

# Comparative Mon-Khmer Linguistics in the 20th Century: Where From, Where To?

Paul Sidwell<sup>1</sup>

Center for Research in Computational Linguistics (Bangkok), &  
Australian National University (Canberra)  
<paulsidwell@yahoo.com>

## 0 Abstract

Just over a century since the foundational studies of Wilhelm Schmidt, and the extraordinary effervescence that the field enjoyed in the 1960s and 1970s, we have again reached an important watershed in the history of Mon-Khmer (MK) comparative-historical studies. 2006 saw the publication of Shorto's *A Mon-Khmer Comparative Dictionary* (MKCD), a wealth of extensive descriptive and lexical data is now available, and the International Conference of Austroasiatic Linguistics (ICAAL) meetings are resuming after a hiatus of three-decades.

At this historic juncture it is appropriate to pause and reflect on this history of comparative MK studies, from its foundations in philological and neogrammarian methods and principles, through the field-work driven structuralist-descriptive phase that has characterized the second half of the 20<sup>th</sup> century. I analyze the strengths and weaknesses of this work, drawing attention to programmatic aspects of the approaches taken by various scholars, especially contrasting their use of comparative and philological methods in historical linguistic reconstruction.

Progress in comparative MK has always come in fits and starts, yet prospects have never been better as we launch into this new era with more data, new tools, and a clarity of purpose gained from 20-20 hindsight.

## 1 Introduction

The MK languages form the oldest and most diverse language family of Mainland Southeast Asia. MK speakers colonized all available ecological niches the region has to offer, from mountain slopes to tropical islands. Along the way they diverged widely; some barely subsisted in near-stone-age conditions, while others built great civilizations still recognized as hallmark achievements of mankind. Enriched by their complex history of contact with other Asian societies, the MK languages present a deeply evocative record of the region that can be revealed by comparative-historical analysis.

But the comparative study of languages does not proceed easily. Scholars must have access to extensive data of suitable quality. They need to understand disparate writing systems and diverse historical and cultural contexts, and must learn to recognize the influences of numerous contact languages. The work, particularly in the pre-computer era, is excruciatingly time consuming, so that investigators must have strong institutional support or extensive private means. And since no one researcher can be an authority on all the languages, there must be qualified collaborators willing to offer data and technical advice. The larger and older the language family, the more daunting these tasks become.

This has certainly been the case in regard to the MK languages, where instead of a century of incremental progress we have seen fits and starts, great leaps, dead ends, and numerous frustrations. The 20th century certainly began well with the grand synthesis of

---

<sup>1</sup> The author gratefully acknowledges the support of the Mon-Khmer Languages Project by the National Endowment for the Humanities. Any views, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect those of the NEH. Thanks are also due to Doug Cooper and Tom Sidwell for their helpful comments on earlier drafts of this paper.

comparative MK attempted by Schmidt and discussed below, yet his work suffered from many of these limitations, and no comparable effort would appear within the century.

Instead, as time went on, scholars focused mainly on specific languages or sub-groupings, and rarely considered them in the broader MK historical context. Genuine cooperation among specialists has been uncommon; indeed the story of comparative MK research over the past a hundred years and more has been one of highly motivated, but usually solitary, individuals. Although they achieved much—often at great personal sacrifice—they also faced barriers that hampered progress, and occasionally made serious errors that reduced the usefulness of their results.

## 2 End of the 19th and beginning of 20th Centuries

The main factor leading to the birth of comparative MK studies was the European colonization of SEAsia, as scholars began to have access to increasingly reliable lexical data developed by colonial authorities, businessmen, and missionaries. At first these sources chiefly produced lexicons useful for translation in support of administration, trading, and prostelyzation, then true dictionaries of languages with written traditions (Mon, Khmer and Vietnamese<sup>2</sup>) gradually became available.

Since the mid 1800's various authors had begun to recognize the genetic relations between languages that we now understand constitute the MK family. As improved sources became available the picture gradually became clearer, and the possibility of comparative-historical reconstruction became real, crystallizing around 1900. Among the most important of the early dictionaries and lexicons were:

- Cambodian (Aymonier 1874, Feer 1877, Moura 1878 etc.)
- Mon (Haswell 1874, Stevens 1896 etc.)
- Nicobarese (Man 1889)
- Palaung (Scott & Hardiman 1900)
- Bahnar (Dourisboure 1889)
- Stieng (Azémar 1886)
- Various Aslian (Morgan 1885; Blagden 1894 etc.)
- Numerous lists compiled during the Pavie Expeditions into Indo-China 1879-1895.

We are not surprised that the beginning of the 20th century saw a great burst of activity that placed the comparative understanding of the MK family on a solid footing. Particularly noteworthy were:

- a series of studies by Schmidt (1901, 1904, 1905, 1906)
- a massive Aslian compilation by Skeat and Blagden (1906)
- a broad, well organized comparative vocabulary by Cabaton (1906)
- the Linguistic Survey of India by Grierson (13 vols. from 1903 to 1928)

The most important of these from the analytical perspective were the works of Schmidt. But before going further, though, it is useful to discuss the use and scope of the term "Mon-Khmer" in this context.

Since the first identification had been made by Mason in 1854, it was apparent that the Munda languages of India could and should group with Mon, Khmer etc. in an Austroasiatic super-family. But to Schmidt and others of his time the term Mon-Khmer was only used to

---

<sup>2</sup> The 1651 Vietnamese-Portuguese-Latin dictionary of de Rhodes being an especially early example.

designate Mon, Khmer, and other languages that appeared to be self-evidently close to them, such as Bahnar and Stieng.

When groups like Palaungic and Aslian were subsequently identified as related, they were not necessarily classified as MK. Rather, they were variously treated as forming other distinct Austroasiatic branches. Indeed, it was only in the late 1960s that the now-received model, in which *all* of the non-Munda languages belong to one large MK branch, really emerged. Throughout this paper the term MK is used to refer to the generally received notion of all Austroasiatic languages other than those of the Munda branch(es).

### 3 Pater Wilhelm Schmidt

The German comparativist and ethnologist Pater Wilhelm Schmidt (1868–1954) effectively established the field of comparative MK studies with a series of four major publications at the beginning of the twentieth century (1901, 1904, 1905, 1906). Taken as a whole these four works form a more or less coherent account of the MK family as it was known at the time.

The first of these (1901) is a monograph length (142 pages) paper that examines the Aslian languages of Malaya, demonstrating that they are genetically related to other MK tongues. The study is substantially lexical, including a vocabulary of 1232 entries, plus some samples of texts. The languages are classified into two major subgroups using distinctive vocabulary tests: Semang (Northern Aslian) and Sakei (a Southern group with further subdivisions). Schmidt also achieves a successful morphological analysis.

Schmidt (1904) discusses Khasi and Palaungic, correctly identifying the latter as an important subgroup. Much of the treatment of Khasi is an extensive morphological analysis that identifies numerous prefixes and infixes, seeking to provide a comprehensive account of word formation. Unfortunately Schmidt's analysis goes somewhat too far, since he decided that all Khasi sesquisyllables were formed by affixation of monosyllabic roots. He was partly stimulated to this view by the fact that there is a phonological tendency for the elision of initial consonants from clusters, so that it was not clear to Schmidt that in some cases he was looking at secondary monosyllables rather than etymological monosyllables.

Schmidt then took his idea further, and suggested that the model be applied to MK generally. This seductively simple but ultimately mistaken model would influence some scholars right up to recent times, such is Schmidt's reputation.<sup>3</sup> This idea would also influence his (1906) attempt to relate MK and Austronesian, since it gave some formal bases for comparing word-forms.

The 1904 paper also prefigures to some extent the methods used in his 1905 opus. Schmidt directly compares Khasi vocabulary to Mon, Khmer, Bahnar and Stieng in an attempt to justify various word-formational formulae. In retrospect, not all the comparisons are valid, or in some cases the interpretations are simply too bold. While the underlying notion of using systematic comparative data to analyze Khasi morphology is a good one, Schmidt was not quite in a position to fully appreciate the significance of the data he had before him.

The third paper (Schmidt 1905, often simply referred to as the *Grundzüge* [roughly, 'Foundations'] from the title) is an extensive comparative treatment of Mon, Khmer, Stieng and Bahnar. In it Schmidt set out to establish regular sound correspondences on the basis of an extensive and fairly reliable data set. To make this study Schmidt used Written Mon and Written Khmer, assuming that their Indic-based spellings faithfully recorded historical

---

<sup>3</sup> For example, Peiros (1996) was influenced heavily by Schmidt's ideas in his unsuccessful morphological analysis of Katuic, and under his influence the present writer's South Bahnaric (Sidwell 1999) and West-Bahnaric (Jacq & Sidwell 2000) reconstructions followed suit.

phonetic values, plus contemporary lexicons of Bahnar and Stieng (Dourisbourne 1889 and Azémar 1886, each recorded in Latin script).

In choosing to work with the two languages that had long written traditions Schmidt followed well-established philological methods, as the comparative investigation of Indo-European had proceeded principally by using Latin, Greek and Sanskrit. He was also correct in assuming that the writing systems of these languages would be conservative, since both Mon and Khmer underwent extensive phonological restructuring after their writing conventions became established.

On the other hand, in some respects Schmidt erred in following the Indic spellings too slavishly, as they include some non-etymological and non-phonetic information, such as use of final voiced stops and retroflex consonants. Yet Schmidt was also well aware of Pali/Sanskrit loans in Mon and Khmer, and discussed many of them. He also concluded (wrongly it turns out) that some MK etyma must be very ancient Indic loans (such as PMK 'water' \**daak* < Sanskrit *udaka*).

Notwithstanding these problems, Schmidt successfully determined the basic outline of Proto-MK consonantism. He was also able to offer morphological analyses that related affixes to various phonetic changes, and he established the importance of the patterning of segmental collocations within rhymes—insights that have been the foundation of successful phonological reconstruction of MK languages to this day.

Schmidt encountered his most serious difficulty in his treatment of the vowels. The biggest problem was his handling of long and short /a/. These are the most common vowels, and occur in some 45% of his comparisons (and, generally, in around one-third of all MK etyma). Schmidt was apparently unaware that Mon and Khmer use similar spellings to represent different vowel phonemes, and consequently he erred in consistently transcribing /a/ and /a:/ according to Indic readings, losing their etymological values, and fundamentally undermining his analysis of the historical systems. The apparent confusion of vowel correspondences created such difficulty for Schmidt that he was only able to posit phonetic equations, and not proper vowel reconstructions.

Despite its various difficulties, the *Grundzüge*, with more than 900 MK lexical comparisons, reconstruction of proto-consonantism, and morphological analyses, laid the foundation for all subsequent comparative work. It is most obviously evident in Shorto's (2006) posthumous MK comparative dictionary, where it is clear that the *Grundzüge* provided the skeleton upon which Shorto more-or-less directly built his edifice.

Schmidt's ultimate work in this series was his (1906) attempt to link MK and Austronesian within a grand Austric macro-family. This bold hypothesis remains controversial. Although it faces serious difficulties in the light of Shutler and Marck's (1975) now generally accepted formulation of a Formosan homeland for Austronesian,<sup>4</sup> it still competes with other models of deep genetic relationship among the various language families of Asia (see Sagart et al. 2005 for recent papers and discussion). Schmidt's evidence was both lexical (more than 200 comparisons) and morphological (including parallels in prefixes and infixes), and has found a small and enthusiastic following, including Shorto, and nowadays notably Laurence Reid (e.g. 1994, 1996, 2005).

The most concrete results to come out of Schmidt (1906) is the discussion of the internal classification of MK languages. It posits three branches as follows:

---

<sup>4</sup> That paper was written specifically to refute Dyen's (1965) classification that suggested a Micronesian homeland for Austronesian. By implication it also refuted the idea, expressed by Shorto from time to time, of an Indo-Chinese homeland, which would have reconciled more easily with the Austric hypothesis.

1.    a) Semang  
      b) Senoi
  
2.    a) Khasi  
      b) Nicobarese  
      c) Wa, Palaung, Riang
  
3.    a) Mon-Khmer (Mon, Khmer, Bahnar, Stieng etc.)  
      b) Munda  
      c) Cham, Rade, Sedang

Group 1 represents the Aslian languages of Malaysia. In group 2, entries a) and c) are Northern MK in today's formulation, with b) Nicobarese being anomalous in that group. In retrospect, group 3 is the most heterogeneous: Munda is arguably separate from all of the languages above, occupying a different Austroasiatic branch, while Cham and Rade are nowadays recognized as Austronesian, although with a large borrowed MK component to their lexicon. Schmidt also included Sedang in this group, a Bahnaric language that has enjoyed considerable Cham and Khmer influence (although perhaps not as much as Bahnar itself).

The above scheme is not as strange as it might first appear to modern eyes. Removing the most serious anomalies in 3 a) and b) we are left with three main branches that resemble in many important respects the three-branch classification of Diffloth (1989 and *passim*.) that now has a more-or-less received status among specialists. The essential difference is that Diffloth would place Nicobarese and Mon into the Group 1 (or Southern) branch alongside Semang and Senoi (Aslian). Neither Diffloth nor any other writer has, to my knowledge, offered in print a detailed proof of this classification (or disproof of any other) based upon a comprehensive comparative reconstruction of the breakup of Proto-MK. Indeed, the most important works of the last century on the issue of MK classification have been based upon lexicostatistics, a method that is generally recognized as only suitable for preliminary investigations.

