Bu-hwan Vocabulary Recorded in 1874:
Comparison with Seediq Dialects[*]

Ochiai, Izumi

Kyoto University

The language called Bu-hwan was recorded in 1874 with 164 vocabulary items. It has a close resemblance to Seediq. Asai (1953) observed that Bu-hwan coincides with the Paran dialect of Seediq; however, he did not provide comparisons with Taroko Seediq to support his opinion. This paper presents the Bu-hwan list with Paran Seediq and Taroko Seediq. Reconstructed Proto-Seediq forms are also provided to show innovations shared by Bu-hwan and Paran Seediq. In addition, Asai's corrections are annotated and further comments on 15 Bu-hwan items are introduced from historical perspectives. Finally, diachronic sound changes from Bu-hwan to Paran Seediq are discussed.

1. Bu-hwan: Background
2. Paran Seediq: Background
3. Bu-hwan list with Paran Seediq, Taroko Seediq, and Proto-Seediq
4. Shared innovations of Bu-hwan and Paran Seediq
5. Corrections on Bu-hwan list by Asai (1953) with annotation
6. Additional notes on the Bu-hwan list
7. Sound changes from Bu-hwan to modern Paran Seediq
8. Conclusion

1. Bu-hwan: Background

Bullock (1874: 40–43) lists 179 words from six Formosan languages: 1. Tsui-hwan (水蕃), 2. Sek-hwan (熟蕃), 3. Bu-hwan (霧蕃), 4. Pepo-hwan (平埔蕃), 5. Favorlang, and 6. Pe-lam-hwan (卑南蕃).[1] With the exception of Favorlang, fieldwork on the five languages was conducted through the collaboration of Thomas Lowndes Bullock, an interpreter of Chinese and English at the British Consulate at Takow,[2] and Joseph

---

[1] The current language names are 1. Thao, 2. Pazeh, 3. Seediq, 4. Siraya, and 6. Puyuma. 5. Favorlang is also known as Babuza.

[2] This place corresponds to today's Gaoxiong (高雄).

Beal Steer, an ornithologist at Michigan University. Bullock states that they worked together on three languages, with Bullock working on one language by himself, and Steer working on one language by himself. However, there is no mention of who worked on which language. Favorlang items were not collected by Bullock but were copied from Campbell (1896).[3]

When they collected words, they first communicated with local interpreters who understood Mandarin, Amoy dialect (Southern Min), or English. Bullock says he sometimes showed them Chinese characters to make himself understood. The local interpreters then spoke with informants in the informant's language. Bullock thus admits that their list may be erroneous since they could not communicate directly with their informants. He was not sure if the interpreters had a correct understanding of their intention, or if the informants had a correct understanding of the interpreters' intention.

Among the 179 words they collected, 104 items are in the Swadesh list devised in the 1940s. Approximately 70 years before the Swadesh list, they succeeded in collecting more than 100 basic vocabulary items from five Formosan languages, which makes this article a valuable resource for Austronesian studies. In addition, this list was compiled in 1874, the period of the Ching dynasty and 21 years before Japan's rule in Taiwan, which began in 1895. Thus, the list is free of Japanese loanwords, which were introduced intensely into Formosan languages under Japan's rule.

As for Bu-hwan,[4] it is the oldest word list of this language as far as the author knows. Asai (1953) has inspected Bullock's vocabulary items and concluded that Bu-hwan corresponds to the Paran dialect of Seediq.

The purpose of this study is three-fold. First, the Bu-hwan list is supplied with Paran Seediq, Taroko Seediq, and Proto-Seediq (Section 3), so as to support the proximity of Bu-hwan to Paran Seediq by shared innovations (Section 4). Second, the revisions of Bu-hwan list proposed by Asai (1953) are reassessed (Section 5), and additional comments are provided on 15 items (Section 6). Third, sound changes from Bu-hwan to Paran Seediq are described (Section 7). Before presenting the Bu-hwan list, a brief description of Paran Seediq is introduced.

2. Paran Seediq: Background

Seediq belongs to the Atayalic subgroup of the Austronesian family. The Atayalic subgroup belongs to one of the first order branches of Proto-Austronesian (Blust 1999). The Atayalic subgroup includes Atayal and Seediq (Figure 1).[5] The Seediq population is approximately 20,000; however, Mandarin is the mother tongue of the majority, espe-

---

[3] This is the English translation of the original manuscript written in Dutch (Happart 1650). One reviewer adds that Campbell's translation contains some errors that need to be corrected.

[4] Bullock says Bu-hwan is also known as Che-hwan or Cheng-hwan, both of which correspond to 生蕃 (raw-savage), meaning a savage who is not civilized, and in this case, not sinicized.

[5] I referred to Ogawa and Asai (1935) for the classification of Seediq into the Paran and Taroko dialects. In this interpretation, Toda dialect is included in Taroko dialect.

cially the youth. Fluent speakers of Seediq are more or less all over 50 years old. Those who were educated in Japanese under Japanese rule, who are now over 80 years old, speak fluent Japanese. People of approximately 70 years old speak less fluent Japanese.

Proto-Austronesian

Atayalic   First order branches

Proto-Seediq   Proto-Atayal

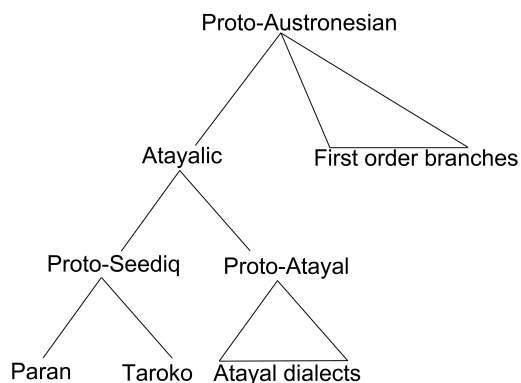Paran   Taroko   Atayal dialects

Fig.1 Seediq in Austronesian

Paran Seediq was spoken around an area called Paran in the indigenous place name, or Wushe 霧社 in the Mandarin Chinese place name. There were several villages scattered around Paran. Paran village, the largest of these, was the center of the Paran tribe (Figure 2). Erin Asai visited the Paran area in 1927 (and published his research in 1953). The Paran dialect was not spoken in the Paran village after 1930, when the Paran tribe rebelled against Japanese authority, known as the Wushe Uprising. Today, there are three Paran villages all in Ren'ai County, Nantou Prefecture. Most Paran people had to migrate to the western edge of Ren'ai County in the 1930s, where they established two villages, Gluban and Nakahara. In the previous Paran territory, the only remaining village of the Paran tribe is Tongan village.

As Mabuchi (1954) observes, Seediq has three tribe groups today: Paran, Toda, and Truku. These villages are in Ren'ai county, Nantou Prefecture. According to Mabuchi, the Seediq people originally lived somewhere around Paran but one group of them began to emigrate to the north approximately 350 years ago, diverging into the Toda and Truku groups. After three tribes had been established, a group from each tribe crossed the mountain ridges and emigrated to the east, to today's Xiulin and Wanrong counties, in Hualian Prefecture. These collateral villages are called Mukuy (originates in Paran), Tausa (originates in Toda), and Taroko (originates in Truku). There are three home groups and three collateral groups, all of which have their origin in Paran, the homeland of Seediq.[6]

---

[6] Another tribe name for Paran is Tugu-daya 'people of the uphill direction' (See Table 1 (159)). Truku and Taroko are treated as the same dialect. This study uses 'Taroko' as in Pecoraro (1977).
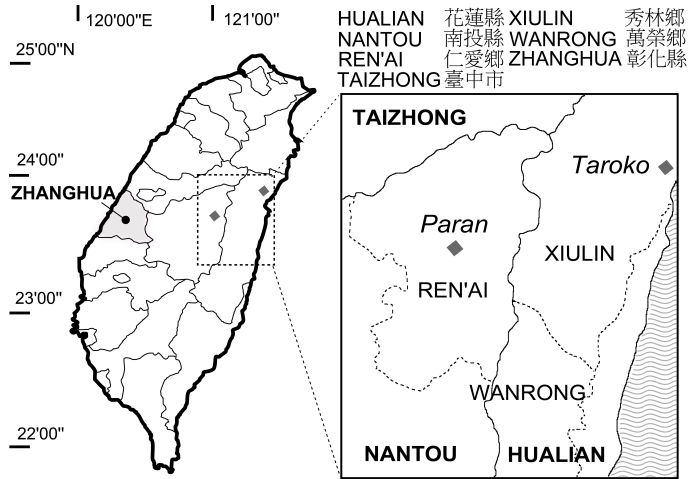
Fig.2   Seediq villages of Paran and Taroko[7]

Yang (1976) has shown that the basic syllable structure of Paran Seediq is CV or CVC. However, CVC is only allowed in final syllables. Yang (ibid.) also describes Paran Seediq as having five vowels (a, e, i, o, u) and 18 consonants (p, t, k, q, b, d, g, m, n, ŋ, s, x, h, ts, l, r, w, j). Holmer (1996) added ʔ to the consonant inventory, and Ochiai (2015a) has added a diphthong u͡i that occurs in a word final position. As Yang (ibid.) says, consonant and vowel length are not contrastive. The orthography used in this paper are the same as the letters shown in IPA above, except for /ts/, /j/, and /u͡i/ which are spelled *c*, *y*, and *uy* in this paper. Asai (1953) says the accent falls on the penultimate syllable. Vowels in pre-accented syllables undergo reduction, which are pronounced in a range from [ə] to [u] (written as *u* in this paper).

3. Bu-hwan list with Paran Seediq, Taroko Seediq, and Proto-Seediq

With regard to Bu-hwan orthography, Bullock (1874) explains "In the following list, 'a', 'e', 'i', 'o', 'eu' are to be sounded as in French, 'u' as 'oo' in 'boot', 'ŭ' as in 'but' 'ai' as in 'aisle' 'ou' as in 'out' 'ao' a mixture of the sounds 'a' and 'ou' given above, same as the Chinese sound spelt by Dr. Morrison. The consonants as in English. Initial 'ch' soft. Final 'ch' much as in the Scotch word 'loch'. Final 'h' a strong aspirate." Based on this description, I assume the following pronunciation for Bu-hwan: *a* [a], *e* [e], *i* [i], *o* [o], *eu* [ø], *u* [u], *ŭ* [ʌ], *ai* [aĭ], *ou* [aŭ], *ch* [tɕ/tʃ] (word initial), *ch* [x] (word final), and *h* [h]. There are no explanations for *ă*, *ŏ*, *ĕ*, or *ĭ*. These symbols are considered to resemble the sounds [a], [o], [e], and [i] in this paper. Unfortunately, Bu-hwan words do not distinguish /k/ and /q/. Both are written as *k*. In Table 1, the

---

[7] It shows the location of a representative Taroko village; however, villages of the Taroko tribe are scattered around Xiulin and Wanrong.

present author changed *ng* of Bu-hwan into ŋ for ease of presentation.[8]

The data of Paran Seediq (the third column from the left in Table 1) were collected by the author during fieldwork conducted intermittently from 2007 to 2015. Mandivaiian (2009), a comparative word list in the three dialects (Paran, Toda, Taroko) with approximately 1000 items, is used as a reference to verify the forms collected in my fieldwork.

