

# Dialectal distinctions in Plngawan Atayal: Current state and history [萬大泰雅語中的方言差異：現況與歷史]\*

Andre Goderich [郭育賢]

National Changhua University of Education [國立彰化師範大學]

## Abstract

Plngawan Atayal can be subdivided into two varieties: Sami'uɔ and Macagis. This subdialectal distinction went mostly unnoticed in linguistic publications barring a short mention by Chen (2012). Despite living in the same village since 1938, elderly speakers of both varieties have retained a number of unique phonological features. These features not only allow us to distinguish contemporary varieties, but also let us identify data from older publications. An examination of works by Ferrell (1969) and Li (1981, 1985) has revealed a prevalence of Sami'uɔ data, with a small amount of Macagis lexemes.

Keywords: Formosan, Atayal, dialectology, historical linguistics

關鍵詞：台灣南島語、泰雅語、方言學、歷史語言學

---

\*The research in this paper was funded by NSTC grant 110-2410-H-018-031. I would like to thank two anonymous reviewers for their suggestions. Any remaining errors are my own.

# 1 Introduction<sup>1</sup>

## 1.1 Background information

Plngawan is a dialect of Atayal spoken in one village (and surrounding hamlets) in Sinyi, Nantou County, Taiwan. It is the southernmost Atayal dialect, and an Atayal exclave, being neighboured by Bunun speakers to the west and south, and Seediq speakers to the north and east.

Plngawan Atayal as it is spoken nowadays can still be further subdivided into two main varieties: Sami’uɿ and Macagis. Despite living in the same village for decades, Sami’uɿ and Macagis speakers still broadly retain their speech patterns, which are readily identifiable from a number of unique sound changes. Both Sami’uɿ and Macagis underwent changes, so neither can be said to be more conservative or innovative.

And yet despite its unique characteristics, Plngawan has been the subject of relatively few linguistic studies when compared with the Squliq and Matu’uwal dialects. Plngawan has most often been studied alongside other Atayal dialects in broader comparative studies, and very little research has been done on the dialect itself. The existence of linguistic variety within Plngawan has only been acknowledged by Chen (2012) (and later by Goderich (2020), citing Chen). This paper will attempt to rectify this omission by addressing the issue of linguistic variety in Plngawan. The two main varieties—Sami’uɿ and Macagis—as well as their defining characteristics, are introduced in section 2. After that, these characteristics are used in section 3 to analyze several 20th century publications featuring Plngawan data, and determine the precise variety of dialect used therein.

### 1.1.1 Subgrouping hypotheses

Because of its isolation from other Atayal dialects and close contact with Seediq, Plngawan has long been considered the black sheep of the family, and researchers have struggled to put it squarely within either Atayal or Seediq (Li 1985, Rau 2004). Within the traditional “Squliq-C’uli” view of Atayal dialectal subgrouping, it has been considered a “C’uli” dialect by virtue of preserving

---

<sup>1</sup>This paper uses the following abbreviations: PA (Proto-Atayal), AV (Actor Voice), PV (Patient Voice), LV (Locative Voice), NAV (Non-Actor Voice, here meaning PV or LV, see Tsuchida 1976).

the <c> (/ts/) phoneme and not having a /q/ sound. More recent research by Goderich (2020) identified Plngawan as a divergent member of the Southern Atayal branch.

The findings in this paper do not affect our understanding of Atayal subgrouping as a whole. The placement of Sami'uɔ and Macagis under the Plngawan umbrella is self-evident, since the two subdialects are fully mutually intelligible. (And with only two varieties of Plngawan, their subgrouping is trivial.)

### 1.1.2 Phonological structure

There are no differences in the number and general quality of phonemes between the Sami'uɔ and Macagis varieties of Plngawan.<sup>2</sup> The consonant phonemes of Plngawan are presented in table 1. The pronunciation of most consonants is the same as their IPA values, unless indicated otherwise.

Table 1: Plngawan consonant phonemes

p	t	k	ʔ
b [b~β]		g [g~ɣ]	
	c [ts]		
	s	x	h
m	n	ŋ	
	l, r [r~r]		
w	y [j], ɹ		

Plngawan preserves all Proto-Atayal consonants except \*q, which merged completely into /ʔ/. However, the distribution of consonants underwent additional changes from Proto-Atayal to Plngawan, (see Goderich 2020:130–133 for details). Plngawan is unique among Atayal dialects for preserving Proto-Atayal \*ɹ as [ɹ], whereas all other Atayal dialects either lost it or merged it into other phonemes, mostly /j/. Thus, it has a total of three liquid phonemes: /r, ɹ, l/.<sup>3</sup>

<sup>2</sup>Future research may reveal measurable phonetic differences between the way some phonemes are pronounced by speakers of the two subdialects. This study did not include phonetic measurements.

<sup>3</sup>Note that the two rhotics (/r/ and /ɹ/) in Plngawan are transcribed differently by different authors. In this paper, I use the closest IPA symbols for ease of comparison and recognition, including when citing data from different sources. Readers should be careful to distinguish the two phonemes when comparing publications.

There is a total of five vowel phonemes in PIngawan, shown in table 2. Unlike most other Atayal dialects, PIngawan completely disallows schwa [ə]. The mid vowels /e/ and /o/ are mostly the result of vowel coalescence, e.g. Proto-Atayal \*riʔax > PIngawan /rex/ ‘day’, or Proto-Atayal \*raʔuŋ > PIngawan /pa-pa-roŋ/ ‘hook’ (Goderich 2020:131).

Table 2: PIngawan vowel phonemes

i	u
e	o
a	

PIngawan is one of three Atayal dialects (alongside Matu’uwal and Matu’aw) that preserve phonemic vowel distinctions outside the final foot, i.e. it does not reduce all prepenultimate<sup>4</sup> vowels. See for example PIngawan /buʔatiŋ/ ‘moon’ < Proto-Atayal \*buʔatiŋ (< Proto-Austronesian \*bulaN), PIngawan /pisaniʔ/ ‘taboo’ < Proto-Atayal \*pisaniq, and PIngawan /hapuniʔ/ ‘ember’ < Proto-Atayal \*hapuniq ‘fire’ (< Proto-Austronesian \*Sapuy).

Syllables must have an onset in PIngawan. Allowed syllable types are CV, CVC, and CGVC (Goderich 2020:44). Phonotactics differ somewhat between the two subdialects. While both varieties allow closed syllables to occur in any part of a word, there is a dispreference for non-final closed syllables in Macagis when compared to Sami’uʔ. This is explained further in section 2.1 and section 2.2, which talk about vowel epenthesis and cluster simplification in Macagis.

Stress in PIngawan is always word-final, as in all Atayal dialects.

### 1.1.3 Sources of information

The PIngawan data used in this paper, as well as linguistic data on other Atayal dialects, is from my own field notes, taken between 2013 and 2022, unless indicated otherwise. Proto-Atayal reconstructions are from Goderich (2020).

A total of five speakers were consulted for the project, three representing Sami’uʔ and two for Macagis. Three of the consultants were female, and two

<sup>4</sup>‘Prepenultimate’ here meaning ‘preceding the penultimate syllable’, i.e. third-to-last, fourth-to-last, etc.

male. The speakers were born between 1939 and 1961, and continue to live in and around the Plngawan area.

## 1.2 Historiography of Plngawan linguistic research

The earliest time I have been able to find Plngawan data in any published work is the word /ciʔuliʔ/ ‘person, other’ in Utsurikawa et al. (1935:sec. 1.1.3), written as *tsiule*. The book is an anthropological survey on the various Formosan villages during the later years of Japanese rule in Taiwan in an attempt to classify the indigenous peoples. Even though the authors of this study were not linguists, they used a rudimentary guesswork-based approach to try and group the “Atayal” peoples on the basis of several linguistic items they had collected and considered characteristic of the dialect as a whole. This resulted in the “Squiliq-C’uli” classification which later made its way into linguistic publications. In reality, because the word ‘person, other’ is so representative of the various sound changes that occurred in different Atayal dialects, even with the imperfect transcriptions in Utsurikawa et al. (1935) it is possible to identify several Atayal dialects in their data (not counting Seediq/Truku). For Plngawan, the main identifying feature is the vowel /i/ following the first consonant in *tsiule*. No other Atayal dialect has this vowel in the first syllable of this word. Utsurikawa et al. (1935) also feature the first use of the name “Plngawan”, written as “Perugawan”.

Plngawan continued to be lumped into the so-called “C’uli” dialect group for many decades to come. Ferrell (1969) includes a comparative dataset of a large number of Formosan languages, among them one he called “Ci’uli Atayal”, which was in fact Plngawan (note again the telltale /i/ in the first syllable). He also called it “Bandai dialect” (Ferrell 1969:75), using the Japanese name for the village (萬大). Ferrell conducted his own fieldwork on Plngawan, and this work is likely the first Plngawan wordlist ever published.

