1. Introduction

The morpheme $UM$ (usually marking actor voice [AV]) in the Austronesian language family has been widely discussed in previous literature (Prince & Smolensky 1993/2004, Orgun & Sprouse 1999, Crowhurst 2001, Klein 2005, Lu 2005). $UM$ most often appears as an infix. However, in stems that contain labial consonants, this infix patterns differently. For example, in Mayrinax (Lu 2005), $UM$ usually shows up as an infix after the first consonant: /tahuq/ $\rightarrow$ /t-um-ahuq/ ‘cook’. When the stem-initial consonant is a labial, however, the labial appears to be lost and replaced by an /m/: /ba.hij/ $\rightarrow$ /m-a.hij/ ‘hit’. In Tagalog (Orgun & Sprouse 1999), the infix $UM$ shows up either after the first consonant or after the consonant cluster: /gradwet/ $\rightarrow$ /g-um-radwet/ $\sim$ /gr-um-adwet/ ‘graduate’. In stems beginning with a sonorant labial, however, the morphological category is simply not realized: /wejI/ $\rightarrow$ */w-um-ejI/ ‘wail’. In this paper, I will investigate the behavior of $UM$ in Thao, an Austronesian language spoken in central Taiwan.

Five realizations of $UM$ in Thao are documented by Blust (2003): infix -um- in stems beginning with two consonants (/t-um-qir/ ‘protest’); infix -m- in stems with a single consonant (/t-m-iu/ ‘to comb’); prefix m- in /ð, r, l/ initial stems (/m-ðai/ ‘tell’) and in vowel-initial stems (/m-ufa/ ‘go’); apparent loss of stem consonant in /p/ initial stems (/pata/ $\rightarrow$ /mata/ ‘write’); and no realization of the morphological category in /b, d, f, m/ initial stems. I will argue that several phonological processes are involved in determining the different realizations of this morpheme, including the preference for words containing a single foot and the requirements of sonority sequencing in syllable onsets. As for the cases in which the morphological category formed with affixing $UM$ is simply not realized, I will argue that this is a paradigmatic gap caused by a phonological restriction, the avoidance of two adjacent labials.

1 Heartfelt thanks to Ellen Broselow, Christina Bethin and Jose Elias Ulloa for all their comments and ideas.

2 Following Klein 2005, I use $UM$, instead of -um-, as the representation of this morpheme due to its various forms.
Gaps in the realization of morphological categories raise some interesting but puzzling questions, especially in a generative framework. Previous literature reports that speakers are unable or in some cases reluctant to create particular inflected forms for particular words (Albright 2003). These gaps are sometimes arbitrary (or due to unfamiliarity according to Albright) and sometimes due to phonological restrictions. The appearance of gaps in Thao belongs to the latter category.

The paper is structured as follows. In section 2, I investigate the behavior of *UM* in Thao. In section 3, I discuss the paradigmatic gap. Section 4 is the conclusion.

### 2. *UM* infixation in Thao

Thao has five realizations of *UM*. First, *UM* appears as an infix in stems beginning with two consonants:

<table>
<thead>
<tr>
<th>(1) -Um- infixation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tqir</td>
<td>t-um-qir</td>
<td>‘protest, get angry and leave’</td>
</tr>
<tr>
<td>b. ŋ̣rak</td>
<td>ŋ̣-um-rak</td>
<td>‘untie’</td>
</tr>
<tr>
<td>c. ŋ̣ð̣aq</td>
<td>ŋ̣-um-ð̣aq</td>
<td>‘stick, adhere’</td>
</tr>
<tr>
<td>d. ŋ̣qa</td>
<td>ŋ̣-um-qa</td>
<td>‘bequeath, leave an inheritance to someone’</td>
</tr>
<tr>
<td>e. θpiq</td>
<td>θ-um-piq</td>
<td>‘thresh grains by beating the stalks against something’</td>
</tr>
<tr>
<td>f. ɬ̣baha</td>
<td>ɬ̣-um-ɬ̣baha</td>
<td>‘float’</td>
</tr>
<tr>
<td>g. ɬ̣fawa</td>
<td>ɬ̣-um-fawa</td>
<td>‘swell’</td>
</tr>
<tr>
<td>h. kriu?</td>
<td>k-um-riu?</td>
<td>‘steal’</td>
</tr>
<tr>
<td>i. ktir</td>
<td>k-um-tir</td>
<td>‘pinch and twist’</td>
</tr>
<tr>
<td>j. ktun</td>
<td>k-um-tun</td>
<td>‘cut, as rope’</td>
</tr>
<tr>
<td>k. kði</td>
<td>k-um-ði</td>
<td>‘pinch between tongs’</td>
</tr>
<tr>
<td>l. qpit</td>
<td>q-um-pit</td>
<td>‘pinch’</td>
</tr>
</tbody>
</table>

In stems beginning with a single consonant, *UM* appears in most cases as *-m-* infixed after the first consonant:

<table>
<thead>
<tr>
<th>(2) -M- infixation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. tiuð</td>
<td>t-m-iuð</td>
<td>‘comb’</td>
</tr>
<tr>
<td>b. sansan</td>
<td>s-m-ansan</td>
<td>‘warm by a fire’</td>
</tr>
<tr>
<td>c. ŋ̣ut</td>
<td>ŋ̣-m-ut</td>
<td>‘measure’</td>
</tr>
<tr>
<td>d. ŋ̣anit</td>
<td>ŋ̣-m-anit</td>
<td>‘cry’</td>
</tr>
<tr>
<td>e. ŋ̣anup</td>
<td>ŋ̣-m-anup</td>
<td>‘bury something’</td>
</tr>
</tbody>
</table>
f.  tüklük  t-m-üklük  ‘hiccup’
g.  kşıkš  k-m-škš  ‘shave’
h.  qusað  q-m-usað  ‘rain’
i.  hafuj  h-m-afuj  ‘chant’

However, when the stem begins with /ð, r, l/ or a vowel, UM appears as a prefixed m-:

(3) M- prefixation: /ð, r, l/ and vowel initials
a.  dāi  m-dāi  ‘tell, advise’
b.  riut  m-riut  ‘encircle, surround’
c.  ripu  m-ripu  ‘surround something, as a crowd of spectators’
d.  riqas  m-riqas  ‘see, examine’
e.  luslus  m-luslus  ‘grate or shred, as vegetables’
f.  ațtu  m-agtu  ‘contemplate, think about’
g.  iup  m-iup  ‘blow with the mouth’
h.  uša  m-uša  ‘go’

When the stem begins with /p/, the stem-initial consonant appears to be lost:

(4) Apparent loss of stem consonant: /p/

a.  pataʃ  mataʃ  ‘write’
b.  pilalaha  milalaha  ‘spread, of the legs’
c.  putun  mutun  ‘break, as a rope’
d.  pataqød  mataqød  ‘carry on a shoulder pole’

(5) The morphological category not realized: /b, d, f, m/

a.  braq  gap  ‘having a hole’
b.  balak  bak  gap  ‘tap the fingers’
c.  da  da  gap  ‘hang something up’
d.  faqaq  gap  ‘understand’
e.  farIW  gap  ‘buy’
f.  muqmuq  gap  ‘speak nonsense’

In stems beginning with /b, d, f, m/, the morphological category is not realized. Blust (1998:14) states that stems beginning in /b, d, f, m/ simply lack an AV form, or use a
different affix or an auxiliary. According to Blust (2003: 45), "these stems simply do not occur as active verbs. Where they do occur, the AV is marked by some other affixes or by an auxiliary, such as /kaj/ ([kaj]) ‘hit’, or /tu/, which is infixed with /-um/-, as with /k-m-ay balakbak/ ([kmaj ñbalakõbak]) ‘tap the fingers’, or /t-m-u barumbum/ ([tmu ñbarumõbum]) ‘to rumble, of distant thunder.’ Following Rice (2005), I call this phenomenon a paradigmatic gap.

Blust identifies the following morphemes as alternative means of realizing the AV function:

\[(6)\]
\[\]
a. /kaj/  "kay: hit, beat, attack, common in serial verb constructions (Blust 2003: 461)"
b. /kał/  "kalh-: a verb prefix attested with three bases, /bariz/ ‘level’, /sun/ ‘gather, collect’, and /tumbuz/ ‘rise up precipitously, as a cliff’ (Blust 2003: 101)"
c. /mia/  "mia-: a verb prefix; it often co-occurs with reduplication of the rightmost foot of the base (Blust 2003: 131)"
d. /ʃu/  "shu-: a verb-forming prefix; with pronominal and deictic bases it forms verbs meaning ‘bring X’ or ‘take X’ (Blust 2003: 176)"
e. /ʃau/  "shau-: a verb prefix; with bases that have an inherently locative sense or temporal sense it means ‘go to X’ or ‘arrive at X’ (Blust 2003: 172)"
f. /ta/  "ta-: a verb prefix attested with just two bases: /braq/ ‘hole’ and /fqat/ ‘shave off’ (Blust 2003: 177)"
g. /tau/  "tau-: a verb prefix with variable functions; combined with concrete nouns it means ‘carry X’ (Blust 2003: 180)"
h. /ti/  "ti-: a prefix, often found in conjunction with the locative suffix; beyond forming active verbs its function is unclear (Blust 2003: 181)"
i. /tu/  "tu-: a verb prefix of uncertain function (Blust 2003: 182)"

Of the 15 /b, ñd, f, m/-initial stems in Blust’s (2003) Thao Dictionary, 2 take kay, 1 takes kalh-, 4 take mia-, 1 takes shu-, 1 takes shau-, 2 take ta-, 1 takes tau-, 1 takes ti- and 2 take tu-. It is clear that the AV meaning of these stems can be expressed by means other than affixation of UM. I will argue for the gap in expressing the AV meaning using UM affixation as a paradigmatic gap, following Rice (2005).

\[3\] Blust (2003: 45) states that "stems that begin with /b/ (b/), /d/ (õd/), a nasal, or a glide do not appear to take the /-um/- infix." However, stems with glide initials /n, j, w/ are stated as having no good examples in Blust (1998: 14). I discuss only the cases with b/, ñd, f, m/ in this paper.
2.1 The um/m alternation

Two alternants of UM as infix are observed, -um- and -m-. Both occur in stems beginning in one of the following consonants: /t, s, θ, t, k, q, h/. However, -um- occurs in stems with initial consonant clusters and -m- in stems with simple onsets. Consider the data excerpted from (1) and (2).

(7) The -um-/m- alternation
a. tiuð t-m-iuð ‘comb’
  b. tqir t-um-qir ‘protest, get angry and leave’
  c. fuṭ f-m-ut
  d. ŋrak ŋ-um-rak ‘untie’
  e. ṭanit θ-m-anit ‘cry’
  f. ŋukluk ŋ-um-faw ‘hiccup’
  g. ṭfawa ŋ-um-faw ‘swell’
  h. ki[i]ki[j] k-m-i[i]ki[j] ‘shave’
  i. kriuʔ k-um-riuʔ ‘steal’
  j. qusað q-m-usað ‘rain’
  k. qpit q-um-pit ‘pinch’
  l. hafuj h-m-afl ‘chant’

A preliminary analysis is that -um- is infixed to avoid creating tri-consonantal clusters (as in [7b, d, g, i, k]) while -m- is used in other cases. This is supported by the fact that Thao tolerates no more than two consonants in onset position (Blust 2003: 20).4

(8) *CCC ‘No triconsonantal clusters.’

The morpheme m, in turn, can appear as a prefix or an infix. Consider the data excerpted from (2) and (3).

(9) M as infix/prefix
a. tiuð t-m-iuð ‘comb’
  b. fuṭ f-m-ut ‘measure’
  c. ki[i]ki[j] k-m-i[i]ki[j] ‘shave’
  d. qusað q-m-usað ‘rain’

4 There is one word with three consonants in initial position recorded in Blust (2003: 23), /ŋkidi/ ‘affect something entirely; leave nothing behind’. Word-medial trilateral consonant clusters are recorded in at least seven forms in the dictionary. No word-final consonant clusters are found.
e.  δai  m-δai  'tell, advise'
f.  ripu  m-ripu  'surround something, as a crowd of spectators'
g.  luslus  m-luslus  'grate or shred, as vegetables'
h.  aŋtu  m-agtu  'contemplate, think about'
i.  iup  m-iup  'blow with the mouth'
j.  uʃa  m-uʃa  'go'

A preliminary analysis is that onset clusters must rise in sonority. Cross-linguistically, the preference for rising sonority in onsets is widely attested (Clements 1990). Sonority Sequencing (10) determines the affixing position of m: onset clusters */ðm/, */rm/ and */lm/ are not allowed because they have the wrong sonority profile; the same constraint penalizes clusters like */mt/, */mʃ/, */mk/, or */mq/. This is supported by the fact that consonants of higher sonority tend not to be the first member of a cluster (Blust 2003: 20).

(10) SONSEQ  ‘The sonority profile of the onset rises.’