The source for Taroko Seediq items is the online Taroko dictionary by the Council of Indigenous Peoples. The forms are cited as they appear in the dictionary with the exception that the *ng* in the dictionary is replaced by ŋ. An additional phoneme is /ɟ/, written as *j*. In this orthography, unaccented vowels are not written, leading to orthographic consonant clusters at the beginning of a word. In addition, the diphthongs ai̯, au̯, ui̯ are written as *ay*, *aw*, and *uy*. Tsukida (2009) explains that unaccented vowels in Taroko Seediq are pronounced as ə; however, sometimes they become voiceless depending on the environment (e.g., between voiceless consonants).

Where the online dictionary lacks the forms, Pecoraro's (1977) dictionary of Taroko Seediq is consulted, although it is problematic in that several phonemes are not distinguished and the high vowels /i/ and /u/ tend to be lowered to *e* and *o*.[9]

In Table 1, bound pronouns are indicated by =, and function as enclitics. A missing word is indicated by ---. Bullock's use of ─ for the possessive pronoun series from (7) to (11) seems to indicate the location of possessed nouns. He analyzed the possessive pronoun as following a possessed noun, which is also the case in Paran Seediq.

The right column contains Proto-Seediq (PS), Proto-Atayalic (PA) (or Proto-Atayal (PAta)), and Proto-Austronesian (PAN) forms. For Proto-Atayalic and Proto-Austronesian, as many samples were collected as the author could find in the previous literature. The Proto-Atayalic forms are from Li (1981), and Proto-Austronesian forms are from Blust and Trussel (2013) unless otherwise noted in footnotes. The Proto-Austronesian forms in parentheses are those with some degree of similarity to Proto-Seediq, but the author is still uncertain about the relation of Proto-Seediq and Proto-Austronesian.

Proto-Seediq is reconstructed as follows. When Paran Seediq and Taroko Seediq show lexical identity, a proto-form is reconstructed from the two forms. When Paran Seediq and Taorko Seediq do not show lexical identity, and Bu-hwan shows a resemblance to one of them, the author looked for Proto-Atayalic reconstructions or Atayal cognates. The reconstructed Proto-Seediq forms are still tentative and some require revision. Many forms require phonological explanation that leads to their reconstruction;

---

[8] Bullock's list contains hyphens, which are used in three ways. One is to indicate that a word is too long to fit in a column. Half of the word ends in the first line with a hyphen, and the rest goes into the second line. Hyphens also indicate a compound, as in *tanach-mahabai* 'yellow' (153). He seemed to know that the first part means 'red', which appears as *matanach* (151). A third use is to indicate the velar nasal. He writes *ng* for the velar nasal, and tries to show that *ng* is one segment by placing a hyphen after *ng*. I made the best guess to decide which of the three is intended in each case. For the first case, I deleted the hyphen. For the second case, I retained the hyphen. For the third case, I deleted the hyphen and changed *ng* into ŋ, except for (157).

[9] He spells both /k/ and /q/ as *k*, both /x/ and /h/ as *h*, and both /r/ and /l/ as *l*.

however, this will be dealt in another occasion due to limited space. A dagger indicates that Asai (1953) made a correction on this item (Section 5).

Table 1 Bu-hwan list with Seediq dialects and Proto-forms

Bu-hwan	Paran	Taroko	Proto-forms
1. yako	yaku	yaku	PS *yaku
'I'			PA *-akuʔ
			PAN *aku
2. issu	isu	isu	PS *isu
'thou'			PA *ʔisuʔ
			PAN *iSu
3. issu	heya	hiya	PS *hiya
'he'			PA *hiyaʔ
			PAN *si ia
4. yamo	yami	yami	PS *yami
'we'			PA *-ami
			PAN *kami[10]
5. ---	yamu	yamu	PS *yamu
'you'			PA *yamu
			PAN *i-kamu[11]
6. abarao	deheya	dhiya	PS *dəhiya
'they'†			PA ---
			PAN ---
7. —yako	=mu	=mu	PS * =mu
'my'			PA ---
			PAN ---
8. —issu	=su	=so[12]	PS * =su
'thy'			PA ---
			PAN ---
9. —nakaga	=na	=na	PS * =na
'his'†			PA ---
			PAN ---
10. —kakashaduk	=nami	=nami	PS * =nami
'our'†			PA ---
			PAN ---

---

[10] This is from Blust (1999: 87).
[11] This is from Blust (1999: 87).
[12] This is from Pecoraro (1977).

Bu-hwan	Paran	Taroko	Proto-forms
11. ―nataha 'their'†	=daha	=dha	PS * =daha[13] PA --- PAN ---
12. edŭk 'man'	seediq	seejiq	PS *s-ʔediq PA --- PAN ---
13. makaidil 'woman'	muqedin	kuyuh	PS *ma-qaidil PA ---[14] PAN ---
14. lakai makaidil[15] 'girl'	weewa	uwa	PS *w-ʔwa PA --- PAN ---
15. lakai risinao[16] 'boy'	riso	risaw	PS *risaw PA --- PAN ---
16. makaidil 'wife'	qedin	kuyuh	PS *(ma-)qaidil[17] PA --- PAN ---
17. lakai 'son'[18]	laqi	laqi	PS *laqi PA *ʔulaqiʔ (PAN *aNak)
18. lakai makaidil 'daughter'	laqi muqedin[19]	laqe kuyuh	PS *laqi maqaidil PA *ʔulaqiʔ (PAN *aNak)
19. nakial 'brother'†	suwai[20]	swai	PS *səwaði PA *suwaiʔ PAN *Suaji
20. swadzu makaidil 'sister'	suwai muqedin[21]	swai	PS *səwaði maqaidil PA *suwaiʔ PAN *Suaji

---

[13] It could be related to PAN *duSa 'two' (171).

[14] Atayal has kneril 'woman' (Egerod 1980: 291), which seems to correspond to the Bu-hwan and Paran forms.

[15] This is a compound of laqi 'offspring' and muqedin 'female adult.' Laqi muqedin 'young female, girl' is used in Paran Seediq as well as weewa 'girl.'

[16] This is a compound of laqi 'offspring' and ruseno 'male adult.' Laqi ruseno 'young male, boy' is used in Paran Seediq as well as riso 'boy.'

[17] Proto-Seediq *(ma-)qaidil seems to mean 'woman, wife.'

[18] A better gloss would be 'offspring.'

[19] A compound of laqi 'offspring' and muqedin 'female person.'

[20] This form means a younger sibling (both male and female). The word for elder sibling is qubusuran in Paran Seediq (Proto-Atayalic is *qəbəsuran (Li 1981)).

[21] A compound of suwai 'younger sibling' and muqedin 'female person.'

Bu-hwan	Paran	Taroko	Proto-forms
21. mama 'uncle'	tama	---	PS *mama (?) PA ---[22] Pᴀɴ ---
22. bubu 'mother'	bubu	bubu	PS *bubu PA --- Pᴀɴ ---
23. taina 'father'†	tama	tama	PS *tama PA --- Pᴀɴ *ta-ama
24. tunuch 'head'	tunux	tunux	PS *tunux PA *tunux Pᴀɴ ---
25. dourŭk 'eyes'	doriq	dowriq	PS *dawriq PA *dawriq Pᴀɴ ---
26. birŭt 'ears'	birac	birat	PS *birat PA ---[23] Pᴀɴ ---
27. mohiŋ 'nose'	muhiŋ	muxeŋ	PS *muhiŋ PA --- (Pᴀɴ *mujiŋ)[24]
28. koak 'mouth'	quwaq	quwaq	PS *quwaq PA --- Pᴀɴ ---
29. rupun 'teeth'	rupun	gupuŋ	PS *gupun PA *gipun Pᴀɴ *nipen[25]
30. padahuŋ 'lips'	pudahuŋ	pdaxoŋ[26]	PS *pədahuŋ PA *padahum Pᴀɴ ---
31. hema 'tongue'	hema	hma	PS *həma PA *həmaʔ Pᴀɴ Sema

---

[22] Atayal has the cognate *mama* 'uncle' (Egerod 1980: 361). Seediq seems to have lost this form.

[23] The Proto-Atayalic form is *caŋiraʔ* (Li 1981). This form corresponds to the Proto-Austronesian *Caŋila* (Blust and Trussel 2013).

[24] This is from Blust (1999: 83).

[25] This is from Blust (1999: 83).

[26] This is from Pecoraro (1977). Taroko Seediq seems to have lost this form. Mandivaiian (2009) has *qwaq* 'lips' for Taroko Seediq, which is evidently the same form as *quwaq* 'mouth' (28).

Bu-hwan	Paran	Taroko	Proto-forms
32. bukului 'chin'	bukeluy	bkluy	PS *bəkəluy PA --- PAN ---
33. muduski 'beard'†	ŋudus	ŋudus	PS *ŋudus PA *ŋudus PAN ---
34. ŭduthiŋ 'throat'†	gelu	glu	PS *gəlu PA --- PAN ---
35. --- 'neck'	nduyuŋ	waru[27]	PS *gəduðuŋ PA *gadiyuŋ PAN ---
36. ahiŋ 'shoulder'	ahiŋ	hiraŋ	PS *hiraŋ PA *qəhiraŋ PAN ---
37. abatha 'arm'†	baga	baga	PS *abaga PA *qabaga[28] PAN *qabaRa
38. puŋuh[29] 'elbow'	hiqun baga[30]	hiqur	PS *hiqur PA *hiquʔ PAN *siku
39. abatha 'hand'†	baga	baga	PS *abaga PA *qabaga PAN *qabaRa
40. tuludiŋ 'finger'†	kehiŋ[31]	tduliŋ	PS *tədiluŋ PA *tiduliŋ PAN ---
41. --- 'thumb'	bubaki[32]	studuliŋ[33]	PS --- PA --- PAN ---

---

[27] Li (1981: 228) has *gduyuŋ* and *duyuŋ* for Taroko Seediq.

[28] This form with the meaning of 'hand, arm' is reconstructed in Ochiai (2015b). The Atayal form is *qbaʔ* 'forearm, hand' (Egerod 1980). This form is considered to be the cognate of the Proto-Austronesian *qabaRa* 'shoulder' (39). Blust and Trussel (2013) also mention the resemblance between the proto-form *qabaRa* 'shoulder' and Taroko Seediq *baga* 'hand, arm.'

[29] *Puŋu* in Paran Seediq only means 'knee' (45).

[30] *Baga* means 'hand, arm' (37, 39).

[31] The Paran form is *təludiŋ*, and the Taroko form is *təduluŋ* in Ogawa and Asai (1935). Li (1981: 293) also has *tludiŋ* for Paran Seediq of Toŋan village. The consonants *d* and *l* underwent metathesis in Paran Seediq and Bu-hwan. Taroko also has a variant as *tludiŋ*.

[32] This is the reduplicated form of *baki* 'grandfather.'

[33] Pecoraro (1977) has *tama*, the word for 'father' (23).

