

Formosan loanwords in four Atayal dialects

Andre Goderich 郭育賢

Abstract

Lexical borrowings in Formosan languages from Japanese and Sinitic languages are frequently discussed in linguistic literature. Borrowings between Formosan languages themselves are more difficult to disentangle. This paper presents loanwords from Formosan languages in four Atayal dialects: Matu'uwal, Matu'aw, Plngawan, and Klesan. All four dialects were found to have been in contact with their immediate neighbours. The donor languages range from distantly related Pazih and Saisiyat (for Matu'uwal), to closely related Seediq (for Plngawan), to other Atayal dialects (for Matu'aw and Klesan). Knowledge of lexical borrowings is useful when reconstructing protolanguages, and also helps understand the cultural history of the Atayal people, and their relationships with their neighbours.

Keywords: Formosan, Atayal, language contact, historical linguistics

語言學文獻時常會提及台灣南島語中的日語或漢語系借詞。然而，台灣南島語之間的借詞難以發現。本篇論文討論四個泰雅語方言之中的台灣南島語借詞，汶水泰雅語、大興泰雅語、萬大泰雅語、宜蘭澤教利泰雅語。此四方言皆與相鄰的台灣南島語有過語言接觸關係。借詞的來源為：關係較遙遠的巴宰語和賽夏語（汶水泰雅）、關係較相近的賽德克語（萬大泰雅）、或其他泰雅語方言（大興泰雅語和宜蘭澤教利）。借詞的研究有助於語言重建，亦可進一步了解泰雅族的文化史及其與他族的接觸史。

關鍵字：台灣南島語、泰雅語、語言接觸、歷史語言學

1 Introduction

Atayal is an Austronesian language spoken in the mountains of northern and north-central Taiwan. There are at least seven dialects of Atayal: Squiliq,

Klesan, S'uli, Matu'aw, Matu'uwal, Plngawan, and Skikun (Goderich 2020). Of these, Squliq is by far the biggest, both in terms of geographical area and the number of speakers, and can be called the prestige dialect of Atayal.

Some minority Atayal dialects are spoken at the fringes of Atayal territory, and close to speakers of other languages: Matu'uwal and Matu'aw in the west, Plngawan in the south, and Klesan in the east. Additionally, the prestige dialect of Atayal—Squliq—has by far the most speakers and the largest geographic distribution, and has influenced other, smaller Atayal dialects.

More recently, the Atayal dialects have come into contact with Sinitic languages (Hokkien, Hakka, and later Mandarin) and Japanese. The phonology of Japanese, and especially that of Sinitic languages, differs significantly from that of Atayal. This makes loanwords from these languages easy to identify, and apparent to native speakers. On the other hand, borrowings from Formosan languages tend to be more opaque, as they generally fit the same phonotactic structure as native vocabulary, and may have been loaned a comparatively long time ago, making them more entrenched in the language. This is also true of loans from other Atayal dialects, where both loaned and inherited words tend to be phonetically similar, and the only way to distinguish the two is by using the comparative method. Although there have been studies of Japanese borrowings into Atayal in the past (e.g. W.K. Wu 2014), there has been no comprehensive research of Formosan loanwords in the language.

The goal of this paper is to identify loans from other Formosan languages and Squliq Atayal in four Atayal dialects: Matu'uwal, Matu'aw, Plngawan, and Klesan. The four dialects in this study were chosen due to their position on the Atayal periphery and history of contact with other Formosan linguistic communities. All four have a number of lexical borrowings from their neighbours, with loanwords in Matu'aw, Plngawan, and Klesan coming from Matu'uwal, Seediq, and Squliq, respectively. Since all the aforementioned dialects belong to the Atayalic subgroup, most borrowings are well-integrated into the phonological systems of the recipient languages. At the same time, their status as loans is more certain thanks to reconstructions at the Proto-Atayalic and Proto-Atayal levels. Loanwords in Matu'uwal originate in Pazih and Saisiyat, which are distant relatives of Atayal, and are more tentative.

Loanwords in this study were identified primarily in two ways: (1) irregular sound correspondences, and (2) uniquely shared lexical items between geo-

graphically adjacent dialects that do not belong to a subgroup. Both of these methods are dependent upon a linguistically sound reconstruction of the proto-language and dialectal subgrouping. The main source for these is Goderich (2020), with additional data from Li (1981).

Since all the languages in this study belong to the Austronesian family, and many are much more closely related, a robust methodology is required to distinguish loanwords from inherited vocabulary. This preliminary groundwork is laid in the first half of the paper, before the list of lexical borrowings is presented. Once regular sound correspondences for all dialects/languages have been established, irregular correspondences become apparent.

I provide sources of data, basic information on the four dialects in the study, and the methodology used to find loanwords in each in section 2. Section 3 introduces the phoneme inventories of the Atayal dialects and donor languages, and provides an overview of regular sound correspondences between Atayalic dialects. Loanwords are provided in an alphabetized list in section 4, separately for each dialect:¹ Seediq loanwords in Plngawan (27 words), Squliq loanwords in Klesan (9 words), Pazih and Saisiyat loanwords in Matu'uwal (22 words), Matu'uwal loanwords in Matu'aw (19 words).

2 Methodology

The four dialects discussed in this paper—Plngawan, Klesan, Matu'uwal, and Matu'aw—are all situated on the periphery of Atayal-speaking territory, as seen on the map in figure 1. In fact, both Plngawan and Klesan are Atayal exclaves, and are completely surrounded by speakers of other languages. Matu'uwal and Matu'aw border each other, but although there are speakers of Squliq and S'uli Atayal in the vicinity, Matu'uwal does not show evidence of close historical contact with either. Likewise, Squliq does not appear to have had any impact on Matu'aw.²

A large cross-dialectal comparative wordlist was used to identify loanwords. The wordlist was compiled using first-hand field data, collected from 2013 to

¹Some Pazih and Saisiyat loans in Matu'uwal Atayal have been reloaned by Matu'aw Atayal. These are listed in the Matu'uwal loanwords section (section 4.3).

²Since Matu'aw and S'uli are very closely related (Goderich 2020), their similarities owe to their genetic proximity.

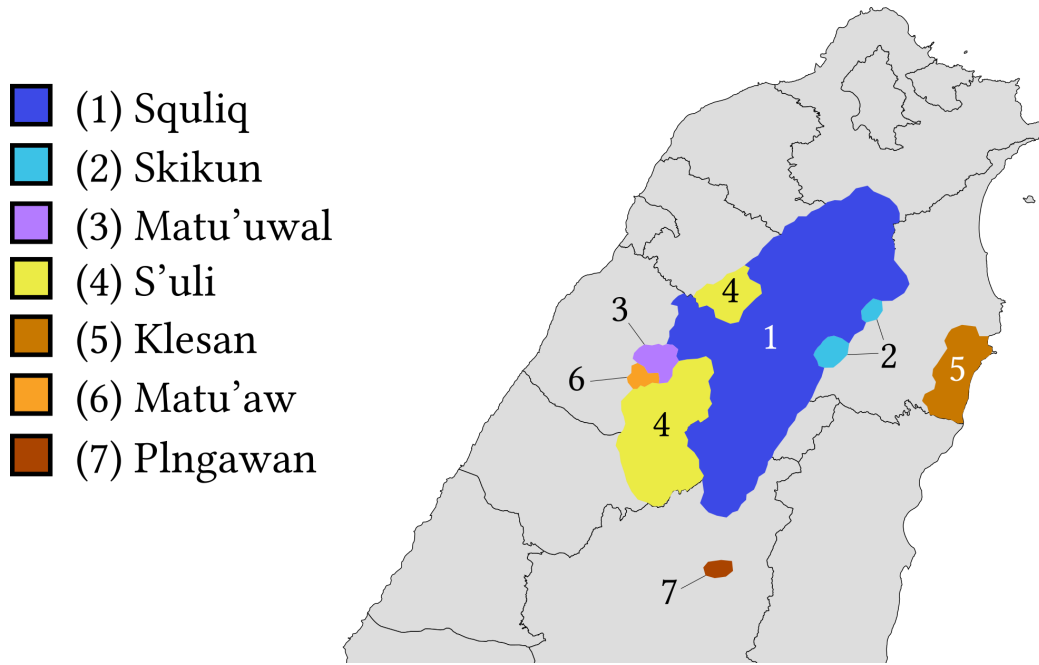


Figure 1: The distribution of Atayal dialects (source: Goderich 2020)

2023. Some basic information on the number of speakers from each village, their gender distribution, and age ranges at the time of fieldwork are provided in table 1. Additional Squliq data was supplemented from several sources: the ILRDF online dictionary³, Egerod (1980), Hetay Payan (2018). Proto-Atayal reconstructions are from Goderich (2020).

Table 1: Language consultants in this study

Dialect	Village	Num. of speakers	Age
Plngawan	Chin-ai (親愛)	4 (3f, 1m)	60-70
Klesan	Pyahaw (碧候)	2 (1f, 1m)	60-65
Matu'awal	Jin-shuei (錦水)	5 (2f, 3m)	70-85
Matu'aw	Maymaralas (南灣)	2 (1f, 1m)	80-85

³Indigenous Languages Research and Development Foundation 原住民族語言研究發展基金會, URL: <https://e-dictionary.ilrdf.org.tw/index.htm> (accessed on 31 July 2023).

This study uses the Atayal dialect classification by Goderich (2020), which divides Atayal into two main groups: Northern and Southern, with further subdivisions in each. The full subgrouping is shown in figure 2.

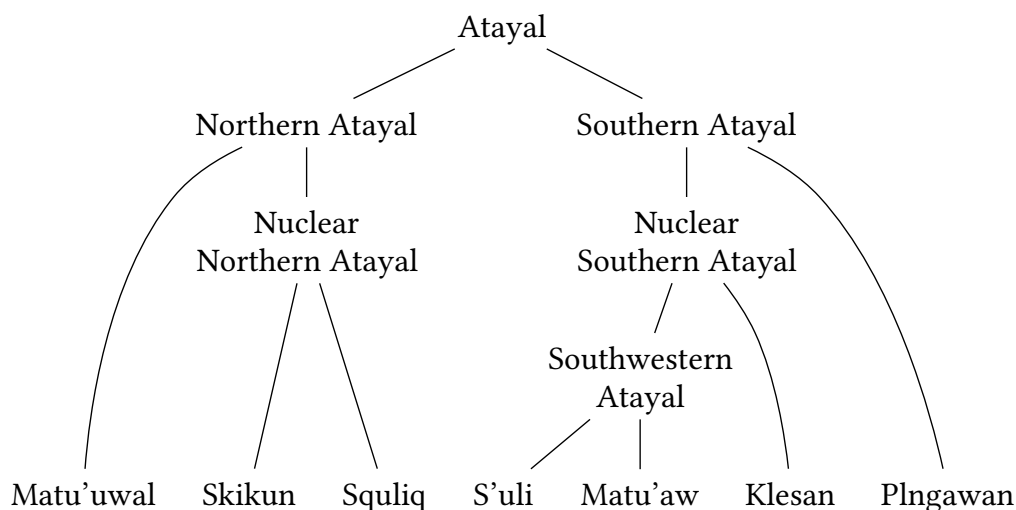


Figure 2: Atayal dialect subgrouping (source: Goderich 2020)

Atayal is famous for its lexical gender register system, which is still preserved by some Matu'uwal speakers but has collapsed in all other Atayal dialects (Li 1982a, 1983). Within the gender register system, men and women use different words for the same concepts. Male register forms are derived from female register (inherited) forms using several different strategies, including suffixation, infixation, segment substitution, segment deletion, etc (Li 1983). The derivations are not predictable, and in fact different dialects will often have different male register forms derived from the same root, which suggests that the system was productive even after the split of Proto-Atayal (Goderich 2020:156–161). The register system features heavily in Atayal protoform reconstructions and dialect comparisons, and has even had an impact on at least one loanword: Matu'uwal *situwinj* 'clothes' is the male register form of *siyatu?*, a Pazih loan (see section 4.3). Within other dialects, cognacy may be obscured by the gender register; I point this out on a case-by-case basis within the paper.

