PHONOLOGICAL EVIDENCE FOR THE SYNTAX OF VOS AND SVO IN HUAVE*

MARJORIE PAK
Emory University

San Mateo Huave (SMH) permits both SVO and VOS word orders, each of which is associated with a different set of phonological properties. First, H-tone never spreads from a preverbal subject to a verb, but very consistently spreads from a verb to a postverbal subject. Second, postverbal subjects are frequently preceded by a 'mystery vowel' (/a/ or /e/). I propose that both of these contrasts are reflexes of a basic syntactic distinction: Preverbal subjects are structurally higher than postverbal subjects in SMH (in Spec,CP instead of internal to TP).

Keywords: Huave, syntax-phonology interface, verb-initial syntax

1 The pattern

In San Mateo Huave (henceforth SMH, alt. Ombeayiüts, a language of the Isthmus of Tehuantepec, Oaxaca), a subject can either precede or follow the predicate:[1]

(1)   a. nehiw   tahawɨw   nakants olam   (SVO)
           3PL.PRO PST.see.3PL red   cane
           'They saw red sugarcane.'
        b. tahawɨw   nakants olam   nehiw   (VOS)
           PST.see.3PL   red   cane   3PL.PRO
           'They saw red sugarcane.'

(2)   a. xike   salmameay   (SV)
           1SG.PRO   1SG.PROG.sleep
           'I'm sleeping.'
        b. almameayiw   nehiw   (VS)
           3PL.PROG.sleep 3PL.PRO
           'They're sleeping.'

(3)   a. aaga nench kiah nechach   (S-Pred)
           DEF   boy   that   stupid
           'That boy is stupid.'
        b. tekech a   las   kam   (Pred-S)
           short   DEF rope this
           'This rope is short.'   (Stairs & Stairs 1981:335)

In addition to the obvious differences in word order shown in (1)-(3), preverbal and postverbal subjects are also distinguished phonologically in SMH. Specifically, (i) postverbal subjects are included with the verb in a single TONE PHRASE (to be defined in §2), while preverbal subjects always form a tone phrase of their own; and (ii) postverbal subjects are frequently preceded by an extra 'mystery vowel' (/a/ or /e/).

The interview excerpt in example (4) demonstrates both phenomena. When presented with the prompt 'Yesterday we painted,' the speaker CG produced two utterances: first an S-Adv-V sentence, then an Adv-V-S sentence. The preverbal subject *xikona* 'we' in the first sentence has a LHL melody, as it would in isolation, indicating that it constitutes a tone phrase by itself. In the second sentence, however, where *xikona* is postverbal, its melody is HHL, indicating that it has grouped together with the preceding verb for the purposes of a H-tone spread rule (see §2). The 'mystery vowel' /a/ also shows up before the subject in CG's second sentence (underlined).

(4)   RN: *[Cómo se dice]* '*Ayer nosotros pintamos*.'   ('How do you say 'Yesterday we painted'?')
        CG: Xìkónà   tím   sànchòmân.   Tím   tànànchòmán   á xíkónà.
             1PL/EX.PRO yesterday paint.1PL/EX   yesterday PST.paint.1PL/EX 1PL/EX.PRO

The mystery vowel also shows up in the last line of the interview excerpt in (5). This example is especially striking because the verb-initial sentence was first presented by the interviewer without the vowel and the SMH consultant repeated it with the vowel, confirming that the interviewer's version was 'okay' without commenting on his own addition of the vowel. (The tone-phrasing contrast is not clearly demonstrated in (5) because this particular subject, *xike* 'I', has HL melody both in its citation form and in tone-spread contexts.)

(5)   MP: [*Cómo se dice*] '*No hice trabajo*.'   ('How do you say 'I didn't do work'?')
        AV: Xíké   ngò   nàráng   náhìɨt. Xíkè ngò nàráng   náhìɨt.
             1SG.PRO NEG   1.SUB.do   work
        MP: *Y* '¿ngo narang nahiɨt xike' está bien tambien?   (And is '*ngo narang nahiɨt xike*' okay too?)
        AV: *Sí, está bien*. ('Yes, that's okay.') Ngò   nàráng   náhíɨt á xíkè.   Ngò nàráng náhíɨt á xíkè.
                                                     NEG   1.SUB.do work   1SG.PRO

The primary claim of this paper is that both the tone-phrasing contrast and the mystery-vowel contrast seen here are reflexes of an underlying syntactic contrast. Specifically, I argue that:

(6)   Proposal: Preverbal subjects are structurally higher than postverbal subjects in SMH
        (in Spec,CP instead of internal to TP).
        [CP Subject_pre [TP [T' …Verb …] Subject_post TP] CP]

Since postverbal subjects are syntactically closer to the verb in SMH, they are included in the same phonological domain as the verb. Preverbal subjects, on the other hand, because of their high clause-peripheral position, are invisible to the verb at the relevant point in spellout and consequently fail to interact with the verb phrase for the purposes of H-tone spread and mystery-vowel insertion.

