Khmer Nominalizing and Causativizing Infixes

Eric Schiller
University of Chicago

Introduction
In this paper I reexamine the infixes of Old and Middle Khmer and attempt to reclassify them using the analytical framework of Autolexical Syntax.¹ I will demonstrate that the infixes cannot be treated merely as a result of metathesis since there are prefixes with the same phonological shape but different functions. The proposed analysis involves treating the infix as a simple affix at the morphosyntactic level, but with morphophonological requirements that vary depending on the specific affix involved.

Background on Khmer
There have been a number of descriptions of the Khmer language dating back to the earlier part of this century, e.g., Maspero 1915; Pinnow 1957; Jacob 1963; Ehrman et al. 1972; Huffman 1970b. The historical stages of Khmer are usually divided into three. The first is Old Khmer, which is divided into Pre-Angkorian (611-801 CE) and Angkorian (802-1431) periods. The second is Middle Khmer (1432 - mid 18th century). Modern Khmer describes the language from then up to the present.

Khmer has no productive derivational morphology. Old Khmer and Middle Khmer, however, had prolific derivational morphology (Jenner & Pou 1981, Jacob 1963, 1976) in the form of prefixes and infixes. Previous studies have proposed a large number of infix shapes, but the present work will show that many of these are merely a result of phonological processes. The infixes of Khmer have been studied by Judith Jacob (1963, 1976), Philip Jenner (1981), Jenner & Saveros Pou (1981), and Long Seam (1989).

Khmer phonology has undergone a number of significant developments, which are fairly easy to trace thanks to a long written history dating back to the 6th century. Scholars generally agree that Old Khmer had a
series of voiced and voiceless initial consonants; that this voicing distinction was replaced by a phonation type distinction of clear and breathy voice in the vowels and that this contrast eventually was lost, leaving behind a highly complex vowel system which displays an unusual "harmony" system. Modern Khmer initial consonants are still divided into two groups, representing the historically voiceless ones referred to as FIRST SERIES, and the historically voiced ones, referred to as SECOND SERIES.

Examples of the infixes which will be discussed are listed in (1-3). There are three distinct infixes classified functionally. One is a nominalizer that creates nouns, usually abstract, from verbs. The second forms causative or transitive verbs, and the third creates nominal agents from verbs.

1) -m- nominalizer
   a. kcəi  
      kəmcəi  'unripe'
      'something unripe'
   b. ?aoy  
      ?ɔəmnaoy 'give'
      'gift'
   c. khəŋ  
      kəmhəŋ 'angry'
      'anger'
   d. cʰəɯɯ  
      cʊmŋɯɯ 'ill'
      'illness'

2) -m- causativizer/transitivizer
   a. kɗau  'warm, hot'
      kəmdaʊ 'to warm, reheat'
   b. ?ɔt  'not'
      ?ɔmŋɔt 'endure privation'
   c. lhæ  'to relax'
      lɔhmhæ 'to amuse someone'
   d. kəɾp  'satisfied'
      kʊmŋdəɾp 'to salute; venerate'
3) -m- agentives
   a. chlak  ‘cut’
      OK: czmlak  ‘sculptor’
   b. kit    ‘think’
      OK: gmit  ‘thinker’ (?)
   c. cam    ‘wait for’
      OK: cmam  ‘sentinel’
   d. Skt: sev  ‘to serve’
      OK smev  ‘servant of god’

Previous Accounts
Jacob (1976) provides the following system of infixes for Old Khmer: -p-, -m-, -mn- -n- -N- and for Middle Khmer, she gives -p-, -m-, -mn- -n- -rn- -N-. Of these, only -rn- is an innovation. Modern Khmer maintains the full Middle Khmer set and shows evidence of an -rm- infix. But it is important to note that there is no productive infixation in Modern Khmer, and that the -rm- infix is therefore a short-lived phenomenon or may have been present in Middle Khmer but is simply unattested in the written sources. She considers the lack of examples of -rm- and -rn- inconsequential. The -rn- infix in Middle Khmer had a semantic function of marking utensils, a function not shared by any other infix. The -p- infix was used for artifacts. These affixes will not play a major role in this paper.