2.1.1 Lexically listed <um, m>

The next step of the analysis is to determine the lexical form of this morpheme. There are several possible ways to analyze the alternation. One possibility is that /um/ is the lexical form, and the vowel /u/ is deleted through some process. Another possibility is that /m/ is the lexical form of the morpheme, and the vowel /u/ is added to avoid tri-consonantal clusters. Still another possibility is that /um/ and /m/ are both listed in the lexicon; the grammar chooses one that fits the context. In the next section I will argue for the third analysis: both /um/ and /m/ are listed in the lexicon.

The first possibility to consider is that /um/ is the only lexical form and the vowel /u/ is deleted through some process. For example, when a stem beginning with a single

---

5 I assume that /ð/, /r/ and /l/ pattern together based on the observation of Li and Blust. "/ð/ sounds somewhat like [l] in word-final... (Li 1976: 223)." "Li states that he initially misheard the phoneme /ð/ as a voiced lateral in final position, .... In my own experience careful checking sometimes resulted in my transcribing /ð/ with [l] and sometimes with [o] in the same morpheme.... To add to the complications, I also recorded variation between /l/ and /r/, and also between /r/ and /ð/ (Blust 2003: 33-34)."

6 These consonants include /d/, /l/, /m/, /n/, /r/, /w/, /j/ and /ð/.

7 Among 686 stems that take UM, 619 of them take either prefix m- or infix -m-, and 67 take -um-.
consonant is affixed with /um/, the vowel /u/ is deleted (*t-um-iu⟩/t-m-iu⟩ ‘comb’); when a stem beginning with two consonants is affixed with /um/, on the other hand, *CCC (8) prevents the vowel from deleting (*t-um-qir⟩/t-m-qir⟩ ‘protest’).

Possible support for this analysis comes from evidence showing that high vowels /i, u/ may be deleted in Thao. Consider the following data:

(11) Medial high vowel syncope (Blust 2003:57)
   a. /l-m-akup/ (AV) [lakp-in] (PV) ‘hold in the two cupped hands’
   b. /s-m-akup/ (AV) [sakp-in] (PV) ‘catch, seize’
   c. /juðup/ (AV) [juðup-an] (LV) ‘heal, of a wound’
   d. /j-m-aðik/ (AV) [jəðk-in] (PV) ‘smell’
   e. /t-m-apiʃ/ (AV) [tap-in] (PV) ‘winnow’

The high vowels /i, u/ are sometimes syncopated, as in (11), while other vowels stay intact. High vowels are also sometimes deleted in particles as in (12) and (13).

(12) Deletion of /u/ in the particle /tu/
   a. /ata  t u agqaqili sa aniamin/ don’t TU carry-on-hip SA thing [atat agqaqili sa aniamin] ‘Don’t carry things on your hip’

   b. /θiθu  uka  t u na-naʃnaʃ/ 3sg not-have TU foresight [θiθu ?ukat nanaʃnaʃ] ‘He/she is ignorant, has no foresight’

(13) Deletion of /i/ in the particle /ti/
   /m-ðai  ti  Ali/ said-AV TI Ali [mðait ali] ‘Ali said (it)’

However, high vowel deletion does not apply across the board in this language if we take into consideration another infix marking realis aspect, -in-.

---

8 Patient voice.
9 Locative voice.
Qualifying Paper

Lu

(14) -In- infixation
   a. /θpíq/ → /θ-in-píq/ ‘thresh’
   b. /θánup/ → /θ-in-ánup/ ‘bury’

(14b) has the same environment (stems with two syllables beginning with a single consonant) as do those in (2), so the vowel /i/ is expected to be deleted. Instead, /i/ is left unaffected. If we were to claim that high vowels are deleted, then high vowel deletion would have to be restricted. On the other hand, if we assume that both <um, m> are listed as the AV marker in the lexicon and only in is listed as the realis marker, then high vowel deletion does not need to be restricted. Deleting the vowel in in will be ruled out by faithfulness constraints while deletion is irrelevant for <um, m>.

The second possibility to consider is that /m/ is the lexical form of the morpheme, and the vowel /u/ is added to avoid tri-consonantal clusters (*/t-m-qir/ → /tu-m-qir/ ‘protest’). However, as I will show in the following analysis, the position of /u/-epenthesis cannot be predicted.

If we assume prefixal infixation in a sense of Prince & Smolensky (1993/2004), the left edge of /m/ is aligned to the left edge of a prosodic word.

(15) ALIGN-L[m] ‘Align the left edge of m with the left edge of a PrWd.’

The alternation between m as prefix (/m-ripu/ ‘surround’) and infix (/t-m-iu/ ‘comb’) is due to the ranking of SONSEQ >> ALIGN-L[m]. DEP (17) penalizes any epenthesized segments. However, an epenthesized vowel is needed if *CCC is violated (*CCC >> DEP). Consider the following tableau:

(16)  

<table>
<thead>
<tr>
<th>/m, tiuð/ ‘comb’</th>
<th>*CCC</th>
<th>SONSEQ</th>
<th>DEP</th>
<th>ALIGN-L[m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. t-m-iuð</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. m-tiuð</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. m-utiðu</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. tu-m-iuð</td>
<td></td>
<td></td>
<td>*!</td>
<td>**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/m, ripu/ ‘surround’</th>
<th>*CCC</th>
<th>SONSEQ</th>
<th>DEP</th>
<th>ALIGN-L[m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>e. m-ripu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. r-m-ipu</td>
<td>*!</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>g. mu-ripu</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>h. tu-m-ipu</td>
<td></td>
<td>*!</td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>
In (16), the constraint *CCC is not effective since there is no tri-consonantal cluster created by affixing \( m \). The affixing position of \( m \) is determined by \( \text{SONSEQ} \) as in candidates (16a, b, e, f). Epenthesis of a vowel to satisfy \( \text{SONSEQ} \) is more costly than violating \( \text{ALIGN-L}[m] \) as in candidates (16c, d, g, h). With the same ranking, we expect an inserted vowel to appear when *CCC is violated. Consider the following tableau:

\[
\begin{array}{|c|c|c|c|c|}
\hline
/m, tqir/ & \text{'protest'} & *\text{CCC} & \text{SONSEQ} & \text{DEP} & \text{ALIGN-L}[m] \\
\hline
\text{a. tu}-m-qir & & * & & & **! \\
\text{b. t-m-qir} & & & \text{!} & & \\
\text{c. m-tqir} & & & \text{!} & & \\
\text{d. m-utqir} & & & & \text{!} & \\
\text{e. tqu-m-ir} & & & \text{!} & & ***! \\
\text{f. tq-m-ir} & & & & & ** \\
\hline
\end{array}
\]

The vowel /u/ can be inserted in several places to avoid *CCC violation in (18b, c, f). The correct output (18a) violates ALIGN-L[m] twice because \( m \) is two segments away from the edge. (18e) encounters the same problem. The undesirable output (18d) is chosen because inserting the vowel after \( m \) does not violate any alignment constraint. That is to say, ALIGN-L[m] incorrectly predicts the position of vowel epenthesis. The alignment constraint needs to be revised to ensure that \( m \) is still an infix when the higher phonological constraints (*CCC and \( \text{SONSEQ} \)) do not decide the position of this affix.