#### **4 Other early 20th Century works**

Contemporary with Schmidt was Skeat and Blagden's (1906) massive comparative Aslian lexicon. It was richly annotated with wider MK comparisons, letting it serve effectively as an etymological dictionary. It is as a useful complement to Schmidt's work, and has been relied upon by comparativists in the decades since. However, Skeat and Blagden did not use their lexical materials within the framework of comparative reconstruction or the genetic theory of language as we know it. Instead, they focused on what they saw as lexical differences, for example a list of 50 Semang words they saw as having no MK parallels, arguing for waves of migration causing partial language shifts among earlier Negrito (non-MK) populations on the peninsula.

Notwithstanding their diffusionist views, Skeat and Blagden did advance our understanding of Aslian beyond the first simple classification offered by Schmidt (1901). They fleshed out the basic North-South divisions, employing both lexical and phonological criteria. Their scheme was later improved by Schebesta's (1926) study, which established the essential classification that is used until today (for a further discussions see Benjamin 1976 and Matisoff 2003).

Another important resource that became available during this highly productive period was the Linguistic Survey of India, with Volume 2 "Mon-Khmer and Siamese-Chinese families" and Volume 4 "Mundā and Dravidian" (Grierson 1904, 1906 respectively) giving

useful lexicons and descriptive information. Contemporaneously, a well-organized data collection that proved influential was Cabaton's 1905 presentation. In this paper he gives some 416 cognate sets for 28 languages, including many small languages of the Annamite Range for which no better sources would be available until the second half of the 20th century. In addition to his analysis of lexical data, Cabaton discusses the phonological systems of Cham, Khmer, Malay, Bahnar, Chrau and Stieng.

However, despite ever-increasing access to data, and a solid foundation analysis, comparative MK studies fell into a prolonged lull. The central difficulty was the lack of a critical mass of scholars interested in pursuing comparative studies in Southeast Asia.

At the time diffusionist views were common (pace Skeat & Blagden 1906) as the wave model of linguistic change, popularized by Johannes Schmidt (1872 etc.), seriously challenged the genetic theory of the Leipzig neogrammarians (a rivalry that periodically resurfaces among scholars who doubt the reality of language families as objects of study). In this context it is not surprising that Wilhelm Schmidt's work attracted severe criticism from some quarters; here is Thomas' tidy summation:

Georges Maspero was very skeptical of the work of Schmidt, and of Schmidt's immediate predecessors. His works on Thai (1911)<sup>5</sup>, Vietnamese and Thai (1920)<sup>6</sup> and Khmer (1915), made his position quite plain. He roundly condemned the 'Mon-Annam' family of Logan, Forbes, Muller, and Kuhn, which included Mon, Khmer, Vietnamese, Cham, etc. Maspero granted Mon-Khmer and Palaungic as being related, but nothing else certainly related.

He felt that the work of Skeat and Blagden had removed the Senoi and perhaps also the Semang from Austroasiatic; if Grierson's report was correct that Khasi had tones, then Khasi was automatically eliminated; and he felt Munda's relationship to Austroasiatic still unproved despite the efforts of Schmidt and Konow. Maspero's skepticism was healthy, as the 'proofs' that had been adduced previously were far from being full proofs.

His denial of Khasi, however, came from his a priori supposition (based on his assertion of a Thai—Vietnamese relationship) that Mon-Khmer languages cannot have tones.

(Thomas 1964:154-155)

With an authority of Maspero's stature to contend with, it would take a brave scholar indeed to plow ahead with overt comparative work within the framework that had been established by Schmidt. Quite predictably, specialists and commentators began to line up against the Austroasiatic hypothesis in its various forms. Perhaps the most important of these was an influential 1942 paper by Sebeok. In it, Sebeok asked "But is there an Austroasiatic sub-group at all?" (p.211), then responded to his own rhetorical question by citing the negative views of Maspero, whom he characterized as "the best authority" on the languages.

#### **4.1 Two transitional figures**

Despite these negative views a solid line of inquiry based on the epigraphic/philological tradition continued at SOAS (London). It was this stream that would keep comparative MK studies alive until an entirely new tradition, based upon fieldwork on living languages, emerged in the 1960s. The key transitional figures were Charles Otto Blagden (1864-1948), and Gordon Luce (1889-1979).

Working at SOAS well into the 20<sup>th</sup> century as Professor of Malay, Blagden had moved on from his earlier work on Aslian to develop a considerable interest in Mon, and had

---

<sup>5</sup> Not listed in Thomas' bibliography.

<sup>6</sup> Not listed in Thomas' bibliography.

prepared a preliminary etymological dictionary on index cards by 1928. Never completed, Blagden's notes informed various research publications, and five publications (four fascicles and a volume of plates) of the *Epigraphia Birmanica* between 1923 and 1936. These notes were eventually passed on to Luce.<sup>7</sup>

An indefatigable field worker, Luce compiled thousands of pages of notes and wordlists for the languages of Burma.<sup>8</sup> Luce clearly meant to carry through Blagden's dream of a Mon etymological dictionary. He expanded and corrected Blagden's corpus of transcribed Mon texts, developed a set of lexical comparisons, and also improved the morphological analyses of the language. Yet Luce's principal linguistic passion was for Burmese and its relatives, so he ultimately passed on the torch of Mon studies to Shorto, who met the challenge admirably.

It is important to consider that despite Luce's enthusiasm for field data, his historical approach was fundamentally philological, favoring languages with long written traditions (as well as architecture and material culture) for their capacity to offer direct witness to history. This comes through strongly in his most important publications, including the three-volume *Old Burma-Early Pagan* (1969-70), the posthumously published two-volume Paris lectures *Phases of Pre-Pagan Burma* (1985) and his *A Comparative Word List of Old Burmese, Chinese and Tibetan* (1981) (the latter having grown out of research for his extensive unpublished draft dictionary of Old Burmese). The role of minor living languages in Luce's method was always a supporting one, helping to clarify and contextualize the histories of the major languages and the cultures they reflect.

## 5 The rebirth of comparative Mon-Khmer linguistics

The mid-20<sup>th</sup> century saw a rebirth of interest in comparative MK studies. Haudricourt (1952, 1953, 1954) breathed new life into the field with his proof that Vietnamese tones could be explained by direct reference to MK etymology, rather than to Thai or Chinese. Haudricourt is the real hero of the era; his studies produced a powerful demonstration of the insight into phonology and proto-history that could be offered by the comparative method. Maspero's Chinese Wall of doubt tumbled, setting the stage for the return of comparative MK reconstruction to academic respectability.

Yet there was still a major hurdle that would not be conquered in the 1950's. In the vast linguistic hotbed of Mainland Southeast Asia, language groups were not necessarily self-evident, and the state of classification was dismal. Aslian and Palaungic were recognized as distinct complex branches, but their real extents were unknown. Nor was there any clear understanding that the scores of minor languages found in the region that stretches from central Cambodia through Isaan and along the Annamite range, up into Yunnan and Sipsongbanna, would eventually be resolved into Bahnaric, Katuic, Pearic, Vietic, and Khmuic. Absent a real delineation of sub-groupings comparative studies floundered.

A way out of the morass came with the application of lexicostatistical methods. Although roundly condemned in many quarters as invalid (especially since Bergsland & Vogt 1962; see also ten pages of vicious denunciation in Campbell 1998), lexicostatistics has enjoyed considerable success in Southeast Asia as a useful classification heuristic.

Preliminary private studies were followed in print by Thomas (1966) and Thomas and Headley (1970). These effectively established the pattern of MK sub-groupings, supported by

<sup>7</sup> It is likely that those index cards were eventually entrusted to Harry Shorto, but were lost when SOAS cleaned out his office after his retirement in the 1980s.

<sup>8</sup> The Luce Papers are now held in the manuscripts collection of the Australian National Library. The Mon-Khmer and Sino-Tibetan lexical materials have been scanned by CRCL and indexed, and are available on-line at <http://archives.sealang.net/luce>.

further studies of Huffman (1976) and Smith (1981). The major early result was that the Bahnaric and Katuic families, which together account for perhaps forty percent of the diversity among MK languages, were correctly delineated, as were some of their most important sub-groupings.

1959 was an especially good year for comparative studies with the appearance of Heinz-Jürgen Pinnow's *Versuch einer historischen Lautlehre der Kharia-Sprache*, as well as the publication of the first volume of the huge, multi-dialectal Bahnar dictionary of Guilleminet and Alberty (1959, 1963).

Pinnow's understated title suggests a tentative historical phonology of Kharia, a Munda language of India. In a more accurate view, in 514 pages of dense text Pinnow attempted to build nothing less than an Austroasiatic etymological dictionary, with some 550 etyma over a representative set of languages. Issues of Kharian phonological evolution were handled in the context of a preliminary Proto-Munda and Proto-Austroasiatic reconstruction. More than 400 cognate sets supported the proto-vocalism, and more than 500 supported the consonantism.

Pinnow made extensive MK comparisons, with the implication that regular Munda : MK correspondences would reliably establish and reconstruct ancient etyma within the MK family. But Pinnow's results were hampered by a lack of data, and by problems of interpreting sources. His book's ultimate impact on the field was not nearly commensurate with the effort that had gone into it (see the witty review by Shafer 1960).

Two other major comparative MK studies also appear at this time: Shafer's *Études sur l'Austroasien/Studies in Austroasian* in 2 parts (1952 and 1965). But like Pinnow's work, these studies did not deliver extensive or useful Proto-MK reconstructions, and thus really did not advance things much beyond the pioneering framework advanced by Schmidt a half a century before.

Nevertheless, the spark had been lit, and as the 1960s and '70s progressed MK research blossomed. A generation of young scholars (many associated with the American based *Summer Institute of Linguistics*) traveled to Indo-China, collected data on many previously little-known and undocumented languages, and began preparing various sub-group-level reconstructions. Significantly many of this new generation were schooled in the Bloomfieldian structural-anthropological approach to linguistics that emerged so strongly, especially in the English speaking countries, post Bloomfield (1930). Among the SIL tendency this morphed into the tagmemic tradition developed by Kenneth Pike, such that a significant stream of SEAsia orientated field linguistics were by-passed by the Chomskyan revolution and its underlying hostility to empirical and comparative-historical methods. Showing extraordinary foresight, in 1964 this same tendency founded the journal *Mon-Khmer Studies* in Saigon (nowadays based at Mahidol University, Thailand).

European interest was also rekindled at this time. For example Michel Ferlus (CNRS, France) began his decades-long commitment to data collection and historical analysis of various languages of Thailand, Laos and Vietnam. Scandinavian scholars such as Kristina Lindell, Jørgen Rischel, Søren Egerod became active, working mainly on languages accessible from Northern Thailand. And in Thailand among local linguistics an admirable tradition of descriptive studies developed, particularly at Mahidol, Chulalongkorn and Thammasat Universities.

Yet it was among the young American trained field workers that an important methodological shift in comparative studies begins to become apparent. Philological methods were effectively eschewed in favor of a purely field data based approach to comparative reconstruction. Not just driven by the increased availability of field data, it reflected the underlying Bloomfieldian biases of the researchers, becoming an explicit programmatic commitment most highly developed in the work of Gerard Diffloth through the 1980s



(discussed below). Yet others would continue with a more European approach to historical linguistics, seeking to get the most from both old and new methods, not necessarily ruling out any particular categories of data (especially, e.g. Michel Ferlus).

The surge in worldwide interest in MK and Austroasiatic linguistics stimulated a major international conference at the University of Hawai'i in 1973 (the first ICAAL); out of which a substantial two-volume set of proceedings was published in 1976 (Jenner et al.). A second ICAAL was held at Mysore (India) that same year.

However, the times were both a highpoint and a turning point for the field. The dramatic political changes in SEAsia after 1975 saw Western field workers locked out of Indo-China, while at home there was a general decline in support for academic work related to the region. There would be no more MK/Austroasiatic themed conferences and instead such work as continued was presented at meetings primarily devoted to other language families, or at conferences of a general SEAsian theme. The tendency for collaborative work also declined, evidenced in the lack of jointly authored comparative studies between the late 1970s and 1990s.

Thus the comparative work that continued to be published after this time, although more developed, was the product of a narrowing research community, one which lacked an organizational or programmatic unity.

## **6 Summary review of Mon-Khmer comparative-historical reconstruction in the second half of the 20th Century**

In the pages that immediately follow I review branch by branch the development of comparative-historical studies, covering as far as possible the whole MK family from the mid-20th century up to more or less the present day. The survey is reasonably comprehensive, although I do not pretend that it is complete.

### **6.1 Bahnaric**

The Bahnaric family is arguably the most extensive and diverse sub-division of MK, with more than 30 distinct languages spread over three branches.<sup>9</sup> Researchers associated with SIL began comparative work in the mid-1960s, and by the 1970s well-respected studies appeared in print. However, despite continued interest and a substantial body of emerging literature, only subgroup reconstructions have been published. No adequate Proto-Bahnaric reconstruction has been presented, despite the efforts of many researchers (including the present writer). Proto-Bahnaric is likely to be an especially difficult nut to crack: the family is split into several distinct language contact areas of great antiquity, and its multiple layers of borrowed vocabulary do not easily yield to analysis.

