

MORA, VOWEL LENGTH, AND DIACHRONY: THE CASE OF ARTA, A PHILIPPINE NEGRITO LANGUAGE

Yukinori Kimoto
Nagoya University
<yk.kimoto@gmail.com>

Abstract

This paper attempts to provide an explanation for the diachronic development of long vowels in Arta, a Negrito language spoken in Nagtipunan, Quirino Province, the Philippines. In Arta, a large number of lexical roots and morphologically complex words have long vowels in them, but the items with a long penultimate vowel which are shared with other Philippine languages that retain an older accentual system are reflected as short vowels. Thus, the long vowels seen in Arta should be separated from inherited accents. It is argued that these vowels developed independently in the language by compensatory lengthening and vowel fusion, after the loss of *k, *q, and *h. Since both compensatory lengthening and vowel fusion crucially involve the principle of mora count conservation, the phonological changes which occurred in Arta indicate that the mora has played a significant role in the language.

Keywords: Arta, vowel length, mora, compensatory lengthening, vowel fusion
ISO 639-3 codes: atz

1 Introduction

One of the interesting typological features observed in many Philippine languages involves the realization of contrastive word accents (or stresses) as vowel length on the penultimate open syllable. This phonetic manifestation shows a clear typological difference from English (stress accent) and Japanese (pitch accent).¹ The following minimal pairs from Tagalog, Ilokano and Bikol illustrate the point.

- (1) Tagalog *áso* [ʔa:so] ‘dog’ vs. *asó* [ʔaso] ‘smoke’
Ilokano *bára* [ba:ra] ‘hot’ vs. *bará* [bara] ‘lung’
Bikol *bága* [ba:ga] ‘ember’ vs. *bagá* [baga] ‘truly!’

(Zorc 1993:18)

This kind of vowel-length contrast is also found among various Philippine languages, including Aklanon, Balangao, Cebuano, Hanunoo, Ibanag, Ifugao, Isnag, Kalinga, Kapampangan, Sambal (Zorc 1979:241). Based on the fact that cognate forms among these languages share the same length on the penult, Zorc (1979) argues that the contrastive word accent system may be attributed to “Proto-Philippines”,² as shown in Proto-Philippines (PPH) *da:Raʔ ‘blood’, cf. Isnag *da:ga*, Ilokano *da:ra*,

¹ This article is a revised version of the following talks I presented: ‘Synchronic and diachronic phonology of the Arta language’ in The 3rd MINPAKU Linguistics Circle, Osaka, Japan, Sept. 2014, ‘Development of the phonological system in Arta’ in Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies, Japan, 2015, and ‘The role of mora in phonology: A case for Arta, a Northern Luzon language’ in 13th International Conference on Austronesian Linguistics, Academia Sinica, Taipei, Taiwan, July, 2015. My thanks go to Ritsuko Kikusawa, Hsiu-chuan Liao, Naonori Nagaya, Lawrence A. Reid, and David Zorc for comments on the presentations and/or an earlier version of the paper. All remaining errors are my own.

² Although Zorc labelled the hypothetical proto-language as “Proto-Philippines” in his previous publications, and this paper cites his reconstructed data with the same label Proto-Philippines for the purpose of comparing reconstructed forms with Arta forms, he no longer holds to a Proto Philippines. (Zorc

Ifugao *da:la*, Kapampangan, Sambal *da:yaʔ*; PPH **da:lan* ‘path, trail’, cf. Aklanon, Balangao, Cebuano, Bikol, Hanunoo, Ibanag, Isnag, Kapampangan, Ilokano *da:lan*.

This is not the case in Arta, a Northern Luzon Negrito language, however. In spite of the abundance of items with long vowels (e.g. *ka:man* ‘big, large’, *bu:ru* ‘new’, and *a:na:* ‘children’), the etyma with long penult in the reconstructed language are all reflected with short vowels, as in (2).

- (2) PPH **si:ku* > Arta *siku* ‘elbow’
PPH **tu:bu* > Arta *tubu* ‘grow’

The aim of this paper is thus to explain the historical development of long vowels in Arta. Three sources of long vowels are identified: (i) onomatopoeia, (ii) borrowing from Ilokano and Yogad, and (iii) sound changes in inherited forms from Proto-Malayo-Polynesian (henceforth PMP). It is argued that the *mora* was responsible for the third pattern, playing a significant role in the historical phonology in Arta.

This paper is organized as follows. §2 provides basic information about the Arta language. In §3, it is argued that the *mora* plays a significant role in synchronic phonology in Arta. §4 deals with the matter as to how long vowels in Arta are developed historically. After the discussion of the first two sources of long vowels in Arta in §4.1 and §4.2, §4.3 observes sound changes and some factors which might have motivated the vowel lengthening, with special reference to the *mora* count. §5 deals with the interactions between the *mora* and other factors which may affect current reflexes in Arta.

2 The Arta language

2.1 *Its speakers and sociolinguistic profile*

Arta is an Austronesian language currently spoken by eleven Arta people living in the municipality of Nagtipunan, Quirino province, the Philippines. The speakers of Arta belong to a larger group called Negrito, who are commonly characterized as having shorter stature, curled hair, and darker skin. They are considered to be a descendant of the people who had settled in the Philippine archipelago over tens of thousands years before the speakers of an Austronesian language migrated into the islands; since Negrito people switched from their original languages to Austronesian languages, the current language that the Arta people speak clearly belongs to the Austronesian family.³

The northern part of Luzon is home to a number of Negrito groups as well as non-Negrito groups (see Fig. 1). The longest river in the Philippines, the Cagayan River, runs from south to north, forming the Cagayan Valley. The valley is sandwiched between the Sierra Madre mountains in the east and mountainous district in the west. The east side of Sierra Madre, that is, the eastern coast of northern part of Luzon, is home to various Negrito groups such as Dupanangan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta. The valley of the Cagayan River is also occupied by other Negrito groups, Pamplona Atta, Faire-Rizal Atta, Pudtol Atta in the northwestern side of the valley, and Central Cagayan Agta in the northeastern side. Quirino Province is located on the upper reaches of the Cagayan River, which is occupied by Nagtipunan Agta as well as Arta (Fig. 2). The area was formerly covered with a thick rainforest, traditionally occupied by Nagtipunan Agta and a non-Negrito group Ilongot (or Bugkalot) as well as Arta, but the area is currently inhabited by a large number of immigrants from outside of the province to reclaim the forested area. The Arta people formerly lived in Cordon and Alicia in Isabela, but they moved to Disubu in the municipality of Aglipay in Quirino around four to five decades ago; they are currently settled in the municipality of Nagtipunan, Quirino.

pers.comm.); in fact, recent studies (Reid 1982, Ross 2005) point out the lack of phonological evidence to justify Proto-Philippines as a distinct proto-language from Proto-Malayo-Polynesian.

³ It is interesting that many of the Negrito languages are conservative in grammatical and/or phonological characteristics. Arta, for example, retains the old pronominal form =*myu* (second person plural genitive form), which could date back to Proto-Northern Luzon (Reid 1979), and perhaps to PMP (see Reid 2009, but Ross 2006 proposes the formatives *=ihu, *=mu-ihu for the second person plural genitive).

Arta is a severely endangered language compared with other Negrito languages which are also seen as “minority” languages (see Headland 2003 for the demographic data on Philippine Negrito languages). The number of fluent speakers of Arta is 10, with 35–45 people barely understanding the language but unable to speak properly; most of the speakers are over 40 years old, except one young fluent speaker at the age of 29. In the last two decades, Arta people moved from Aglipay and/or Maddela to Nagtipunan, being merged into the speech communities whose majority are Nagtipunan Agta.⁴ This seems to have caused the increase of the intermarriage between Arta and Nagtipunan Agta, to the extent that there is currently no “pure” Arta family. This social structure inevitably forces the Arta language, a “minority” language in the community, to be out of use; in my fieldwork, I have been unable to find any family in which Arta is spoken. They usually communicate in Nagtipunan Agta within the community, in Ilokano outside the community, and in Arta with an older generation whose first language is Arta; in fact, some Arta people use the language when they talk to their siblings, but not to their children.

Figure 1: Northern part of Luzon



Figure 2: Quirino Province (seen from the south)



2.2 Subgrouping relationship and language contact

The genetic subgrouping of Arta was studied by Reid (1989). He concludes that the language is an isolate within the Northern Luzon (NLZN) subgroup of Malayo-Polynesian (MP). Northern Luzon languages are widely distributed in the northern part of Luzon, surrounded by Bashiic languages spoken in Batanes islands in the north, and by Central Luzon languages in the south. Northern Luzon languages include Ilokano; Meso-Cordilleran languages such as Bontok, Ifugao and Kalinga; Cagayan Valley languages such as Yogad, Ibanag and Gaddang; and North-eastern Luzon languages including Dupanangan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta.⁵ One piece of linguistic evidence for the subgrouping as NLZN is the sporadic metathesis between *t ... s > /s ... t/, which is shared exclusively by NLzn languages, as in PMP *tañis > PNLZN *sanjit (Reid 2006). Although Reid

⁴ Nagtipunan Agta is another Negrito group occupying the area. Few ethnographic or linguistic studies are conducted on Nagtipunan Agta except Robinson and Lobel (2013), a comparative study on East coast Negrito languages. Nagtipunan Agta is mutually intelligible with Casiguran Agta (or Casiguran Dumagat; Headland and Headland 1974, Headland and Healey 1974). It is still unclear as to how long they have been occupying the areas in Disimungal, or how different it is from Casiguran Agta.

