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

Yukinori Kimoto
Kyoto University
<*kimoto@hi.h.kyoto-u.ac.jp*>

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 it, but the items with a long penultimate which are shared with other Philippine languages which retain old accentual system, are reflected as short, 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 the mora count conservation, the phonological changes which occurred in Arta indicate that the mora played significant roles in the language.

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

1 Introduction

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

(1) Tagalog *áso* [ʔaːso] 'dog'   vs. *asó* [ʔaso] 'smoke'
     Ilokano *bára* [baːra] 'hot'   vs. *bará* [bara] 'lung'
     Bikol   *bága* [baːga] 'ember' vs. *bagá* [baga] 'truly!'

(Zorc 1993: 18)

This kind of vowel-length contrast is also found in diverse Philippine languages, including Aklanon, Balangao, Cebuano, Hanunoo, Ibanag, Ifugao, Isneg, Kalinga, Kapampangan, Sambal (Zorc 1978: 241). Based on the fact that cognate forms among these languages share the same length on the penult, Zorc (1978) argues that the contrastive word accent system may be attributed to "Proto-Philippines",[1]

---

[1] Although Zorc presumes a hypothetical proto-language "Proto-Philippines", and this paper also tentatively uses the label "Proto-Philippine" for the purpose of comparing reconstructed forms with Arta forms, this does not mean that the author agrees with the idea; rather, recent studies (Reid, Ross) point out the weakness of the phonological evidence as a distinct proto-language from Proto-Malayo-Polynesian.

as shown in PPʜ *da:Raq cf. Isneg *da:ga*, Ilokano *da:ra*, Ifugao *da:la*, Kapampangan Sambal *da:ya?*; PPʜ *da:lan cf. Aklanon, Balangao, Cebuano, Bikol, Hanunoo, Ibanag, Isneg, 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 cognate forms with long penult in the reconstructed language are all reflected with short vowels:

(2)   PPʜ *si:ku > Arta *siku* 'elbow'
        PPʜ *tu:bu > Arta *tubu* 'grow'

The aim of this paper is thus to explain the historical development of long vowels in Arta. It is shown here that there are three sources where long vowels are found: (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 on the Arta language. In §3, it is argued that the mora plays significant roles in synchronic phonology in Arta. The next section deals with the matter as to how the long vowels in Arta historically developed. After discussing that the first two sources are relevant to the items with long vowels in Arta, §4 observes sound changes and some factors which 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 characteristic of their 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 ago before the speakers of an Austronesian language migrated into the islands; since Negrito people switched their original languages to Austronesian languages, the current language that the Arta people speak clearly belongs to the Austronesian family.[2]

---

[2] 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 =*muyu* (second person plural genitive form), which could date back to Proto-Northern Luzon (Reid 1989), and perhaps to PMP (see Reid 2009, but Ross 2006 proposes the formatives *=ihu, *=mu-ihu for the second person plural genitive).

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, is running from south to north, forming Cagayan Valley. The valley is sandwiched between Sierra Madre mountains in the east

Figure 1. *Northern part of Luzon*

Figure 2 *Quirino province (seen from south)*

side, 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 Dupaningan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta. The valley of 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 4–5 decades ago; they currently settled down in Disimungal in Nagtipunan, where there are two communities, Pulang Lupa and Kalbo.

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 of 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.[3] 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 useless; in my fieldwork, I have been

---

[3] 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 with 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.

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.

2.2 Subgrouping relationship and language contacts

The genetic subgrouping of Arta was studied in Reid (1989), in which 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 northern part of Luzon, surrounded by Basiic 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, North-eastern Luzon languages including Dupaningan Agta, Pahanan Agta, Casiguran Agta, and Dinapigue Agta.[4] 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 *taŋis > PNLZN *saŋit. Although Reid (1989, 2013) do not provide such evidence for the subgrouping of Arta (thus the argument for placing Arta as a primary branch of NLZN was insufficient), the metathesis does exist:

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

Since *ʔ was lost and *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 Northern Luzon.

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 from Ilokano, which is still difficult to subgroup with, partially because most of the forms are not uniquely shared. Even the forms which is expected to share the reflex of *R are reflected differently: *bəRŋaw > Arta: *biriŋaw*, Ilk. *bərŋaw* '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 reflex of *R as /r/ than Ilokano). In Arta, PMP *j consistently changed into /d/, thus merged with *d, which exhibits a further difference from Ilokano, where *j is reflected as /g/ (Reid 1989).