Note that in this paper, the weak vowel [ə] is written out in all instances. Many studies on Atayal choose to leave it out, since in some dialects its placement is

predictable.⁴ However, since syllabic structure varies from dialect to dialect, I chose to explicitly transcribe all occurrences of weak vowel for ease of dialectal comparison.

While making cross-dialectal comparisons, I looked for two likely indicators of loanwords: (1) irregular sound correspondences, and (2) unique lexical items. The second group includes forms uniquely shared by two dialects that do not form a subgroup (cross-dialectal borrowing), as well as words that only appear in one Atayal dialect but are also used by their non-Atayal neighbours. The contact situation of each of the four dialects in this study is distinct. Each required a slightly different methodological approach based on the most likely Formosan loans identified during initial observation. The rest of this section provides additional background information on each dialect, and detailed descriptions of the methods used to identify the loanwords therein.

2.1 Plngawan

Plngawan Atayal is spoken one village and its surrounding hamlets in Ren'ai township, Nantou county (historically, there used to be a number of villages, which were conglomerated under Japanese rule). It is bordered by Takituduh Bunun in the south and west, and by various Seediq dialects in the north and east. Crucially, no other Atayal dialects are spoken in the vicinity of Plngawan. The speakers maintain a close relationship with the surrounding Seediq, and intermarriage is common. There appears to have been no significant linguistic contact with their Bunun-speaking neighbours—unsurprising given their history of animosity.⁵

Seediq and Atayal are closely related to each other (Ferrell 1969, Blust 1999b). Together, they form the Atayalic subgroup within the Austronesian family, and share many lexical and phonological innovations (Li 1981). Plngawan shows a higher degree of mutual intelligibility with Seediq (higher than Plngawan with Squliq or Squliq with Seediq), though this is likely due to frequent linguistic contact (Rau 2004:75–77). The nature of the linguistic contact between Plngawan and Seediq is that Plngawan is the smaller group, and the contact is one-sided, with vocabulary being borrowed from Seediq into Plngawan. Li

⁴But not in others, see Huang (2015b, 2018).

⁵Even so, at least one Bunun loan is attested: Plngawan *saylaw* 'bitter melon,' from Bunun *saylav* 'vegetable.'

(1985:710) identified 11 words that he believed were Seediq loans in Plngawan (because they were shared between Plngawan and Seediq, but not found in other Atayal dialects). Here we expand Li's list using both irregular sound correspondences and exclusively shared vocabulary as evidence of borrowing.

There are three closely related Seediq dialects: Tgdaya (Paran), Toda, and Truku (Tsukida 2005, Sung 2018). All three are spoken in the vicinity of Plngawan, and all three can serve as a source of loanwords. In some cases, it is possible to identify a specific Seediq dialect from which Plngawan loaned a word. Where this is not possible, I use 'Seediq' as a general term.

The lexical influence of Seediq on Plngawan is significant. Loanwords exhibit both regular and irregular sound correspondences, which suggests that some of them were borrowed before the relevant sound changes took place, and some after these sound changes; thus, the language contact has been sustained for quite some time. The Plngawan contact situation is relatively simple. Seediq loanwords in Plngawan are plentiful, and in many cases can easily be matched back to the source, as there have been few sound changes in both Seediq and Plngawan after the split of their common ancestor, Proto-Atayalic. Both irregular sound correspondences and uniquely shared vocabulary have been helpful in identifying loanwords in this case. Certain words can be traced back not just to Seediq as a whole, but to a specific Seediq dialect, based on a sound change or a specific aberrancy that was loaned into Plngawan.

Since Atayal and Seediq are closely related, they share a large number of cognates. Some of the Seediq loanwords in Plngawan in fact have cognates in other Atayal dialects. However, they can be identified as borrowings using irregular sound correspondences as a gauge.

In a few words, the directionality of borrowing is ambiguous if the word is examined on its own (see for example *karetan* 'coin' in section 4.1). These words are still assumed to be loaned from Seediq into Plngawan for the following reasons: (1) Plngawan is a small dialect surrounded by Seediq speakers, and the relative importance and prestige of the two languages is skewed heavily in Seediq's favour, (2) we already see a lot of Seediq borrowings in Plngawan, but so far to borrowings in the opposite direction have been reported.

2.2 Klesan

Klesan Atayal is spoken in five villages on the coast of Yilan county in north-eastern Taiwan. However, this coastal location is not their original home.

Klesan speakers used to live in the upper reaches of the Heping river basin (和平溪),⁶ around Nanhu mountain (南湖大山), some 40 km inland and to the south, and moved to the coast in the first half of the 20th century. They had originally split from the B'ala clan in northern Nantou county, moving to the Heping river upstream sometime in the 18th century (Mabuchi 1953-54:95). Not long after that, a group of Squliq speakers joined them after their dispersal from Nantou, constituting just under half the population (Mabuchi 1953-54:92). Thus, the settlers of the coastal Yilan were not homogeneous, but were a mixture of Klesan Atayal, Squliq Atayal, and also Seediq speakers, who used to preserve their linguistic identities (Mabuchi 1953-54:94). While the Klesan dialect as it is currently spoken is a distinct form of Atayal, it shows considerable influence from Squliq, unsurprising given their history.

Klesan is most closely related to the S'uli and Matu'aw dialects, and the three form the Nuclear Southern Atayal subgroup (Goderich 2020:193). Squliq, on the other hand, is a much more distant relative, belonging instead to the Norther Atayal branch. This means that lexical items uniquely shared by Squliq and Klesan are prime candidates for being lexical borrowings. Despite the presence of Seediq ancestry in Klesan, there does not appear to be any significant amount of linguistic influence from the former to the latter.

The biggest hurdle when working with Klesan to find irregular correspondences is Japanese influence: while it is an Atayal dialect and not a creole language like that of several Atayal villages in Yilan (Chien and Sanada 2010), it has been heavily influenced by Japanese, with many lexical items of Atayal origin being replaced by their Japanese counterparts. This influence decreased the pool of potential borrowings from other languages. The focus of this study was limited to Formosan loanwords, and all Japanese borrowings were ignored.

⁶Also called Dazhuoshui river (大濁水溪).

2.3 Matu’uwal

Matu’uwal Atayal is spoken in three villages in Tai’an township, Miaoli county. Saisiyat is spoken directly to the north of Matu’uwal, and Pazih used to be found to the west and southwest. The latter is no longer spoken in the area, although some heritage speakers of Kaxabu, a closely related language, preserve their linguistic heritage in a few villages around Puli township, Nantou county. Another Austronesian language, Taokas, was spoken even further west, on the coast of Miaoli county, but it has long been extinct, and the only resources for it are a few very short wordlists collected from heritage speakers around the turn of the 20th century (Tsuchida 1982). As such, the most likely sources for loanwords in Matu’uwal are Pazih and Saisiyat.

Matu’uwal was influenced by Formosan languages outside the Atayalic subgroup, and thus differs the most from other Atayal dialects both in terms of phonology and lexicon. The lowest common ancestor of Pazih, Saisiyat, and Atayal may well be Proto-Austronesian (PAN)—we do not have a universally agreed upon subgrouping of Austronesian languages, but it is clear that the level of linguistic divergence in Formosan languages is very high.

Mabuchi (1953-54:103) writes that in the patrilineal family histories of the Matu’uwal, the status of some of the ancestors is uncertain, suggesting they could be of non-Atayal origin. He further surmises that these outsider ancestors may be the reason for Matu’uwal being so distinct from other Atayal dialects, both lexically and phonologically. Unlike the dispersal of the Sqliq, which can be dated to the mid 18th century, Mabuchi does not provide an approximate time for the arrival of the Atayal people in Miaoli, but does mention that the movement of “C’uli” Atayal along the Da-an river basin (大安溪) happened earlier than that.⁷ He goes on to say that inland Miaoli north of the Da-an river used to be populated by the Saisiyat until they were displaced by the Atayal.

While there is comparatively little lexical influence of Pazih and Saisiyat in Matu’uwal, it is considerably higher than in other Atayal dialects (barring Matu’aw, who most likely loaned these words via Matu’uwal). Linguistic evidence also tells us that Matu’uwal is quite distinct from both S’uli and Matu’aw Atayal (Goderich 2020), belonging to a completely different subgroup. Unlike other Atayal groups, Matu’uwal speakers moved in not just from the south

⁷Mabuchi (1953-54) puts all non-Sqliq Atayal groups into a single entity, “C’uli”.

and east, but also from the north and even west, sometimes intermingling with other populations (Mabuchi 1953-54:103). Language data recorded in 1901 from a heritage speaker in Touwu township, Miaoli county (苗栗縣頭屋鄉) points to other Atayal dialects spoken in the area, closely related to, but distinct from, Matu'uwal (Goderich 2021).

The amount of PAN cognates between Matu'uwal-Pazih or Matu'uwal-Saisiyat is low. There is the additional issue of circularity: protoforms are reconstructed based on current linguistic data, but they may include borrowed items that linguists did not account for. Thus, PAN may include reconstructions for words that originated in only a single branch and spread to other branches through language contact. Using PAN protoforms to support an argument must always be done with care.

With that said, there are still instances where PAN protoforms can be helpful. They are useful as a starting point, to identify an initial batch of potential loans. In this study, I compared Matu'uwal not only with other Atayal dialects, but also with reconstructed PAN forms. Where Matu'uwal disagreed with the other dialects, PAN reconstructions were used to check for irregular reflexes of proto-phonemes. This led to the identification of several borrowings.

Matu'uwal is lexically quite divergent from the rest of the Atayal dialects. Despite this, irregular sound correspondences in it are rare. Instead, Matu'uwal often just uses completely different vocabulary from other Atayal dialects. My main method of identifying loanwords in Matu'uwal was collecting these words and trying to match them to either Pazih or Saisiyat. The results were mixed: I did indeed find some plausible candidates in both of the source languages, but there would oftentimes be a discrepancy in one or two segments. These discrepancies prevent us from claiming with certainty that Pazih and Saisiyat are the direct sources of these loans in Matu'uwal, yet at the same time both the form and the meaning of these words are similar enough to arouse suspicion of a connection. The fact that there are quite a few very similar-sounding words with identical meanings in languages spoken so close to each other is indicative of a borrowing relationship, but the phonological discrepancies cannot yet be adequately explained. In some cases, it is unclear where a given lexical item originated. The direction of borrowing from Pazih to Matu'uwal makes sense for introduced cultural items and New World flora, since Matu'uwal speakers live further inland, and would likely encounter such innovations later than their plains-dwelling Pazih neighbours.

2.4 Matu'aw

Matu'aw is a very small dialect spoken in just two tribal villages in Daxing Village, Tai'an Township, Miaoli County. This dialect is most closely related to S'uli, who are their neighbours to the south and east. Matu'aw and S'uli together form the Southwestern branch of the Southern Atayal subgroup (Goderich 2020:210). To the north, Matu'aw borders Matu'uwal Atayal, a member of the Northern Atayal subgroup. Phylogenetically speaking, Matu'aw and Matu'uwal are clearly distinct, but their geographical proximity has spurred a contact relationship. Inter-marriage between the two communities is also quite common.

The presence of close cultural contact between Matu'aw and Matu'uwal naturally arouses suspicions of linguistic contact as well. Matu'uwal is lexically divergent, and as such many borrowings from it into Matu'aw should be readily identifiable.

My main approach in identifying loanwords was checking for any lexical items shared uniquely between Matu'aw and Matu'uwal, and crucially, not being found in S'uli as spoken in Miaoli county. A few of these words have irregular sound correspondences in Matu'aw, which provide further evidence of lexical borrowing.

A small number of lexical items are uniquely shared between Matu'aw, Matu'uwal, and either Pazih or Saisiyat. Due to evidence of close contact between Matu'uwal and Matu'aw on the one hand, and the presence of a larger number of Pazih and Saisiyat loans in Matu'uwal on the other hand, the most likely scenario is that these words were also loaned from Matu'uwal into Matu'aw, having first been loaned into Matu'uwal from other languages.

