

## Proto-Austronesian \*lima revisited: From archaic “hand” in Atayalic languages\*

**Izumi Ochiai**  
**Hokkaido University**

**Abstract:** This study argues that Proto-Austronesian (PAN) \*lima, which is supposed to mean “hand” and “five,” originally meant “hand” and that its meaning is extended to “five” on the basis of the historical investigation of \*lima in Atayalic languages and semantically related forms such as “shoulder,” “to wash hands,” and “to wash oneself.” Atayalic languages lack \*lima meaning “hand.” However, this paper proposes that they are retained in a derived form meaning “to wash oneself” in which the reflexes of \*lima are attached with a prefix (*m-* in Atayal and *tu-* in Seediq). The connection between “to wash oneself” and “to wash hands” is verified as a near-cognate in other Formosan languages such as Saisiyat, Pazih, Bunun, Paiwan, and Rukai. The innovative word for “hand” in Atayalic languages comes from PAN \*qabaRa “shoulder.” Atayalic languages retained \*lima with the meaning of “hand” in a derived form. In Atayalic languages, \*lima as “hand” became a kind of fossilized root. This suggests that the meaning of “hand” is older than “five.”

**Key words:** Austronesian, \*lima, etymology, Atayal, Seediq

### 1. Introduction —“hand” and “five” in Formosan languages (Endo 2011)

There are approximately 20 Austronesian languages in Taiwan: Atayalic subgroup (Atayal and Seediq), Saisiyat, Taokas, Babuza, Pazih, Bunun, Hoanya, Siraya, Tsouic subgroup (Tsou, Kanakanavu, Saaroa), Rukai, Paiwan, Puyuma, Amis, Kavalan, and Basay. These languages are collectively referred to as Formosan languages and comprise all the Austronesian languages spoken in Taiwan, barring Yami.<sup>1</sup>

In Proto-Austronesian, the word for “hand” is reconstructed as \*lima (Blust 1999a: 82)<sup>2</sup>, which is reflected in many present-day Austronesian languages, including Formosan languages.<sup>3</sup> PAN \*lima means “hand” as well as “five”<sup>4</sup> (Dahl 1981:49). The semantic connection between the two is evident given that a hand has five fingers.

Endo (2011) investigated the forms of “hand” and “five” in Formosan languages. He produced the following linguistic maps specifying variant forms and their location: distribution of various forms of “hand,” distribution of “hand” originating in PAN \*lima,

distribution of various forms of “five,” and distribution of a combination of the forms of “hand” and “five.” These maps show that most languages exhibit the forms that reflect PAN \*lima either for “hand” or “five.” However, languages that show different forms for either “hand” or “five” rather than \*lima are in existence. Table 1 presents the list of words representing “hand” and “five” in Formosan languages.<sup>5</sup> In the table, forms that do not reflect PAN \*lima are highlighted in boldface.

**Table 1. Forms of “hand” and “five” in the Formosan languages**

Language	“hand”	“five”
Atayal	<b><i>qəba</i></b>	<i>ima-gal</i>
Seediq	<b><i>baga</i></b>	<i>rima</i>
Saisiyat	<i>ima/Rima</i> <sup>6</sup>	<b><i>asəb/Rasəb</i></b>
Taokas	<i>lima</i>	<b><i>hasap</i></b>
Pazih	<i>rima</i>	<b><i>xasəp</i></b>
Babuza	<i>rima</i>	<b><i>achab</i></b>
Tsou	<b><i>mucu/emucu</i></b>	<i>eimo</i>
Saaroa	<b><i>lamoco</i></b>	<i>k[ima</i>
Kanakanbu	<b><i>lamucu</i></b>	<i>[ima</i>
Rukai	<i>[ima/a]ima</i>	<i>[ima</i>
Paiwan	<i>[ima</i>	<i>[ima</i>
Puyuma	<i>[ima</i>	<i>[ima</i>
Amis	<b><i>kamay/kayam</i></b>	<i>[ima</i>
Bunun	<i>ima</i>	<i>ima/hima</i>
Kavalan	<i>lima</i>	<i>lima</i>
Basay	<i>tsima</i>	<i>tsjima</i>
Papora <sup>7</sup>	<i>rima</i>	<i>rima</i>
Hoanya <sup>8</sup>	<b><i>pila, pilat, pilas</i></b>	<i>lima</i>
Thao	<i>rima</i>	<i>rima</i>
Siraya	<i>rima</i>	<i>rima</i>

This study aims to provide additional data to Endo (2011) on the basis of the internal reconstruction of archaic “hand” in Atayalic languages (Section 2) and its comparison with other Formosan languages (Section 3). Moreover, this paper discusses that the innovative forms of “hand” in the Atayalic languages originate in \*qabaRa “shoulder” and suggests the historical changes that have happened to \*lima, that is, “hand” is semantically extended to

“five,” and in some languages, “hand” is replaced by an innovative form to disambiguate “hand” and “five” (Section 4). In addition, this paper revises Endo’s (2011) data and analysis for the various forms of “hand” relating to languages such as Atayal, Saisiyat, Babuza, Basay, Siraya, Amis, and Tsouic. This study also provides a linguistic map that is different from the four maps of Endo (2011). In this map, a language is classified by whether \*lima means “hand” only, “five” only, or both “hand” and “five” (Section 6). Many languages have both meanings, while some mean “hand” only and others mean “five” only.

## 2. “Five” and “to wash oneself” in the Atayalic languages

In the Atayalic languages, the words used to indicate “hand” and “five” are distinct. The Atayalic form of “five” is *ima-gal* (Egerod 1965:210), which has the added suffix *-gal* (Tsuchida 1975:258). Therefore it is the reflex of PAN \*lima (Tsuchida 1975:231). In Seediq, *rima* is used to indicate “five,” which is also the reflex of PAN \*lima.

