Pre-nominal *a* in San Mateo Huave[*]

Marjorie Pak
mgpak@emory.edu
Emory University
November 2010

DRAND – *Comments welcome. Please do not circulate without permission.*

1 Introduction

In the dialect of Huave spoken in San Mateo del Mar, Oaxaca, a definite noun phrase (NP) can be preceded by either the determiner *aaga* or the vowel *a*, or it can be unmarked. These three options are illustrated by the underlined NPs in the following excerpt from *Cuentos huaves III* (2006:46):[1]

1) a. Tajlüy   chük nop naxey nop najtaj. Ajiürüw chük mikombüliw   nejiw ...
        there.was EVID one man   one woman 3PL.have EVID 3PL.friend.POS 3PL.PRO
        'They say that there was once a man and a woman who had a friend...'

    b. Noik nüt tamb chük kiaj **a** <u>kombül</u>; tajaw   tengial chük arangüw nitiül.
        one day went EVID then   friend PST.see PROG   EVID 3PL.make tamales
        'One day the friend came and saw that they were making tamales.'

    c. Tajaw   tengial chük mapeax   tixor   **Ø** <u>najtaj</u> **aaga** <u>nitiül</u>.
        PST.see PROG   EVID SUB.put in.pot woman   tamales
        'The woman was putting the tamales in a pot.'

    d. **Ø** <u>Xor</u> alchük tojliüm tiül nüt
        olla   EVID   on.back on   ground
        'The pot was face-up on the ground.'

    e. Kos **Ø**   nataxey xowüy chük ind.atün.manguiay, kiaj chük tapiüng:
        because old.man very EVID ask   then EVID PST.say
        'Since the old man was very nosy, he asked...'

    f. Ngineay apmajün **a** <u>nitiül kiaj</u>?
        how   FUT.cook   tamales there
        ''How are those tamales going to cook?''

---

[1] The orthographic conventions used in Stairs and Stairs (1981) are adopted here in a slightly modified form: **x** is a voiceless alveopalatal fricative, **j** is a glottal fricative, **ü** is a high central unrounded vowel, **ch** is a voiceless alveopalatal affricate, **rr** is an alveolar trill, **ng** is a prenasalized /g/, **y** is a palatal glide, and other letters have their normal IPA values. Abbreviations used in example glosses: EVID=evidential, FUT=future, PL=plural, POS=possessive, PRO=pronoun, PRT=particle, PROG=progressive, PST=past, SG=singular, SUB=subordinate verb/clause.

According to the *Diccionario huave de San Mateo del Mar* (Stairs & Stairs: 3,310), *aaga* is a definite article ('the') and *a* is its apocopated form*; that is, *a* and *aaga* are slightly different instantiations of the same basic vocabulary item. If this is true, all else being equal, we should expect *a* and *aaga* to be used in the same kinds of syntactic and semantic contexts. And to a certain extent, they do appear to be interchangeable: in a demonstrative NP like 'those tamales' in (1f), for example, the choice between *a* and *aaga* does not appear to affect meaning or grammaticality in any way.

However, when *a* and *aaga* are systematically compared, they turn out to have a very different distribution. I report two principal findings from a corpus study of Huave narratives: (i) *a,* unlike *aaga*, can precede a pronoun; and (ii) *a*, unlike *aaga*, shows a striking tendency to appear in *postverbal subject position*. To explain these results, I propose in that in addition to the apocopated form of *aaga*, Huave has a homophonous but unrelated version of *a* that is the result of a syntax-sensitive phonological rule of epenthesis – specifically, *a* is optionally inserted at the end of an XP that is non-final within a clause. We will see that this same syntactic context is involved in conditioning a phonological rule in Welsh, and that it accounts for the basic distribution of *a* while remaining consistent with previous assumptions about Huave syntax and phonology.

2   Distribution of *a* and *aaga*: a corpus study

As noted above, the first sign that *a* cannot just be an apocopated form of *aaga* is that *a* occurs in contexts where *aaga* is impossible – for example, before a pronoun.

2) a. `tajiüntsas` **`a`** `xike`
      1SG.PST.cry   I
      'I cried.'
  b. *`tajiüntsas` **`aaga`** `xike`

3) a. `ngineay apmekwiür` `a` `neje`
      how   FUT.run   he
      'How will he run?'
  b. *`nguineay apmecuiür` `aaga` `neje`

According to native-speaking consultants, the *a* that appears before the subject pronoun in examples like (2a) and (3a) is not obligatory and does not change the meaning of the sentence in any perceivable way.