3 Phonological aspects

3.1 Phoneme inventories

3.1.1 Atayal

3.1.1.1 Consonants The consonant inventory of Atayal is presented in table 2. Three phonemes are found in some dialects, but not others. These are: /q/, /c/, and /ɿ/. The consonant inventory shown in the table is essentially

that of Proto-Atayal. Various dialects lost from one up to three contrasts in this system, discussed below. Additionally, Squliq Atayal exhibits glide fortition, with /y/ fortitioning into [ʒ], written as <z> or <zy>. In some varieties of Squliq, this fortitioned [ʒ] has developed quasi-phonemic status (Huang 2015a).

Table 2: Common Atayal consonant inventory (following Goderich 2020)

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

The phonetic realization of the phonemes /b/ and /g/ depends on the dialect and the speaker. In most dialects, they are voiced fricatives in all positions.⁸ Plngawan /b/ and /g/ are usually plosives in word-initial position or following consonants, and realized as fricatives intervocalically.

The liquid /r/ can be a tap, or sometimes a trill. In Plngawan, it is distinguished from the alveolar approximant /ɹ/, written in the current orthography as <ɹ>. In this paper, I use the IPA symbol directly in order to avoid confusion.

The presence of the phonemes /q/, /c/, and /ɹ/ in various Atayal dialects is shown in table 3. Plngawan, Klesan, and Matu'aw lack /q/. Matu'aw and Squliq merged Proto-Atayal *c into /s/.⁹ Only Plngawan has a second rhotic /ɹ/. This distribution originates in the historical changes of Proto-Atayal *q, *c, and *ɹ, discussed in section 3.2.

⁸Only Matu'uwal allows /b/ and /g/ to occur in word-final position. In other dialects, they can only be found word-initially and word-medially.

⁹Squliq has a phonemic or quasi-phonemic /c/ (Li 1980, Egerod 1999), however it does not originate from Proto-Atayal *c and its distribution is much more limited.

Table 3: The presence of phonemic /q/, /c/, and /ɹ/ in Atayal dialects

	q	c	ɹ
PIngawan		V	V
Klesan		V	
Matu'uwal	V	V	
Matu'aw			
Squliq	V	(V)	

It is important to note that Matu'uwal itself has minor subdialectal differences. The Matu'uwal speakers in Qing'an Village, which is geographically the closest to Matu'aw, also lack a /c/ phoneme, having merged it with /s/. It is possible that this merger occurred due to influence from Matu'aw. The Qing'an variety of Matu'uwal is the most likely source of Matu'uwal loanwords in Matu'aw, due to their geographic proximity. However, we cannot be sure that the merger in Qing'an Matu'uwal occurred before any words with historical /c/ were loaned into Matu'aw. Even if some loans happened before the merger, /s/ is a likely target to map /c/ in a language that does not have it as a phoneme.

3.1.1.2 Vowels There are greater differences between the vowel systems of various Atayal dialects than between their consonant inventories. They are summarized in table 4, with detailed discussion of individual dialects below. This section is sourced from Li (1980), Huang (2006, 2018), Chen (2011), and Goderich (2020).

Table 4: Vowel phonemes in Atayal dialects

Dialect	Vowels
PIngawan	/a, i, u, e, o/
Klesan	/a, i, u, e, o, (ə)/
Squliq	/a, i, u, (e), (o), (ə)/
Matu'uwal	/a, i, u, (ə)/
Matu'aw	/a, i, u/

There are five vowels in PIngawan: /a/, /e/, /i/, /o/, /u/. The mid vowels come from coalesced vowel sequences and diphthongs, but can appear in

closed syllables: Plngawan *useʔ* ‘water,’ *baɿok* ‘pig,’ compare Matu’uwal *qusiyaʔ* ‘water,’ *bawwak* ‘pig.’ The fricative /h/ also has a lowering effect on high vowels, so that they are perceived as mid vowels: Plngawan *gaʔiloh* ‘banana,’ cf. Matu’uwal *guqiluh*. Schwa in Plngawan is not phonemic and never appears in surface forms.

Klesan permits six phonetically distinct vowel sounds to occur in the surface representation: [a], [i], [u], [e], [o], [ə]. Which of these are phonemic is a matter of phonological analysis, and outside the scope of this paper. It should be noted that the mid vowels [e] and [o] are the result of either the monophthongization of historical *ay and *aw, or else the result of vowel lowering in certain environments, much like the case with Plngawan. Vowels outside the rightmost foot are weakened to [ə].

Squliq is the biggest Atayal dialect both in terms of geographical spread and the number of speakers, and as such it shows variation in its phonology. The vowel systems of different Squliq dialects may include mid vowels /e/ and /o/, or preserve the more conservative pronunciations /ay/ and /aw/ (Li 1980), but they all share the phonemes /a/, /i/, /u/. Vowels outside the rightmost foot are weakened to [ə]. The schwa itself is at best only marginally phonemic (Huang 2018), and cannot be stressed.

Matu’uwal has three phonemic vowels: /a/, /i/, and /u/. A central vowel [ə] can only appear in the initial syllable of a word, e.g. [bəhut] ‘squirrel’ or [təhawnak] ‘to sit.’ The syllable structure of Matu’uwal is CV(C).(C)V(C), meaning that codas are allowed in all syllables, but onsets are optional only in the rightmost syllable. The latter means that Matu’uwal allows hiatuses, but only in the rightmost foot, e.g. /kaal/ [ka.'al] ‘sky’ is disyllabic. There are no vowel-initial words. Words that are customarily written with an initial vowel have a phonemic glottal stop preceding it, e.g. /itaal/ [ʔi.ta.al] ‘indigenous person, Atayal.’ Matu’uwal allows heterorganic consonant clusters across syllable boundaries, for example /waqcug/ [waq.ʔsuy] ‘throat.’ Tautosyllabic clusters are disallowed.

The vowel system of Matu’aw is even simpler than that of Matu’uwal, having just three vowels: /a, i, u/. Unlike Matu’uwal, Matu’aw does not allow /ə/ in the surface representation, and instead uses the low vowel /a/ as a default vowel for epenthesis. Matu’aw often has an intervening vowel where its Matu’uwal cognate has a heterosyllabic consonant cluster: compare Matu’uwal *talɰaʔ* and Matu’aw *talaɰaʔ*, both from Proto-Atayal *taləɰaʔ ‘k.o. snare.’ One exception

to this restriction is glides: the glides /w/ and /y/ can occur as word-internal codas freely, e.g. Matu’aw *rawyí?* ‘eye’ and *wayluŋ* ‘chicken.’

Historical *ə became *u in the final syllable even before Proto-Atayalic: e.g. PAN *qaRəm ‘pangolin’ > Proto-Atayalic *qagum, whence Matu’uwal *qagum*, Squliq *qom*, Plngawan *ʔoŋ*, Seediq *ʔaruŋ* (Li 1981). This distribution has not changed to this day, and no Atayal dialects have [ə] in the final syllable, even though in some dialects it can appear in other positions.

Stress in all Atayal dialects normally falls on the final syllable (Li 1980, Chen 2011, Goderich 2020).

3.1.2 Seediq

Atayal and Seediq are closely related languages, and their phonemic inventories are quite similar. The main difference is the development of Proto-Atayalic *d and *r, which remained /d/ and /r/ in Seediq, but shifted into *r and *ɹ in Proto-Atayal. As mentioned in section 2.1, Seediq consists of three dialects: Tgdaya (Paran), Toda, and Truku.

The common consonant system of Seediq is shown in table 5. Truku does not have /c/ as a phoneme, but the other two dialects do. The rest of the consonants are the same among the three dialects.

Table 5: Seediq consonants (following Tsukida 2005, and Sung 2018)

p	t	k	q	ʔ
b	d	g		
	(c [ts])			
	s	x		h [h]
m	n	ŋ		
	l,r [r]			
w	y [j]			

Seediq differs from Plngawan in having /d/ and /q/, and lacking /ɹ/. The voiced obstruents /b/ and /d/ are plosives, but Tsukida (2005:292) describes /g/ as a fricative in Truku.

Truku Seediq has four vowel phonemes: /a/, /i/, /u/, and /ə/. Just like in Atayal, the central vowel /ə/ does not appear in the final syllable. Like in most

dialects of Atayal (but not Plngawan), all vowel distinctions beyond the rightmost foot are lost. Stress is penultimate, thus /ə/ can also be stressed.

Tgdaya Seediq, on the other hand, has a five-vowel system: /a, i, u, e, o/ (Sung 2018:13). The vowel /ə/ is not present in the dialect, instead, the reduced vowel (i.e. the vowel that is inserted automatically outside the final foot) is /u/. The mid vowels /e/ and /o/ are the result of the monophthongization of /ay/ and /aw/; additionally, historical *ə became /e/ in the penultimate syllable. In my transcriptions of Seediq, as with Atayal, I write out the weak vowels for the reader's convenience.

3.1.3 Pazih

The Pazih consonant inventory is shown in table 6, (source: Li and Tsuchida 2001). It has a few key differences with Matu'uwal, namely: (1) the presence of /d/, (2) the absence of /q/, (3) the presence of /z/, and (4) the absence of /c/.

Table 6: Pazih consonant inventory

p	t	k	ʔ
b	d	g	
	s	x	h
	z		
m	n	ŋ	
	l, r		
w	y		

Pazih has four vowel phonemes: /a/, /i/, /u/, and /ə/. Unlike Matu'uwal, the central vowel /ə/ can be stressed, and its distribution is not restricted. Stress in Pazih is word-final.

The canonical structure of Pazih roots is CVCVC (Blust 1999a). The only heterosyllabic clusters allowed are either homorganic -NC- sequences (/bin.tun/ 'star'), or else -GC- sequences in reduplicated monosyllables (/taw.taw/ 'peanut').

3.1.4 Saisiyat

The phoneme inventory of Saisiyat differs considerably from Atayal and Pazih (source: Zeitoun and C. Wu 2005, Yeh 2018). There are two dialects of Saisiyat: Ta'ai (also called "Northern Saisiyat") and Tungho (or "Southern Saisiyat") (Li 1978). The Saisiyat dialect discussed here is Tungho, which is spoken in Nanchuang township, Miaoli county, and is geographically closer to Matu'uwal Atayal than the Northern Ta'ai dialect. The consonantal inventory of Tungho Saisiyat is shown in table 7, with symbols in square brackets marking the phonetic value where it differs from the conventional orthography. Tungho is marked by (1) the presence of interdental fricatives /s/ [θ] and /z/ [ð], (2) the lack of a /g/ phoneme, (3) a complete lack of velar fricatives.

Table 7: Tungho Saisiyat consonant inventory

p	t	k	ʔ
	s [θ], s [ʃ]		h
b [β]	z [ð]		
m	n	ng [ŋ]	
	l, r		
w	y [j]		

The vowel inventory of Saisiyat, shown in table 8, is also unlike those of Pazih and Matu'uwal. It has six vowels, the phonemes /æ/ and /œ/ adding to the four vowel system of Proto-Austronesian.

Table 8: Tungho Saisiyat vowel phonemes

i	
ə, œ	o
æ	a

According to Zeitoun and C. Wu (2005), the phonemes /æ/ and /œ/ are actually more central than their orthography would imply.

The maximum syllable in Saisiyat is CVC (Zeitoun and C. Wu 2005). Like Matu'uwal, Saisiyat allows hiatuses: *tiiS* [ti.iʃ] 'to wipe.' Unlike Matu'uwal,

Saisiyat hiatuses are not limited to the final two syllables: *paatol* [pa.a.tol] ‘to sing.’

Saisiyat allows heterorganic consonant clusters across syllable boundaries: [rim.ʔan] ‘broad’ (Huang 2015b:56). Vowels can be long in Saisiyat, but only in the Tungho dialect, where they are an artifact of a historical lateral flap (Li 1978 : 139): Tungho *talkaa* [tal.ka:] ‘table,’ cf. Ta’ai *talkaL* . (Tungho is spoken in Nanchuang township, Miaoli county, and is geographically closer to Matu’uwal than the northern Ta’ai dialect).

3.2 Regular sound correspondences

This section provides regular sound correspondences between Seediq and the various Atayal dialects under discussion.

The ancestor of both Atayal and Seediq is called Proto-Atayalic. It split into the two aforementioned branches, which in turn split into the various dialects we see today.