⁵ Northern and Southern Alta are also Negrito languages subgrouped within Northern Luzon (see Reid 1989, 1991).

(1989, 2013) does not provide such evidence for the subgrouping of Arta, the metathesis does exist, as shown in (3).

- (3) PMP *ditaʔas > *disat* ‘high (the sun)’
 PMP *təRas ‘hardwood, hard’ > *sarat* ‘narra wood’

Since PMP *ʔ was lost and PMP *R changed into /r/ in Arta, the items in (3) are in accordance with regular sound changes that occurred in the language. The cognates presented above seem to provide strong evidence for positing that Arta is subgrouped within other Northern Luzon languages.

Another important sound change that occurred in Arta is *R > /r/, which provides strong evidence for the subgrouping of NLzn languages. As Reid (1989) states, Arta and Ilokano reflect *R as /r/, Meso-Cordilleran languages /l/, and North-eastern and Cagayan Valley languages /g/. Arta has a different reflex from the latter two subgroups, except Ilokano, which is still difficult to subgroup with, partially because most of the forms are not uniquely shared. Even the cognate words that would be expected to have shared innovations (*R > /r/) are reflected differently, e.g. *bəRŋaw > Arta *biriŋaw*, Ilk. *bəŋaw* ‘fly (n.)’; *huRas > Arta *uras*, Ilk. *u:gas* ‘wash’; *kaRat > Arta *arat*, Ilk. *kagat* ‘bite’; *Raŋu > Arta *raŋu*, Ilk. *gaŋu* ‘wither’; *Rapu > Arta *rapu*, Ilk. *gapu* ‘be from’ (it seems that Arta has more coherent reflexes of *R as /r/ than Ilokano). In Arta, PMP *j consistently changed into /d/ and thus merged with *d, which exhibits a further difference from Ilokano, where *j is reflected as /g/ (Reid 1989).

This subgrouping, however, reflects only one aspect of the linguistic history of Arta; the speakers of Arta have undergone several major periods of language contact. The ancestors of Philippine Negrito groups, including Arta, are considered to have dispersed into the Philippine archipelago in the Palaeolithic age over 20,000 years ago. It is suggested by current genetic and archaeological studies that, after common ancestral populations of modern humans dispersed “out-of-Africa”, probably through south Asia along the coastal side, the populations settled in the Southeast Asia as a second dispersal no later than 25–38,000 years ago, after the first dispersal into New Guinea, Melanesia, and Australia ~62–75,000 years ago (Rasmussen et al. 2011, Reyes-Centeno et al. 2014). The populations that settled in the Philippines must have spoken non-Austronesian languages, although there seems to be no clear evidence for reconstructing the details, except probable non-Austronesian lexical residuals in some current Negrito languages (Reid 1994).

The early Austronesians residing in Taiwan migrated into the Philippines probably via the Batanes islands around 4,000 years ago (see Reid 2013, Ko et al. 2014, see further Blust 1999, 2013). The ancestors of Arta are considered to have switched their languages to an Austronesian language which may currently be called Proto-Northern Luzon. Furthermore, after the language shift, the Arta language was influenced by Ilokano and Cagayan Valley languages such as Yogad, Gaddang, and Ibanag. Yogad, in particular, was influential on Arta; in my count, Arta shares with Yogad at least 98 lexical items. This is probably because the Arta people resided in Alicia until four to five decades ago, and they were bilingual in Arta and Yogad until recently. Speakers of Arta claim that they are similar to Yogad as “relatives”.

Following a basic description of the sociolinguistic and historical profiles of the Arta language, the next section will provide the definition of the mora count and its application to the Arta synchronic phonology. It is argued that the mora is at work in various phonological aspects in Arta, which will in turn be a requisite for explaining diachronic change in vowel length.

3 The mora in synchronic phonology

3.1 Phonological status of vowel length in Arta

Arta has sixteen consonant phonemes: seven stops /p, b, t, d, k, g, ʔ/, three nasals /m, n, ŋ/, two fricatives /s, h/, two liquids /l, r/, and two glides /y, w/, and has six vowels /i/, /e/, /a/, /o/, /u/, and /ə/. In this paper, the symbol “y” refers to the glide /j/ in the phonemic and orthographic descriptions. Vowel length is phonemically distinctive, as in *bi:ləg* ‘bracelet’ vs. *biləg* ‘fastness’, *ana:* ‘child’ vs.

a:na: ‘children’. /i/, /a/, /u/, and /ə/ have long and short phonemic contrasts;⁶ however, /e/ and /o/ always appear as long vowels /e:/ and /o:/ unless they are reduced as short vowels within the CVC syllable template (as discussed in §3.2, long vowels cannot appear in CVC syllables).

How should long vowels be interpreted phonologically? This may be seen as vowel length as it is, or as the phonetic manifestation of the stress (or accent) system as in the case of other Philippine languages. Nevertheless, the present study demonstrates that it is phonologically the length of a vowel, rather than the phonetic manifestation of the stress system in two respects. First, if the long vowel is really a manifestation of lexical stress, the place of the stress may shift after affixation, as in English: *phótophraph* > *photógraphy* > *photográphic*. This is the case in Ilokano, as in *ba:sa* > *basa:-en* ‘read’ and *la:wa* > *ka-lawa:-en* ‘to widen’ and in Tagalog as in *gu:gol* > *gugu:l-in* ‘to spend’ and *tu:ro* > *turo:-an* ‘to teach’, but this kind of shift is not observed in Arta, as shown in (4):

- (4) *pe:nas* > *pe:nas-an*, ****pena:s-an** ‘wipe’⁷
di:muy ‘bathing’ > *pandi:muy-an*, ****pandimu:y-an** ‘bathing place’
ka:lig > *ka:lig-ən*, ****kali:g-ən** ‘carry, transfer something’

The second defining feature is that each word (not a *root*) has one stress nucleus. Even if two stresses are observed within a single word, they have a phonetic asymmetry to the extent that one stress is considered to be primary, and the other secondary (e.g. *dòcuméntary*, *téxt-bòok*). This is not the case in Arta. Several words in Arta do have more than one long vowel without such phonetic asymmetry. Consider the following forms in (5):

- (5) More than one long vowel can occur within one word
ta:me:ta ‘different’ > *ta:ta:me:ta* ‘quite different’
na:na:b (*na:na:bən*) ‘remember, recall’
no:no:t (*no:no:tən*) ‘think’
pe:be:bu:d=u ‘I am asking’

Since neither of these facts favor the long-vowel-as-stress analysis, this contrast will be treated more appropriately as the phonological long-short distinction of vowels.

3.2 Mora and synchronic phonology

In many of the world’s languages, prosodic and/or metric structures are sensitive to a different “weight” of syllables defined by the vowel length and other syllable features. The unit of the syllable weight is called *mora*, and a monomoraic syllable is defined as an open syllable with a short vowel. Languages may differentiate a monomoraic syllable (light syllable) with a short vowel (CV), a bimoraic syllable (heavy syllable) with a long vowel (CV:), and, in some languages, a trimoraic syllable (superheavy syllable) with a superlong vowel (CV::). Many languages such as Japanese and Latin count a coda consonant as having one mora. This means that there are two patterns for classifying syllables in terms of mora count, as shown in (6) and (7).

- (6) Pattern A (coda consonants are counted)
 i. monomoraic syllable: CV
 ii. bimoraic syllable: CV:, CVC
 iii. trimoraic syllable: CV::, CV:C

⁶ /ə/ does not appear as a long vowel except in one item *də:gi* ‘strip or string used for carrying a basket’.

⁷ The presence of double asterisks (**) before a form indicates that the form in question is “unacceptable” or “illegitimate”.

- (7) Pattern B (coda consonants are NOT counted)
 i. monomoraic syllable: CV, CVC
 ii. bimoraic syllable: CV:, CV:C
 iii. trimoraic syllable: CV::, CV::C

In the case of Arta, moraic classification of syllables can be shown in (8) and generalized in (9).

- (8) Arta (coda consonants are counted)
 i. monomoraic syllable: CV
 ii. bimoraic syllable: CV:, CVC
 iii. *trimoraic syllable: CV::, CV:C
- (9) Mora constraint on syllables: The syllable must not exceed 2 moras

First, as to be illustrated below, the syllables in Arta behave in a way similar to pattern A, that is, a syllable with a coda consonant and a syllable with a long vowel behave in the same way. The language is not allowed to have trimoraic syllables such as CV:: (superlong vowel) and CV:C (long vowel and coda consonant). If morphological conditions require such syllable structure, these syllables should be reduced to CV: and CVC types, respectively. This constraint can be paraphrased as (9): “The syllable must not exceed 2 moras.” Now let me introduce some mora-sensitive phenomena in Arta phonology, which support the statements in (8) and (9).

