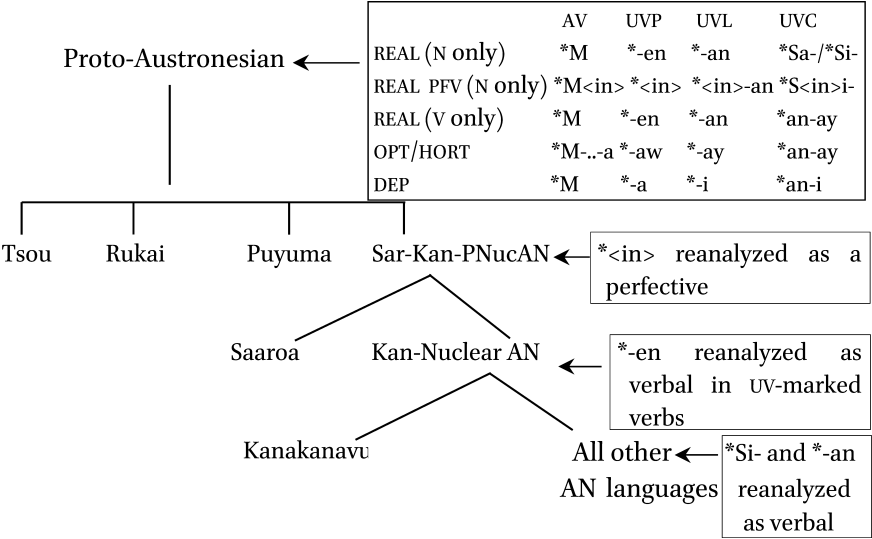


FIGURE 30.10    Zeitoun & Teng’s revision of Ross’s subgrouping hypothesis  
                      BASED ON ZEITOUN & TENG 2016, P. 195

30.3.4.3    Zeitoun & Teng (2016)

Zeitoun & Teng (2016) take as a point of departure Ross’s (2009) subgrouping hypothesis and show that while Kananavu and Saaroa are subsumed under PNucAN, identified on the basis of the “nominalization-to-verb” innovation, both languages have only partially reanalyzed second-generation suffixes: in Kananavu, the reflex of \*-en was reanalyzed as a verbal marker encoding UV, and the reflex of \*<in> is a perfective and a UV voice marker/nominalizing formative. In Saaroa, the reflex of \*<in> ‘PFV’ can only occur in AV-marked verbs. It is otherwise a nominalizing formative. In both languages, reflexes of \*-an and \*Si- are only used as nominalizers and were never reanalyzed as verbal affixes.

30.3.4.4    Aldridge (2015, 2016, 2021)

While Ross (2009) infers that PAN had ergatively aligned clauses, both Starosta (1995), within the Lexicase theory, and Aldridge (2015, 2016 and 2021), based on the Minimalist framework, assume that Rukai, the first offshoot of the Austronesian language family, was an accusative language and preserves the alignment that must be reconstructed at the PAN level (see Figure 30.11). She argues that the emergence of ergativity was the result of the reanalysis of irrealis clauses through a detransitivization process in “Proto-Ergative-AN” (PEAN). More specifically, Aldridge (2016) claims that ergative alignment arises when a transitive verb, unable to structurally license an object, selects it as a subject if the external argument does not take case from T.