The semantic and pragmatic constraints on preverbal and postverbal subjects have not been investigated to nearly the same extent in Huave as in other languages, e.g. Mayan (Aissen 1992, Avelino 2008, Gutiérrez-Bravo 2011, Skopeteas & Verhoeven 2009). The proposal advanced here (6) rests on the assumption that there is a strong association between phrasal phonology and underlying syntactic structure – strong enough to justify using primarily *phonological* evidence to support initial hypotheses about the syntax of an understudied language like SMH. Extensive cross-linguistic research on the syntax-phonology interface supports this idea. I will be assuming a model where phonological rules apply directly to chunks of syntactic structure (Pak 2008; see also Kaisse 1985, Seidl 2001), but my analysis could be adapted to fit a model where phonological rules apply to derived prosodic constituents like Phonological Phrases, Intonational Phrases, etc. (Selkirk 2003, Nespor & Vogel 1986, among

others). The common assumption in both types of theories is that the phonology normally does not deviate wildly from the underlying syntax, but rather is constrained by it in very regular ways.

2 H-tone spread and the separateness of preverbal subjects

The analysis of tone assignment presented here is taken from Noyer 1991, which in turn is based on data from Pike & Warkentin 1961. Their observations have been corroborated by data from multiple sources; see note 1, Appendix, and Pak 2008.

In SMH, the stressed syllable of a word pronounced in isolation will have H tone (which may be part of a HL contour tone); all other syllables are L. In phrasal contexts, however, words may be grouped together into TONE PHRASES, each of which forms a domain for H-TONE ASSIGNMENT and H-TONE SPREAD.

(7)   a. H-TONE ASSIGNMENT: H tone is assigned to the stressed syllable of the head of a tone phrase. (Trivially, a word in isolation is the head of a tone phrase.)
        b. H-TONE SPREAD: A H tone assigned to a word by (7a) spreads rightward onto any following words in the tone phrase, up to (or through) the last syllable of the tone phrase.[2]

The example in (8) constitutes a single tone phrase: a verb plus its following object and modifier. The verb serves as the head of this tone phrase, so H docks on the final (stressed) syllable of the verb and spreads rightward to the end of the utterance. In this example and subsequent examples, I use parentheses to indicate tone-phrase boundaries.

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

This example demonstrates a fairly well-established pattern seen with vP-like constituents in SMH. Typically, the verb groups together with any following objects or adverbials (including PPs) to form a single tone phrase, and the verb serves as the head of the tone phrase.

(9)   a. ...(màlòpɨ́w   tíɨl nérráár káfèy)
            3PL.SUB.soak.3PL in hot   coffee
            '[Then...] they soak them in hot coffee.'   (Cuturí 2009:25)
        b. (nèhîw) (tàhchɨ́w   nérráár yów námbéór íchweàìk)
            3PL.PRO PST.give.3PL   hot   water black   monkey
            'They gave hot water to a black monkey.'

As noted above, postverbal subjects are also included in the same tone phrase as the verb in SMH. This means that postverbal subjects are not distinguished tonally from postverbal objects in SMH – in contrast to postverbal subjects in languages like English, which typically have distinct intonational melodies associated with 'right-dislocation' or 'afterthoughts' (Grosz & Ziv 1998).

(10)   (tàhàwɨ́w   nákánts ólám néhìw)
        PST.see.PL red   cane 3PL.PRO
        'They saw red sugarcane.'

---

[2] Whether the H-plateau extends to the end of the phrase (as in (8)) or only through the penult (as in (9)) depends on the lexical tone of the final word. If the final syllable has lexical L tone, the H-tone plateau will stop short of that syllable. If the final syllable does not have lexical L tone, H-plateau will extend all the way to the end of the phrase.

On the other hand, if a verb is followed by a clausal complement or clausal modifier, its H tone does *not* spread across the clause boundary. I take this as an initial sign that utterances are spelled out in chunks –with chunks corresponding roughly to CPs – so that phonological rules like H-tone spread can only 'see' one chunk at a time.

(11) a. (tàhàwɨ̂w)   (nàkánts ómeááts á ólám)
          PST.say.3PL SUB.red body   cane
          'They said the sugarcane was red.'
      b. (sàndɨ̂m) (nàndeák   ómbeáyìɨts)   (pòrkè   màhneâh)
          1SG.want   1SUB.speak language.1P/INC   because SUB.beautiful
          'I want to speak Huave because it's beautiful.'