To ensure that \( m \) is always an infix, we could try assuming that the left edge of the morpheme is aligned to the left edge of the first consonant.

\[
\begin{array}{|c|c|c|c|c|}
\hline
/m, tqir/ & \text{'protest'} & *\text{CCC} & \text{SONSEQ} & \text{DEP} & \text{ALIGN-L}[m] \\
\hline
\text{a. tu-m-qir} & & & * & & **! \\
\text{b. t-m-qir} & & & \text{!} & & \\
\text{c. m-tqir} & & & \text{!} & & \\
\text{d. m-utqir} & & & & \text{!} & \\
\text{e. tqu-m-ir} & & & \text{!} & & ***! \\
\text{f. tq-m-ir} & & & & & ** \\
\hline
\end{array}
\]

Constraint (19) encodes the proposal of Yu (2007) that an affix can align its edges to specific phonological units. The set of ‘edge pivots’ is after the first consonant, vowel or syllable, or before the final consonant, vowel or syllable. The set of ‘prominence pivots’ is before/after the stressed foot, stressed syllable, or stressed vowel. However, we will still need high-ranked \( \text{SONSEQ} \) to predict the prefix/infix alternation of \( m \). Consider the following tableau:
ALIGN[m, L, C1, R] is irrelevant in (20). Losing candidates have been ruled out by SONSEQ and DEP. With this revised alignment constraint, however, we still encounter the same problem with the epenthesized vowel. Consider tableau (21):

This alignment constraint favors candidate (21f) over (21a) with the vowel /u/ inserted after m since the alignment constraint requires m to be next to the first consonant.

Another possibility is ALIGN[m, R, V1, L] since m always shows up to the left of the first vowel.

(22) ALIGN[m, R, V1, L]  ‘Align the right edge of m to the left edge of the first vowel.’

This alignment constraint has no effect in the following tableau because the placement of the affix is determined by SONSEQ. There is no need for an inserted vowel. Losing candidates in (23) have been ruled out by SONSEQ and DEP.
But again, the alignment constraint incorrectly predicts the position of the epenthesized vowel in tableau (24) when *CCC comes into play.

ALIGN [m, R, V₁, L] favors candidate (24e) over the correct output (24a). To correctly predict the position of the inserted vowel, we would have to assume that the right edge of m is aligned to the left edge of the second consonant, so that the inserted vowel would come before the affix (/tu-m-qir/). However, we would need two different alignment constraints for the same affix (note that m does not align to the second consonant in /t-m-iuð/ or /m-ripu/).

Having both <um, m> listed in the lexicon, on the other hand, solves this theoretical problem. With two phonological constraints, ONSET and SONSEQ, ranking over an alignment constraint, ALIGN-L[UM], the position of the affix is predicted.

---

10 Note that infixation to the second consonant is not an option in Yu’s system.

11 An alternative analysis to avoid having both forms listed in the lexicon is brought to my attention by Jose Elias-Ulloa (personal communication). If the morpheme bears a mora, the position of the epenthesized vowel can be predicted. [tum.qir] is predicted because /m/ is moraic; *[m.u.qir], *[mu.tqir] and *[tqu.mir]
(25) a. ONSET ‘Every syllable must have an onset.’
   b. ALIGN-L[UM] ‘Align the left edge of UM with the left edge of the prosodic word.’

Um as infix is predicted by the ranking of ONSET over ALIGN-L[UM], as in candidate (a) in the following tableau. M as prefix or infix is predicted by the ranking of SONSEQ over ALIGN-L[UM], as candidates in (26c, e).

(26) <um, m>

<table>
<thead>
<tr>
<th>/&lt;um, m&gt;, tqir/</th>
<th>ONSET</th>
<th>SONSEQ</th>
<th>ALIGN-L[UM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. t-um-qir</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. um-tqir</td>
<td></td>
<td>!</td>
<td></td>
</tr>
<tr>
<td>/&lt;um, m&gt;, ripu/</td>
<td>ONSET</td>
<td>SONSEQ</td>
<td>ALIGN-L[UM]</td>
</tr>
<tr>
<td>c. r-m-ipu</td>
<td></td>
<td>!</td>
<td>*</td>
</tr>
<tr>
<td>d. m-ripu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/&lt;um, m&gt;, hafuj/</td>
<td>ONSET</td>
<td>SONSEQ</td>
<td>ALIGN-L[UM]</td>
</tr>
<tr>
<td>e. h-m-afuj</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>f. m-hafuj</td>
<td></td>
<td>!</td>
<td></td>
</tr>
</tbody>
</table>

The arguments above show that um is not predictable from m nor the other way around. Both <um, m> need to be listed as basic forms. I will present an analysis for the phonological conditions determining the choice among two allomorphs in the next section.

2.1.2 Phonologically conditioned allomorphs <um, m>

As for the phonological conditions of these two allomorphs, I propose that this has to do with the stress pattern of this language. In this section, I will present the phoneme inventory and the stress pattern related to the choice of allomorph.

Thao has five vowels /i/, /u/, /a/, /o/ and /e/, with the latter two found only in personal names. The consonant inventory is in (27).

---

are out because onset is not moraic. Interactions between /m/ being moraic and other prosodic constraints favor /m/ as onset in [tmiuð] ([tum.iuð], [um.tiuð], ONSET), and in [mripu] ([um.ripu], [rum.ipu]). This alternative is not explored in this paper because there is no evidence showing that coda consonants in Thao project a mora.

Blust uses [o] and [e] as variants of /u/ and /i/ as well.