Bu-hwan	Paran	Taroko	Proto-forms
42. kukuh 'finger-nail'	kukuh	kokox	PS *kukuh PA --- (PAN kuSkuS)
43. turŭŋ 'breast'	teeraŋ	brah[34]	PS *tʔeraŋ PA --- PAN *tageRaŋ
44. papŭk 'leg'	papak	qaqay	PS *qaqay PA ---[35] PAN *qaqay
45. puŋuh 'knee'	puŋu	puŋu	PS *puŋu PA --- PAN ---
46. kapal papŭk 'foot'	qapan papaq[36]	qaqay	PS *qapal qaqay(?) PA --- PAN ---
47. jitio papŭk 'toe'†	kehiŋ	tdoliŋ	PS *tədiluŋ[37] PA *tiduliŋ PAN ---
48. --- 'toe-nail'	kukuh	kukuh	PS *kukuh[38] PA *tiduliŋ PAN ---
49. dara 'blood'	dara	dara	PS *dara PA *dagaʔ PAN *daRaq
50. tamabahak 'heart'†	toma bahaq	tama baraq	PS *tama baraq PA *baraq 'lung' PAN *baRaq 'lung'
51. churuk 'fish'†	qucurux	qsurux	PS *qəcurux PA *qucurux PAN ---
52. tamat 'deer'†	ruqenux	rqnux	PS* rəqənux PA *raqə-nux[39] (PAN *qaNuaŋ[40])

---

[34] Paran Seediq has a cognate, *berah* 'front.'
[35] Atayal has *kakai* 'foot, leg, hind leg' (Egerod 1980: 252).
[36] I have not so far found a cognate of Proto-Seediq *qapal* in Atayal or Taroko Seediq.
[37] It is the same form as 'finger' (40).
[38] It is the same form as 'fingernail' (42).
[39] Li (1981) recognizes that this proto-form has an additional suffix *-nux*. Another example is Proto-Atayalic *batu-nux* 'stone' (93).
[40] The form means 'large ruminant species: carabao, water buffalo (?)' (Blust and Trussel 2013).

Bu-hwan	Paran	Taroko	Proto-forms
53. dapa 'cow'	dapa	kaciŋ	PS *dapa PA *dapaʔ Pᴀɴ ---
54. rouduch 'hen'	rodux	rudux	PS *rawdux PA --- Pᴀɴ ---
55. hulin 'dog'	huliŋ	huliŋ	PS *huliŋ PA --- Pᴀɴ *asu
56. niao 'cat'	ŋiyo	ŋiyaw	PS *ŋiyaw PA *ŋiyaw Pᴀɴ ---
57. baluŋ 'egg'	baluŋ	baluŋ	PS *baluŋ PA --- Pᴀɴ ---
58. bahani 'bird'	qubeheni[41]	qbhni	PS *qəbəhəni PA *kabah-niq[42] Pᴀɴ ---
59. ubal 'feather'[43]	uban	ubal	PS *ubal PA --- Pᴀɴ ---
60. babui 'pig'	babuy	babuy	PS *babuy PA *babuy Pᴀɴ *babuy
61. kuzu 'snake'	quyu	quyu	PS *quðu PA --- Pᴀɴ ---
62. kăhoni 'tree'	quhuni	qhuni	PS *qəhuni PA *kahu-niq Pᴀɴ *kaSiw
63. hazi kăhoni 'leaf'†	waso	rnabao	PS *abaw PA *abag[44] Pᴀɴ ---

---

[41] There is a variant as *beheni*, which is the form that Bullock (1874) recorded.

[42] Li (1981) recognizes that this proto-form has an additional suffix *-niq*. Other examples are Proto-Atayalic *kabah-niq* 'bird' (58), *kahu-niq* 'tree, wood' (62), *hapu-niq* 'fire' (88). Proto-Seediq *bəhəniq* 'bow' (100) also includes this suffix.

[43] It means body hair as well as feather.

[44] This Proto-Atayalic form is reconstructed by the present author from the forms for 'leaf' in Atayal and Seediq dialects reported in Li (1981: 287).

Bu-hwan	Paran	Taroko	Proto-forms
64. raparap 'root'†	gamin	gamil[45]	PS *gamil PA *gamil PAN *RamiS
65. --- 'flower'	pehepah	phpah	PS *pahəpah PA *pahəpah PAN ---
66. --- 'fruit'	hei[46]	hiyi	PS *hiði PA *hiiʔ (PAN *Sesi/isi[47])
67. shuduh 'grass'	sudu	spleq[48]	PS *sudu/səpəriq PA ---[49] PAN ---
68. bĕlĭbeul 'banana'	bulebun	blbul	PS *buləbul PA --- PAN *beNbeN
69. buŋa 'sweet potato'	buŋa	buŋa	PS *buŋa[50] PA --- PAN ---
70. barass 'rice'†	beras	buwax	PS *buwax/beras PA *buwax PAN *beRas
71. timu 'salt'	timu	cimu	PS *timu PA *timuʔ PAN *timu
72. --- 'sugar'	sibus[51]	sibus	PS *sibus PA --- PAN *Cebus
73. kasia 'water'	qusiya	qsiya	PS *qəsiya PA --- PAN ---

---

[45] This is from Pecoraro (1977).

[46] This means 'body, fruit.'

[47] This means 'meat, flesh' (Blust and Trussel 2013).

[48] This is from Pecoraro (1977).

[49] Atayal has *kʔman* 'grass' (Egerod 1980: 312). It is probably related to *mukeeman* 'dark as grass grows so high' and *keeman* 'night' in Paran Seediq.

[50] Blust (1995: 470–471) explains that the form for 'sweet potato' such as *buŋa* suggests introduction of this plant from the Northern Philippines.

[51] This means 'sugarcane.' Sugar is a recently introduced product. It is expressed by a loanword from Japanese *sato* (砂糖).

Bu-hwan	Paran	Taroko	Proto-forms
74. dzadzuŋ 'river'	yayuŋ	yayuŋ	PS *ðaðuŋ PA --- PAN ---
75. chiluŋ timu[52] 'sea'	guciluŋ	gsiluŋ	PS *gəciluŋ PA *-ciluŋ PAN ---
76. chiluŋ 'lake'	guciluŋ	gsiluŋ	PS *gəciluŋ PA *-ciluŋ PAN ---
77. kuzuch 'rain'	quyux	quyux	PS *puðux PA --- PAN *quzaN
78. ruluŋ 'clouds'	pulabu	luluŋ	PS *ruluŋ PA *ruluŋ (PAN *lemlem[53])
79. mabarua 'thunder'	muburuwa	bruwa	PS *ma-buruwa PA --- PAN ---
80. sassaina 'lightning'	tugilaq	liwaq	PS --- PA --- PAN ---
81. hidao 'sun'	hido	hidaw	PS *hidaw[54] PA ---[55] PAN ---
82. idaas 'moon'	idas	idas	PS *idas PA ---[56] PAN *qiNas
83. --- 'stars'	puŋerah	pŋrah	PS *pəŋərah PA --- PAN ---
84. kushun 'day'†	diyan	jiyan	PS *diyan PA *diyax PAN ---
85. babien 'night'	bubiyan	gbiyan	PS *gəbiyan PA *gabiyan PAN *Rabi-an

[52] This form is a compound of *gucilung* 'pond, lake' and *timu* 'salt' (71).
[53] This is from Blust (1999: 86).
[54] This form could be related to the Proto-Austronesian *siNaR* 'light' reconstructed in Tsuchida (1976: 187).
[55] The Proto-Atayalic form is *wagiʔ* (Li 1981), which is different from the Proto-Seediq form.
[56] The Proto-Atayalic form is *bural* (Li 1981), which is different from the Proto-Seediq form.

Bu-hwan	Paran	Taroko	Proto-forms
86. lemedeuch 'light'	lumedax[57]	rmdax	PS *ləmedax PA --- PAN ---
87. măkuun 'dark'	mukuunŋ	kuuŋ	PS *ma-kuuŋ PA *mana-kuʔum PAN ---
88. hăpunek 'fire'	puniq	puniq	PS *hapuniq PA *hapu-niq PAN *Sapuy
89. kareŋeul 'smoke'	qureŋun	qrŋul	PS *qəreŋul PA --- PAN ---
90. makaluch 'ashes'[58]	qubulic	qbulit	PS *qəbulit PA *qabu-lit[59] PAN *qabu
91. haruŋ[60] 'wood'†	haruŋ	haruŋ	PS *haruŋ PA *haruŋ PAN *saleŋ
92. dagizak paru[61] 'mountain'	dugiyaq	dgiyaq	PS *dəgiðaq PA --- PAN ---
93. batunuch 'stone'	butunux	btunux	PS *bətunux PA *batu-nux PAN *batu
94. banakail 'sand'	bunaquy	bnaqig	PS *bənaqig PA *bunaqig (PAN *qenay)
95. pilat 'silver'	pila[62]	pila	PS *pila PA *pilaʔ PAN ---
96. hilui 'iron'	hiluy	xiluy	PS *hiluy PA --- PAN ---

[57] This form *l<um>edax* is a verb meaning 'to shine.' It is infixed by *<um>*, which is the actor voice present morpheme.

[58] The identical form is in 'black' (156). *Muqalux* in Paran Seediq means 'black' not 'ash.' *Qubulic* in Paran Seediq means 'ash, dust' (104).

[59] Li (1981) recognizes that this proto-form has an additional suffix *-lit.*

[60] The gloss should be 'pine tree,' not 'wood.'

[61] *Paru* means 'big' (121).

[62] This form in Paran Seediq and Taroko Seediq means 'money.'

Bu-hwan	Paran	Taroko	Proto-forms
97. sinmadat 'knife'	hulumadac	hrŋudux	PS --- PA --- PAN ---
98. simbaraŋan 'spear'	puyan[63]	sblaŋan[64]	PS *səbraŋan PA *sbraŋan PAN ---
99. sinmadat kanadish[65] 'sword'	hulumadac	puciŋ	PS *putiŋ (?) PA *putiŋ PAN ---
100. bahĕnek 'bow'	beheniq	bhniq	PS *bəhəniq PA *bahug PAN *busuR
101. budi 'arrow'	budi	budi[66]	PS *budi PA --- PAN ---
102. sinuzuk 'cord'	sunuyuk	snuyuk[67]	PS *sənuðuk PA --- PAN ---
103. haluŋ 'gun'	haluŋ	puniq[68]	PS *haluŋ (?) PA ---[69] PAN ---
104. kabulit[70] 'powder'	qubulic	qbulit	PS *qəbulit PA *qabu-lit PAN *qabu
105. bali 'shot'	bubali[71]	bali	PS *bali PA --- PAN ---
106. sapah 'house'	sapah	sapah	PS *sapah PA --- PAN ---

---

[63] Mandivaiian (2009) has *snbraŋan* for Paran Seediq.

[64] This form is from Pecoraro (1977).

[65] The second word *kanadish* corresponds to *qunedis* 'long' in Paran Seediq (119).

[66] This is from Pecoraro (1977).

[67] The Taroko Seediq form is used as a verb 'to twist fibers into ropes.'

[68] This is the same as 'fire' (88). It seems that in Taroko Seediq, the word for 'fire' has extended its meaning to signify 'gun.'

[69] Atayal has *patus* 'rifle, gun; to shoot' (Egerod 1980: 457).

[70] The more accurate gloss is 'dust.' 'Ash' and 'dust' are expressed by the same form in Seediq. *Nemu* is the form for 'powder' in Paran Seediq.

[71] This means 'bullet.' It shows the reduplication of Proto-Seediq *bali*.

Bu-hwan	Paran	Taroko	Proto-forms
107. denamuch 'roof'	dunamux[72]	dnamux	PS *dənamux PAta *rinamug[73] P_{AN} ---
108. răheŋun 'door'	ruheŋun	l'xŋun[74]	PS *rəhəŋun PA --- P_{AN} ---
109. ashu 'canoe'	asu	asu	PS *asu PA *qasu P_{AN} ---
110. tataku 'paddle'[75]	taku	taku	PS *taku PA *takuʔ P_{AN} ---
111. malup 'good'	malu	malu	PS *malu PA --- P_{AN} ---
112. nakach 'bad'	naqah	naqih	PS *naqih PA --- P_{AN} ---
113. sasibuss 'sweet'	musibus[76]	ssibus	PS *sasibus PA --- P_{AN} ---
114. tabashi 'sour'	tubasi[77]	basi[78]	PS * tabasi PA --- P_{AN} ---
115. maŋihul 'bitter'	muŋihun	mnihur	PS *maŋihur PA *ma-ŋihur P_{AN} ---
116. sĕsĕchaun 'ugly'	naqah qutaan	mkamul	PS --- PA --- P_{AN} *ma-Seyaq[79]

---

[72] This is derived from a verb *d<um>amux* 'to thatch.' The infix *<un>* in *d<un>amux* is a past tense marker of an undergoer voice.