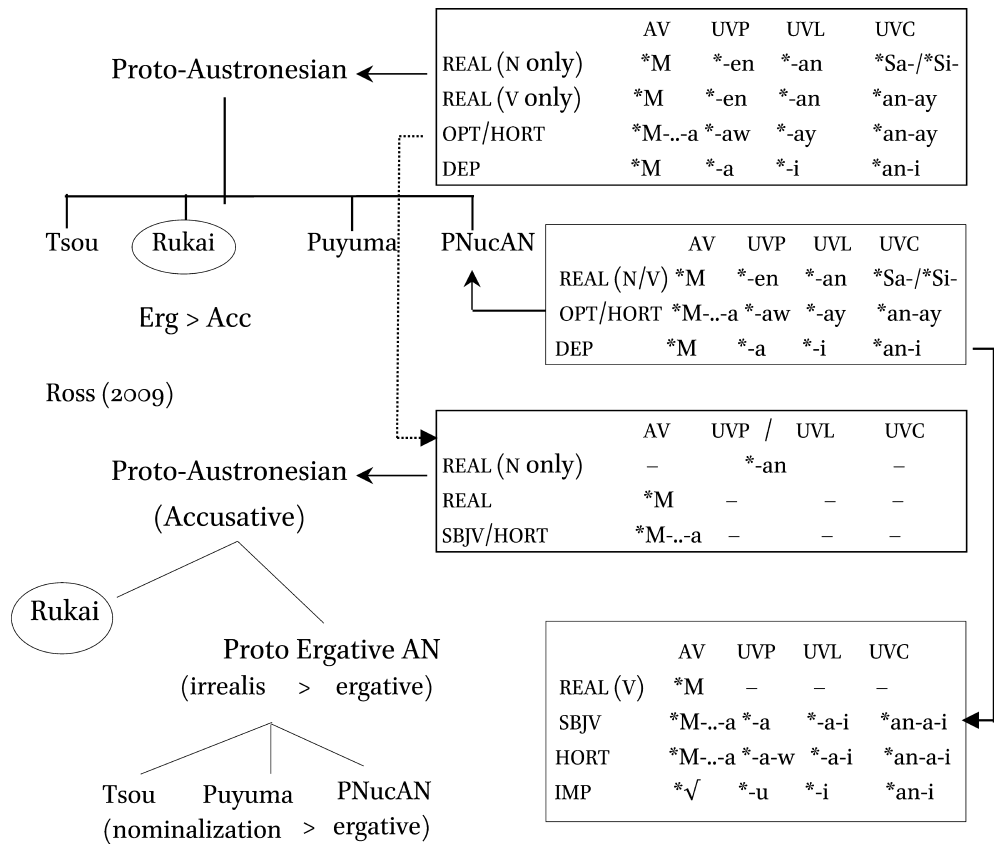


FIGURE 30.11 Aldridge's (2016) revision of Ross's subgrouping hypothesis

30.3.5 Numeral-Based Classification

Sagart (2004, p. 415) assumes that the PAN numeral system was quinary, with “stable words for numerals up to ‘5.’” According to him, there were no stable words for ‘6,’ ‘7,’ ‘8,’ and ‘9,’ and these were coined using additive, multiplicative, and subtractive strategies, e.g., 5 + 1 = 6, 4 × 2 = 8. On the basis of this assumption, Sagart proposes that PAN split into a language ancestral to Pazeh, Saisiyat, and Luilang, which formed, along with “Pituih” (comprising Atayalic, Thao, Favorlang, Taokas, Siraya, Papora, and Hoanya), primary branches; the other Formosan languages fell into nonprimary groups, called respectively, “Enemish” (a language ancestral to Siraya), “Walu-Siwaish” (ancestral to Tsouic, Paiwan, Rukai, Puyuma, Amis, and Bunun), and Muish (ancestral to all other Austronesian languages), with each of these groups reflecting the innovations for the numerals ‘7,’ ‘8,’ and ‘9,’ derived from the longer additive forms 5 + 2, 5 + 3, 5 + 4, retained in Pazeh, along with two lexical innovations: the politeness

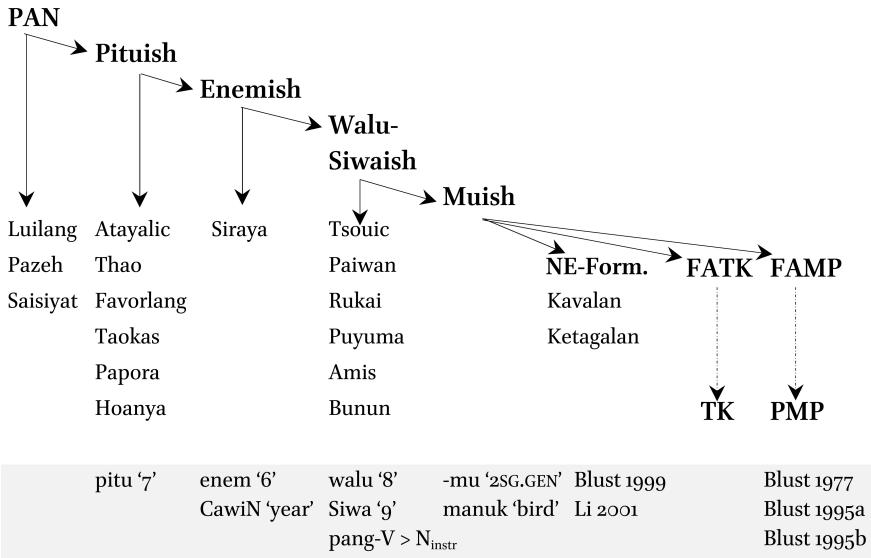


FIGURE 30.12 Sagart's (2004, p. 431) higher Austronesian phylogeny

shift (PAN -mu '2PL.GEN' > '2SG.GEN') and the replacement of PAN \*qayam by \*manuk 'bird'. Sagart's (2004, p. 431) subgrouping hypothesis is reproduced in Figure 30.12.

Sagart's (2004) phylogeny is not unproblematic, as discussed by Blust (2013, pp. 752–754). It is based on a reconstruction of numerals, supposedly forming a quinary system in PAN, based on only two languages, Pazeh and Saisiyat. However, all the other Formosan languages have a decimal or modified decimal systems, including Rukai, Tsou, and Puyuma, treated as first offshoots in Blust (1999b) and Ross (2009).

30.3.6 Summary

As shown in the previous sections, the theoretical and analytical assumptions and methodologies of the scholars concerned, as well as the viability of the types of evidence brought forward to support each proposal, make it difficult to reach a consensus regarding the classification of the Formosan languages. But probably every scholar writing in this century who has referred to Formosan classification has mentioned Blust (1999b) because, following the Indo-Europeanist example, shared phonological innovations are widely regarded as the soundest foundation for subgrouping.

### 30.4 Discussion of Controversial Subgroups

In this section, we discuss four controversial subgrouping hypotheses: Tsouic (§ 30.4.1), Atayal-Seediq (§ 30.4.2), Amis-Sakizaya (§ 30.4.3), and Sirayaic (§ 30.4.4).

#### 30.4.1 *Tsouic*

The term “Tsouic” is a linguistic construct. Though it was accepted *de facto* for years, from the very beginning, it was never shown to be a viable subgroup. Ogawa & Asai (1935, p. 3ff.) treat Kanakanavu and Saaroa as dialects of Tsou. Dyen (1963, p. 263), however, argues that “their relation is that of closely related languages [rather] than of dialects of the same language”. On the basis of Japanese-era ethnolinguistic data and analysis, he also suggests that “whether they form a group or not, the [...] comparisons [that he gives] suggest a connection between the three languages” (ibid., p. 266).

The term “Tsouic” was conventionalized by Ferrell (1969), who showed that Tsou, Kanakanavu, and Saaroa differ not only from other Formosan languages in their overall linguistic and cultural complexity but also from one another (ibid., p. 36). He mentions that extensive borrowing must have taken place among Saaroa, Siraya, and Rukai: “Saaroa is lexically as near to the (Paiwanic) Siraya as to Tsou, although the known close contacts between Saaroa, Siraya and Rukai make it most likely that extensive vocabulary resemblances with these languages are due to borrowing by Saaroa rather than indicating that Tsouic and Paiwanic languages are directly linked genetically” (Ferrell 1969, p. 39). He also insists that Tsou is structurally more different from Kanakanavu and Saaroa than they are from each other: “Grammatically, lexically and phonologically, Tsou is by far the most aberrant of all Formosan languages, leading us to suspect that its separation from the ancestors of the other Formosan languages was at a very remote period indeed [...] Tsou linguistic peculiarities are shared [...] to a limited extent by Kanakana[v]u and Saaroa” (ibid.).