3.2.1 Allomorphs of enclitics

The mora count is necessary for generalizing some allomorphic conditions. Two enclitics, =*di* ‘already, just now’ and =*pa* ‘just, try -ing, do a little’ are sensitive to the mora count of the preceding syllable to which they attach. If the preceding word ends with a bimoraic syllable, then =*di* and =*pa* must appear, but if the preceding word ends with a monomoraic syllable, =*d* and =*p* must be used instead. The following example with =*di* illustrates the point.

- (10) CVC + *di* *awan*=*di* (NEG=already) ‘It does not exist already.’
 CV: + *di* *mebbuyu*:=*di* (bad-smelling=already) ‘It has already become bad-smelling.’
 CV + *di* *nappati*=*d* (died=already) ‘S/he has already died.’

The condition under which different allomorphs may occur cannot be explained solely in terms of vowel length nor only according to whether the preceding syllable is open or closed. Rather, this fact indicates that =*p* and =*d* may appear if the preceding word ends with a monomoraic syllable, whereas =*pa* and =*di* may appear if the preceding word ends with a bimoraic syllable, with coda consonants counted as one mora in the language.

3.2.2 Vowel length alternation

The moraic constraint on the maximal weight of the syllable structure, explicated in (8) and (9), predicts vowel length alternations. If a long vowel appears in a closed syllable, it is supposed to be realized as a reduced short vowel, whereas it is realized as a long vowel if it appears in an open syllable. This alternate realization is not applied only to the native lexemes (11a), but also to Spanish and English loan words (11b).

- (11) a. /latto:ŋ/ > *lattoŋ* vs. *latto:ŋ=i*⁸ ‘outside’
 /sa:y/ > *mas-say* vs. *sa:y-an* ‘ride on’
 /dage:t/ > *man-daget* vs. *dage:t-an* ‘sew’
 /du:t/ > *dut* vs. *du:t=i* ‘firewood’
 b. /hapo:n/ > *hapon* vs. *hapo:n=i* ‘Japan’
 /pantalo:n/ > *pantalon* vs. *pantalo:n=i* ‘pants’
 /hamba:g/ > *hambag* vs. *hamba:g=i* ‘handbag’
 /bolpe:n/ > *bolpen* vs. *bolpe:n=i* ‘ballpoint pen’

Each lexeme, which potentially has a long vowel, exhibits different realizations of the vowel according to the type of syllable in which it occurs. The first pattern blocks the realization of long vowels because of the presence of a coda consonant, whereas the second pattern, without a coda consonant, allows the vowel to be realized as long. This clearly comes from the moraic constraint on the syllable structure, in which any trimoraic syllable like CV:C is not allowed to occur.⁹

A similar case can be seen in nominalizing circumfixes *paC-* *-ən* and *paC-* *-an*, which cause the lengthening of the following vowel.¹⁰ This is illustrated in (12), where *tapik*, *bisag*, *lagip*, and *sirit* undergo the lengthening of the first syllable by the affixation of *paC-* *-ən* or *paC-* *-an*. However, the circumfixes do not trigger the lengthening of the vowel if the syllable in question has a coda consonant. The syllable structure again preempts vowel lengthening, because of the moraic-constraint on the syllable structure in which a trimoraic syllable such as CV:C is allowed to occur in Arta.

- (12) *pat-ta:pik-ən* ‘slapping something’
pab-bi:sag-ən ‘breaking something’
pal-la:gip-an ‘telling something, story’
pas-si:rit-an ‘defecating somewhere’
- (13) *pag-gimt-ən* ‘making/doing something’
pap-pissay-ən ‘tearing something’
pat-tuttud-an ‘sitting on, something to sit on’
pag-gusgus-an ‘scratching something’

All of the above data illustrate that the mora count is at work at least in the synchronic phonology in Arta. In addition, the mora count also explains the development of the short-long contrast in vowels as shown in §4.3.

4 The development of long vowels in Arta

As mentioned in the introductory section, all of the forms inherited from PPH lost their vowel length contrasts. A fuller list of items is shown in (14).

⁸ =*i* is a post-nominal specifier. See Table 5 in Appendix (and compare it with determiner sets shown in Table 4 in Appendix).

⁹ Another account for this vowel length alternation would be that the enclitic =*i* and the suffixes *-an* and *-ən* might trigger the lengthening of a preceding vowel. These formatives, however, do not have an ability to lengthen a preceding vowel; in fact, the following realizations do not include any lengthening: /asuk/ > *asuk* vs. *asuk=i*, /pabay/ > *pabay* vs. *pabay-an*, /idut/ > *idut-ən*.

¹⁰ *paC-* *-ən* and *paC-* *-an* are the progressive and nominalizing forms of *-ən* (patient-transitive) and *-an* (location-transitive), respectively (see Table 6 in Appendix).

- (14) PPH *ku:tu[h] > *utu* ‘lice’
 PPH *si:ku > *siku* ‘elbow’
 PPH *tu:bu > *tubu* ‘grow’
 PPH *su:su > *susu* ‘breast’
 PPH *qu:lu > *ulu* ‘head’
 PPH *bu:lan > *bulan* ‘moon’
 PPH *ŋa:jan > *ŋadin* ‘name’
 PPH *ka:yuh > *ayu* ‘tree’

A comparison with other Philippine languages which are considered to retain the old accentual system also suggests that Arta lost penultimate long vowels (the following data from Isnag, Bontok, Ifugao, Ilokano, Cebuano, and Bikol are collected from Blust and Trussel (In progress)).

- (15) Arta: *dudun* ‘locust’
 cf. Isnag, Bontok, Ifugao, Ilokano *du:dun*, Cebuano *du:lon*, Bikol *du:ron*
 Arta: *lutu* ‘cook’
 cf. Isnag, Bontok, Ifugao, Ilokano *lu:tu*, Bikol *lu:to?*
 Arta: *mula* ‘to plant’
 Isnag, Bontok, Ilokano *mu:la*
 Arta: *pusəd* ‘navel’
 cf. Isnag *pu:sag*, Bontok, Ilokano *pu:səg*, Ifugao *pu:hog*, Bikol *pu:sod*

These data strongly indicate that the long vowels in Arta are not the short-long contrast inherited from a proto-language common to languages retaining the old accent system. In what follows, it is shown that there are three types of lexemes with long vowels which should be treated separately: onomatopoeic words, loanwords, and inherited forms. After observing long vowels seen in onomatopoeic words and loanwords (the reason why onomatopoeic words should be treated separately will also be mentioned in §4.1), the most puzzling case, inherited forms containing long vowels, is discussed. It is shown that long vowels in inherited forms developed independently as a result of the loss of PMP consonants and subsequent compensatory lengthening or vowel fusion, both of which, arguably, involve the mora count conservation.

4.1 Long vowels in onomatopoeic words

Some ideophones, more specifically, expressions imitating sounds (*onomatopoeia*), contain long vowels. The following five items with a long vowel are found in my corpus.

- (16) *tattara:kot* ‘cock-a-doodle-doo (a cry of roosters)’
ku:tak ‘cluck (a cry of hens)’
be:w ‘a cry of deer’
ku:rək ‘chicken’
pi:yək ‘chick’

The first three words are the imitations of an animal’s call. The rest of them are the names of animals probably via the semantic shift metonymically from the typical sound they emit. From a methodological perspective, these words should be treated separately from other non-onomatopoeic words. It is often the case that onomatopoeic words tend to develop differently in terms of sound change. This in fact occurred in the history of the Japanese language; Komatsu (1989) argues that some onomatopoeic words retained the sound /p/ as in /pitoku pitoku/ (a sound of a small bird’s crying) even after /p/ changed into /ϕ/ in Classical Japanese. The important point is that even if the above items had been inherited from a proto-language, these are not counterexamples. They might

have developed subsequently after the loss of the old accentual system, or might have retained a long vowel in it independently of the loss of accentual system.¹¹

4.2 Long vowels in borrowed items

A large number of items containing long vowels seem to be loanwords from Yogad and Ilokano (and Spanish and English via these languages).¹² The following items share the same surface forms with those in Yogad. The items containing /k/ are clearly loans because *k was lost in Arta. The formative *ta:* in *ta:gatut* ‘one hundred’, *ta:hulu* ‘ten’, and *ta:ribu* ‘thousand’, which probably underwent the change from *sa ‘one’ > /ta:/, exhibits *s > /t/, a sound change characteristic of Cagayan Valley languages, not of Arta. The sources of *ka:ya* and *la:ku* are unclear because Ilokano also has the same forms. Considering the large number of lexical items shared with, and thus possibly borrowed from, Yogad, it seems that there is a smaller number of loanwords with a long vowel than expected. This may reflect a historical change which occurred in Yogad, one of the languages which lost the old accentual system, with a large number of lexical items having a short penult.