Jenner (1981) lists /-vmn-/ separately as a syllabic infix. He properly considers the assimilating nasal /-N/- only as part of complexes involving further prefixation but includes /-h/- as an allomorph of /-?-/-, for which I consider the evidence to be inadequate, and indeed, doubt that /-?-/- functions as an infix at all. With regard to the liquid infix /-L-/, it is by no means clear that there is any semantic effect, and the relationships claimed by Jenner seem to be stretching the point, e.g., /sαα/ ‘to be white, clear’, /sλαα/ ‘stew’ (Jenner has ‘to clarify, stew’, but there is no evidence of the semantics of clarification. We will see, however, that even within the nasals he proposes far too many infixes, as most are just phonologically conditioned allomorphs.

Syllable Structures of Khmer
In order to understand the morphophonology of Khmer, it is necessary to have a picture of the basic phonology
of Khmer. Most Khmer words are monosyllabic or sesquisyllabic. All disyllabic words are either borrowed or the result of affixation via non-productive morphological processes.

Underlying Syllable Structures are CV:, CV(:)C, CCVC. Surface forms are CV:, CVC, CVCVC and derived CVCCVC. All native Khmer words are stress final. The complex Khmer orthography reflects a surface phonological level. Morphophonological operations include prefixation, infixation, partial and full reduplication.

With these preliminaries out of the way we can begin our examination of the Khmer infixed. Example (4) involves infixation into an unaspirated first series cluster. In this case the affix forces a syllabification which splits the root. The phonotactically acceptable \( C_1C_2VC \) combines with the infix \( m \) to produce \( C_1mC_2VC \), which is not well formed at the morphophonological level. An epenthetic vowel must be inserted to create the well-formed string \( C_1VmC_2VC \). Because \( C_1 \) belongs to the first series, the vowel will be /α/.

4) \( k\alpha\epsiloni \) ‘unripe’ + -m- → \( k\alpha\epsiloni\epsiloni \) ‘something which is not ripe’

\[
\begin{array}{ll}
\text{Morphosyntax:} & X^0 \quad X^{-1} [Af m + X^0 k\epsiloni] \\
\text{Morphophonology:} & \#k\epsiloni\# \quad \#k\epsiloni\epsiloni\# \\
\text{Phonology:} & /k\epsiloni/ \quad /k\epsiloni\epsiloni/ \\
\text{Phonetics:} & [k\epsiloni] \quad [k\alpha\epsiloni\epsiloni]
\end{array}
\]

The next example (5) involves infixation into an unaspirated first series cluster where the second element is a liquid. The same analysis applies.

5) \( t\epsilonmr\epsilonp \) ‘to imitate’ + -m- → \( t\epsilonmr\epsilonp \) ‘example, model; manner of behaviour’

\[
\begin{array}{ll}
\text{Morphosyntax:} & X^0 \quad X^{-1} [Af m + X^0 \epsilonm\epsilonp] \\
\text{Morphophonology:} & \#\epsilonm\epsilonp\# \quad \#\epsilonm\epsilonp\epsilonp\# \\
\text{Phonology:} & /\epsilonm\epsilonp/ \quad /\epsilonm\epsilonp\epsilonp/ \\
\text{Phonetics:} & [\epsilonm] \quad [t\epsilonm\epsilonp]
\end{array}
\]
Ironically, less complex initials require more complicated analysis. With a simple CVC root the affix again forces a syllabification which splits the root. The phonotactically acceptable CVC combines with the infix m to produce CmVC, which would be well formed at the morphophonological level (cf. khmae), except for the requirement that this particular affix must go into the coda of the first syllable, which applies to the nominalizing and causativizing affixes but not to the agentive (cf. 3.c). CVmC and CVCm would both be ill formed regardless of the licensing requirement. So now an epenthetic consonant must be inserted in the onset of the second syllable to create the string CmCVC. Since gemination of consonants is forbidden by the phonotactics of Khmer, dissimilation takes place. It might be that the choice of velar nasal or alveopalatal nasal is determined by the backness of the vowel nucleus (cf. cumpuuu in the first example set), but this might also be the residue of a deletion of /h/, a possibility which is explored further below. In order to be well-formed, a vowel is still needed in the first syllable, and because C₁ belongs to the first series, the vowel will be /α/.