Finally, as noted in §1, preverbal subjects pattern very differently from postverbal subjects: essentially, they behave like NPs pronounced in isolation. While an H-tone on a verb spreads onto a postverbal subject, a H-tone on a preverbal subject never spreads onto the following verb. This asymmetry is highlighted in the example pairs below.
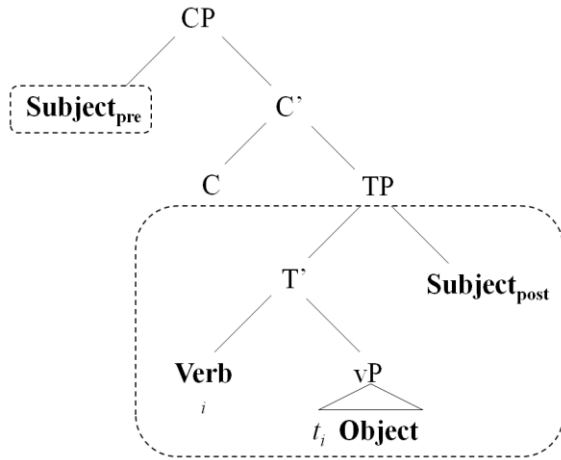
(12) a. (S)(VO):   (nèhîw) (tàhàwɨ́w   nákánts ólám)
                        3PL.PRO PST.see.3PL red   cane
      b. (VOS):   (tàhàwɨ́w   nákánts ólám   néhìw)
                        PST.see.3PL red   cane 3PL.PRO
                        'They saw red sugarcane.'

(13) a. (Adv)(S)(V): (nìngɨ́y) (òlám) (ngò màtâng)
                         here   cane   NEG SUB.grow
      b. (Adv)(VS):   (nìngɨ́y) (ngò màtáng   ólám)
                         here   NEG SUB.grow cane
                         'Sugarcane doesn't grow here.'

(14) a. (S)(AdvV):   (nêh)   (làndóh áxèèb)
                         3SG.PRO already bathe
      b. (AdvVS):   (làndóh áxééb é nèh)
                         already bathe   3SG.PRO
                         'S/he already bathed.'

(15) a. (S)(V):   (xíkè)   (tàxèhpɨ́ɨs)
                         1SG.PRO PST.bathe.1SG
      b. (VS):   (tàxèhpɨ́ɨs   á xíkè)
                         PST.bathe.1SG   1SG.PRO
                         'I bathed.'

In Pak 2008, I analyze these patterns as follows. I assume a model where syntactic structures are spelled out in chunks, or phases, instead of all at once (Chomsky 2001 et seq.), and where different phonological rules apply directly to spelled-out chunks of structure at various points in PF. Within this model, I propose that preverbal and postverbal subjects in SMH have the basic syntax sketched in (16)[3] – i.e., preverbal subjects are in Spec,CP while postverbal subjects are internal to TP. Since postverbal

---

[3] The structure in (16) has the postverbal subject as a right-specifier of TP. An alternative possibility worth considering is that the the postverbal subject is a *left*-specifier, and that the verb+object constituent raises (by predicate-movement) to some position above the subject but below C. While such an analysis is possible, its execution is complicated. Tense and Mood are affixed to verbs in Huave, suggesting that V raises to T. The moved predicate, then, would need to be at least as big as T', but the position this constituent raises to would have to be below C. We would therefore need to posit at least one undefined layer of functional structure between CP and TP that was otherwise unmotivated. See Pak 2007 for more discussion.

subjects are below the phase head C, they are spelled out as part of the same phonological domain as the verb, while preverbal subjects are spelled out in a separate phonological domain.[4]

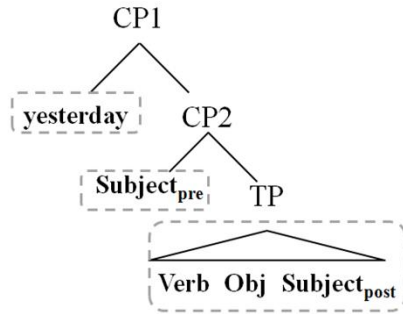(16)   Syntax of preverbal and postverbal subjects in SMH (phonological domains in dotted lines)

    (nèhîw)   (tàhàwɨw   nákánts ólám)
                    (tàhàwɨw   nákánts ólám néhìw)
    3PL.PRO   PST.see.3PL red   cane 3PL.PRO
    'They saw red sugarcane.'

One more generalization needs to be explained before moving on: In sentences with multiple topic-like constituents in the left periphery – e.g. a preverbal subject as well as a fronted locative or temporal adverb, as in (13)a and (17) – each preverbal constituent forms its own tone domain. I take this as evidence for recursive CPs, with each C triggering spellout of its complement.

(17)   (òxêp)   (áágá náxéy kiàh) (àpmɨɨch ákókiáw chípín míkwál   xékómbɨl)
     tomorrow DEF man there FUT.give five   tomato 3/POSS.son 1/POSS.friend
       'Tomorrow that man will give five tomatoes to my friend's son.'

---

[4] One question that may arise at this point is why C is the only head that seems to trigger spellout in SMH, given Chomsky's (2001) claim that C as well as *v* are phase heads. This question is not unique to the current analysis, of course; there are a number of competing proposals about what counts as a phase head (e.g. any *v* vs. only strong *v* vs. any category-defining head (Embick 2010, Marvin 2001), etc.). Although I cannot do justice to this question here, I suspect that a solution may lie in a more articulated definition of what spellout entails. For example, the closing off of a domain for the purposes of syntactic movement (phase-impenetrability effects) may involve only *partial* spellout – placing a sub-structure in a 'holding bin' before it actually undergoes the full range of PF operations. The larger CP-sized domains that delimit SMH tone-spread (as well as Luganda tone-spread; see Pak 2008:ch4) might then correspond to merged and chained substructures from the holding bin, which are subjected to all or nearly all PF operations (vocabulary insertion, phrasal phonological rules, etc.).