This subgrouping however reflects only an aspect of the linguistic history of Arta; the speakers of Arta have undergone several major language contacts. The ancestors of Philippine Negrito groups including Arta, are considered to have dispersed into the Philippine archipelago in the Paleolithic age over 20,000 years ago. It is suggested by current genetic and archaeological studies that, after a 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 New Guinea, Melanesia, and Australia ~62–75,000 years ago (Rasmussen et al. 2011, Reyes-Centeno et al. 2014). These populations who settled down 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 remaining in some current Negrito languages (Reid 1994).

---

[4] Northern and Southern Altas are also the Negrito languages subgrouped within Northern Luzon (see Reid 1989, 1991).

The early Austronesians residing in Taiwan migrated into the Philippines probably via 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. In particular Yogad was influential on Arta: they share at least 98 lexical items. This is probably because the Arta people had been residing in Alicia until 4–5 decades ago, and they were bilingual between Arta and Yogad until recently. The speakers of Arta claim that they are similar like "we are relatives".

Following a basic description on 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 language. It is argued that the mora is at work in various phonological aspects seen 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/. Vowel length is phonemically distinctive as in *bi:ləg* 'bracelet' vs. *biləg* 'fastness', *ana:* 'child' vs. *a:na:* 'children', *bubuy* 'pig', vs. *bo:buy* 'grandmother'. /i/, /a/, /u/, and /ə/ have long and short phonemic contrasts[5]; however, /e/ and /o/ always appear as long vowels /e:/ and /o:/ unless they are reduced as a short vowel within the CVC syllable template (as discussed in 3.2, long vowels cannot appear within CVC).

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 a vowel length rather than 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 by affixation as in English: *phótograph > photógraphy > photográphic*. This is the case in Ilokano, as in *ba:sa > basa:-en* and *la:wa > ka-lawa:-en*, but this kind of shift is not observed in Arta:

(4)   *pe:nas > pe:nas-an* (not *pena:s-an*) 'wipe'
        di:muy 'bathing' > *pandi:muy-an* (not *pandimu:y-an*) 'bathing place'
        ka:lig > ka:lig-ən (not *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, the other secondary (e.g. *dòcuméntary, téxt-bòok*). This is not the case in Arta. Several words do have more than one long vowels without such phonetic asymmetry. Consider the following forms.

(5)   More than one long vowels 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'

---

[5] /ə/ does not appear as a long vowel except in one item *də:gi* 'band for carrying a basket'.

pe:be:bu:d=u 'I am asking'

Since none of these two facts favour 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 monomoraic syllable defined as the open syllable with a short vowel. Languages may differentiate monomoraic syllable (light syllable) with a short vowel (CV), bimoraic syllable (heavy syllable) with a long vowel (CV:), and, in some languages, 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 below:

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

(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 as follows:

(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 excess 2 moras

First, as 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::, or CV:C. If morphological conditions require such syllable structure, these syllables should be reduced into CV:, and CVC, respectively. This constraint can be paraphrases as (9): "The syllable should not excess 2 moras." Now let me introduce some mora-sensitive phenomena in Arta phonology, which support the statement in (8, 9).

3.2.1 Allomorphs of enclitics

The mora count is necessary for generalizing some allomorphemic 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 respectively, 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 close. 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 1 mora in the language.

3.2.2 *Vowel length alternation*

The moraic constraint on the maximal weight of the syllable structure, explicated in (8, 9), predicts vowel length alternations. If a long vowel appears in a close 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:

(11) /latt**o:**ŋ/ > *latto**ŋ*** vs. *latt**o:**ŋ=i*[6] 'outside'
      /s**a:**y/ > *mas-s**ay*** vs. *s**a:**y-an* 'ride on'
      /dag**e:**t/ > *man-dag**et*** vs. *dag**e:**t-ən* 'sew'
      /d**u:**t/ > *d**ut*** vs. *d**u:**t=i* 'firewood'

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, where any trimoraic syllable like CV:C is not allowed to occur.[7]

A similar case can be seen in nominalizing circumfixes *paC- -ən* and *paC- -an,* which cause the lengthening of the following vowel.[8] 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 a vowel lengthening, because of the moraic-constraint on the syllable structure where 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-**gim**t-ən* 'making/doing something'
      *pap-**pis**say-ən* 'tearing something'
      *pat-**tut**tud-an* 'sitting on, something to sit on'

---

[6] *=i* is a post-nominal specifier. See Appendix 3 (and compare it with determiner sets shown in Appendix 2).