One could suggest that the output of an infixation process must conform to a given template, say CVCCVC, but this would account only for the nominalizing and causativizing examples, and the template would not apply to the agentive infix (sf. 3).

6) infixation into an unaspirated first series initial ?aoy ‘to give’ + -m- → ?αmnaoy ‘gift’

<table>
<thead>
<tr>
<th>Morphosyntax:</th>
<th>X₀</th>
<th>X⁻¹ [Af m + X₀ ?δy]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphophonology:</td>
<td>#?δy#</td>
<td>#?mōy#</td>
</tr>
<tr>
<td>Phonology:</td>
<td>/?δy/</td>
<td>/?mōy/</td>
</tr>
<tr>
<td>Phonetics</td>
<td>[?aoy]</td>
<td>[?αmnaoy]</td>
</tr>
</tbody>
</table>

At this point it is instructive to examine the phenomenon as applied to a second series initial consonant. The only difference is the quality of the epenthetic vowel, now
[ɔ] or [u] instead of [α]. If this process is synchronic, vowel harmony is taking place, but there are complications.

While "register harmony" has frequently been discussed among Khmerists, unfortunately I know of no published survey of the phenomena. Gérard Diffloth (p.c.) suggests that /h/ is permeable to register unless the preceding consonant is a liquid or nasal. This can account for the fact that there is no register harmony in /mhoup/ which would be [mhu:p] if the second register nature of /m/ were affective, but there is register harmony in [tiʔhia:n] which would be [tiʔhaːn] if /h/ blocked register harmony.

7) kit ‘to think’ + -m- → kəmnit ‘thought’

<table>
<thead>
<tr>
<th>Morphosyntax:</th>
<th>X⁰</th>
<th>X⁻¹ [Af m + X⁰ kit]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphophonology:</td>
<td>#kit#</td>
<td>#kəmnit#</td>
</tr>
<tr>
<td>Phonology:</td>
<td>/kit/</td>
<td>/kəmnit/</td>
</tr>
<tr>
<td>Phonetics:</td>
<td>[kit]</td>
<td>[kəmnit]</td>
</tr>
</tbody>
</table>

One should compare (7) with the Old Khmer form <gmit> in (3.b). This illustrates an important difference between the nominalizing infix and the agentive infix, a point which was mentioned above. The agentive ones can have the infix in any of three positions: breaking up an initial cluster, in the coda of the first syllable, or in the onset of the second syllable, e.g. pamañ ‘soldier’ from pañ ‘shoot’.⁴ Actually, it is not entirely clear that it can appear in the coda of the first syllable, but might even be in complementary distribution with the nominalizing infix. The only problematic example in my corpus is the one in (3.a), which has alternative meaning of ‘sculpture’ (Inscriptions numbered K.21 and K.165) and ‘to engrave’ (K.728). In other words it is found with both the causativizing and nominalizing uses as well, and the only example of it as an agentive is in a slavename (K.192).

The next example involves infixation into an aspirated first series initial.⁵ Again, the affix forces a syllabification which splits the root, but this time in a more interesting way. The initial consonant loses its aspiration, which shows up as a segmental consonant in the
onset of the second syllable. The phonotactically acceptable C^hVC combines with the infix m to produce C[-asp]mhVC, which is not well formed at the phonological level. An epenthetic vowel must be inserted to create a well-formed string C[-asp]Vmhc. Because the initial consonant belongs to the first series, the vowel will once again be /α/.