The regular reflexes of Proto-Atayalic (PAic) **c*, **q*, **d*, **r* and **y* are shown in table 9. The most important sound correspondences are marked in bold. The main difference between Atayal and Seediq is a chain shift in the former: PAic **r* > PA **ɹ*, then PAic **d* > PA **r*. Plngawan, Klesan, and Matu’aw merged PA **q* into /ʔ/. Matuaw and Squliq merged **c* into /s/. Proto-Atayal **ɹ* merged with /y/ in Klesan, Matu’aw, and Squliq, and in Matu’uwal it was either deleted or became /w/ (depending on the environment and the subdialect). As mentioned in section 3.1.2, Truku Seediq does not have /c/, having merged PAic **c* into /s/.

Table 9: Reflexes of Proto-Atayalic **c*, **q*, **d*, **r* and **y*

Proto-Atayalic	<i>*c</i>	<i>*q</i>	<i>*d</i>	<i>*r</i>	<i>*y</i>
Seediq	c/s	q	d	r	y
Proto-Atayal	<i>*c</i>	<i>*q</i>	<i>*r</i>	<i>*ɹ</i>	<i>*y</i>
Plngawan	c	ʔ	r	ɹ	y
Klesan	c	ʔ	r	y	y
Squliq	s	q	r	y	y or z
Matu’uwal	c	q	r	Ø/w-	y
Matu’aw	s	ʔ	r	y	y

There are several irregular sound correspondences that can help us identify Seediq loans in Plngawan. These are:

- Plngawan /k/ corresponding to Seediq /q/ instead of the regular /ʔ/ to /q/.
- Plngawan /r/ corresponding to Seediq /r/ instead of the regular /ɹ/ to /r/.
- Plngawan and Seediq sharing an aberrant vowel correspondence in the final syllable.

In loans from Seediq which have a /ə/ vowel, Plngawan substitutes a cardinal vowel. This is most commonly /a/, but sometimes also /u/. Part of these are vowels in the AV infix *-um-* (*-əm-* in Truku and Toda), but some are root-internal. The reason for choosing /u/ may be related to the pronunciation of weakened vowels in different Seediq dialects.

The biggest differences between Klesan and Squliq are the changes of Proto-Atayal *c and *q. The irregular correspondences indicative of borrowing are thus:

- Klesan /k/ corresponding to Squliq /q/ instead of the regular /ʔ/ to /q/.
- Klesan /s/ where a /c/ is expected because of reflexes in other dialects (e.g. Matu'uwal or Plngawan).

Irregular sound correspondences are not the main tool for identifying Matu'uwal loanwords in Matu'aw. Instead, we have to rely on uniquely shared lexical items that cannot be reconstructed to Proto-Atayal. However, we do see one irregular correspondence in several loanwords:

- Matu'aw /k/ corresponding to Matu'uwal /q/ instead of the regular /ʔ/ to /q/.

4 Loanwords

4.1 Seediq loanwords in Plngawan

The following words can be identified as loans from Seediq based on one or more of the sound correspondences mentioned in section 3.2, or by being exclusively shared between Plngawan and Seediq, sometimes even a specific dialect of Seediq.

apatu ‘frog.’ Truku *qəpatur*, cf. Tgdaya *qupatur*. Both the /ʔ/ to /q/ and the /ɹ/ to /r/ correspondences are regular, but the word itself appears to be limited only to Truku. The correspondence between the final consonants in Truku and Tgdaya is irregular. Other Atayal dialects use reflexes of Proto-Atayal *takaɹ: cf. Matu’uwal *taka*, S’uli and Klesan *takay*.

bibiy ‘duck.’ Tgdaya *bibi*. Ducks are not native to Taiwan, and the lack of a common Atayal term for this bird serves as evidence. Cf. Truku *ruru*, Squliq *guruʔ*, Matu’uwal *sababul*, Matu’aw *sababun* (the latter two likely loaned from Pazih, see section 4.3).

calak ‘paddy.’ Tgdaya *celaq*, Toda *cəlaq*, Truku *səlaq*. Cf. also Matu’uwal *cəlaq*, Squliq *səlaq*, Klesan *cəlaʔan*. Irregular correspondence of /k/ to /q/.

camuʔ ‘iron nail.’ Truku *samu*. Most other Atayal dialects use a loan from Japanese *kugi*: Matu’uwal *kugiʔ*, S’uli *kugi*, Skikun *kugiʔ*, Klesan *bali* or *kugi*. This word may have been borrowed from Truku before the merger of *c and *s.

ciyak ‘cucumber.’ Tgdaya *ciyak*. Cf. Matu’uwal *tabuwil* (discussed in section 4.3), Squliq *təbwil* or *təmiʔ*, Skikun *təbil*.

cumilak ‘to cut up.’ Compare Tgdaya *cumilaq* ‘to rip apart,’ cf. Truku *səmilaq* ‘to break tree branches.’ Irregular correspondence of /k/ to /q/. Other Atayal dialects do not have cognates of this lexeme.

inabal ‘wall.’ Tgdaya *qunabil*, Toda, Truku *qənabil*. /ʔ/ to /q/ is regular, but the vowel correspondence is not. However, Seediq has other irregular correspondences of *bil~bal* and *pil~pal* in the final syllable, e.g. Tgdaya *gumabil* vs Truku *gəməbal* ‘to pull out.’¹⁰ Cf. Matu’uwal *qiniriyəŋ*, S’uli *inryəŋ*, Squliq and Skikun *qənəryəŋ*.

kakaluŋ ‘wood-ear mushroom.’ Tgdaya *qeluŋ*, Truku *qəluŋ*. Cf. also Matu’uwal *qaqluŋ*, Squliq *qəluŋ*. The /e/ in Tgdaya is a regular reflex of earlier *ə in penultimate position. This is an Atayalic cognate, but identified as a loan because of the irregular correspondence of Plngawan /k/ to /q/ in other dialects.

karetan ‘coin.’ Tgdaya *kuretan*. This form seems to exist only in Plngawan and Tgdaya: cf. Truku *buhug* or *həbəŋan*, Matu’uwal *habəŋan*, Squliq and S’uli *həbəŋan*. The vowel /e/ in the penultimate syllable is additional evidence of the

¹⁰I thank Walis Hian-chi Song for bringing this to my attention.

word originating in Tgdaya. However, the weak vowel in Tgdaya is /u/, which does not correspond with Plngawan.

kulabuy ‘paper.’ Tgdaya *kulabuy*. The vowel /u/ in the initial syllable may be due to borrowing from Tgdaya. Cf. Matu’uwal *ruwas* (a Saisiyat loan, see section 4.3), Squliq *bərəbar*, Skikun *bərəbal*, S’uli *rəban*.

kuncik ‘to lend, borrow.’ Cf. Matu’uwal *kabaux*, Squliq *kəsyuw*, Skikun *kəsiyux*, Klesan *kəsiyu*. Truku has *kəməsiyuk* meaning ‘to lend, borrow,’ but also ‘to answer.’ Tgdaya *kumuciyuk* only has the meaning ‘to resist, fight back,’ and not ‘to lend.’ Plngawan uses *cumik* to mean ‘to answer,’ which is a retention from Proto-Atayal. This means that the word was likely loaned from Truku before it merged *c into *s. The change of -iyu- to -i- in Plngawan is regular (Goderich 2020:131).

liwas ‘large cooking pot.’ Truku, Tgdaya *liwas*. Cf. Matu’uwal *aybaw*, S’uli *zibaw*, Klesan *təbali*. There are usually 2-3 different words in Atayal dialects that refer to different kinds of pots, but none have a cognate of *liwas*.

masnukan ‘drunk.’ Tgdaya *busukan*, Truku *bəsukan*. Atayal dialects use two related, but different forms of the root for this meaning: (1) Matu’uwal *mabusuk*, Squliq *məbusuk*, Klesan *busuk*; (2) Matu’uwal *businuk*, Skikun *məsinuk*. Note that the final -an the Plngawan form is not a Locative voice suffix: *masnukan* is an Actor Voice form (its negation is *ini pasnukan*). Because the aberrant -an (unrelated to voice morphology) is shared between Plngawan and Seediq, this is classified as a loan.

mastamak ‘broken, spoiled.’ Tgdaya *musutemaq*, Truku *məsətəmaq*. Cf. also Squliq, Skikun *sətəmaq* or *sətəmah*, Klesan *sətəmah*. This lexical item is phonetically similar across Atayalic, however the irregular correspondence between Plngawan *k* and Seediq *q* indicates that it is a loan.

ɲiraw ‘mushrooms.’ Truku and Toda *ɲiraw*, Tgdaya *ɲiro*. Note the irregular /r/ to /r/ correspondence. The vowel /o/ in Tgdaya does not necessarily exclude it as a source, as the word may have been borrowed before monophthongization in that dialect. The reconstructed Proto-Atayal form is *tVqinug (initial vowel uncertain): Matu’uwal *təqəqinug*, Squliq *təqinuw*, S’uli *təʔinu*, Matu’aw *taʔaʔinuw*.

papak ‘foot, leg.’ Truku, Tgdaya *papak*. Cf. Matu’uwal *kukuy*, Squliq, Skikun, S’uli, Klesan *kakay*. The homophonous word *papak* means ‘ear’ in all Atayal

dialects except Matu'uwal and Plngawan.

piluw 'sparrow.' Tgdaya *piilo*, compare Truku *pisaw*. Cf. Matu'uwal *piit*, Squliq *pəzit*, Matu'aw *payit*, all from PA **pə.ɬit*. This is an almost perfect match with Tgdaya.

raral 'relatives.' Tgdaya *dadān*. The correspondences of /r/ to /d/ are regular. Final /l/ in Tgdaya regularly becomes /n/ (Li 1991:164). Cf. Matu'uwal *matrarumaʔ*, Squliq and Klesan *ləpyuŋ*.

rumigaw 'to walk around.' Tgdaya *rumigo*, Truku *rəmigaw*. Irregular correspondence of /r/ to /r/. Other Atayal dialects do not have this word.

ɰakarit or **nakarit** 'bat.' Tgdaya *pulale karic*, Toda *bəkaric*, Truku *kəlaway karic*. Note the irregular /r/ to /r/ correspondence in the final syllable. The meaning of *pulale* in Tgdaya is 'butterfly.' The initial syllable in Plngawan can be either *ɰa-* or *na-*, depending on the subdialect. This syllable was likely added after borrowing, but its function is unclear (it also appears in *ɰakarit~nakarit* 'spider,' which is homophonous with 'bat,' and *ɰapiriy~napiriy* 'butterfly').

ɰakeh/ɰakah 'bad.' This is an interesting correspondence. Tgdaya has the doublets *naqah~naqih*, just like Plngawan, and the irregular correspondence of /k/ to /q/ is suggestive of borrowing, but the initial consonants do not match, and Plngawan has /ɰ/ instead, same as Proto-Atayal **ɰaqih* 'bad' (the expected form in Plngawan would be ***ɰaʔeh*). However, a wordlist by Tashiro Antei from 1896 records 'bad' as <lakkai> in a Tgdaya dialect in what is now Hualien.¹¹ While Tashiro's transcription is far from perfect, his use of <l> here suggests that Hualien Tgdaya may have had an initial liquid in this word in the past. Cf. also Matu'uwal *aqih*, Squliq *yaqih*, Klesan *yaʔih*.

sapit 'footwear.' Tgdaya *sapic*, cf. also Hoanya 1-3 *sapit*, Taokas 3-2c *sapit* (Tsuchida 1982:80).¹² This was likely a loan into both Plngawan and Tgdaya from a Western plains language. The reconstructed Proto-Atayalic form for 'footwear' is **ramil*: see Truku *ramil*, Matu'uwal *wamil*, Squliq *yamil*.

¹¹田代安定〈臺東殖民地豫察巡回日誌〉, URL: <https://dl.lib.ntu.edu.tw/s/Tashiro/item/714782>, page 8 (accessed on 28 July, 2023). I thank Walis Hian-chi Song for bringing this document to my attention.