#### ***Blood (1966)***

Henry Blood produced a reconstruction of Proto-Muong and a preliminary Proto-South Bahnaric with his 1966 MA thesis (SIL has re-released this work, and it is sometimes dated to 1968 or 1976). He also has the unfortunate distinction of dying as a result of his devotion to research; Blood (along with other SIL researchers and missionaries) was captured by North Vietnamese troops around the time of the 1968 Tet Offensive, and ultimately perished during a forced march to the north under horrific conditions.

Ostensibly Blood's study is an attempt to deal with the issues raised by Thomas (1964), drawing our attention to the following quote:

---

<sup>9</sup> Palaungic may be comparable in diversity if, as appears likely, the Mangic/Pakanic languages of Vietnam and China belong to the Waic sub-group (Nguyễn Văn Lợi, pers. com. 8/10/07).

Still another difficulty, one that besets all Mon-Khmer comparativists, is the complexity of the vowel shifting that has taken place in Mon-Khmer, making it very difficult to establish regular patterns. Schmidt, after a careful survey of the situation, had to content himself with just statements about possible general trends, establishing no sound-laws. Other comparativists have stated flatly that regular sound-laws simply do not exist in Mon-Khmer vowels, and, indeed, no one has yet succeeded (in print, anyway) in establishing a regular pattern in Mon-Khmer vowel comparisons.  
(Thomas 1964:160-161)

Using data from a half-dozen Mnong dialects, Blood assembled 428 cognate sets, then matched these with comparisons from Chrau, Stieng and Sre where available.

Blood's reconstruction proved the possibility of establishing regular vowel correspondences, provided the data are recorded in accurate phonemic (or at least broad phonetic) notation, and the sub-groupings well understood. This was done by direct field methods, using data collected by himself and trusted colleagues, in stark contrast to earlier studies by Shafer, Schmidt, Pinnow, and others who worked from secondary sources. Thus, Blood's great contribution went well beyond his own specialty: his publication helped to launch a new tradition of empiricism in Southeast Asian comparative linguistics that set the agenda for the next 40 years.

### ***Thomas & Smith (1967)***

David Thomas and Marilyn Smith also explicitly took on the problem of vowel reconstruction in MK languages. They attempted a very low-level reconstruction based on a straightforward binary comparison of two North Bahnaric languages that are close enough to form a self-evident sub-grouping: Jeh-Halang. In their own words:

... the study is significant in that it demonstrates the possibility of reconstruction of vowels in Mon-Khmer language comparisons, provided one starts with languages closely enough related and has data on dialect trends; ...  
(Thomas & Smith 1967:175)

However, I would rather see the significance of the study in the fact that the investigators continued the pattern begun by Henry Blood. Thomas and Smith both had direct experience working with related languages, and for their study used recent field data collected by trusted colleagues in circumstances that they understood.

The actual phonological reconstruction offered in the paper is not so significant. It does not address any of the real problems of North Bahnaric phonology, in particular the nature of the register system. The methodology employed simply posited no vowel change when forms were in agreement, and assigned an arbitrary form when they were not, without reference to a general theory of vocalic development.

### ***Smith (1967/1972)***

At the same time that David Thomas and Marilyn Smith were collaborating, Marylin's husband Kenneth was working on Sedang dialects. In addition to much fine descriptive work, Smith produced a draft reconstruction under the initial title *A Phonological Reconstruction of Proto Central North Bahnaric* (a carbon copy is available for inspection at the David Thomas Library, Bangkok), and later (1972) published a revised version as a short monograph in the SIL *Language Data* series as *A Phonological Reconstruction of Proto-North-Bahnaric*.

The study combines the fieldwork-driven empiricism of the SIL milieu of the day with a proper, recognizable, comparative reconstruction methodology. Data are regularized into

comparable phonemic representations, languages are considered within a theory of their genetic relations, and a serious theory of the evolution of the vowel systems is elaborated within the context of proper distributional statements of segmental collocations. Formal correspondences are established for Bahnar, Proto-Jeh-Halang, Proto-Hre-Sedang, Hre, and internally reconstructed Early-Sedang.

Smith explicitly elaborated the key methodological breakthrough that set the stage for solving the otherwise intractable vowel problem that had bedeviled comparative MK studies since the time of Schmidt:

It has been found both here and in other studies of the Mon-Khmer languages of South-Vietnam that the vowels are difficult to determine apart from the final consonant. For this reason many analysts construct "rhyming dictionaries" during their initial phonological study. Some final consonants occur with only a limited number of vowels; few, if any, occur with all vowels (Thomas 1966). In this paper all reconstructed vowels and final consonants are brought together so that the overall effect on the vowels by a given final consonant can be clearly seen. (Smith 1972:4)

As Smith notes, SIL field investigators tasked with creating dictionaries and phonemic writing systems had turned to the Tang Dynasty Chinese innovation of the rhyming dictionary, which solves the problem of the distributional collocations of vowels and finals.

Smith extended the application of these methods from synchronic description to comparative-historical analysis, and thus demonstrated in principle and practice a solution to the MK "vowel problem". Concretely, Smith presented 571 etymologies and proto-forms, listed in blocks according to their rhyme correspondences.

The reconstruction is, thus, completely transparent, with the regularity and reliability of all correspondences immediately accessible and open to challenge. This is Smith's achievement, and is the basis of his enduring contribution to comparative MK studies: not the specific results of his reconstructions, but rather his methodological exemplar.

Interestingly Smith erred in including Bahnar within North Bahnaric, a possibility that he rightly acknowledged himself (on page 11). And in this respect Smith made the most serious mistake in his study. In treating Proto-North-Bahnaric as a register language (probably correct), he assumed that Bahnar lacked register because the feature had been lost. He did not consider the complementary possibility that register was an innovation of the North Bahnaric group. But this error can be readily forgiven. Unlike Khmer, Mon, West Katuic, etc., the North Bahnaric registers do not have a transparent correlation with initial voicing, and thus appear to be archaic in the absence of another explanation for their origin.

### ***The 1970s and 1980s***

After Smith's (1967/1972) breakthrough study, Gregerson, Smith & Thomas (1976) offered a substantial paper looking into the place of Bahnar within Bahnaric, coming to the conclusion that Bahnar belongs in a Central Bahnaric sub-group, distinct from either North Bahnaric or South Bahnaric. In that same paper Gregerson, Smith & Thomas list a bibliographical reference to a draft PhD *A Reconstruction of Proto-South-Bahnaric* by Richard Phillips (no date, and thus far unseen by this writer).

The SIL archives hold several minor unpublished comparative Bahnaric studies from this period. Sheldon's (1979) 62-page essay *A Reconstruction of Proto-South Bahnaric* applies basic computational techniques to the sorting of 224 cognate sets. Leitch's (1981) 32-page essay essentially fits Rengao data into the framework established by Smith (1967/1972), listing Rengao cognates for 561 sets.

***Efimov (1990)***

Efimov (1990, in Russian) is a monograph-length reconstruction of Proto-South Bahnaric that presents 675 etymologies, including some discussion of wider MK comparisons. The most remarkable fact concerning this study is that it is completely out of character with his earlier dissertation on Proto-Katuic (1983), which was more or less a model of reasonable reconstruction methodology.

For Proto-South-Bahnaric Efimov posits a complicated historical phonology, unlike anything attested in the known typology of the daughter languages. The result is a proto-vocalism of 31 members characterized by four degrees of aperture, and an overly large consonantal inventory. This is achieved by ascribing a distinct proto-phoneme to each phonological correspondence without attempting to make distributional analyses of their complementary patterning, more or less as Peiros (1996) would later do for Proto-Katuic. Similarly Efimov reconstructs no final velar stop, instead all examples of **-k** are derived from **\*-ʔ**, also seen in Peiros (1996). Peiros is named as one of the editorial board who approved the publication, and would seem to have exercised some influence in its writing.

***Phraya Prachakij-karacak (1995)***

This is a translated and annotated edition of a study originally published in Thai. It includes wordlists for various West and Central Bahnaric languages, and for decades was the only published source for some of these. The translators, David Thomas and Sophana Srichampa, transcribe the wordlists from Thai characters into IPA, and offer several hundred impressionistic Proto-West-Bahnaric reconstructions. The volume makes an important historiographic contribution, but the data is limited in its usefulness by a lack of certainty as to the phonetic representations.

***Sidwell (1998)***

Sidwell's PhD thesis (University of Melbourne, supervisor Ilia Peiros) is a reconstruction of Proto-Bahnaric based on a comparison of five criterion languages: Sedang, Jeh, Bahnar, Stieng and Chrau. 780 etymologies and reconstructions are compiled. Data from up to 25 other Bahnaric languages and nine other MK branches are included in the comparative lexicon, but not used in the formal reconstruction.

Methodologically the thesis attempts to make a very specific point concerning the combined application of internal and comparative reconstruction methods, using Bahnaric data for the demonstration. The criterion languages are first subjected to internal reconstruction to produce strictly morphophonemic representations of the data, and are then subjected to comparative reconstruction. The analysis included the faulty model of monosyllabic root structure inherited from Schmidt (1904 etc.) and advocated by Peiros (1996).

In retrospect the thesis was not successful: the model of word-structure did not allow for correct analysis of initial clusters, the internal analysis of vocalism conflated a number of vowel oppositions, and it was not recognized that the chosen criterion languages had all been subject to strong Chamic influence (which may have been avoided if West Bahnaric data had been analyzed). In retrospect these weaknesses may have been minimized if a broad consultation with relevant scholars had been pursued, rather than narrowly sticking to advice from one principal supervisor.

**Sidwell (2000)**

The South Bahnaric component of Sidwell's PhD thesis (1998) was extracted, augmented and reanalyzed to provide the text for a monograph-length study, *Proto-South-Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*, published by Pacific Linguistics.

The study compares internally reconstructed Stieng, Chrau and Sre (Koho) with Central Mnong, East Mnong, and Ma, compiling 829 etymologies and reconstructions. It is somewhat more successful than Sidwell (1998), but it still suffers especially from lack of an appropriate morphological model, and a failure to appreciate the role of Chamic lexical influence in the history of South Bahnaric (and Bahnaric more generally).

**Jacq & Sidwell (2000)**

Recognizing the severe problems caused by failure to appreciate the West Bahnaric languages in Sidwell (1998), Jacq and Sidwell in 1998 began several years of joint fieldwork in the Lao PDR, which ultimately contributed to 2 important projects, a reconstruction of Proto-West-Bahnaric and a descriptive grammar of Loven/Jruq (the latter as an MA Thesis at the Australian National University Jacq completed in 2001).

*A Comparative West Bahnaric Dictionary* (Jacq & Sidwell 2000) is the first of two attempts by these collaborators to produce a comprehensive Proto-West-Bahnaric reconstruction. 1023 West-Bahnaric etymologies and reconstructions are presented. Regrettably the results of the project were compromised by several factors: the monosyllabic root theory was applied leading to a misanalysis of initial clusters, an overly phonemic approach was applied which oversimplified the proto-vocalism, and a defective lexicostatistical matrix used in the discussion of classification. Fortunately these issues were addressed in the follow-up study by Sidwell and Jacq (2003).

**Theraphan L-Thongkum (2001)**

Chapters Five and Six of this substantial study deal with the Bahnaric family, and offer reconstructions for Proto-Bahnaric, Proto-West-Bahnaric and Proto-North-West-Bahnaric.

Chapter Five discusses Bahnaric internal classification at length, including both lexical and phonological criteria, and divides the family into five branches: South-, Central-, North-West-, North- and West-Bahnaric. Phonological inventories and morphological systems of West- and North-West-Bahnaric languages are presented and discussed. The basis of the phonological reconstructions is briefly discussed, although tabled correspondences are only given for the diphthongs (principally to deal with the defective treatment of diphthongs in Jacq & Sidwell 2000).

Chapter Six begins with a list of 262 Proto-Bahnaric reconstructions, followed by presentation of approximately 2,300 West- and North-West-Bahnaric etymologies and reconstructions. As with the earlier chapters on Katuic in the same volume, the data are organized semantically. The reconstructions lack explicit correspondences justifying their formulations, but the primary data itself is of the highest quality and extremely useful.

The most problematic aspect of the study is the reconstruction of a North-West-Bahnaric sub-group proposed to consist of Kasseng, Tariang, Yaeh and Alak. Theraphan's theory has received some support, and Diffloth (2005) included the North-West-Bahnaric branch in his most recent classification (for a review of Bahnaric classification see Sidwell 2002).

In my alternative analysis North-West-Bahnaric is not a genetic grouping as such. Rather, Kasseng (a Lao designation lumping various peoples of the Upper Sekong Valley) and Tariang are one language, properly called Tariang, which may be classified as West-Central-Bahnaric. Yaeh is a dialect of Jeh (North Bahnaric) spoken in Laos, and finally Alak forms a separate branch that I call North-Central-Bahnaric.

***Sidwell & Jacq (2003)***

*A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric* (Pacific Linguistics 2003) reflects new fieldwork data obtained and analyzed in order to address the problems of Jacq & Sidwell (2000). Much of the new analysis followed on from wide ranging discussions with relevant scholars, especially Michel Ferlus.

The study compiles 1094 etymologies and reconstructions. The phonological history of the sub-group is more or less completely worked out, and includes a discussion of the effects of areal contact with Katuic, which manifests as traces of a glottalization feature in the phonology. The internal classification of the West Bahnaric languages is determined on the basis of comparative phonology. This generates a model of the West Bahnaric homeland location, and the path of splits and migrations leading to the present distribution of daughter languages. The full extent of the West Bahnaric languages is established, with maps and demographic data given for each. The book is intended to be the first of a short series on the Bahnaric family culminating in a reconstruction of Proto-Bahnaric to appropriately supersede the model offered by Sidwell (1998).