8) <khəŋ> ‘to be angry’ + -m- → <kəməŋ> ‘anger’

Morphosyntax: X⁰ X⁻¹ [Af m + X⁰ khəŋ]
Morphophonology: #khəŋ# #kəməŋ#
Phonology: /kəŋ/ /kəməŋ/
Phonetics: [kʰəŋ] [kəməŋ]

But this is not the only case of strange behavior when an infix follows an aspirated consonant. Consider the following words, both seen in the inscription K.292:

(9) chpan ‘to fight’ → cəmpan ‘war’

Here the aspiration, which might, ultimately, be carried over from a form <chan>, disappears entirely! There is a similar problem with <jhum> (I.d). What happened to the /h/, and could this be responsible for the velar nasal in /cumŋum/?

I am assuming in this discussion that orthographic aspirated consonants in Old Khmer reflected aspiration rather than a segmental /h/. Were the latter to be the case, then the aspiration simply falls in line with other initial clusters. The question of whether aspiration or a segmental /h/ is involved is a thorny one which has not yet been resolved for Modern Khmer, let alone for the older languages where native speakers are not available. The phonetic realization of this segment or feature varies. As Huffman (1970b:8) notes: “When the stops /p t c k/ occur as the first member of two-place initial consonant sequences /CC/-, they are unaspirated before /r s h/, slightly aspirated before voiceless stops and all continuants other than /r/ (except in the homorganic sequence /kn/-), and released with slight vocalism before /q⁶b d/
and in the sequence /kn-/ . When the consonants /qm l/ occur as the first consonant of /C-/ sequences, they are released with slight vocalism .

Gérard Diffloth (p.c) suggests that aspiration in initial clusters where the second element is a liquid or nasal (e.g . in chlak) is a post-Angkorian development, but that there are certainly items in Old Khmer which were either aspirated or had a segmental /h/ as part of a cluster. These were written as single glyphs, but that can hardly be taken as conclusive evidence.

In addition to the simple infixation seen in the previous examples, Khmer also shows evidence of double infixation and combinations of prefixation and infixation. In the case of double infixation, each of the infixes should be placed in the coda of the initial syllable, but this can only be accomplished via resyllabification. In the following example, the alternation between implosives and plosive unaspirates is predictable. The CVC root is combined with an instrumental infix to yield CpvVC. This infix is not specified for location (codas of first syllable), so the output with the initial cluster is acceptable. The addition of the nominalizing infix -m- to this structure, however, creates the form CmpVC since CpmVC or CpvCm are both ill-formed and create additional inter-modular mismatches - a topic which we will return to below. The addition of the standard epenthetic vowel yields CVmpVC, which is well-formed.

10)  dot 'to grill' + -b/p- [INS] → <tbot> 'to hold in cooking sticks' (PJ) 'to secure (meat, fish, etc .) with tongs prior to grilling' (TH)+ -m- → tǎmdot 'cooking sticks'

Morphosyntax:  \(X^0\) Af \(X^0\) \(X^{-1}[\text{Af m + } [X^0]]\)
Mor-phonology:  #tót# -b- #tót# #tmbót#
Phonology:  /dót/ /-b/- /tót/ /tmbót/
Phonetics:  [d] [t] [tfôt] [tǎmôt]

Combining prefixation with infixation is not problematic, requiring only the insertion of the epenthetic vowel to meet the relevant phonological conditions. An example is presented in (11).
11) \( <liəy> \) 'melted, mixed' + \(<r->\) \( → \) \( rəliəy \) 'melted, fluent' + \(-m-\) \( → \) \( rumliəy \) 'to destroy'

\[
\begin{array}{llll}
<r-> & <\text{-m}> & <\text{liəy}> & <\text{rumliəy}> \\
\text{Morphosyntax:} & \text{Af} & \text{Af} & \text{X}\text{₀} & \text{X}\text{⁻¹} \\
\text{Mor-phonology:} & \text{r-} & \text{-m-} & \#\text{liəy}\# & \#\text{rumliəy}\# \\
\text{Phonology:} & /\text{r-/} & /\text{-m-/} & /\text{liəy/} & /\text{rumliəy/} \\
\text{Phonetics:} & [\text{ᵣə ≈ ru}] & [\text{m}] & [\text{liəy}] & [\text{rumliəy}] \\
\end{array}
\]

Morphosyntax/Morphophonology Interface

We have seen that infixes require discrepant representations on the morphophonological and morphosyntactic dimensions. They are simple affixes at the morphosyntactic level but are categorized by their positional licenses at the morphophonological level. Given a morphosyntactic affix outside the stem, and a morphological infix inside the stem there must always be a mismatch between these two levels.