¹²The numbers and letters in "Hoanya 1-3" and "Taokas 3-2c" refer to the sources of data. The first number refers to the "original village" of the speaker (or their dialect). The second number refers to the source, and letters signify different informants (see Tsuchida 1982:15 for a full list).

suŋkanux ‘to smell [AV].’ Tgdaya *sumukenux*. Cf. Matu’uwal *sumauk*, Squliq, Skikun, S’uli, Klesan *səmək* (< PA *s<um>aʔuk < PAN *Sajek). The Seediq form is likely related to Atayal, but with an additional suffix, most probably a remnant of the now defunct gender register system.

supux ‘cockroach.’ Truku *supug~supur*, Tgdaya *pucupux*. The /s/ in the first syllable points to Truku, although the final consonant is irregular (judging from the variation it could be a later innovation). All other Atayal dialects have forms descended from Proto-Atayal *hipux: e.g. Matu’uwal *hahipux*, Squliq *hipux*, Skikun *kəhepux*.

tamurak ‘pumpkin.’ Tgdaya *tumurak*, Truku *təmurak* ‘cucumber.’ Irregular correspondence of /r/ to /r/. Cf. Matu’uwal *rilum*, Matu’aw *baun*, S’uli *bon* or *abaŋ*, Squliq and Skikun *qabaŋ*, Klesan *kabaŋ*. Squashes are a New World plant, and lexical borrowing for this word is common.

umpix ‘to push down, press.’ Truku *epix*, cf. Matu’uwal *umpux*, Squliq and Klesan *məpux* (< PA *ʔ<um>əpux). This word is an Atayalic cognate, but the vowel in the final syllable of the root is irregular. This aberrant vowel is shared between Plngawan and Truku. I have collected two separate LV forms of this verb: *apix-an* and *apx-an* (the latter is the expected form). In LV, both the inherited and the borrowed form are attested. In AV, it appears that the loaned *umpix* has outcompeted the native pronunciation (expected to be *umpux*).

yageh ‘baby girl.’ Truku *agih*. Most Atayal and Seediq dialects do not have separate words for babies of different genders. Matu’aw has *yulah* ‘baby boy’ (a cognate of Plngawan *yulah* ‘baby boy’), but uses *yunay* for ‘baby girl.’ The initial glide in *yageh* may be the result of the nominative case marker *i* fusing with the root, as happened in some kinship terms and personal names that start with a glottal stop (Ochiai 2019).

4.2 Squliq loanwords in Klesan

Klesan is marked by the loss of Proto-Atayal *q and the retention of *c, as opposed to Squliq, which preserves /q/ but merged PA *c into /s/. The loanwords collected in this section have for the most part been identified using the irregular correspondence of Klesan /k/ to Squliq /q/.

cipok ‘a little.’ Squliq *cipoq*. This word is unique to Squliq, the equivalent in Matu’uwal and S’uli is *tikay*. Apart from that, it still demonstrates the irregular

/k/ to /q/ correspondence. Note the mid vowel in the final syllable, lowered in Squliq due to a following /q/, but preserved as a mid vowel in Klesan despite the lack of this consonant phoneme.

cira ‘spindle.’ Squliq *cira?*. Cf. also Matu’uwal *matisa?*, Plngawan *matira?*, Matu’aw *matisa?* ‘to turn spindle,’ all from Proto-Atayal **tisa?* ‘spindle.’ Both **ti* > *ci* and rhotacism of **s* in this specific position are characteristic of Squliq (Goderich 2020:118, 132). The expected reflex in Klesan would have been ***tisa*.¹³

kabəŋ ‘pumpkin.’ Squliq *qabəŋ*. Cf. Matu’uwal *rilum* (loaned from a neighbouring language, see section 4.3), S’uli *abəŋ*. Note that even though squashes are a New World plant, the sound correspondence between S’uli *abəŋ* and Squliq *qabəŋ* is regular.

kasu ‘boat.’ Squliq *qasu?*. Cf. Matu’uwal *qacu?*, Plngawan *asu?*, S’uli *asu*, Skikun *kasu?*, from a possible Proto-Atayal form **qacu?*, for which the expected reflex in Klesan would be ***ʔacu*.¹⁴ This word can be connected to Squliq on the basis of not just one, but two irregular sound correspondences.

kenu ‘mushrooms.’ Squliq *təqinu*. Cf. Matu’uwal *təqəqinug*, S’uli *təʔinu*. Note that Klesan has the mid vowel [e] in this word. High vowel lowering in Squliq occurs next to /h/ and /q/, and the mid vowel in Klesan may be an attempt by Klesan speakers to mimic the lowering effect of /q/.

kəbubu ‘hat.’ Squliq *qəbubu?*. Cf. Matu’uwal *qəbubu?*, S’uli *bubing*.

kəcyan mit ‘Siberian weasel’ (*Mustela sibirica*). One of the terms for this animal in Squliq is *qəcyan mit*, which can be further analyzed as *qəcyan* ‘buttocks, anus’ and *mit* ‘goat,’ meaning ‘goat buttocks’ (the weasels are so named due to their apparent propensity to attack ensnared animals, including large ruminants, by entering through the rectum and eating the animal’s entrails). The word for ‘buttocks, anus’ in Klesan is *byux*, a completely unrelated form. The part *kəcyan* is indicative of borrowing from Squliq, with the irregular correspondence of /k/ to /q/.

¹³Note that not all Klesan words with *-ira-* or *-iru-* are borrowings. Some are directly inherited from Proto-Atayal, e.g. **miray* > Klesan *miray* ‘to turn’. Only instances of irregular /s/ to /r/ correspondences are considered loanwords.

¹⁴This reconstruction is somewhat problematic, because Plngawan *asu?* and Skikun *kasu?* are also irregular. The forms are all very similar, however there appears to have been widespread borrowing.

kəmasu ‘to divide.’ Squliq *qəmasuw*. Cf. Matu’uwal *qumasug*, S’uli *əmasu*.

zəzik ‘deep.’ Squliq *zəzik*. This word has an inherited doublet in Klesan: *yəyik*. Fortified glides are a trademark feature of Squliq Atayal.

4.3 Pazih and Saisiyat loanwords in Matu’uwal

There are not many solid candidates for borrowings from Saisiyat and Pazih into Matu’uwal. The sound correspondences are not impeccable, but at the same time the similarities in both form and meaning are too great to be explained by chance resemblance alone. These loanwords may have come into Matu’uwal from a different language, e.g. Taokas, but there is extremely little data on Western plains languages. Those were already moribund with only a few heritage speakers by the time they were recorded, and the only surviving data is field notes by Japanese linguists and anthropologists, elicited from speakers who were neither native nor fluent (Tsuchida 1982).

Most possible loanwords in Matu’uwal have imperfect phonological correspondences, which makes the comparisons somewhat less convincing. There is still a number of words for which Pazih or Saisiyat is the only plausible candidate.

A small number of words are shared with Matu’uwal and either Pazih or Saisiyat. These represent a subset of Pazih/Saisiyat loans in Matu’uwal, and Matu’aw does not appear to have other Pazih or Saisiyat loanwords not found in Matu’uwal. This means that for these words, the most likely path of borrowing is Pazih/Saisiyat > Matu’uwal > Matu’aw. These instances are all individually mentioned in this list.

anan ‘right (hand side).’ Compare Squliq, S’uli, Klesan *ləlaw*, Skikun *ləlax*, Plngawan *anali?*. Both Matu’uwal and Plngawan reflexes appear to be connected to PAN *wanaN ‘right,’ however the Matu’uwal reflex of PAN *N is irregular, the regular reflex being /l/.¹⁵ The Matu’uwal lexeme is identical with Pazih *anan*, which is the most likely candidate for a source.

balii? ‘flute, musical instrument.’ The main musical instrument of the Atayal has traditionally been the mouth harp (Proto-Atayalic *lubug, see (Li 1981)).¹⁶

¹⁵The extra syllable in Plngawan is a male register affix, also found in *wakili?* ‘strap,’ cf. Matu’uwal *wakil*.

¹⁶An anonymous reviewer points out that it’s not just Atayal and Seediq. Other Formosan terms for ‘mouth harp’ are reconstructable as PAN *NubəR > Budai Rukai *ləbər*, Paiwan *la-luber*.

Most Atayal dialects use one or two roots to refer to musical instruments, usually just mouth harps (some also have a separate root for the meaning ‘drum’). Skikun has *pənəgoʔ* ‘flute,’ which may be related to Seediq *pwau* and Truku *pgagu* ‘flute.’ In my field notes, Matu’uwal is the only Atayal dialect with three distinct words for musical instruments: *lubug* ‘mouth harp,’ *ginugun* ‘drum,’ and *baliʔ* ‘flute.’¹⁷ The Matu’uwal lexeme is likely related to Saisiyat *bo:liʔ*, however Saisiyat lacks a hiatus here, and the initial vowels do not match.

ginuʔ ‘winnowing basket,’ also Matu’aw **ginuʔ**. This refers to a shovel-like basket used specifically for winnowing grain, deep and rounded on one end, flat on the other. This word cannot be reconstructed to Proto-Atayal. Other Atayal groups do not appear to have used winnowing implements of this shape, and utilized flat round woven baskets (Proto-Atayal **balukuʔ*) for winnowing, among other uses. Blust initially reconstructed PAN **niRu* (Blust 1999a) based on Philippine reflexes, but later changed his reconstruction to **Rinu* (Blust and Trussel Ongoing) because of the Pazih reflex *xinu*, with the caveat that Philippine languages and Pazih do not agree on consonant position. Pazih, Matu’uwal, and Matu’aw appears to be the only Formosan languages that have reflexes of this word. Just like in *tigubuʔ* ‘sugarcane,’ Pazih /x/ is mapped to Matu’uwal /g/.

hamhum ‘cloud.’ All other dialects have a reflex of Proto-Atayal **ɹulun*: Suliq, Klesan, Skikun *yulun*, PIngawan *ɹɹulun*. The ILRDF online dictionary gives the form Saisiyat *homom* ‘cloud,’ but Li (1978) has *əməm* for Tungho and *LəmLəm* for Ta’ai. Blust’s Austronesian Comparative Dictionary¹⁸ reconstructed Saisiyat *LəmLəm* to PAN **lemlem* ‘dark, overcast,’ also using Thao *ma-rumrum* ‘dim’ and Bunun *humhum* ‘twilight’ as evidence. The latter Bunun form appears very similar to Matu’uwal *hamhum*, however I have not been able to find it in other sources, and Blust did not provide a source dialect. The word for ‘cloud’ is *luhum* in most Bunun dialects (per Shibata 2020), and the ILRDF online dictionary has Isbukun Bunun *madumdum* ‘dark.’ The Matu’uwal form is a reduplicated monosyllable which can be represented underlyingly as /həm.həm/, with both vowels being the regular realization of an underlying schwa in their respective positions (see also Matu’uwal *marahum* ‘bruise’ in this section). The presence of /ə/ vowels in Saisiyat precludes borrowing from

an (Ferrell 1982), Puyuma *luber*.

¹⁷Matu’aw also has *ginugun* meaning ‘drums,’ which is probably a loan from Matu’uwal.

¹⁸<https://www.trussel2.com/ACD/>, accessed on February 1, 2024.