Like his predecessors, Tsuchida (1976, pp. 1–10) holds the assumption that Kanakanavu and Saaroa form a distinctive subgroup called “Tsouic” and notes that Kanakanavu and Saaroa are more closely related to each other and form the “Southern Tsouic” subgroup (as opposed to Tsou, commonly referred to as “Northern Tsou”).

Ross (2012, p. 1303ff.) further reassesses the innovations taken by Tsuchida (1976) as evidence for Tsouic and concludes that most of them are not valid: they may have been borrowed or developed independently in different members of the group. Thus, he concludes that they cannot be taken as evidence for a Tsouic subgroup. Sagart (2014), responding to Ross, argues that the Tsouic lan-

guages exclusively share at least one sporadic change and one irregular sound change. The first consists of the metathesis of PAN \*pataS ‘tattoo, write’ as PT tapaSə, reflected as Kanakanavu *tapásu*, Saaroa *taa-tapa-a*, Tsou *ta-tpos-a* ‘pattern, design’. The second involves the split of PAN \*C into PT \*t and \*C, cf. PAN \*Caqi ‘excrement’ > PT \*táʔ<sub>3</sub>i (reflected as Kanakanavu *táaʔi*, Saaroa *tiiʔi*, Tsou *tʔee* ‘excrement’). Sagart (2014, p. 876) also lists 57 lexical items that are exclusively shared by Tsou, Kanakanavu, and/or Saaroa, and are not found in any other language.

The foregoing discussion shows that to date, the linguistic and cultural relationships of these three ethnolinguistic groups, their migration, and their history remain unclear. No convincing exclusively shared phonological innovations have been found among these three languages that would characterize “Tsouic” as a valid subgroup. Exclusively shared phonological innovations between Kanakanavu and Saaroa show that they are more closely related; morphosyntactic evidence indicates that they do not form a subgroup (Ross 2009, 2012, Zeitoun & Teng 2016). Despite the fact that there are few shared innovations between Kanakanavu and Tsou on the one hand and Saaroa and Tsou on the other, we cannot currently exclude the possibility that these three languages might be closely related, but the situation is rather complex, as there is a strong possibility that there was extensive borrowing within the putative “Tsouic” group and between Southern Tsouic and adjacent languages.

30.4.2 Atayal-Seediq

Atayal is most closely related to Seediq. The main linguistic evidence for the Atayalic subgroup is the distinction between male and female forms of speech, which are found in all the dialects of these two languages (P. Li 1982) and are further illustrated in (10).

(10) Male and female forms of speech in Atayalic (Li 2004, p. 1483ff.)

Gloss		‘fire’		‘louse’		‘eat’		‘tree’	
PAN		*Sapuy		*kuCuh		*kaen		*kaSuy	
		M	F	M	F	M	F	M	F
Atayal	Squliq	<i>puniq</i>	–	<i>kuhiŋ</i>	–	<i>maniq</i>	–	<i>qhoniq</i>	–
	C’uli’	<i>hapuniq</i>	<i>hapuy</i>	<i>kuhiŋ</i>	<i>kucuʔ</i>	<i>maniq</i>	–	<i>kahuniq</i>	<i>kahuy</i>
Seediq	Tgdaya	<i>puniq</i>	–	<i>quhiŋ</i>	–	–	<i>mekan</i>	<i>qhuni</i>	–
	Toda	<i>puniq</i>	–	<i>quhiŋ</i>	–	–	<i>məkan</i>	<i>qhuni</i>	–
	Truku	<i>puniq</i>	–	<i>quhiŋ</i>	–	–	<i>məkan</i>	<i>qhuni</i>	–



The classification by linguists may be different from the concept that indigenous people have of an ethnic group, however. After the lifting of martial law in 1987, there was a proliferation of indigenous rights movements in Taiwan. In the mid-1990s, the Truku Seediq, located in Hualien County, launched a “Name Rectification Campaign”, with the aim to separate from the Atayal, with whom they thought they had been identified for too long (Wang 2008, p. 6). They were not able to coordinate with the Seediq of Nantou County, so while the movement resulted in the official recognition of the Truku as a separate ethnic group in 2004, the Seediq were not recognized until 2008.

### 30.4.3 *Amis-Sakizaya*

Amis stretches from Hualien to Taitung and is divided into five major groups: Sakizaya, Northern, Tavalong-Vata'an, Central, and Southern, which are lexically distinct. Phonological innovations are very sporadic. For example, the change *su-* > *hu-* applies only to a few lexical forms and provides crucial evidence for dialectal distinction. It is questionable whether Amis can be divided into two main groups—North (which includes Sakizaya on the one hand and Northern and Tavalong-Vata'an on the other) and South (subsuming Central and Southern)—or whether Sakizaya should be treated as the first offshoot (see Li & Tsuchida 2022).

Linguistically speaking, Sakizaya is definitely a dialect of Amis (*contra* Lin 2022), even though it was officially recognized as a separate ethnic group by the Council of Indigenous Peoples in 2007.

### 30.4.4 *Sirayaic*

Siraya, Taivuan, and Makatau were the Formosan languages or dialects formerly spoken in the southwestern plains of Taiwan. Roughly speaking, Siraya was spoken in the coastal area of Tainan Plain and Taivuan mostly inland from Tainan Plain to the north, while Makatau was spoken in Kaohsiung and Pingtung prefectures to the south. These languages or dialects probably became extinct in mid-19th century but were recorded by Dutch missionaries between 1624 and 1661 (see Adelaar, this handbook, Chapter 57).

The three speech varieties have different reflexes of PAN \*l, \*N, \*D, \*-k-, and \*-S-/\*-R- (Li 2009, pp. 400, 402, 404, 405).