[73] This Proto-Atayal form is reconstructed by the present author based on the Atayal forms for 'roof' provided in Li (1981: 291).

[74] This is from Pecoraro (1977). The apostrophe indicates a small pause between segments.

[75] 'Ladle, spoon' would be a better gloss than 'paddle.' The Seediq people rarely use canoes and paddles nowadays.

[76] This form is derived from *sibus* 'sugarcane' (72).

[77] *Mutubasi* is also accepted.

[78] This is *mbasi* in Pecoraro (1977).

[79] The Bu-hwan form reflects this proto-form which means 'shy, ashamed' (Blust and Trussel 2013).

Bu-hwan	Paran	Taroko	Proto-forms
117. matatiluch tidao 'hot'	mutilux	mtilux[80]	PS *ma-tilux PA --- Pᴀɴ ---
118. mabatunuch 'pretty'	mubutunux	malu taan[81]	PS *ma-butunux PA *mag-baytunux Pᴀɴ ---
119. kanadish 'long'	qunedis	qjis[82]	PS *qənədis PA --- Pᴀɴ ---
120. dĕhakuh 'short'	lulebu	lbu	PS *ləlebu (?) PA *la-ləbu Pᴀɴ ---
121. paru 'big'	paru	paru	PS *paru PA --- Pᴀɴ ---
122. tikŏh 'little'	tikuh[83]	cikuh	PS *tikuh PA --- Pᴀɴ ---
123. mǎriemuch 'round'†	mutumun	mtumun	PS *ma-tumun PA *limuk 'pot'[84] Pᴀɴ ---
124. mǎsassaput balai[85] 'square'	sepac papak	---	PS --- PA --- Pᴀɴ ---
125. matatiluch[86] 'warm'	muxun	muxul	PS *muxul PA --- Pᴀɴ ---
126. mǎsekuĭch 'cold'	musekuy	mskuy	PS *ma-səkuy PA --- Pᴀɴ ---

---

[80] This is from Pecoraro (1977).

[81] This is from Pecoraro (1977). Paran Seediq also uses *malu qutaan*, meaning 'good-looking.' *Malu* is 'good' (111), and *qutaan* is 'to see (undergoer voice, location subject, present tense).'

[82] This is *kndus* in Pecoraro (1977).

[83] This means 'small (in quantity).'

[84] This corresponds to Bu-hwan (See the explanation in Section 5).

[85] This is a compund of 'four' and 'true' (See 'four' (173) and 'yes' (128)). The form for 'four' is prefixed by *ma-* and a reduplicated segment (it was probably be *ma-səsepat*).

[86] This is identical to *mutilux* 'hot' (117).

Bu-hwan	Paran	Taroko	Proto-forms
127. ukach 'no'	uka[87]	oŋat	PS *uka/uŋat PA *ukaʔ/uŋat PAN *uka
128. balaiwa 'yes'†	bale[88]	balay	PS *balay PA --- PAN ---
129. mătugessa[89] 'run (verb)'	tumalaŋ	tmalaŋ	PS *təmalaŋ PA *t-um-alaŋ PAN *taNam
130. makan tamako 'smoke (verb)'	mekan tumaku[90]	mqan tbako[91]	PS --- PA --- PAN ---
131. manakamakan[92] 'eat (verb)'†	mekan	mkan	PS *məkan PA *kan PAN *kaen
132. nimah 'drink (verb)'†	mimah	mimah	PS *mimah PA *mimah PAN *mimah (?)[93]
133. makakaisa walk (verb)	mukukesa	ksa	PS *ma-kakaisa PA --- (PAN *s<um>akay)
134. taiăkarăk lie down (verb)	turuqeraq	mksaraw	PS *ma-kəsaraw PA *ma-si-ka-kig' PAN ---
135. matakai sleep (verb)	mutaqi[94]	taqi	PS *ma-taqi PA --- PAN ---

---

[87] This means 'there is none.'

[88] This means 'true.'

[89] The corresponding form in Paran Seediq would be *mutugesa*. This does not mean 'to run' but 'to teach (future tense).' I have no idea so far why Bullock has collected 'to teach' when he asked for 'to run'. *Tumalaŋ* 'to run, try' in Paran Seediq is the reflex of the Proto-Austronesian form meaning 'to taste, try.'

[90] This literally means 'to eat tobacco' (*mekan* 'eat' (131)).

[91] This is from Pecoraro (1977).

[92] One reviewer suggested that this could be a typo of the intended form *ma-maka-makan* with the reduplication of the two syllables (*maka-*) of the stem *makan*. It is possible that Bu-hwan had such reduplication. However, Paran Seediq verbs do not reduplicate two syllables (a few nouns have reduplication of two syllables). Some verbs reduplicate the first consonant of a root (*mu-du-dayo* 'to help each other' (root is *dayo* 'help')). The root in this case is *ʔekan*. The reduplicated form is *meʔekan* (< probably from /m-ʔ-ʔekan/) 'to eat (future tense).'

[93] This is from Blust (1999: 83).

[94] *Taqi* is also acceptable.

Bu-hwan	Paran	Taroko	Proto-forms
136. mahŏkal 'die'	muhuqin	mhuqil	PS *ma-huqil PA *mana-hawqil PAN ---
137. musha 'go'	musa	musa	PS *musa PA *m-a-usa PAN *uSa
138. maidzach 'come'	meyah	miyah	PS *maiðah PA --- PAN ---
139. --- 'buy'	mari[95]	barig[96]	PS *barig[97] PA *ma-barig' PAN *baliw
140. --- 'sell'	gunbari	gmbarig	PS *gəmbarig PA --- PAN *baliw
141. lĕminish cry (weep)†	lumiŋis	lmiŋis	PS *ləmiŋis PA *liŋis PAN *Caŋis
142. mahulish 'laugh'	muhulis	mhulis	PS *ma-hulis PA --- PAN ---
143. măhoyesh 'sing'	muuyas	meuyes	PS *mʔuyas PA --- PAN ---
144. maraŋao[98] 'talk'	rumeŋo	rmŋaw	PS *rəməŋaw PA --- PAN ---
145. kŏmŏpach 'work'	kumeepah	qeepah[99]	PS *kəmʔepah PA --- PAN ---
146. papurai 'roast'†	hunpure	hapuy	PS *hapuy[100] PA *hapuy PAN *Sapuy

---

[95] *Mari* is the citation form (actor voice present) of the root is *bari(g)*.
[96] This is *mali* in Pecoraro (1977).
[97] 'Buy' and 'sell' (139–140) share the root *barig*.
[98] This corresponds to *mureŋo* 'talk (future tense)' in Paran Seediq.
[99] This is *kmpax* in Pecoraro (1977).
[100] This is the same as 'fire' (88).

Bu-hwan	Paran	Taroko	Proto-forms
147. hamaŋut 'boil'	humaŋuc	hmaŋut	PS *humaŋut PA --- PAN ---
148. --- 'fish'[102]	pulaliŋ[101]	pŋahi	PS --- PA --- PAN ---
149. --- 'hung'	qumeya	luwil	PS *qəmeya PA *pa-qeyaʔ PAN ---
150. --- 'fight'	tugudiyan	tmdiyal[103]	PS *təgədiyal (?) PA --- PAN ---
151. matanach 'red'	mutanah	embanah	PS *ma-tanah PA *ma-tanah[104] PAN *ma-tanah (?)[105]
152. bahagai 'white'	behege	bhgay	PS *bəhəgay PA --- PAN ---
153. tanach-mahabai 'yellow'	muguciyas	mxeboŋ[106]	PS *ma-hibuŋ[107] PA --- PAN ---
154. masama 'blue'	mubulawa	mqliyaŋ	PS --- PA --- PAN ---
155. --- 'green'	mugusama	mgpajiq	PS --- PA --- PAN ---
156. makaluch 'black'	muqalux	mqalux	PS *ma-qalux PA *ma-qalux PAN ---
157. tungarĕt 'north'†	narac 'right'	narat 'right'	PS *taga-narat PA --- PAN *wanaN

[101] This is formed by *pu-* 'causative' and *laliŋ* 'rod.'
[102] This is meant to be a verb 'to fish.'
[103] This is from Pecoraro (1977).
[104] This is reconstructed by the present author based on the forms in Atayal and Seediq dialects provided in Li (1981: 290).
[105] This is from Blust (1999: 86).
[106] This is from Pecoraro (1977).
[107] This is reconstructed based on an Atayal cognate. Atayal has *mhibuŋ* 'yellow' (Egerod 1980: 370).

Bu-hwan	Paran	Taroko	Proto-forms
158. tagaēril 'south'†	irin 'left'	iril 'left'	PS *taga-iril PA *ʔiril Pan *wiRi
159. daia 'east'	daya 'uphill'	daya 'uphill'	PS *daya PA *daya Pan *dayaʔ[108]
160. hunat 'west'	hunac 'downhill'	hunat 'deep'	PS *hunat/rahut[109] PA --- Pan *lahud[110]
161. --- 'part'	icin	isil	PS *icil PA --- Pan ---
162. egu 'many'	egu	lala	PS --- PA --- Pan ---
163. ramut 'all'	kana	kana	PS *kana PA --- Pan ---
164. bilŭk 'few'	biciq[111]	bilaq	PS *bilaq PA --- Pan ---
165. kanuan 'when'	kunuwan	knuwan	PS *kanu(w)-an PA *kanuwan Pan ---
166. kanahainu 'how many'	kunuhenu[112]	khnu	PS *kanuhainu PA --- Pan ---
167. mua 'where'†	inu	inu	PS *inu PA *ʔinuʔ Pan *inu
168. ima 'who'	ima	ima	PS *ima PA *ʔimaʔ Pan *ima/si-ima

---

[108] This means 'towards the interior' (Adelaar 1997: 53).

[109] Both forms mean 'downhill.' *Rahut* is not used anymore, but it is still seen in a derived form *tugu-rahut* 'towards the lower side of a slope.'

[110] This means 'towards the sea' (Adelaar 1997: 53).

[111] This means either 'small in size' or 'small in amount.' The second meaning is usually expressed by *tikuh* (122).

[112] Another form for 'how many' in Paran Seediq is *piya.*

Bu-hwan	Paran	Taroko	Proto-forms
169. chĕchekach 'half'	cuceka[113]	kska	PS *cəceka PA *cəkaʔ PAN ---
170. kial 'one'	kiŋan	kingal	PS *kiŋal PA --- PAN ---
171. daha 'two'	daha	dha	PS *daha PA *duSaʔ PAN *duSa
172. tĕru 'three'	teru	tru	PS *təru PA *təruʔ PAN *telu
173. sŭpŭt 'four'	sepac	spat	PS *səpat PA *səpat PAN *Sepat
174. rima 'five'	rima	rima	PS *rima PA *rimaʔ PAN *lima
175. mataru 'six'	mumuteru[114]	mataru	PS *ma-təru PA *ma-təruʔ PAN ---
176. pitu 'seven'	mpitu	empitu	PS *(ma-)pitu PA *ma-pituʔ PAN *pitu
177. mŭssupat 'eight'	mumusepac	maspat	PS *ma-sepat PA *ma-səpat PAN ---
178. --- 'nine'	muŋari	maŋari	PS *maŋari PA --- PAN ---
179. nahal 'ten'†	maxan	maxal	PS *maxal PA --- PAN *masehaN[115]

---

[113] This is a reduplicated form of *ceka* 'broken into two (?).'