Matu'uwal to Saisiyat (we would expect Saisiyat to retain Matu'uwal vowels in case of borrowing). If borrowed from Saisiyat into Matu'uwal as a reduplicated monosyllable, the schwa vowels would need to be repaired to conform with Matu'uwal phonotactics, and CaCCuC is the expected template in that case. The /h/ is not adequately explained in this scenario; one possibility is that Matu'uwal borrowed the form from Tungho when the phoneme /L/ was weakened, but not completely lost, and reinterpreted it as /h/.

kapuwa 'guava.' Saisiyat *lapowar*, Pazih *laapuwat*. Cf. Plngawan *uti?-gilun* (lit. 'chicken faeces'), Matu'aw *rapwal*, S'uli *puwan*, Squliq *sebwal*, Skikun *para?* (< Southern Min *puát-á*), Klesan *labas*. Guavas, like squashes, are a New World plant. Note the open final syllable in Matu'uwal, which comes from the loss of final *ɿ. Saisiyat has a final /r/, which may have been mapped to /ɿ/ in an earlier stage of Matu'uwal, presumably because word-final /r/ is much rarer. The only problem is the initial consonant. Saisiyat does have the variant form *rapowar*, so there may have been other variants in the past. Matu'uwal does not use *ka-* productively as a nominal prefix, so it must have been present before borrowing. Cf. also Southern Min *puát-á*, which may be related.

lalai? 'k.o. cicada (*Cryptotympana holsti*),' also Matu'aw **lalay?**. Cf. S'uli *ryay*, Plngawan *caciret*, Squliq *katata?*, Skikun *ɲale?*, Klesan *semi* (< Japanese *semi*). This lexeme is a loan in 'many Formosan languages' (Blust 1999a), cf. for example Pazih *lalay*. Note that Matu'uwal has a hiatus and a final glottal stop, which is identical to Saisiyat *lala:i?* 'cicada.' While Matu'aw *lalay?* and S'uli *ryay* are phonetically somewhat similar, the liquid correspondence is irregular. Neither Saisiyat nor other Atayal dialects distinguish between different kinds of cicadas, but both Matu'uwal *lalai?* and Matu'aw *lalay?* refer specifically to cicadas of the *Cryptotympana* genus, which are the largest cicadas in Taiwan. Other cicadas are called *lalat* in Matu'uwal and *lalan* in Matu'aw (possibly also loaned).

lihpiq 'thin.' This comparison is mystifying. On the one hand, it is just one or two segments off compared to other Atayal dialects: Squliq *ləhəmiq*, Plngawan *lahmi?*.¹⁹ Saisiyat *lihpih* matches both vowels and the medial /p/, but the /q/ to /h/ correspondence is neither regular nor indicative of a borrowing relationship. The Matu'uwal form looks like an amalgamation of the two, but so

¹⁹One or two depending on whether the vowel correspondence in the initial syllable is regular. Atayalic vowel correspondences beyond the final foot have not been discussed in linguistic literature.

far there is not enough evidence to match it with Saisiyat or prove that it is an inherited item that just happens to resemble the Saisiyat form.

macay ‘ripe.’ Other Atayal dialects use a completely different root for this meaning: Plngawan *mahuʔil*, S’uli *məhuʔin*, Squliq *məhuqil*, Skikun *nuqil*, Klesan *məhoʔij*; from these we can safely reconstruct Proto-Atayal **mahuqil* (also meaning ‘dead’). Saisiyat *asʔasay* seems to be connected, but probably not as a source of borrowing. There are no cognates of Matu’uwal *macay* in other Atayal dialects, or in Seediq, so it is unlikely to be an inherited form. Saisiyat /s/ and Matu’uwal /c/ are both reflexes of PAN *C. The Matu’uwal lexeme must have been borrowed from a language that reflected PAN *C as /c/, so neither Pazih nor Saisiyat.

marahum ‘bruise,’ also Matu’aw **marahum**. Saisiyat *rahəm*. Most other Atayal dialects have some reflex of Proto-Atayal **lasug*: Plngawan *tulasuw*, Squliq *lasuw*, Skikun *lasux*. Saisiyat has a /ə/ in the final syllable, which is repaired in Atayal. However, its presence is attested by the alternation in the Matu’aw suffixed form *mataraham-an* ‘to bruise s.o. [LV],’ where the /u/ vowel alternates with /a/ after suffixation, as expected of historic *ə. This is in contrast with the phonetically similar inherited Matu’aw *parahum* or *parahuman* ‘lips’ (< PA **parahum*), where the /u/ vowel is retained after suffixation.

mug(u)malak ‘young man.’ This word could potentially be a retention of PAN **aNak* ‘child,’ however no other dialect appears to have any cognate, and thus it cannot be reconstructed to Proto-Atayal. However, it appears to share both form and meaning with PAN **aNak* ‘child,’ thus it is very likely a loan. Saisiyat *alʔalak* ‘young’ is a possible source. The *mugu-* (and perhaps also *m-*) part may be a prefix, cf. *mag-* in *magbatunux* ‘pretty, beautiful’ and *magnabalay* ‘fat.’

parnah ‘boat,’ also Matu’aw **paranah**. Although the Atayal live inland, their settlements are usually situated near creeks and rivulets. There is a separate term for boats that may be reconstructable to Proto-Atayal **qacuʔ*: Matu’uwal *qacuʔ*, Plngawan *asuʔ*, S’uli *asu* (also *tipay* ‘raft,’ possibly a loan from Taiwanese Southern Min [TSM] 竹排 *tik-pâi*), Squliq *qasuʔ*, Skikun *kasuʔ*, Klesan *kasu* (the latter most likely a loan from Squliq, see section 4.2). The more commonly used term in Matu’uwal appears to be *parnah*, although the semantic distinction between the two is unclear. Matu’uwal *parnah* has no cognates in other Atayal dialects (except for Matu’aw), but is phonetically very similar to Pazih *paranah*, albeit with an irregular correspondence of the medial vowel. The Matu’aw medial vowel need not be a retention of Pazih: Matu’aw requires all non-final syl-

lables to be open, and the default epenthetic vowel is /a/.

rawlu? ‘tortoise,’ also Matu’aw **rawlu?**. This word is not reconstructable to Proto-Atayal, and is not found in other Atayal dialects. Pazih *raul* is a perfect match (note that Pazih has an unwritten final glottal stop); cf. also Saisiyat *rakolo?*, which is likely more distantly related. Cf. Plngawan *kame* (< Japanese *kame*), S’uli *sipa*, Squliq *kəsipa?*, Klesan *rakus*. Note that this refers to land tortoises, see also Matu’aw *kahat* ‘softshell turtle,’ which is itself a loan from Matu’uwal.

rilum ‘pumpkin.’ Saisiyat *lilom*. Cf. Plngawan *tamurak*, Matu’aw *baun*, S’uli *bon* or *aban*, Squliq and Skikun *qaban*, Klesan *kaban*. (Both Plngawan and Klesan loaned their respective terms, see section 4.1 and section 4.2). Squashes are a New World plant, and could have only appeared in Taiwan during the last several centuries. It would be expected for this word to be a loan. The liquid correspondences are irregular: Saisiyat does have an /r/ phoneme but it does not appear in this lexeme. Matu’uwal does allow two identical liquids in the same morpheme, so this is not a case of liquid dissimilation in Matu’uwal. One possibility is that this word was borrowed from a different language into both Saisiyat and Matu’uwal. Another is that Saisiyat underwent a liquid assimilation process after this word was borrowed into Matu’uwal.

ruwas ‘paper.’ Formosans did not use paper until recent times, and did not have words for it. The Matu’uwal lexeme is likely connected to Saisiyat *zinowas* ‘paper,’²⁰ which may be etymologically related to *zowas* ‘Hibiscus taiwanensis.’

sababul ‘duck,’ also Matu’aw **sababun**. Pazih *sibabun*. Ducks are not native to Taiwan, and were introduced after Han colonisation.²¹ Cf. Plngawan *bibiy* or *kukuk*, S’uli *babun* or *kawayay*, Squliq and Skikun *guru?*, Klesan *guru*. We can identify two groups within Atayal: the *sababun* cluster (Matu’uwal, Matu’aw, S’uli), and the *guru?* cluster (Squliq, Skikun, Klesan). Note that the dialects within each of these two clusters do not form subgroups, but are geographically close to each other. The forms in the *sababun* cluster likely originate with Pazih *sibabun*. S’uli and Matu’aw are spoken south of Matu’uwal, and S’uli bordered Pazih in the past. The final /l/ in Matu’uwal is irregular, and could be a later

²⁰This is the Tungho form. In the Ta’ai dialect, the word is *zinoLas*. I thank an anonymous reviewer for sharing this with me.

²¹Chen Di, *Dong Fan Ji*, (1603): “There are no horses, donkeys, cattle, sheep, geese, or ducks [in Taiwan]” (translation mine). (Sourced from Chou 2003:44–45).

hypercorrection. The initial vowel is different between the Atayal forms and Pazih, which is an issue, see also *tigubu?* ‘sugarcane’ for a similar example.

salaman ‘bowl.’ This word is phonetically identical to Pazih *salaman*. Among other Atayal dialects, a similar form is found only in Klesan *salaman*, but meaning ‘large jar.’ Klesan is currently spoken in Yilan, far away from both Pazih and Matu’uwal. However, the ancestors of Klesan speakers lived in northern Nantou, geographically close to Pazih/Kaxabu, before migrating to Yilan some 200-300 years ago (Mabuchi 1953-54:95). It is therefore plausible that Klesan loaned this word before their migration, and it subsequently shifted in meaning.

siyatu? ‘clothes.’ Cf. Squliq, Skikun, Plngawan, Klesan *lukus*, S’uli *lukus* or *latan*, Matu’aw *balatan*. Matu’uwal *siyatu?* is phonetically identical to Pazih *siatu*. Interestingly, this word also has a male register form in Matu’uwal: *situwiŋ*. Not only did this loan completely replace the original Atayalic lexeme, it was nativized to such an extent as to enter the domain of male-female register dichotomy. This also proves that the lexical gender register system was productive in Matu’uwal even after it split from the Northern Atayal subgroup.

tabuwil ‘cucumber.’ This is a perfect match with Saisiyat *tabowil*. Some Atayal dialects have similar terms: Skikun *təbil*, Squliq *təbwil* or *təgwil*. Cf. Plngawan *ciyak*, Matu’aw *kalibuy* or *tahabu?*, S’uli *kəlibuy*, Klesan *kəmi*, also Truku *təbuwir*. Note that although Matu’uwal, Skikun, and Squliq form a subgroup within Atayal (called Northern Atayal), the sound correspondences in Skikun are not regular, and this word cannot be reconstructed to Proto-Atayal level. At the same time, the Northern Atayal dialects share phonetically similar forms, even though they are not confined to one geographical area. There are at least three possible borrowing scenarios: (1) a loan at an earlier stage, from Saisiyat into the Northern Atayal group before its split; or (2) a loan from Saisiyat into Matu’uwal at a later stage, disseminated separately to Squliq and Skikun. (3) a later loan from Matu’uwal into Saisiyat. Cucumbers began to be cultivated in Taiwan relatively recently, so a migration route from a coastal area further inland is more likely, as with other agricultural produce. The first scenario is hampered by the irregular reflex in Skikun. The second scenario is possible, however the distribution of this isogloss aligns too well with the Northern Atayal subgroup, instead of other Atayal dialects which are geographically much closer to Saisiyat, such as S’uli or Matu’aw. In the third scenario, if we assume an early loan into Northern Atayal that was later loaned into Saisiyat,

we do not have any plausible candidates for the source (the Truku form is similar, but it might just as well be a loan from Atayal, otherwise we would expect it to be loaned to Plngawan first). A connection here is self-evident, however in this case the question of directionality cannot be easily answered. I believe a loan into Northern Atayal is more likely due to a clear distribution inside the clade.

tatupun ‘k.o. mushroom.’ This is an almost perfect match with Pazih *tatupun*. None of the other Atayal dialects seem to have this word, although the names of wild mushrooms are somewhat specialized vocabulary and difficult to elicit. If these forms were inherited from a common ancestor, we would expect the place of articulation of the nasal to be identical. Atayal has few readily identifiable Austronesian cognates, and, due to the gender register system, many roots have been unpredictably mutated (Li 1983). However, identifiable cognates faithfully preserve the place of articulation for nasals at the Proto-Atayal level, including in coda position: PAN *zalan > Proto-Atayal *raʌan ‘road,’ PAN *tinun > Proto-Atayal *t<um>inun ‘to weave cloth,’ PAN *luNuŋ > Proto-Atayal *ʌluŋ ‘cloud, fog,’ PAN *Nusuŋ > Proto-Atayal *luhuŋ ‘mortar’ (more examples for PAN > Proto-Atayalic can be found in Li (1981:273–275)). All Atayal dialects preserve Proto-Atayal(ic) *n and *ŋ unchanged (Li 1981:266–267, Goderich 2020:89–90).²² Likewise, Pazih preserves PAN *n and *ŋ as /n/ and /ŋ/ (Li and Tsuchida 2001:6–9). The lack of a cognate in other Atayal dialects coupled with the discrepancy in the final nasal are indicative of a borrowing relationship.

tigubu? ‘sugarcane.’ The reconstructed PAN protoforms are *CebuS/*tebuS. Other Atayal dialects appear to have regular reflexes: cf. Plngawan *cabilus*, Suliq and S’uli *bilus*, Skikun *libus* (with metathesis), Klesan *cyubus*. Plngawan (and possibly also Klesan) point to an initial *c, in which case the Proto-Atayal would be *c[æ]bilus.²³ The Matu’uwal form has initial /t/ instead, and is thus not compatible with the other Atayal dialects. The Pazih word for sugarcane is *tubus*, with a derived term *tuxubus* ‘sweet,’ which is semantically compatible and phonetically similar. Pazih /x/ was mapped to Matu’uwal /g/ [ɣ], presumably because the former mostly occurs word-finally. The fact that both phonemes are regular reflexes of PAN *R is incidental. There are still

²²With the possible exception of Klesan (Goderich 2020:89).