The name “Plngawan” is an exonym for the village, but it is now used by the speakers themselves to refer to their dialect. As is the case with many peoples, Plngawan speakers would generally refer to themselves using the word /ʔitaaʔ/ ‘human, person’ (cognate of “Atayal”), but this nomenclature never caught on as the name of the group. Li (1980) is the first linguistic work to use the name “Plngawan”, mentioning it briefly when discussing non-Squiliq dialects of Atayal. Li returned to writing on Plngawan on several occasions, usually within a comparative context (Li 1981, 1985), juxtaposing it with other

Atayal and Seediq/Truku dialects.

After the turn of the 21st century, occasional works on various aspects of Plngawan have been published: these include sociolinguistics (Rau 2004), phonetics-morphosyntax interface (Shih 2008), morphosyntax (Huang 2006), and morphophonology (Chang 2012, Chen 2012). Of these, Chen (2012) is notable because it is the only work which mentions sub-dialectal distinctions in Plngawan. Chen (2012:2–4) lists a few lexical differences between Macagis and Sami’u, and mentions that Macagis has an epenthetic /a/ where Sami’u has heterosyllabic consonant clusters.

## 2 Dialectal distinctions in Plngawan

Even though nowadays Plngawan is spoken in a single village with a few outlying hamlets, historically it used to be more widely dispersed. The area used to be dotted with smaller settlements, including the Macagis<sup>5</sup> and Sami’u<sup>6</sup> villages, which were the largest of those, and whose residents formed the bulk of the current unified tribal village. The people from these villages were forcibly relocated to Cin’ay village by the Japanese authorities around the year 1938 (close to the end of the Japanese colonial period), according to local oral histories.

As a result, the Plngawan dialect as it is nowadays spoken by Cin’ay elders is a mixture of several closely related dialects. This amalgamation presents a challenge to any linguistic study, much more so a phonological one focused on synchronic alternations, since speakers may readily switch between several variant forms of the same lexeme, or utilize forms originating from different Plngawan subdialects. It is often the case when doing fieldwork on Formosan languages that there is no such thing as a “pure” form of the local variety, and this is even more pronounced in Plngawan.

Plngawan speakers are aware of their ancestral origins, and still preserve their distinctive speech patterns. After almost a century living in the same village, we might expect there to be no variation left at all, let alone a pronounced distinction between variants, since almost everyone still alive today was born after the relocation. Nevertheless, despite the prolonged language contact be-

---

<sup>5</sup>Called ‘Makcagis’ by Sami’u speakers.

<sup>6</sup>Called ‘Samiuw’ by Macagis speakers.

tween the Sami'u and Macagis variants of Plngawan, many distinguishing features are retained by the speakers, which are listed and described below.

## 2.1 Vowel epenthesis in Macagis

Chen (2012:3) noted that for words with a CVC.CVC structure in Sami'u, Macagis will have CV.Ca.CVC instead, inserting an /a/ vowel to break up the heterosyllabic cluster. For some of these lexical items, we have Proto-Atayal reconstructions based on comparative data. Table 3 shows examples of such words, with Proto-Atayal given for comparison. In all of the words in the table, Macagis has a vowel /a/ where Sami'u has a consonant cluster, and the corresponding segment in Proto-Atayal is \*ə.

Table 3: Vowel epenthesis in Macagis, with Proto-Atayal roots

Proto-Atayal	Sami'u	Macagis	Gloss
*qalətiŋ	ʔaltiŋ	ʔalatiŋ	board; plank
*bəliŋ	ma-bliŋ	ma-baliŋ	to have a hole
*masəʔaŋ	masʔaŋ	masaʔaŋ	to scold
*marəŋuʔ	marŋuʔ	marəŋuʔ	dry
*təhaɯ	si-thaɯ	si-tahaɯ	left over
*tVgəliq	tagliʔ	tagaliʔ	waterfall
*gVhəɯaq	gahɯaʔ	hahɯaʔ	cold
*tuləqiŋ	tulʔiŋ	talaʔiŋ <sup>7</sup>	to hide (intr.)
*t<um>əʔətuʔ	t<un>tuʔ	t<um>atuʔ	to chop

Some verbs with the vowel /u/ in the final syllable of the root exhibit a weakening phenomenon, whereby the vowel is deleted in Sami'u and changes to /a/ in Macagis. This phenomenon is observed in all Atayal dialects, and is related to the historical origins of the vowel, which is a reflex of earlier \*ə (Goderich 2020:66–67). In the final syllable, this vowel became /u/ as early as Proto-Atayalic, but in suffixed forms it remained \*ə, and only changed in individual dialects. Examples of these verbs in Sami'u (Sm) and Macagis (Mc) are

<sup>7</sup>Note the antepenultimate vowel /a/ in Macagis /talaʔiŋ/ 'to hide' does not correspond to /u/ in Sami'u /tulʔiŋ/, nor to Proto-Atayal \*tuləqiŋ. See section 2.10 for more examples of irregular correspondences of antepenultimate vowels.

given in table 4. Just like the data in table 3, all the instances of alternating vowels are reflected as  $\emptyset$  in Sami'u and /a/ in Macagis after suffixation.

Table 4: Verbs with historical \*ə in PIngawan

Proto-Atayal AV~PV/LV	Sm/Mc AV	Sm PV/LV	Mc PV/LV	Gloss
*-caqərug~caqərəg-an	ma-caruw	carg-an	carag-an	to stand
*t<um>əʔətuʔ~təʔətəʔ-un	t<un>tuʔ/t<um>atuʔ	tatʔ-un	tataʔ-un	to chop
*l<um>əpug~ləpəg-un	l<um>puw	lapg-un	lapag-un	to count
(*g<um>əbul~gəbəl-un)	ʔ<um>bul	ʔabl-an	ʔabal-an	to bury
*ʔ<um>əbug~ʔəbəg-an	ʔ<um>buw	ʔabg-an	ʔabag-an	to soak

With the above data alone, it is not certain that this phenomenon is indeed epenthesis in Macagis, and not a change of Proto-Atayal \*ə > Macagis /a/ but Sami'u  $\emptyset$  in this environment. However, the examples in table 5 give support to the epenthesis hypothesis.

Table 5: Additional evidence of vowel epenthesis in Macagis

Macagis AV/root	Macagis	Gloss
lukus	pa-lakus-an	clothing
ma-gu.uuw	pa-ga.ɔg-an	to become lost

Macagis /pa-lakus-an/, meaning ‘clothing’ or ‘things that are worn’, is derived from the root /lukus/ ‘clothes’, which is a cross-dialectally attested Atayal lexeme. The corresponding Sami'u lexeme is /pa-lkus-an/, with vowel syncope. The same antepenultimate vowel changes to /a/ in Macagis. If we assume that the data in tables 3, 4 can be explained with a simple change of Proto-Atayal \*ə > Macagis /a/, we would expect cardinal vowels to retain their values. Instead, we see a change from /u/ in /lukus/ ‘clothes’ to /a/ in /pa-lakus-an/ ‘clothing, what is worn’ in Macagis, implying that a weakening process took place, where that vowel was either reduced to /ə/ or deleted entirely, and was then fortitioned to /a/ in Macagis. A similar phenomenon happens in the Macagis verb ‘to become lost’ (AV /ma-gu.uuw/, LV /pa-ga.ɔg-an/): the vowel following /g/ is /u/ in the strong penultimate position, but becomes /a/ after the verb is suffixed and the vowel shifts to a weaker antepenultimate position.



The reason why the vowel /a/ is chosen is likely due to it being the least marked vowel in the phonological inventory of Plngawan. The vowel system, introduced in section 1.1.2, is the same in both subdialects: /a, i, u, e, o/. Lacking a mid central vowel, Macagis opts for the low vowel /a/ in epenthesis instead.

Vowel epenthesis occurs even in loanwords. For example, Sami'uɹ /kamceʔ/, Macagis /kamaceʔ/ 'sugar, sweets' is a loan from Taiwanese Southern Min 甘蔗 *kam-tsià* 'sugarcane'. While both dialects monophthongize the diphthong in the second syllable, Macagis inserts an epenthetic /a/ to break up a heterosyllabic consonant cluster.

All the epenthetic vowels in the examples up to this point have been /a/. However, in a few rare examples, shown in table 6, there may be a different vowel breaking up two consonants. So far only three such examples have been found, and this break in the pattern may have different explanations.