(18)   Syntax of sentences with multiple preverbal constituents in SMH

```
                    CP1
                   ╱╲
         ┌─────────┐  CP2
         ┆yesterday┆  ╱╲
         └─────────┘ ╱  ╲
              ┌───────────┐ TP
              ┆Subject_pre ┆ ╱╲
              └───────────┘╱  ╲
                   ┌────────────────────────┐
                   ┆ Verb  Obj  Subject_post ┆
                   └────────────────────────┘
```

       With this analysis of SMH tone domains in mind, we can turn to the distribution of the 'mystery vowel' described in §1 (see (4)-(5), (11)a, (14)b, (15)b). We will see that the domains proposed for H-tone spread in (16) also play an important role in restricting the distribution of this mystery vowel.

3 Vowel epenthesis: phonological domains within domains

3.1 Distinguishing the mystery vowel from the definite article

Before we examine the distribution of the mystery vowel, it is important to distinguish this vowel from the homophonous definite article *a*, an apocopated form of *aaga* ('the') (Stairs & Stairs 1981:3,310).

       In many contexts the mystery vowel and the apocopated definite article are indistinguishable. For example, both (underlined) instances of *a* in (19) are ambiguous, because they each precede a discourse-old, non-prominal NP/DP:

(19)   sapi̵ng   xow̵y   lahneah   <u>a</u>   cielo, aw   <u>a</u>   poh
       1SG.think   very   beautiful   DEF?   sky   say   DEF?   turtle
       '"I think the sky is very beautiful," says the turtle.'   *(Cuentos huaves)*

However, as I show in Pak 2010, there are a number of contexts where *a* can be used while *aaga* cannot, which would not be expected if *a* were always a reduced form of *aaga*. First, *a* can precede a pronoun, but *aaga* cannot:

(20)   a.   tàxèhpí̵s   á   xíkè
             PST.bathe.1SG   1SG.PRO
             'I bathed.'
       b.   *taxehpi̵s aaga xike

Second, *a* can co-occur with *aaga*. This is most easily observed when the vowel undergoes harmony with a front vowel in the preceding syllable and surfaces as /e/. Obviously, the extra vowel in such examples cannot itself be a reduced form of *aaga*.[5]

(21)   tàxééb   <u>é</u>   áágá   náxéy
       PST.bathe   DEF   man
       'The man bathed.'

---

    [5] The fact that the vowel is interacting phonologically (harmonizing) with the preceding verb, rather than with the following noun, might itself be seen as evidence against the idea that the vowel is a determiner. See §3.3 for more discussion of this apparent morphological mismatch.

Third, *a* and *aaga* show very different distributions in the *Cuentos huaves*, a collection of transcribed oral narratives (see Appendix). Table 1 shows that 52% of all tokens of prenominal *a* in this corpus precede postverbal subjects, compared to only 25% of tokens of *aaga* ($p < .0001$).

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

	a		*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%

Table 2 shows the distribution of *nataxey* 'old man' and *nench* 'boy' after their first mention in a given story (i.e., when are discourse-old). When these nouns are used as preverbal subjects, they are marked with *a* 19% of the time and with *aaga* 16% of the time. When they are used as postverbal subjects, however, they are marked with *a* 61% of the time and with *aaga* only 5% of the time ($p < .0001$).

Table 2. *Nataxey* 'old man' and *nench* 'boy' in the *Cuentos huaves II-III*

	nataxey/nench as pre-V subject		*nataxey/nench* as post-V subject	
Preceded by *a*	6	19%	57	61%
Preceded by *aaga*	5	16%	5	5%
Unmarked	20	65%	31	34%
	31	100%	93	100%

The skewed distributions of *a* and *aaga* in Tables 1 and 2 would not be expected if all instances of *a* were apocopated *aaga*, and lend further support to the idea that there is an independent source for the mystery vowel.

3.2 The distribution of the (unambiguous) mystery vowel: initial generalization

When we focus exclusively on unambiguous instances of the mystery vowel – e.g. vowels that precede pronouns or *aaga* – several generalizations become apparent. First, unambiguous mystery vowels only show up between words, never in utterance-initial or utterance-final position. Moreover, mystery vowels do not show up between just any two words, but only between words that are 'visible' to each other by virtue of being in the same phonological domain.

(22)   Generalization (*to be revised*): Unambiguous mystery vowels appear only in the context (…W1__W2…), where the parentheses delimit a single phonological domain.

Crucially, the phonological domains referred to in (22) can be shown to be the same as those that were invoked in my analysis of H-tone spread (§2).
        Recall that H-tones in SMH never spread from a preverbal subject to a verb; this was the main motivation for the idea that preverbal subjects are at the edge of a phase (in Spec,CP) and thus belong to a separate phonological domain. Correspondingly, mystery vowels never show up between preverbal subjects and verbs.

(23)   a.   (áágá náxéy) (*e/a) (tàxéèb)
                 DET man   PST.bathe
                 'The man bathed.'