The form of “hand” in the Atayalic languages (*qəba* in Atayal and *baga* in Seediq as in Table 1) does not reflect PAN \*lima “hand.” However, this paper proposes that the reflex of \*lima “hand” is existent in the Atayalic languages in derived forms. The words in point are *mima* (Egerod 1965:219) in Atayal and *turima* in Seediq, which mean “to wash oneself.” The hypothetical root for these forms is *ima* in Atayal and *rima* in Seediq. These forms are analyzed to have prefixes attached to them. In Atayal, the prefix is *m-*, which possibly functions as a verbalizer. In Seediq, the prefix is *tu-*, which indicates low transitivity and probably dates back to PAN \*ta- (Ochiai 2016b).

If PAN \*lima “hand” was retained in Proto-Atayalic, then the reflex would be \*rima because according to Li (1981:275) because PAN \*l becomes Proto-Atayalic \*r. Moreover, if this is reflected in the Atayalic language of the present-day (which it is not), its form would be *ima* in Atayal and *rima* in Seediq because according to Li (1981, 275), \*r is lost word-initially in Atayal.<sup>9</sup> The hypothetical roots, *ima* in Atayal and *rima* in Seediq, in the forms of “to wash oneself” represent the same forms as the regular reflexes of PAN \*lima “hand.” This paper proposes that this accordance is not coincident and that the hypothetical roots are actually the reflexes of PAN \*lima “hand,” which have barely survived in derived forms, through investigating near-cognate forms in the other Formosan languages discussed in the next section.

Table 2 exhibits the forms of “five” and “to wash oneself” with tentatively reconstructed Proto-Atayalic forms.

**Table 2. Forms of “five” and “to wash oneself” in the Atayalic languages**

	“five”	“to wash oneself”
Atayal	<i>ima-gal</i>	<i>m-ima</i>
Seediq	<i>rima</i>	<i>tu-rima</i>
Proto-Atayalic	* <i>rima</i>	* <i>m-rima</i> /* <i>ta-rima</i>

### 3. “To wash hands” in the other Formosan languages

Notably, the word for the phrase “to wash hands” is derived from the reflexes of PAN \**lima* “hand” in the Formosan languages, such as Saisiyat, Pazih, Bunun, Paiwan, and Rukai as shown in Table 3.<sup>11</sup>

**Table 3. Reflexes of PAN \**lima* in the Formosan languages**

	“hand”	“to wash hands”
Saisiyat	<i>Rima</i>	<i>ti-Rima</i>
Pazih	<i>rima</i>	<i>pana-rima</i>
Bunun	<i>ima</i>	<i>tal-ima</i>
Paiwan	[ <i>ima</i>	<i>mi-[ima</i>
Rukai	[ <i>ima</i>	<i>tawa-[ima</i>

The semantic connection between “hand” and “to wash hands” seen in Table 3 is evident. However, the semantic connection between “five” and “to wash oneself” in the Atayalic languages is not evident. Their connection becomes evident if “five” is assumed to mean “hand.” A similar word formation in Table 3 is observed in the Atayalic languages. However, the Atayalic languages differ from the other Formosan languages in Table 3 in that the part to be washed was extended from hand to whole body. These derived forms in the Atayalic languages could mean “to wash hands,” just like the other Formosan languages in the earlier time of Proto-Atayalic, but later, they are likely to have shifted meanings to “to wash oneself.” In the Atayalic languages of present-day, “to wash hands” is expressed as *mima qəba* in Atayal and *turima bəga* in Seediq (see Section 4).

It follows that the Atayalic reflexes of PAN \**lima* “hand” exist in derived forms, *mima* in Atayal and *turima* in Seediq, and both means “to wash oneself.” In these derived forms, the reflexes of \**lima*, that is *ima* in Atayal and *rima* in Seediq, have become fossilized roots. The original meaning of them is synchronically unrecoverable.

#### 4. “Hand” in the Atayalic languages and “shoulder” in Proto-Austronesian

Endo (2011) cited *kaba* as an Atayal form of “hand” although it should be *qaba* with *q* as an initial consonant as it is so in present-day Atayal. The first vowel appears to have undergone weakening to a schwa. Endo cited *baga* as a Seediq form of “hand” and analyzed that *kaba* (*qaba*, accurately) in Atayal and *baga* in Seediq come from the same word and that either of them underwent metathesis. Even though the segment *ba* (the second syllable in Atayal form and the first syllable in the Seediq form) are in common, it is difficult to explain the difference of the other consonant, *q* in Atayal and *g* in Seediq.

The two forms are actually related but not related by metathesis. They date back to a PAN \*qabaRa, which means “shoulder.”<sup>13</sup> In earlier documentation of the Seediq vocabulary recorded by Bullock (1874), “hand” is recoded as *abatha*. The *th* probably represents a kind of fricative, such as [ɣ], which may have been a variant of [g] at that time. Therefore, an earlier Seediq form for “hand” was *abaga*. Using earlier Seediq form *abaga* and Atayal cognate *qəba*, Ochiai (2016a:315–316) reconstructed Proto-Atayalic \*qabaga “hand” and stated that it is a reflex of PAN \*qabaRa “shoulder.” She also stated that the Atayalic reflex of PAN \*lima “hand” was replaced by the Proto-Atayalic reflex of PAN \*qabaRa “shoulder.” Table 4 summarizes the phonological changes from PAN \*qabaRa “shoulder” to the forms in present-day Atayal and Seediq. Atayal deleted the final syllable \*Ra from the PAN form. Another possibility is that \*R was sporadically deleted resulting in the sequence of two like vowels, *qabaa*, and then one of the like vowels was deleted. Seediq first deleted initial consonant *q* resulting in *abaga* and then deleted initial vowel *a*.<sup>14</sup>

**Table 4. Phonological and semantic change of \*qabaRa from Proto-Austronesian to the Atayalic languages**

Proto-Austronesian	*qabaRa “shoulder”
Proto-Atayalic	*qabaga “hand”
Atayal	<i>qəba</i> “hand” (< <i>qaba</i> < <i>qabaa</i> < *qabaga)
Seediq	<i>baga</i> “hand” (< <i>abaga</i> < *qabaga)

In the Atayalic languages, \*qabaRa replaced the reflexes of \*lima, *ima* in Atayal and *rima* in Seediq, meaning “hand.” This replacement was probably motivated by the need to disambiguate “hand” and “five” at the time when both meanings were expressed by the identical form \*lima. Later, this form was reserved for “five” and lost the meaning of “hand” although this meaning is retained in the derived forms that means “to wash oneself” (Section 2).