Our recorded-speech corpus contained several examples that were structurally similar to (2) and (3), but with /e/ rather than /a/ preceding the pronominal subject:

4) a. `saxeeb`   `e` `xike`
      1SG.bathe   I
      'I'm bathing.'
  b. `taxeeb` `e aaga naxey`
      PST.bathe   man
      'The man bathed.'

In all of the examples of this type that we collected, the final syllable of the preceding word contained a front vowel (here, the /e/ in *-axeeb*). I assume therefore that the underlined vowel in (4) is underlyingly *a*

(as in (2)-(3)), and that *a* harmonizes with a front vowel in the preceding syllable. If this is correct, then we can view (4b) as further evidence that *a* cannot be simply an apocopated form of *aaga* – since this example has (a form of) *a* immediately followed by *aaga*.

In all of the examples we have seen so far, *a* shows up between a verb and a postverbal subject. This turns out to be by far the most common position for *a*-marked NPs to occupy. Out of 76 examples of pronouns preceded by *a* in the *Cuentos huaves* and the *Nuevo testamento*, all of the pronouns were subjects and none of them were in clause-initial position. Furthermore, out of 306 non-pronominal and pronominal *a*-marked NPs in the *Cuentos huaves*, over half are postverbal subjects. *Aaga*, in contrast, is more likely to appear in *preverbal* subject position - although its distribution is in general far less skewed than that of *a*:

Table 1. Distribution of *a* and *aaga* in the *Cuentos huaves II-III*

	Distribution of *a*		Distribution of *aaga*	
Postverbal subject	159	**52%**	27	**25%**
Preverbal subject	16	**5%**	37	**34%**
Postverbal object	72	24%	24	22%
Other contexts	58	19%	20	19%
	306	100%	108	100%

Recall that not all definite NPs are marked by *a* or *aaga* – it is possible for a definite NP to be unmarked (see e.g. (1c)-(1e)). Table 2 shows that while the majority of *preverbal* definite subjects are unmarked (65%), the majority of *postverbal* definite subjects are marked with *a* (61%). Together with the data in Table 1, this contrast points to a strong tendency for *a* to appear in postverbal subject position.

Table 2. Morphological marking of the subjects *nataxey* 'old man' and *nench* 'boy' in the *Cuentos huaves II-III*

	nataxey/nench as a postverbal subject		*nataxey/nench* as a preverbal subject	
Preceded by *a*	57	**61%**	6	19%
Preceded by *aaga*	5	5%	5	16%
Unmarked	31	33%	20	**65%**
	93	100%	31	100%

Accordingly, our spoken corpus included several examples of minimal or near-minimal pairs, in which the speaker uttered a SV(O) sentence without *a* and then, when asked to produce a V(O)S version of the same sentence, spontaneously inserted the vowel *a*:

5) a. xike ngo narang najiüt
        I   not sᴜʙ.do work
        'I'm not doing work.'

    b. ngo narang najiüt <u>a</u> xike
        not sᴜʙ.do work   I

6) a. `pet antsorr`
      dog bark
      'The dog is barking.'

   b. `antsorrüw `<u>`a`</u>` pet`
      3PL.bark   dog
      'The dogs are barking.'

To review, the data presented in this section have shown us that:

   i) There are some instances of Huave *a* that cannot be apocopated forms of the definite determiner *aaga*, since they occur in contexts where *aaga* is forbidden (e.g. in front of pronouns).

   ii) This alternative version of *a* does not appear to be obligatory, and does not seem to change the semantics of the sentence in any way.

   iii) Huave *a* strongly favors one particular syntactic context – the postverbal subject position.

The questions I turn to next are: What is this alternative version of *a*? And why does it show up in the postverbal subject position in particular?

3   Proposal and Analysis

3.1   *A phrasal phonological rule*

Since the theory of *prosodic phonology* was developed in the 1980s (Nespor and Vogel 1986, Selkirk 1986, Zec and Inkelas 1990, among others), it has been widely recognized that there are certain kinds of phonological rules whose domains of application are at least partly determined by the syntactic structure. English stress retraction, for example, has been argued to apply between a noun and a following word only if the two words are in a relatively 'close' syntactic relationship, like the possessor and possessee in (7b). If two words are separated by a strong syntactic boundary like the subject-predicate boundary in (7c), however, stress retraction does not apply (Inkelas and Zec 1995):

7) a. Isolation form: Annemar**íe**
   b. (**Ánne**marie's hérd)
   c. (Annemar**íe)** (héard about it)

Syntax-sensitive phrasal phonological rules have been attested in a wide range of languages, and the question of just what kind of syntactic information plays a role in influencing phonological rule application is an active area of research. Let us look at another example, which will end up being more directly relevant for our purposes. In Welsh (Tallerman 2006), the first consonant of a word X is 'mutated' if the immediately preceding word belongs to a phrasal constituent (YP) that does not include X. In (8b), for example, the first consonant of *beic* is mutated (>*ḟeic*) because the preceding word *ddynes* belongs to a DP that does not include *beic*.