7

b. (nòp nèlòp   chíy) (*e/a) (teàmɨ́ɨt   sòx)
     one hungry mouse   PROG.eat grass
     'A hungry mouse is eating grass.'

It could be argued that there is an independent reason why the mystery vowel is ruled out in (23): perhaps
the mystery vowel is a morpheme associated with NPs/DPs (e.g. some kind of case-marker), and thus
would never be expected to show up before verbs in the first place. I will return to this possibility in
§3.3.2. In the meantime, we can show that even in pre-NP/DP contexts, the distribution of the mystery
vowel still constrained by the phonological domains that delimit H-tone spread in SMH.

   We saw that in sentences with multiple topic-like constituents at the left periphery, H-tone does
not spread from one preverbal constituent to another – e.g., from a fronted adverb onto a preverbal
subject. For this reason, I proposed in §2 that such sentences have recursive CP structures, with each C
triggering spellout of its complement, effectively creating a separate phonological domain for each topic-
like constituent (see (18)). As expected under this analysis, unambiguous mystery vowels do not show up
between fronted adverbs and subjects in the left periphery:

(24)   a. (òxêp)   (*e/a) (áágá náxéy kiàh) (àpmɨ́ɨch ákókiáw chípín míkwál   xékómbɨ́l)
            tomorrow   DEF man there FUT.give five   tomato 3/POSS.son 1/POSS.friend
            'Tomorrow that man will give five tomatoes to my friend's son.'
      b. (tím)   (*e/a) (xíkè)   (tàhàwás   námbeór kóy   tílám)
            yesterday   1SG.PRO PST.see.1SG black   rabbit in.river
            'Yesterday I saw a black rabbit in the river.'

   On the other hand, unambiguous mystery vowels *do* show up before *postverbal* subjects –
whether the preceding word is a verb (as in (25)a) or an object (as in (25)b). This is fully expected under
the central proposal advanced in this paper: that postverbal subjects are structurally lower than preverbal
subjects and are therefore part of the same phonological domain as the verb (6).

(25)   a. (làndóh áxééb é̱ nèh)
             already bathe   3SG.PRO
             'S/he already bathed.'
      b. Ngò nàráng   náhɨ́ɨt á̱ xíkè.
             NEG 1.SUB.do work   1SG.PRO
             'I didn't do work.'

   Evidence from our text corpora (see Appendix) suggests one more context for the mystery vowel
that is correctly ruled in by the '(...W1__W2...)' context defined in (22). The adverbial *at* 'also' is
sometimes written as *ata* or *at a* when it precedes (and modifies) a preverbal subject. Unlike the fronted
time/place adverbs in (24), *at* can plausibly be analyzed as attaching *internal* to the subject NP/DP (as
shown with the bracketing in (26)). If this is correct, *at* is predicted to belong to the same phonological
domain as the subject (pro)noun. The two words *at* and *xike* in (26)a (and *at* and *ike* in (26)b) thus provide
a context for the mystery vowel to be inserted, resulting in the variant pronunciation *ata*.

(26)   a. [DP[AP At a̱] xike]   ngome sanasah ik   nikwahind.
                  also   1SG.PRO NEG   1SG.say 2SG.PRO nothing
             'I, too, won't say anything to you.'   (*Cuentos huaves*)
      b. [DP[AP At a̱] ike]   xekwal apmambar.
                  also   2SG.PRO 1.son FUT.go.DUAL
             'You too, my son, we'll go together.'   (*Cuentos huaves*)

The contrast between (24), (25) and (26) is noteworthy. It is true that the mystery vowel generally shows up before subjects, but it does not show up before just any subject. Rather, it can only be inserted if the subject is preceded by another word within the same phonological domain, as stated in (22).

The generalization in (22) cannot be left as it stands, however. The problem is that it is too general: it suggests that mystery vowels can show up between *any* two words that belong to the same phonological domain, e.g. between heads and complements. But unambiguous mystery vowels are unattested in these positions.

(27)　a.　andi̵m (*a) xik
　　　　　love　　　1SG.PRO
　　　　　'S/he loves me.'
　　　b.　wi̵x (*a) neh
　　　　　about　　3SG.PRO
　　　　　'about him/her'

How can the generalization in (22) be replaced with an analysis that correctly rules out examples like (27)? I will consider two possibilities – (i) that the mystery vowel is the product of (structurally restricted) phonological epenthesis, and (ii) that the mystery vowel is an allomorph of the [NOM] morpheme, whose insertion is also structurally conditioned. I lay out some of the consequences of each proposal as well as predictions to be tested in future work.

3.3 The distribution of the mystery vowel: further restrictions

3.3.1 A phonological account: The mystery vowel is epenthetic

The first proposal I consider is that the mystery vowel is the product of an optional phonological rule of epenthesis. This phonological rule must crucially be structurally restricted, as specified below, to rule out its application in head-complement configurations like (27).

(28)　Epenthesis (optional): $\emptyset \rightarrow$ a / (...W1__ $]_{XP}$ W2…), where parentheses delimit a single phonological domain.