## 5. Revision of “hand” in the Formosan languages cited in Endo (2011)

Endo (2011) used Ogawa (2006)—a comparative vocabulary of the Formosan languages—as the main source of the data for “hand” and “five” in the Formosan languages. The original data of Ogawa (2006) comes from Ogawa’s manuscripts recorded in the early 20<sup>th</sup> century which contain Formosan words which were collected by himself or cited from other sources. These manuscripts of Ogawa were compiled and published by Paul Jen-kuei Li and Masayuki Toyoshima.

Ogawa (2006) needed to be treated with care because some mistakes were found in the presentation of the data. In some cases, the data do not belong to the language they supposed to belong to.<sup>15</sup> In other cases, the original source of the data must be referred to for identifying typographical errors. The data of Endo (2011) that are based on those of Ogawa (2006) also contain a few mistakes. Further, the analysis based on these data by Endo (2011) should be modified in some cases. This section provides the revisions of Endo (2011).

### 5.1. “Arm” in Atayal

Endo (2011:5) observed a form *kakimaan* “hand” in one of the Atayal data and verified that this may be a cognate with *kamay* “hand” in Amis. However, the form is related not to *kamai* but to \**lima*. The original source of the Atayal form is a manuscript by Nakanishi (1900). This form in Nakanishi was from an Atayal village situated north of Wenshui (汶水) river. The Atayal dictionary of Ogawa (1931:45) includes a similar word, *kiiman*, which means “wrist” rather than “hand.” In addition, another similar form *kapaimaʔan* (Li 1978:160) is seen in Saisiyat, and it also means “wrist” or “lower arm.” Therefore, the Atayal form *kakimaan* recorded by Nakanishi possibly meant “wrist.” These forms above (*kakimaan* and *kiiman* in Atayal and *kapaimaʔan* in Saisiyat) are derived from the reflexes of PAN \**lima* “hand,” which appears to be *ima* in Atayal and Saisiyat (see Table 1 for the Saisiyat variants, *Rima* and *ima*).<sup>16</sup> In *kakimaan*, *kaki-* is probably a prefix, and *-an* is a suffix (*kaki-ima-an* > *kak-ima-an*). Likewise, in *kiiman*, *ki-* is a prefix, and *-an* is a suffix (*ki-ima-an* > *kiima-n*). This form is identical to *kakimaan*, except for the initial syllable *ka*. In the Saisiyat form, *kapaimaʔan*, *kapa-* is a prefix, and *-ʔan* is a suffix. All forms share similar word formations that include the root *ima* and prefix (*kaki-*, *ki-*, or *kapa-*) and suffix (*-an* or *-ʔan*). These Atayal words for “arm” show another case that retains archaic *ima* “hand” in derived forms other than *mima* “to wash oneself.”

## 5.2. Babuza or Basay

Endo (2011:5) reported that *kakama* “hand” is a word in Babuza. However, according to Ogawa (2006:65), this form is classified in Basay. This form is exceptional among the Basay data. In 17 Basay forms of “hand” cited from different sources in Ogawa (2006:65), all forms show *tsima* although there were some variations in transcription. The original source of *kakama* is uncertain, but according to Ogawa (2006:xiv), the data were collected by Ino (probably, Ino Kanori) in a village called Xiulang (秀朗). Incidentally, Ino (1898:491) published a journal paper dealing with Basay, in which he recorded some basic vocabulary collected in Xiulang village and “hand” is recorded as *tsima*. Therefore, a supposedly Basay form *kakama* in Ogawa (2006:65), which is supposed to have been collected by Ino Kanori in Xiulang village, appears to be a mistake.

## 5.3. Siraya or Amis

Endo (2011:5) states that “hand” is represented as *kayam* in Jiayi (嘉義) and Taidong (臺東) without specifying the language.<sup>17</sup> The comparison of the Endo’s maps revealed that the form in Jiayi belongs to Siraya and that in Taidong belongs to Amis. The *kayam* in Siraya was collected by Ogawa, and it is cited in Tsuchida, Yamada, and Moriguchi (1991:55), in which they explain that the area in which this data was collected was Dazhuang (大庄) in Taidong.<sup>18</sup> A group of Siraya migrated from their homeland in southwestern plain to Taidong in the late 19th century according to Tashiro (1900:68–69). Endo was unaware that *kayam* in Siraya was collected in Taidong; therefore, he plotted this data in Jiayi, near Siraya homeland, in his map. This form, *kayam*, is not usually seen in other Siraya areas in which they have “hand” as *rima*, the reflex of PAN \*lima. It is likely to be due to borrowing from *kayam* “hand” in Amis situated in the southeastern coast, as it is also pointed out by Tsuchida, Yamada, and Moriguchi (1991:55).

## 5.4. Tsouic languages

Endo (2011:5) affirmed that the forms for “hand” in the Tsouic languages, such as *ramucu* in Saaroa or *muca* in Tsou, are the cognates of PAN \*lima on the basis of one of the Tsou forms, *limtsu* in Ogawa (2011:62).<sup>19</sup> He highlighted that the segment *lim* are in common in PAN \*lima and Tsou *limtsu*, and analysed that *tsu* is an added element in the Tsouic languages. However, the cognate relationship is rejected according to the explanation by Sagart (2013:482). Sagart considered Proto-Rukai-Tsouic form \*ramuCu “hand” reconstructed by Tsuchida (1975:179) and suggested that the original meaning of this form was “finger” rather than “hand” on the basis of its meaning in the Mantaaran

dialect of Rukai, *ramucu* “finger.” Sagart goes on to say that the reflex of PAN \*lima “hand” was replaced by an innovative form \*ramuCu, which originally meant “finger,” in the Tsouic languages.