[114] This form is derived form *teru* 'three.' Paran Seediq has doubled the prefix *ma-* (Pre-Paran Seediq *mamateru* > Paran Seediq *mumuteru*). The same pattern is observed in *mumusepac* 'eight,' which is derived from *sepat* 'four.'

[115] This is an approximation for Proto-Austronesian 'ten' reconstructed by Sagart (2004: 424). Tsuchida (1976: 182) reconstructed *masʔaL* 'ten' for Proto-South-Formosan.

4. Shared innovations of Bu-hwan and Paran Seediq

Asai (1953) states that the "Bu-hwan (霧番?)[116] vocabulary collected by J. B. Steere and Bullock in China Review III 1874–1875 coincides with the Paran-dialect." However, he does not mention how he reached this conclusion. This section tries to follow Asai's line of thought to explain that Bu-hwan is more closely related to Paran Seediq than Taroko Seediq with the aid of reconstructed Proto-Seediq forms.

Asai's conclusion seems to partly depend on the correspondence of the character '霧' in 霧社 (Wushe), a Chinese place name for Paran village, as well as the geographical information that Bullock (Ibid. p.39) described as "near the central range due east from Chang-hwa (彰化)" (Figure 1).

His conclusion must have drawn from the similarities in vobabulary between Bu-hwan and Paran Seediq. In fact, he conducted research on both Paran and Taroko dialect (Ogawa and Asai 1935); however there has not been an attempt to compare Bu-hwan to Paran Seediq and Taroko Seediq. This paper presents shared innovations between Bu-hwan and Paran Seediq and innovations that occur only in Taroko Seediq. This observation supports the grouping of Bu-hwan and Paran Seediq on the one side and Taroko Seediq on the other.

Table 2 Shared innovations of Bu-hwan and Paran Seediq

	Bu-hwan	Paran	Taroko	Proto-Seediq
29	**r**upun 'teeth'	**r**upun	gupuŋ	*gupun
50	tamaba**h**ak	toma ba**h**aq	tama baraq	*tama baraq
85	**b**abien 'night'	**b**ubiyan	gbiyan	*gəbiyan
146	papu**r**ai 'roast'	hunpu**r**e	hapuy	*hapuy
36	**ahiŋ** 'shoulder'	**ahiŋ**	hiraŋ	*hiraŋ
44	**papŭk** 'leg'	**papak**	qaqay	*qaqay
134	taiăkarăk	turuqeraq	mksaraw	*ma-kəsaraw

Segment changes are seen in (29), (50), and (85) for both Bu-hwan and Paran Seediq. In (29), the segment *g* in Proto-Seediq *gupun 'teeth' changes to *r* in Bu-hwan and Paran Seediq, whereas Taroko Seediq retains the *g*. In (50), the segment *r* in Proto-Seediq *baraq 'lung' changes to *h* in Bu-hwan and Paran Seediq. In (85), the segment *g* in Proto-Seediq *gəbiyan changes to *b*, probably due to assimilation to the following consonant, in Bu-hwan and Paran Seediq.

In (36), segments are inserted in Bu-hwan and Paran Seediq. Proto-Seediq *hapuy had extra segments *ra* inserted between *hapu* and *y*, i.e., *hapu-ra-y*,[117] in the early

---

[116] These are transcribed as 霧 *bū* and 番 *hoan* in Amoy vernacular (Cambell 1913).

[117] Li (1985: 258–259) discusses sporadic affixation in Atayalic languages. One of these affixes is the infix <*ra*>.

stages of Bu-hwan and Paran Seediq. The actor voice present form is *h\<um>apuray*. In Paran Seediq, *h\<um>apúray* became *h\<um>púray* (the accent is indicated), then *ay* underwent monophthongization (Ochiai 2015a) and became *humpure*. The infix *\<um>* tends to change into *\<un>* in modern Paran Seediq. Now, *humpure* has variants as *hunpure* or even *pure*. Bu-hwan *papurai* looks like a reduplicated form of *puray*. Taroko Seediq, on the other hand, retains the form of Proto-Seediq.

In (44), Proto-Seediq *qaqay is lost in Bu-hwan and Paran Seediq, being replaced by *papak*. Taroko Seediq retains the proto-form. In (134) also, Proto-Seediq *ma-kəsaraw is lost in Bu-hwan and Paran Seediq is replaced by *taiăkarăk* and *turuqeraq*.

Table 3 Innovations only in Taroko Seediq

	Bu-hwan	Paran	Taroko	Proto-Seediq
71	timu 'salt'	timu	**c**imu	*timu
122	tikŏh 'little'	tikuh	**c**ikuh	*tikuh
151	matanach 'red'	mutanah	em**b**anah	*ma-tanah
13	makaidil 'woman'	muqedin	**kuyuh**	*ma-qaidil
34	ŭduthiŋ 'throat'	gelu	glu	*gəlu
35	--- 'neck'	nduyuŋ	**waru**	*gəduðuŋ
43	turŭŋ 'breast'	teeraŋ	**brah**	*tʔeraŋ
53	dapa 'cow'	dapa	**kaciŋ**	*dapa

There are segment changes in (71), (122), and (151). In (71), Proto-Seediq *t* in *timu 'salt' changes into *c* in Taroko Seediq only. In Bu-hwan and Paran Seediq, it remains as the proto-form. Similarly, in (122), *t* in Proto-Seediq *tikuh changes into *c* in Taroko Seediq. In (151), the *t* in Proto-Seediq *ma-tanah changes into *b* in Taroko Seediq. In (13), (35), (43), and (53), Taroko Seediq lost the Proto-form and replaced it with another form. In (13), Taroko Seediq lost *ma-qaidil and replaced it with *kuyuh*. Similarly, Proto-Seediq *gəduðuŋ 'neck' is lost and replaced by *waru* in (35). Proto-Seediq *tʔeraŋ is lost in (43). *Brah* 'front' became the word for 'breast, chest.'[118] In (53), Taroko Seediq lost Proto-Seediq *dapa and replaced it with *kaciŋ*.

---

[118] There is another possibility that *brah* is a retention of the meaning 'chest' from Proto-Seediq. The Proto-form *tʔeraŋ could have meant something different from 'chest,' for example 'ribs' (Tsuchida (1976: 227, 257) reconstructed Proto-Hesperonesian *təgəRaŋ with the meanings 'chest, breast' as well as 'ribs').

Table 4 Innovations either in Bu-hwan and Paran Seediq or only in Taroko

	Bu-hwan	Paran	Taroko	Proto-Seediq
70	barass 'rice'	beras	buwax	*buwax/beras
127	ukach 'no'	uka	oŋat	*uka/uŋat
67	shuduh 'grass'	sudu	spleq	*sudu/səpəriq
120	dĕhakuh 'short'	lulebu/dehekko[119]	lbu	*ləlebu (?)
103	haluŋ 'gun'	haluŋ	puniq	*haluŋ (?)
97	sinmadat 'knife'	hulmadac	hrŋudux	---
162	egu 'many'	egu	lala	---

In (70) and (127), two forms are reconstructable as Proto-Seediq. Bu-hwan and Paran Seediq have one of the proto-forms, and Taroko Seediq has the other form. In (67), Taroko Seediq *spleq* 'grass' has a cognate in Paran Seediq as *superiq.* However, Taroko Seediq seems to lack *sudu.* It is not clear if *sudu* is an innovation in Bu-hwan and Paran Seediq or if it is lost in Taroko Seediq. In (120), 'short' in Taroko Seediq seems to lack the cognate. In (103), Taroko Seediq *puniq* is also the word for 'fire' and this form means 'fire' only in Paran Seediq. It is not clear if *haluŋ* is an innovation in Bu-hwan and Paran Seediq or if this form existed at an earlier stage of Seediq but Taroko has lost this form. In (97) and (162), Proto-Seediq is unreconstructable due to lack of cognates in Atayal. However, the pattern of Bu-hwan and Paran Seediq on the one side and Taroko Seediq on the other is obvious throughout Tables 2, 3, and 4.[120]

5. Corrections on Bu-hwan list by Asai (1953) with annotation

6. **abarao** 'they'   Asai says *abarao* is 'they,' but this form, *hubaraʋ*, means 'many, multitude.' The form today is *hubaro*, with a final monophthong.

9. **nakaga** 'his'   There is no similar form with the meaning of a pronoun. Asai assumes that *nakaga* is *na gaga* 'of there'. *Na* is a possessive marker, and *gaga* is a distal demonstrative. Thus, *na gaga* means 'of that thing' or 'of that person.'

10. **kakashaduk** 'our'   There is no similar form with the meaning of a pronoun. Asai assumes that *kakashaduk* is [gaga ɕeɖiq], literally 'there man.' The corresponding form is *gaga seediq* 'a person is there.' Bu-hwan list has a word for 'man' as *adŭk* (12) . This form matches with the *-aduk*, in the last half of *kakashaduk* (See Section 6 'Person').

11. **nataha** 'their'   Asai analyzes it as *na daha* 'their.' *Na* is a possessive marker. As Asai (1953: 42) shows, *daha* is a bound pronoun for the third person plural. The

---

[119] The second form is from Ogawa and Asai (1935).

[120] The single exception to shared innovation between Bu-hwan and Paran Seediq is (164), where Bu-hwan *bilŭk* and Taroko Seediq *bilaq* share the same form but Paran Seediq *biciq* underwent sporadic sound change of *l* to *c.*

problem is that, at least synchronically, the possessive marker *na* is not followed by a bound pronoun but by a free pronoun. The free pronoun for the third person plural is *deheya*. *Na deheya* 'their' is acceptable (e.g., *neepah na deheya* 'their field'), but *na daha* is not (e.g., **neepah na daha* 'their field'). Asai (1953: 42) has a genitive form for the third person plural, *nɯɯdəhiða*, which corresponds to *na deheya*. One explanation could be that *nataha* meant to be *na deheya*. Another explanation could be that *na daha* was actually the form in use at that time. *Daha* is a polysemous word. It also means 'two.' Yet another possibility is that the informant gave the meaning of 'two people' for *na daha*.

19. **nakial** 'brother' Asai suggests that *nakial* originates in the numeral 'one,' and it means 'of one.' The *na* of *nakial* is a possessive marker. The rest, *kial*, means 'one' (170). He goes on to say that a southern dialect of Atayal uses *qottoħ* 'one' for siblings. The forms are either *qottoħ java* or *qottoħ jaja*. These literally mean 'one father' and 'one mother,' respectively. *Nakial* seems to have been lost in modern Paran Seediq. Now, 'sibling' is either *qubusuran* 'elder sibling regardless of gender' or *suwai* 'younger sibling regardless of gender.' If there is a need to specify the gender, *ruseno* 'man' or *muqedin* 'woman' follows these words (e.g., *swai muqedin* 'a younger female sibling').

23. **taina** 'father' Asai suspects that *taina* is a misprint and gives *tama* as a word for 'father.' It is possible that Bullock had correctly transcribed the form as *tama*; however, his handwriting of *m* could have been misinterpreted as having two parts, *i* and *n*, when he re-transcribed it.