(11)	PAN	Siraya	Taivuan	Makatau	Gloss
a.	*l	r	Ø ~ h	r	
	*telu	<i>туру</i>	<i>too, toho</i>	<i>toru</i>	'three'
b.	*N	l	l	n	
	*(qa)Nuang	<i>luang</i>	<i>lowan</i>	<i>noang</i>	'cow'
c.	*D	s	r ~ d	r ~ d	
	*Daya	<i>saija</i>	<i>raijsa</i>		'east'
d.	*-k-	-k-	Ø	-k- ~ Ø	
		<i>akosai</i>	<i>ausaij</i>	<i>akusaij</i>	'not have'
		<i>tarokaij</i>	<i>taroaej</i>	<i>tarauwei</i>	'given name'
e.	*-S-, *-R-	-g-	Ø	-	
		<i>dagogh</i>	<i>daoh</i>		'price'
		<i>ligig</i>	<i>liih</i>		'sand'

The sound change PAN \*D, \*d, \*Z > s in Siraya, > r ~ d in Taivuan and Makatau must have taken place prior to the Dutch occupation of Taiwan (1624–1662), as the phonological difference is manifested in the Dutch missionary documents: s is found in van der Vlis (1842) for Siraya vs. r ~ d in Taivuan in Gravius (1661, 1662).

The change \*-S-, \*-R- > x (written as *g, gh, h*) or Ø may have started in the early 17th century, because the rule applies to some lexical forms, but not to others containing the consonant even in the same set of language data as recorded by the Dutch missionaries.

PAN \*l is reflected as *r* rather than *h* or Ø, as shown in (12), from Saint Matthew, e.g., \*lahuD > *raour* 'west', \*piliq > *peri* 'to choose', \*kalih > *k<m>ari* 'to dig'. PAN \*l was still reflected as *r* in the mid-17th century, when the Gospels of Saint Matthew and Saint John were translated. In fact, it was still reflected as *r* in a Wanli (Taivuan) text dated 1770 (Li 2010, p. 565), as in *raur* 'west' < \*lahuD, and as *h* in a Makatau (Taivuan) text dated 1781, as in *mi-likoh* 'to return' < \*likuD (Li 2010, p. 538). It was not lost until much later, when the Japanese started to investigate the languages of the southwestern plains in 1897. The sound change \*l > *h* or Ø was a late innovation limited to Taivuan.

30.5 Internal Relationships of Formosan Languages

In this section, we review the internal relationships of five Formosan languages, Atayal (§ 30.5.1), Bunun (§ 30.5.2), Rukai (§ 30.5.3), Paiwan (§ 30.5.4), and Puyuma (§ 30.5.5). We exclude from this discussion Saisiyat, Tsou, and Pazeh-Kaxabu because although each of these three languages can be divided into two or three dialects, these dialects show very few phonological and/or lexical distinctions, as discussed in the grammatical sketch of each language in this handbook.

30.5.1 Atayal

Atayal has two major dialects, Squliq and C'uli', which are distinguished lexically, phonologically, and morphosyntactically (Li 1998).

Most Atayal dialects share essentially the same vocabulary stock except for a few dozen items that are completely or partially different, as shown in (12).

(12)	'chicken'	'shoulder'	'sweat'	'swallow'	'lie'	'fish'	'tumor'
Squliq	<i>ngta'</i>	<i>qhiyang</i>	<i>yabux</i>	<i>mqum</i>	<i>brus</i>	<i>qulih</i>	<i>pangih</i>
C'uli'	<i>wailung</i>	<i>hngali'</i>	<i>rinang</i>	<i>qmtam</i>	<i>'ihuy</i>	<i>qcyux</i>	<i>qilis</i>

Phonologically, Squliq and C'uli' exhibit three major differences:

1. Squliq /s/ corresponds to C'uli' /c/  
                  'sweet'                  'person'          'that'          'I'  
Squliq   *sbing*                  *squliq*   *hasa*          *saku'*  
C'uli'    *cbing*                  *cquliq*   *haca*          *caku'*
2. Squliq /r/ corresponds to C'uli' /s/  
                  'how many'          'later'          'nine'  
Squliq   *pira'*                  *kira'*   *qiru'*  
C'uli'    *pisa'*                  *kisa'*   *qisu'*
3. Squliq /-ʔ/ corresponds to C'uli' /-t/ or /-c/  
                  'drift'                  'ash'          'leopard'   'rat'  
Squliq   *mqlui'*                  *qbuli'*   *rkli'*   *qoli'*  
C'uli'    *mqliut*                  *qbulit*   *rklit*   *qolit*

Squliq and C'uli' exhibit at least two major morphosyntactic differences, (1) Squliq *ku'*: C'uli' *cu* or *ci* 'I (NOM)' and (2) Squliq *nyux/cyux*: C'uli' *kia'*/*hani'an* 'PROG' (with a near/remote distinction).

Goderich (2020) proposes an entirely different division of the Atayal dialects, based on phonological and lexical innovations. He proposes dividing them into two main groups, Northern and Southern. The Northern subgroup comprises Matu'uwal (also known as Mayrinax), Squliq, and Skikun, and is supported by the merger of Proto-Atayal word-final \*-lit and \*-liʔ. The Southern group consists of Pngawan, Klesan, S'uli, and Matu'aw, which share the merger of Proto-Atayal \*q and \*ʔ. The main difference between Tsuchida's (1980) and Goderich's (2020) proposals concerns the position of Skikun and Matu'uwal, which are treated as C'uli' dialects by Tsuchida (1980) but as closer to Squliq by Goderich (2020).