[7] Another account for this vowel length alternation would be that the enclitic *=i* and suffixes *-an* and *-ən* might be triggering 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/ > *idutən*.

[8] *paC- -ən* and *paC- -an* are the progressive and nominalizing forms of *-ən* (patient-transitive) and *-an* (location-transitive), respectively (see Appendix 6).

*pag-**gus**gus-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 what follows.

4 The development of long vowels in Arta

As mentioned in the introductory section, all of the cognate forms inherited from PPH lost its vowel length contrast. A fuller list of the items is shown here:

(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 for Isneg, Bontok, Ifgaw, Ilokano, Cebuano, and Bikol are collected from Blust and Trussel (2010–)):

(15)   Arta: *dudun* 'locust'
        cf. Isneg, Bontok, Ifugaw, Ilokano *du:dun*, Cebuano *du:lon*, Bikol *du:ron*
        Arta: *lutu* 'cook'
        cf. Isneg, Bontok, Ifugaw, Ilokano *lu:tu*, Bikol *lu:toʔ*
        Arta: *mula* 'to plant'
        Isneg, Bontok, Ilokano *mu:la*
        Arta: *pusəd* 'navel'
        cf. Isneg *pu:sag*, Bontok, Ilokano *pu:səg*, Ifugaw 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 lexeme with long vowels which should be treated separately: onomatopoeic words, borrowings, and inherited forms. After observing long vowels seen in onomatopoeic words and borrowings (where the reason why onomatopoeic words should be treated separately will also be mentioned), the most puzzling case, inherited forms containing long vowels is discussed. It is shown that long vowels in inherited forms independently developed by the loss of PMP consonants and subsequent compensatory lengthening or vowel fusion, both of which, arguably, involves the mora count conservation.

4.1 Long vowels in onomatopoeic words

Some idephones, more specifically, expressions imitating sounds (*onomatopeoia*) 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 animal's crying. 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, even after /p/ changed into /ɸ/ in Classical Japanese, some onomatopoeic words retained the sound /p/ as in /pitoku pitoku/ (a sound of a small bird's crying). The important point is that even if these 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.

4.2 Long vowels in borrowed items

A large number of items containing long vowels seem to be borrowings from Yogad and Ilokano (and Spanish and English via these languages).[9] The following items share the same surface forms with those in Yogad. The items containing /k/ are more likely to be borrowings because Arta changed *k into zero. The formative *ta:* in *ta:gatut, ta:hulu,* and *ta:ribu*, which probably underwent the change from *saga- > /ta:/, exhibit a sound change *s > /t/ characteristic of Cagayan Valley languages, not of Arta. The sources of *ka:ya* and *la:ku* is unclear because Ilokano also have the same forms. Considering the large amount of lexical items shared with, thus possibly borrowed from, Yogad, it seems that there are a smaller amount of borrowings with a long vowel than is 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 borrowings from Yogad

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

(18) shows a list of identical forms shared with Ilokano. These forms are likely to be borrowings 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. Borrowings from Spanish and English are also found possibly via Ilokano or Yogad, as shown in (19, 20):

(18) Possible borrowings from Ilokano

badu:ya 'kind of cake'	*bu:ya* 'watch TV'

---

[9] The following sources were used for examining diachronic changes in phonology: Blust and Trussel (2010 –) 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 explicitly mentioned.

  inda:yun 'hammock'   *tu:luy* 'continue'
  ta:wa 'window'g.   *bi:lin* 'order'
  ba:sa 'read letter'   *uga:li* 'habit, custom'

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

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

As the above data indicate, a large amount of items used in Yogad, Ilokano, Spanish, and English were borrowed into Arta. 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 interpreted as long vowels, as in *kla:se:* 'class', /bo:l**pe:**n/ (e.g. *bolpe:n=i* 'a specific ballpoint pen'), /hap**o:**n/ (e.g. *hap**o:**n=i* 'a specific Japanese person').