### 5.5. Hoanya

Endo (2011:6) cited two Hoanya forms for “finger” recorded in Ino (1907:109): *pira* in the Arikun dialect (this form is transcribed as *pila*, *pilat*, or *pilas* by Ogawa [2006:66] in the data collected by Ogawa himself in Beitou [北投] village) and *pira mait* in the Lloa dialect. Further, Endo cited the comment of Ino that *pira* in Arikun is also the word for “hand.” Hoanya tribes have been influenced by Chinese immigrants so much that they lost many native words, leading to the confusion of “hand” and “finger.” With regard to the form in the Lloa dialect, the second word *mait* is recorded to mean “small” in the Hoanya vocabulary (Tsuchida 1982:122). “Finger” in the Lloa dialect literally means “small hand.” Ogawa (2006:66) suspected the connection of “hand” in Hoanya and “finger” in Babuza. After the variant Hoanya forms *pila*, *pilat*, or *pilas*, Ogawa wrote a note saying “cf. Favorlang “finger”.” Favorlang is another name for Babuza, and “finger” in Babuza is recorded as *appuru* or *apillo* in Ogawa (2006:70). If this is cognate with the Hoanya forms, then the forms of “hand” in Hoanya, such as *pila*, *pilat*, or *pilas*, are possibly used to mean “finger.”

### 5.6. Amis

Endo (2011:5) noted that *kamay* “hand” in Amis may be a cognate with *hamai* “claw, fingernail” in Puyuma, which is recorded in Ferrell (1969:133, 209) and that the original meaning of *kamay* in Amis could be “fingernail.” Although the two forms look similar, the initial consonants do not show a regular sound correspondence. The *k* in Amis should also be *k* in Puyuma. Ogawa (2006:72) had another similar form *gamai* in Rukai meaning “fingernail.” However, the initial consonant *g* does not correspond to that of either the forms. Therefore, whether *kamay* in Amis, *hamay* in Puyuma, and *gamai* in Rukai are in a cognate relationship is uncertain. According to the comparative list in Blust and Trussel (2010), *kamay* in Amis has many cognates outside the Formosan languages (Table 5).



**Table 5. Cognate set of Proto-Austronesian \*kamay “hand” in Blust and Trussel (2010)**

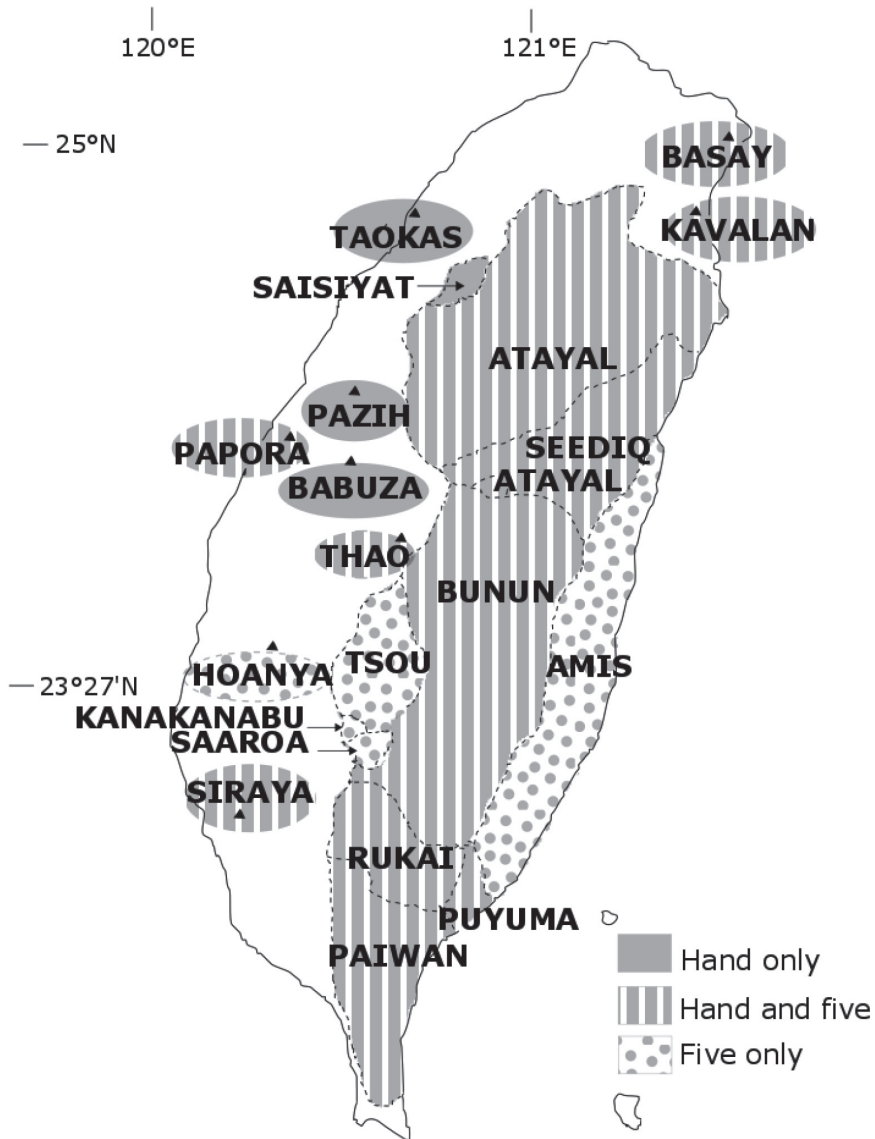
PAN	*kamay	“hand”
Amis	<i>kamay, kayam</i>	“hands, including the arm”
Yami	<i>ka-kamay</i>	“finger, toe”
Itbayaten	<i>ka-kamay</i>	“finger, limbs (of crustaceans)”
Tagalog	<i>kamay</i>	“hand”
Cebuano	<i>kamay</i>	“summon by a wave of the hand”
Taboyan	<i>kamɿy</i>	“hand”
Minyafuin	<i>kame</i>	“hand”
Kokota	<i>kame-na</i>	“hand”
Zabana	<i>kame</i>	“hand”

The cognates in Yami and Itbayaten mean “finger” like the quasi-cognates in Puyuma and Rukai above although the meaning in PAN is reconstructed as “hand” by Blust and Trussel (2010). Whether the original meaning of \*kamay was “hand” or “finger” is unclear.<sup>20</sup> In the Formosan languages, the reflex of \*kamay is only seen in Amis (or the quasi-cognates are seen in geographically continuous areas, Puyuma and Rukai, in southeast Taiwan); others are seen in some Malayo-Polynesian languages. Along this line, \*kamay can be classified as a post- PAN innovation that appeared later than PAN \*lima.