8) a. <u>`beic`</u>   'bicycle'
   b. `prynodd `_{DP}`[y  ddynes] `**`feic`**
      bought   the woman bicycle
      'The woman bought a bicycle.'

Welsh consonant mutation occurs not only between the subject and the object in a VSO sentence, but in other contexts where a word is immediately preceded by the end of an XP – e.g. *tranc* (>*dranc*) in (9b), which is immediately preceded by the right-boundary of a PP:

9) a. t̲r̲anc   'death'

   b. yr   oedd Prys yn   rhagweld $_{PP}$[yn 1721] d̲ranc yr   iaith   Gymraeg
       PRT went Prys PROG foretold   in 1721   death the language Welsh
       'Prys foretold the death of the Welsh language in 1721.'

However, in contexts where a word immediately follows a head that does *not* end an XP – e.g. where a word X is the complement of the immediately preceding word –consonant mutation does not apply:

10)   roedd y   ddynes yn   prynu beic   (*feic)
       was   the woman PROG buying bicycle
       'The woman was buying a bicycle.'

Welsh consonant mutation does not apply between heads and complements, but only between one word that belongs to an XP and an immediately following word outside of that XP. In other words, this rule appears to signal contexts where two words are separated by a right XP boundary: Y $]_{XP}$ Z. A similar context has been found to play a role in Modern Greek, where certain kinds of vowel elision are argued to apply between two words separated by a right XP boundary (Condoravdi 1990).

Given that the Huave pre-nominal *a* favors certain syntactic contexts (postverbal subjects rather than preverbal subjects or postverbal objects), and that it does not appear to have any discernable semantic content, it is plausible that this *a* is the product of a syntax-sensitive phrasal phonological rule like English stress retraction or Welsh consonant mutation. In fact, we will see that its distribution can be very well accounted for if we assume that it applies in the same basic syntactic environment as Welsh consonant mutation.

Specifically, I will pursue the following hypothesis:

11) **Proposal:** In addition to the *a* that is an apocopated form of *aaga*, Huave has a version of *a* that is inserted by a phonological rule of epenthesis. This rule applies at the end of a word Y that immediately precedes a word Z, only if Y belongs to a phrasal constituent XP that does not include Z.[2]

   $\emptyset \rightarrow$ a / (...__)$_Y$]$_{XP}$ (...)$_Z$

In the following subsections we will see how this proposal accounts for the distribution of *a* in various syntactic contexts.

---

[2] The epenthetic /a/ could in theory be inserted either at the end of word Y or at the beginning of word Z. I have chosen to represent /a/ at the end of word Y based on the vowel-harmony facts reviewed in example (4), the idea being that if the domain for vowel harmony is a word, the epenthetic vowel must belong to the preceding word rather than the following word. However, nothing hinges on this assumption.

3.2 VOS contexts

Let's begin with a VOS sentence like (5b), 'I didn't do work' (repeated below). I assume that Huave V(O)S sentences are derived from a basic SVO syntax by *predicate movement* –raising of VP or TP to a position above the subject. If this is correct, then the subject *xike* will be preceded by *multiple* right-XP boundaries: the boundary that closes off the object NP as well as the boundaries that close off the predicate VP, TP and NegP. The condition for the rule in (11) is therefore amply satisfied, and the vowel /a/ is inserted at the end of the object NP.

12) $_{XP}$[$_{TPi}$[ngo $_{VP}$[narang $_{NP}$[najiüt]]]$_{TPi}$ xike t_i]   --> ngo narang najiüt<u>a</u> xike
            not   SUB.do   work   I
              'I didn't do work.'

It is also possible that instead of being derived by predicate movement, VOS word order is the base structure for Huave sentences, with the subject merged as a right-specifier of VP or TP (see Rizzi 1990, Aissen 19xx for similar proposals). The context for the rule in (11) will still be met under this assumption, and we will correctly predict that /a/ will show up between the object and the subject.