   Even though we consider the influences 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 Proto-Malayo-Polynesian 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. The list of sound correspondences between PMP and Arta is shown in Table 1:

Table 1: *Reflexes of PMP*

PMP[10]	Arta	PMP	Arta
*p	/p/	*n	/n/
*t	/t/	*ŋ	/ŋ/
*k	Ø ~ /k/[11]	*h	Ø
*q	Ø	*w	/w/
*b	/b/	*y	/y/
*d	/d/	*a	/a/
*j	/d/	*i	/i/
*g	/g/	*u	/u/

---

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

[11] *k is sporadically reflected as /k/, such as *bəRək > *bərək,* *anak > *kanakannak* 'child, kid', *=ku > =*ku* 'I' (1SG.NOM).

*R	/r/	*ə	/ə/
*l	/l/	*a+i	/e:/
*m	/m/	*a+u	/o:/

In Arta, *k, *q (glottal stop) and *h are reflected as zero, *R as /r/, *j as /d/, and, as discussed later, the two vowel sequences *a+i and *a+u are reflected as two new vowels /e:/, /o:/, respectively. It is argued here that there are two types of process under which inherited forms underwent vowel lengthening, 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 of the following consonant, as in Latin *kasnus > ka:nus 'gray' (Hayes 1989: 260). The following items in Arta are also the case in point:[12]

(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 so many pieces of 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 is depicted in (22), by dropping a coda consonant, here /k/, 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).

(22)
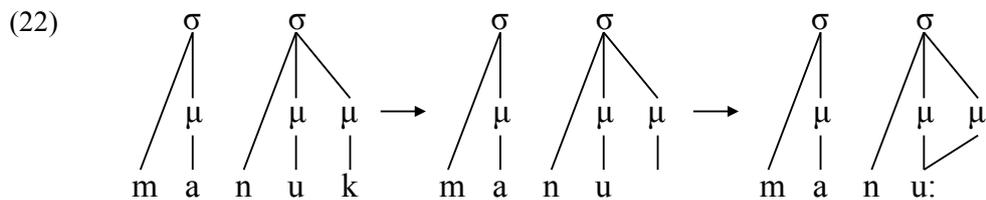
A more complex pattern of compensatory lengthening is found in personal pronouns. The following enclitic forms of personal pronouns underwent the loss of *k (see Appendix 1. for full set of personal pronouns):

(23)   PNLzn (Reid 1979) > Arta

---

[12] Other forms which might involve compensatory lengthening are: *pura:* 'white hair' cf. Ilokano *purakrak* '(white things are) brilliant' and *adu:yu* 'far, distant' (cf. Isnag *adayyu* with vowel harmony).

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

The above pronouns are synchronically peculiar as well. They differ from other pronouns in that they exhibit a complex morphophonemic alternation conditioned by the type of a segment to which they attach. (24) illustrate three conditions under which the pronouns are realized differently:

(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:na****yam/yami*** 'You/we will go'

If the preceding word ends with a consonant (except /n/) as in (A), 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 (B), /n/ is velarized, 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 (C), it is just followed by the enclitic in case of =*ku*, =*ita* and =*itam*, and, in the case of =*am* and =*ami*, /y/ is inserted between the host word and enclitic.

    How should this complex set of morphophonemic alternations be interpreted in terms of a historical development? This synchronic fact seems to be subsumed into "double-flop" (Hayes 1989: 265), which is a subtype of compensatory lengthening. A double flop may occur when there are 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 a 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):

(25)
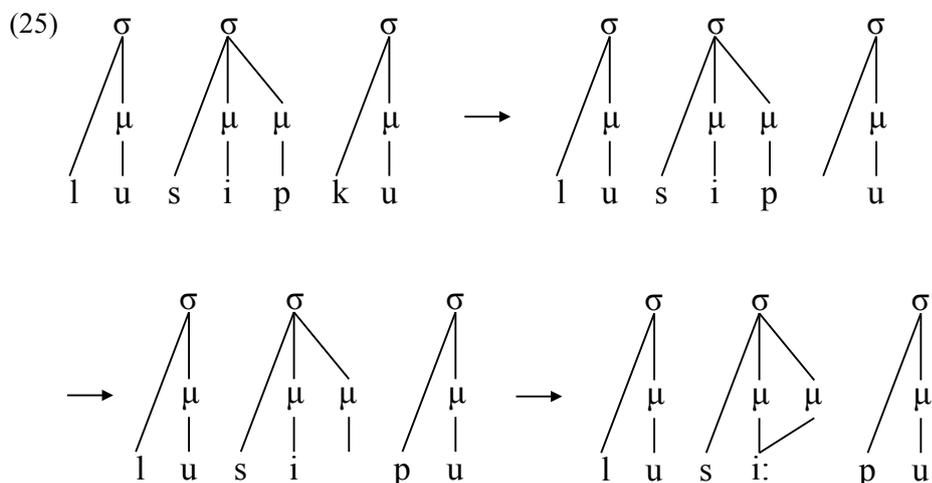
In this case, after the loss of the onset consonant /k/, the preceding segment /p/ is re-interpreted as the onset consonant probably by a universal principle of stable, unmarked syllable structure. This leads to the floating of the mora by the shift of /p/, thus being resolved by the lengthening of the preceding vowel /i/ to bear the mora. The velarization of /n/, in (B), as in *buka:gan* 'woman' + *=ami* > *buka:ga:ŋ=ami* 'we are women', must be a result of assimilation to /k/; thus it seems that this change occurred at the initial stage of (25), that is, before the loss of /k/.[13]