33. **muduski** 'beard' Asai gives the form as *ŋudʊs*. This form corresponds to *mudus*, the first half of Bullock's form. The final segments *ki* should be a demonstrative 'that.' In the modern Paran Seediq, the demonstrative referring to an object near to the hearer is *kiya*, and its shortened form is *kii*. Bullock may have been pointing to his beard when he was asking for a word for 'beard.' The informant seemed to reply '*ŋudʊs kii*' (That is a beard).

34. **ŭduthing** 'throat' Asai gives the form as *nədɯɯðʊŋ*. The present form is *nduyuŋ* [n̩.ˈdu.juŋ] with a syllabic *n* at the initial.[121]

37. **abatha** 'arm' Asai gives this form as *baga*. The *th* in Bu-hwan corresponds to *g* in Asai's. Bullock might have misheard the consonant *g* as *θ*. The initial vowel, which was present in Bu-hwan, was lost before 1927 as it is absent in Asai's form.

39. **abatha** 'hand' As Asai states, Paran Seediq makes no distinction between 'arm' and 'hand.' The forms for 'arm' (37) and 'hand' (39) are identical (See Section 6 'Hand, Five, and Shoulder').

---

[121] There are variants such as [nu.ˈdu.juŋ] and [un.ˈdu.juŋ].

40. **tuludiŋ** 'finger'   Asai says that the form does not mean 'finger' but it is a part of a compound *təlʊɖiŋ baga* 'index finger.' According to Asai, *təlʊɖiŋ* means 'to point at someone.' However, this form has not been observed in modern Paran Seediq as far as the present author knows (See Footnote 31). The same meaning is expressed by *tumiyu* 'to point at someone (actor voice present form[122]).' *Baga* means hand as in (39).

47. **jitio papŭk** 'toe'   The second word *papŭk* is 'leg' (44). Asai reports that *jitio* is not in use. The word for 'toe' at his time was either *katɕhil pappak* or *wawa pappak*. In modern Paran Seediq, the first word in the first compound corresponds to *kehiŋ* 'finger' (40). *Wawa* means either 'baby,' 'meat,' or 'finger' (metaphorically, babies of a hand).

50. **tamabahak** 'heart'   Asai proposes that *tama*, the first half of *tamabahak*, is *tama* 'father,' but he does not give the meaning for *bahak*. He adds that this form is not in use but has been replaced by *rəmmʊnaĭ*. However, the corresponding form, *rumune*, in modern Paran Seediq means 'kidney.' 'Heart' in modern Paran Seediq is *toma bahak*. *Toma* means 'under' and *bahak* means 'lung.' Thus, *toma bahak* means '(an organ) under the lung.' Taroko Seediq is *tama baraq*. Bu-hwan and Taroko Seediq both have *tama* as the first element of the compound, so Paran Seediq seems to have replaced *tama* with *toma*.

51. **churuk** 'fish'   The final consonant is *k*, a stop, in Bu-hwan. Asai reports this form as *katʊrʊħ*, with a final fricative. The present form is *qucurux*.

52. **tamat** 'deer'   Asai says this means 'animal' not 'deer.' His form for 'animal' is *tsamat* and for 'deer' is *waqqinʊħ*. The present forms are *camac* and *ruqenux*.

63. **hazi kăhoni** 'leaf'   Asai states that *hazi kăhoni* means 'fruit,' and not 'leaf.' The word for 'fruit' in modern Paran Seediq is also *hei quhuni*. *Hei* means body, and *quhuni* means 'tree' (62).

64. **raparap** 'root'   Asai reports that he could not find this form; instead, *gamil* was obtained. Nor could the present author find a form similar to *raparap*.

70. **barass** 'rice'   Asai explains that *barass* refers to hulled rice or hulled millet. He adds that rice that has not been hulled is called *paðaĭ*. In modern Paran Seediq, these forms are *beras* and *paye*.

84. **kushun** 'day'   Asai says the corresponding form, *kuɕʊn*, does not mean 'day' but 'tomorrow.' *Kusun*, the modern form, also means tomorrow. The present forms for 'day' are *diyan* (as in day and night), and *ali* (as in Monday, etc.).

91. **haruŋ** 'wood'   Asai says the corresponding form, *harʊŋ*, means torch or wood used for a firebrand. 'Wood' in modern Paran Seediq is *quhuni*, the same form for

---

[122] The base is *tiyu*. The infix <*um*> is inserted after the initial consonant.

'tree' (62). 'Torch' in the modern form is *haruŋ*, which is also a word for 'pine tree' (91).

123. **mărïemuch** 'round'   Asai proposes that the corresponding form is *məlimʊk*, but it means 'brassy or golden.' The corresponding form in Paran Seediq would be *limuk*, which means 'bullet case' but this tool is not used anymore. Proto-Atayalic *limuk 'pot' (Li 1981) should be the cognate. Perhaps, a pot was made of metal and usually round, brassy, and golden. 'Round' in the modern form is *mutumun*.

128. **balaiwa** 'yes'   Asai analyzes *balaiwa* as made up of two parts, *balaĭ* and *wa*. He says the two words as a whole mean 'true.' As Asai suggests, the modern form is also *bale wa* '(You may not believe it but,) it is true.' *Bale* means 'true' and *wa* is a final particle. *Un* is a word for 'yes, it is so' in Paran Seediq.

131. **manakamakan** 'eat'   Asai analyzes this form as made up of two parts *meɳak mekkan* 'come to eat?' The corresponding form in modern Paran Seediq would be *meniq mekan*. The first word, *meniq*, however, does not mean 'come' (*meyah* is the form for 'come'). *Menaq* means 'to stay, to be in a place.' If this analysis is correct, the informant gave 'to stay and to eat' for the meaning of 'eat,' the part for 'to stay' seems unnecessary. It is possible that *meniq* is used as a kind of tense aspect (e.g., progressive) marker; however, it does not have that function in the modern Paran Seediq. The progressive is marked by the demonstratives *gaga* 'that' and *nii* 'this' or the verb *gisu* 'to come towards a speaker' (Lin 2005).

132. **nimah** 'drink'   Asai says the form for 'drink' should be *mimaħ*. *Nimaħ* indicates a past tense of *mimaħ*.

141. **lĕminish** 'cry (weep)'   Asai comments that the form he collected is *lumiŋis*, with a velar nasal in the middle, not an alveolar nasal.

146. **papurai** 'roast'   Asai says it is a causative form of the verb 'to cook.' Along the line of Asai's analyses, the base *pure*[123] may be prefixed by the causative marker *pu-*. Still another possibility is that *pu-* is a reduplicant of the first consonant of the base *pure*. However, reduplication of verbs is not productive in the modern language; therefore, there is little evidence to decide which is the correct analysis.

157. **tuŋarĕt** 'north'   Asai says the form should be *tʊ gɯ-narats*, which means 'on the right (side).' *Tugu-* is a prefix with the meaning of 'towards' and *narac* is 'right.' The exact transcription of Bu-hwan is *tungarĕt*. He might have written *tugnarĕt*, but later the *n* and *g* in the middle were switched accidentally while transcribing into other scripts, resulting in *tungarĕt*.[124]

---

[123] The root are either *hupure* or *pure* (See Section 4, the explanation for (36)).
[124] Syuntarô Tida has pointed out this possible explanation to the author.

158. **tagaēril** 'south'   Asai says the form should be *tɯɡɯ-iril*, which means 'on the left.'

167. **mua** 'where'   Asai explains that *mua* should correspond to *hamuwa* 'how.' The corresponding forms in the modern languages are *humuwa* 'how do.' Asai says the form for 'where' is *inno*.

179. **nahal** 'ten'   Asai suggests that *nahal* is a misprint and gives *mahhal* as a correction. The tendency for Bullock to spell *m* as *n* is also seen in (23) (*ta**m**a* as *tai**n**a*) and (132) (***m**imah* as ***n**imah*).

6. Additional notes on the Bu-hwan list

Person   Bullock reports 'man' as *edŭk* (12), and this form, as Asai has correctly pointed out, appears in Bu-hwan *kakashaduk* 'our' (10), which is analyzed into two parts *kaka* and *shaduk*. The latter corresponds to *seediq* 'person' in the modern Paran Seediq. Then, Bu-hwan *edŭk* corresponds to *shaduk* in meaning although the latter has an initial segment *s*. These forms suggest that Bu-hwan had *ediq* 'person' as well as *seediq*.[125] There is no form such as *ediq* in modern Paran Seediq. *Seediq* is inseparable; the prefix-like segment, *s-*, is fused into one word. This prefix could be a personal article related to the prefix **si-* in **si-ima* 'who' in Proto-Austronesian.[126]

Hand, **Five**, and **Shoulder**   'Hand' in the modern Paran Seediq is *baga*. Bu-hwan *abatha* (39) has an extra vowel *a*. It can be assumed that the form was *abaga*, based on the forms in Paran Seediq and Taroko Seediq. Ochiai (2015b) explains that this form has a close resemblance to Proto-Austronesian **qabaRa* 'shoulder.' **R* in Proto-Austronesian appears either as *r* or *g* in Seediq (Li 1981). In addition, **q* in Proto-Austronesian tends to be lost at word-initial position in Paran Seediq (Li 1981). Thus, the change from **qabaRa* to Paran Seediq *abaga* is plausible. Furthermore, 'hand' in Atayal is *qbaʔ* (Egerod 1980), the cognate of *baga*. Atayal still retains the initial *q*. Then, 'shoulder' (36) in Proto-Austronesian became 'hand' in Atayal and Seediq.

Ferrell (1969) lists the word for 'hand' in the Formosan languages (Kanakanavu, Saaroa, Rukai, Favorlang, Thao, Pazeh, Saisiyat, Taokas, Paiwan, Puyuma, Amis, Bunun, Kavalan, Siraya, Atayal, and Seediq).   By investigating this list, Ochiai (2015b) remarks that it is clear that the forms for all languages are derived from Proto-Austronesian **lima* 'hand,' except for the Atayalic languages.[127] She goes on to say that the Atayalic languages could also have used **lima* for 'hand.' As **lima* 'hand' acquired a new meaning as 'five' in the Atayalic languages, 'hand' was replaced by

---

[125] As Holmer (1996: 215) points out, a stem that begins with a vowel actually has a glottal stop as onset. So, in this case, it should be *ʔediq* or *s-ʔediq*. The latter is pronounced as [seʔediq] or [seediq].

[126] For instance, Bunun has *símaʔ* 'who' (Ferrell 1969).

[127] Tsou is also the exception in which the word for 'finger' acquired the meaning of 'hand' (Sagart 2013: 482–483).

another form, which used to mean 'shoulder,' in order to avoid the confusion between 'hand' and 'five.'[128]

Cloud   Bu-hwan is *ruluŋ* (78), and Taroko Seediq is also *ruluŋ*. However, Paran Seediq has a different form, *pulabu*. From Bu-hwan and Taroko Seediq, it is clear that the older form for 'cloud' is *ruluŋ*.[129] The origin of *pulabu* seems to be *labu* 'to wrap, cover.' *Pu-* is a causative prefix. 'Cloud' in Paran Seediq might have replaced *ruluŋ* with *pulabu* 'causing (the sky) to be covered.'

Lightning   Bu-hwan is *sassaina* (80), which seems to have been lost in Paran Seediq and Taroko Seediq. However, Taokas, an extinct Formosan language spoken in the Northwestern plains, has a similar form to Bu-hwan. The form in Taokas is *saisanad* (Ino 1998: 118). According to Ino, the meaning is not 'lightning' but 'star.' They could be cognates.

