17th Annual Meeting of the Southeast Asian Linguistics Society (SEALSXVII) University of Maryland, College Park, August 31st to September 2nd, 2007.

Proto-Mon-Khmer Vocalism: moving forward from Shorto's 'alternances'

Dr. Paul Sidwell
Centre for Research in Computational Linguistics
& Australian National University
<paulsidwell@yahoo.com>

Introduction

While we have had a century of more-or-less consensus views on the nature of the Proto-Mon-Khmer (PMK) consonant inventory, cries of exasperation have accompanied consideration of PMK vocalism. Writing in the first issue of MKS David Thomas reminded us that "… comparativists have stated flatly that regular sound-laws simply do not exist in Mon-Khmer vowels".

Harry Shorto developed a reconstruction of PMK vocalism which he defended at the 1973 Hawaii Austroasiatic conference, and elaborated in his postumously published A Mon-Khmer Comparative Dictionary (2006). Shorto based his analysis on a binary comparison of Old Mon and Written Khmer, and believed that in the correspondences he could discern a pattern of variation which may have reflected ancient system of vowel gradation. The principal types of variation he postulated were (i) between short and long vowel: **u**/**uu**, etc.; (ii) between simple vowel and diphthong: **ii**/**iə**, **uu**/**uə**; and (iii) between diphthong and **ə** : **iə**/**ə**, **uə**/**ə**. In the application of his reconstruction Shorto effectively set up a hierarchy in which, if the correspondences did not unambiguously point to a single proto-value, the presence of a diphthong reflex presumed a long-high proto-vowel (e.g. **uə** < ***uu**), and the presence of a long-low vowel presumed a proto-diphthong (e.g. **ɔɔ** < ***uə**). This approach greatly skews the reconstruction, making it particularly 'top heavy' (lacking low vowels).

Considering a wider range of cognates from MK languages it is evident that long-low monophthongs actually have a strong tendency to diphthongise within the family, and thus several of the sound-changes reconstructed by Shorto may be better reversed.

2. Shorto's method

In pursuing his phonological reconstruction of a language family that was (and still is) far from adequately documented, Shorto followed the well established procedure of establishing sound correspondences for several criterion languages for which extensive and reliable sources were available. In this case he selected two languages, Old Mon (for which he had compiled a dictionary) and Khmer as represented in the standard writing system (which was presumed to more or less faithfully reflect historical pronunciation).

Table 1: *Mon : Khmer correspondences*

PMK	Old Mon orthography	Old Mon phonology	Khmer orthography
***i**	*i, u, a, ī, ū, e, ui, ei*	**ø**	*i, ɯ*
***-iʔ**	*-i(')*	**-iʔ**	*-ī*
***ii**	*ī, i*	**i**	*i, ɯ*$^{(1)}$, *ī*$^{(1)}$, *e*$^{(2)}$
***-iiʔ**	*-ey*	**ɔy**	*-ai*
***e**	*e*	**e**	*ɛ*$^{(3)}$, *e*$^{(4)}$
***ee**	*ī, i*	**i**	*ɛ*$^{(5)}$, *e*$^{(6)}$
***a**	*a*; *e*$^{(7)}$	**a**$^{(8)}$	*a*, *ā*$^{(9)}$, *ai*$^{(10)}$, *e*$^{(11)}$, -$^{(12)}$
***aa**	*ā*; *e*$^{(13)}$	**ai**$^{(8)}$, **a**	*ā*
***ə**	*i* etc.; *a*$^{(14)}$, *o*$^{(12)}$	**ø**; **ɔ**$^{(12,14)}$	*ɔ*; *ɔ̄*$^{(9)}$; *a*$^{(2)}$
***əə**	*i* etc.; *u*, *ū*$^{(10)}$	**ø**; **u**$^{(10)}$	*ɤ*
***ɔ**	*o*$^{(8)}$, *a*	**ɔ**	*ɔ*; *ɔ̄*$^{(9)}$; *a*$^{(2)}$
***ɔɔ**	*o*$^{(8)}$, *a*	**ɔ**	*ɔ̄*; *ūv*$^{(15)}$
***o**	*u, ū, o*$^{(8)}$; *i* etc.	**ɯ**$^{(8)}$, **ø**	*o*
***oo**	*o*; *u*$^{(15)}$	**o**; **u**$^{(15)}$	*o*
***u**	*u, ū, o*$^{(8)}$; *i* etc.; *ū*, *u*$^{(10)}$	**ɯ**$^{(8)}$, **ø**; **u**$^{(10)}$	*u*; *ūv*$^{(15)}$, *ū*$^{(16)}$
***uu**	*ū, u*	**u**	*ū*, *u*$^{(12)}$; *o*$^{(2)}$
***-uuʔ**	*-ow*	**ɔw**	*-au*
***iə**	*e*; *a*$^{(17)}$	**ei**; **ɔ**$^{(17)}$(?); **iə**$^{(18)}$	*iə*
***uə**	*o*	**o**	*uə*
***ai**	*a*; *e*$^{(13)}$	**ai**$^{(8)}$, **a**	*ɛ*