**6.2 Katuic**

The Katuic languages have been subject of several articles and book length reconstructions since the 1960s. We can now say that the history of the group is quite well understood, and the proto-lexicon is extensively documented.

***Thomas (1967/1976)***

The first attempt at a Katuic comparative reconstruction is Dorothy ('Dot') Thomas's master's thesis, submitted to University of North Dakota in August 1967, and published as a Summer Institute of Linguistics Workpaper in 1976. I was fortunate to obtain a duplicate of the original carbon copies from the David Thomas Library, Bangkok.

Thomas characterizes the work as a reconstruction of Proto-East-Katuic. Three languages are compared: Bru, Pacoh and Katu (note that a contemporary description of Katuic may list more than 20 named languages). These are treated as East-Katuic in a default classification, in contrast to a western sub-branch consisting of Kui (as its diverse dialects). Nowadays we would treat Bru as sub-grouping with Kui, so that technically the languages Thomas used are sufficient to reconstruct Proto Katuic. Using then-unpublished wordlists, Thomas assembled 667 etymologies. Comparisons with non-Katuic languages were not made.

Thomas' phonological reconstruction was rather complex, positing four series of initial stops, 22 monophthongs and 14 diphthongs. The analysis of vocalism effectively treated the Bru system as conservative, and derived the Pacoh and Katu vowels by various shifts and mergers. Vowel registers were not analyzed, but merely left as an open question.

Despite subsequent work by other researchers, Thomas' reconstruction remained influential well into the 1990s. It significantly informed Peiros (1996) and was used by Thurgood (1999) in his analysis of loans in Chamic. The analysis of the initials was not taken up by later researchers, as it has both typological and etymological difficulties. External comparisons show that Thomas' aspirated series correspond to EMK plain voiced stops, while her voiced and imploded series both reflect a single EMK imploded series.

In retrospect Thomas' main error lay in neglecting the analysis of registers, so that she did not connect this important aspect of the vocalism with the historical voicing structure of the consonants. This oversight can be forgiven, however, if we recognize that at the time, philological analysis of Mon and Khmer was not yet widely discussed or understood among Southeast Asian linguists whose primary focus was fieldwork (and brings all the more credit to Ferlus, below).

**Ferlus (1974)**

This paper discusses the Katuic language Ong, analyzing its historical phonology, and offering a model for the development of its glottalized registers. Although the Ong registers are structurally different from the Mon and Khmer types, Ferlus attempts to relate them by extending the general model of register formation derived from the philological tradition. Here we see the beginning of an important subsequent characteristic of the research of Ferlus: applying insights from the history of written languages to the interpretation of field data of minor languages.

**Diffloth (1982)**

Diffloth's 1982 essay is an excellent analysis of the problems of Katuic historical phonology. It specifically focuses on the different paths of register development vis-à-vis West Katuic and Pacoh, and offers a summary Proto Katuic reconstruction supported by 138 numbered etymologies.

The paper is not an attempt at a comprehensive account of the history of the family. Rather it gives two main findings. On the one hand, the phonetic history of Kui and Bru parallels that of Khmer in many respects, with a 'normal' path of devoicing, register formation, and vowel splitting. On the other hand, Pacoh is a "registrogénèse hérétique", the problem being that one cannot connect its registers with consonant devoicing. Diffloth's solution to the Pacoh issue is both elegant and insightful, correlating register with various vowel shifts and the consequent phonemization of voice quality features related to vowel height/openness.

Beyond Pacoh, the great contribution of Diffloth's paper to theory is to clearly establish that there is no single universal process underlying the formation of vowel registers in MK languages. Concretely he sketched out the phonemic system of Proto Katuic, which is only moderately affected by later work; I would offer a slightly different analysis of phonetic values of some proto-vowels, and the number of proto-vowels involved (see Sidwell 2005).

**Efimov (1983)**

Efimov's *Kandidat* thesis (in Russian) was completed in Moscow; recently Alexander Kassian kindly sent me a scan of the carbon copy held in the library of the *Far Eastern Institute*.

Efimov worked at the same time as Diffloth and consequently did not have the benefit of the latter's essay. Efimov's thesis makes a sincere attempt at a comprehensive account of Katuic historical phonology, and is a remarkably good piece of work given the limited data to which he had access. Efimov worked with five criterion languages: Katu, Suoei, Bru, Pacoh and Kui. Only Katu is represented by an extensive source—Costello's 1971 multi-dialectal dictionary. The other languages were represented by fragmentary sources, mainly taken from journal articles. In this context Efimov assembled 407 etymologies, and presented a reconstruction of Proto-Katuic vocalism and consonantism.

Given his data set, the results he presents are very reasonable. Efimov correctly identified the analytical errors in Thomas' 1967 thesis, showing that three series of initial stops are adequate to account for shifts in the consonantism, and he offered a model of vocalism that respectably takes Katu as the historical model. Some of the credit for Efimov's good results rests on his very appropriate use of several papers by Ferlus (specifically Ferlus 1971, 1974, 1977).

**Gainey (1985)**

Jerry Gainey wrote his MA thesis, *A Comparative Study of Kui, Bruu and So Phonology from a Genetic Point of View*, at Chulalongkorn University under the supervision of Theraphan L-Thongkum. The thesis ostensibly investigates the genetic relations between the three West Katuic languages, but achieves much more than this. Gainey's methodology involves the identification of shared phonological innovations, so there is much attention to working out details of the phonetic evolution of West Katuic. The study is crowned with a comparative lexicon of 570 items between Kui, Bru and So.

For whatever reason Gainey declined to take the next logical step and formally present his work as a reconstruction of Proto-West-Katuic. Nevertheless, this is essentially the resultant effect of this under-appreciated thesis.

**Shorto (no date, manuscript)**

When the late Harry Shorto was preparing his *Mon-Khmer Comparative Dictionary* he used data from only one Katuic language, Kui. He was at first under the impression that Kui shares a close, even sisterly, genetic relation with Khmer, and was quite unaware of the larger Katuic family. Consequently he compiled a 50 page draft Khmer-Kui historical phonology, with 176 lexical comparisons, and discussion of phonological correspondences.<sup>10</sup>

Around 1981 Shorto obtained the recently published (1980) Bru dictionary of See Puengpa and Theraphan L-Thongkum, and began to rework his theories, developing the idea of a "Bruan" (= West-Katuic) subgroup and reconstruction. An undated manuscript that came into my possession in 2005 is a 52-page draft of a detailed comparative analysis. The historical phonology describes the development of Kui and Bru as register languages from a non-register proto-Bruan, and the lexical comparisons are indexed to the relevant Proto-MK reconstructions throughout.

After completing his draft Proto-Bruan, Shorto obtained two more important Katuic dictionaries: the Pacoh dictionary of Watson et al. (1979), and Nancy Costello's (1971) multi-dialectal Katu dictionary. Shorto redrafted his corpus of comparisons to include the Pacoh and Katu data, creating a total of 794 Katuic etymologies and reconstructions. The historical phonology was revised, especially in respect of the proto-vocalism. The pages on which the correspondences were recompiled have not been found, although there is a passing reference to them in the existing notes.

In effect Shorto drafted a monograph-length Proto-Katuic reconstruction, with special emphasis on the development of West Katuic ("Bruan"). His results and conclusions effectively parallel those reached independently by Diffloth (1982) of which Shorto was apparently unaware. It is truly a great pity that this study did not appear in print when it was completed in the mid 1980s.

**Peiros (1996)**

Iliia Peiros moved from Moscow to Melbourne in the early 1990's, and began lecturing in historical linguistics at the University of Melbourne (including mentoring the present writer). To produce his own Proto-Katuic reconstruction Peiros took four principle sources of Katuic data: the dictionaries of Kui (Sriwises 1978), Bru (See Puengpa and Theraphan L-Thongkum 1980), Pacoh (Watson 1979) and Katu (Costello 1971). Methodically comparing every entry of each source pair-wise with the others, he produced an exhaustive comparative lexicon of 1241 putative cognate sets.

---

<sup>10</sup> Electronic copies of the unpublished papers of Harry Shorto mentioned in this paper can be obtained directly from Paul Sidwell (paulsidwell@yahoo.com).



Unfortunately, a series of errors undermines his work. Peiros misunderstands the register distinction in Bru and Kui, incorrectly describing it as being between "breathy" and "lax" voice, rather than between "breathy" and "clear". He wrongly analyzes the notational conventions of the sources for final consonant **-jh**, **-jʔ**, **-wh**, **-wʔ**, taking them as indicating **-h**, **-ʔ** suffixes in final clusters, apparently harking back to Schmidt (1906). Under the apparent influence of Schmidt (1904), Peiros interpreted all phonological minorsyllables as prefixes, thus incorrectly analyzing the entire lexicon as having underlying monosyllabic roots. And Diffloth (1982) is mischaracterized and misunderstood.

Problems continue in the phonological reconstruction. Peiros correlates breathy vowels and voiced proto-initials, but goes too far by reconstructing a voiced/voiceless opposition among all the initial sonorants, reconstructing a cumbersome and typologically marked consonant inventory. In respect of vowels Peiros implements an elaborated version of the approach of Thomas (1967), positing a Bru-like vocalism for the proto-language consisting of 15 simple vowels and 22 diphthongs. No word-final velar stop is reconstructed, instead all modern examples are derived from a glottal stop.

Overall, the work is deeply flawed and rarely cited. All of the errors noted above would certainly have been avoided if a draft version of the reconstruction had been circulated and discussed at one of the various Southeast Asian linguistics conferences meeting regularly at the time (such as SEALS, ICSTLL) where such a paper would have been welcomed. It is a most salutary lesson in the risks of holding back major analytical work from collegial review.

### ***Theraphan L-Thongkum (2001)***

Comparative Katuic research took a great leap forward at the turn of the century with the appearance of Theraphan's substantial volume. Reporting on field research conducted in remote areas of Sekong Province (Lao PDR), Theraphan extended our knowledge of the Katuic family with a substantial comparative lexicon that included original data for Kantu, Ta'Oi, Kriang, Chatong, Triw and Dakkang. The latter three languages, perhaps more appropriately characterized as Katu dialects, had not previously been documented in print. There is also a chapter on Bahnaric (discussed above).

The main text is in Thai, but data are given in IPA transcription with English glosses and are readily accessible to a wide audience. Data are presented in 1406 comparisons, reflecting more than 1300 distinct etymologies, organized semantically into a thesaurus, but lacking a lexical index. The presentation includes a comparative Katuic reconstruction and discussion of classification (but see below). Proto forms are offered for each entry, and each is labeled (PK, WK, NEK, CEK and/or SEK) according to the subgroup it reflects in Theraphan's classification.

The reconstruction is problematic because it is not explicitly justified in the text or in tabular correspondences. As data are presented in semantic groups rather than by phonological order, correspondences for each proto-phoneme must be extracted and compiled individually. In my own phoneme-by-phoneme survey, I was unable to infer a consistent formal analysis for the reconstruction as a whole. Publication of this would greatly enhance Theraphan's work; but even without it, the book makes a substantial amount of new and reliable data available, and is an excellent lexical resource.

### ***Sidwell (2005)***

Sidwell presents 1395 Katuic etymologies, a reconstruction of the historical phonology and a phonologically motivated classification of the languages. The results are in principle consistent with Diffloth (1982), but are more extensive by an order of magnitude. Data for some 16 named languages are utilized, drawing upon all published dictionaries, SIL wordlists,

published reconstructions and other sources that could be obtained at the time (including my own fieldwork in Laos).

The major lacunae of the work, also consistent with Diffloth (1982), is that a complete explanation is not offered for the phenomenon of glottalized rhymes attested in some Ta'Oï dialects (but discussed in Diffloth 1989 and Ferlus 1974). This phenomenon still awaits a proper account.

### 6.3 Pearic

Although Pearic was one of the first MK groups to be documented, the languages are highly endangered and thoroughly affected by contact with Khmer and Thai. In both principal and practice it is difficult to do comparative reconstruction on this grouping. Nonetheless a preliminary reconstruction has been offered, and some recent descriptive and linguistic salvage work has been pursued at Mahidol University. Useful MA dissertations from Mahidol include Siripen Ungsitipoonporn (2001), which contains an extensive comparative lexicon of two Chong dialects with more than 1400 entries, and Noppawan Thongkham (2003), which has a comparative lexicon of Chong and Kasong; it presents 281 numbered entries but a significant number of lacunae.

#### *Headley (1978)*

Robert Headley produced a preliminary Proto-Pearic reconstruction and internal genetic classification. He relied mostly upon the Baradat vocabulary of 1941 and a manuscript Chong vocabulary provided by Franklin Huffman. From the perspective of comparative phonology the reconstruction is rather incomplete. Unfortunately, his sources did not reliably distinguish the four registers, so Headley simply noted the phenomenon of "prefinal glottals" and decided to leave the question "to future linguists". The result is a compilation of 410 etymologies and reconstructions. It is expected that the addition of more recent field data collected by Thai linguists will encourage more fruitful efforts at reconstruction.