Ugly   Paran Seediq seems to lack a word for 'ugly.' The Bu-hwan form *sĕsĕchaun* (116) corresponds to *susiqaun* in modern Paran Seediq, which means 'feel embarrassed, ashamed.' It could either mean that an ugly person feels embarrassed to see others or others feel embarrassed to see an ugly person. Ugliness is expressed by a metaphorical extension of 'embarrass.' In Paran Seediq, *naqah qutaan* 'it is not good to see' is also used as an equivalent of 'ugly.' *Naqah* means 'bad' (112), and *qutaan* means 'to see (undergoer voice, location subject, present tense).'

Hot   In Bu-hwan *matatiluch tidao* (117), the first word corresponds to *mutilux* in Paran Seediq. It means 'hot' by itself. The second word is not used in Paran Seediq; however, it shows a close resemblance to Bu-hwan *hidao* 'sun' (81). In Paran Seediq, there is a form *tihido*, which is derived from *hido* 'sun,' by prefixing *ta-*. It is a prefix described as having a passive function in Ogawa and Asai (1935) (*tahido* becomes *ti-hido* (See Table 11)). *Tihido* means 'The sun blazes.' Bullock's imformant might have meant 'It is hot. The sun is blazing.'

Yellow   The Bu-hwan form is *tanah-mahabai*. The first part is the root of *matanach* 'red' (151). In modern Paran Seediq, 'red' is *mutanah*. *Mu-* is a prefix characteristic of stative/adjectival expressions in Paran Seediq. However, modern Paran Seediq has lost the second word, so its meaning is not clear. Still, it is clear that Seediq people regarded yellow as a kind of red. Taroko Seediq has *mxeboŋ* for 'yellow,' which looks like a cognate of *mahabai*. Modern Paran Seediq replaced the older form with *muguciyas*, although the origin of this word is unknown.

Blue   The Bu-hwan form is *masama*. This is derived from *sama* 'vegetables.' Today, Bullock's form, *musama*, is not used; instead *mugusama* is used but with a slightly

---

[128] The etymology of Proto-Seediq **hiraŋ* 'shoulder' (36) is not yet known.

[129] The form remains marginally in a personal name (e.g., *Walis Ruluŋ* (literally *Walis* of *Ruluŋ*; *Ruluŋ* is the father)), but the meaning of 'cloud' has been lost.

different meaning. *Mugusama* is used for 'green' as in (155), and *mubulawa* is used for 'blue.' There is a similar form, *Mutulawa*, which means 'bruise.' These share the same root *lawa*.

East and **West**   In Paran Seediq, *daya* (159) means 'uphill.' The Seediq people used to live on the western side of the central mountain ridges. The peak is the direction of *daya*, which coincides with east. However, a series of immigrations of three tribe groups made the situation complicated. The word for 'east' was created recently as *hreyan hido* 'the place where the sun shines' (Mandivaiian 2009: 118).[130]

 Hunac (160) in Paran Seediq means 'downhill.' The plain is the direction of *hunac*, which coincides with the west. Proto-Seediq is reconstructed as *hunat or *rahut. The latter is not used anymore, but it is still seen in a derived form *tugu-rahut* 'towards the lower side of a slope.' *Hunat may have undergone a few sound changes including metathesis of *ra* and *hu* (*hurat*), and *r* to *n* (*hunat*). It is *rahut that corresponds to Proto-Austronesian *lahud. The word for 'west' was created recently as *gqiyan hido*[131] 'the place where the sun inserts (itself into the layers of mountain ridges)' (Mandivaiian 2009).

Part   Bullock did not collect this item (161), and the author has difficulty finding words for 'a part' as opposed to 'whole.'[132] The approximation for 'a part' would be *icin*, which means a pair of things, such as body parts. Both parts are expressed by *kana* 'all,' such as both hands, both legs, and both eyes. A pair is expressed by *icin* (e.g., *icin papak* 'a leg').[133]

All   Bullock's list has *ramut* (163), but this form is not found in Paran Seediq, which has *kana* 'all' instead. However, Tsuchida's (1982) list of Western plain languages shows that Taokas has an identical form, *ramut* 'all.'[134] Seediq might have had *ramut* 'all' at the time Bullock collected vocabulary but since lost it.

Nine   Bullock's list has all the numerals from one to ten except for nine (178).[135] However, in all other languages (Tsui-hwan, Sek-hwan, Pepo-hwan, and Pelam-hwan) except for Favorlang,[136] Bullock and Steer collected all the numerals from one to ten. In Paran Seediq, 'nine' is *muŋari*. Taroko Seediq has nearly the same form *mŋari* [məŋari].

 Pecoraro (1977) reports that 'nine' in Taroko Seediq is *maŋali*. He suggests that this

---

[130] *Hreyan* may be equivalent to *huriyan* 'a shiny place,' which is the locative voice of *h<um>iru* 'to light up.' In Mandivaiian (2009), vowels before penultimate syllables are not written.

[131] *Gqiyan* (or *guqiyan*) is derived from *g<um>eqi* 'to insert.'

[132] *Mureemux* seems to mean 'whole.'

[133] *Icin* also means 'somewhere else.'

[134] Taokas is the language with which Bu-hwan seems to share another cognate, *sassaina* 'lightning' (80).

[135] It is only a speculation, but 'nine' may have been expressed by 'eight' and 'one,' or 'by adding one, it becomes ten.'

[136] Ino (1996: 177) interviewed Favorlang descendants in 1897, and said "With regard to numerals, no one seems to know how to express 'nine' (translated by the present author)."

form is related to *ŋali* (the imperative form of *ŋal* 'to take'), and 'nine' originally meant 'take one (from ten objects)!' This explanation is slightly problematic. Taroko Seediq *mŋari* 'nine' has *r*, not *l*. On the contrary, *maŋal* 'take' in Taroko Seediq has *l*, not *r*.[137]

7. Sound changes from Bu-hwan to modern Paran Seediq

This section compares Bu-hwan and modern Paran Seediq and describes sound changes. These changes are: *au̯ > o* (Table 5), *ai̯ > e* (Table 6), *l > n* word finally (Table 7), *t > ts* word finally (Table 8), *ð > j* (Table 9), and pre-stress neutralization (Table 10). In addition, vowel harmony mediated by *h* (Table 11) and redundant final segments in Bu-hwan (Table 12) are explained.

Bu-hwan's *ao*, which corresponds to *au̯* in Asai (1953), becomes *o* in modern Paran Seediq (Table 5).

Table 5   *au̯ > o*

	Bu-hwan	Paran Seediq
6	abar**ao** 'many'	hubar**o**
15	laqai risin**ao** 'boy'	laqi rusen**o**
81	hid**ao** 'sun'	hid**o**
144	maraŋ**ao** 'talk'	mureŋ**o**

Table 6 shows examples of *ai* in Bu-hwan and *e* in modern Paran Seediq.

Table 6   *ai̯ > e*

	Bu-hwan	Paran Seediq
128	bal**ai**wa 'yes'	bal**e** wa 'It is true.'
133	makak**ai**sa 'walk'	mukuk**e**sa
146	pupur**ai** 'roast'	hunpur**e**
152	bahag**ai** 'white'	beheg**e**
166	kanah**ai**nu 'how many'	kunuh**e**nu

Table 7 shows that word final *l* in Bu-hwan become *n* in modern Paran Seediq.

---

[137] Another possible explanation could be that Paran Seediq *muŋari* is related to *nuŋari* 'leftover'; however, this explanation is not unproblematic either.

Table 7 *l* > *n* word finally

	Bu-hwan	Paran Seediq
59	uba**l** 'feather'	uba**n**
89	kareŋu**l** 'smoke'	qureŋu**n**
115	maŋihu**l** 'bitter'	muŋihu**n**
136	mahŏka**l** 'die'	muhuqi**n**
170	kia**l** 'one'	kiŋa**n**
179	naxa**l** 'ten'	maxa**n**

The words that end in *t* in Bu-hwan have *c* [ts] in modern Paran Seediq (Table 8). However, as Yang (1976: 619) has noted, the final [ts] is observable only when it is pronounced separately or when it is at the end of an utterance. If a word follows, it becomes [t]. For example, *miric* [mirits] 'goat' is pronounced as *miric lumiqu* [mirit lumiqu] 'goat (of a mountain).'

Table 8 *t* > *ts* word finally

	Bu-hwan	Paran Seediq
52	tama**t** 'deer'	cama**c**
97	sinmada**t** 'knife'	hukmada**c**
104	kabuki**t** 'powder'	qubuli**c**
147	hamaŋu**t** 'boil'	humaŋ**c**
157	tuŋarĕ**t** 'north'	tugunara**c** 'right'
173	sŭpŭ**t** 'four'	sepa**c**
177	mŭssupa**t** 'eight'	mumusepa**c**

Bullock wrote *d*, *dz*, and *th*, which correspond to [ð] in Asai (1953). This segment has changed into *y* [j] in modern Paran Seediq (Table 9).

Table 9 ð > j

	Bu-hwan	Paran Seediq
20	swa**dz**u makaidil 'sister'	suwa**y**i[138]
74	**dz**adzuŋ 'river'	**y**ayuŋ
138	mai**dz**ah 'come'	me**y**ah
61	ku**z**u 'snake'	qu**y**u
63	ha**z**i kăhoni 'fruit'	he**y**i quhuni
77	ku**z**uch 'rain'	qu**y**ux
92	dagi**z**ak paru 'mountain'	dugi**y**ak
102	sinu**z**uk 'cord'	sunu**y**uk
34	ŭdu**th**iŋ 'throat'	ndu**y**uŋ

The forms recorded by Asai (1953) in 1927 indicate that the sound changes in Tables 5 to 10 had not yet occurred. The forms reported in Yang (1976) indicate that the changes were already completed; therefore, these sound changes were in progress between 1927 and 1976.[139]

Another sound change is pre-stress neutralization. Vowels *a*, *ă*, *e*, *ě*,[140] *i*,[141] and *ĭ* before the penultimate syllable (the stressed syllable), undergoes neutralization to *u*. Stress is shown as an acute accent in Table 10.[142]

Table 10 Pre-stress neutralization

	Bu-hwan	Paran Seediq
6	**a**barao 'many people'	h**u**báro
13	m**a**kaidil 'woman'	m**u**qédin
30	p**a**dahuŋ 'lips'	p**u**dáhuŋ
73	k**a**sia 'water'	q**u**síya
79	m**a**b**a**rua 'thunder'	m**u**b**u**rúwa
85	b**a**bien 'night'	b**u**bíyan
89	k**a**reŋeul 'smoke'	q**u**réŋun
93	b**a**tunuch 'stone'	b**u**túnux
94	b**a**nakail 'sand'	b**u**náquy
104	k**a**bulit 'ash'	q**u**búlic

---

[138] It is, in fact, difficult to determine if it is [suwai] or [suwaji].

[139] Ochiai (2015a) asserts that Paran Seediq has undergone monophthongization of word final *ai̯* and *au̯* to *e* and *o* during 1927 and 1976.

[140] There are four items with pre-stress *e* or *ě* in Bu-hwan; three of them are related to infixes <um> (actor voice, present tense) or <un> (undergoer voice, patient subject, present tense). These are l<um>edax 'shine' (86), l<um>iŋis 'weep' (141), d<un>amux 'roof, literally, what was thatched' (107).

[141] Bu-hwan pre-stress *i* follows *s* (15, 97, 102).

[142] With regard to Bu-hwan accent, Bullock (1874: 40) explains that "In that of Bu-hwan the accent is in most words laid on all the syllables equally."