## 6. Conclusion—from “hand” to “five”

Map 1 illustrates the location of the Formosan languages and the meanings the reflexes of PAN \*lima in each language. Every Formosan language has \*lima either with the meaning of “hand” or “five” or both. The majority of languages have both meanings, and a few languages have one or the other.

Map 1. Semantic distribution of PAN \*lima in the Formosan languages



Those languages that seem not have undergone the semantic extension of \*lima “hand” to “five” are Taokas, Saisiyat, Babuza, and Pazih in northwestern Taiwan. They lack \*lima meaning “five,” and their forms for “five” (Taokas *hasap*, Saisiyat *asəb/Rasəb*, Babuza *achab*, and Pazih *xasəp*) are considered to be cognates by Ogawa (1944:480–482), which is reconstructed as \*RaCəp by Sagart (2004:416).<sup>21</sup> This form should be an innovation that happened only in the four languages in northwestern neighbors. If \*RaCəp “five” is reconstructed to Proto-Austronesian, then its reflexes should be more widely observed

among Austronesian languages; however, its distribution is confined to northwestern Taiwan. Therefore, this form of “five” could be an areal feature that involves borrowing. That is, this form could be innovated in a certain (proto-)language, and it was borrowed to neighboring languages because the four languages do not form a single subgroup according to Blust’s (1999a) hypothesis for the first-order subgrouping of PAN (Taokas and Favorlang are in one group, and Saisiyat and Pazih are in the other group).

Ochiai (2015) reexamined \*RaCəp “five” and proposed that this could rather be reconstructed as \*xasəp on the basis of Atayalic numerals “one” and “four.” According to her, the former part, *xə*, reflects Proto-Atayalic \*xə “one,” and the latter part, *səp*, reflects the first half of the Proto-Atayalic \*səpat “four.”<sup>22</sup> Moreover, she argued that Proto-Austronesian had numerals from “one” to “four” and that numeral “five,” whether being \*xasəp “five (one-four)” or \*lima “hand,” is an innovation in a later stage.

Additionally, she asserted that \*xasəp originating in Atayalic spread to neighboring languages, Pazih, Saisiyat, Babuza, and Taokas, with slight adjustments to the form to accommodate to their own phoneme inventory. For instance, the initial consonant was changed to *R* in Saisiyat (or lost in a dialect of Saisiyat), *h* in Taokas, and lost in Babuza. The medial consonant *s* is changed to “ch,” which represents an orthography by a Dutch who recorded this language; hence, it probably represented a phonetic [x]. The vowel in the final syllable is changed to *a* in Taokas.

The final consonant is changed to *b* in Saisiyat and Babuza. As Sagart (2004:416) pointed out, this voicing of *p* to *b* is seen in Pazih numerals from “six” to “nine,” which comprise *xasəp* “five” followed by numerals from “one” to “four,” i.e., *xasəp-uzə* (five-one) “six.” On the basis of phonological rules in Pazih (Blust 1999b:326), a voiceless stop before a morpheme boundary is voiced when followed by a vowel. The forms from “seven” to “nine” are *xasəb-i-dusə* “seven,” *xasəb-i-təru* “eight,” and *xasəb-i-supat* “nine” (Li and Tsuchida 2001). *xasəb* with the intervocalic voicing has probably been introduced to neighboring languages such as Saisiyat and Babuza.

This hypothesis is not without flaw. In this hypothesis, Atayalic is considered a source of \*xasəp “five.” However, this form is absent in the Atayalic languages. As aforementioned, “five” in the Atayalic languages reflects \*lima “hand.” Presumably, Atayalic had \*xasəp “five,” but it was replaced with the reflexes of \*lima “hand.” However, this cannot be proved.

These four languages that lack “five” originating in \*lima share a common feature in higher numerals (from “six” to “nine”). This feature is an imperfect decimal. Some higher numerals lack their own form but are derived from lower numerals. An imperfect decimal is

seen not only in these four languages but also in other Formosan languages, such as Atayal, Seediq, Thao, and Siraya.

Table 6 summarizes the imperfect decimals in these Formosan languages, as pointed out by Ogawa (1944). For instance, the numeral “six” to “nine” in Pazih shows the additive system  $5 + 1$  to  $5 + 4$ . There is also a subtractive system  $10 - 1$  for “nine” as seen in Saisiyat, Babuza, Taokas and Thao. Additionally, there are multiple systems wherein “six” are derived from “three” as seen in Thao, Atayal, and Seediq and “eight” are derived from “four” as in Saisiyat, Babuza, Taokas, Thao, Atayal, Seediq, and Siraya.

For comparison, the numerals in Bunun, which are perfect decimals reflecting Proto-Austronesian \*ənəm “six,” \*pitu “seven,” \*walu “eight,” and \*Siwa “nine,” are also presented.<sup>23</sup> The numerals of imperfect decimal are shaded in the table, and the composition of the numeral is shown in the second line.<sup>24</sup> The data for “five” is also added; however, the analysis of  $1 + 4$ , as proposed by Ochiai (2015) for the four languages from the top, are still tentative.