---
(27) Thao consonant inventory (Blust 2003:18)

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Dental</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>p</td>
<td>q</td>
<td>t</td>
<td>?d</td>
<td>k</td>
<td>q</td>
</tr>
<tr>
<td>Nasals</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td>f</td>
<td>s, θ</td>
<td>δ</td>
<td>f</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>Laterals</td>
<td>l14</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flap</td>
<td>r(r)15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glides</td>
<td>w</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stress appears on the penult in stems with more than one syllable, and in monosyllabic stems, stress falls on the only available vowel. It is reported that the affixed forms of such monosyllables retain stress on the stem (Blust 2003:37). Stress remains on the stem when monosyllabic stems are prefixed. Consider the following data:

(28) Stress pattern of monosyllabic stems

a. /kán/ ‘eat’ /ma-kán/ [makán] ‘eat (future)’
b. /rúni/ ‘prepare’ /kaʃi-rúni/ [kaʃiʃíni] ‘get something ready’
c. /júdì/ ‘become more intense’ /puʃ-ʃúd-ak/ [puʃʃúdak] ‘made something more extreme’

Although the stress generally appears on the penult in this language, when monosyllabic stems are prefixed, the stress remains on the stem, as in (28a-b). I propose that the penultimate stress pattern in Thao reflects a trochaic foot at the right edge of a word. A monosyllabic stem forms a stress foot by itself. The stress pattern interacts with the choice of UM allomorph. Consider the following data:

(29) Foot structure selects m/um

a. (t-m-ʃūd) * t-u(m-ʃūd) ‘comb’
b. (ʃ-m-út) * ʃ-u(m-út) ‘measure’
c. (θ-m-ánit) * θ-u(m-ánit) ‘cry’
d. (k-m-ʃįkij) * k-u(m-ʃįkij) ‘shave’
e. (h-m-ʃuʃ) * h-u(m-ʃuʃ) ‘chant’
f. (m-ʃipu) * r-u(m-ʃipu) ‘surround something’
g. (m-şúslus) * l-u(m-şúslus) ‘grate or shred, as vegetables’

---

13 The velar nasal seems to be phonemic only in names, onomatopoeic words, and loans (Blust 2003: 18).
14 This is a voiceless fricative.
15 A trill [r] was recorded in Li 1976 while a flapped [ɾ] was recorded in Blust 2003 (p. 33).
16 ‘Stress appears contrastively on the last vowel in some 2-3% of base forms.’ (Blust 2003: 19).
h. (m-ţíntu) * u(m-ţíntu) 'contemplate, think about'
i. (m-ţúp) * u(m-ţúp) 'blow with the mouth'

If we assign the stress accordingly, we find that the allomorph m is chosen when the vowel /u/ is outside of the stress foot. It seems that there is a preference for words with only one foot containing no unparsed syllables. This preference is confirmed by the frequent word shapes in this language: CVC, CVCV, CVCCV, CVCCVC, CCV, and CCVC (Blust 2003: 19).

The allomorph um is chosen to avoid a triconsonantal cluster (the segments in bold). Consider the following data:

(30) *CCC selects m/um
b. t-um-(qír) * (t-m-qír) 'protest'
d. ţ-um-(rák) * (ţ-m-rák) 'untie'
f. ţ-um-(piq) * (ţ-m-piq) 'thresh grains by beating the stalks'
h. k-um-(ríúʔ) * (k-m-ríúʔ) 'steal'

The stress pattern is predicted by FT-TROCHAIC (31a) and STRESSSTEM (31b). The ranking of ALLFTRIGHT (31c) over PARSE (31d) indicates that it is more important to have only one stress than to avoid unparsed syllables.

(31) a. FT-TROCHAIC 'Foot must be trochaic.'
b. STRESSSTEM 'Stems must be stressed.'
c. ALLFTRIGHT 'All feet must be aligned with the right edge of PrWd.'
d. PARSE 'Syllables are parsed by feet.'

With these constraints, m is chosen in situations where the vowel /u/ is outside of the stress foot.

(32) <um, m>

<table>
<thead>
<tr>
<th>/&lt;um, m&gt;, hafuj/ ‘chant’</th>
<th>FT-TROCHAIC</th>
<th>STRESSSTEM</th>
<th>ALLFTRIGHT</th>
<th>PARSE</th>
<th>ALIGN-L[UM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (hmáfu/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. hu(máfu/)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. (humá)fu/)</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. (hu)(máfu/)</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. (húma)fu/)</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Choosing the allomorph um produces a violation of PARSE, as in (32b, g). Assigning a head final foot violates FT-TROCHAIC, as in (32c, h). Having more than one foot in order to satisfy PARSE violates ALLFTRIGHT, as in (32d, i) and having the stress elsewhere on the affix violates STRESSSTEM, as in (32e, j).

The high-ranked phonological constraint *CCC, as defined earlier, favors the allomorph um over m, even though PARSE is violated, as in candidates (a) and (e) in the following tableau. And SONSEQ determines the placement of the allomorph m, as in (33f, g, j, k).

(33) <um, m>
2.2 Coalescence

We can now turn our attention to cases in which the stem-initial consonant appears to be lost when the stem begins with /p/.

(34) Apparent loss of stem consonant: /p/
   a. pataf / mataf / ‘to write’
   b. pilalaha / milalaha / ‘to spread, of the legs’
   c. putun / mutun / ‘to break, as a rope’
   d. pataqad / mataqad / ‘to carry on a shoulder pole’

UM as apparent loss of stem consonant shown in (34) can be accounted for by a feature co-occurrence restriction (Obligatory Contour Principle, or OCP) against two labial stops—Thao disfavors two adjacent labials on the consonantal projection created by the initial labial consonant and /m/ in UM.17

(35) OCP-L ‘No adjacent labials in consonantal projection.’

The question now is whether (34) should be analyzed as deletion of the stem /p/ or as coalescence of /p/ and /m/. Notice that stems with other initial labials /m, f, ʰb/ simply do not occur with this morpheme (the appearance of gaps will be discussed in detail in the next section).18 To explain why this affix behaves differently with /p/ as opposed to other labials /m, f, ʰb/, I propose that the loss of the stem consonant results from coalescence of /p/ and /m/ rather than deletion of the stem /p/ due to co-articulation. Coalescence of /p/ and /m/ is allowed because both are labial stops. Segments of different places of articulation cannot be coalesced (/<um₁, m₁,>, t₂iuð/ → *[n₁₂iuð] ‘comb’). Other labials,

---

17 Note that the restriction has to be two labials on consonantal tier because OCP is found in the following data as well.

18 There is one consonant left unexplained in the gap category—/d/. In other Austronesian languages, there are some cases of the OCP spreading to nasal (Kanakanvu [Wu 2006]), and sonorant (Chamorro [Klein 2005]); and there are some cases of the OCP restricting to only sonorant labial (Tagalog [Orgun & Sprouse 1999]). My speculation is that the speakers over-generalized the category being affected from labial to nasal since the target segment is /m/ in UM, which is also a nasal; the same over-generalization happened to cases where sonorant is affected. For cases where the OCP is narrowed down to only sonorant labials, the speakers under-generalized the category. My speculation for /d/ being affected is that the speakers over-generalized from the pre-glottalized labial /b/ to /d/.