### 6.4 Monic

The Monic branch essentially consists of just two languages: Mon and Nyahkur. This limits the extent to which comparative reconstruction can be applied to investigate the history of the branch. Fortunately, Mon has a long written history that dates back some 1500 years. It has received much attention from philological/epigraphic studies, the most important of these being Shorto's (1971) *A Dictionary of the Mon Inscriptions from the sixth to the sixteenth centuries*. One attempt at a purely comparative reconstruction of Monic has been made, but the most successful efforts have combined a range of methodologies.

#### *Ferlus (1983)*

In this 90-page paper Ferlus uses both comparative and philological methods to reconstruct the phonetic history of Mon and offer a phonological reconstruction of Proto-Mon (in his formulation the stage before written Old Mon).

Particular attention is paid to the evidence of loan words from Thai, Khmer, Burmese and Indic to inform the phonetic interpretation of Old Mon inscriptions. The analysis is illustrated with extensive explicit tables showing the development of Mon rhymes leading to modern forms. Methodologically the paper elaborates the principals employed by Ferlus (1975) in respect of Vietic and reflected in his work on other MK branches.

Regrettably the paper does not offer results in the form of a consolidated comparative lexicon or index of reconstructions, which users would certainly welcome. Instead the reader must extract these from the hundreds of examples distributed throughout the discussion.

Methodologically Ferlus perfects a modern version of the approach pioneered so masterfully by Henri Maspero in his (1912) analysis of the history of Vietnamese (which used dialect and literary evidence to conclude that the language originated from an ancient mixing of Tai and MK elements, later overlain with Chinese). Thus we see that Ferlus' work derives from a well developed and respected French tradition that draws upon various classes of evidence to inform the analysis.

### ***Diffloth (1984)***

While prepared at approximately the same time as Ferlus' equivalent study, Diffloth's monograph uses a completely different approach to historical reconstruction. Developing upon the approach seen amongst the SIL tendency, Diffloth explicitly privileges the evidence of data from contemporary fieldwork for comparative reconstruction over and above evidence from historical written sources. On the basis of the extensive dialect data Diffloth offers some 680 etymologies, and provides reconstructions stratified into Proto-Mon, Proto-Nyakhkur and their common ancestor Proto-Monic.

The book makes a bold case against philological methods. To illustrate this Diffloth gives the example of Dvaravati Old Mon <srañ> "silver", which would seem to unambiguously indicate a Proto-Monic form \*sraŋ. However, since no expected modern Nyakhkur form [chreŋ] is found in the data, the principled decision is taken to not reconstruct \*sraŋ "silver". In effect Diffloth claims that (then) contemporaneous written evidence is not acceptable in the absence of confirmatory fieldwork data recorded some 1000 or 1500 years removed from the fact. A different but related argument favoring purely comparative methods is advanced in Diffloth (1980), discussed below (§6.7).

## **6.5 Khmeric**

The Khmeric branch is represented by a single language, Khmer, which shows three main dialects: Standard Khmer, Northern or Surin Khmer, and Western or Cardamom Khmer. Apparently all the known dialects derive from Khmer of the Angkorian period or later, and thus are younger than the oldest written sources. Therefore strictly comparative studies based on the dialects alone would not yield a reconstruction of Proto-Khmer, the ancestor of Old Khmer.

Existing dictionaries of Old Khmer include Long Seam (2000), Pou (1992, revised 2004), and Jenner (1980-86); a new dictionary by Jenner is in preparation. Studies that have investigated the history of Khmer by analysis of the epigraphic record include Ferlus (1992), which also extensively uses evidence of loan words to reconstruct the phonetic history of the language.

## **6.6 Khmuic**

The Khmuic branch is primarily represented by the vast Khmu' dialect chain that spreads across Northern Laos. Another ten or so member languages, all quite different from each other, are located around the periphery of the Khmu' area. The Khmu' dialects are now well documented (e.g. Suwilai Preamsirat 2002) but published sources in regard to the other Khmuic languages remain mixed. So far there has been no attempt at a consolidated Khmuic lexicon and Proto-Khmuic reconstruction, although some limited historical studies do exist. A bibliography of Khmuic studies is given by Proschan (1996).

### ***Filbeck (1978)***

Filbeck's 1971 PhD dissertation, published by *Pacific Linguistics* in 1978, is a comparative study of the Mal and Pray dialects known collectively as T'in. The Proto-T'in phonological system is reconstructed on the basis of a modest but well-chosen comparative lexicon. The

discussion includes consideration of the role of ancient loans from Thai, and there is significant discussion of the classification of T'in, incorporating tables of lexical comparisons with Khmu, Palaungic, Khmer and Mon.

***Egerod (1984)***

In a paper presented to the 17th International Conference of Sino-Tibetan Languages and Linguistics Søren Egerod presented a brief comparative study of Mlabri with comparisons establishing phonological correspondences with Kammu, Lamet, Mal, Proto-Palaung, Proto-Waic and Muong.

***Unchalee Singnoi (1988)***

This unpublished MA thesis at Mahidol University, *A Comparative Study of Pray and Mal Phonology* develops much of the work in Filbeck (1978), including presentation of an extensive Mal-Pray lexicon (approx. 1250 entries).

**6.7 Palaungic**

The Palaungic (or Palaung-Wa) branch is widely scattered geographically, and is commensurately very diverse internally. With more than a century of accumulated data the family is now quite well known. Since the 1980s in particular various grammars and dictionaries have been produced by Chinese scholars. Various sub-groups have been subject to important comparative-historical studies and Gerard Diffloth is known to be working on a much-anticipated consolidated Proto-Palaungic. Proschan (1996) includes a select bibliography of Palaungic studies.

***Shafer (1952)***

The first of Robert Shafer's two *Études sur l'austroasién* deals specifically with Palaungic. Building directly on the framework established by Schmidt (1904), Shafer sets forth more than a hundred comparisons for 11 Palaungic tongues, establishing regular correspondences for the consonants and vowels. The results include a modest lexicon of 88 Proto-Palaungic items. A substantial part of the paper (~ 25 pages) is devoted to discussing lexical parallels with Sino-Tibetan, Thai, and Khasi.

***Shorto (1963)***

The paper gives a brief account of word-formational patterns, particularly prefixation, in Palaung and related languages. Although not offering historical reconstructions as such, this study is immensely informative for comparative linguistics, yet this type of work has been somewhat neglected, and needs to be fostered.

***Luce (1965)***

In a paper ostensibly concerning Danaw, Luce provides a list of 245 lexical comparisons between Mon, Danaw, Riang, Palaung and Wa, plus an appendix of further MK and wider comparisons. It is apparent that this paper and related notes (such as preserved in the Luce archive at <http://archives.sealang.net/luce>) subsequently informed Shorto (1971, 2006 etc.)

***Mitani (1977)***

Mitani's paper treats 283 comparisons between Palaung, Rumai, Ra-ang and Darang, so it is effectively restricted to the Palaung-Riang sub-group. Phonological correspondences for consonants and vowels are established and proto-values reconstructed. A proto-lexicon is not offered, but one could readily derive the forms implied by the phonological formulae offered in the paper.

**Mitani (1979)**

In a follow-up to Mitani (1977), Riang data is worked into the analysis to further confirm Mitani's reconstruction of Proto-Ralaung-Riang.

**Diffloth (1980)**

Filling an entire issue of *Linguistics of the Tibeto-Burman Area*, this monograph-length study is to date the most important published contribution to comparative-historical studies of the Palaungic languages. Diffloth marshaled all of the then-available sources for languages of the Waic sub-branch (6 main sources, plus 28 lexicons of varying quality) to propose a comprehensive reconstruction of Proto-Waic. 544 etymologies are compiled and organized into a hierarchy of Proto-Wa-Lawa (PWL) and Proto-Waic (PW). This reflects the division of Waic into two sub-branches, Wa-Lawa and Samtao. Where Wa and Lawa reflexes are present a PWL form is reconstructed. Where additionally there is a reflex found in the Samtao sub-branch a PW form is also reconstructed. No distinct Proto-Samtao reconstructions are offered.

Generally the study is a model of comparative reconstruction, and Diffloth takes the opportunity to offer it as an exemplar of his preferred methodology. The study opens with a the following explicit statement of Diffloth's programmatic perspective:

Among the fourteen or so extant branches of the Mon-Khmer family, only three or four have developed and preserved enough differentiation today to yield proto-branch reconstructions of great antiquity. They are: the Bahnaric, the Aslian, the Palaungic and probably the Viet-Muong branches. It is mostly from these reconstructions that we will be able some day to cast a glance at Proto-Mon-Khmer and beyond. The Katuic, Khmuic and Nicobarese branches, while extremely useful, do not appear to be as diversified as the first four; Monic and Khmeric, in spite of their written records and resulting prestige, ironically rank lower in this respect. (p.3)

**Svantesson (1988)**

This descriptive and historical paper goes into great detail analyzing and explaining the odd phonetic history of U, with extensive comparisons to other Palaungic data (especially with Hu and Lamet).

**Paulsen (1989/1991)**

Originally an MA thesis (1989), this study was subsequently published as a long paper in *Mon-Khmer Studies* (1991). It nicely complements Diffloth (1980), since it is effectively a reconstruction of the Samtao sub-branch of Waic, which lacks a distinct reconstruction in Diffloth's treatment. The results include a clear historical phonology, and an etymological lexicon of 556 entries and reconstructions, comparing Plang, Shinman and Pangloh, drawn from the author's own field data.

Curiously the study only mentions Diffloth (1980), but appears not to make use of it. Thus the Plang data and etymologies are analyzed without reference to already known Waic context, and consequently a number of Thai and other loans are erroneously treated as Plang.

**Diffloth (1992)**

A short follow-up to Diffloth (1980), this paper discusses data from the recently surveyed Bulang sub-branch of Waic, with reference to his earlier Waic reconstruction. It is a pity that Paulsen's (1989/1991) related results are not discussed.

## 6.8 Aslian

The Aslian branch, with a score or more distinct languages, is comparable to Katuic in its internal diversity, and may thus make a significant contribution to Proto-MK reconstruction. Exciting early progress at the beginning of the 20th century was followed by a long lull. The Malayan Emergency (1948-1960) meant that circumstances were not ripe for a revival of fieldwork until the 1960s; then a surge of effort and some comparative studies appeared.

More recent fieldwork initiatives are now helping to complete our knowledge of the family (e.g. Kruspe 2004 on Semelai, and now working on Ceq Wong and Mah Meri; Burenhult 2005 on Jahai, and now working on various North and Central Aslian languages). Consequently the volume and quality of published and unpublished Aslian lexicon that could be marshaled by an effective collaborative effort would go a long way towards facilitating a comprehensive Proto-Aslian reconstruction.

A selective bibliography of Aslian studies is provided by Bishop & Peterson (1994), and there is long descriptive article on the Aslian branch by Matisoff (2003).

### *Diffloth (1968)*

This short paper marks the rebirth of comparative Aslian studies in the second half of the 20th century. A modest amount of data from four Semai dialects (collected in the field by the author) is presented in support of regular correspondences that lead to a partial reconstruction of Proto-Semai phonology.

### *Benjamin (1976)*

This is a major essay that discusses the classification and disposition of the Aslian family. The remote history of the Aslians is reconstructed to some extent with reference to migrations, contact and borrowings. The Skeat & Blagden materials are discussed and a very useful concordance of language names with contemporary equivalents is presented. A basic vocabulary of 146 terms for 22 named languages crowns the paper.

### *Diffloth (1975)*

In this 19-page paper the classification of Aslian into three sub-branches, originally proposed on lexical grounds, is further confirmed by historical phonology. With a minimal number of illustrative etymologies, outlines of the phonological histories of the Northern, Central and Southern sub-branches are sketched.

### *Diffloth (1977)*

This paper is the better-developed successor to Diffloth (1968). It is based on comprehensive data collection over many years in the Semai speaking communities. 284 numbered etymologies illustrate phonological correspondences and a very detailed historical phonology. The paper is copiously illustrated with diagrammatic explanations of the historical vowel changes, including a large unified, flow chart of the chronology of vowel innovations. The author mentions having collected an Aslian lexicon in excess of 26,000 items; it is regrettable that this resource remains unpublished/inaccessible.

### *Diffloth (1979)*

This short paper puts the results of Diffloth (1977) into broader MK perspective, and includes a very useful comparison, focused on pronominal systems, between Aslian and other MK languages.

***Phillips (2005)***

Within these unpublished but widely circulated 115 pages, Phillips reports on extensive survey of Semai dialects, and elaborates Diffloth's (1977) reconstructions with several hundred new reconstructions.

**6.9 Vietic**

The Vietic branch, and in particular the Viet-Muong<sup>11</sup> sub-branch, has attracted considerable attention in the form of comparative research, especially since the breakthrough studies by Haudricourt in the 1950s. With the increasing availability of field data on minor Vietic languages since the 1970s, a better understanding of the extent and typology of the branch has emerged. Draft consolidated Proto-Vietic reconstructions are now circulating among specialists. Barker (1993) provides a bibliography of Vietic studies.

***Cuisinier (1951)***

In a paper documenting Muong ritual texts, Cuisinier presents considerable dialect data, and sets forth phonological correspondences between Vietnamese and Muong, although without attempting reconstruction.