### 30.5.2 *Bunun*

Bunun has five dialects: Takituduh, Takibakha, Takbanuaz, Takivatan, and Isbukun. They form three main groups (Ogawa & Asai 1935, Li 1988, Shibata 2020): Takituduh and Takibakha, which are known as the Northern dialects; Takbanuaz and Takivatan, the Central dialects; and Isbukun, the most divergent dialect, with dialectal varieties spoken in Nantou, Kaohsiung, and Taitung. Takituduh and Takibakha retain the distinction between Proto-Bunun (PB) \*c and \*s, which has been lost in the three other dialects. On the basis of phonological and lexical evidence, it can be hypothesized that Isbukun split off from Proto-Bunun first. It has undergone three phonological changes: PB \*q is reflected as /h/; PB \*ʔ is lost intervocally; and PB \*h has become /ʔ/ or is lost after a consonant.

### 30.5.3 *Rukai*

The internal relationships of the Rukai dialects have, to date, not been completely clarified. Since Li's (1977) reconstruction of Proto-Rukai (PR), it has generally been accepted that there are two distinct groups of dialects: Tanan, Labuan, and Budai on the one hand and Maga and Tona on the other. The position of the Mantauren dialect of Rukai, however, remains controversial, because the structure of the language has been obscured by drastic phonological and syntactic changes. Li (1977) shows that phonologically, PR voiced stops have been weakened and spirantized. Hence, PR \*b, \*d/\*d, and \*g correspond to Mt Rukai /v/, /ð/, and /h/, respectively. Syntactically, Mantauren exhibits mostly bound pronouns and has developed verb-object agreement. At least three hypotheses have been proposed: (1) Li (1977) suggests that Mantauren subgroups immediately with Maga and Tona because they exclusively

share 94 cognates against 44 cognates with Budai and Tanan; (2) Li (1996) follows Tu (1994) in assuming that Mantauran is the first offshoot of Proto-Rukai; and (3) Zeitoun (2007, forthcoming) claims that Mantauran groups with Labuan/Tanan and Budai based on semantic and morphosyntactic innovations, including, among others, CVV-reduplication to mark the habitual aspect and express comparatives and superlatives, the occurrence of the first-person singular pronoun *nao*= to express a volitional/modal agent, the occurrence of the first-person plural pronoun =*nai*, and verbal modification by *toramoro* ‘very’.

### 30.5.4 *Paiwan*

Paiwan is spoken in Pingtung and Taitung Counties in southern Taiwan. It includes various dialects, which have diverged not only phonologically but also lexically.

Ho (1978) proposes that on the basis of phonology, the dialects of Paiwan divide into two groups. The “Northwestern group” (or so-called dental group), which includes Se Paiwan and Timur, is characterized by the merger of the palatals *tj* and *dj* with their dental counterparts, *t* and *d*, respectively. The “Southeastern group” (or palatal group) includes Butanglu, Tjuabar, and Tjuvuali, which have retained the two palatals *tj* and *dj* as distinct from the dentals *t* and *d*, respectively.

Cheng (2016a, 2016b) generally follows Ho (1978) in considering the presence or absence of the palatals *tj* and *dj* as a primary feature allowing us to divide Paiwan into two major groups, “Northern” and “Southern”. He considers the absence of /q/ and /k/ and the change of the retroflex /ɭ/ to a velar fricative /ɣ/ as secondary changes. He also lists eleven areal sound changes and seven features found in a specific communalelect. He shows that the Northern group divides into five branches: Paridrayan, Timur, Ulaljac, and Eastern—none of which exhibits /q/—and a fifth branch that retains /q/. The Southern group divides into three types: the first has /ɣ/ in place of the lateral retroflex /ɭ/ and includes a western and eastern branch; the second lacks /k/ and is divided into three branches (western, eastern, and southern); and the third, which he calls the Sinvaudjan-Kuljaljau type, is the most conservative. Overall, this classification mixes synchronic and diachronic criteria and may not be valid.

Ferrell & Tjakisuvung (forthcoming) warn that Paiwan dialect classifications are currently based on lexical data from only a handful of the 100+ traditional Paiwan settlements, but that complex earlier migration patterns, along with the extensive relocation of villages away from their traditional regions since the 1950s, make geographical dialect classifications largely useless. They propose a tentative four-way phonological classification: (i) “core dialect group”,

(ii) “t-merger dialect group”, (iii) “k-to-glottal stop dialect group”, and (iv) “uvular fricative dialect group”.

Communalects of the “core dialect group” are considered conservative because they show no distinctive mergers of Proto-Paiwan (PPai) phonemes. The best documented members of this group are Kuljaljau in the central area and Tjuaqatsiljay in the south.

The “t-merger dialect group” is characterized by the merger of PPai *\*c/\*t* (< PAN *\*t/\*s*) as /t/, and PPai *\*d/\*ʃ* (< PAN *\*j/\*z*) as /d/ and include the dialects of Stimul, Makazayazaya, Se Paiwan, and Kazangilan. All of these dialects belong to the original Vutsul (or Butsul) subgroup of the northwest.

PPai *\*k* has become a glottal stop /ʔ/ in Tjala'avus (Tjalakavus) and Tja'u'uvu'u'vul (Tjakuvukuvul), which belong to the “k-to-glottal stop dialect group”. However, they all retain /k/ in a handful of words typically including *dikitj* ‘short’, *iku* ‘tail’, *m-ekel* ‘run’, *kapaz* ‘root’, *kedri* ‘little’, *kuka* ‘chicken’, and *teku* ‘below’.

In the “uvular fricative dialect group”, the PPai alveolar trill *\*r* is reflected as a uvular fricative [ʁ]. In some dialects, the flap *\*l* is realized as a velar fricative [ɣ], and in Laleklek, the PPai alveolar stop *\*d* has become a dental fricative [ð]. Members of this group include Butanglu, Pacaval (Daniao), and Tjaridik.

### 30.5.5 Puyuma

Ting (1978) classifies the Puyuma dialects into two main groups, Nanwang vs. all the other dialects based on a phonological innovation; Nanwang retains the voiced stops /b, d, ɖ, g/, whereas all other dialects reflect these as /β~v, ð, ɗ, h/, respectively.