16
/m, f, ʰb/, do not coalesce with /m/ because two consonants with different manner features cannot be coalesced. Stem initial /f, ʰb/ and UM cannot coalesce because they are not similar enough in terms of manner. IDENT-[MANNER] is motivated to capture the fact that /f, ʰb/ fail to coalesce with /m/ in UM because manner of articulation needs to be kept. In contrast, stem initial /m/ and UM cannot coalesce because they are identical; if coalescence occurred, there would be no recoverable exponent of the AV morpheme. REALIZE-MORPH (cf. Prince & Smolensky 1993/2004, Orgun & Sprouse 1999) prevents /m/ from undergoing coalescence because of recoverability of the morpheme.¹⁹

(36) a. IDENT-[MANNER] ‘Manner feature of an input segment must be identical in the output.’
   b. REALIZE-MORPH ‘A morpheme must have some realization in the output.’

If the apparent loss of /p/ were to be taken as deletion instead of coalescence, we would then have to allow random deletion of stem initial consonants of any place/manner of articulation. Thus, MAX must outrank UNIFORMITY.

(37) a. MAX ‘No deletion.’
   b. UNIFORMITY ‘No coalescence.’

(38) OCP violation on labial

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(m₁,ataʃ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. (m₁,ataʃ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. p₁-u(m₁-ataʃ)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. (p₁m₁-ataʃ)</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. (m₁,ataʃ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. (p₁ataʃ)</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The optimal candidate (38a) only violates lower-ranked UNIFORMITY. (38b, c) violate the higher ranked OCP-L. Keeping only the affix violates MAX, as in (38d), and keeping only the stem-initial consonant violates REALIZE-MORPH, as in (38e).

¹⁹ This constraint is sometimes used as a cover term ‘REALISE µ’ or ‘M-PARSE’ to regulate that the morphological components must have some realization in phonological segments.
2.3 Gaps
The next question is why stems with other labial initials /m, f, ʰb/, in response to OCP-L, simply do not occur with UM.

We successfully prevent stems beginning with /m, f, ʰb/ from coalescing with UM to escape the OCP-L violation by using IDENT-[MANNER] and REALIZE-MORPH in the previous section. However, the constraint ranking that allows the coalescence candidate to be chosen does not successfully choose the gap candidate. As a matter of fact, the general assumption of OT that all constraints are violable, and that candidates that violate the constraints less seriously win, it should be impossible to favor ‘nothing’ over other less faithful candidates. The appearance of gaps requires additional mechanisms to be introduced into the theory. In the following section, I consider mechanisms that have been proposed to account for systematic gaps.

3. Paradigmatic gaps
In Thao, gaps triggered by the OCP are found only in UM concatenation; other morphemes with labials do not trigger gaps. Furthermore, only stem-initial consonants trigger gaps in Thao while labials elsewhere in a stem do not.

Gaps triggered by the OCP effect only appear when UM is concatenated with stems beginning in /m, f, ʰb/ initials (see also fn.18), repeated below in (39):

(39) The morphological category not realized
   a. ʰbraq  gap  ‘having a hole’
   b. ʰbalak ʰbak  gap  ‘tap the fingers’
   c. ʰda'da  gap  ‘hang something up’
   d. fazaq  gap  ‘understand’
   e. fariw  gap  ‘buy’
   f. muqmuq  gap  ‘speak nonsense’

There are other morphemes with labials in Thao that do not show gaps in stems with /m, f, ʰb/ initials. Consider the following examples, which include the prefixes ma- ("Ma₁: a highly productive prefix marking stative verbs, or occasionally nouns derived from stative verbs" [Blust 2003:110]) and mu- ("Mu-: a high frequency prefix with various functions. It is most commonly used to derive verbs of motion" [Blust 2003:136]).
No gaps in prefixation with /ma-, mu-/ in labial stems

a. pruq ‘earth; down’ mu-pruq ‘go down, descend’
b. magkaci ‘other side’ mu-magkaci ‘go to the other side’
c. maða ‘shallows’ mu-maða ‘enter the shallows’
d. "balaj"baj ‘slack’ mu-"baraj"baj ‘go slack’
e. fafaw ‘top’ mu-fafaw ‘go to the top’
f. "braq ‘hole’ ma-"braq ‘having a hole’

Furthermore, the OCP effect seems to be restricted to stem-initial consonants. Any adjacent labials caused by UM infixation in other positions in the stem do not trigger gaps.

Infixation before labials in non-initial positions

a. qpit q-um-pit ‘pinch’
b. hafuj h-m-afuj ‘to chant’
c. "baha l-um-"baha ‘to float’
d. fawa l-um-fawa ‘to swell’

I will focus on these facts in the following subsections.

3.1 Morpheme-specific gaps

The first point of interest is that the appearance of gaps triggered by the OCP effect is morpheme specific. When both UM and ma-, which marks stative actor voice, concatenate to the same stem, a gap occurs with the former but not with the latter. Such forms are relatively rare because generally a verb stem does not take both UM, marking a more dynamic action and ma-, marking a more stative action.

Ma-/UM- asymmetry

<table>
<thead>
<tr>
<th>Stem</th>
<th>ma-/</th>
<th>UM</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;braq</td>
<td>ma-&quot;braq ‘having a hole’</td>
<td>gap (*b-um-raq/*b-m-raq)</td>
</tr>
<tr>
<td>fafaw</td>
<td>ma-fafaw ‘understand’</td>
<td>gap (*f-um-azaq/*f-m-azaq)</td>
</tr>
<tr>
<td>fariw</td>
<td>ma-fariw ‘buy’</td>
<td>gap (*f-um-ariw/*f-m-ariw)</td>
</tr>
</tbody>
</table>

It is clear that a markedness constraint (here, OCP-L) should be violated by any marked structures. However, the marked structure created by ma- is accepted, but the one created by UM is left unformed. In comparing the two marked structures, we see no difference aside from the order of the two labials. The realization of ma- on a given stem seems to be of high priority than realization of UM.
I adopt Rice’s (2005) $\text{MAX}\{\text{CAT}\}$ constraints, set within an Optimal Paradigms approach, to deal with this problem. Rice proposes, following McCarthy (2005), that candidates are evaluated in paradigms. If a verbal paradigm requires the realization of six morphological categories, then there are six $\text{MAX}\{\text{CAT}\}$ constraints in charge of the realization of each category. With the different rankings of $\text{MAX}\{\text{CAT}\}$ constraints and faithfulness/markedness constraints, several typological predictions are made. Within a paradigm, if both markedness constraints and faithfulness constraints rank higher than constraints that require the expression of a certain morphological category, then the grammar chooses a gap over realizing the morphological category.