Multiple Dimensions and the Interface

Now I would like to discuss some of the theoretical consequences of the analysis. I will have to be brief, even sketchy here. Schiller (1992) contains a much fuller treatment of the Autolexical analysis of the same data, revised and simplified in Sadock & Schiller (1993).

To date, Autolexical theorists have concentrated primarily on the interaction of pairs of representations, with an occasional glance at trigrams that are created by the interaction of syntax, logico-semantics, and what was then a single-level morphology (Sadock 1991). But since logico-semantics involves only constituency relations without internal linear ordering, the complications are not as robust as when dimensions with rules pertaining to the linear ordering of elements are involved.

Although it has generally been assumed that each dimension in Autolexical theory is a context free phrase stucture grammar, or at least no more powerful a grammar than that, I take the mophosyntactic dimension to be a more or less finite-state grammar (more on this hedge below), thus providing a reasonable basis for the Relative Abstractness of Levels (REAL) principle of Schill-
The term "more-or-less," when applied to a formal grammar, may well be uncomfortable to many formal grammarians. But there is solid evidence that our formal devices for natural language analysis actually require a bit more, but not much more, power than that provided by our enumerated mathematical models. As Victor Yngve has pointed out on numerous occasions (1959, 1973, 1975), center-embedding has its limits. As far back as 1959 he formulated his "depth hypothesis" which still rings true. It may even be the case that our entire view of the appropriateness of the hierarchy of grammars assumed in linguistic theory is incorrect, as argued by Manaster-Ramer (1987).

I assume that the morphophonological dimension is also a finite-state grammar. Thus, we can reformulate the REAL hierarchy by the simple statement (12)

12) in the event of a conflict, the more constrained grammars take precedence over the less constrained grammars.

This should be understood as referring to the actual grammar as constrained by language specific rules, not the potential generative power as defined by the mechanism. If I may borrow and mangle a phrase of Michael Benedikt (1991), I would suggest that our principles are perhaps not hardwired into the brain (cf. Chomsky 1992) but are at the very least felicitous conventions for linguistic theory, derived from the constraints and opportunities that language seems to have chosen for itself. If we weaken the position concerning the finite state status of morphosyntax, and allow that in the case of compounding and rare center-embedding a context free grammar is required, the grammar which describes the morphosyntactic component is still less powerful than the one which describes the syntactic component in the languages (here: English and Khmer) under discussion.
Against Metathesis

At the start of the paper it was suggested that it is not possible to treat infixes as mere metathesis of prefixes. There is empirical support for this position in yet another Khmer labial affix. There is an affix /m/ which is phonetically identical to the one we have been discussing. It is always a prefix and never an infix. It has the meaning of the numeral “one” (a reduction of /muʔ/), and combines with nouns and classifiers (13).

13) mkhaaŋ ‘one side’ (cf. khaaŋ ‘side, direction)

If infixation were a mere phonological process then there is no reason why metathesis should be blocked, given that /mkʰəŋ/ is a perfectly acceptable form, and the difference between the vowels has no significance in the phonotactics of Khmer. But /mkʰəŋ/ is possible, I claim, because the affix has no syntactic reality in this word. If we analyzed the clitic m- in a manner similar to that of the infixes discussed below, there would be a second crossing of association lines, as shown in the following figure, which graphically represents the difference between the nominalizing infix and the simple m- prefix. This is illustrated in Figure 1 below. Note that if the affix were a prefix we would have no such complications, as is evident in the Figure 2.
Figure 1: Infix causing multiple violations of linearity
Figure 2: Prefix with no violations of Linearity Constraint

So the simple claim is that the numeral clitic cannot be an infix because it would create additional structural discrepancies. That may be so, but there is nothing in Auitolexical theory as formulated at present to prevent any number of discrepancies in a single linguistic unit. With our governing principles referring to node admissibility conditions at an individual level or on the pairing of representations from two different dimensions, there is nothing to suggest that the overall system has some sort of threshold for just what can be out of sync.