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:

(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.[14] Among them, /e:/ and /o:/ are interesting in that they are the phonemes that PMP did not have, and which 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/)

---

[13] The item *la:səm* 'sour' (cf. cf. Ilk. *alsəm*) might be considered to involve another kind of compensatory lengthening. If the Ilokano form *alsəm* could be an older form, and the metathesis between /a/ and /l/ occurred, the bimoraic syllable should be compensated for by lengthening the vowel, resulting in *la:səm* in Arta.

[14] *i+i > /i:/ has not been attested so far, but this seems to come from the fact that Arta shows the low percentage of PMP cognates.

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'.

This 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-* (stative verb prefixes, see Appendix 6.) followed by /a/-initial bases, trigger a vowel fusion (** represents a pre-Arta form).

(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 disppearing. Apart from PMP *q which was lost at an early stage, Arta has another glottap stop, which is inserted before 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'.[15] However, as shown above, many items have free variations in terms of the presence or 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. The number of moras after the resyllabification remains the same unless the structure may violate the moraic constraint on the syllable structure. As mentioned above, the reason why *dukut and *sakay are realized with a short vowel *dut* and *say* respectively, is that a trimoraic syllable such as du:t and sa:y is not allowed by that synchronic constraint.

(31)
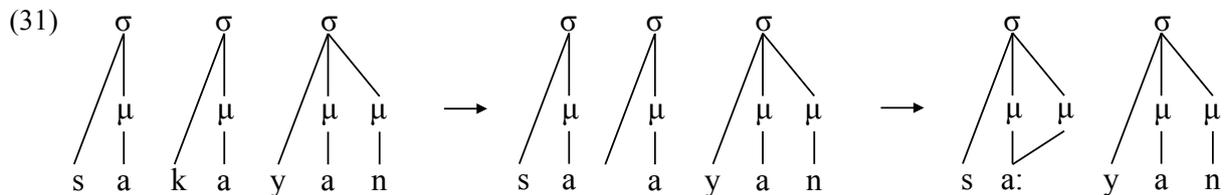
Vowel fusions are also responsible for the development of two new phonemes /e:/ and /o:/. /e:/ emerged by the loss of *h and *k in the following items:

(32)   *ahi, *aki > /e:/

---

[15] There has been a discussion over the phonological status of the glottal stop in many Philippine and Formosan languages, which I will not enter into in this paper. The author, however, does not take the stance that it is just a *phonetic* variant of /zero/, nor that all the apparent vowel-initial lexemes actually have a word-initial glottal stop. It is better to consider the glottal stop as a phonologically inserted phoneme before a vowel-initial base word. First, 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 phoneme (or consonant). Second, the occurrence of glottal stop is fully predictable, and there is no phonemic contrast between *ʔa:du* vs. *a:du*; there is no necessity to list the occurrence of glottal stop in the lexicon.

*unahik > *une:*[16] 'climb a mountain'
*bahi > *beb~be:* 'aunt' (with a reduplication)
*laki > *lel~le:* 'uncle' (with a reduplication)
*maki- > *me:-, mi:-* (comitative[17])

(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-)

a+i > /e:/ and a+u > /o:/ are again observed in synchronic morphophonemic alternations. If a base word begins with /i/ or /u/, and if *ma-/maka-* or *pa-* (causative marker) is prefixed to the base, 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 other variation with /ʔ/ except the case in (36).

(34)   *maka-/ma- (stative), 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'

(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 where *patay is reflected as /pati/, *ay is reflected as /ay/.