(43) A ranking which favors gaps (Rice 2005: 7)

\[
\text{MARKEDNESS, FAITHFULNESS} \gg \text{MAX}\{\text{CAT}\}
\]

When two $\text{MAX}\{\text{CAT}\}$ constraints rank higher than faithfulness constraints, both categories undergo repairs.

(44) Same problem, same repair (Rice 2007)

\[
\text{MARKEDNESS} \gg \text{MAX}\{\text{CAT}_1\}, \text{MAX}\{\text{CAT}_2\} \gg \text{FAITHFULNESS}
\]

When two $\text{MAX}\{\text{CAT}\}$ constraints rank lower than markedness constraints and faithfulness constraints, both morphological categories are left unformed.

(45) Same problem, both gaps (Rice 2007)

\[
\text{MARKEDNESS, FAITHFULNESS} \gg \text{MAX}\{\text{CAT}_1\}, \text{MAX}\{\text{CAT}_2\}
\]

When two $\text{MAX}\{\text{CAT}\}$ constraints flank faithfulness constraints, the higher ranked $\text{MAX}\{\text{CAT}_1\}$ undergoes repairs, while the lower ranked $\text{MAX}\{\text{CAT}_2\}$ is unformed. Rice terms this phenomenon same problem, different solutions.

(46) Same problem, different solutions: one repair, one gap (Rice 2005: 7)

\[
\text{MARKEDNESS} \gg \text{MAX}\{\text{CAT}_1\} \gg \text{FAITHFULNESS} \gg \text{MAX}\{\text{CAT}_2\}
\]

The appearance of gaps in $\text{UM}$ but not in $ma$-concatenation illustrates the same problem, different solution pattern. The constraint requiring the expression of $ma$- ($\text{MAX}\{\text{STATIVE-AV}\}$) ranks higher than the markedness constraint (OCP-L), so that it is not restricted by the marked structure. The constraint requiring the expression of $\text{UM}$ ($\text{MAX}\{\text{DYNAMIC-}$
AV}) is dominated by both markedness constraints and faithfulness constraints so the grammar favors gaps in this category. The data in Thao show another typological prediction of \( \text{MAX}\{\text{CAT}\} \) constraints.

(47) Same problem, one gap and one unaffected
\[
\text{MAX}\{\text{CAT}_1\} \gg \text{MARKEDNESS, FAITHFULNESS} \gg \text{MAX}\{\text{CAT}_2\}
\]

Although both morphemes are subject to the same OCP-L constraint, the requirement of realizing \( ma- \) (\( \text{MAX}\{\text{STATIVE-AV}\} \)) is higher than the constraint requiring realization of \( UM \) (\( \text{MAX}\{\text{DYNAMIC-AV}\} \)). The markedness constraint, OCP-L, ranks lower than the \( \text{MAX}\{\text{CAT}\} \) constraint on realizing \( ma- \) since no repair occurs to the adjacent labials. Consider the following tableau (other members in the paradigm are not listed for the sake of convenience).

(48) \( Ma- \) and \( UM \) affixation

<table>
<thead>
<tr>
<th></th>
<th>( \text{MAX}{\text{STATIVE-AV}} )</th>
<th>OCP-L</th>
<th>( \text{MAX}{\text{DYNAMIC-AV}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. f-um-ariw ma-fariw</td>
<td>**!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. f-um-airw</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. ma-fariw</td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d.</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(49) a. \( \text{MAX}\{\text{STATIVE-AV}\} \) ‘Realize the expression of the stative actor voice morpheme \( ma- \).’

b. \( \text{MAX}\{\text{DYNAMIC-AV}\} \) ‘Realize the expression of the dynamic actor voice morpheme \( UM \).’

Expressing both \( UM \) and \( ma- \) in a paradigm violates OCP-L twice, as in (48a); realizing only \( UM \) violates the higher ranked \( \text{MAX}\{\text{STATIVE-AV}\} \), as in (48b). Thus, the candidate paradigm without \( UM \) wins. The two behaviors of \( UM \) triggered by OCP, coalescence and gaps, can now be explained (I omit other members in the paradigm for convenience).
Having both the stem-initial consonant and the /m/ in UM violates OCP-L, as in (50c, d, i, j). Deleting any of the two labials violates MAX, as in (50b, g, k, l). Stem initial /p/ is similar enough to /m/ to undergo coalescence. Thus, candidate (50a) does not violate OCP-L and MAX{DYNAMIC-AV}. On the other hand, the higher-ranked IDENT[MANNER] forbids the coalescence of /f/ and /m/, and thus realization of UM is sacrificed, as in (50f).

### 3.2 Restriction of the OCP to stem initial consonants

In Thao, the OCP effect is restricted to stem-initial consonants. Any adjacent labials created by UM inflixation in other positions in the stem do not trigger gaps. Klein (2005) proposes that the domain of the OCP effect is the initial bimoraic foot in the base of UM-affixation in three Austronesian languages, Tagalog, Chamorro and Toba Batak, where no consecutive labial consonants are allowed within this domain. This solution is not ideal for Thao because this foot does not coincide with Thao’s prosodic foot (stress is formed by a trochaic foot on the right), and the only function of the foot is to govern the OCP effect.

I propose, following Beckman (1997), that this restriction is due to the special status of stem-/root-initial consonants. Stem-/root-initial material is generally assumed to have a privileged status because stem-/root-initial material facilitates lexical access and language
processing (Freedman & Landauer 1966, and Nooteboom 1981). Beckman (1997) provides a substantial number of examples showing the asymmetrical behavior of stem/root-initial vs. non-initial syllables (Beckman 1997:6): phonological processes do not apply to the same degree in the two positions. Thus, stem/root-initial positions are stronger positions, and thus are less likely to undergo alteration such as assimilation, dissimilation, and neutralization. 20 Adopting Walker’s (Walker 2000) segmental correspondence (as opposed to feature spreading), I propose that there is a correspondence relation between a segment in a strong position, here the initial consonant, and an adjacent sound bearing the same feature.21

Segmental correspondence relations, originally proposed by Walker (2000) to deal with long-distance consonant assimilation, involve similar segments in the output. Consider the following configuration:

\[(51) \quad \begin{array}{c|c|c|c} C_x & V & C_x & V \\ \hline [\alpha F] & [\alpha F] \end{array} \quad {\text{(Rose & Walker 2004)}} \]

Similar segments, represented here by a subscript \( x \), stand in correspondence, and are subject to correspondence constraints.

\[(52) \quad \text{Corr-C} \leftrightarrow \text{C} \]

Let \( S \) be an output string of segments and let \( C_i, C_j \) be segments that share a specified set of features \( F \). If \( C_i, C_j \in S \), then \( C_i \) is in a relation with \( C_j \); that is, \( C_i \) and \( C_j \) are correspondents of one another.