Table 6: Apparently irregular epenthetic vowels in Macagis

Sami'uɹ	Macagis	Gloss
ʔamgal / ʔamugal	ʔamagal	flea
ma-stuw	ma-sutuw	to fall
ma-spel	ma-sipel	to dream

I have elicited three different forms for the word 'flea': /ʔamgal/, /ʔamagal/, and /ʔamugal/. The latter was elicited from a Sami'uɹ speaker, but could perhaps represent a form from a third dialect, which by now has lost its linguistic identity.<sup>8</sup> Cognate forms in other Atayal dialects indicate that the penultimate vowel was historically most likely \*ə: cf. Klesan /məŋin/, Matu'aw /ʔamaŋal/.<sup>9</sup> Klesan regularly merges word-final /l/ into /n/, but the vowel in the final syllable is still unexpected. Nevertheless, the penultimate vowel in Klesan is /ə/. Matu'aw changed all historical \*ə into /a/, so the penultimate /a/ may indicate either \*a or \*ə. The two dialects together suggest \*ə in the penult, meaning that the forms /ʔamagal/ and /ʔamugal/ in Plngawan are likely the result of vowel epenthesis.

<sup>8</sup>Ferrell (1969), whose wordlist is clearly identifiable as Sami'uɹ, has <ʔamugal> for 'flea'. See section 3.1 for further information.

<sup>9</sup>The vowel /i/ in the final syllable in Klesan, as well as the correspondence of /ŋ/ in Klesan and Matu'aw to Plngawan /g/ are irregular.

The remaining two verbs in table 6 both have their own idiosyncracies. Sami'uɔ /mastuw/ and Macagis /masutuw/ 'to fall' are very likely cognate with S'uli /məsəhutaw/ and Klesan /səhotaw/, which have the same meaning. However, both S'uli and Klesan appear to have an extra syllable which Plngawan lacks, so it is unclear what the penultimate vowel should be. Likewise, Sami'uɔ /ma-spel/ and Macagis /ma-sipel/ 'to dream' are cognate with Matu'uwal /s<um>apiyal/.<sup>10</sup> Interestingly, the PV forms I elicited were /sipel-un/ for Sami'uɔ and /sapel-un/ for Macagis speakers, with the vowel /i/ appearing in Sami'uɔ in PV. Plngawan is the only Atayal dialect with /i/ in the first syllable of the root in this verb, and the verb itself is highly irregular (at least in part due to dialect admixture), and does not lend itself to systematic analysis.

The vowel breaking up heterosyllabic clusters in Macagis is overwhelmingly /a/, although several roots use /u/ or /i/ instead. The apparent exceptions could be loans from related dialects where a different vowel was used for epenthesis, or sporadic changes.

## 2.2 Cluster simplification in Macagis

Consonant cluster avoidance in the Macagis variant manifests in various ways. In section 2.1, we saw how epenthetic vowels can be used to repair infelicitous clusters. Another strategy employed by Macagis is cluster simplification by means of deleting one of the consonants in the cluster. Here, I divide the data into two groups, which are discussed separately: prefixal clusters, and Cɔ-clusters.

'Prefixal' clusters are identified by a Sami'uɔ form which begins with *mak-*, *pak-*, *mas-*, or *pas-*, followed by a consonant.<sup>11</sup> Examples are shown in table 7. These are all derivational prefixes, and most of the words in the table can be analysed as having a discrete prefix and root. Some, like /makturuʔ/ 'leech' or

<sup>10</sup>Cf. also Squiliq /mə-səpiʔ/, Klesan /mə-səpi/ 'to dream'. The forms in Plngawan and Matu'uwal have a male register suffix, and are derived from female register forms in Proto-Atayal. For more on the lexical gender register in Atayal, see Li (1982, 1983).

<sup>11</sup>Note that there are no vowel-initial stems, or indeed syllables, in Plngawan Atayal. Words that are transcribed in other works as vowel-initial have a phonemic glottal stop in the beginning. These mostly behave as other roots, e.g. /mas-ʔabulit/ 'to burn to ashes' (< /ʔabulit/ 'ash'), or /pak-ʔabagan/ 'spring' (< /ʔabagan/ 'summer'). There appear to be some exceptions, like Sami'uɔ /mas-utiʔ/ 'to defecate', with glottal stop deletion (< /ʔutiʔ/ 'faeces'), or Sami'uɔ /maku-ʔaraʔ/ and Macagis /paka-ʔaraʔ/ 'to put on pants' (< /ʔaraʔ/ 'pants'), with an additional vowel in the prefix.

/pakbabuy/ ‘masked palm civet’,<sup>12</sup> should perhaps be thought of as monomorphemic under a synchronic analysis, with the prefix as a fossilized morpheme. The first consonant of the cluster is deleted in all instances in Macagis. This means that the Sami’uɔ prefixes /mak-/ and /mas-/ get merged as /ma-/ in Macagis, and likewise for Sami’uɔ /pak-/ and /pas-/ to Macagis /pa-/. The change affects the name of the place name (and its corresponding dialect) itself, which is called /makcagis/ by Sami’uɔ speakers, but /macagis/ by Macagis speakers. (Note also the irregular vowel correspondence in the initial syllable of Macagis /pubabuy/ ‘masked palm civet’.)

Table 7: Simplification of prefixal clusters in Macagis

Sami’uɔ	Macagis	Gloss
makturu?	maturu?	leech
paktamuku?	patamuku?	to wear a hat
makcagis	macagis	(place name)
maskuɔula?	makuɔula?	to molt; to shed skin
mastanux	matanux	to relieve oneself
masgali?	magali?	to tear (intr.)
pascik	pacik	to turn over
pakbabuy	pubabuy	masked palm civet

A second type of cluster simplification happens in word-medial clusters where the second consonant is /ɹ/ (Cɹ-clusters), shown in table 8. Unlike prefixal clusters, where Macagis deletes the first consonant of the cluster, in Cɹ-clusters it is always the /ɹ/ segment which is deleted.

Table 8: Simplification of Cɹ-clusters in Macagis<sup>13</sup>

PA root	Sami’uɔ	Macagis	Gloss
*qarɹəɹat	ʔarɹat	ʔarɹat	diligent

<sup>12</sup>Readers familiar with other Austronesian languages may recognize Proto-Austronesian \*babuy ‘pig’ in the Pngawan word /pakbabuy/ ‘masked palm civet’. Pngawan (or indeed any Atayal) does not use the word *babuy* to refer to pigs, but calls them /baɹok/. However, neighbouring Seediq has /babuy/ meaning ‘pig’. Whether Pngawan /pakbabuy/ is a remnant of an older Atayal form or a newer loan remains to be determined.

PA root	Sami'u	Macagis	Gloss
	gagjaʔ	gagaʔ	seeds
	malapɔw	malapow	ten
	magɔŋ	maguŋ	to snap (e.g. a twig)
*kuɔahil	k<un>ɔahil	k<um>ahil	to skin

With Cɔ-clusters, the most common repair strategy in Macagis is consonant deletion, however there is one counterexample. In table 3, Sami'u /gahjaʔ/ 'cold' corresponds to Macagis /hajaʔ/. In this instance, Macagis opted for vowel epenthesis instead. This is the only case where a Cɔ-cluster was repaired this way, and vowel epenthesis here was also coupled with the deletion of the initial syllable.

Cluster simplification may also have an impact on allomorphs. The standard Austronesian AV infix *-um-* has the allomorph *-un-* in Pngawan when followed directly by a coronal consonant. This process is seen in Macagis /k<um>ahil/ vs Sami'u /k<un>ɔahil/ 'to skin', a verb formed from the nominal root /kuɔahil/ 'skin' with the addition of the infix *-um-*. After affixation, Sami'u deletes the first vowel of the root, forming a consonant cluster, and the nasal consonant in *-um-* assimilates in place to /ɔ/, giving *-un-*. On the other hand, the Macagis form has no cluster, and thus no place assimilation, and instead deletes /ɔ/ completely (along with the root vowel /u/).

Cluster simplification will occasionally happen in environments other than prefixal clusters and Cɔ-clusters. For example, Sami'u /nabkis/ and Macagis /nakis/ 'old person'. These, however, are less regular than the two cases identified above: an identical *-bk-* cluster is instead repaired with epenthesis in Macagis /ma-bakaʔ/ 'to break (intr.)', cf. Sami'u /ma-bkaʔ/.