***Haudricourt (1952, 1953, 1954)***

As discussed elsewhere in this paper, these ground-breaking papers established the solid MK ancestry of Vietnamese, and set the agenda that has guided the reconstruction of Vietic linguistic history over since. Crucial to the argumentation was the use of comparative data from Khmu' that permitted the reconstruction of finals \*-ʔ, \*-h and \*-s, supporting the new model of tonogenesis.

***Barker (1963)***

This modest paper (nine pages) is the first in a series by Milton Barker (and also later Muriel Barker) contributing to the reconstruction of Proto-Viet-Muong. The Barker's had done field work on Muong dialects in the early 1960s, stimulating their interest in to the relation with Vietnamese. This paper discusses the reconstruction of initial labial consonants.

***Barker (1966)***

A large amount of data is packed into this paper's 16 pages. Some 600 Vietnamese-Muong lexical comparisons are given, illustrating the tonal correspondences between Vietnamese and Muong. The paper demonstrates that their systems share a common origin, with the five-tone Muong system readily derived from the six tones preserved in Northern Vietnamese.

***Barker & Barker (1970)***

On the basis of 210 illustrative comparisons between Vietnamese and Muong the Proto-Viet-Muong finals and vowels are reconstructed. Some ambiguity is unresolved concerning the phonetic values of several vowel phonemes, reflecting the fact that the paper does not take the analysis further by invoking wider Vietic comparisons.

***Ferlus (1975, 1982, 1991, 1992a, 1996, 1997, 1998, 2001, 2005)***

In a series of related papers since the early 1980s Ferlus has continued to develop his reconstruction of Vietic history. He progressively takes into account more data from Minor Vietic languages, and develops a comprehensive theory concerning the early influence of Chinese, and the emergence and nature of tones and registers in Vietic. From the beginning

---

<sup>11</sup> Perhaps better called "Muong-Viet", to eliminate confusion with "Vietic".

Ferlus has carried on and developed the great methodological traditions of Henri Maspero, whose early 20th Century work has clearly been very influential.

In the first of these papers (1975) Ferlus sets forth a comprehensive account of the phonological history of Viet-Muong within the broader MK context, given then-available data. Each aspect of the historical phonology is dealt with in a systematic way, invoking wider MK comparisons, and, for the first time, data from other minor Vietic languages that were just starting to be documented (especially Sáčh, Thavung and some Pakatan). The interpretation of philological data, especially de Rodes' Middle Vietnamese dictionary, also plays an important role. We also see the very beginnings of Ferlus' *spirantization* theory that would subsequently be so influential in his comparative reconstruction. Typical of Ferlus' style, the paper does not contain a consolidated comparative vocabulary, but features scores of illustrative items throughout the text.

The subsequent papers build on and improve the model, especially as more data from minor Vietic languages is integrated. Since 1991 Ferlus has been circulating a consolidated Proto-Vietic lexicon in various drafts, and is now preparing a "final" version with over 3000 entries. The latter is keenly anticipated for the fundamental contribution it is expected to make to deeper MK reconstruction.

#### ***Thompson (1976)***

This substantial paper attempts to reconstruct Proto-Viet-Muong as extensively as the available data would permit. Thompson provides a consolidated comparative vocabulary of approximately 700 entries, with Vietnamese and three Muong dialects. Reconstructed forms are offered for about 600 of these. There are 49 pages of discussion and analysis, which are based entirely upon comparative reconstruction.

#### ***Huffman (1977)***

In this paper Huffman revisits the thorny issue of the place of Vietnamese within MK, and sets forth a comparative basic vocabulary with illustrative entries from most branches of MK.

#### ***Sokolovskaia (1978)***

This very important paper presents a reconstruction of Proto-Vietic, utilizing a total of 27 sources. They include numerous Muong dialects, minor Vietic languages such as Thavung, Pakatan and Sáčh, and Vietnamese dialects. The bulk of the text is found in its etymological lexicon of approximately 650 entries. The methodology is entirely comparative, and wider MK data are not discussed. I understand that Sokolovskaia subsequently produced an extensively revised and expanded treatment, the manuscript of which was offered to *Mon-Khmer Studies* for publication in the 1980s. However, Sokolovskaia passed away, and the only hard copy of her manuscript was misplaced and has not been seen since.

### **6.10 Nicobaric, Khasic**

Unfortunately there is no effective tradition of published comparative Nicobaric or Khasic studies of which I am aware, beyond the inclusion of various lexical comparisons in works of wider scope such as Shorto (2006).

### **6.11 Proto-Mon-Khmer**

It was quite an extraordinary disappointment that in the years which immediately followed the descriptive and comparative effervescence of the 1960s and '70s, no major work appeared that surveyed, consolidated and analyzed the newly available field data and reconstructions, relating them to the frameworks sketched out by Schmidt, Shafer, Pinnow etc. Lacking such



an authoritative canonical reference work and the programmatic impetus it would have provided, the field drifted and stagnated.

None the less some individuals did pursue their own compilations and analyses. In this respect the highest expectations fell upon Gerard Diffloth. He ceased publishing branch level reconstructions after 1984, although did not halt his comparative research. Since the late 1970s he has been actively compiling a consolidated MK comparative lexicon and reconstruction of Proto-MK. Many of his notes from the time still exist and can be seen at the Cornell University Library. By 1980 the project was already so well developed that he applied for and received National Science Foundation funding for *A Mon-Khmer Etymological Lexicon*. The application included a draft chapter reconstructing the Proto-MK faunal lexicon, indicating that even then the project was certainly at a stage that would have been useful and welcomed by the wider research community. A decade later a related project was also funded (*Khmer (Cambodian) Etymological Lexicon*, NSF 1988-1990 and NEH 1989-1991), predicated on the claim that a sufficient MK etymological lexicon had already be compiled and analyzed to inform etymological dictionaries of individual languages. The wider academic community is still waiting to gain the benefit of these publicly funded resources.

In Moscow through the 1980s Ilia Peiros was also preparing his own MK comparative lexicon and reconstruction. In Peiros (1998) he describes the methods he used, and mentions a set of 1500+ etymologies compiled on ledger cards. More than 100 of the reconstructions formed part of his 1989 paper in which he seeks to link SEAsian language families with the Nostratic and Sino-Tibetan reconstructions of Sergei Starostin. Recently Peiros has made his comparative data and results available on-line through the *Tower of Babel* project site (<http://starling.rinet.ru/>), including some 2,246 Proto-Austroasiatic reconstructions (by my latest count). It is very welcome to have this material available on-line, although the underlying analysis is difficult to assess. There is no supporting documentation justifying the reconstruction, such as statements of correspondences or distributions, or other explanatory text. Generally the approach has been to compile as extensive a list as possible based on a few criterion languages/sub-groups (especially Written Khmer), such that most sets invoke data from only one, two or three MK branches. The same site also provides comparative data and reconstructions for various MK branches, although the bulk of these appears to be reproduced from publications other authors.

Both Diffloth and Peiros consistently shy away from philological methods, preferring to use the most reliable and recent field data they can obtain. By contrast, Harry Shorto pursued his comparative MK project solidly on the basis of more traditional methods which especially privilege epigraphic data. He believed that written evidence that dates back 1,000-1,500 years—and is thus a much closer witness to the proto-language—is inherently preferable over modern forms, which are that much further removed from the mother tongue (Shorto 1965 elaborates on the theory and method of using epigraphic data). On that basis Shorto went ahead to produce his own version of the canonical reference that he hoped would do justice to the field, but which unfortunately would have to wait 30 years after the 2nd ICALL and 12 years after his own passing to the appear in print.

### ***Shorto(2006)***

Shorto's passion was Mon philology. Having taken up the torch from Blagden and Luce after accepting a lectureship at SOAS in 1948, Shorto devoted himself to the study of this Southeast Asian language with the longest and most complete written tradition. He published his *A Dictionary of Modern Spoken Mon* in 1962, and followed this up with *A Dictionary of the Mon Inscriptions* in 1971. The latter included extensive etymological commentary (largely drawn from Schmidt 1905, Skeat & Blagden 1906 and Gordon Luce's notes), which he

subsequently reworked to form the basis of his MKCD, the results of which are summarized below.

The proto-MK consonants presented by Shorto are shown below. They match the table offered by Diffloth in his 1974 *Encyclopaedia Britannica* article (and elsewhere), reflecting a more or less consistent view that can be traced back to the foundation laid by Schmidt:

*/	<b>p</b>	<b>t</b>	<b>c</b>	<b>k</b>	<b>ʔ</b>	
	<b>b</b>	<b>d</b>	<b>j</b>	<b>g</b>		
	<b>ɓ</b>	<b>ɗ</b>				
	<b>m</b>	<b>n</b>	<b>ɲ</b>	<b>ŋ</b>		
	<b>w</b>	<b>r l</b>	<b>y</b>			
			<b>s</b>	<b>h</b>	<b>/</b>	

Additional proto-segments **\*t<sub>2</sub>**, **\*d<sub>2</sub>**, **\*n<sub>2</sub>** were added by Shorto in his second draft of the MKCD. The first two of these are posited (in most cases) to account for the correspondence of prenasal s in Northern MK and Munda with **t** in other MK languages (e.g. **\*t<sub>2</sub>ɲiiʔ** 'sun, day' on the basis of forms such as Palaung **səŋi** and Mundari **singi** versus Khmer **thŋay** and Old Mon *tiēy*). **\*n<sub>2</sub>** is suggested to explain a correspondence of prevocalic laterals in Northern MK and nasals elsewhere (e.g. **\*bn<sub>2</sub>iəs** 'spear' on the basis of such forms as Riang-Lang **\_ples** and Old Mon **bnəs**).

Shorto reconstructs PMK initial consonant clusters based upon the following relation between the registers of Mon and Khmer:

PMK	Mon	Khmer
voiceless + voiceless	head register	head register
voiceless + voiced	chest register	head register
voiced + voiceless	head register	chest register
voiced + voiced	chest register	chest register

This cluster reconstruction appears to hold up fairly robustly, although necessarily it will not recover features that have been lost due to parallel developments in the two criterion languages.

The proto-vowel inventory suggested by the MKCD is as follows:

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

Shorto tables the correspondences justifying this reconstruction (excepting the bracketed item) in Table 1 of Part 1 of the MKCD. The system is strictly MK in the sense that it is derived by comparison of just those two languages. The resulting reconstruction is generalized to the family as a whole, invoking various vowel alternations to account for irregular or unexpected correspondences.

Shorto's lexical reconstruction yields 2,246 groups of proto-MK lexemes, plus several hundred Palaungic and South Bahnaric etymologies (in appendices) for which provisional PMK antecedents are posited. In most cases the PMK forms are justified by the presence of reflexes in at least Mon and Khmer, or one MK language and a convincing wider comparison (in Shorto's estimation) with Munda and/or Austronesian.

Although not spelt out explicitly in the surviving notes, the method Shorto employed in compiling more than 2,500 etymologies and 30,000 lexical citations is more or less clear. It began with extracting the nearly a thousand comparisons first presented by Schmidt in his various foundational studies. To these were added additional data from other comparative studies, specifically Skeat and Blagden (1906), Schafer (1965), Barker (1963, 1966), Blood (1966) and Smith (1972). Reliance on the later two sources accounting for the fact that some 146 items in his PMK lexicon are posited with Bahnaric reflexes alone.

The framework thus established was further fleshed out with data from published lexicons and dictionaries, and some unpublished field data contributed by colleagues, especially from individuals with whom he had direct contact thanks to the ICAAL meetings in Hawai'i and Mysore. However, personal factors acted to limit the cooperative sharing of data to Shorto's disadvantage. According to his daughter Anna (pers. com.), Harry was reluctant to have dealings with "missionary types". Consequently he did not seek to cultivate personal contacts with SIL researchers, and even declined to submitted work to *Mon-Khmer Studies*. Instead he worked more and more alone

In total the MKCD presents some 30,000 lexical citations from numerous languages, in a well organized and accessible fashion, making it a marvelous resource for comparative studies. However, given the basing of the historical analysis principally upon epigraphic Mon and Written Khmer, and the failure to use important field data that was unpublished but arguably available, the reconstructions he offers are skewed and deserve a careful reworking. For a detailed constructive critical review of the reconstruction see Sidwell (2006).

## 6.12 Concluding remarks of history of comparative reconstruction

In reviewing the last half century or so of comparative MK linguistics, we discern a striking pattern that might be crudely characterized as the dichotomy of **philology versus field work**. In its extreme realizations, we see the approach that Shorto took in reconstructing Proto-MK counterposed to the strictly 'contemporary data only' approach of Diffloth (1980, 1984 etc.), which explicitly downgrades the data that Shorto most relied upon.

Shorto's MKCD is the ultimate test case of the first approach, and it does have many evident weaknesses. The very method that made the reconstruction possible—the reliance on philological/epigraphic data—also limited the scope and accuracy of the results. Yet it allowed Shorto to draw a broadly accurate picture whose flaws could be readily understood, and which may be corrected sooner rather than later.

Diffloth, in contrast, has been the primary proponent of the non-philological approaches that have been relatively dominant since the 1960s. Although his published work has been limited to branch-level reconstructions, Diffloth still has the opportunity to prove his larger case, if and when he publishes his own long-awaited reconstruction of proto MK.

But as may be most appropriate for the study of Southeast Asia, the middle road is also being successfully navigated. The great exemplar in this respect is Ferlus, whose work takes into account a wide range of field data collections and analyses, yet is also profoundly influenced by his deep knowledge of classical languages and written traditions. His distinguished analyses of the history of Mon (1983), Khmer (1992b) and Vietic (1975, 1982, 2001 etc.) have been based upon a readiness to apply all types of available evidence, weigh them accordingly, and apply them to solving the problems at hand.