²³The Plngawan reflex of the vowel in the initial syllable could come from either *a or *ə. The extra syllable in the Proto-Atayal form is an infix used to derive the male register form, see Li (1982a) for more information.

two problems with this comparison: (1) the vowels in the initial syllables do not match, and (2) Matu’uwal lacks word-final /s/. So far Pazih is the closest match to Matu’uwal, but it still has issues that need to be resolved.

tunu? ‘brain.’ Most Atayal dialects, and Seediq as well, use the same word for both ‘brain’ and ‘marrow’: Plngawan and Klesan *lu?iŋ*, Seediq *luqi?*. Matu’uwal uses *tunu?* for the specific meaning of ‘brain,’ and it is identical to Pazih *tunu* and Saisiyat *tono?*. Some Squliq speakers also use *tunu?* to mean ‘brain,’ but in light of evidence from other dialects and Seediq, it is likely to be a loanword there as well.

xuxu? ‘milk, breasts.’ This word is phonetically similar to Saisiyat *hoehoe?*, although with velar /x/ instead of pharyngeal /h/ for Saisiyat glottal /h/. All other Atayal dialects use *bubu?*, and both Seediq and Pazih use *nunuh* for this meaning, so it is highly unlikely that Matu’uwal *xuxu?* is an inherited form. It does bear a resemblance to PAN **susu*, however PAN **s* is only reflected as velar /x/ word-finally, and as pharyngeal /h/ in all other positions (Goderich 2020:143). Matu’uwal only has two words with initial /x/, the other being *xuwil* ‘dog,’ which makes it even more peculiar. There are no other languages in the area with a reflex of PAN **susu* for this meaning, including the plains languages Taokas, Babuza, and Papora.

4.4 Matu’uwal loanwords in Matu’aw

Matu’aw has borrowed a number of lexical items from Matu’uwal, many of them animal and plant names, but also cultural items and other miscellanea. An additional six lexical items than are borrowings in both Matu’uwal and Matu’aw are listed in section 4.3.

butul ‘masked palm civet.’ Matu’uwal *butul*. This could be a loanword, or a shared retention of a possible Proto-Atayal form. Cf. Plngawan *bitux*, Squliq *qəbux* or *butul*, Skikun *qəbux* or *bətuɣux*, Klesan *təli*, S’uli *bətəli*. Both Skikun *bətuɣux* and S’uli *bətəli* (together with Klesan *təli*) appear to have different male register suffixes: *-yux* in Skikun and *-i* in S’uli (*-i?* or *-iq* in other dialects).²⁴ The original Proto-Atayal female form could be **butul*, which conforms to the canonical Austronesian root shape CVCVC (Blust 2013:212). There are two pos-

²⁴These suffixes can also be seen in e.g. Squliq *ramu?* vs Matu’aw *ramuyux* ‘blood,’ and Matu’uwal *s<um>bil* vs Plngawan *s<un>bili?* ‘to leave over (AV).’ For more examples, see Li (1982a, 1983).

sible scenarios that explain the Matu'aw form: (1) it is a retention of the original Proto-Atayal form, or (2) it is a loan from Matu'uwal that displaced an earlier suffixed form. Since Matu'aw, S'uli, and Klesan together form the Nuclear Southern Atayal subgroup (Goderich 2020:193), and since both S'uli *bətəli* and Klesan *təli* share a male register suffix, the second scenario appears just as likely as the first.

kahat 'softshell turtle.' Matu'uwal *kahat*. No other Atayal dialect appears to distinguish softshell turtles from other kinds of turtles.²⁵ Cf. S'uli *sipa*, Squliq *kəsipaʔ*, Plngawan *kame* (from Japanese *kame*), also Klesan *binuk* 'sea turtle' and *rakus* 'tortoise.' See also Matu'aw and Matu'uwal *rawluʔ* 'tortoise' in section 4.3.

kakatiʔ 'ferret-badger.' Matu'uwal *qaqtiʔ*, cf. Squliq *qəciʔ* 'Siberian weasel.' Irregular correspondence of /q/ to /k/. Note the epenthetic vowel in Matu'aw.

kakusun 'small shrimp.' Matu'uwal *kakusun*. This is distinct from the normal word for 'shrimp,' which is *kakuwas* in Matu'uwal and *bawluŋ* in Matu'aw. Cf. S'uli *bolun*, Squliq and Klesan *kəbolun*, Plngawan *balulun* 'shrimp.' There appear to be no cognates in other Atayal dialects.

kalaynan 'dragonfly.' Matu'uwal *qaqlaynan*. There is an irregular correspondence in the initial consonant: Matu'uwal /q/ to Matu'aw /k/. Note that Matu'uwal has Ca-reduplication (common in many animal names in Matu'uwal), but Matu'aw does not. S'uli *leŋun* might be related, but the vowel in the final syllable does not match. Cannot be reconstructed to Proto-Atayal: cf. Plngawan *lalosan*, Squliq *kəlosan*, Skikun *tun*, Klesan *tombo* (the latter loaned from Japanese *tonbo*).

kapinapinas 'bird sp. (black bulbul).' Matu'uwal *kapinapinas*, cf. S'uli *kəsipin*, Squliq *kəsəypil*, Plngawan *cacipil*. Perfect match with Matu'uwal, uniquely shared.

kasahuy 'snake sp. (Taiwanese krait).' Matu'uwal *qacahuw*. Cf. S'uli *ləmiʔuy*, Klesan *niʔuy*, Squliq *kinpahuw*. Irregular correspondence of /q/ to /k/. The final glides in Matu'uwal and Matu'aw point to a historical *ɿ. This may be an early loan, from a time when Matu'uwal still preserved Proto-Atayal *ɿ as /ɿ/.²⁶ Note that Matu'uwal *qacahuw* has a second meaning, 'internal organs, guts,' for

²⁵Klesan has separate words for sea turtles and tortoises, which is a different distinction.

²⁶An extinct variety of Atayal very closely related to Matu'uwal had different reflexes of Proto-Atayal *ɿ (Goderich 2021:469–470), so this scenario is plausible.

which Matu'aw has a cognate *asahuy*, with a regular correspondence of /q/ to /ʔ/.

kawbay 'spider.' Matu'uwal *kawbay*. Perfect match with Matu'uwal. This word cannot be reconstructed to Proto-Atayal: cf. S'uli *kyukit*, Plngawan *nakarit*, Squliq *yokay*, Skikun *kasuʔ*, Klesan *kuy koyay* (note that these are all generic terms for spiders; there appear to be no distinctions between various kinds of spiders in any Atayal dialect).

kawran 'k.o. bamboo (*Dendrocalamus latiflorus*). Matu'uwal *qawran*. Irregular correspondence of /q/ to /k/ indicates a borrowing relationship. Cf. S'uli and Squliq *takan*, Plngawan *batakan*. Some S'uli speakers also use *oran* to refer to a similar variety of bamboo. The latter is a cognate of Matu'uwal *qawran*, with a regular /ʔ/ to /q/ correspondence.

kumwak 'spicy.' Matu'uwal *kumuwak*. This root is shared exclusively between Matu'uwal and Matu'aw. Cf. S'uli and Skikun *meluh*, Squliq and Klesan *meloh*, Plngawan *mañihu* or *saʔuik*. The lack of a S'uli cognate points unambiguously to a Matu'uwal → Matu'aw borrowing directionality.

mahalay 'lame.' Matu'uwal *mahlay*. Uniquely shared between Matu'aw and Matu'uwal. The epenthetic /a/ in Matu'aw repairs an infelicitous syllabic structure. Cf. S'uli *məpika*, Plngawan *mapikaʔ*, Squliq *pikaʔ*, also Seediq *pika*.²⁷

manabalay 'fat, stout.' Matu'uwal *magnabalay*. This word, uniquely shared with Matu'uwal, is used to describe people. Cf. S'uli *kəʔəhuy*, Plngawan *katuh*, Squliq and Skikun *qəʔəhuy*, Klesan *təhuy*, meaning 'thick,' which have cognates in Matu'uwal *kithuw* and Matu'aw *katahuy* (but in the latter two are normally reserved for inanimate objects). The lack of a S'uli cognate points unambiguously to a Matu'uwal → Matu'aw borrowing directionality.

mar(u)bawak 'weasel sp.,' possibly Siberian weasel. Matu'uwal *marbawwak*. Irregular correspondence of Matu'uwal genitive /w/ (which indicates a co-occurrence of Proto-Atayal *ɿ and *w). Note the regular correspondence in Matu'uwal *bawwak* and Matu'aw *baywak*, from Proto-Atayal *baɿwak 'pig,' and likely the root in this word.

pilaw 'next year.' Matu'uwal *pilag*. Other Atayal dialects do not have a specific

²⁷Despite the similarity of *pikaʔ* in form and meaning to TSM 跛跛 *pái-kha* 'lame', the reflexes in Atayal and Seediq point to Proto-Atayalic *mapikaʔ. Since the form can be reconstructed to Proto-Atayalic, the resemblance with TSM is likely accidental.

word for this, and instead use either a periphrastic construction meaning ‘the coming year,’ such as S’uli *kayan ʔənyan*, Squliq *kawas ʔənyal*, Klesan *mwah na kawas*, or just use the word for ‘year,’ with its meaning understood from context: Skikun *kawas*, Plngawan *inʔkaʔalan*. Note that Matu’uwal *pilag* is used together with a temporal marker in two situations: *cu pilag* ‘last year,’ and *i pilag* ‘next year.’ Matu’aw uses *pilaw* only with the meaning ‘next year,’ whereas ‘last year’ is expressed using the periphrastic construction *kayal ka wayal*. Matu’aw used to have /g/ in word-final position, but has since lenited it to /w/ (Li 1981, 1982b).

pilapilaw ‘a while, [for] a period of time.’ Matu’uwal *pilapilag*. This is a reduplication of *pilaw* ‘next year.’ Other Atayal dialects tend to use periphrastic constructions here, such as S’uli *misu wayan*, Squliq *ini kəzəŋaʔ*, Skikun *ini na ga*, Klesan *micu ci* (cf. also Plngawan *mingagos*). Alternatively, they might use the versatile adverb *tikay* or *cikay* ‘a little,’ which is extremely common and used in a variety of situations.

rinwanj ‘sweat.’ Matu’uwal *rinuwanj*, cf. S’uli *rənanj*, Plngawan *rinanj*, Skikun *rinanj*, Klesan *rənan*. While the forms in all dialects are obviously related, Matu’aw and Matu’uwal uniquely share a /w/ that is not found anywhere else.

sapanj ‘old (of things).’ Matu’uwal *capanj*. Uniquely shared with Matu’uwal. Cf. S’uli *səswin* or *səmurān*, Plngawan *cacil*, Squliq *səmural*, Skikun *səmurān*, Klesan *muranj*.

siŋak ‘food debris,’ i.e. food particles stuck between teeth. Matu’uwal *ciŋaq*. Cf. Plngawan *ciŋas*, S’uli *siŋas* < Proto-Atayal **ciŋas*, itself from Proto-Austronesian **Ciŋas*. Matu’uwal has a completely unrelated final consonant due to it being a derived male register form (the corresponding female form appears to have been lost). Crucially, there is also an irregular correspondence of Matu’uwal /q/ to Matu’aw /k/.

tatabul ‘plough.’ Matu’uwal *tatabul*. Cf. S’uli *kawin*. Cattle and horses were introduced to Taiwan after Han colonization.²⁸ Ploughs were likewise unknown before that time, and no word for ‘plough’ can be reconstructed to Proto-Atayal. Some Atayal dialects chose various roots meaning ‘hoe’ to designate ploughs as well, e.g. Plngawan *ot na wanux*, lit. ‘cow hoe.’²⁹ Matu’uwal

²⁸See footnote 21.