This rule states that a vowel /a/ is optionally inserted between two words, W1 and W2, iff (i) W1 and W2 are both contained inside the same phonological domain, and (ii) W1 belongs to a phrasal constituent (XP) that does not include W2. It is part (ii) of this description that rules out epenthesis in configurations like (27). Unlike H-tone spread (§2), which applies blindly throughout a large, CP-sized phonological domain, vowel epenthesis applies only at the juncture between the end of an XP and a following word *within* a large phonological domain – producing a kind of 'domain within domain' effect. This contrast is shown schematically below: the double-slash represents the boundary between large phonological domains, within each of which H-tone spread applies, and the single-slash represents the juncture between smaller domains, where epenthetic vowels can be inserted.

(29)　a.　*big domain:*　(Subj$_{pre}$)　// (Verb　Obj　Subj$_{post}$)　　*H-tone spread*
　　　b.　*small domain:*　(Subj$_{pre}$)　// (Verb Obj) / (Subj$_{post}$)　　*Epenthesis*

Notice that the '__ $]_{XP}$' juncture condition in (28) also correctly rules in vowel-epenthesis between a DP-internal adverb like *at* 'also' and the following noun in examples like (26).

But where does this XP-juncture condition come from? Does it have any independent motivation, or does it simply have to be stipulated to account for the SMH pattern? In Pak 2008, I hypothesize that the linearization of syntactic structures is completed in stages, and that different phonological rules apply at different stages in PF and thus have access to different amounts of (linearized) material. SMH H-tone

spread is a relatively late rule, applying throughout a spell-out domain after its contents have been fully linearized into a 'chain.' Epenthesis, however, might apply at an earlier stage in PF, when the sub-parts of a chain are being concatenated. The end of an XP constituent is a point where concatenation must 'pop back' to a higher level of structure before continuing; this is when Epenthesis applies, creating the juncture effect.

If this idea is on the right track, we should expect to find similar juncture effects in other languages. And in fact such cases are attested (although they seem to be somewhat uncommon for reasons that I do not yet understand). In Welsh (Tallerman 2006), for example, the first consonant of a word W2 is 'mutated' if the preceding word W1 belongs to a phrasal constituent (XP) that does not include W2. Compare the underlined initial consonant in the citation forms below ((30)a and (31)a) to the mutated initial consonants in phrasals context ((30)b and (31)b).

(30) a. <u>b</u>eic   'bicycle'
      b. *prynodd* DP[*y ddynes*] <u>f</u>eic
         bought   the woman   bicycle
         'The woman bought a bicycle.'

(31) a. <u>t</u>ranc '*death*'
      b. *yr oedd Prys yn   rhagweld* PP[*yn 1721*] <u>d</u>ranc yr iaith   Gymraeg
         PRT went Prys PROG foretold   in 1721 death the language Welsh
         'Prys foretold the death of the Welsh language in 1721.'

Crucially, initial consonants in Welsh are not mutated in just any phrasal context, but only if the 'juncture condition' is met – i.e., if the word is preceded by a right-XP boundary. Since Welsh is a VSO language, this condition will normally be met by direct objects (since the preceding subject belongs to its own XP). However, in sentences with Aux-SVO order, direct objects are preceded by the verb and thus do not undergo mutation.

(32) *roedd y   ddynes yn   prynu* <u>b</u>eic   (*feic*)
      was   the woman PROG buying bicycle
      'The woman was buying a bicycle.'

In other words, like SMH vowel epenthesis in (27), Welsh mutation fails to apply between a head and a complement.

Let us conclude this subsection by reviewing some predictions of a phonological account of the SMH mystery vowel. First, if the mystery vowel is phonologically inserted, it should not have any associated meaning. This prediction seems to be borne out: the speaker in (4) did not seem to be aware that he was adding the vowel in a sentence-repetition context, and when we asked several consultants to compare two sentences, one with and one without the vowel, each said the sentences 'meant the same thing.'

A second prediction of the phonological account, which may turn out to be problematic, is that unambiguous epenthetic vowels should show up not just before subjects but in other syntactic contexts that meet the conditions in (28) – e.g. between two NP objects, or between an object NP and a prepositional phrase:

(33) *tàxòmás*   DP[*nóts kóchíl*] (?<u>e</u>) *sálín*
      PST. find.1SG   one knife   Salina.Cruz
      'I found a knife in Salina Cruz.'

So far the mystery vowel has not been attested in contexts like these. Since our spoken corpus has only a few tokens with these structural properties, however, this is an important area for future work.

3.3.2 A morphological account: The mystery vowel is a case-marker

We have seen that the mystery vowel cannot be a definite article (§3.1). It could, however, be a morpheme realizing some other syntacticosemantic feature associated with subjects – e.g. nominative case. Analyzing the mystery vowel as a nominative-case marker would explain why the mystery vowel usually precedes subjects, why it never appears unambiguously between heads and complements (27), why it can co-occur with pronouns and the determiner *aaga* (§3.1), and why it is semantically light.