On the basis of their morphosyntactic properties, including case syncretism, NP conjunction, and pronominal paradigms, Teng (2009, 2011, and 2015) suggests that at least three dialect groups should be distinguished: (1) Nanwang, (2) Katripul and Kasavakan, and (3) Ulivelivek, which consists of the variants as spoken in Ulivelivek, Tamalakaw, Rikavung, Pinaski, and Alipay.

### 30.5.6 Summary

Increased in-depth documentation of each of these five languages will allow us to better understand the internal relationships of their dialects. It is unfortunate that the Formosan languages are undergoing changes and disappearing at an unprecedented rate, and more research is urgently needed to understand the complexities of the extant languages.

### 30.6 Conclusion

In this chapter, we have reviewed 11 of the more than 25 subgrouping hypotheses regarding the Formosan languages, trying to outline different methodologies and findings. We have also discussed four controversial subgrouping hypotheses (Atayal-Seediq, Tsouic, Amis-Sakizaya, and Sirayaic) and examined the internal relationships of the dialects of five languages: Atayal, Bunun, Rukai, Paiwan, and Puyuma.

### References

- Adelaar, Alexander (2009). Review of “Laurent Sagart, Roger M. Blench, and Alicia Sanchez-Mazas (Eds.), *The peopling of East Asia: Putting together archaeology, linguistics, and genetics*. London, New York: RoutledgeCurzon, 2005, xxiii + 323 pp. ISBN 0415322421. Price: USD 200.00 (hardback).” *Bijdragen tot de Taal-, Land- en Volkenkunde* 165(2–3), 401–408.
- Adelaar, Alexander (this handbook). Siraya. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Aldridge, Edith (2015). A Minimalist approach to the emergence of ergativity in Austronesian languages. *Linguistics Vanguard* 1(1), 313–326.
- Aldridge, Edith (2016). Ergativity from subjunctive in Austronesian languages. *Language and Linguistics* 17(1), 27–62.
- Aldridge, Edith (2021). Syntactic conditions on accusative to ergative alignment change in Austronesian languages. *Journal of Historical Linguistics* 11(2), 214–247.
- Asai, Erin (1936). A study of the Yami language: An Indonesian language spoken on Botel Tobago Island. PhD dissertation. Leiden: University of Leiden.
- Bellwood, Peter (1991). The Austronesian dispersal and the origin of languages. *Scientific American* 265(1), 88–93.
- Blust, Robert (1977). The Proto-Austronesian pronouns and Austronesian subgrouping: A preliminary report. *University of Hawai'i Working Papers in Linguistics* 9(2), 1–15.
- Blust, Robert (1984–1985). The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26(1), 45–67.
- Blust, Robert (1993). \*S metathesis and the Formosan/Malayo-Polynesian language boundary. In Oyvind Dahl (Ed.), *Language—a doorway between human cultures: Tributes to Dr. Otto Chr. Dahl on his ninetieth birthday*, (pp. 178–183). Oslo: Novus.
- Blust, Robert (1995a). Sibilant assimilation in Formosan languages and the Proto-Austronesian word for ‘nine’. *Oceanic Linguistics* 34(2), 443–453.
- Blust, Robert (1995b). The position of the Formosan languages: Method and theory

- in Austronesian comparative linguistics. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng (Eds.), *Austronesian Studies Relating to Taiwan*, (pp. 585–650). Symposium Series of the Institute of History and Philology, Academia Sinica No. 3. Taipei: Academia Sinica.
- Blust, Robert (1996). Some remarks on the linguistic position of Thao. *Oceanic Linguistics* 35(2), 272–294.
- Blust, Robert (1999a). Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2), 321–365.
- Blust, Robert (1999b). 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*, (pp. 31–94). Symposium Series of the Institute of Linguistics (Preparatory Office), Institute of Linguistics, No. 1. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert A. (2013). *The Austronesian languages*. Asia-Pacific Linguistics Open Access Monographs A-PL 008. Canberra: The Australian National University. (revised ed.). Retrieved from <https://openresearch-repository.anu.edu.au/bitstream/1885/10191/6/Blust-2013-AustronesianLanguages.pdf>.
- Chen, Victoria, Jonathan Kuo, Maria Kristina S. Gallego and Isaac Stead (2022). Is Malayo-Polynesian a primary branch of Austronesian? *Diachronica* 39(4), 449–489.
- Cheng, Chung-hua 鄭仲樺 (2016a). Paiwanyu fangyan yanjiu 排灣語方言研究 [A study of Paiwan dialects]. Boshi lunwen 博士論文 [PhD dissertation]. Beijing 北京 [Peking]: Beijing daxue 北京大學 [Peking University].
- Cheng, Chung-hua 鄭仲樺 (2016b). Fangyan dili he yuyanxue shijiao de paiwan zuqun fenlei 方言地理和語言學視角的排灣族群分類 [The classification of Paiwan from a dialect geography and linguistic perspective]. *Taiwan yuanzhuminzu yanjiu jikan* 臺灣原住民族研究季刊 [*Taiwan Journal of Indigenous Studies*] 9(2), 55–89.
- Clark, Ross (1987). Austronesian languages, in Bernard Comrie (Ed.), *The world's major languages*, (pp. 899–912). New York: Oxford University Press.
- Dahl, Otto Christian (1973). *Proto-Austronesian*. Oslo: Studentlitteratur.
- Dahl, Otto Christian (1981). *Early phonetic and phonemic changes in Proto-Austronesian*. Oslo: The Institute of Comparative Research in Human Culture.
- Dyen, Isidore (1963). The position of the Malayopolynesian languages of Formosa. *Asia Perspectives* 7(1–2), 261–271.
- Dyen, Isidore (1965). Formosan evidence for some new proto-Austronesian phonemes. *Lingua* 14, 285–305.
- Dyen, Isidore (1971). The Austronesian languages of Formosa. In Thomas A. Sebeok (Ed.) *Linguistics in Oceania*, (pp. 168–199). Current Trends in Linguistics 8. The Hague: Mouton.