3.3 VS contexts

Now let us consider the derivation of VS sentences like (2), repeated below. There are (at least) two possible structures for this sentence:

13)   The VP/TP that contains the verb *tajiüntsas* could raise to a position above the subject via
       phrasal movement – the same predicate movement that applies in (12).

              $_{XP}$[[tajiüntsas]$_{TPi}$ <u>**a**</u> xike t_i]
                  PST.cry.1SG   I
                    'I cried.'

14)   The verb *tajiüntsas* could raise to a position above the subject via **head-movement** (e.g. V could
       raise to C while the subject remains in a position below C).

              $_V$tajiüntsas$_i$ <u>**a**</u> xike [... t_i]
                  PST.cry.1SG   I
                    'I cried.'

Notice that in (14), the word that precedes the subject *xike* does *not* end an XP; i.e. the condition for the rule in (11) is not met. Recall also that the *a* in this example cannot be an apocopated form of *aaga*, since it precedes a pronoun. In order to preserve the hypothesis in (11), then, we must assume that VS sentences in Huave are derived by phrasal movement (13) rather than head movement (14).

As it turns out, there are independent reasons to believe that Huave VS sentences are derived by phrasal movement rather than head movement. Unlike Welsh, Irish, Spanish, and other languages in which the verb has been argued to raise above the subject via head movement, Huave does not appear to productively allow VSO order in verb-initial sentences with objects; rather, VOS order is preferred. VSO order occurs in texts primarily when the object is a quotation or complement clause, but is otherwise very infrequent:

(3)   ngom majaw   naxey kam [tiül mintaj akiiub]
        not sub.see man   this   if   wife   accompany
        'This man didn't realize that his wife was with him.' (Radin 1929:4)

If the verb could raise above the subject via head movement, then we would expect VSO order to be freely allowed, as it is in Welsh, Irish and Spanish. The fact that it is not suggests that verb-initial sentences involve predicate movement rather than head movement.

3.4   VO contexts

If my proposal is on the right track, the Huave 'epenthetic *a'* should *not* get inserted between a head and its complement, since this context does not meet the condition in (11). Therefore, all of the examples we find in which *a* shows up between a verb and a direct object (e.g. 13) must be instances of the apocopated *aaga*.

15)   lachük amb majaw   a nüx   kiaj
       EVID   go SUB.see   young.lady there
       'He went to see that young lady.'

And as it turns out, *a* and *aaga* occur between a verb and its direct object with almost the same frequency (22% and 24%, respectively; see Table 1). I will assume, then, that examples like (13) do not present a challenge to the hypothesis in (11), since they represent instances of the apocopated *aaga* rather than the epenthetic *a*.

3.5   SV contexts

In all of the examples we have seen so far, *a* immediately precedes a noun – we never find it in front of a verb, adjective or adverb. However, the rule in (11) makes no reference to the category of either word Y or Z, and as such it predicts that *a* should be inserted before *any* word that immediately follows the right boundary of an XP – whether the word is a noun or some other category. We might expect, then, to find the epenthetic a between a preverbal subject and a verb:

16)   [aaga najtaj]_{FN} (*a) taxeeb
       the   woman   bathed
       'The woman bathed.'

But /a/ never shows up in this context. Does this mean that the rule in (11) needs to be revised?

There may be a hint of an explanation for the absence of /a/ in SV contexts in another phonological phenomenon in Huave. As shown by Pike and Warkentin (1961) and Noyer (1991), there is a system of *phrasal tone* in San Mateo Huave. Utterances are divided into *tone phrases*, each of which receives a high (H) tone. This H tone extends to the end of each phrase, producing a kind of pitch plateau. The following utterance, for example, contains a single tone phrase, with a H tone starting on the verb and extending to the end of the sentence:

17)   (taxomás   nóts kóchíl sálín)
       1SG.PST.find one   knife Salina.Cruz
       'I found a knife in Salina Cruz.'

7

As reported in Pak 2008, the way in which utterances are divided into tone phrases is heavily influenced by the underlying syntactic structure. Moreover, there turns out to be a striking contrast between subject-initial and verb-initial sentences: postverbal subjects always group together with the verb, while preverbal subjects always form their own tone phrases:

18)   a.   (taxejpíús   á xíke)
             1SG.PST.bathe   I
             'I bathed.'

    b.   (xíke) (taxejpíüs)
             I   1SG.PST.bathe
             'I bathed.'

19)   a.   (ngo matáng   ólám)
             not SUB.grow cane
             'Cane doesn't grow.'

    b.   (ólám) (ngo matâng)
             cane not SUB.grow
             'Cane doesn't grow.'