The lexical entries for the various infixes discussed this paper is presented in (14).

14) nominalizing causativizing agentive
Syntax: nil nil
Log-Sem: $f^-' \rightarrow f^-'[\text{abs}]$ $f^-' \rightarrow f^-'$ $f^-' \rightarrow f^-'$
Mor-syn: [VBL$\gg$NML] [VBL$\gg$VBL] [VBL$\gg$NML]
Mor-phon: Coda-lic.Cl Coda-lic. Cl Cl
Phonology:/m/ /m/ /m/
Disc-Fun: nil nil nil
Interface: -lc -lc -lc
The Locality Principle
I proposed in Schiller (1992) that there is in fact a constraint which limits the load a single unit can bear.

15) An atomic unit on a level with a finite-state grammar may not violate the Linearity Constraint more than once.

Under the recent reformulation of the Homomorphism Constraints this is no longer necessary. The requirement that the language conform to the Generalized Interface Principle (Sadock & Schiller 1993) accounts for the inadmissibility of the infix when the infix has syntactic reality.

Conclusion
In this paper I have provided a unified analysis of Old and Middle Khmer infixation, involving a restriction on licensing of the bilabial nasal infix in the coda of the initial syllable in the case of the nominalizing and causativizing infixes but with no such restriction on the agentive infix. I have also sketched a general treatment of infixes at the morphosyntax/morphophonology interface. While I consider the descriptive and analytical parts of the paper to be of greater importance here, I do wish to point out that the Autolexical framework provides a useful mechanism for discussing this data in a formal, and, I hope, clear manner.

Notes

1 For discussion of the Khmer data, I would like to thank Gérard Diffloth, Channy Sak-Humphrey and Sokhoum Khek. On the theory side, Bill Croft, Bill Darden, San Duanmu, John Goldsmith, Rich Janda, David Kathman, Kyunghwan Kim, Steve Lapointe, Karen Peterson, Jerry Sadock, Stephen Spackman and Arnold Zwicky all contributed useful discussion and ideas. All of the aforementioned should be held harmless from any misinterpretation or other errors on my part. There are no co-conspirators in this paper.
Orthographic conventions: <word> = transliteration, word = broad transcription citation form, #word# = morphophonological form, /word/ = phonological form (this level will be defined more clearly below), [word] = phonetic form, 'word' = gloss. In the case of Old Khmer, the relevant inscriptions are also identified.

This is only attested in a slavename, so the semantics may be conjecture on Jenner's part.

Gérard Diffloth (p.c.) pointed out an article by Goslier (which I have not yet tracked down) wherein it is suggested that slavenames often reflected the items for which a given slave was responsible. Thus at present there is little justification for considering (3.a) to be agentive.

The problem is the same for Modern Khmer, as has been noted by Henderson (1952) and Pinnow (1958), among others. Future research on Khmer and related languages may help to solve the puzzle by providing data on loanwords. For now, keeping an open mind is the best approach.

Huffman uses /q/ for [?].

Gérard Diffloth (p.c.) suggests that /muəy/ might be a combination of /mu/, attested as the form of numeral 'one' in Lamet, and /-i/. He offers a superior example in the Khmer word for 'mother', /mdaay/ which is a combination of a prefix /mee-/ 'female' with a word /daay/. The same arguments apply to this example as to the /m-/ numeral prefix.

This was pointed out to me by Stephen Spackman.

References
Carden, Guy. 1983. The non-finiteness of the word formation component. Linguistic Inquiry 14: 537-547
Culy, Christopher. 1985. The Complexity of the Vocab-
ulary of Bambara. Linguistics and Philosophy 8:345-351.
Lapointe, Steven G. 1988. Distinguishing Types of Mor-