The conclusion to be drawn from reviewing the past century of work on the MK languages is surely that no single methodology or resource (and certainly no individual researcher!) is sufficient to reconstruct the history of this extraordinarily ancient and diverse language family. But by the same token, no method, resource, or researcher should ever be seen as being marginal enough to be safely ignored. Again and again it is the combination of method, resources, and individual drive that ultimately leads to lasting results.

## 7 Where to now?

We have now reached a stage in the course of comparative MK linguistics that is remarkably favorable. The extent of the family is now fairly well known; there are reliable and extensive lexical data available for all branches of the family, and there are unlikely to be a score of new MK languages still to be found, let alone any new branches. The most diverse branches—Bahnaric, Katuic, Palaungic, Aslian, Vietic—have already been the subject of comparative studies of varying extents (discussed above), and comprehensive reconstructions are either in train or at least already technically feasible. Thus the program briefly sketched out by Diffloth (1980:3) is not only practical, but with a serious effort could be effectively implemented over the next few years.

However, I do not suggest that we follow Diffloth's prescriptions to the letter and pay less regard to the evidence of smaller branches, especially Mon and Khmer. Shorto unambiguously demonstrates the usefulness and relevance (although arguably not the primacy) of these well documented and historically analyzed languages for comparative reconstruction. It thus seems self-evidently reasonable that the insights achieved by Shorto (2006), and Ferlus (especially 1983, 1992), should be appropriately integrated with the results of comparative studies dealing with the most diverse MK branches.

In addition we should look to utilize Khmu' and Pearic lexicon in reconstruction. Despite the fact that extensive new data is required to recover Proto-Khmuic, the multi-dialectal Khmu' lexicon of Suwilai Premasri (2002) represents a substantial reliable source that can be applied directly to the reconstruction of PMK. Regarding Pearic, Suwilai is presently consolidating a comparative lexicon based on two decades of field work conducted from Mahidol University. Hopefully this will usefully inform a Proto-Pearic reconstruction that will rework and enhance the framework established by Headley (1985). Although the Pearic data are inherently limited due to their endangered status, there are indications that Pearic may be especially important to understanding PMK phonology (if not lexicon<sup>12</sup>). As Diffloth (1989) discusses, there is the unresolved issue of whether to reconstruct a voice quality distinction, perhaps a tense/creaky register, in PMK. The relevant branches are Vietic, Katuic and Pearic. It is imperative that we investigate the historical origins of the tense/creaky registers in these branches, determining as far as the data will allow the extent to which they have arisen secondarily, or must be treated as archaic.

Another crucial aspect of Pearic concerns its ambiguous classification within MK. There are no clear indications of Pearic sub-grouping with any other branch. Therefore, Pearic potentially reflects the highest branching node in the tree (below Munda) and may thus be indispensable to reconstructing PMK. By contrast, if Nicobaric sub-groups with Monic, and Khasic sub-group with Palaungic, these two branches are arguably less important. Their data may be better dealt with latter, after attention to the larger branches has further informed our understanding of the past and refined our recognition of the gaps. Currently available published data for these two branches is rather limited:

- Khasic: there are dictionaries of Standard Khasi (e.g. Singh 1906, Bars 1973), and contemporary descriptive studies such as Nagaraja (1985), but we need extensive dialect data to facilitate a Khasic reconstruction (I am sure that more exists than I am aware of),
- Nicobaric: there are some older dictionaries (e.g. Man 1888/1889, Röpstroff 1884, Whitehead 1925), and descriptive analyses (e.g. Radhakrishnan 1981), but

<sup>12</sup> The Pearic languages are all well into late stages of endangerment, such that it is not possible to collect extensive lexicons unaffected by contact languages, restricting their usefulness for lexical reconstruction.

more descriptive work is urgently required, especially since the recent devastating tsunami.

Given that the technical conditions for substantial progress in comparative reconstruction are largely satisfied, the remaining variable to consider is the human factor. The sheer scope of the challenge of dealing with a family of 150 or so languages is such that no one can or should try to do everything; collaboration is ultimately essential. This has been the weakest link in our field over the last 30 years, characterized by a lack of dedicated conferences, joint publications, and the lack of a clearly articulated programmatic vision around which efforts can be coordinated and motivated.

Can we work together to accelerate the pace of progress? Yes, by pursuing any strategy that promotes collaboration between researchers, including sharing data, cooperating in field work, hosting regular conferences and publications, and committing to open discussion of the methodological and programmatic aspects of our work.

With an appropriate collaborative effort I am sure that we can see the emergence of a consolidated Proto-MK reconstruction and comprehensive etymological dictionary in years, rather than in decades. It is my earnest wish that the revival of these ICAAL meetings we are participating in today will contribute concretely to the realization of such a project.

## References

- Aymonier, Etienne F. 1974. *Dictionnaire français-cambodgien*, Saigon, Imprimerie Nationale.
- Azémar, H. 1886. Dictionnaire Stieng. Recueil de 2,500 mots fait à Brolâm en 1865. *Excursions et Reconnaissances* 12:99-146, 251-344.
- Baradat, R. 1941. *Les dialectes des tribus sâmrê*. Paris, Manuscript.
- Barker, Milton E. & Muriel A. Barker 1970. Proto-Vietnamuong (Annamuong) final consonants and vowels. *Lingua* 24.3:268-285.
- Barker, Milton E. 1963. Proto-Vietnamuong initial labial consonants. *Văn-hoa Nguyệt-san* 12.3:491-500.
- Barker, Milton E. 1966, Vietnamese-Muong tone correspondences, in *Studies in Comparative Austroasiatic Linguistics*, ed. N.H. Zide (ed.) Studies in comparative Austroasiatic linguistics pp. 9-25. The Hague, Mouton & Co.
- Barker, Milton E. 1993, Bibliography of Mường and other Vietic language groups, with notes, *Mon-Khmer Studies* 23:197-243
- Bars, E. 1973. *Khasi-English Dictionary*. Shillong, Meghalaya, Don Bosco.
- Benjamin, Geoffrey. 1976. Austroasiatic Subgroupings and Prehistory in the Malay Peninsula. In Jenner et al. (eds.) *Austroasiatic Studies*, Oceanic Linguistics Special Publications No. 13, University of Hawaii Press, Honolulu. 37-128.
- Bergsland, Knut and Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology*, 3:115-153.
- Bishop, N. and Peterson, M. 1994. A selective Aslian bibliography, *Mon-Khmer Studies*, 24:pp. 161-169
- Blagden, Charles Otto. 1894. Early Indo-Chinese influence in the Malay Peninsula, as illustrated by some of the dialects of the aboriginal tribes. *Journal of the Straits Branch of the Royal Asiatic Society*. 27:21-56.
- Blagden, Charles Otto. 1923. *Epigraphia Birmanica*. Vol. 3, part 1: Môn inscriptions IX-XI. Rangoon, Government Printing.

- Blagden, Charles Otto. 1928. *Epigraphia Birmanica*. Vol. 3, part 2: Môn inscriptions. Section 2: The medieval Môn records. XII: The inscriptions of the Kalyân?nîsimâ, Pegu. Rangoon Government Printing
- Blagden, Charles Otto. 1934. *Epigraphia Birmanica*. Vol. 4, part 1: Môn inscriptions XIII-XVIII; cover dated as part 2 [1936]; edited by U Mya. Rangoon, Government Printing.
- Blagden, Charles Otto. 1936. *Epigraphia Birmanica*. Vol. 4, part 2: Môn inscriptions XIII-XVIII; plates 1-11 in portfolio. Rangoon, Government Printing.
- Blood, Henry F. 1966. *A Reconstruction of Proto-Muong (Including Tentative Reconstruction of Proto-South-Bahnaric)*. M.A. Thesis, Department of Linguistics Indiana University.
- Burenhult, Niclas. 2005. *A grammar of Jahai*. Canberra, Pacific Linguistics.
- Cabaton, Antoine. 1905. Dix dialectes indochinois recueillis par Prosper Oden'hal. Etude linguistique par Antoine Cabatone. *Journal Asiatique*, Dixième série, tome V:265-344.
- Costello, Nancy A. 1971. *Ngũ-vũng Katu: Katu vocabulary*. Vietnam Montagnard Language Series 5. Saigon, Department of Education (available via Summer Institute of Linguistics, Dallas).
- Cuisinier, Jeanne. 1951. *Prières accompagnant les rites agraires chez les Muong de Man duc*. Publications de l'École Française d'Extrême-Orient. Hanoi, Vol. 33.
- de Rhodes, Alexandre. 1651. *Dictionarium Annamiticum Lusitanum, et Latinum*, Rome.
- Dempwolff, Otto, 1938. *Vergleichende Lautlehre des austronesischen Wortschatzes*. 3. Band: Austronesisches Wörterverzeichnis. Beihefte zur ZES. Berlin: Dietrich Reimer.
- Diffloth, Gérard. 1968. Proto-Semai Phonology. *Federation Museums Journal* (new series), 13: 65-74.
- Diffloth, Gérard. 1974. Austro-Asiatic Languages. *Encyclopaedia Britannica*. Chicago/London/Toronto/Geneva, Encyclopaedia Britannica Inc. pp 480-484.
- Diffloth, Gérard. 1975. Les langues Mon-Khmer de Malaisie: classification historique et innovations. *Asie du Sud-Est et Monde Insulindien* 6.4: 1-19.
- Diffloth, Gérard. 1977. Towards a History of Mon-Khmer: Proto-Semai Vowels. *Tônan Ajia Kenkyû (Southeast Asian Studies)* 14.4:463-95.
- Diffloth, Gérard. 1979. Aslian languages and Southeast Asian prehistory. *Federation Museums Journal*. 24ns:3-16.
- Diffloth, Gérard. 1980. The Wa Languages. *Linguistics of the Tibeto-Burman Area*. Vol. 5.2. Berkeley: University of California.
- Diffloth, Gérard. 1982. Registres, dévoisement, timbres vocaliques: leur histoire en Katouique. *Mon-Khmer Studies* 11:47-82.
- Diffloth, Gérard. 1984. *The Dvaravati-Old Mon Language and Nyah Kur* (Monic Language Studies 1). Bangkok, Chulalongkorn University Printing House.
- Diffloth, Gérard. 1989. Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies* 15:139-154.
- Diffloth, Gérard. 1990. What Happened to Austric? *Mon-Khmer Studies* 16-17:1-9.
- Diffloth, Gérard. 1991. Palaungic vowels in Mon-Khmer perspective. In Davison ed. *Austroasiatic Languages, essays in honour of H. L. Shorto*. SOAS, University of London.
- Diffloth, Gérard. 1992. *Indigenous languages of Cambodia*. 25th ICSTLL, Berkeley, CA, Oct. Unpublished manuscript.
- Diffloth, Gerard. 1994. The lexical evidence for Austric so far. *Oceanic Linguistics* 33(2):309-321.
- 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.
- Dourisbourne, Le Père P.-X. 1889. *Dictionnaire bahnar-français*. Hongkong, Imprimerie de la Société des Missions Etrangères 45.

- Efimov, Aleksandr. 1983. *Problemy fonologicheskoy rekonstrukcii proto-katuicheskogo jazyka*. Kandidat Dissertation, Institute of Far Eastern Studies Moscow.
- Efimov, Aleksandr. 1990. *Istoricheskaja Fonologija Juzhnobaxnaricheskix Jazykov*. Moskva, Nauka.
- Egerod, S. 1984. *The language of the Mlabri or Spirits of the Yellow Leaves: preliminary observations on its affiliation*. Paper read at the 17th International Conference of Sino-Tibetan Languages and Linguistics.
- Feer, Léon. 1877. *Etudes cambodgiennes*. La collection Hennecart de le Bibliothèque Nationale. *Journal Asiatic*, 161-234.
- Ferlus, Michel. 1974. La langue Ong, mutations consonantiques et transphonologisations. *Asie du Sud-Est et Monde Insulindien* 5.1: 113-121.
- Ferlus, Michel. 1975. Vietnamien et proto-viet-muong. *Asie du Sud-Est et Monde Insulindien* 4.4: 21-55.
- Ferlus, Michel. 1982. Spirantisation des obstruantes médiales et formation du système consonantique du vietnamien. *Cahiers de Linguistique, Asie Orientale* 11.1:83-106.
- Ferlus, Michel. 1983. Essai de phonétique historique de môn. *Mon-Khmer Studies* 12:1-90.
- Ferlus, Michel. 1991. *Vocalisme du Proto-Viet-Muong*. Paper circulated at the Twenty-fourth ICS-TL&L. Chiang Mai University, Oct. 10-11, 1991.
- Ferlus, Michel. 1992a. Histoire abrégée de l'évolution des consonnes initiales du vietnamien. *Mon-Khmer Studies* 20:111-27.
- Ferlus, Michel. 1992b. Essai de phonétique historique du khmer (Du milieu du premier millénaire de notre ère à l'époque actuelle), *Mon-Khmer Studies* 21: 57-89.
- Ferlus, Michel. 1996. Langues et peuples viet-muong. *Mon-Khmer Studies*, 26:7-28.
- Ferlus, Michel. 1997. Problèmes de la formation du système vocalique du vietnamien. *Cahiers de Linguistique, Asie Orientale*, 26.1.
- Ferlus, Michel. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15:1.1-27.
- Ferlus, Michel. 2001. *The Origin of Tones in Viet-Muong*. Paper presented at the XI meeting of the Southeast Asia Linguistics Society, Mahidol University, Bangkok, May 16-18 2001.
- Ferlus, Michel. 2005. *Proto viet-muong (Proto-Vietic)*. Manuscript provisional reconstruction.
- Filbeck, David. 1978. *T'in: a historical study*. Pacific Linguistics Series B-49. Canberra, Australian National University.
- Gainey, Jerry. 1985. *A comparative study of Kui, Bruu and So phonology from a genetic point of view*. Bangkok, Thailand: Chulalongkorn University thesis.
- Gregerson, Kenneth J., Kenneth D. Smith, and David D. Thomas. 1976. The place of Bahnar within Bahnaric. In Philip N. Jenner, Laurence C. Thompson and Stanley Starosta (eds.), *Austroasiatic Studies*, part 1, pp:371-406. Oceanic Linguistics Special Publication, 13. Honolulu: University of Hawaii Press.
- Grierson, G.A., 1903-1228 (13 vols.). *Linguistic Survey of India*. Calcutta, Office of the superintendent of government Printing.
- Guilleminet, Paul and R.P. Jules Alberty. 1959-63. *Dictionnaire bahnar-français*. Hanoi, Paris, Publications de Ecole Française d'Extrême Orient.
- Haswell, J.M. 1874. *Grammatical notes and vocabulary of the Peguan language*. American Baptist Press, Rangoon.
- Haudricourt, André-Georges. 1952. L'origine môn-khmèr des tons en vietnamien. *Journal Asiatique* 240:264-265.
- Haudricourt, André-Georges. 1953. La place du vietnamien dans les langues austroasiatiques. *Bulletin de la Société de Linguistique de Paris* 49.1:122-128.