(37)   *ay > /ay/ (cf. *ay > /ɛ/ Casiguran Agta)
        *aNay > *aŋay* 'go' (cf. Casiguran Agta *aŋɛ*)
        *anay > *anay* 'termite' (cf. Casiguran Agta *anɛ*)
        *wasay > *wasay* 'ax, hatchet'
        *patay > *pati* 'die'

(38)   *aw > /aw/
        *ulitaw > *ulitaw* 'unmarried man'
        *tahaw > *di-taw* 'inside' (with the fossilized prefix *di-*)
        *biRŋaw > *biriŋaw* 'fly (n.)'
        *lətaw > *lətaw* 'float'

---

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

[17] See Appendix 6. for the full list of verbalizing affixes.

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 clusters 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: 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. At the first stage, /e:/ and /o:/ were seen exclusively in the items which underwent the vowel fusion. But in the process of interpreting the borrowing items containing [ɛ] and/or [ɔ] as the instances of the two existing vowels /e:/ and /o:/, the feature [+long] was also penetrated into the borrowed items.

5 Competing motivations

The above section discussed that long vowels in Arta are observed in inherited forms from PMP, and they are captured by compensatory lengthening and vowel fusions, both of which are subsumed under the principle of the mora-count conservation. This suggests that the mora-based principle played significant roles in Arta, as well as the moraic constraint on the eligible syllable structure. However, this does not imply that the mora is affecting 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 will 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 *track* 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) [trak] ~ [tărak] ~ [tarak]
      [traba:hu] ~ [tăraba:hu] ~ [taraba:hu]

The onset cluster consisting of /t/ and /r/ may or may not have the insertion of [a] with different lengths. It is difficult to assume abstract phonological representations for each lexical items, and difficult to posit any specific syllable structure or mora count. Rather, this should be considered to be the case in which the application of mora count is instable after adopting an unstable structure by 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, *tiyan, respectively. However, their reflexes again show instability in terms of the mora count and syllable structure:

(40) *tiyan > [ti(j)an] ~ [tʃan]
      *diya > [di(j)a] ~ [dʒa]

Each item may be pronounced like a bisyllabic word, or may be pronounced 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 below [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 are undergoing the change in the mora count, and it is shown in §4.3 that the mora count was conserved, rather than reduced, on the process of phonological changes. A more crucial factor applying to this particular case would be attributed to phonetic aspects. This seems to involve the fact that the two segments *iy shares the same phonetic features, thus 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 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:[18]

(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/, with the mora which would otherwise be borne by other vowels, is not associated with the schwa. Furthermore, the following items containing schwa in the final syllable, exhibit a irregular type of reflex. By the loss of final coda consonants, CVC syllables are reflected as monomoraic syllable with a short /a/.[19]

(42)   *abək > *aba* 'mat'
         *utək > *uta* 'brain'
         *pədək > *pəda* 'step on'
         *mutəq > *muta* 'gummy section of eyes'
         *luhəq > *luwa* 'tears'
         *basəq > *bisa* 'wet'

The relationship between the syllable structure and the mora count differs from the case in other vowels. Phonologically /Cə/ is treated as 0 mora, and /CəC/ as 1 mora. Note that this generalization is established from a phonological, or emic perspective. Seen from an objective, or etic perspective, the

---

[18] 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 shfting to the coda position), which could a target of compensatory lengthening by the loss of coda *q.

[19] 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].

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 1 mora fewer than the syllables with other vowels. It seems obvious however that this phonological treatment comes from the phonetic shortness of schwa. In explaining the relationship between stress and schwa in Malay, Blust (2013) argues 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 for applying the mora to the synchronic phonology in Arta, where it is argued that some allomorphimic distribution is conditioned depending on whether the preceding syllable is monomoraic or bimoraic, and that a vowel length alternation is explainable by the moraic constraint on the syllable structure. There are three cases in which long vowel occur in Arta: onomatopoeias, 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 the mora-count conservation as seen in compensatory lengthening adn 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 following forms listed in (43, 44) have a long vowel, which probably need to resort to some other phonological and/or morphological mechanisms.