Taylor (2008) extends the correspondence account from assimilation to dissimilation: similar segments are in a correspondence relation and are driven to become less similar. Along the same lines, I further propose that strong positions, here the initial consonant, bear correspondence to adjacent sounds with similar features. Consider the following configuration:

\[
20 \text{ Other strong positions are stressed syllables, syllable onsets and root syllables.}
\]
\[
21 \text{ Adjacency is formalized through a proximity constraint (Rose & Walker 2004): PROXIMITY} \quad \text{Correspondent segments are located in adjacent syllables.}
\]
Segmental correspondence (cf. Walker 2000)

a. f\textsubscript{\textbeta} -m\textsubscript{\textbeta} ari w
   \[\text{[lab]} \text{[lab]}\]

b. h\textsubscript{\textbeta} -m\textsubscript{\textbeta} afuj
   \[\text{[lab]} \text{[lab]}\]

The stem-initial labial in (53a) corresponds to the adjacent labial, the infix \textit{-m-}, and thus these labials violate the OCP constraint. On the other hand, the stem-initial consonant in (53b) bears different features from the adjacent consonant, and thus they are not in a correspondence relation; the OCP constraint is not violated in this case.

I use a positional correspondence constraint \textit{Corr-}\textsubscript{\textsigma\textsubscript{1}}\leftrightarrow\textsubscript{C} to establish the correspondence between the stem-initial segment and its adjacent segments. This specific \textit{Corr-}\textsubscript{\textsigma\textsubscript{1}}\leftrightarrow\textsubscript{C} constraint ranks higher than the general \textit{Corr-}\textsubscript{C} \leftrightarrow\textsubscript{C}.

\hspace{5mm}
(54) \textit{Corr-}\textsubscript{\textsigma\textsubscript{1}}\leftrightarrow\textsubscript{C} (cf. Rose & Walker 2004)

Let \textit{S} be an output string of segments and let \textit{Ci}, \textit{Cj} be segments that share a specified set of features \textit{F}. If \textit{Ci}, \textit{Cj} \in \textit{S}, and \textit{Ci} is in stem-initial position, then \textit{Ci} is in a relation with \textit{Cj}; that is, \textit{Ci} and \textit{Cj} are correspondents of one another.

In this case, only segments in correspondence are subject to OCP-L. Consider the following tableau.
Qualifying Paper

Lu

(55) Initial/non-initial asymmetry

<table>
<thead>
<tr>
<th>/&lt;um, m&gt;, fariw/ 'buy'</th>
<th>OCP-L</th>
<th>Corr-σ₁↔C</th>
<th>MAX{DYNAMIC-AV}</th>
<th>Corr-C↔C</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. f₃₅-m₃-ariw</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. f₃₅-m₃-ariwₓ</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. f₃-m-ariw</td>
<td></td>
<td>!</td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/&lt;um, m&gt;, hafuj/ 'chant'</td>
<td>OCP-L</td>
<td>Corr-σ₁↔C</td>
<td>MAX{DYNAMIC-AV}</td>
<td>Corr-C↔C</td>
</tr>
<tr>
<td>e. h₃₅-m₃-af₃₅uj</td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. h₃₅-m-afuj</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>g.</td>
<td></td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

(55d) wins in the first part of the tableau. It violates only the comparatively lower-ranked MAX{DYNAMIC-AV}. (55a) violates OCP-L once because of the correspondent segments /fx/ and /mx/; it also violates the general correspondence constraint because the stem-final labial is not in correspondence. With all the labials in correspondence, (55b) violates OCP-L twice. (55c) is out because there is no correspondence relation among all the labials (note that there is no subscript next to the labial segments). When labials come later in the stem, the specific Corr-σ₁↔C is irrelevant, and (55f) is chosen.

The proposal of positional correspondence constraints establishes a domain for the OCP effect. Relevant cases from Thao are limited to stem-initial positions. It would be worth researching if other prominent/strong positions show the same locality effect.

In this section, I have discussed two problems raised by the appearance of gaps in UM concatenation relating to the OCP effect in Thao. The ‘same problem, different solution’ phenomenon is captured by ranking the markedness constraint in between two different OP constraints, MAX{STATIC-AV} and MAX{DYNAMIC-AV}, so that ma- is realized everywhere while UM causes a gap under the condition of two adjacent labials. As for the initial/non-initial asymmetry, I use a positional correspondence constraint, Corr-σ₁↔C, to
ensure the stem-initial segment is in correspondence with its adjacent segments with similar features.

4. Conclusion
The five realizations of $UM$ in Thao are: infixation of $-um-$ and of $-m-$, prefixation of $m-$, coalescence, and gaps. I argued for a dual lexical form for $UM$, <um, m>, and argued that this approach is superior to a competing analysis having only $m$. The alternation of $um/m$ is attributed to the stress pattern of this language; the position of $UM$ is predicted by an alignment constraint, ALIGN-L[UM] and two prosodic constraints, ONSET and SONSEQ. Coalescence and gaps are the results of OCP-L violations. The constraint ranking pertaining to the realizations of $UM$ is:

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Position} & \text{um/m} & \text{Coalescence} & \text{Gap} \\
\hline
\text{ONSET, SONSEQ, *CCC, FT-TROCHAIC, STRESSSTEM, ALL-FR-T} & \text{IDENT[MAN], MAX, OCP-L} & \text{Corr-$\sigma_1$}$\leftrightarrow$C \\
\text{PARSE} & \text{MAX{DYNAMIC-AV}} & \text{Corr-C}$\leftrightarrow$C \\
\text{ALIGN-L[UM]} & \text{UNIFORMITY} & & \\
\hline
\end{array}
\]

To account for the appearance of gaps in $UM$ but not in $ma$- concatenation, I propose different MAX{CAT} constraints, MAX{STATIVE-AV} and MAX{DYNAMIC-AV}, flanking a markedness constraint (OCP-L). This re-ranking of MAX{CAT} constraints with markedness/faithfulness constraints adds another typological prediction to Rice’s proposal. For initial/non-initial asymmetry, I propose a positional correspondence constraint to Walker’s segmental correspondence to ensure the stem-initial segment is in correspondence with its adjacent segment.
References