²⁹Since cattle were unknown in Proto-Atayal, they are referred to in modern dialects by reflexes of *wVqanux, meaning ‘sambar deer.’

and Matu'aw follow a similar strategy, choosing the verb *t<um>abul* 'to hoe, to till,' and deriving from it an instrumental noun using Ca-reduplication. Since Matu'aw speakers live in a more isolated location, and because S'uli uses a different word for 'plough,' this is likely a Matu'awal innovation that was readily adopted by Matu'aw speakers who already had this root in their lexicon.

5 Conclusion

Loanword research often begins with the more obvious borrowings: words that can be easily identified by the researcher or the speakers as originating in a different language. Lexical borrowings are especially apparent when they come from a separate language family, and the phonology, phonotactics, and morphology make identification simple. Loanwords between closely related languages are a completely different proposition, even more so when they were borrowed some time ago, and are no longer recognized as loanwords by the speakers. It should come as no surprise that the biggest influence on each of the four dialects in the study was from their immediate neighbours.

It is not just the fact of borrowing itself, but also the kinds of words that were being borrowed. Cultural vocabulary, such as new technologies, implements, and tools, is expected to be borrowed regularly. The same can be said for newly introduced flora and fauna, such as New World agricultural items like pumpkins or guavas, or previously unknown domesticated animals like ducks. Crucially, in all four dialects we see borrowing of words for native flora and fauna, mostly species names but also a few general terms; there are even a few terms for body parts. While the majority of the loans are nouns, verbs (active or stative) were borrowed in every dialect as well.

The loanwords in Matu'aw and Klesan are the most straightforward. In part, they are easier to recognize because there are obvious diagnostics for both dialects, making irregular sound correspondences apparent. In both cases they come from within the Atayal group itself, and are probably comparatively recent. Both dialects belong to the Southern Atayal subgroup, while their donors are from the Northern subgroup, so lexical borrowings can be double-checked against other Southern Atayal dialects. For these reasons, the certainty of borrowing in the case of both Matu'aw and Klesan is very high.

Pngawan is an in-between case. On the one hand, the influence of Seediq on this dialect is undeniable: in fact, it is so profound that it necessitated academic papers to ascertain Pngawan as an Atayal dialect (Li 1985). On the other hand, Pngawan lexical items that match Seediq sometimes have regular sound correspondences, even though they are not found in any other Atayal dialects. I infer that this dichotomy between regular and irregular sound correspondences in exclusively shared vocabulary is a product of a long history of contact between Pngawan and nearby Seediq varieties.

Many lexical items in the Matu'uwal list are frustratingly inconclusive. They bear a clear resemblance in both form and meaning to the proposed source forms in Pazih or Saisiyat, but there are slight imperfections in the phonetic match-up. Approximately half of the items are somewhat tentative. Nevertheless, there is still merit in shedding light upon these comparisons, flawed though they may be. Matu'uwal and Pngawan are the two most lexically divergent Atayal dialects, and one of the sources of this divergence must have been language contact. There are several possible reasons for the imperfections in the loans into Matu'uwal:

1. A greater time depth than other contact situations, with sound changes in the source or the recipient language.

This is a possibility for words like *tigubu?* 'sugarcane' or *hamhum* 'cloud,' where the currently recorded forms do not phonetically align with Matu'uwal. However, there is a ceiling for this time depth: it must necessarily follow the split of Matu'uwal from the rest of the Northern Atayal subgroup.

2. Being filtered through an intermediate language, or loans originating with a different language and being borrowed into Matu'uwal and Pazih/Saisiyat separately.

This may account for inexplicable mismatches in words like *tatupun* 'k.o. mushroom' or *kapuwa* 'guava.' Unfortunately, the only candidates for this are the Western Plains languages, which are extinct, with very little (and rather low-quality) data left.

3. Phonetic adaptations and the lexical gender register in Matu'uwal.

Since the phonologies of Pazih, Saisiyat, and Matu'uwal are rather distinct, lexical borrowing would necessitate a certain amount of phonolog-

ical adaptation. This is the case with, e.g., the vowel *marahum* ‘bruise.’ Additionally, the lexical gender register was still productive during the time that Matu’uwal was borrowing words from Pazih, as evidenced by *siyatu?~situwiŋ* ‘clothes.’ A similar process may have happened in *lihpiq* ‘thin,’ with the original form being lost.

However, the above is speculation, and somewhat unsatisfying. What we can know for sure is, these lexical items are confined to a single geographical region, and not shared with other Atayal dialects, including the words with flawed sound correspondences. Whatever their precise nature, there can be little doubt that they are all the result of language contact in one form or another.

These findings have implications for how we perform higher-level linguistic reconstructions with Austronesian data. Of the borrowings in this study, we can date loans into Plngawan, Klesan, and Matu’aw to the last few hundred years. (Borrowing into Matu’uwal may have occurred earlier.) Taiwan has been home to Austronesian speakers for about 5000 years, and, small as the island is, linguistic contact and lexical borrowings have permeated every single one of the Formosan languages. Since essentially nothing is off-limits when it comes to language contact, we have to be especially careful when collecting and analyzing Formosan data, because much of what we see now has been obscured by millennia of change, contact, and replacement.

Linguistic reconstruction is perforce a cyclical process. Initial phoneme inventories and protoform glossaries will undeniably be contaminated by lexical borrowing, and require further refinement down the road. Since nine out of ten primary branches of Austronesian are found only in Taiwan (Blust 1999b, 2013), loanwords between Formosan languages can severely skew Proto-Austronesian reconstructions, and our picture of the family as a whole. Working from the ground up to weed out lexical borrowings in Formosan languages will lead to a more accurate reconstruction of Proto-Austronesian.

Acknowledgements

The research in this paper was funded by NSTC grant 111-2410-H-018-027. An early version of this paper was presented at the 2022 Historical Relationships of East and Southeast Asian Languages Conference (HRESAL), National Tsing

Hua University, Hsinchu, where I received a number of helpful comments. I would also like to thank anonymous reviewers for their suggestions and questions, which have helped improve the paper tremendously. Additional thanks go to Walis Hian-chi Song for lending me his expertise in Seediq. Any remaining errors are my own.

References

- Blust, Robert. 1999a. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38.2:321–365.
- Blust, Robert. 1999b. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Taipei: Academia Sinica.
- Blust, Robert. 2013. *The Austronesian Languages* (Asia-Pacific Linguistics Open Access Monographs). Canberra: Australian National University, College of Asia and the Pacific.
- Blust, Robert, and Stephen Trussel. Ongoing. Austronesian comparative dictionary, web edition.
- Chen, Christina Yin-Ling. 2011. *Issues in the Phonology of Ilan Atayal*. PhD thesis, National Tsing Hua University.
- Chien, Yuehchen, and Shinji Sanada. 2010. Yilan creole in Taiwan. *Journal of Pidgin and Creole Languages* 25:350–357.
- Chou, Wan-Yao. 2003. 陳第〈東番記〉—十七世紀初臺灣西南地區的實地調查報告 [Chen Di's "Dong Fan Ji": An on-site investigation report in southwest Taiwan from the beginning of the 17th century]. *National Palace Museum Monthly of Chinese Art* 241:22–45.
- Egerod, Søren. 1980. *Atayal-English Dictionary* (Scandinavian Institute of Asian Studies Monograph Series 35). London: Curzon Press.
- Egerod, Søren. 1999. *Atayal-English Dictionary* (Historisk-Filosofiske Skrifter 20). Ed. by Jen Østergaard Petersen. 2nd edition. Copenhagen: The Royal Danish Academy of Sciences; Letters.
- Ferrell, Raleigh. 1969. *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification* (Institute of Ethnology, Academia Sinica, Monograph), vol. 17. Taipei: Academia Sinica.
- Ferrell, Raleigh. 1982. *Paiwan Dictionary* (Pacific Linguistics C-73). Canberra: Australian National University.

- Goderich, Andre. 2020. *Atayal Phonology, Reconstruction, and Subgrouping*. PhD dissertation, National Tsing Hua University, Hsinchu.
- Goderich, Andre. 2021. Taokas-10 revisited: Taokas or Atayal? *Oceanic Linguistics* 60.2:447–473.
- Hetay Payan. 2018. *Encyclopedia of Atayal Culture [in Chinese]*. Taipei: Hetay Payan.
- Huang, Hui-chuan J. 2006. Squliq Atayal syllable onset: Simple or complex? *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, 489–505. Taipei: Institute of Linguistics, Academia Sinica.
- Huang, Hui-chuan J. 2015a. The phonemic status of /z/ in Squliq Atayal revisited. *Capturing Phonological Shades Within and Across Languages*, ed. by Yuchau E. Hsiao and Lian-Hee Wee, 243–265. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Huang, Hui-chuan J. 2015b. Syllable types in Bunun, Saisiyat, and Atayal. *New Advances in Formosan Linguistics* (Studies on Austronesian Languages 003), ed. by Elizabeth Zeitoun, Stacy F. Teng, and Joy J. Wu, 47–74. Asia-Pacific Linguistics.
- Huang, Hui-chuan J. 2018. The nature of pretonic weak vowels in Squliq Atayal. *Oceanic Linguistics* 57.2:265–288.
- Li, Paul Jen-kuei. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 49.2:133–199.
- 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. 1982a. 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. 1982b. Atayalic final voiced stops. *Papers from the Third International Conference on Austronesian Linguistics* (Pacific Linguistics C-75), ed. by Amran Halim, Lois Carrington, and Stephen Wurm, vol. 2: Tracking the travellers, 171–185. Canberra: Australian National University.
- 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.

- Li, Paul Jen-kuei. 1991. Vowel deletion and vowel assimilation in Sediq. *Currents in Pacific Linguistics: Papers on Austronesian Languages and Ethnolinguistics in Honour of George W. Grace* (C 117), ed. by Robert Blust, 163–169. Pacific Linguistics. doi:10.15144/PL-C117.163.
- Li, Paul Jen-kuei, and Shigeru Tsuchida. 2001. *Pazih Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Mabuchi, Tōichi. 1953–54. 〈高砂族の移動および分布〉 [Migration and distribution of the Formosan native tribes]. *Minzokugaku Kenkyū* 18.1/2:123–154, 18.4:23–72. Translated into Chinese and annotated by 楊南郡 as 《台灣原住民族移動與分布》, 2014, Taipei: SMC Books.
- Ochiai, Izumi. 2019. Fossilized person marker in Atayal. Paper presented at the 29th Annual Meeting of the Southeast Asian Linguistic Society (SEALS-29), 27–29 May 2019, Tokyo, Japan.
- 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.
- Shibata, Kye. 2020. *A Reconstruction of Proto-Bunun Phonology and Lexicon*. Master's thesis, National Tsing Hua University.
- Sung, Li-May. 2018. *An Introduction to Seediq Grammar*. 2nd edition. New Taipei City: Council of Indigenous Peoples.
- Tsuchida, Shigeru. 1982. *A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan Part I: West Taiwan* (Memoirs of the Faculty of Letters 7). Tokyo: University of Tokyo.
- Tsukida, Naomi. 2005. Seediq. *The Austronesian Languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 291–325. Abingdon: Routledge.
- Wu, Wendy Ko-shin. 2014. *Language Contact and Lexical Borrowings in Atayal*. Master's thesis, National Tsing Hua University, Hsinchu.
- Yeh, Marie Meili. 2018. *An Introduction to Saisiyat Grammar*. 2nd edition. New Taipei City: Council of Indigenous Peoples.
- Zeitoun, Elizabeth, and Chen-huei Wu. 2005. Saisiyat reduplication revisited. *Concentric: Studies in Linguistics* 31.2:31–56.