(43) *hadu > *a:du* 'many, much' (cf. *ádu* Northern & Southern Alta)
      *ikan > *i:yan* 'fish'
      *zuRu > di:ru 'soup'
      *ikəj 'cough' > *i:yər* 'phlegm'[20]
(44) *u > /o:/ in the reduplicated forms
      *kutkut > /ko:tko:t/ 'dig'
      *pukpuk > /po:kpo:k/ 'beat (with a hammer)'
      *tuktuk > /to:kto:k/ 'top, summit'
      *həyup > /yo:pyo:p/ 'blow on'

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

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

---

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

 de:kət 'stickyness'   *=te:* 'only'

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

Appendix: Sets of grammatical forms in Arta

1. Personal pronouns

	TOP	ABS	GEN/ERG	OBL
1SG	*tən*	*=tən*	*=ku*	*dən*
1PL.EX	*tami*	*=ami*	*=mi*	*dami*
1PL.DU	*tita*	*=ita*	*=ta*	*dita*
1PL.INC	*titam*	*=itam*	*=tam*	*ditam*
2SG	*taw*	*=taw*	*=mu*	*daw*
2PL	*tam*	*=am*	*=muyu*	*dam*
3SG	*siya*	–	*=na*	*diya*
3PL	*tidu*	*=tid*	*=di*	*did, didu*

2. Demonstratives

		TOP	ABS	GEN/ERG	OBL	LOC
PROXIMAL	SG	*si:yəy, sate:yi:*	*a:yi:*	*ni/na a:yi:/ayni =ni*	*ti/ta a:yi:*	*(ti/ta) ayti, =ti*
	PL	*satidi:*	*(ay)tidi a:yi:*	*(ay)didi a:yi:*	*(ay)didi a:yi:*	
MEDIAL	SG	*sayna, sate:na*	*a:yina, =ina*	*ni/na ayna =nina*	*ti/ta ayna*	*(ti/ta) aytina =tina*
	PL	*satidi:na*	*(ay)tidi:na*	*(ay)didi:na*	*(ay)didi:na*	
DISTAL	SG	*saya:, saddya:, satadde:ya:*	*a:ya:, =ya:*	*ni/na a:ya:*	*ti/ta a:ya:*	*(ti/ta) ayta, =ta*
	PL	*satiddya:*	*(ay)tiddya*	*(ay)didi a:ya:*	*(ay)didi a:ya:*	

3. Determiners

			TOP/ABS	GEN/ERG	OBL
DEFINITE	SG	PERSONAL	*ti*	*ni*	*ni*
		COMMON	*i*	*ni*	*ti*
	PL		*tidi*	*didi*	*didi*
INDEFINITE			*Ø*	*na*	*ta*

4. Specifiers

SPECIFIC[21]	=*i*
SPECIFIC (PAST)	=*ti*
SPECIFIC (PLURAL)	=*tidi*
NON-SPECIFIC	=*Ø*

5. Interrogatives

WHAT	*a:nu*
WHO	*tatin*
WHEN	*tanakan*
WHERE	*adin*
	(past: *adinti*, present/future: *adi:ni*)
WHERE (PAST)	*adinti*
WHY	*ataʔay ~ adaʔay*
HOW (MANNER)	*kassandi*
HOW MANY/MUCH	*sa:ŋan*

6. Verbs

VERBS	SIMPLE NONPAST	SIMPLE PAST	PROGRESSIVE/ NOMINALIZATION
INTRANSITIVES (ACTOR-VOICE)			
PUNCTUAL (*um)	*<um>*	*<in><um>*	–
DURATIVE I[22] (*maN-)	*maN-*	*(mi)naN-*	*paC-*
DURATIVE II (*maR-)	*maC-*	*(mi)naC-*	*paC-*
COMITATIVE (*maki)	*me:-, meC-*	*ne:-, neC-*	*pe:-*
TRANSITORY STATE I[23]	*tiC-*	–	–
TRANSITORY STATE II	*maŋa:-*	–	–
OWNERSHIP	*makaN-*	–	–

---

[21] 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 Lambrechet (1994).

[22] *maC-* (durative I) indicates that the process being described is internally more homogeneous, *maN-* (durative II) more heterogeneous; *maN-*prefixation does not trigger the deletion of stem-initial consonants.