It is important to note that Macagis speakers do not avoid adjacent consonants entirely, and will readily produce forms with heterosyllabic consonant clusters. Some examples of forms in clusters in Macagis are given in table 9. These

<sup>13</sup>As evidenced by the Proto-Atayal reconstructions in table 8, both Sami'u and Macagis underwent vowel deletion, reducing the number of syllables in the root by one. This deletion most likely occurred in Proto-Pngawan (the immediate ancestor of Sami'u and Macagis), which led to the appearance of Cɔ-clusters that can still be seen in Sami'u. These clusters were then simplified in Macagis in a separate stage: Proto-Atayal \*VCVɔV > Proto-Pngawan \*VCɔV > Macagis \*VCV. The same process is assumed for correspondences with no Proto-Atayal reconstructions.

include morpheme boundary clusters after infixation with AV *-um-* or perfective *-in-*, but root-internal clusters can also be found: e.g. /rahkal-un/ ‘to boil water (PV)’ or /lahg-an/ ‘to sharpen (LV)’. It is possible that the modest amount of heterosyllabic clusters in Macagis is the result of sustained language contact with Sami’u, where a much wider variety of clusters is allowed.

Table 9: Heterosyllabic consonant clusters in Macagis

Macagis	Gloss
r<um>ahkal, rahkal-un	to boil water
k<un>loh, k<in>lah-un	to reap; to harvest (rice)
k<un>lakah, kilkah-an	to tread
k<un>lalabah	to weed; to hoe
s<un>liluj	to poison fish
c<um>bu?	to shoot
h<un>tu	to be blocked
l<un>haw, lahg-an	to sharpen (a knife)

Despite the occasional presence of heterosyllabic clusters in Macagis, there is a general tendency in the dialect of consonant cluster avoidance. Two main strategies—vowel epenthesis and consonant deletion—are used in Macagis to repair consonant clusters. Consonant deletion applies in more specific environments, whereas vowel epenthesis does not appear to have a specific environment and is most likely the default option. The lexical items where consonant clusters have been repaired can be identified through comparison with Sami’u, and sometimes with other Atayal dialects.

### 2.3 Liquid assimilation in Macagis

Goderich (2020:90) notes that most Atayal dialects have a historical liquid assimilation rule, where /r.l/ becomes /l.l/, but only if both liquids are syllable onsets. Plngawan is said to be one of a few dialects that did not undergo liquid assimilation. As can be seen in table 10, this is true for Sami’u speakers, whereas Macagis does in fact have /r/ assimilate to a following /l/.

Table 10: Liquid assimilation in Macagis

Proto-Atayal	Sami'uɿ	Macagis	Gloss
	riluŋ	liluŋ	poisonous vine
	raliʔ	laliʔ	the past
*raluʔ	raluʔ	laluʔ	name
*r<um>iliq	r<um>iliʔ	l<um>iliʔ	to lift (AV)

Of the words in the table, two can be traced back to Proto-Atayal: \*raluʔ ‘name’ and \*r<um>iliq ‘to lift’. Sami'uɿ /raliʔ/ (Macagis /laliʔ/) ‘the past’ is likely a cognate of Matu'aw Atayal /raraliʔ/, cf. also Squliq /raral/, with the same meaning. The lexeme /riluŋ~linuŋ/ ‘poisonous vine’ does not appear to have cognates in other Atayal dialects, but is assumed to also be a case of liquid assimilation in Macagis. This is because we have no evidence of liquid *dissimilation* in Sami'uɿ, and in fact plenty of words have two identical liquids in the same root, such as /luluŋ/ ‘river’, /ʔalileh/ ‘tick (insect)’, or /halilaw/ ‘fast, quick’, all of which are identical between Sami'uɿ and Macagis.

Like other Atayal dialects, words with /r..l/ where the /l/ is in a syllable coda do not undergo liquid assimilation in Macagis: e.g. /r<um>ahkal/ ‘to boil water’ or /r<in>ol/ ‘bud, shoot’.

Other times, the correspondences of liquids between Plngawan dialects (or between one of the dialects and Proto-Atayal) does not appear to follow any regular pattern. Some examples of these irregularities are presented in table 11, with the reconstructed Proto-Atayal form given for reference.

Table 11: Irregular liquid correspondences in Plngawan

Proto-Atayal	Sami'uɿ	Macagis	Gloss
*raɿuhiŋ	ɿaɿuhiŋ	raruhiŋ	nest fern
*lamiquɿ	lamiʔuɿ/ramiʔuɿ	lamiɿuw	Miscanthus
*r<um>əhag	r<un>haw	l<un>haw	to sharpen (AV)

In Proto-Atayal \*raɿuhiŋ ‘nest fern’, the two rhotics are distinct, however they get assimilated in different ways in Sami'uɿ and Macagis. For Proto-Atayal

\*lamiqua ‘Miscanthus’, I have elicited two different forms from Sami’ua speakers: one with the expected reflex of \*l, and one where the initial liquid unexpectedly becomes /r/. Similarly for Proto-Atayal \*r<um>əhag ‘to sharpen’, Sami’ua has the expected reflex of \*r, whereas Macagis unexpectedly changes it to /l/. In these examples, changes occur in both Sami’ua and Macagis, but there does not seem to be a conditioning environment, nor do the changes themselves appear to be regular.

A similar irregular change can also be observed in the conjunctive particle /lu/ ‘and’ (identical between Sami’ua and Macagis), which has the form /ru/ in all other dialects of Atayal, and thus must have been \*ru in Proto-Atayal.

## 2.4 A prefix in animal names

An unusual distinction between the two dialects is the choice between /ɿa-/ and /na-/ for a prefix used in the names of some small animals. There are few examples of this prefix, but it is noticeably more regular than the distinctions in section 2.11. As shown in table 12, the choice in Sami’ua is generally /ɿa-/, while Macagis speakers opt for /na-/. (The words for ‘bat’ and ‘spider’ are identical.)

Table 12: Animal name prefix in Sami’ua and Macagis

Sami’ua	Macagis	Gloss
ɿapiriy	napiriy	butterfly
ɿakarit	nakarit	bat
ɿakarit	nakarit	spider

There is reason to believe that the syllables /ɿa-/ are /na-/ form a fossilized prefix rather than part of the root, even though they cannot be used productively. The word ‘butterfly’ has cognates in many Atayal dialects: S’uli, Klesan /piray/, Squliq, Skikun /kəpiray/. Even though Squliq and Skikun Atayal do have an extra syllable in the beginning of the word, it cannot be said to be related to either /ɿa-/ or /na-/ in Plngawan. The Plngawan word for ‘bat’ is most likely a loanword from Seediq *plale karic* ‘bat’ (taking only the second half), which again has nothing resembling either of the Plngawan prefixes.

Note, however, that this is not a regular correspondence of all initial /ɿa-/ in

Sami'uɔ to initial /na-/ in Macagis. When either /ɔa-/ or /na-/ is part of the root, it is identical in both dialects, e.g. Sami'uɔ /ɔaŋlit/ and Macagis /ɔaŋalit/ 'housefly', or Sami'uɔ and Macagis /nanahiʔ/ 'wilderness'. In both of these cases, /ɔa-/ and /na-/ cannot be analyzed as prefixal.

## 2.5 Merger of -ʔuɔ and -ɔuw in Macagis

Another extremely specific correspondence between the two Plngawan varieties can be found in word-final position in a few words. As shown in table 13, Sami'uɔ word-final /-ʔuɔ/ regularly corresponds to Macagis /-ɔuw/. There are only a few examples of this correspondence, but the pattern is quite clear.

Table 13: Word-final -ʔuɔ and -ɔuw in Plngawan dialects

Sami'uɔ	Macagis	Gloss
buliʔuɔ	burɔuw	cataract; gout
lamiʔuɔ/ramiʔuɔ	lamɔuw	Miscanthus
Samiʔuɔ	Samɔuw	(place name)

This phenomenon is most likely a merger in Macagis: i.e. earlier /-ʔuɔ/ and /-ɔuw/ merged together as /-ɔuw/ in that dialect. The evidence for this is that Sami'uɔ still preserves a distinction between the two syllables, as evidenced by such examples as /ciɔuw/ 'warbling white-eye (a type of bird)' and /maguɔuw/ 'to be lost (AV)', which have /-ɔuw/ in both Sami'uɔ and Macagis. Additionally, there is external evidence from Matu'uwal /lamiquw/ and Matu'aw /lamiʔuy/ 'Miscanthus' pointing to Proto-Atayal \*lamiquɔ, of which /lamiʔuɔ/ would be the expected reflex in Plngawan.

## 2.6 Monophthongization in Sami'uɔ?

There is a correspondence between Sami'uɔ and Macagis in some lexical items, where a glide + vowel sequence in Macagis may correspond to a monophthong in Sami'uɔ. Several examples of this correspondence are given in table 14.



Table 14: Examples of monophthongization in Sami'uɹ

Proto-Atayal	Sami'uɹ	Macagis	Gloss
*ragiyax	ragex	ragyax	shin; mountaintop
*[ʔq]uciyux	ʔucix / ʔucyux	ʔucyux	fish
	taɲahen / taɲahyan	taɲahyan	nose

As is demonstrated in table 14, Sami'uɹ speakers will readily produce forms with monophthongs as well as those with a glide + vowel sequence. Conversely, Macagis speakers seem to disprefer the monophthongized forms. The fact that Macagis speakers choose only one of the variant forms may indicate that this change is specific to Sami'uɹ. However, there is very little data on this phenomenon, so it is difficult to make definitive conclusions.