(17) Possible loanwords from Yogad

<i>illa:yug</i> ‘long’	<i>li:nis</i> ‘clean’
<i>i:lug</i> ‘egg’	<i>no:not</i> ‘think’
<i>ka:ya</i> ‘can, be able to’	<i>ta:gatut</i> ‘one hundred’
<i>ki:gad</i> ‘until’	<i>ta:hulu</i> ‘ten’
<i>ki:bu</i> ‘mix’	<i>ta:ribu</i> ‘thousand’
<i>la:ku</i> ‘buy’	<i>tu:rak</i> ‘write’
<i>le:but</i> ‘walk around’	<i>u:bi</i> ‘violet/purple yam’

(18) shows a list of forms shared with Ilokano. These forms are likely to be loanwords because Ilokano is a language which retains the old accentual system, with a large number of items having a long penult, which would otherwise have been shortened if they should be inherited forms from PMP. Loanwords from Spanish and English are also found possibly via Ilokano or Yogad, as shown in (19) and (20).

(18) Possible loanwords from Ilokano

<i>badu:ya</i> ‘kind of cake’	<i>bu:ya</i> ‘watch TV’
<i>inda:yun</i> ‘hammock’	<i>tu:huy</i> ‘continue’
<i>ta:wa</i> ‘window’	<i>bi:lin</i> ‘order’
<i>ba:sa</i> ‘read letters’	<i>uga:li</i> ‘habit, custom’

(19) Loanwords from Spanish

<i>amerika:no:</i> ‘American’ (< <i>Americano</i>)	<i>kalsa:da</i> ‘paved road’ (< <i>calzada</i>)
<i>antipa:ra</i> ‘goggles’ (< <i>antiparras</i>)	<i>kande:la</i> ‘candle’ (< <i>candela</i>)
<i>ari:na</i> ‘flour’ (< <i>harina</i>)	<i>kasape:gu</i> ‘matches’ (< <i>casa de fuego</i>)
<i>binta:na</i> ‘window’ (< <i>ventana</i>)	<i>kla:se:</i> ‘class’ (< <i>clase</i>)
<i>bisi:ta</i> ‘guest’ (< <i>visita</i>)	<i>pantalon</i> (/pantalo:n/) ‘pants’ (< <i>pantalon</i>)
<i>hapon</i> (/hapo:n/) ¹³ ‘Japan’	<i>sakripi:syu</i> ‘sacrifice’ (< <i>sacrificio</i>)
<i>ka:da-</i> ‘each, every’ (< <i>cada</i>)	<i>taraba:hu</i> ‘work’ (< <i>trabajo</i>)

¹¹ See also Zorc (1990), which points out that Kalamianiac languages show the irregular retention of *k in sound-symbolic monosyllabic roots and their derivatives.

¹² The following sources were used for examining diachronic changes in phonology: Blust and Trussel (In progress) for Proto-Austronesian (PAN) and Proto-Malayo-Polynesian (PMP), Reid (1979) for Proto-Northern-Luzon pronominal forms, Rubino (2000) for Ilokano, Davis and Mesa (2000) for Yogad, Headland and Headland (1974) for Casiguran Agta. Unless other sources are referred to, the references for each data will not be explicitly mentioned.

¹³ See the description in (11b) for the explanation of the short realization of the potentially long vowel.

- (20) Loanwords from English
ba:bay ‘good-bye’
basket (/baske:t/) ‘basket’
bolpen (/bolpe:n/) ‘ballpoint pen’
hambag (/hamba:g/) ‘handbag’
me:kap ‘makeup’

As the above data indicate, Arta has borrowed a large number of items used in Yogad, Ilokano, Spanish, and English. Original stresses of the items in the source languages were re-interpreted as a long vowel when borrowed into Arta. And interestingly, two vowels *o* and *e* are always reinterpreted as long vowels, as in *kla:se*: ‘class’, /bo:lpe:n/ (e.g. *bolpe:n=i* ‘a specific ballpoint pen’), /hapo:n/ (e.g. *hapo:n=i* ‘a specific Japanese person’), regardless of whether the vowel has originally a stress or not. The possible reason for the reinterpretation will be considered in §4.3.2.

Having considered the effects of borrowings, we still find other lexical items with long vowels. These items lead us to consider the third case, in which inherited forms from PMP have a long vowel as a result of several kinds of sound changes.

4.3 Long vowels in inherited forms from PMP

Arta underwent some phonological changes from PMP. A list of PMP reconstructed phonemes and their reflexes in Arta is shown in Table 1 (see Kimoto 2017 for the fuller discussion of the sound changes in Arta).

Table 1: Reflexes of PMP phonemes

PMP ¹⁴	Arta	PMP	Arta	PMP	Arta
*p	/p/	*s	/s/	*a	/a/
*t	/t/	*R	/r/	*i	/i/
*k	Ø ~ /k/ ¹⁵	*l	/l/	*u	/u/
*q	Ø	*m	/m/	*ə	/ə/
*b	/b/	*n	/n/	*a+i	/e:/
*d	/d/	*ŋ	/ŋ/	*a+u	/o:/
*j	/d/	*h	Ø		
*z	/d/	*w	/w/		
*g	/g/	*y	/y/		

In Arta, *k, *q (glottal stop) and *h are reflected as zero, *R as /r/, *j as /d/, and, as to be discussed later, the two vowel sequences *a+i and *a+u are reflected as two new vowels /e:/ and /o:/, respectively. It is argued here that there are two types of process under which inherited forms underwent vowel lengthening: compensatory lengthening and vowel fusion. Both of which involve the conservation of the mora count.

4.3.1 Compensatory lengthening

The first type of process in which vowel lengthening occurs is COMPENSATORY LENGTHENING. Compensatory lengthening is a phonological process in which the loss of a coda consonant triggers the lengthening of an adjacent segment. A typical case is the vowel lengthening triggered by the loss

¹⁴ The representation of proto-phonemes follows a conventional style of Austronesian linguistics: PMP *R is considered to have been pronounced as [ɣ], and PMP *q as [ʔ].

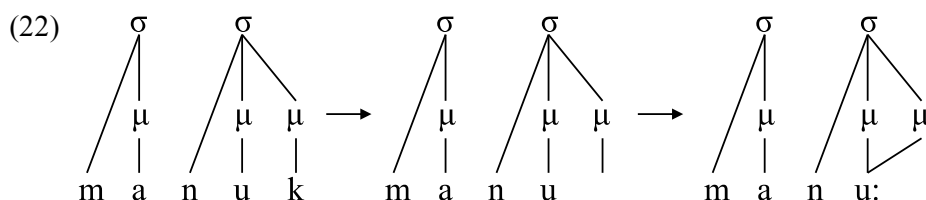
¹⁵ *k is sporadically reflected as /k/, such as *bəRək > bəRək, *anak > kanakannak ‘child’, *=ku > =ku ‘I’ (1SG.GEN).

of the following consonant, as in Latin **kasnus* > *ka:nus* ‘gray’ (Hayes 1989:260). The following items in Arta also illustrate the same point:¹⁶

- (21) **manuk* > *manu:* ‘bird’
**anak* > *ana:* ‘child’
**buyuk* > *buyu:* ‘bad-smelling’
**abak* > *abi:* ‘body’ cf. AltaS: *abek*
**buliq* > *buli:* ‘buttocks’

The development of long vowels seems to result from the loss of **k* in the case of *manu:*, *ana:*, *buyu:*, and *abi:*, and the loss of **q* in *buli:*. Since the language has the lowest percentage of retentions of reconstructed PMP vocabulary of Philippine languages, 27% (Reid 1989), not much evidence can be provided. However, in all the cases in which a coda consonant is lost, the preceding vowel is lengthened.

Compensatory lengthening can be explained in terms of the conservation of mora count, as argued by Hayes (1989, 1995). Remember that in Arta both CV: and CVC are treated as bimoraic. As depicted in (22), by dropping a coda consonant, /k/ in this case, a mora becomes empty, which is assigned to the preceding vowel by its lengthening (in the following illustration, σ represents a syllable, and μ the mora).



A more complex pattern of compensatory lengthening is found in person forms. The following enclitic pronominal forms underwent the loss of **k* (see Table 2 in Appendix for a full set of pronominal forms).

- (23) PNLzn (Reid 1979) > Arta
**=ka* > *=a* (2SG.ABS)
**=kamuyu* > *=am* (2PL.ABS)
**=kami* > *=ami* (1PL.ABS)
**=kita* > *=ita* (1+2SG.ABS)
**=kitam* > *=itam* (1+2PL.ABS)
**=ku* > *=ku* ~ *=u* (1SG.GEN)

The above forms are synchronically peculiar as well. They differ from other paradigmatically-related items in that they exhibit a complex morphophonemic alternation conditioned by the type of segments to which they attach. More specifically, if the preceding word ends with a consonant (except /n/) as in (24a), a vowel before the consonant is lengthened, and, in the case of *=ku*, /k/ is dropped. If the preceding word ends with /n/ as shown in (24b), /n/ changes to a velar, and the preceding vowel lengthened and, in the case of *=ku*, the deletion of /k/ occur. If the preceding word ends with a vowel as shown (24c), it is just followed by the enclitic in the case of *=ku*, *=ita* and *=itam*, and, in the case of *=am* and *=ami*, /y/ is inserted between the host word and enclitic.