**Table 6. Imperfect decimal in the Formosan languages**

	Five	Six	Seven	Eight	Nine
Pazih	<i>xa-səp</i> (1+4?)	<i>xasəb-uza</i> 5+1	<i>xasəb-i-dusa</i> 5+2	<i>xasəb-i-turu</i> 5+3	<i>xasəb-i-supat</i> 5+4
Saisiyat	<i>Ra-səb/a-səb</i> (1+4?)	<i>saibusiR</i>	<i>saibusiR-o-aha</i> 6+1	<i>maika-spat</i> 4×2	<i>R-aʔha</i> 10-1
Babuza	<i>a-chab</i> (1+4?)	<i>na-taap</i>	<i>na-ito</i>	<i>maa-spat</i> 4×2	<i>ta-nnacho</i> 10-1
Taokas	<i>ha-sap</i> (1+4?)	<i>tahap</i>	<i>y-weto</i>	<i>ma-halpat</i> 4×2	<i>ta-nasu</i> 10-1
Thao	<i>rima</i>	<i>ka-turu</i> 3×2	<i>pitu</i>	<i>ka-shpat</i> 4×2	<i>ta-nacu</i> 10-1
Atayal	<i>ima-gal</i>	<i>cziuʔ</i> 3×2	<i>pituʔ</i>	<i>spat</i> 4×2	<i>qeru</i>
Seediq	<i>rima</i>	<i>ma-taru</i> 3×2	<i>pitu</i>	<i>ma-səpat</i> 4×2	<i>məɣari</i>
Siraya	<i>rima</i>	<i>nəm</i>	<i>pitu</i>	<i>kui-xpa</i> 4×2	<i>matuda</i>
Bunun	<i>hima</i>	<i>nuum</i>	<i>pitu</i>	<i>vauʔ</i>	<i>sivaʔ</i>

Those languages that lack \*lima meaning “hand” are Hoanya, Tsouic languages, and Amis, which are geographically discontinuous. The Hoanya and Tsouic languages are on the one side, and Amis is on the other side, blocked by Bunun. Further, the forms of “hand” in these languages differ: *pilat/pilas/pila* in Hoanya, \*ramuCu in the Tsouic languages,

and *kamay/kayam* in Amis. Therefore, these forms are likely to have developed in each language independently.

The Atayalic languages also lack the reflexes of \*lima meaning “hand.” However, they are retained in derived forms through prefixation *m-ima* in Atayal and *tu-rima* in Seediq, and both forms mean “to wash oneself.” The reflexes of \*lima in the Atayalic languages, *ma-gal* in Atayal and *rima* in Seediq, are used as the words for “five.” That the “hand” is only seen in derived forms suggests its oldness because lost roots are retained in derived forms in some cases. In the Atayalic languages of the present-days, “hand” is expressed by words that reflect PAN \*qabaRa “shoulder”: *qaba* in Atayal and *baga* in Seediq. Accordingly, in the Atayalic languages, “hand” is an old meaning of \*lima, but this meaning was replaced by an innovative meaning “five,” and the original forms for “hand,” i.e., \*lima, were replaced by other forms, probably triggered by a necessity to avoid the ambiguity between “hand” and “five.” The semantic extension of \*lima from “hand” to “five” may have happened in the other Formosan languages.

Even among languages that have the reflexes of \*lima meaning both “hand” and “five,” slight differences emerge between their forms of “hand” and “five” as exemplified in Bunun and Rukai forms in Table 1 (*ima* “hand” as opposed to *hima* “five” in Bunun and *arima* “hand” as opposed to *rima* in Rukai).<sup>27</sup> In short, some Formosan languages replaced \*lima “hand” with another word, while other languages modified either (“hand” or “five”) forms of \*lima, and both changes were triggered for the purpose of semantically disambiguating “hand” and “five.”

In conclusion, PAN \*lima originally meant “hand.”<sup>28</sup> Later, it was semantically expanded to signify “five.” In some languages such as the Atayalic languages, Hoanya, Tsouic languages, and Amis, the forms of “hand” were innovated to avoid semantic ambiguity between “hand” and “five.” In the Atayalic languages, the innovated forms of “hand” are derived from a word which originally meant “shoulder”; however, the older form of “hand” originating from \*lima is retained in words meaning “to wash oneself.” This claim that \*lima “hand” is semantically extended to “five” in the Austronesian languages is also supported by a cross-linguistic observation of “hand” developing into “five” as described in Heine and Kuteva (2002:166).

## Notes

- \* A part of Section 4 in this paper was presented at the 25<sup>th</sup> meeting of Southeast Asian Linguistic Society, May 27–29, 2015. I would like to thank Mitsuaki Endo who took interest in my presentation and directed me to this study by providing me his paper on the related topic.