[23] A state described by *tiC-* (transitory state I) involves the intention (e.g. *tit-tuttud* 'sitting', *tip-pulot* 'wear a loincloth'), whereas a state described by *maŋa:-* (transitory state II) is a temporary state which is construed as accidentally (e.g. *maŋa:-la:ŋin* 'being sad', *maŋa:-sigid* 'being born', *maŋa:-dəgnin* 'shivering with a cold')

TRANSITIVES (UNDERGOER VOICES)			
PATIENT-TRANSITIVE	*-ən*	*<in>*	*paC- -ən*
LOCATION-TRANSITIVE	*-an*	*<in> -an*	*paC- -an*
CONVEYANCE-TRANSITIVE	*i-*	*(i)ni-*	*paC-*
ma-/maka- POTENSIVES			
INTRANSITIVE	*maka-*	*naka-*	*–*
TRANSITIVE *-ən*	*ma-*	*(mi)na-*	*–*
-an	*ma- -an*	*(mi)na- -an*	*–*
i-	*ma-, me:-*	*(mi)na-, (mi)ne:-*	*–*
CAUSATIVE FORMS			
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-*

7. Negator, Existentials and quantifiers

SENTENTIAL NEGATION	*awan*
EXISTANTIAL	*atti:* (+ABS)
EXISTENTIAL NEGATION	*awan* (+OBL)
MANY, MUCH	*meʔʔa:du*
FEW, LITTLE	*killək*

8. Second-position enclitics

ASPECTUAL ENCLITICS	
AGAIN	*=mandi*
ALREADY, SOON	*=di*
STILL	*=tep*
JUST DO, DO A FEW	*=pa*
EVALUATIONAL ENCLITICS	
ONLY	*=te:*
PRAGMATIC ENCLITICS	
REQUEST	*=nen*
HYPOTHETICAL	*=mina*

MIRATIVE	*=sika, =mat*
HEARSAY	*=an*
META-COMMUNICATIVE[24]	*=hug, =ay*

References

Blust, Robert. 2013. *The Austronesian Languages.* Revised edition. ANU, ebook.

Blust, Robert, and Stephen Trussel. 2010 –. Austronesian Comparative Dictionary, Web Edition. www.trussel2.com/ACD

Davis, Philip W., and Angel D. Mesa. 2000. *A dictionary of Yogad.* Vol. 17. Bow Historical Books.

Du Bois, John W. 1985. Competing Motivations. In Haiman, John, (ed.) *Iconicity in Syntax.* John Benjamins Publishing, 1985: 343-365.

Hayes, Bruce. 1989. Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry* 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* 6:1–54.

Komatsu, Hideo. 1989. Nihongo no Rekishi. On-in. [The History of Japanese: Phonology] Kamei, Takashi, Rokuro Kono, and Ei-ichi Chino (eds.) *The Encyclopaedia of Linguistics.* Vol.2 Tokyo: Sansei-do. pp.1634–53.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents.* Cambridge: Cambridge University Press.

Lobel, Jason William. 2010. Manide: An Undescribed Philippine Language. *Oceanic Linguistics* 49 480–512.

Prince, Ellen F. 1981. Toward a Taxonomy of Given-New Information. In Cole, Peter (ed.) *Radical pragmatics.* New York: Academic Press: 223–255.

Rasmussen M, 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.

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. 1989. Arta, Another Philippine Negrito Language. *Oceanic Linguistics* 28(1) 47–74.

---

[24] 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 the redundant information which should have been in the common ground, because the information is already mentioned (*=hug*), or it can be inferred from the linguistic/non-linguistic context (*=ay*).

Reid, Lawrence A. 1991. The Alta Languages of the Philippines. In Harlow, Ray (ed.) *VICAL 2, Western Austronesian and Contact Languages: Papers from the Fifth International Conference on Austronesian Linguistics*. 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

Robinson, Laura C. 2008. Dupaningan Agta: Grammar, Vocabulary, and Text. Ph.D. dissertation, University of Hawai'i.

Robinson, Laura C, and Jason William Lobel. 2013. The Northearstern 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. *Streams Converging into an Ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th Birthday*: 521–63.

Rubino, Carl Ralph Galvez. 2000. *Ilocano Dictionary and Grammar.* University of Hawaii Press.

Zorc, David R. 1978. On the Development of Contrastive Word Accent: Pangasinan, A Case in Point. In Liem, Nguyen Dang (ed.) *Southeast Asian Linguistic Studies*. 3 241-258. Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Zorc, David R. 1983. Proto Austronesian Accent Revisited. *Philippine Journal of Linguistics* 14(1) 1–24.

Zorc, David R. 1993. Overview of Austronesian and Philippine Accent Patterns. In Edmondson, Jerold A, and Kenneth J Gregerson (eds.) *Tonality in Austronesian Languages.* 17-24. University of Hawaii Press.