In other words, preverbal subjects, unlike postverbal subjects, form their own phonological domains, separate from the rest of the sentence. To explain this contrast, I proposed in Pak 2008 that preverbal subjects occupy a higher position in the syntactic structure than postverbal subjects, as has been proposed for several other pro-drop languages (Alexiadou and Anagnastopolou 2001). Specifically, a postverbal subject is in Spec,TP (a), but a preverbal subject is in Spec,CP (b):

1)   a.   VOS: TP[VP[**verb object**]$_i$ TP[**subject** ~~VP[verb object]$_i$~~]]

    b.   SVO: CP[**subject** TP[VP[**verb object**]$_i$ TP[*pro*$_j$ ~~VP[verb object]$_i$~~]]

Adopting the ideas underlying Chomsky's (1999) *phase theory*, I proposed that the material at the edge of CP – i.e. Spec,CP and C – is in a separate 'spell-out domain.' Therefore, when tone phrases are formed, the preverbal subject and the verb cannot 'see' each other because they belong to completely separate spell-out domains. If correct, this analysis automatically gives us an explanation for the absence of *a* between preverbal subjects and verbs: the rule of epenthesis in (11) applies only *within* a spell-out domain. When epenthesis applies within TP, the preverbal subject is simply not visible. The context for the rule in (11) is therefore not met, and /a/ is not inserted.

4   Conclusion

The project discussed here is a corpus study of Huave narratives, focusing on the distribution of the (orthographic) word *a*. Although *a* appears at first sight to be a reduced form of the definite determiner *aaga*, I have shown that the distribution of *a* and *aaga* are actually quite distinct: (i) *a* can precede pronouns, and (ii) *a* favors the postverbal subject position. I propose that in addition to the apocopated form of *aaga*, Huave has another version of *a* that is inserted by a phonological rule of epenthesis. This rule is argued to be a syntax-sensitive phrasal phonological rule that applies in a very specific structural

context (which has also been shown to play a role in Welsh and Modern Greek): between two words Y and Z, where Y belongs to an XP that does not contain Z. This analysis explains why *a* occurs in the following contexts, where the (boldfaced) preceding word ends an XP:

[verb **object** __]$_{VP}$ subject
[**verb** __]$_{VP}$ subject

...but not in these contexts, where the first word does not end a phrase:

[**verb** __ object]$_{VP}$
[**preposition** __   object]$_{PP}$

This analysis of Huave *a* allows us to maintain Pak's (2008) proposal that Huave VS syntax involves predicate movement rather than head movement, and sheds further light on previous findings about syntax-phonology relations in Huave (Pak 2008, Noyer 1991).

References

Alexiadou, Artemis, and Elena Anagnostopolou. 1998. Parametrizing Agr: word order, verb-movement and EPP-checking. *Natural Language and Linguistic Theory* 16:491-539.

Condoravdi, Cleo. 1990. Sandhi rules of Greek and prosodic theory. In Sharon Inkelas and Draga Zec (eds.), *The phonology-syntax connection*. Chicago: University of Chicago Press.

Cuentos huaves II. 2006. 2a edición. Instituto Lingüístico de Verano, A.C. Electronic version.

Cuentos huaves III. 2006. 2a edición. Instituto Lingüístico de Verano, A.C. Electronic version.

Jayats Nanderac wüx Miteatiiüts Jesucristo (El Nuevo Testamento de nuestro Señor Jesucristo en el huave de San Mateo del Mar). 2009. La Liga Bíblica. Electronic version.

Nespor, Marina and Irene Vogel. 1985. *Prosodic phonology.* Dordrecht: Foris Publications.

Noyer, Rolf. 1991. Tone and stress in the San Mateo dialect of Huave. *Proceedings of ESCOL '91*, 277-288.

Pike, Kenneth L. and Milton Warkentin. 1961. Huave: a study in syntactic tone with low lexical functional load. In Benjamin F. Elson and Juan Comas (eds.), *A William Cameron Townsend en el vigisimoquinto aniversario del Instituto Lingüístico de Verano*, 627-42. SIL, Mexico.

Stairs, Emily F. and Elena E. de Hollenbach. 1981. Gramática huave. In Stairs and Stairs 1981.

Stairs Kreger, Glenn Albert and Emily Florence Scharfe de Stairs. 1981. *Diccionario huave de San Mateo del Mar.* SIL, Mexico.

Tallerman, Maggie. 2006. The syntax of Welsh 'direct object mutation' revisited. *Lingua 116*, 1750-1776.