- Dyen, Isidore (1990). Homomeric lexical classification. Philip Baldi (Ed.), *Linguistic change and reconstruction methodology*, (pp. 211–230). Trends in Linguistics Studies and Monographs 45. Berlin: Mouton de Gruyter.
- Ferrell, Raleigh (1969). *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica Monograph No. 17. Taipei: Institute of Ethnology, Academia Sinica.
- Ferrell, Raleigh (1980). Phonological subgrouping of Formosan languages. In Naylor Paz B. (Ed.), *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*, (pp. 241–254). Michigan Papers on South and Southeast Asia, Center for South and Southeast Asian Studies 15. Ann Arbor: University of Michigan.
- Ferrell, Raleigh and Drungdrung Tjakisuvung (forthcoming). Paiwan texts. MS.
- Goderich, Andre (2020). Atayal phonology, reconstruction and subgrouping. PhD dissertation. Hsinchu: National Tsing Hua University.
- Gravius, Daniël (1661). *Het Heylige Evangelium Matthei en Johannis. Ofte Hagnau ka D'llig Matiktik. Ka na sasoulat ti Mattheus, ti Johannes appa. Overgeset inde Formosaansche tale, voor de Inwoonders van Soulang, Mattau, Sinckan, Bacloan, Tavocan en Tevorang* [*The Holy Gospel according to Matthew and John or the True and Holy Gospel, According to the Books of Matthew and John. Translated into the Formosan language, for the inhabitants of Soulang, Mattau, Sinckan, Bacloan, Tavocan and Tevorang*]. Amsterdam: Michiel Hartogh.
- Gravius, Daniël (1662). *Patar ki Tna-'Msing-An ki Christang, Ka Tauki-papatar-en-ato tmæu'ug tou Sou ka Makka Si-Deia. Ofte 't Formulier des Christendoms, Met de Verk-laringen van dien, Inde Sideis-Formosaansche Tale* [*A Formulary of the Christian Faith, with Explanations in the Sideis-Formosan Language*]. Amsterdam: Michiel Hartogh.
- Harvey, Mark (1982). Subgroups in Austronesian. In Halim Amran, Lois Carrington and S.A. Wurm (Eds.), *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, (pp. 47–99). Pacific Linguistics C-75. Canberra: The National Australian University.
- Ho, Dah-an 何大安 (1983). Lun lukaiyu de qinshu guanxi 論魯凱語的親屬關係 [On the position of Rukai in the Formosan languages]. *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 [*Bulletin of the Institute of History and Philology, Academia Sinica*] 51(1), 121–168.
- Ho, Dah-an 何大安 (1998). Taiwan nandaoyu de yuyan guanxi 臺灣南島語的語言關係 [The genetic relationships among the Formosan languages]. *Hanxue yanjiu* 漢學研究 [*Chinese Studies*] 16(2), 141–171.
- Ho, Dah-an, and Hsiu-fang Yang 何大安、楊秀芳 (2000). Daolun: Nandaoyu yu taiwan nandaoyu 導論：南島語與臺灣南島語 [Introduction: Austronesian languages and Formosan languages]. In Lillian Meei-jin Huang 黃美金 (Ed.), *Taiwan nandao yuyan*

- 1–13 臺灣南島語言 1–13 [Formosan languages 1–13], (pp. 1–25). Taipei 臺北 [Taipei]: Yuanliu chubanshe 遠流出版社 [Yuan-Liou Publishing Co.].
- Huang, Lillian M. (1995). The syntactic structure of Wulai and Mayrinax Atayal: A comparison. *Bulletin of the National Taiwan Normal University* 40, 261–294.
- Huang, Wei-chen (a.k.a. Ng, I-sin) (2012). A study of verbal morphology in Puljetji Pawan. MA thesis. Hsinchu: National Tsing Hua University.
- Klaproth, Julius H. (1822). Sur la langue des indigènes de l'île de Formose [On the language of the Aborigines of Formosa]. *Journal Asiatique*, 193–202.
- Li, Li-ying Lilian (2018). A grammar of Isbukun Bunun. PhD dissertation. Hsinchu: National Tsing Hua University.
- Li, Paul Jen-kuei (1977). The internal relationships of Rukai. *Bulletin of the Institute of History and Philology, Academia Sinica* 48(1), 1–92.
- Li, Paul Jen-kuei (1982). Male and female forms of speech in the Atayalic group. *Bulletin of the Institute of History and Philology, Academia Sinica* 53(2), 265–304.
- Li, Paul Jen-kuei (1985). The position of Atayal in the Austronesian family. In Andrew Pawley and Lois Carrington (Eds.) *Austronesian Linguistics at the 15th Pacific Science Congress*, (pp. 257–280). Pacific Linguistics C-88. Canberra: Pacific Linguistics.
- Li, Paul Jen-kuei (1988). A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 59(2), 479–508.
- Li, Paul Jen-kuei (1990). Classification of Formosan languages: Lexical evidence. *Bulletin of the Institute of History and Philology* 61(4), 809–835.
- Li, Paul Jen-kuei (1998). The dialectal differences in Atayal in I-Lan. In Shuanfan Huang (Ed.), *Selected papers from the Second International Symposium on Languages in Taiwan*, (pp. 49–78). Taipei: The Crane Publishing Co., Ltd.
- Li, Paul Jen-kuei (2003). The internal relationships of six Western Plains languages. *Bulletin of the Department of Anthropology, National Taiwan University* 61, 39–51.
- Li, Paul Jen-kuei (2004). *Selected papers on Formosan languages, vols 1 et 2*. Language and Linguistics Monograph Series C3. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei (this handbook). Historiography of Formosan languages from 1895 until 1990. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Li, Paul Jen-kuei and Shigeru Tsuchida (2022). Subclassification of Amis dialects. *Journal of Taiwanese Languages and Literature* 17(1), 93–116.
- Liao, Hsiu-chuan (this handbook). Formosan languages and Proto-Austronesian phonology. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.
- Lin, Ting-yu (2022). Amis-Sakizaya phonology, reconstruction and subgrouping. MA thesis. Hsinchu: National Tsing Hua University.