	Bu-hwan	Paran Seediq
114	**ta**bashi 'sour'	**tu**bási
115	**ma**ŋihul 'bitter'	**mu**ŋíhun
117	**ma**tatiluch tidao 'hot'	**mu**tílux
118	**maba**tunuch 'beautiful'	**mubu**túnux
119	**ka**nadish 'long'	**qu**nédis
133	**maka**kaisa 'walk'	**muku**késa
135	**ma**takai 'sleep'	**mu**táqi
136	**ma**hŏkal 'die'	**mu**húqin
142	**ma**hulish 'laugh'	**mu**húlis
144	**ma**raŋao 'talk (future tense)'	**mu**réŋo
151	**ma**tanach 'red'	**mu**tánah
154	**ma**sama 'blue'	**mu**gusáma 'green'
156	**ma**kaluch 'black'	**mu**qálux[143]
158	**ta**gaēril 'south'	**tu**guírin
165	**ka**nuan 'when'	**ku**núwan
166	**ka**nahainu 'how many'	**ku**nuhénu
175	**ma**taru 'six'	**mu**mutéru
134	**tai**ăkarăk 'lie down'[144]	**turu**qéraq
62	**kă**honi 'tree'	**qu**húni
87	**mă**kuun 'dark'	**mu**kúunŋ
88	**hă**punek 'fire'	**pu**níq
108	**ră**heŋun 'door'	**ru**héŋun
126	**mă**sekuĭch 'cold'	**mu**sékuy
129	**mă**tugessa 'teach (future tense)'	**mu**tugésa
143	**mă**hoyesh 'sing'	**mu**úyas
86	**le**medeuch 'light'	**lu**médax 'shine'
107	**de**namuch 'roof'	**du**námux
68	**bĕ**lĭbeul 'banana'	**bu**lébun
141	**lĕ**minish 'weep'	**lu**míŋis
169	**chĕ**chekach 'half'	**cu**céka
15	lakai **ri**sinao 'boy'	laqi **ru**séno
97	**si**nmadat 'knife'	**hu**lumádac
102	**si**nuzuk 'cord'	**su**núyuk

In three items, pre-stress *a* of Bu-hwan becomes *e* of Paran Seediq as in Table 11. This is due to a sound change mediated by *h* in a stressed syllable. An acceenied vowel

[143] There is a variant *muguqalux*.

[144] Paran Seediq form suggests that the third segment of Bu-hwan could be *r*, instead of *i*. Bullock could have mistaken *r* as *i* when he was transcribing *tarăkarăk*. The expected form, **tar**ăkarăk, contains the prefix *ta-*, which indicates passivity (Ogawa and Asai 1953, See Section 6 'Hot') or involuntariness (Ochiai 2015c).

is copied to the preceding syllable when the accented vowel is preceded by *h* (Yang 1976: 668).

Table 11 Vowel harmony mediated by *h*

	Bu-hwan	Paran Seediq
58	b**a**hani 'bird'	qubeh**é**ni
100	b**a**hĕnek 'bow'	beh**é**niq
152	b**a**hagai 'white'	beh**é**ge

In addition to these sound changes, there are eight cases of irregular segment addition at the end of a word in Bu-hwan froms, where modern Paran Seediq lacks such segments (Table 12). Bullock might have heard a non-phonemic noise at the end of an utterance as part of a word, such as a puff of air (written as *h* or *ch*), closing of the lips (written as *p*), or contact with the alveolar ridge (written as *t* or *l*).

Table 12 Redundant final segments in Bu-hwan

	Bu-hwan	Paran Seediq
45	puŋu**h** 'knee'	puŋu
67	shudu**h** 'grass'	sudu
94	banakai**l** 'sand'	bunaquy
95	pila**t** 'silver'	pila
111	malu**p** 'good'	malu
126	mǎsekuĭ**ch** 'cold'	musekuy
127	uka**ch** 'no'	uka
169	chĕcheka**ch** 'half'	cuceka

It is also possible that these segments in Bu-hwan have been lost in modern Paran Seediq. However, at least for (94) and (126), it is clear that these final segments are redundant based on the syllable structure of Seediq.[145] Ochiai (2015a) observes that in Paran Seediq, a diphthong is allowed only in a final syllable and it cannot be followed by another consonant (CVV is acceptable as final syllable but not CVVC). In the Bu-hwan items (94) *banakail* and (126) *mǎsekuĭch*, diphthongs *ai̯* and *ui̯* are followed by another consonant. These forms are inconsistent with the syllable structure. This also suggests that other final segments in Bu-hwan may be redundant.

There still are exceptional sound correspondences between Bu-hwan and Paran Seediq, such as *lak**ai*** : *laq**i*** 'offspring' (17), *matak**ai*** : *mutaq**i*** 'sleep' (135), and *mahŏk**al*** : *muhuq**in*** 'die' (136). The *ai* (17, 135) or *a* (136) of Bu-hwan is preceded

---

[145] It is assumed that the basic syllable structures are the same for Bu-hwan and modern Paran Seediq.

by *q* of Paran Seediq, where Bu-hwan has *k*. Bu-hwan does not distinguish /k/ and /q/. These segments in question should be *q* based on the Paran Seediq data. It was this *q* that caused the lowering of the following vowel /i/, and it was heard by Bullock as [ai] or [a].

8. Conclusion

This paper reconstructed Proto-Seediq forms for Bu-hwan items, based on Paran Seediq, Taroko Seediq, Atayal, Proto-Atayalic, and Proto-Austronesian (Table 1), and showed that Bu-hwan is more closely related to Paran Seediq than Taroko Seediq with shared innovations in Bu-hwan and Paran Seediq as evidence. Table 1, however, lacks many forms for Proto-Atayalic. In fact, many cognates in Seediq and Atayal remain unreconstructed. Proto-Atayalic forms need to be complemented in the future.

References

Adelaar, Alexander. 1997. "An exploration of directional systems in west Indonesia and Madagascar". In Gunter Senft (ed.) *Referring to space: Studies in Austronesian and Papuan languages*. Oxford: Clarendon Press. pp.53–81.

Asai, Erin. 1953. *The Sedik Language of Formosa*. Kanazawa: Cercle Linguistique de Kanazawa.

Blust, Robert. 1995. "The prehistory of the Austronesian-speaking peoples: A view from language". *Journal of World Prehistory* 9. pp.453–510.

Blust, Robert. 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*. Taipei: Institute of Linguistics (Preparatory Office), Academica Sinica. pp.31–94.

Blust, Robert and Stephen Trussel. 2013. "The Austronesian comparative dictionary: A work in progress, web edition".
http://www.trussel2.com/acd/. Retrieved 2015/08/13.

Bullock, Thomas Lowndes. 1874. "Formosan dialects and their connection with the Malay". *China Review* : *Or Notes and Queries on the Far East* 3. pp.38–46.

Campbell, William (ed.). 1896. *The articles of Christian instruction in Favorlang-Formosan, Dutch and English from Vertrecht's manuscript of 1650: With Psalmanazar's Dialogue between a Japanese and a Formosan, and Happart's Favorlang vocabulary*. London: Kegan Paul, Trench, Trübner.

Campbell, William. 1913. *A dictionary of the Amoy vernacular spoken throughout the prefectures of Chin-Chiu, Chiang-Chiu and Formosa*. Tainan: Taiwan Jiaohui Guangbaoshe.

Council of Indigenous Peoples, Taiwan [原住民委員會]. Taroko online dictionary [太魯閣語詞典]. In Online dictionaries of Taiwan indigenous languages [原住民族族語線上詞典]. http://e-dictionary.apc.gov.tw/index.htm. Retrieved 2015/01/30.

Egerod, Søren. 1980. *Atayal-English dictionary*. London: Curzon.

Ferrell, Raleigh. 1969. *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Taipei: Academia Sinica.

Happart, Rev. Georgius. 1650. Woord-boek der Favorlangsche taal, waarin het Favorlangs voor, het Duits achter gesteld is. Manuscript [in Dutch][Translated in Campbell (1896)].

Holmer, Arthur. 1996. *A parametric grammar of Seediq*. Lund: Lund University Press.

Ino, Kanori. 1988. Xun Tai Ri Cheng [巡臺日乘]. In Tsunekazu Moriguchi (ed.) *Ino Kanori Fanyu Diaocha Shouce* [蕃語調査手冊]. Taipei: Southern Materials Center. pp.13–201.

———. 1996. *Taiwan fieldwork diary, vol.1* [in Mandarin][台灣踏査日記〈上〉], translated by Nan-jun Yang. Taipei: Yuan-Liou.

Li, Paul Jen-kuei. 1981. "Reconstruction of Proto-Atayalic phonology". *Bulletin of the Institute of History and Philology, Academia Sinica* 52. pp.235–301.

———. 1985. "The position of Atayal in the Austronesian family". In Andrew Pawley and Lois Carrington (eds.). *Austronesian Linguistics at the 15th Pacific Science Congress*. [Pacific Linguistics C-88]. Canberra: Pacific Linguistics. pp.257–280.

Lin, Hsiu-hsu. 2005. "The grammaticalization of tense/aspect auxiliaries in Seediq". *Concentric: Studies in Linguistics* 32. pp.111–132.

Mabuchi, Toichi. 1954. "Migration and distribution of the Formosan aborigines [in Japanese]". *Minzokugaku Kenkyu* 18, 1–2 pp.123–54, 18, 4 pp.23–72.

Mandivaiian, Aping (ed.). 2009. *The illustration dictionary of Sediq* [In Mandarin]. Nantou: Nantou County Aboriginal Administration.

Ochiai, Izumi. 2015a. "Sedekkugo Paran hōgen no nijūboin nitsuite [Notes on the diphthong in Paran Seediq]". *Proceedings of the 150th Meeting of the Linguistic Society of Japan*. Kyoto: Linguistic Society of Japan. pp.392–397.

———. 2015b. "Origin of the numeral 'five' in Northwestern Formosan languages: Revision of the Proto-Austronesian numeral system". The 25th Annual Meeting of the Southeast Asian Linguistic Society, Payap University, Chiang Mai, Thai. May 27–29.

———. 2015c. "A prefix *tu*- for involuntariness in Paran Seediq: Further evidence for *ta- in Proto Austronesian". The 48th International Conference on Sino-Tibetan Languages and Linguistics, University of California, Santa Barbara. August 21–23.

Ogawa, Naoyoshi and Erin Asai. 1935. *The myths and traditions of the Formosan native tribes*. Taipei: Taihoku Imperial University.

Pecoraro, Ferdinando. 1977. *Essai de dictionnaire Taroko-Français*. Cahier d'Archipel 7. Paris: Sociéte pour l'Etude et la Connaissance du Monde Insulindien.

Sagart, Laurent. 2004. "The higher phylogeny of Austronesian and the position of Tai-Kadai". *Oceanic Linguistics* 43. pp.411–444.

Sagart, Laurent. 2013. "Is Puyuma a primary branch of Austronesian? A rejoinder". *Oceanic Linguistics* 52. pp.481–492.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

———. 1982. "A comparative vocabulary of Austronesian languages of sinicized ethnic groups in Taiwan, part I. West Taiwan". *Memoirs of the Faculty of Letters, University of Tokyo* 7. Tokyo: Faculty of Letters, Tokyo University.

Tsukida, Naomi. 2009. "Sedekkugo (Taiwan) no bunpo [A grammar of Seediq (Taiwan)]". Ph.D diss, University of Tokyo.

Yang, Hsiu-fang. 1976. "The phonological structure of the Paran dialect of Seediq". *Bulletin of the Institute of History and Philology, Academia Sinica* 47. pp.611–706.