¹⁶ Other forms which might involve compensatory lengthening are: *pura:* ‘white hair’, cf. Ilokano *purakrak* ‘(white things are) brilliant’ and *adu:yu* ‘far, distant’ (cf. Isneg *adayyu* with vowel harmony).

(24) Morphophonemic alternations

a. after a consonant:

babakat ‘old woman’ + =*ami* > *babaka:t=ami* ‘we are old women’

lusip ‘nail’ + =*ku* > *lusi:p=u* ‘my nails’

b. after /n/:

buka:gan ‘woman’ + =*ami* > *buka:ga:ŋ=ami* ‘we are women’

bunbun ‘house’ + =*ku* > *bunbu:ŋ=u* ‘my house’

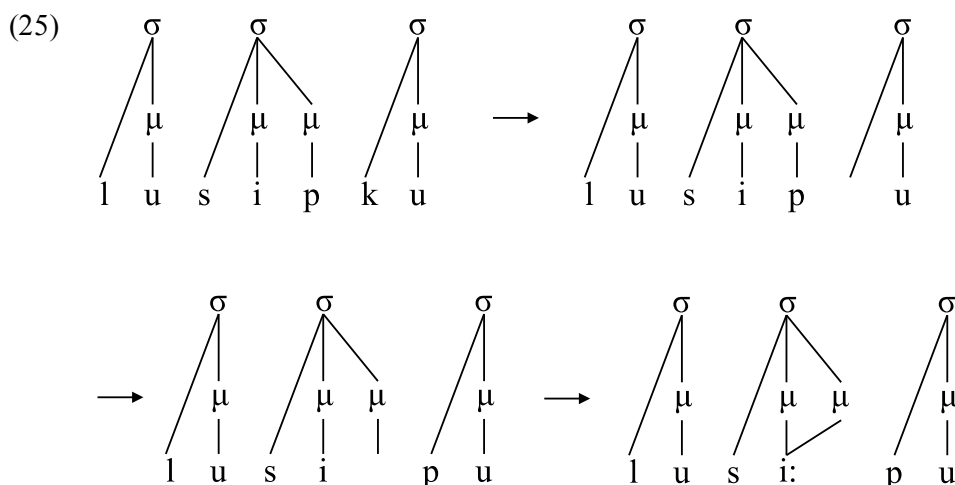
c. after a vowel (in case of =*ku*, =*ita*, =*itam*)

mata ‘eye(s)’ + =*ku* > *mata=ku* ‘my eyes’

after a vowel: insertion of /y/ (in case of =*am*/=ami)

me:na ‘go’ + =*am*/=ami > *me:nayam/yami* ‘You/we will go’

How should this complex set of morphophonemic alternations be interpreted in terms of a historical development? This synchronic fact seems to be subsumed in a “double-flop” (Hayes 1989:265), which is a subtype of compensatory lengthening. A double flop may occur when there is a cluster of consonants consisting of a syllable coda and a following onset consonant. After the loss of the onset consonant, the preceding coda consonant is re-interpreted as an onset consonant of the following syllable, and the empty mora carried by the original coda consonant is borne by the preceding vowel by lengthening. Consider the case of *lusip* ‘nail’ + =*ku* > *lusi:p=u* ‘my nails’, following the illustration in (25):



In this case, after the loss of the onset consonant /k/, the preceding segment /p/ is re-interpreted as the onset consonant probably by a universal principle of stable, unmarked syllable structure. This leads to the floating of the mora by the shift of /p/, thus being resolved by the lengthening of the preceding vowel /i/ to bear the mora. The change of /n/ to a velar nasal, in (24B), as in *buka:gan* ‘woman’ + =*ami* > *buka:ga:ŋ=ami* ‘we are women’, must be a result of assimilation to /k/; thus, it is more likely that this change occurred at the initial stage of (25), that is, before the loss of /k/.¹⁷

Positing the mora as an explanatory apparatus is further supported by the asymmetrical relation between the loss of coda and onset consonants. The moraic account predicts that the loss of onset consonants does not trigger a compensatory lengthening because the onset consonants do not bear a mora, unlike the coda consonant. Consider the following phonological changes involving the loss of onset consonants.

¹⁷ The item *la:səm* ‘sour’ (< *alsəm) involves another kind of compensatory lengthening. After the metathesis of /a/ and /l/ occurred, the bimoraic syllable seems to have been compensated for by lengthening the vowel, resulting in *la:səm* in Arta.

- (26) *kutu > *utu* ‘lice’
 *kua > *wa* ‘what-cha-ma-call-it’
 *kulit > *ulit* ‘bark, skin’
 *kaRat > *arat* ‘bite’
 *kan > *an* ‘eat’

The items shown in (26) demonstrate that a long vowel was not developed after the loss of an *onset* consonant; in fact, all the items with the loss of an onset do not exhibit vowel lengthening. This suggests that vowel lengthening in inherited forms involves the mechanism of the mora count as an explanatory apparatus.

4.3.2 Vowel fusion

The second process in which long vowels emerged is vowel fusions caused by the loss of intervocalic consonants. Vowel clusters *a+a, *u+u, *a+i, and *a+u lead to the emergence of four long vowels /a:/, /u:/, /e:/, and /o:/, respectively.¹⁸ Among them, /e:/ and /o:/ are interesting in that they are the phonemes that PMP did not have and that were developed by the very process.

The following five items exhibit vowel fusions *a+a > /a:/ and *u+u > /u:/.

- (27) *a+a > /a:/
 *sakay > /sa:y/ ‘ride on’: *mas-say* (INTR-ride), *sa:y-an* (ride-LV) ‘ride on’
 *ka-ama-ən > *ka:man* ‘big’
 *di *tahaw > /dita:w/ ‘outside’: *ditaw* ‘outside’, *dita:w=i* ‘the specific outside space’
- (28) *u+u > /u:/
 *dukut > /du:t/ ‘fire’ (cf. *dut* ‘fire’ vs. *du:t=i* ‘the fire’)
 *baqəRu > *bu:ru* ‘new’ (with a vowel harmony /a/ > /u/)

Note that, although *sakay, *tahaw and *dukut exhibit vowel fusions caused by the loss of intervocalic *k, given the mora constraint on syllable structure, a long vowel appears only if the word-final consonant is resyllabified as an onset consonant as in *sa:y-an* ‘ride on’ and *du:t=i* ‘the fire’.

The process of *a+a > /a:/ and *u+u > /u:/ is also observed in synchronic variations in some cases. A reduplication of a vowel-initial base, and the prefixation of *ma-* and *maka-* (potentive verb prefixes, see Appendix 6) followed by /a/-initial bases, trigger a vowel fusion (* represents a pre-Arta form in (29) and (30)).

- (29) *ma-ʔalap > *ma:lap* or *maʔalap* ‘can get, succeed’
 *maka-ʔaŋay > *maka:ŋay* or *makaʔaŋay* ‘can go into’
- (30) *ʔa-ʔana: (reduplication of *ana:*) > *a:na:* ‘children’
 *ma-ʔarawat-an > *ma:rawatan* ‘can grasp, understand’

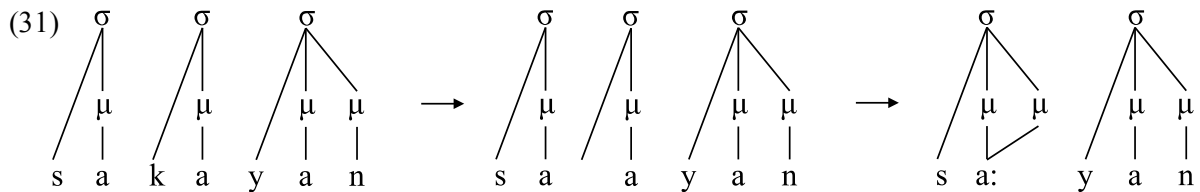
This involves the fact that an intervening glottal stop between two adjoining morphemes is synchronically on the verge of disappearing. Apart from PMP *q, which was lost at an early stage, Arta has another glottal stop, which is inserted before a vowel-initial base word regardless of whether the base word is realized by itself /ʔana:/ ‘child’, or undergoes further derivations /ʔa-ʔana:/ ‘children’.¹⁹ However, as shown above, many items have free variations in terms of the presence or

¹⁸ *i+i > /i:/ has not been attested so far, but this seems to come from the fact that Arta shows a low retention rate of PMP etyma.

¹⁹ There has been a discussion over the phonological status of the glottal stop in many Philippine and Formosan languages, which I will not get into in this paper. The author, however, consider vowel-initial glottal stop to be a phoneme rather than just a *phonetic* variant of /zero/. The glottal stop can be a target of geminates like other consonants (e.g. *meC- subəg* > *messubəg*, and *meC- a:du* > *meʔʔa:du*), thus it has a phonological status as a phonemic consonant.

absence of /ʔ/ as in (29), while some of the items do not allow the presence of /ʔ/, as in (30). And the loss of /ʔ/ triggers a vowel lengthening.