## 2.7 Debuccalization in Sami'uɹ

Some Sami'uɹ speakers will occasionally debuccalize syllable-final /s/ into /h/. This debuccalization is sporadic, and does not happen all the time. It should best be viewed as an optionally applied phonological rule. The same speaker will readily produce forms with and without debuccalization. Table 15 shows some examples of forms where I encountered debuccalization. All cases of debuccalization appear in word-internal codas.

Table 15: Examples of debuccalization in Sami'uɹ

Regular	Debuccalized	Gloss
kisliʔ	kihliʔ	to like
pasbaʔ	pahbaʔ	to teach
paspun	pahpun	full
pascikun	pahcikun	to turn over
maskakiy	mahkakiy	to hike

An interesting corollary of debuccalization is that speakers will occasionally hypercorrect forms with /h/ into /s/ in the same environment where debuccalization occurs. Two examples of this phenomenon are given in table 16.

Table 16: Hypercorrection of /h/ into /s/ in Sami'uɹ

Proto-Atayal root	Regular	Hypercorrected	Gloss
*haɹutiʔ	ta-hɹutiʔ	ta-sɹutiʔ	to slip
*haŋaliq	ma-hŋaliʔ	ma-sŋaliʔ	to carry on shoulder

Here, hypercorrection occurs in prefixed forms, because the prefixes trigger vowel epenthesis which in turn puts /h/ in coda position. This hypercorrection may be indicative of a broader merger of /h/ and /s/ in this environment.

The environment where debuccalization could occur may have been broader in the past. See section 3.1.2 for examples of debuccalization in other positions found in Ferrell (1969).

## 2.8 Merger of /awa/ and /uwa/ in Sami'uɹ

Normally, PA \*-uwa- coalesced into /o/ in both Sami'uɹ and Macagis, e.g. PA \*buwax > /box/ 'rice seeds', PA \*suwagiʔ > /sogiʔ/ 'sister-in-law'. The same monophthongization process is also seen in the AV column of table 17, where the historical PA \*-uwa- sequence directly follows the AV infix <um> (the AV forms are the same between Sami'uɹ and Macagis). However, this sound change was blocked if the sequence \*-uwa- was directly preceded by PA \*q (Goderich 2020:131). This can be seen in the non-infixed forms in table 17: neither Sami'uɹ nor Macagis have monophthongization in the bare stem and suffixed forms. Goderich (2020:131) writes that Proto-Atayal \*quwa- became Pngawan /awa/ without an intervening infix, however that is only true for Sami'uɹ, as seen in table 17. The Macagis forms of these words have /uwa/, without the characteristic vowel change.

Table 17: /awa/-/uwa/ correspondence in Pngawan

PA root	Sam/Mac AV	Sami'uɹ	Macagis	Gloss
*quwalax	ʔ<um>olax	ʔawalax	ʔuwalax	rain
*quwax	ʔ<um>ox	ʔawax-an	ʔuwax-an	to wash dishes

Thus, Proto-Atayal roots beginning with \*quwa- did not normally undergo monophthongization, but inserting an infix would break up \*q and the

sequence \*-uwa-, allowing the latter to coalesce. Both roots in table 17 have monophthongization when infixed with AV <um>, but monophthongization does not occur in bare stem or suffixed forms. In cases where this coalescence was blocked, Sami'uɿ has /awa/ to Macagis /uwa/, suggesting that this is a change in Sami'uɿ.

The above is in contrast with the Proto-Atayal sequence \*-awa-, which is reflected as /awa/ in both Sami'uɿ and Macagis, as shown in table 18.

Table 18: Historical \*-awa- sequence reflexes in Pngawan

Proto-Atayal	Sami'uɿ/Macagis	Gloss
*ɿawaʔ	ɿaɿawaʔ	basket
*gawah-an	gawah-an	to open (LV)
*malawaʔ	malawaʔ “to call”	to invite (AV)

The Proto-Atayal sequence \*-uwa- when directly preceded by \*q, as well as the sequence \*-awa-, are both reflected as /awa/ in Sami'uɿ, showing a phonologically conditioned merger. On the other hand, Macagis preserves the distinction between /uwa/ and /awa/, which Sami'uɿ has lost in this environment.

## 2.9 Final liquid merger in Sami'uɿ

Some Sami'uɿ speakers may merge word-final /l/ and /ɿ/. This is likely an ongoing merger, but it is near ubiquitous in those speakers that have it. Both /l/ and /ɿ/ are pronounced identically in word-final position, as a lateral, probably with some retroflexion. (The precise phonetic details of this allophone will require further study.)

This merger creates phonological ambiguity, which can be resolved in verbal roots by eliciting a suffixed form. Table 19 presents several examples of verbs where the final liquid in the root is ambiguous (represented here as <L>), but is revealed to be /ɿ/ in suffixed forms.

Table 19: Word-final liquid allophone in Sami’ua

AV	NAV	Gloss
t<um>akuL	takuɔ-an	to push over
h<un>tuL	hataɔ-un	to block
ʔ<un>tuL	ʔatɔ-un	to ball hand into fist
ʔ<uŋ>kuL	ʔakɔ-un	to fold

This merger appears to be confined to Sami’ua speakers among the fluent elderly Plngawan population. The Macagis speakers I interviewed distinguished word-final /ɔ/ and /l/ quite reliably. (Younger speakers may struggle with liquids in general, but this is probably due to language attrition.)

This merger can be found in recent publications on Plngawan. Rau (2004) includes two comparative wordlists of Atayalic dialects, both featuring Plngawan: one containing 326 lexical items, and the other 190 lexical items. The two lists overlap, but different Plngawan speakers were consulted for each list. In the longer list, no Plngawan transcriptions whatsoever end in /l/, including words that etymologically should contain it, such as <ramagar> ‘five’, <rahar> ‘earth’, or <gamir> ‘root’. While Rau (2004) does confuse the rhotics /r/ and /ɔ/, these are only confused with the lateral /l/ when occurring word-finally. The shorter list rectifies the transcription of some /l/-final words, but not all of them, and at the same time transcribes several words with etymological word-final /ɔ/ as <l>, e.g. <lihul> ‘forehead’, <alihul> ‘wings’, and <tahal> ‘leftover’. Since confusion between rhotics and the lateral only occurred in word-final position, it is likely that Rau encountered the Sami’ua final liquid merger phenomenon.

## 2.10 Prepenultimate vowels

Plngawan is one of three Atayal dialects (alongside Matu’uwal and Matu’aw) that preserve vowel distinctions in the third-to-last syllable and beyond (Goderich 2020:102–109). This makes it valuable for Proto-Atayal reconstructions. In some cases, the three Atayal dialects do not agree on the vowel in the third-to-last syllable, including instances where the vowel in Plngawan is different from the other two dialects (Goderich 2020:108).

Occasionally, Plngawan will even disagree with itself on what the vowel should

be. Table 20 contains examples of words which have a different third-to-last or fourth-to-last vowel in Sami'u and Macagis.

Table 20: Prepenultimate vowels in Pngawan

Sami'u	Macagis	Gloss
baahu?	baahu?	Taiwan barbet
kinpuahun	kanpaahun	a kind of hornet
pakbabuy	pubabuy	Masked palm civet
hiri?-un	huri?-un	to destroy (PV)
umala?	rimala?	vegetable fern
tasasi?	tusasi?	to seek shade
saminalu?	siminalu?	to pity
ʔosaŋi?	ʔasaŋi?	rabbit

The AV form of the verb ‘to destroy’ is /h<um>iri?/ in both dialects (cf. also Squliq /h<əm>iriq/, Klesan /h<əm>iri/). The Macagis PV form having the vowel /u/ is therefore unexpected.

Sami'u /ʔosaŋi?/, Macagis /ʔasaŋi?/ ‘rabbit’ is a loanword from Japanese /usagi/. In Japanese, [ŋ] is a common allophone of /g/, so the consonantal correspondences are regular. The high back vowel in Japanese is unrounded /u/, but in Macagis it is a low vowel /a/, which is further from /u/ articulatorily than Sami'u /o/. Japanese was spoken in Taiwan mostly during the period of Japanese rule (1895–1945), so a vowel difference between the two dialects in a Japanese loanword is telling.