Table of correspondences from Shorto (2006)

This use of only two languages stands in contrasts to the more common practice of comparing at least four languages to determine phonological correspondences, evidenced in such canonical works as Schmidt (1905), Dempwolff (1938), Li Fangkuei (1977) etc. It is also notable that these scholars characteristically assisted their interpretation of the correspondence sets by considering relevant available data from other related languages, a methodological necessity if one is to distinguish phonological history otherwise obscured by parallel changes that may have occurred among the selected criterion languages.

In this case however, Shorto implemented a novel approach - first he determined his reconstruction based on the binary comparison of Mon and Khmer, and then he applied the results to his wider data set, in a sense testing if it was sufficient. What he found confirmed is expectations - a substantial proportion of regularly reflexes that could be accounted for without difficulty, and a sizable minority of apparently irregular correspondences that did not immediately sit with the preliminary reconstruction.

How did he deal with this? At this point Shorto took a crucial step - he supposed that among the problematic correspondences he could discern regular patterns that suggested an explanation which would allow him to maintain his preliminary model more or less without revision. This patterning was of the following kind: where he may have expected, for example, to see a reflex of ***u**, he instead sometimes saw what appeared to be a reflex of ***uu**, where he expected a reflex of ***uu**, he instead sometimes saw what appeared to be a reflex of ***uə**, and so forth, suggesting a pattern of vowel gradation with PMK along the lines of ***u** > ***uu** > ***uə** > ***ɔɔ**, and similarly for the front vowels. Assuming that there were co-occurring forms of the same etymon with various vowel grades within PMK, reflecting perhaps some ancient morphophonemic processes, one could posit alternate forms or "alternances" as Shorto called them, without needing to posit additional proto-phonemes or complicated sound laws to account for the more problematic correspondences. Consequently when one browses Shorto's dictionary a veritable plethora of alternate reconstructions are noted.

Discussion

The argument that I will mount today is not necessarily a refutation of the underlying premise of vowel gradation in PMK - I don't yet presume to understand PMK vocalism so well - but I will defend the claim that by reanalysing the correspondences on the basis of certain additional language data that was used by Shorto may readily yield a ore reasonable reconstruction. The reanalysis requires only a small change to Shorto's PMK vowel inventory yet eliminates many of the alternate forms. That only a small change in the phonological system could indicate a dramatic simplification of the proto-lexicon is, on the principle of Occam's razor, an advance. To illustrate this I will now discuss several examples, with particular reference to the relationship between ***uə** and ***ɔɔ**.

Referring back to Table 1, you will note the otherwise unremarkable correspondence of Old Mon orthographic **o** to Written Khmer **uə** and **ɔ**, **ɔ̄** (and parallel correspondence of Old Mon orthographic **e** to Written Khmer **iə** and **ɛ**). Shorto interprets this as reflecting as a merger in Mon, while Khmer is archaic. The straightforward consequence is that wherever the Khmer reflex diphthonged, so the PMK reflex is presumed to be.

Now it turns out that when Shorto began assembling MK cognate sets, he did so by first extracting the data assembled by Schmidt (1905), being some 900 comparisons of Mon, Khmer, Bahnar and Stieng, the latter two being closely related within one sub-branch of Bahnaric. Among these Shorto noted that for proportion of etyma for which Khmer has **uu** and **uə**, a goodly proportion of Bahnar and Stieng reflexes show **ɔɔ** (or similar regular reflexes). Applying his theory of alternances, Shorto took this to indicate that Bahnar and Stieng **ɔɔ** reflects a regular development from PMK ***uə** - in some cases directly from a primary PMK ***uə**, in others from an **uə** alternant of PMK ***uu**. A neat example as is seen here [retrieve examples online at htp://sealang.net/monkhmer/dict.htm]:

review item 822 "to spit, transfix"

So confident was Shorto that variously reconstructed PMK *__uə__ to explain correspondences of Old Mon **o** to Bahnar and/or Stieng **ɔɔ** even when a Khmer reflex was lacking, e.g.:

review items 280 "egret", 485 "wasp"

And in some cases when a Mon reflex is missing:

review items 878 "flow"

Parallel considerations also apply to his treatment of *__ii__, *__iə__ such that Bahnar/Stieng etc. **ɛɛ** is frequently treated as a reflex of PMK *__iə__ even in the absence of any diphthonged forms.

review items 731 "forehead", 1010 "lorikeet"

I don't find this state of affairs to be satisfactory. Certainly there are some unambiguous cases of Bahnar having monopthong reflexes of original diphthongs after /r/, e.g.