This kind of vowel lengthening by vowel fusions can be illustrated as follows. By the loss of an intervocalic consonant, two vowels adjoin each other. This leads to the re-interpretation of two homogeneous vowels as one long vowel. Note that this process again involves the conservation of mora count. After the resyllabification, the number of moras remains the same unless the structure violates the moraic constraint on the syllable structure. As mentioned above, the reason why *dukut and *sakay are realized with a short vowel *du*t and *sa*y, respectively, is that a trimoraic syllable such as ***du:t* and ***sa:y* is not allowed by that synchronic constraint.



Vowel fusions are also responsible for the development of two new phonemes /e:/ and /o:/. /e:/ emerged after the loss of *h and *k in the following items.

- (32) *ahi, *aki > /e:/
 *unahik > *une*:²⁰ ‘climb a mountain’
 *bahi > *beb-be*: ‘aunt’ (with a reduplication)
 *laki > *lel-le*: ‘uncle’ (with a reduplication)
 *maki- > *me:-, mi:-* (comitative)²¹

- (33) *a+u > /o:/
 *dahun > /do:n/ ‘leaf’ (e.g. *do:n=i*)
 *lahud > /di-lo:d/ ‘downstream’ (e.g. *dilo:d=i*) (with the fossilized prefix *di-*)

The changes *a+i* > /e:/ and *a+u* > /o:/ are again observed in synchronic morphophonemic alternations. If *ma-/maka-* or *pa-* (causative marker) is prefixed to a base beginning with /i/ or /u/, vowel fusions occur, realized as /me:/, /make:/, and /pe:/ on the one hand, and /mo:/, /mako:/, and /po:/ on the other. This again seems to involve the loss of the intervening glottal stop. This type of prefixation, however, hardly exhibits the variant with /ʔ/ except the case in (36).

- (34) *maka-/ma- (potentive), pa- + i > /make:/, /me:/, /pe:/
 *maka-ʔidəm > *make:dəm* ‘sleepy’
 *maka-inum > *make:nom* ‘drunk’
 *ma-ʔidəm > *me:dəm* ‘sleepy’
 *i-pa-ʔita (see) > *ipe:ta* ‘show’

- (35) *ma-ʔuras-an > *mo:rasan* ‘can be cleaned’
 *ma-ʔulit-an > *mo:litan* ‘can be peeled’
 *pa-ʔudiŋ-ən > *po:diŋən* ‘darken’

²⁰ The phonological change in *unahik > *une*: involves the reduction of mora; *nahik contains three moras, whereas *ne*: contains two moras. This reduction seems to come from the moraic constraint on the syllable, that is, a trimoraic syllable such as *ne::* is not allowed in the language.

²¹ See Table 6 in the Appendix for a full list of verbalizing affixes.

- (36) *ma-ʔune: > *maʔune:* ~ *mo:ne:* ‘can climb’

It should be noted that vowel fusions did not occur in the case of vowel + glide clusters, such as *ay and *aw. In Casiguran Agta, *ay was fused into /ɛ/ (Headland and Healey 1974), but that is not the case in Arta. With one exception in which *patay is reflected as /pati/, *ay is reflected as /ay/.

- (37) *ay > /ay/ (cf. *ay > /ɛ/ Casiguran Agta)
 *aNay > *anay* ‘go’ (cf. Casiguran Agta *aŋɛ*)
 *anay > *anay* ‘termite’ (cf. Casiguran Agta *ane*)
 *wasay > *wasay* ‘ax, hatchet’
 *patay > *pati* ‘die’

- (38) *aw > /aw/
 *ulitaw > *ulitaw* ‘unmarried man’
 *tahaw > *di-taw* ‘inside’ (with the fossilized prefix *di-*)
 *biRɲaw > *biriɲaw* ‘fly (n.)’
 *lɔtaw > *lɔtaw* ‘float’

The traditional analysis in which /y/ and /w/ constitute part of a diphthong is not applied to Arta. /y/ and /w/ in the language are treated as consonants in synchrony as well. Pronominal enclitics such as =*ku* are realized differently depending on whether the preceding word ends with a consonant (realized as =*u*), or with a vowel (=ku). /y/ and /w/ exhibit the same pattern as other consonants (e.g. *tataw* > *tata:w=u*, not *tataw=ku* ‘I know’). There is no reason that /y/ and /w/ in the coda position should be analyzed separately from other consonants, and it is inappropriate to consider the vowel-glide sequences as diphthongs.

In §4.2, it is observed that /e:/ and /o:/ in borrowed items are always long vowels unless the length is overridden by the constraint on the syllable structure. It seems difficult to understand the reason why only these two vowels are always long. However, the above discussion on the historical development of /e:/ and /o:/ provides one reasonable account. That is, the feature of length which the two vowels acquired via the fusion of *a+i and *a+u must be applied to the vowels in loanwords. In the first stage, /e:/ and /o:/ were seen exclusively in the items which underwent vowel fusion. But in the process of interpreting loanwords containing [ɛ] and/or [ɔ] as the instances of the two existing vowels /e:/ and /o:/, the feature [+long] also penetrated borrowed items.

5 Competing motivations

The above section showed that long vowels in Arta are observed in inherited forms from PMP, and they are captured by compensatory lengthening and vowel fusion, both of which are subsumed under the principle of mora-count conservation. This suggests that the mora-based principle played a significant role in Arta, as well as the moraic constraint on the eligible syllable structure. However, this does not imply that the mora affects the phonetic and phonological organization in the language. Rather, a more naturalistic view would be that it is at work interacting, and sometimes competing, with other factors which may affect the organization (cf. “competing motivations” (DuBois 1985)). Some apparent exceptions to the mora principles tell us the interactions between the mora and other factors such as sociolinguistic and phonetic aspects.

The first case in which the regularity of the mora may be affected involves borrowings. The two lexical items *trabajo* in Spanish and *truck* in English were borrowed into Arta, meaning ‘work’ and ‘car’, respectively. The peculiar feature with the two items is that they have an onset cluster *tr*. This structure seems unstable in the phonology of Arta; in fact, they exhibit the following phonetic variations with/without vowel insertions.

- (39) Spanish *trabajo* > Arta [traba:hu] ~ [t̄araba:hu] ~ [taraba:hu] ‘work’
 English *truck* > Arta [trak] ~ [t̄arak] ~ [tarak] ‘car’

The onset cluster consisting of /t/ and /ɾ/ may or may not have the insertion of [a] with different lengths. It is difficult to assume abstract phonological representations for each lexical item, and difficult to posit any specific syllable structure or mora count. Rather, this should be considered to be a case in which the application of mora count is unstable after adopting an unstable structure through borrowing.

A more puzzling case is found in inherited forms. The following two items are not explained by the borrowing of “unstable structure” from outside of the language, because they seem to be reflexes of PMP *diya and *tiyan, respectively. However, their reflexes again show instability in terms of the mora count and syllable structure.

- (40) *tiyan > [tʰi(j)an] ~ [tʰʌn] ‘belly’
 *diya > [dʰi(j)a] ~ [dʒa] ‘to him/her/it’ (3SG.OBL)

Each item may be pronounced like a bisyllabic word or with the palatalization of /t/ and like a monosyllabic word ([j] is parenthesized because whether a glide should be recognized is difficult to judge purely in terms of a phonetic observation). Furthermore, even when there is no palatalization of /t/, [i] is pronounced as shorter (hence the diacritic above [i]), and it is not obvious as to whether the words should be regarded as monosyllabic or bisyllabic words. This of course means the indeterminacy of the mora count as well.

This indeterminacy might be interpreted as the case in which the items are on the way of a gradual language change from bimoraic to monomoraic words. This, however, seems overgeneralized. It does not explain exactly why these particular items undergo the change in the mora count, and it is shown in §4.3 that the mora count was conserved, rather than reduced, in the process of phonological change. A more crucial factor applying to this particular case would be attributed to phonetic aspects. This seems to involve the fact that a sequence of two segments *iy shares the same phonetic features: front, close, and unrounded, thus it is easy to produce with a reduction. And more crucially, the articulatory movement from /t/ or /d/ to /a/ constitutes a single unidirectional movement from a complete closure to a fully opened position. The pressure by the ease of this particular articulatory movement may facilitate a reduction to monosyllabic words by kicking out the principle of the mora-count conservation.

As Blust (2013) discusses, mid-central vowel /ə/, or schwa has a systematic irregularity in terms of the mora count. The following four items in Arta show the irregularity, in which *ə is reflected with no mora value.²²

- (41) *tuqəlan > /tulan/ ‘bone’
 *pahəzam > /padam/ ‘lend’
 *bituqən > /bitun/ ‘star’
 *ka-ama-ən > /ka:man/ ‘big’

In the above cases, *tuqə is reflected as /tu/, *pahə as /pa/, *tuqən as /tun/, and *ma- -ən as /man/. The mora, which would be borne by other vowels, is not associated with the schwa. Furthermore, the following items containing schwa in the final syllable exhibit an irregular reflex. By the loss of final coda consonants, CVC syllables are reflected as monomoraic syllable with a short /a/.²³

²² Note that the reflex of PMP *baqəRu is *bu:ru* ‘new’ with a long vowel. This reflex exhibits an exceptional pattern. This might involve an intermediate stage *baqRu (by the loss of schwa with *q shifting to the coda position), which could be a target of compensatory lengthening after the loss of the coda *q.