It is important to note that Sami'u and Macagis agree on vowel quality in the vast majority of words. However, words with different vowels do exist, and these differences are largely confined to antepenultimate syllables (barring other, regular correspondences such as vowel epenthesis in section 2.1). No obvious pattern emerges from the data, and it is likely that these distinctions are sporadic. They may even indicate a possible breakdown of prepenultimate vocalic distinctions altogether, much like what has already happened in Seediq and most Atayal dialects.<sup>14</sup> As of today, Pngawan appears to be unique among Atayal dialects in how it treats vowels outside the final foot. While it did not

<sup>14</sup>Seediq does not distinguish vowel phonemes outside the final two syllables (Tsukida 2005:292). Of the Atayal dialects, only Pngawan, Matu'uwal, and Matu'aw distinguish vow-

fully neutralize prepenultimate vowel distinctions (like Squliq, S’uli, Klesan, and Skikun), it is sprinkled with occasional sporadic changes and irregular correspondences.

## 2.11 Sporadic segment differences

In addition to the various regular correspondences described above, some words that are clearly cognates between Sami’ua and Macagis have sporadic irregular distinctions in their vowels, consonants, or both.

A small number of words have different vowels in one or both of the final two syllables. All of the examples from my data are presented in table 21. These are fewer in number than words with irregular prepenultimate vowels. Part of the reason for this is likely the fact that the final foot in an Atayal word is a strong position, and cardinal vowels are never reduced in this environment in Plngawan, or any other Atayal dialect.

Table 21: Sporadic vocalic differences in Plngawan dialects

Sami’ua	Macagis	Gloss
keweh/kewah	kaweh	sickle
sebeh	subeh	close; near
ʔakeh	ʔakah	bad; dislike
mubabox	mabubox	naked

Of the four words in table 21, two have Proto-Atayal reconstructions in Goderich (2020): \*sawbih ‘close, near’ and \*ʔaqih ‘bad, dislike’. Proto-Atayal did not have a phonemic mid vowel /e/, and this vowel in the final syllable of /sebeh, subeh/ is the result of centralization adjacent to Proto-Atayal \*h, whose phonetic value was [h] (however, it is no longer pharyngeal in Plngawan). The expected reflex of the Proto-Atayal sequence \*aw in Plngawan is /o/, which neither dialect has in the penult. However, Macagis /u/ is less surprising than Sami’ua /e/.

els outside the final foot. For example Proto-Atayalic \*bunaqig ‘sand’ > Toda Seediq /bənaqi?/, Truku Seediq /bənaʔig/, Squliq Atayal /bənaqiy/, but Matu’uwal Atayal /bunaqiy/, Plngawan /bunaʔiy/ (from Li 1981:291).

For Proto-Atayal \*raqih ‘bad, dislike’, Macagis /ɹakah/ has an irregular vowel in the final syllable, whereas in Sami’uɹ /ɹakeh/ the vowels are regular. However, the expected reflex of Proto-Atayal \*q in Plngawan in /ʔ/, so the Plngawan lexemes may not even be direct descendants.

Irregular consonantal correspondences can appear in any position: initial, medial, or final. Although there do not seem to be any patterns or constraints to their distribution, the data so far includes a rather large number of cases including either glottal segments (/h/ and /ʔ/) or glides (/w/, /y/, and /ɹ/). Examples of these irregular correspondences can be seen in table 22.

Table 22: Sporadic consonantal differences in Plngawan dialects

Sami’uɹ	Macagis	Gloss
ɹibul	libuɹ	skirt
maɹusun/maɹlusun	mamusun	horsefly
teʔuɹ	teɹuɹ	crow
nawiluɹ	naɹiluɹ	wire
ʔumakas	humakas	greedy
ɹakuʔ-an	ɹakuc-an	to scoop up (LV)
haɹeʔ-an	huɹeg-an	to pour (LV)
wasiluɹ	ʔasiluw	thin
masuruɹ	masuruʔ	overcast, cloudy
sunɹaʔiɹ	sunyaʔiɹ	to envy
ʔunɹaʔil	sunɹaʔil	to bully

These irregular correspondences affect verbal as well as nominal roots. Interestingly, in verbal roots, they appear to affect not the whole paradigm, but only one or a few forms, meaning that a regular verb in one dialect corresponds to an irregular one in the other.

Sami’uɹ /ɹakuʔ-an/ and Macagis /ɹakuc-an/ ‘to scoop up (LV)’ both have /pa-ɹakut/ as their AV form, derived from the noun /ɹakut/ ‘net’. Here, Macagis /c/ in the suffixed form is not unexpected, since /c/ and /t/ merge in coda position, however Sami’uɹ /ʔ/ is highly irregular.

Sami’uɹ /haɹeʔ-an/ and Macagis /huɹeg-an/ ‘to pour (LV)’ both have the AV form /h<un>ɹiʔ/. The reconstructed Proto-Atayal form of this verb is

\*h<um>əiʔ in AV and \*həiʔ-an in LV, thus making the Macagis LV form irregular. (The vowels in the initial syllable in LV also do not match, however in this case they descend from Proto-Atayal \*ə, for which we do not yet have a complete picture in Plngawan.)

Some lexical items have more than two variant forms, for example /maŋusun/, /maŋlusun/, and /mamusun/ for ‘horsefly’. A more extreme example of this is ‘canine; fang’, for which I have recorded the forms /ʔawaʔit/, /wawaʔic/, /yawaʔic/, and /ʔawaʔit/.<sup>15</sup>

We may have expected reflexes for some lexical items, based either on comparative data or on verbal paradigms. Unexpected forms appear not just in one of the dialects, but in either, for various forms, without any consistent pattern. This means that for cases where the dialects have slightly different reflexes of the same word that are not explainable by regular sound changes, we cannot predict the ‘original’ form of a word based solely on data from Sami’ua and Macagis, since both these dialects are amenable to sporadic changes.

The origin of these sporadic changes cannot be determined with certainty. They are somewhat reminiscent of the gender register system in Atayal (see Li 1982, 1983). However, the gender register system employed a variety of derivational strategies, including infixation, suffixation, segment substitution, segment deletion, and even suppletion. Of these, infixation and suffixation were the most productive. Unlike the register system, we only see segment substitution (and perhaps deletion, if initial glottal stops are counted as epenthetic) in these sporadic sound changes in Plngawan, meaning that it is probably an unrelated phenomenon.

### 3 Dialects in older linguistic works

Despite the appearance of Plngawan in numerous publications, none of them mention subdialectal distinctions except Chen (2012). However, when looking back at the various works with significant amounts of Plngawan data, we can use the diagnostics from section 2 to quite reliably identify the Plngawan variant used in each paper. As it turns out, the representation of Sami’ua has been

<sup>15</sup>The expected reflex of Proto-Atayal \*waqit ‘fang’ would be an unattested form \*\*waʔit. All the Plngawan forms have a prefix, and /ʔawaʔit/ additionally has an irregular consonant correspondence.



much greater than that of Macagis. Sometimes, both Sami'uɔ and Macagis lexemes will appear within the same publication, which may unnecessarily complicate matters for the unaware researcher.

A list of works in which Plngawan has made a significant appearance is given in section 1.2. Of these, Ferrell (1969) and Li (1981, 1985) are some of the earliest, and include a large amount of Plngawan data. These works are examined in detail in section 3.1 and section 3.2, respectively. The result of this examination is that data in both researchers' papers points very clearly to Sami'uɔ, despite a small amount of variation and dialect admixture.

### **3.1 Ferrell 1969**

Ferrell (1969) is the first ever published wordlist of Plngawan vocabulary. Unlike Li, who was extremely familiar with various Atayal dialects and thus more prone to writing phonological representations based on cross-dialectal comparisons, Ferrell's data is transcribed phonetically (however imperfectly at times), which gives us a glimpse at the phonological and phonetic processes that have been happening in Plngawan at the time.

Much of Ferrell's data, taken over 60 years ago now, has the same tendencies that can be observed in Plngawan today. On the one hand, this is not particularly surprising, since 60 years is a very short time for a language to evolve. On the other hand, we can assume that Ferrell's language consultants were not influenced by Sinitic languages to the same extent that Formosan speakers are at present. This means that any sound changes found in Ferrell's data cannot be put down to language attrition.

The sound changes found in Ferrell (1969) include debuccalization of /s/, confusion of antepenultimate vowels, monophthongization of /-yu-/ after consonants, and the merger of /uwa/ and /awa/ sequences. Most of these changes point specifically to a Sami'uɔ dialect speaker.

Ferrell marks stress on some Plngawan words, sometimes penultimate and other times ultimate (and in some cases has no stress marking at all). This is somewhat unexpected since he also states that stress does not appear to be phonemic in Atayalic (Ferrell 1969:76). More recently, Atayal has been reported to always have word-final stress in all dialects (Goderich 2020:31). Ferrell's transcription was rather narrow, but his stress marking does not

appear to correspond to any phonological phenomena found in PIngawan, and can be safely disregarded.