## 14 Izumi Ochiai

- 1 Yami is spoken on the outlying Orchid Island and belongs to the Malayo-Polynesian subgroup of the Austronesian family.
- 2 The alternative PAN from Blust (1999a:82) reconstructed for “hand” is \*qalima. However, throughout this paper, \*lima is used for the form of “hand,” considering that a basic vocabulary like “hand” should show a typical syllable structure pertaining to the Austronesian languages, which is disyllabic.
- 3 The map of Endo (2011:2) that introduces the names of languages lacks one language, Siraya. Siraya should be added to the last of the list of language names right after Thao.
- 4 The data sources are the following: Atayal (Egerod 1965:211, 210); Saisiyat (Li 1978:160, 196); Pazih (Li and Tsuchida 2001); Taokas, Babuza, Kavalan, Thao, and Siraya (Ferrell 1969:217, 411); Tsou, Saaroa, Kanakanabu, Rukai, Paiwan, Puyuma, Amis, and Bunun (Ogawa and Asai 1935); Basay (Tsuchida, Yamada, and Moriguchi 1991:231, 239); Papora (Tsuchida 1982:49); and Hoanya (Ogawa 2006:66 and Tsuchida 1982:49). The forms cited from previous studies are slightly modified in this paper. Seediq data are taken from the field note of the author throughout this paper, which represents the Paran dialect of Seediq.
- 5 The phoneme represented by “R” is described as a voiced uvular fricative in Ogawa and Asai (1935:109) but as an alveolar–palatal liquid in Li (1978:138).
- 6 Papora data are taken from Tsuchida (1982:49).
- 7 Hoanya data for “hand” are taken from the data numbers 142b.1 and 142b.3 in Ogawa (2006:66). Hoanya data for “five” are taken from Tsuchida (1982:49).
- 8 More precisely, Li (1981:295) affirmed that *r* is deleted word-initially in Proto-Atayalic, including Seediq. He provided only one example to verify this, which was \*lima meaning “five.” This initial consonant deletion is only observed in Atayal as he gave examples \*lima > \*rima > *ima-gal* “five,” but not in Seediq in which “five” is *rima* with an explicit initial consonant.
- 9 Further, the reflexes of \*lima (*ima* in Atayal and *rima* in Seediq) appear in the word for “fifty,” which is *m-ima-l* in Atayal and *mu-rima-n* in Seediq. Their roots are attached with a circumfix *m-...-l* in Atayal and *mu-...n* in Seediq, which can be reconstructed as Proto-Atayalic \**ma-...l*. The reconstruction of the circumfix to Proto-Austronesian is proposed in Zeitoun, Teng, and Ferrell (2010:864-868).
- 10 The Ogawa and Asai (1953) forms have been slightly modified by the author in Table 2.
- 11 Data on languages other than Pazih were obtained from Ogawa and Asai (1935: Appendix 4, 46). Pazih was obtained from Li and Tsuchida (2001:251).
- 12 The PAN form for “shoulder” is taken from Blust and Trussel (2010), which contains a cognate sets, such as Saisiyat *ʔæbaLaʔ*, Pazih *ʔabazaʔ*, Amis *ʔafala*, and Ilokano *abága*.

- 13 Li (1981:249) says \*q is sometimes deleted in Seediq. In addition, with regard to Paran Seediq, Yang (1976:640, 664–665) observed that there is no syllable structure such as VCVCV. She explained that the antepenultimate vowel in the underlying structure is deleted, resulting in CVCV. Seediq *baga* “hand” is likely to be the result of these two phonological changes (the sporadic loss of *q* and the loss of antepenultimate vowel *a*).
- 14 For instance, the data numbered 19 and 26 in Ogawa (2006) are classified in Seediq. However, Ochiai (2016c) clarified that they belong to Atayal.
- 15 Naoyoshi Ogawa noticed that the medial *ima* in *kiiman* comes from PAN \*lima because he segmented it a *s ki-ima-n*, which indicates that *ki-* is a prefix, *-n* is a suffix, and *ima* is a root.
- 16 Endo (2011:5) highlighted that *kayam* is a metathesized form of *kamay* (see the data for Amis that have *kamay* and *kayam* in Table 1).
- 17 This place is called Dongli (東里) nowadays and situated in the south of Hualian county.
- 18 Endo (2011:5) has *lamutsu* as a Saaroa form.
- 19 Citing the reflexes of \*kamay in some Malayo-Polynesian languages that mean either “hand” and “finger,” Dahl (1981:49) observed the semantic similarity in expressions like “doing something with the hand” and “doing something with the fingers.”
- 20 According to Ogawa (1944), an extinct language, Luilang located in northern plains of Taiwan also had “five” different from \*lima. This form for “five” is *(na-)lup*, and its origin is uncertain. The higher numerals in Luilang were not concerned in this paper because they are quite different from those of other Formosan languages. Higher numerals in Luilang need to be investigated in the future.
- 21 The reconstruction of this numeral “one” in Atayalic is mentioned in Ochiai (2019) and Ochiai (2020). The form of Proto-Atayalic numeral “four” is supplemented by the present author.
- 22 Proto-Austronesian numerals from six to nine are from Blust and Trussel (2010). To date, Proto-Austronesian is considered to be a perfect decimal. However, Ochiai (2015) claimed that Proto-Austronesian had numerals up to “four” and that Proto-Austronesian \*lima “five” to \*puluq “ten” should be considered as innovations in post-Proto-Austronesian. The forms for “ten” are not included in this table. However, none of these languages in Table 6 have the reflexes of PAN \*puluq “ten.”
- 23 With regard to the forms that are non-imperfect decimals, “seven” in Babuza, Taokas, Thao, Atayal, Seediq, and Siraya reflects PAN \*pitu, and “six” in Siraya reflects PAN \*ənəm. The origin of other forms (“six” in Saisiyat, Taokas, and Babuza and “nine” in Atayal, Seediq, and Siraya) is uncertain. It is clear that “six” in Taokas and “six” in Babuza are cognates (Ogawa 1944:483). Additionally, Tsukida (2009:239) suggested that “nine” in Seediq may be related to *səŋari* “left over.”

- 24 The data is taken from Sagart (2004:414), except for Saisiyat that is taken from Ogawa (1944) and Seediq that is taken from Rakaw et al. (2006). The segmentation of the words is provided by the present author.
- 25 For “nine” in Babuza, Taokas, and Thao, Ogawa (1944:488) analyzed the word initial *ta-* as the form indicating “one.” The analysis of “nine” in Saisiyat is based on Li (2011:201). The second part *aʔha* is taken as the form indicating “one.”
- 26 This point is presented as a conclusion in the paper of Endo (2011:10–11), in which he gave examples from Kavakan and Rukai. However, the Kavalan forms of “hand” and “five” show no difference in Ferrell (1969:217, 411). Possibly, Basay forms in Table 1, *tsima* and *tsjima*, show phonetic differences (e.g., lengthening of the first vowel in the latter form). Dahl (1981:49) also provided the possible cases of the similar phonological disambiguation in two Malayo-Polynesian languages, Ngaju Dayak and Fijian, in which the nasal in \*lima changed into *ŋ* in the forms of “hand,” whereas the forms of “five” remained unchanged.
- 27 Sagart (2013:482) made the same point based on the analysis on Tsouic and Rukai words for “hand,” “finger,” and “five.”