Case is (otherwise) manifested in SMH only on first- and second-person pronouns, which have vowel-final forms when they are used as subjects and consonant-final forms when they are used as objects. I assume, following Stairs & Stairs 1981:296, that these forms correspond to nominative and objective case respectively.

(34)

	NOM	OBJ
1SG	*xike*	*xik*
2SG	*ike*	*ik*
1/2DU	*ikora*	*ikor*
1PL/IN	*xikona*	*xikon*
2PL	*ikona*	*ikon*
1PL/EX	*ikootsa*	*ikoots*

It is important to note that the nominative forms are used for both preverbal and postverbal subjects, while the objective forms are reserved for objects. In other words, even though preverbal subjects are high in the left periphery in SMH, they still get assigned nominative case (unlike left-dislocated subjects in languages like French, which get default accusative case).

(35) a. <u>xike</u>   taxehpiis   (preverbal subject)
           1SG.PRO   PST.bathe.1SG
           'I bathed.'
       b. taxehpiius   a <u>xike</u>   (postverbal subject)
           PST.bathe.1SG   1SG.PRO
           'I bathed.'
       c. ambeol <u>xik</u>   (object of verb)
           help   1SG.PRO
           'S/he helps me.'
       d. wɨx <u>xik</u>   (object of preposition)
           on   1SG.PRO
           'on/about me'

(36) *Moi je vais   mettre ça   comme Pol.*   (De Cat 2004:53)
       me I go.1SG put.INF that like   Pol
       'I'll put it like Pol.'

If the mystery vowel /a/ is a case morpheme, then the apparent 'double-marking' of nominative case in examples like (35)b would need to be attributed to some kind of case agreement. Furthermore, we would need to assume that case-marking with /a/ is optional, in order to explain its absence in e.g. (37).

(37) (tàhàwɨw   nákánts ólám néhìw)
       PST.see.3PL   red   cane 3PL.PRO
        'They saw red sugarcane.'

Finally – and more problematically – we need to explain why this nominative-case marker never appears on *sentence-initial* subjects:

(38) a. (*a) xike   taxehpiɨs
           1SG.PRO PST.bathe.1SG
           'I bathed.'
      b. *cf.* taxehpiɨs a̱ xike

The relevant observation is that the mystery vowel needs to have something on its left within the same H-tone spread domain (as stated in our initial generalization in (22)). In order to encode this requirement within an analysis of /a/ as a case marker, we could posit an allomorphy rule like (39):

(39) [NOM] ↔ a / X ⌢___
             ↔ Ø elsewhere

The first part of this allomorphy rule states that the nominative-case marker [NOM] is spelled out as /a/ iff it is concatenated (linearized) with some word X on its immediate left. When this condition is not met, a null morpheme is inserted instead. Presumably, this rule would be constrained to apply within each phonological domain, as defined in §2 for H-tone spread, so that the Ø allomorph would be inserted even when the subject was preceded by a fronted time/place adverb (see (18)).

While the phonological domains for H-tone spread proposed in §2 certainly play a role in the distribution of the mystery vowel, they introduce a problem when they are incorporated into an allomorphic treatment like the one just sketched. Simply put, these domains are too big for an allomorphic alternation. Unlike phonological rules, whose domains range from the very small (word-internal rules) to the very large (across-the-board rules), allomorphy is usually assumed to be a strictly word-internal phenomenon, and even within words it appears to be subjected to strict locality constraints (Embick 2010). The allomorphy rule in (39) is suspiciously *non*-local: it depends on [NOM] being able to 'see' across a word-boundary – and potentially across one or more XP boundaries as well, as in (40) – in order to establish whether there is another word on its left.

(40) Ngò [vₚ nàráng [DP náhíɨ́t]] [DP á̱ xíkè̱].
      NEG   1.SUB.do   work   1SG.PRO
      'I didn't do work.'

While cases of 'external' or 'phrasal allomorphy' have been proposed in other languages, they are usually still structurally quite constrained, involving a head needing to see its complement or a part of its complement, e.g. the English *a/an* alternation or the alternation between *en* and *l'* in the Catalan personal definite article (Mascaró 2007; see also Hayes 1990). The proposed allomorphy rule in (39) would need to be much more powerful than this, and would not have any cross-linguistic precedent to my knowledge.

Recall moreover that the mystery vowel can harmonize with a front vowel in the *preceding* word:

(41) tàxééb   é̱ áágá náxéy
      PST.bathe   DEF man
      'The man bathed.'

Suppose we took vowel-harmony in these examples as evidence that [NOM] cliticizes to the preceding word. Then we could say that the allomorphy rule in (39) applies word-internally, after cliticization, thus obviating the need for non-local allomorphy. This kind of solution, unfortunately, would only be shifting the problem from one part of the grammar to another: now, instead of non-local allomorphy, we would have non-local cliticization. As discussed at length in Embick & Noyer 2001, Embick 2007, etc., morphological readjustments are also subjected to strict locality conditions; among other things, they are

12

confined to phases that are much smaller than the CP-sized phases that phrasal phonological rules like H-tone spread are constrained by. Observe, for example, the availability of full contraction (with vowel reduction) in English in (42)a but not (42)b, where *you* and *will* are in a non-local relationship.