### 3.1.1 Merger of /awa/ and /uwa/

In at least one word in Ferrell’s data, we see the Sami’uḷ-specific merger of /uwa/ into /awa/: <ʔawaláh> ‘rain’, corresponding to Sami’uḷ /ʔawalax/ and Macagis /ʔuwalax/ in my own field notes. As noted in section 2.8, this is a merger in Sami’uḷ, because Macagis still distinguishes the two sequences.

### 3.1.2 Debuccalization

Ferrell recorded debuccalization in a number of words in his data, some of which are presented in table 23. As mentioned in section 2.7, debuccalization in Sami’uḷ is optional, and the same lexical item is often produced by the same speaker with and without debuccalization at different times.

Table 23: Debuccalization in Ferrell (1969)

Ferrell	My data	Gloss
mahkaki	maskakiy	walk
hinúnuh	sinunux	hair
hi:níu	sinyuw	rope/cord
hibih	sebih	near

What is unusual about Ferrell’s data, is that debuccalization could apply not only in coda position as today, but even in onset position. One of the reasons why debuccalization appears to have receded may be contact with other PIngawan varieties, whose speakers do not have this phenomenon.

Not all of Ferrell’s data undergoes debuccalization. Counterexamples include <paskišili?> ‘breathe’ and <paspun> ‘full’, where /s/ is retained despite being in coda position. This suggests that, just like today, debuccalization was optional at the time Ferrell conducted his fieldwork.

### 3.1.3 Monophthongization

Ferrell recorded at least one instance of Sami'uᐱ-*specific monophthongization* in the word <ʔucih> ‘fish’ (cf. Macagis /ʔucyux/). However, ‘nose’ is still given as <taḡahi'an>, compared to a possible monophthongized Sami'uᐱ variant /taḡahen/.

It was noted in section 2.6 that Sami'uᐱ speakers may produce variants of the same lexical item with or without monophthongization (in contrast to Macagis speakers, who tend to eschew it in the same words). It is possible that Ferrell's language consultants chose a non-monophthongized variant sporadically, or that it was simply the only one they regularly used.

### 3.1.4 Syllable structure

One of the most salient distinctions between Sami'uᐱ and Macagis is the difference between their syllable structures. This tendency to avoid heterosyllabic clusters by inserting the vowel /a/ is discussed in section 2.1. In Ferrell (1969), we see a number of lexical items with consonant clusters that would be repaired in Macagis. Some examples of these words are given in table 24.

Table 24: Heterosyllabic consonant clusters in Ferrell (1969)

Pᐅḡawan	Gloss
cackaʔ	half
malahḡan	night
paspun	full

In addition to words transcribed with clusters, <ʔapənúx> ‘tooth’ was written by Ferrell with a schwa. The vowel [ə] is not a phoneme in Pᐅḡawan, and in fact does not appear in surface representations at all. Ferrell likely heard a vocalic segment in a carefully enunciated cluster and transcribed it as such. In a case of true vowel epenthesis, we would expect to see a cardinal vowel in this position.

### 3.1.5 Antepenultimate vowels

Section 2.10 describes how Sami'uᐱ and Macagis may sometimes disagree on vowel quality in the third-to-last or fourth-to-last syllable. Occasionally, the

two variants will agree with each other, but not with other Atayal dialects. Ferrell's data suggests this is not a new phenomenon. Table 25 gives examples of irregular antepenultimate vowels in Ferrell's Plngawan transcriptions, with my own Matu'aw Atayal data given for comparison.

Table 25: Irregular vowels in Ferrell (1969)

Plngawan	Matu'aw	Gloss
mulíkul/malíkul	mamalikuy	man
sinbaɿaŋan	sinbuyaŋan	spear
ramagal	yimagal	five

The word 'man' is reconstructed to Proto-Atayal as \*malikuɿ (cf. Matu'uwal /mamalikuw/, Matu'aw /mamalikuy/). My own field notes also have the vowel /a/ in the first syllable of this word in Plngawan: /malikuɿ/. Ferrell actually recorded the same lexical item twice: once on p. 174 as <mulíkul> 'man', and again on p. 185 as <malíkul> 'young man'.

Ferrell transcribes <sinbaɿaŋan> for 'spear', with the vowel /a/ in the antepenult. I have elicited Plngawan /sinbuɿaŋan/ for the same lexical item from a Sami'ua speaker, which is also supported externally by Matu'aw /sinbuyaŋan/.

The last item on the list is <ramagal> 'five'. Here, my own transcriptions agree with Ferrell's on the vowel in the first syllable: /ɿamagal/.<sup>16</sup> However, this vowel is unexpected: cf. Matu'uwal /ʔimagal/, Matu'aw /yimagal/, both of which suggest Proto-Atayal \*ɿimagal, ultimately from Proto-Austronesian \*lima (cf. also Plngawan /maɿimal/ 'fifty', in which the vowel /i/ is retained). The sporadic change in this lexical item had already been completed by the time Ferrell conducted his fieldwork.

These examples demonstrate that sporadic changes in antepenultimate vowels in Plngawan started before any significant language attrition could have taken place. They may indicate a general weakening of vocalic distinctions outside the final foot, which is characteristic of many Atayal dialects and also Seediq, with whom Plngawan speakers are neighbours.

<sup>16</sup>However, there appears to be an error in Ferrell's transcription of the initial consonant.

### 3.1.6 Vowels in the final foot

While irregularities in antepenultimate vowels are somewhat common in Pngawan, vowels in the final foot generally tend to be more regular, both within the main varieties of Pngawan and when compared to other Atayal dialects. Occasional exceptions still arise, as noted in section 2.11. I have been able to find just two examples in Ferrell's data, which are presented in table 26.

Table 26: Irregular final foot vowels in Ferrell (1969)

Pngawan	Gloss
ʔamugal	flea
ɿakeh	bad

As noted in section 2.1, the Macagis form for 'flea' is /amagal/. I have collected both /ʔamgal/ and /ʔamugal/ from Sami'u speakers. Even though /ʔamugal/ appears to have an epenthetic vowel, it is not the one generally used in Macagis. Thus Ferrell's transcription can be considered yet another indicator of Sami'u.

The word 'bad' has two possible forms in Pngawan: /ɿakeh/ and /ɿakah/ (see also section 2.11). Macagis speakers prefer the latter, while Sami'u speakers are more likely to choose the former.

### 3.1.7 Law of liquids

In section 2.3, evidence was presented for liquid assimilation in the Macagis variety. Ferrell's data, shown in table 27, does not have liquid assimilation.

Table 27: No liquid assimilation in Ferrell (1969)

Pngawan	Gloss
ráluʔ	'name'
raliʔ	'former'

The lack of liquid assimilation by itself would be insufficient to identify the variety as Sami'u, since we do not know when liquid assimilation in Macagis took place. However, it is still useful as corroborating evidence alongside the other features listed here.

### 3.1.8 Differences from modern speech

Most of the distinctive features found in Ferrell's data can still be observed today in Sami'ua speech. However, in a couple of instances the language patterns appear to have undergone slight changes.

First, as noted in section 3.1.2, Ferrell recorded instances of debuccalization in onset position in a number of words, however this has not been reported for Pngawan anywhere else. In my own fieldwork, I have only recorded debuccalization in syllable codas. Since in both Ferrell's and my own data the debuccalization is an optional process, it is not irreversible. This is what appears to have happened in Sami'ua: an innovation that spread debuccalization to all positions was reverted, possibly due to pressure from Macagis.

An important phenomenon that can be observed today but is not found in Ferrell's transcriptions is the Sami'ua final liquid merger, whereby word-final /l/ and /ɺ/ are pronounced identically (see section 2.9). While Ferrell did struggle with /r/ and /ɺ/, his transcriptions of word-final liquid categories match our etymological expectations (i.e. /ɺ/ was sometimes confused for /r/ but never for /l/). The only exception appears in <mulíkul> 'man' and <malíkul> 'young man', which, as noted in section 3.1.5, is actually a single lexical item transcribed twice with different antepenultimate vowels. Because Ferrell's data follows Sami'ua features so closely, the lack of a general word-final liquid merger is important. Based on this data, we may hypothesize that the Sami'ua liquid merger is a very recent phenomenon that was not present in older Sami'ua speakers at the time Ferrell was conducting his research.

Another difference between Ferrell's transcriptions and modern Pngawan data lies in conjunctive markers. Ferrell (1969:396) has <ruʔ> and <cukáʔ> for 'and', when used with non-human and human subjects, respectively. These particles are formally different from today's Pngawan: the non-human conjunctive marker is /luʔ/, with a lateral approximant, and in the human conjunctive marker /cikaʔ/ the vowel in the first syllable is /i/ (Huang 2006). Other Atayal dialects all have /ruʔ/ for non-human conjunction, so Pngawan /luʔ/ appears to be a very recent innovation, especially considering Ferrell's transcriptions.