²³ It is not obvious why the reflex of *CəC is *Ca*, rather than any other vowel. However, this reflex may be relevant to the fact that when schwa occurs in the position which is required to be lengthened, the segment is phonetically realized as a sound similar to [a].

- (42) *abək > *aba* ‘mat’
 *utək > *uta* ‘brain’
 *pədək > *pəda* ‘step on’
 *mutəq > *muta* ‘gummy section of eyes’
 *luhəq > *luwa* ‘tears’
 *basəq > *bisa* ‘wet’ (with low vowel fronting (LVF): *a > /i/)

The relationship between the schwa and the mora count differs from the case in other vowels. Phonologically, /Cə/ is treated as zero mora and /CəC/ as one mora. Note that this generalization is established from a phonological or emic perspective. Seen from a phonetic or etic perspective, the length cannot be zero because it must have a physical length. Considering the internal logic of the language according to which syllables containing schwa are treated phonologically, it can be argued that the language treats the syllables with schwa as having one mora fewer than the syllables with other vowels. However, it is likely that this phonological treatment comes from the phonetic shortness of schwa. In explaining the relationship between stress and schwa, Blust (2013) claims that “[t]his behaviour derives from subphonemic differences of length in the vowels of PAN and their reflexes in many daughter languages, in which the schwa appears to be extra short” (ibid.:256). It seems also to be the case in Arta. The irregular reflex in the language may come from the phonetic peculiarity of schwa and historical reason which dates back to PAN.

6 Conclusions

This paper attempted to argue that the development of long vowels in Arta, which once lost the old accentual system, is largely explained by the notion of the mora. The first part of this paper discussed the validity of applying the mora to the synchronic phonology in Arta, in which it is argued that some allomorphic distribution is conditioned depending on whether the preceding syllable is monomoraic or bimoraic, and that vowel length alternation is explainable by the moraic constraint on the syllable structure. There are three cases in which a long vowel occurs in Arta: onomatopoeia, borrowings, and inherited forms from PMP. It is shown that vowel length in the inherited forms could be explained at least by assuming the principle of mora-count conservation as seen in compensatory lengthening and vowel fusion, as well as the moraic constraint on the syllable structure.

There are several items which seem not to be explained by compensatory lengthening or vowel fusion. The forms listed in (43) and (44) have a long vowel, which probably requires some other phonological and/or morphological mechanisms to explain.

- (43) *hadu > *a:du* ‘many, much’ (cf. Northern & Southern Alta *ádu*)
 *ikan > *i:yan* ‘fish’
 *zuRu > *di:ru* ‘soup’
 *ikəj ‘cough’ > *i:yər* ‘phlegm’²⁴
- (44) *u > /o:/ in doubled monosyllables
 *kutkut > /kɔ:tkɔ:t/ ‘dig’
 *pukpuk > /pɔ:kpo:k/ ‘beat (with a hammer)’
 *tuktuk > /to:kto:k/ ‘top, summit’
 *həyup > /yo:pyo:p/ ‘blow on’

²⁴ This item may not be an inherited form because *j is reflected as /r/, not /d/ (see Table 1).

The following items are the ones whose sources remain to be identified:

- | | |
|---|--|
| (45) <i>alilyo:gən</i> ‘kind of ghost’ | <i>di:muy</i> ‘bathing’ (cf. Southern Alta <i>di:muy</i>) |
| <i>atti:</i> ‘exist’ (cf. Ilokano <i>adda</i>) | <i>diso:no:</i> ‘inside’ |
| <i>aydi:</i> ‘and’ | <i>dupu:</i> ‘old man’ |
| <i>ba:kəw</i> ‘grain, corn’ | <i>o:gip</i> ‘lying’ |
| <i>baruwa:si</i> ‘clothes’ | <i>pulot/pulo:t/</i> ‘loincloth’ (cf. Maranao: <i>mampoyot</i>) |
| <i>be:kut</i> ‘ghost’ | <i>pu:nəd</i> ‘rain’ |
| <i>bi:lat</i> ‘python’ | <i>sa:ŋan</i> ‘how many’ |
| <i>bidi:yu</i> ‘buri palm’ | (cf. Central Cagayan Agta <i>haŋan</i>) |
| (cf. Casiguran Dumagat <i>bidi:yo</i>) | <i>si:paŋ</i> ‘one’ |
| <i>biyu:ŋət</i> ‘night’ | <i>ta:me:ta</i> ‘different’ |
| <i>de:kət</i> ‘stickyness’ | = <i>te:</i> ‘only’ |

In fact, the sources of many items in Arta are not identified yet because of its low retention rate of PMP etyma. Further studies need to be conducted to reveal how long vowels are developed in the items observed.

Appendix: Sets of grammatical forms in Arta

Table 2: Pronominal forms

	TOP	ABS	GEN/ERG	OBL
1SG	<i>tən</i>	= <i>tən</i>	= <i>ku</i>	<i>dən</i>
1PL	<i>tami</i>	= <i>ami</i>	= <i>mi</i>	<i>dami</i>
1+2SG	<i>tita</i>	= <i>ita</i>	= <i>ta</i>	<i>dita</i>
1+2PL	<i>titam</i>	= <i>itam</i>	= <i>tam</i>	<i>ditam</i>
2SG	<i>taw</i>	= <i>taw</i>	= <i>mu</i>	<i>daw</i>
2PL	<i>tam</i>	= <i>am</i>	= <i>muyu</i>	<i>dam</i>
3SG	<i>siya</i>	= <i>siya</i> , \emptyset	= <i>na</i>	<i>diya</i>
3PL	<i>tidi ~ tidu</i>	= <i>tid</i>	= <i>di</i>	<i>did, didu</i>

Table 3: Demonstratives

		TOP	ABS	GEN/ERG	OBL
PROXIMAL	SG	<i>si:yəy</i>	<i>a:yi:</i> = <i>i</i>	<i>ni/na a:yi:/əyni</i> = <i>ni</i>	<i>ti/ta a:yi:</i> = <i>ti</i>
	PL	<i>satidi:</i>	(<i>ay</i>) <i>tidi a:yi:</i>	(<i>ay</i>) <i>didi a:yi:</i>	(<i>ay</i>) <i>didi a:yi:</i>
MEDIAL	SG	<i>sayna</i>	<i>a:yina,</i> = <i>ina</i>	<i>ni/na ayna</i> = <i>nina</i>	<i>ti/ta ayna</i> = <i>tina</i>
	PL	<i>satidi:na</i>	(<i>ay</i>) <i>tidi:na</i>	(<i>ay</i>) <i>didi:na</i>	(<i>ay</i>) <i>didi:na</i>
DISTAL	SG	<i>saya</i>	<i>a:ya:, =ya:</i>	<i>ni/na a:ya:</i>	<i>ti/ta a:ya:</i> = <i>ta</i>
	PL	<i>satiddya:</i>	(<i>ay</i>) <i>tiddya</i>	(<i>ay</i>) <i>didi a:ya:</i>	(<i>ay</i>) <i>didi a:ya:</i>

Table 4: Determiners

			TOP/ABS	GEN/ERG	OBL
DEFINITE	SG	PERSONAL	<i>ti</i>	<i>ni</i>	<i>ni</i>
		COMMON	<i>i</i>	<i>ni</i>	<i>ti</i>
	PL		<i>tidi</i>	<i>didi</i>	<i>didi</i>
INDEFINITE			\emptyset	<i>na</i>	<i>ta</i>

Table 5: Specifiers

SPECIFIC ²⁵	= <i>i</i>
SPECIFIC (PAST)	= <i>ti</i>
SPECIFIC (PLURAL)	= <i>tidi</i>

Table 6: Verbs

VERBS	SIMPLE NONPAST	SIMPLE PAST	PROGRESSIVE/ NOMINALIZATION
DYNAMIC INTRANSITIVES (ACTOR-VOICE)			
PUNCTUAL (*um)	<um>	<in><um>	–
DISTRIBUTIVE ²⁶ (*maN-)	maN-	(mi)naN-	paC- ²⁷
DURATIVE (*maR-)	maC-	(mi)naC-	paC-
COMITATIVE (*maki-)	me:-, meC-	ne:-, neC-	pe:-
DYNAMIC TRANSITIVES (UNDERGOER VOICES)			
PATIENT-TRANSITIVE	-ən	<in>	paC- -ən
LOCATION-TRANSITIVE	-an	<in> -an	paC- -an
CONVEYANCE-TRANSITIVE	i-	(i)ni-	paC-
<i>ma-/maka-</i> POTENTIVES			
INTRANSITIVE	maka-	naka-	–
TRANSITIVE -ən	ma-	(mi)na-	–
-an	ma- -an	(mi)na- -an	–
i-	ma-, me:-	(mi)na-, (mi)ne:-	–
STATIVES			
AGENTIVE STATIVE	tiC-	–	–
PATIENTIVE STATIVE	maŋa:-	–	–
POSSESSIVE STATIVE	makaN-	–	–
CAUSATIVE DERIVATIONS			
INTRANSITIVE CAUSATIVE	mampa-, mama-	(mi)nampa-, (mi)nama-	pappa-
TRANSITIVE CAUSATIVE			
-ən	pa- -ən	pina-	pappa- -ən
-an	pa- -an	pina- -an	pappa- -an
i-	ipa-, pe:-	nipa-	pappa-

²⁵ Definiteness refers to a given-new distinction varying depending on the hearer's knowledge assumed by the speaker (Prince 1981), whereas specificity is another grammatical distinction relevant to a speaker's knowledge as to whether s/he can identify the referent or not, as is well-known in the two distinct readings of the following sentence: *I want a car*. See Lambrecht (1994).