- Marsh, Mikell Alan (1977). The Favorlang-Pazeh-Saisiat subgroup of Formosan languages. PhD Dissertation. Pullman, WA: Washington State University.
- Ogawa, Naoyoshi and Erin Asai 小川尚義、淺井惠倫 (1935). *Gengo ni yoru taiwan takasago zoku densetsu shū* 原語による臺灣高砂族伝説集 [*The myths and traditions of the Formosan native tribes (texts and notes)*]. Taihoku 臺北 [Taipei]: Taihoku Teikoku Daigaku 臺北帝國大學 [Taihoku Imperial University].
- Reid, Lawrence (1982). The demise of Proto-Philippines. In Halim Amran, Lois Carrington, and S.A. Wurm (Eds.), *Papers from the Third International Conference on Austronesian Linguistics*, vol. 2: *Tracking the travellers*, (pp. 201–216). Pacific Linguistics C-75. Canberra: The National Australian University.
- Ross, Malcolm (2009). Proto Austronesian verbal morphology: A reappraisal. In Alexander Adelaar and Andrew Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, (pp. 295–326). Pacific Linguistics 601. Canberra: The Australian National University.
- Ross, Malcolm (2012). In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13(6), 1253–1330.
- Ross, Malcolm (2015). Reconstructing Proto Austronesian verb classes. *Language and Linguistics* 13(3), 279–315.
- Sagart, Laurent (2004). The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2), 411–444.
- Sagart, Laurent (2012). In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic. *Language and Linguistics* 15(6), 859–882.
- Shibata, Kye (2020). A reconstruction of Proto-Bunun phonology and lexicon. MA thesis. Hsinchu: National Tsing Hua University.
- Starosta, Stanley (1995). A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng (Eds.), *Austronesian Studies Relating to Taiwan*, (pp. 683–726). Symposium Series of the Institute of History and Philology, Academia Sinica No. 3. Taipei: Academia Sinica. Subsequently published in Elizabeth Zeitoun (Ed.), (2009). *Formosan Linguistics: Stanley Starosta's contributions*, (pp. 737–769). Language and Linguistics Monograph Series C6-2. Taipei: Institute of Linguistics, Academia Sinica.
- Starosta, Stanley, Andrew Pawley and Lawrence 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. Canberra: Australian National University. Reprinted in Elizabeth Zeitoun (Ed.) (2009). *Formosan linguistics: Stanley Starosta's contributions*, (pp. 297–328). Language and Linguistics Monograph Series. Taipei: Institute of Linguistics, Academia Sinica.
- Teng, Stacy Fang-ching (2008). *A grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics 595. Canberra: The Australian National University.

- Teng, Stacy Fang-ching (2009). Case syncretism in Puyuma. *Language and Linguistics* 10(4), 819–844.
- Teng, Stacy Fang-ching (2011). Noun phrase conjunction in three Puyuma dialects. *Language and Linguistics* 12(1), 171–204.
- Teng, Stacy Fang-ching (2015). A comparative study of the personal pronominal systems in three Puyuma dialects. In Elizabeth Zeitoun, Stacy Fang-ching Teng and Joy Jing-lan Wu (Eds.), *New Advances in Formosan Linguistics*, (pp. 407–429). Asia-Pacific Linguistics 017 / Studies in Austronesian Linguistics 003. Canberra: Pacific Linguistics.
- Ting, Pang-hsin 丁邦新 (1978). Gubeinanyu de nice 古卑南語的擬測 [Reconstruction of proto-Puyuma phonology]. *Zhongyang yanjiuyuan lishi yuyan yanjiusuo jikan* 中央研究院歷史語言研究所集刊 [Bulletin of the Institute of History and Philology, Academia Sinica] 49(3), 321–392.
- Tu, Wen-chiu (1994). A synchronic classification of Rukai dialects in Taiwan: A quantitative study of mutual intelligibility. PhD dissertation. Illinois, IL: University of Illinois at Urbana-Champaign.
- Tsuchida, Shigeru (1976). *Reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series No. 5. Tokyo: Tokyo Gaikokugo Daigaku.
- Tsuchida, Shigeru (1980). Linguistic position of Skikun and Manawyan: Linguistic bases of subgrouping of Sqolyeq and Ts'ole dialects in Atayal. ms.
- Van der Vlis, Christianus Jacobus (1842). Formosaansche woorden-lijst, volgens een Utrechtsch handschrift: Voorafgegaan door eenige korte aanmerkingen betreffende de Formosaansche taal [Formosan wordlist, according to a Utrecht manuscript: Preceded by some short observations on the Formosan language]. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen; dl. 18*, [Proceedings of the Batavian Society for the Arts and Sciences, part 18], (pp. 437–488). Batavia: Bataviaasch Genootschap van Kunsten en Wetenschappen.
- Wang, Mei-hsia (2008). The reinvention of ethnicity and culture: A comparative study on the Atayal and the Truku in Taiwan. *Journal of Archaeology and Anthropology* 68, 1–44.
- Zeitoun, Elizabeth (1992). A syntactic and semantic study of Tsou focus system. MA thesis. Hsinchu: National Tsing Hua University.
- Zeitoun, Elizabeth and Stacy F. Teng (2016). The position of Kananavu and Saaroa within the Formosan languages revisited. *Oceanic Linguistics* 55(1), 163–198.

*Online Resources and Open Access Data*

Blust, Robert and Stephen Trussel (2020). *Austronesian comparative dictionary*. Retrieved from at [www.trussel2.com/ACD](http://www.trussel2.com/ACD).

Council of Indigenous Peoples 原住民族委員會 (2021). *Yuanzhuminzu yuyan xian-shang cidian 原住民族語言線上辭典* [Online dictionaries of Indigenous languages]. Retrieved from <https://e-dictionary.ilrdrf.org.tw/>.