Proto-Bahnaric		Bahnar	
*__ruat__	>	**rɔt**	"to buy"
*__ruay__	>	**rɔɔy**	"fly"
*__riah__	>	**rəh**	"root"

This is associated with a deeper trend of syllable restructuring in Bahnar involving a shift in stress placement within syllables, e.g.

*__puan__	>	**pwan**	"four"
*__ciam__	>	**hjɛm**	"to feed"

However, there are plenty of cases of Bahnar **ɔɔ** corresponding to **ɔɔ** in other MK branches, corresponding to Old Mon graphic **o**, suggesting PMK *__ɔɔ__:

review items 407 "plant upright", 412 "squirrel", 547 "handle", 1127 "child"

And of course there are examples that point to an opposition with *__uə__, e.g.:

review item 846 "citrus", 1166 "four", 1534 "fly"

Shorto's analysis of the relevant phonological correspondences can be represented as follows:

	Old Mon	Written Khmer	Bahnar	Stieng	PMK (Shorto)
1	**o**	**o**	**ɔɔ**	**ɔɔ**	***ɔɔ**
2	**o**	**uə**	**ɔ(ɔ)**	**ɔɔ**	***uə**
3	**o**	**uə**	**ɔ(ɔ) ~ wa**	**uə**	***uə**

Lines 1 and 3 above are straightforward enough, but line 2 requires further consideration. The question reduces to whether the PMK reconstruction should be ***uə** or ***ɔɔ** (or something else?), depending upon which of Khmer or Bahnaric is the innovator.

In the absence of an obvious conditioning factor it is not enough data here to decide. All other things being equal, it may be suggested that it is as likely that Khmer merged ***uə** and ***ɔɔ** to **uə** as it is that Bahnaric merged ***uə** and ***ɔɔ** to **ɔɔ**. However, not all things are equal.

Shorto's PMK vowel inventory is as follows:

*/	**i**		**u**	**ii**		**uu**	
	e	**ə**	**o**	**ee**	**əə**	**oo**	
		a	**ɔ**		**aa**	**ɔɔ**	
				iə	**[ɯə]**	**uə**	
					ai		/

Note the following asymmetries:

- The complete lack of low from vowels despite the frequent fact of such a contrast in MK languages,
- The imbalance in frequency between Shorto's reconstruction of 365 cases of ***uə** versus only 80 cases of ***ɔɔ**.

If the line 2 correspondences are treated as reflecting ***ɔɔ**, supposing a Khmer merger of ***uə** and ***ɔɔ** to **uə**, and by implication a parallel merger of ***iə** and ***ɛɛ** to **iə**, the revised proto-vocalism would be rather more balanced and consistent with the typology of existing conservative or "unrestructured" MK languages (applying the terminology of Huffman 1985).

There is no need to posit a new back vowel phoneme to account for the line 2 correspondences, and positing a ***ɛɛ** merely fills a rather odd gap in an otherwise more or less normal inventory. More data is required to determine if a specific conditioning environment can be identified for the specific restricted mergers identified for Khmer in this revision. However, a broad survey of of Shorto's ***uə** comparisons identifies that reflexes in other (e.g. Northern MK etc.) languages are more often ***ɔɔ** rather than ***uə**, lending significantly more weight to the revision suggested here (specific counts will be offered in the publication version of this paper).

Conclusion

Shorto most likely erred in only basing his vocalism on the comparison of two languages. In my view, if he had used the four languages as laid out in his most important source (Schmidt 1905), he could have avoided the apparently excessive application of his theory of "Alternances", and offered a more reasonable reconstruction. As it stands the phonological and lexical reconstruction offered by Shorto (2006) is skewed and in serious need of revision. Even within the limits of the data organised and presented by Shorto it is possible to move more or less quickly to address these issues and produce a much more satisfactory account of PMK vocalism.

References

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-asiatic. In: Sagart, Laurent , Roger Blench and Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. Routledge/Curzon. pp79-82.

Huffman, Franklin E. 1985. Vowel Permutations in Austroasiatic Languages. *Linguistics of the Sino-Tibetan Area: The State of the Art.* Pacific Linguistics Series C-No.87. Canberra: Australian National University, pp141-45.

Schmidt, Wilhelm. 1905. "Grundzüge einer Lautlehre der Mon-Khmer-Sprachen." *Denkschrift der Akademie der Wissenschaften, Wien, Philologisch-Historische Klasse* 51:1-233

Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary.* Canberra, Pacific Linguistics 579.

Thomas David. 1964. A survey of Austro-asiatic and Mon-Khmer comparative studies. *Mon-Khmer Studies* 1:149-163.