²⁶ *maN-* prefixation does not trigger the deletion of stem-initial consonants (e.g. *taradtad* > *man-taradtad* 'escape, run away', *pili* > *mam-pili* 'choose').

²⁷ *maN-* and *maC-* show syncretism in progressive/nominalization form.

Table 7: Interrogatives

what	<i>a:nu</i>
who	<i>tatin</i>
when	<i>tanakan</i>
where	<i>adin</i> (past: <i>adinti</i> , present/future: <i>adi:ni</i>)
why	<i>ataʔay ~ adaʔay</i>
how (manner)	<i>kassandi</i>
how many/much	<i>sa:ŋan, saŋan</i>

Table 8: Negator, existentials and quantifiers

SENTENTIAL NEGATION	<i>awan</i>
EXISTENTIAL	<i>atti: (+ABS)</i>
EXISTENTIAL NEGATION	<i>awan (+OBL)</i>
many, much	<i>meʔa:du</i>
few, little	<i>killək</i>
one (of)	<i>gissa</i>
some (of)	<i>a:duwan</i>
all (of)	<i>attanān, atanan</i>

Table 9: Second-position enclitics

ASPECTUAL ENCLITICS	
again	<i>=mandi</i>
already, soon	<i>=di ~ =d</i>
still	<i>=tep</i>
just do, do a few	<i>=pa ~ =p</i>
EVALUATIONAL ENCLITICS	
only	<i>=te:</i>
PRAGMATIC ENCLITICS	
REQUEST	<i>=nen</i>
HYPOTHETICAL	<i>=mina</i>
MIRATIVE	<i>=sika, =mat</i>
HEARSAY	<i>=an</i>
META-COMMUNICATIVE ²⁸	<i>=hug, =ay</i>

References

- Blust Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In *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*. Revised edition. ANU. URL: <http://hdl.handle.net/1885/10191>
- Blust, Robert, and Stephen Trussel. In progress. *Austronesian Comparative Dictionary*, Web Edition. www.trussel2.com/ACD.
- Davis, Philip W., and Angel D. Mesa. 2000. *A dictionary of Yogad*. München: Lincom.

²⁸ The formatives *=hug* and *=ay* which we call here “meta-communicative” are used to embed into the utterance the implicature that the utterance is conveying redundant information which should have been in the common ground because the information is already mentioned (*=hug*), or because it can be inferred from the linguistic/non-linguistic context (*=ay*).

- Du Bois, John W. 1985. Competing motivations. In *Iconicity in syntax*, ed. by John Haiman, 343-365. Philadelphia: John Benjamins Publishing.
- Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20:253-306.
- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press.
- Headland, Thomas N. 2003. Thirty endangered languages in the Philippines. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 47:1-12.
- Headland, Thomas N., and Janet D. Headland. 1974. *A Dumagat [Casiguran]-English dictionary*. Canberra: Pacific Linguistics.
- Headland, Thomas N., and Alan Healey. 1974. Grammatical sketch of Casiguran Dumagat. *Papers in Philippine Linguistics* No.5, 1-54. Pacific Linguistics Series A 43. Canberra: Pacific Linguistics.
- Kimoto, Yukinori. 2017. A grammar of Arta: A Philippine Negrito language. Ph.D. dissertation, Kyoto University.
- Ko, Albert Min-Shan, Chung-Yu Chen, Qiaomei Fu, Frederick Delfin, Mingkun Li, Hung-Lin Chiu, Mark Stoneking, and Ying-Chin Ko. 2014. Early Austronesians: Into and out of Taiwan. *The American Journal of Human Genetics* 94:426-436.
- Komatsu, Hideo. 1989. Nihongo no Rekishi: On-in [The History of Japanese: Phonology]. *The encyclopaedia of linguistics*, Vol.2, ed. by Takashi Kamei, Rokuro Kono, and Ei-ichi Chino, 1634-53. Tokyo: Sansei-do.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In *Radical pragmatics*, ed. by Peter Cole, 223-255. New York: Academic Press.
- Rasmussen, Morten, et al. 2011. An aboriginal Australian genome reveals separate human dispersals into Asia. *Science* 334(6052):94-98.
- Reid, Lawrence A. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication 8. Honolulu: University of Hawaii Press.
- Reid, Lawrence A. 1979. Towards a reconstruction of the pronominal systems of Proto-Cordilleran, Philippines. In *South-East Asian linguistic studies* 3, ed. by Nguyen Dang Liem, 259-275. Pacific Linguistics C-45. Canberra: Australian National University.
- Reid, Lawrence A. 1982. The demise of Proto-Philippines. In *Papers from the Third International Conference on Austronesian Linguistics, vol. 2: Tracking the travellers*, ed. by Amran Halim, Lois Carrington, and S. A. Wurm, 201-216. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 1989. Arta, another Philippine Negrito language. *Oceanic Linguistics* 28(1):47-74.
- Reid, Lawrence A. 1991. The Alta languages of the Philippines. In *VICAL 2, Western Austronesian and contact languages: Papers from the Fifth International Conference on Austronesian linguistics*, ed. by Ray Harlow, 265-297. Auckland: Linguistic Society of New Zealand.
- Reid, Lawrence A. 1994. Possible non-Austronesian lexical elements in Philippine Negrito languages. *Oceanic Linguistics* 33(1):37-72.
- Reid, Lawrence A. 2006. On reconstructing the morphosyntax of Proto-Northern Luzon, Philippines. *Philippine Journal of Linguistics* 37:1-64.
- Reid, Lawrence A. 2009. The reconstruction of a dual pronoun to Proto Malayo-Polynesian. In *Discovering history through language: Papers in honour of Malcolm Ross*, ed. by Bethwyn Evans, 461-477. Canberra: Pacific Linguistics.
- Reid, Lawrence A. 2013. Who are the Philippine Negritos? Evidence from language. *Human Biology* 85:329-358.

- Reyes-Centeno, Hugo, Silvia Ghirotto, Florent Détroit, Dominique Grimaud-Hervé, Guido Barbujani, and Katerina Harvati. 2014. Genomic and cranialphenotype data support multiple modern human dispersals from Africa and a Southern route into Asia. *PNAS* 111(20):7248–7253.
- Robinson, Laura C, and Jason William Lobel. 2013. The Northeastern Luzon subgroup of Philippine languages. *Oceanic Linguistics* 52(1):125–168.
- Ross, Malcolm. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2):1–24.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*, ed. by Henry Y. Chang, Lillian M. Huang, and Dah-an Ho, 521-563. Language and Linguistics Monograph Series Number W-5. Taipei: Institute of Linguistics, Academia Sinica.
- Rubino, Carl Ralph Galvez. 2000. *Ilocano dictionary and grammar*. Honolulu: University of Hawaii Press.
- Zorc, R. David. 1979. On the development of contrastive word accent: Pangasinan, a case in point. In *South-East Asian linguistic studies* 3, ed. by Nguyen Dang Liem, 241-258. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Zorc, R. David. 1990. The Austronesian monosyllabic root, radical or phonestheme. In *Linguistic change and reconstruction methodology*, ed. by Philip Baldi, 175–194. Berlin: Mouton de Gruyter.
- Zorc, R. David. 1993. Overview of Austronesian and Philippine accent patterns. In *Tonality in Austronesian languages*, ed. by Jerold A. Edmondson and Kenneth J Gregerson, 17–24. Honolulu: University of Hawaii Press.

© 2017 University of Hawai'i Press

All rights reserved

OPEN ACCESS – Semiannual with periodic special publications

E-ISSN: 1836-6821

<http://hdl.handle.net/10524/52405>



Creative Commons License

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

JSEALS publishes fully open access content, which means that all articles are available on the internet to all users immediately upon publication. Non-commercial use and distribution in any medium is permitted, provided the author and the journal are properly credited.

Cover photo courtesy of Alexander Smith.

JSEALS Special Publication No. 1

**ISSUES IN AUSTRONESIAN
HISTORICAL LINGUISTICS**



Edited by Hsiu-chuan Liao



UNIVERSITY of
HAWAII
PRESS