(42)   a.   <u>You'll</u> (/juḷ/, /jəl/) never walk alone.
          b.   [The guy [next to <u>you</u>]]<u>'ll</u> (/juḷ/, */jəl/) never walk alone.

       Since vowel harmony is generally a word-internal phenomenon in SMH, the fact that the mystery vowel harmonizes with the previous word may itself present a problem for the [NOM] morpheme analysis sketched above. This case is reminiscent of cases of purported 'ditropic' clitics in e.g. Kwakwala (see Embick & Noyer 1999). Under the phonological account in §3.3.1, however, the vowel-harmony facts are unproblematic: the epenthesis rule simply states that /a/ is inserted at the end of the first word under the appropriate conditions.

       I leave the choice between the phonological and morphological accounts laid out here for future work. The basic tension is that the mystery vowel seems to be associated with morphosyntactic properties of the object to its right (a subject), but at the same time demands the presence of and interacts phonologically with a word on its left. Its basic status as a product of the phonology or the morphology is therefore unclear. Interestingly, similar questions arise in the literature on Welsh consonant mutation (see Tallerman 2006), and may suggest that the mystery vowel is undergoing reanalysis.

4   Conclusion

I have used phonological evidence to advance a specific claim about the syntax of SMH – that preverbal subjects are higher than postverbal subjects (Spec,CP instead of Spec,TP). I showed that the phonological domains for SMH tone-spread also play a role in the distribution of a 'mystery vowel' /a/ that shows up before postverbal subjects. While tone-spread applies blindly throughout these phonological domains, the mystery vowel is subjected to additional structural constraints. I examined the distribution of this vowel in depth, paying particular attention to whether it is morphological or phonological in nature.

Appendix A   Corpus texts

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.

References

Aissen, Judith L. 1992. Topic and Focus in Mayan. *Language 68*, 43-80.
Avelino, Heriberto. 2009. Intonational patterns of Topic and Focus constructions in Yucatec Maya. *In MIT Working Papers in Linguistics 59: New Perspectives on Mayan Linguistics*, pp. 1-21. MIT.
Chomsky, Noam. 2001. Derivation by Phase. In *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, 1–52. Cambridge, Mass.: MIT Press.
Cuturí, Flavia G. 2009. *Nüeteran ikoots naw San Mateo del Mar* (*Comida ikoots de San Mateo del Mar*). México: Instituto Nacional de Lenguas Indígenas.
De Cat, Cécile. 2004. Apparent non-nominative subjects in L1 French. In Philippe Prévost & Johanne Paradis (eds.), The acquisition of French in different contexts: Focus on functional categories (pp. 51-88). Amsterdam: John Benjamins.

Embick, David. 2007. Linearization and local dislocation: derivational mechanics and interactions. *Linguistic Analysis 33*, 303-336.

Embick, David & Noyer, Rolf. 1999. Locality in post-syntactic operations. *MIT Working Papers in Linguistics 34*, 265-317.

Embick, David & Noyer, Rolf. 2001. Movement operations after syntax. *Linguistic Inquiry 32*, 555-595.

Gutiérrez-Bravo, R. 2011. External and Internal Topics in Yucatec Maya. En *Representing Language: Linguistic Essays in Honour of Judith Aissen*, Eds. R. Gutiérrez-Bravo, L. Mikkelsen & E. Potsdam. Lexington, KY, Lingustics Research Center, pp. 105-119

Hayes, Bruce. 1990. Precompiled phrasal phonology. In Sharon Inkelas & Draga Zec (eds.), *The phonology-syntax connection* (pp. 85-108). Chicago: University of Chicago Press.

Kaisse, Ellen. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Academic Press.

Mascaró, Joan. 2007. External allomorphy and lexical representation. *Linguistic Inquiry, 38*(4), 715-735.

Marvin, Tatjana. 2002. *Topics in the stress and syntax of words*. Doctoral dissertation, MIT.

Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

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

Pak, Marjorie. 2007. Phrasal tone and syntax in San Mateo Huave. Unpublished manuscript.

Pak, Marjorie. 2008. *The Postsyntactic Derivation and its Phonological Reflexes*. Unpublished Ph.D. dissertation, University of Pennsylvania.

Pak, Marjorie. 2010. Pre-nominal *a* in Huave. (Spanish version also available.) In *Memorias de la I Jornada de Estudios Huaves*.

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 Linguístico de Verano*, 627-42. SIL, Mexico.

Seidl, Amanda. 2001. *Minimal Indirect Reference: A Theory of the Syntax-Phonology Interface*. Routledge.

Selkirk, Elisabeth. 2003. The Prosodic Structure of Function Words. in J. McCarthy, ed. *Optimality Theory in Phonology: A Reader,* Blackwell Publishing.

Skopeteas, Stavros & Verhoeven, Elisabeth. 2009. Distinctness Effects on VOS Order: Evidence from Yucatec Maya. In Heriberto Avelino et al. (eds), *New Perspectives on Mayan Linguistics* (157-174). Cambridge, MA: MITWPL.

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.