## 3.2 Li 1981 and 1985

Paul Li has done extensive research on various Atayal dialects. Pngawan features extensively in two of his publications: Li (1981), which is a reconstruction of Proto-Atayal phonology, and Li (1985), which classifies Pngawan as an Atayal dialect on the basis of lexical evidence. In section 3.2.1, we take a look at the lexical items in Li's data which can be reliably classified as Sami'ua or Macagis. In section 3.2.2, we see the ways in which other evidence may have been obscured by Li's transcriptions.

### 3.2.1 Pngawan variants in Li's publications

Both papers include a large amount of Pngawan vocabulary, and some lexical items can be identified as Sami'ua. Table 28 presents lexical items with Sami'ua features from Li's 1981 paper. All of the words in this table have heterosyllabic consonant clusters, which in Macagis are broken up with epenthetic vowels.

Table 28: Sami'ua lexemes in Li (1981)

Pngawan	Gloss
miŋkuŋ	dark
maspel	dream
matŋiʔ	full
lalbuʔ	low/short
cackaʔ	middle
tagliʔ	waterfall

Li (1981) still includes a couple of Macagis words. One of these is <ɲaŋalic> 'fly',<sup>17</sup> with the telltale epenthetic /a/ (cf. Sami'ua /ɲaŋlit/, Matu'uwal /ʔaŋriʔ/, Squliq /zəŋəliʔ/). Another potential Macagis word is <lumuhuw> 'to thread a needle (AV)'. There is a (possibly Sami'ua) variant /lunhuw/, but cross-dialectal evidence for the word-medial vowel is mixed: cf. Matu'uwal /lumhug/, but Squliq /ləmuhuw/, Skikun /ləmuhux/, Matu'aw /lumuhuw/. In this case, the vowel quality and mixed evidence from other dialects do

<sup>17</sup>There is no distinction between /t/ and /c/ in Pngawan in word-final position. Some speakers pronounce it as a plosive, while others use an affricate some or all of the time. Throughout this paper, I transcribe my own data with /t/, to prevent confusion.

not constitute sufficient evidence to place this word within either Sami'uɔ or Macagis.

In Li (1985), there is a more varied selection of Sami'uɔ lexemes, shown in table 29. Here, we see the *-uwa-* > *-awa-* merger in <ʔawalax> 'rain', Cɔ-cluster retention in <gahɔaʔ> 'cold', and prefixal clusters in the remaining three lexical items. There is also debuccalization in <paskihliʔ> 'to breathe', discussed in section 3.2.2.

Table 29: Sami'uɔ lexemes in Li (1985)

Li	Gloss
ʔawalax	rain
gahɔaʔ	cold
makturuʔ	leech
malhituɣ	moccasin snake
paskihliʔ	to breathe

Li (1985) also included a couple of Macagis items: <nakis> 'old', with a simplified cluster (cf. Sami'uɔ /nabkis/), and <laluʔ> 'name', with liquid assimilation (Sami'uɔ /raluʔ/). Thus Li's 1981 and 1985 publications appear to be mostly Sami'uɔ, but a small amount of Macagis data managed to creep in.

### 3.2.2 Li's transcriptions

An important caveat when examining Li's Pngawan transcriptions is his familiarity with other Atayal dialects. Li's interest in Pngawan was primarily from a comparative perspective, which appears to have influenced the way he transcribed it.

For example, while both Li (1981) and Li (1985) appear to be based primarily on Sami'uɔ data, we do not see any debuccalization. Compare that with Ferrell (1969), where debuccalization featured prominently. Since Li was comparing Pngawan to other Atayal dialects, he was able to puzzle out the words in which debuccalization occurred and use a more abstract underlying representation with /s/. We can deduce this based on his transcription of <paskihliʔ> 'to breathe' (Li 1985:711), which Li did not recognize as an Atayal cognate at the time, and transcribed with <h> in the root. This root features in other



Atayal dialects with the meaning ‘to like’ or ‘mood’: cf. *Matu’uwal* /kɪsɪq/, *Squliq* /qəsəliq/, *Klesan* /kəsəli/. Had Li made that connection, he may have transcribed /s/ in this word as well, which, while useful for cross-dialectal comparisons, would have completely obscured the debuccalization process that he undoubtedly encountered.

In a similar fashion, we do not see a merger of word-final /ɹ/ and /l/ in Li’s data. Li was intimately familiar with Atayal and Seediq, and had done extensive comparative research, so he was able to accurately resolve any ambiguities resulting from a word-final liquid merger in Pngawan. We can infer that he still encountered the merger based on his transcription of the word <ʔamoɹ> ‘what’ (Li 1985:712), which does not have any known cognates. Being unable to use comparative data, Li transcribed what he heard: a merged liquid, which he decided to categorize as /ɹ/. Compare that with Ferrell’s <ʔaməl> ‘what’.<sup>18</sup> We know that Ferrell’s transcription was narrow (more so than Li’s), and that he distinguished liquids word-finally (see section 3.1.8). It is possible that other word-final liquids in Li’s data merged together but were transcribed using their underlying or historical values using internal and external comparisons.

In short, Li’s transcriptions are broad and abstracted to a high level, using various methods to determine the underlying representation. This is useful when making cross-dialectal comparisons or conducting historical linguistic research, like Li was doing. However, it may also obscure synchronic dialectal variation or even ongoing sound changes.

## 4 Conclusion

Being able to distinguish between Sami’uɹ and Macagis data is especially helpful when making phonological analyses (both synchronic and diachronic), since many apparent irregularities in the data can be explained as regular correspondences between the two varieties. This is doubly important since the various Sami’uɹ and Macagis diagnostics identified in this paper may not necessarily come from different speakers. The same speaker may use lexical items from both varieties, and even produce both Sami’uɹ and Macagis versions of the same word.

---

<sup>18</sup>Ferrell (1969) sometimes transcribed Pngawan /o/ as <ə>. His reasons for doing so are unclear.

An examination of past publications on Plngawan Atayal has revealed a clear prevalence of Sami'ua data. The relative scarcity of Macagis data in linguistic papers needs to be addressed in future research. Any future publications on Plngawan should ideally utilize data from both varieties, or at the very least acknowledge their existence and specify which variety they are examining.

## References

- Chang, Yiyang Ann. 2012. Pronominal ordering in Plngawan Atayal. *Proceedings of ConSOLE XIX*, 177–190. Groningen, Netherlands: University of Groningen.
- Chen, Joyce Yi-Jie. 2012. *Affixation Induced Phonological Variations in Plngawan Atayal*. MA thesis, National Tsing Hua University, Hsinchu.
- Ferrell, Raleigh. 1969. *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification* (Institute of Ethnology, Academia Sinica, Monograph), vol. 17. Taipei: Academia Sinica.
- Goderich, Andre. 2020. *Atayal Phonology, Reconstruction, and Subgrouping*. PhD dissertation, National Tsing Hua University, Hsinchu.
- Huang, Lillian M. 2006. Case marking system in Plngawan Atayal. *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-Kuei Li on His 70th Birthday* (Language and Linguistics Monograph Series W-5), ed. by Henry Yungli Chang, Lillian M. Huang, and Dah-an Ho, 205–238. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei. 1980. The phonological rules of Atayal dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 51.2:349–405.
- Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52:235–301.
- 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* 53:265–304.
- Li, Paul Jen-kuei. 1983. Types of lexical derivation of men's speech in Mayrinax. *Bulletin of the Institute of History and Philology, Academia Sinica* 54.3:1–18.
- Li, Paul Jen-kuei. 1985. Linguistic criteria for classifying the Atayalic dialect groups. *Bulletin of the Institute of History and Philology, Academia Sinica* 56.4:699–718.
- Rau, Der-Hwa Victoria. 2004. Lexical similarity, sound change and intelligibility of Atayalic dialects. *Papers on Austronesian Subgrouping and Dialectology* (Pacific Linguistics), ed. by John Bowden and Nikolaus Himmelmann, 37–96. Canberra: Australian National University.

- Shih, Cindy Peiru. 2008. *Interrogative Constructions in Pngawan Atayal*. MA thesis, National Taiwan Normal University, Taipei.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology* (Study of Languages & Cultures of Asia and Africa Monograph Series 5). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Tsukida, Naomi. 2005. Seediq. *The Austronesian Languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 291–325. Abingdon: Routledge.
- Utsurikawa, Nenezō, Tōichi Mabuchi, and Nobuto Miyamoto. 1935. *The Formosan Native Tribes: A Genealogical and Classificatory Study*, vol. 1. Taihoku: Institute of Ethnology, Taihoku Imperial University.