## References

- Blust, Robert. (1999a) Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In *Selected papers from the eighth international conference on Austronesian linguistics*, edited by Elizabeth Zeitoun and Paul Jen-kuei Li, 31-94. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert. (1999b) Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38: 321-365.
- Blust, Robert and Stephen Trussel. (2010) *Austronesian comparative dictionary, Web edition*. <http://www.trussel2.com/ACD/>, accessed May 18, 2019.
- Bullock, Thomas L. (1874) Formosan dialects and their connection with the Malay. *China Review, or Notes and Queries on the Far East* 3:38-46.
- Dahl, Otto C. (1981) Austronesian numerals. *NUSA —Linguistic Studies of Indonesian and Other Languages in Indonesia*: 46-58.
- Egerod, Søren. (1965) An English-Atayal vocabulary. *Acta Orientalia* 29:203–20.
- Endo, Mitsuaki. (2011) Linguistic maps of Formosan languages in the early twentieth century: “Hand” and “five.” *Journal of Taiwanese Languages and Literature* 6(2):1-14.
- Ferrell, Raleigh. (1969) *Taiwan aboriginal groups: problems in cultural and linguistic classification*. Taipei: Institute of Ethnology, Academia Sinica.
- Heine, Bernd and Tania Kuteva. (2002) *World lexicon of grammaticalization*. Cambridge:



Cambridge University Press.

- Ino, Kanori. (1898) Taiwan tsusin (dai 26 kai) [Taiwan newsletter No. 26]. *The Journal of the Anthropological Society of Tokyo* 13(150):486-491.
- Ino, Kanori. (1907) Taiwan dohan no shusi nitsukiteno gainen [Concepts of hand and fingers in Taiwan aborigines]. *The Journal of the Anthropological Society of Tokyo* 23(261):105-109.
- Li, Paul Jen-kuei. (1981) Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology, Academia Sinica* 52(2):235-301.
- Li, Paul Jen-kuei. (1978) A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 49:133-199.
- Li, Paul Jen-kuei (2011) *Taiwan nandao Minzu de zuqun yu qianxi* [The groupings and migrations of Taiwan’s Aboriginal peoples]. Taipei: Avanguard Publishing.
- Li, Paul Jen-kuei and Shigeru Tsuchida. (2001) *Pazih dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Ochiai, Izumi (2015) Origin of the numeral “five” in northwestern Formosan languages: Revision of the Proto- Austronesian numeral system. Paper presented at the 25th Annual Meeting of the Southeast Asian Linguistic Society, 27 May 2015, Payap University, Chiang Mai.
- Ochiai, Izumi. (2016a) Bu-hwan vocabulary recorded in 1874: Comparison with Seediq dialects. *Asian and African Languages and Linguistics* 10:287-324.
- Ochiai, Izumi. (2016b) Sedekkugo Paran hōgen no bunpō kijutsu to hiishisei settouji no hikaku gengogakuteki kenkyū [A descriptive grammar of Paran Seediq and a comparative study of non-volitional prefixes], Ph. D diss, Kyoto University.
- Ochiai, Izumi. (2016c) Taiwan bango shūroku no Shabogala to Baiban : Sedekkugo dewanaku Atayarugo toshiteno bunrui [The Position of Shabogala and Ka’aran as Atayal not Seediq: Reflection on Naoyoshi Ogawa’s Comparative Vocabulary]. *Tokyo University Linguistics Papers* 37: 171-190.
- Ogawa, Naoyoshi. (1931) *tayaru goshū* [Atayal word lists]. Taihoku: Taiwansōtokufu.
- Ogawa, Naoyoshi. (1944) Indoneshiago ni okeru taiwan takasagogo no ichi [The place of the Formosan languages in the Indonesian languages]. In *Taiheiyōkoku-minzoku to nunka* vol.1, edited by Hirano Yoshitaro, 451-502. Tokyo: Kawade Shobo.
- Ogawa, Naoyoshi. (2006) *A Comparative vocabulary of Formosan languages and dialects*, compiled by Paul Jen-kuei Li and Masayuki Toyoshima. Tokyo: Research Institute for Languages and Cultures in Asia and Africa.
- Ogawa, Naoyoshi and Erin Asai. (1935) *The myths and traditions of the Formosan native*

- tribes*. Taipei: Taihoku Imperial University. [In Japanese].
- Rakaw, Lowsi, Jiru Haruq, Yudaw Dangaw, Yuki Lowsing, Tudaw Pisaw, and Iyuq Ciyang ed. (2006) *Tailugezuyu jianyi zidian* [Concise Dictionary of Taroko Seediq]. Xiulin: Xiulin Township Administrative Office. [In Mandarin].
- Sagart, Laurent. (2004) The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2):411-444.
- Sagart, Laurent. (2013) Is Puyuma a primary branch of Austronesian? A rejoinder. *Oceanic Linguistics* 52(2): 481-492.
- Tashiro, Yasusada. (1900) *Taitō shokuminchi yosatsu hōbun* [A preliminary report on the colony in Taito]. Taihoku: Taiwan Sotokuhu Minseibu. [In Japanese].
- 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, University of Tokyo* 7:1-166.
- Tsuchida, Shigeru. (1975) Reconstruction of Proto-Tsouic phonology. Ph. D. diss, Yale University.
- Tsuchida, Shigeru, Yukihiro Yamada, and Tsunekazu Moriguchi (1991) *Linguistic materials of the Formosan sinicized populations I : Siraya and Basai*. Tokyo: The University of Tokyo, Linguistics Department.
- Yang, Hsiu-fang. (1976) The phonological structure of the Paran dialect of Sediq. *Bulletin of the Institute of History and Philology, Academia Sinica* 47: 611-706. [In Mandarin].
- Zeitoun, Elizabeth, Stacy, Fang-ching Teng, Raleigh Ferrell. (2010) Reconstruction of ‘2’ in PAN and related issues. *Language and Linguistics* 11(4):853-884.