- Haudricourt, André-Georges. 1954. De l'origine des tons en vietnamien. *Journal Asiatique* 242:69-82.
- Headley, Robert K. Jr. 1985. Proto-Pearic and the Classification of Pearic. In Suriya Ratanakul and David Thomas (eds.) *Papers in Southeast Asian Languages and Linguistics in Honour of André-Georges Haudricourt*. Bangkok, Mahidol University. pp 428-478
- Huffman, Franklin E. 1977. An Examination of Lexical Correspondences between Vietnamese and Some Other Austro-Asiatic Languages. *Lingua* 43:171-198.
- Jacq, Pascale & Paul Sidwell. 2000. *A Comparative West Bahnaric Dictionary*. München, Lincom Europa.
- Jenner, Philip N., Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies*, Vols. 1 and 2. Oceanic Linguistics Special Publication #13. Honolulu, University of Hawaii (Oceanic Linguistics, Special Publication, No. 13)
- Jenner, Philip Norman. 1980-86. *A Chrestomathy of Pre-Angkorian Khmer*. Southeast Asia Paper No. 20, Parts 1-4, Honolulu, University of Hawaii.
- Kruspe, Nicole D. 2004. *A grammar of Semelai*. Cambridge, Cambridge University Press.
- Leitch, Myles. 1981. *A Study of Rengao in Proto-North Bahnaric*. ms. Copy held at Summer Institute of Linguistics Library Bangkok.
- Long Seam. 2000. *Dictionnaire du Khmer Ancien (D'après les inscriptions du Cambodge du I<sup>e</sup>. – VIII<sup>e</sup>. siècles)*. Phnom Penh.
- Luce, Gordon H. 1965. Danaw, a dying Austroasiatic language. *Lingua* 14:98-129.
- Luce, Gordon H. 1969-70. *Old Burma—early Pagán*. Vols. I-III. New York, J. J. Augustin.
- Luce, Gordon H. 1981. *A Comparative Word List of Old Burmese, Chinese and Tibetan*. London, SOAS.
- Luce, Gordon H. 1985. *Phases of Pre-Pagán Burma: Languages and History*. 2 vols. Oxford, Oxford University Press.
- Man, E.H. 1888/1889. *A Dictionary of the Central Nicobarese Language*. Reprint: Delhi: Sanskaran Prakashak, 1975
- Man, E.H. 1889. *A Dictionary of the Central Nicobarese Language*. W.H. Allen & Co. London.
- Mason, Francis. 1854. The Talaeng Language. *Journal of American Oriental Society*, 4: 277-288.
- Maspéro, Georges. 1915. *Grammaire de la langue khmère (cambodgien)*. Paris, Imprimerie Nationale.
- Maspéro, Henri. 1912. Etude sur la phonétique historique de la langue annamite. Les initiales. *Bulletin de l'Ecole Française d'Extrême Orient* 12.1-27.
- Matisoff, James A.. 2003. Aslian: Mon-Khmer of the Malay Peninsula. *Mon-Khmer Studies* 33:1-58.
- Mitani, Yasuyuki. 1977. Palaung dialects: a preliminary comparison. *Tônan Ajia Kentyû (South East Asian Studies)*, 15.2: 193-212.
- Mitani, Yasuyuki. 1979. Vowel Correspondences between Riang and Palaung. *Studies in Thai and Mon-Khmer Phonetics and Phonology in Honour of Eugénie J.A. Henderson*. Chulalongkorn University Press, Bangkok, pp142-50.
- Morgan, Albert. 1885. Moers, coutumes et langages des négritos de l'intérieur de la presqu'île malaise. Sakayes et Sömañs. *Bulletin société normande géographique*. 7:411-55.
- Moura, Jean. 1878. *Vocabulaire français-cambodgien et combodgien-français*. Paris, Challamel Ainé.
- Nagaraja, K.S. 1985. *Khasi - A Descriptive Analysis*. Pune: Deccan College, Post-Graduate and Research Institute.
- Nai Pan Hla. 1976. A Comparative study of Old Mon Epigraphy. In Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.). *Austroasiatic Studies*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13). Part II, pp.891-918.



- Nissor Singh, U. 1906. *Khasi-English dictionary*. Shillong: E. Bangal and Assam Secretariat Press
- Noppawan Thongkham, *The phonology of Kasong at Khlong Saeng Village, Danchumphon sub-District, Bo Rai District, Trat Province*. MA thesis, Mahidol University, 2003.
- Paulsen, D. 1991, A phonological reconstruction of Proto-Plang, *Mon-Khmer Studies*, vol. 18-19, pp. 160-222.
- Peiros, Ilija. 1989. Dopolnenie k gipoteze S.A.Starostina o rodstve nostraticheskix i sinokavkazskix jazykov. *Lingvisticheskaja rekonstrukcija i drevnejshaja istorija vostoka: Materialy k diskussijam na mezhdunarodnoj konferencii (Moskva, 29 Maja - 2 Ijunja 1989g.)* Chast' 1. Moskva, Nauka.
- Peiros, Ilija. 1996. *Katuic comparative dictionary*. Canberra, Pacific Linguistics C-132.
- Peiros, Ilija. 1998. *Comparative Linguistics in Southeast Asia*. Pacific Linguistics Series C-142. Canberra, Australian National University.
- Phillips, Timothy C. 2005. *Linguistic Comparison Of Semai Dialects*. Unpublished manuscript. Economic Planning Unit, Prime Minister's Department, Malaysia.
- Pinnow, Heinz-Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden, Otto Harrassowitz.
- Pou, Saveros. 1992. *Dictionnaire vieux khmer-français-anglais: An Old Khmer-French-English Dictionary*. Paris, Centre de documentation et de recherche sur la civilisation khmère.
- Pou, Saveros. 2004. *Dictionnaire vieux khmer-français-anglais: An Old Khmer-French-English Dictionary*. Paris, Harmattan.
- Prachakij-karacak, Phraya. 1995. *Some Languages of Siam*. Translated and annotated by David Thomas & Sophana Srichampa. Bangkok, Mahidol University.
- Prasert Sriwises. 1978. *Kui (Suai)-Thai-English dictionary*. Bangkok, Chulalongkorn University.
- Proschan, Frank. 1996. A Survey of Khmuic and Palaungic Languages in Laos and Vietnam. *Pan-Asiatic Linguistics*, 3: 895-919.
- Reid, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33(2):323-344.
- Reid, Lawrence A. 1996. The current state of linguistic research on the relatedness of the language families of East and Southeast Asia. In Ian C. Glover & Peter Bellwood (eds.), *Indo-Pacific Prehistory: The Chiang Mai Papers*, Volume 2, Bulletin of the Indo-Pacific Prehistory Association 15. Canberra, Australian National University. pp. 87-91.
- Reid, Lawrence A. 2005. The current status of Austric: A review and evaluation of the lexical and morphosyntactic evidence. In Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas (eds.), *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London, Routledge Curzon. pp. 81-106.
- Röepstroff, F.Ad.De. 1884. *Dictionary of the Nancowry Dialect of the Nicobarese Language, in Two Parts: Nicobarese - English, and English - Nicobarese*. Ed. by Mrs. De Röepstroff. Calcutta
- Sagart, Laurent, Roger Blench and Alicia Sanchez-Mazas (eds.). 2005. *The peopling of East Asia: putting together archaeology, linguistics and genetics*. London, Routledge Curzon.
- Schebesta, Paul. 1926. The jungle tribes of the Malay peninsula. *Bulletin of the School of Oriental and African Studies*. 4:269-278.
- Schmidt, Pater Wilhelm. 1901. Die Sprachen der Sakai und Semang auf Malacca und ihr Verhältnis zu den Mon-Khmer-Sprachen. *Bijdragen tot de Taal-, Land-, en Volkenkunde van Nederlandsch-Indië*, 52:399-583.
- Schmidt, Pater Wilhelm. 1903. The Sakai and Semang languages in the Malay Peninsula and their relation to the Mon-Khmer languages. *Journal of the Straits Branch of the Royal Asiatic Society*, 39:38-45.

- Schmidt, Pater Wilhelm. 1904. Grundzüge einer Lautlehre der Khasi-Sprache in ihren Beziehungen zu derjenigen der Mon-Khmer-Sprachen. Mit einem Anhang: die Palaung-Wa-, und Riang-Sprachen des mittleren Salwin. *Abh. Bayrischen Akademie der Wissenschaft*, 1.22.3:677-810.
- Schmidt, Pater Wilhelm. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie*, Braunschweig, 5:59-109.
- Schmidt, Wilhelm. 1905. Grundzüge einer Lautlehre der Mon-Khmer-Sprachen. *Denkschrift der Akademie der Wissenschaften*, Wien, Philologisch-Historische Klasse 51:1-233
- Scott, James George, and J. P. Hardiman. 1900. *Gazetteer of Upper Burma and the Shan States*, Part 1, Vol. 1. Rangoon, Superintendent, Government Printing.
- Sebeok, Thomas A. 1942. An examination of the Austro-Asiatic Language family. *Language*, 18: 206-217.
- See Puengpa and Theraphan L-Thongkum. 1980. *A Bruu-Thai-English dictionary*. Bangkok, Chulalongkorn University Printing House
- Shafer, Robert. 1952. Études sur l'Austroasian. *Bulletin de la Société de Linguistique de Paris*, 48:111-158.
- Shafer, Robert. 1960. Review of Pinnow, 1959, Versuch einer historischen Lautlehre der Kharia-Sprache. *Word* 16:424-433.
- Shafer, Robert. 1965. Studies in Austroasian II. *Studia Orientalia edidit societas orientalis fennica* 30.5:3-69.
- Shelden, Howard. 1979. *A Reconstruction of Proto-South Bahnaric*. Ms. Essay submitted in partial fulfillment of course requirements in non-Indo-European languages, University of Texas at Arlington. Copy held at Summer Institute of Linguistics Library Bangkok.
- Shorto, Harry L. 1962. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press.
- Shorto, Harry L. 1963. The Structural pattern of northern Mon-Khmer languages. In H.L. Shorto, (ed.) *Linguistic Comparison in South-East Asia and the Pacific*. pp 45-61.
- Shorto, Harry L. 1965. The interpretation of archaic writing systems. *Lingua* 14. 88-97.
- Shorto, Harry L. 1971. *A dictionary of the Mon inscriptions, from the sixth to the sixteenth centuries, incorporating materials collected by the late C.O. Blagden* (London Oriental Series 24). London, Oxford University Press.
- Shorto, Harry L. 1976a. The Vocalism of Proto-Mon-Khmer. Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta (eds.) *Austroasiatic Studies*. Honolulu: University of Hawaii (Oceanic Linguistics, Special Publication, No. 13). Part II, pp.1041-1067.
- Shorto, Harry L. 1976b. In Defense of Austric. *Computational analyses of Asian and African languages* 6:95-104.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics.
- Shutler, Richard, Jr. and Jeff Marck. 1975. On the dispersal of the Austronesian horticulturalists, *Archaeology and Physical Anthropology in Oceania*, 10.2:81-113.
- Sidwell, Paul & Pascale Jacq. 2003. *A Handbook of Comparative Bahnaric: Volume 1, West Bahnaric*. Canberra, Pacific Linguistics 551.
- Sidwell, Paul. 1998. *A Reconstruction of Proto-Bahnaric*. University of Melbourne PhD thesis.
- Sidwell, Paul. 2000. *Proto South Bahnaric: a reconstruction of a Mon-Khmer language of Indo-China*. Canberra, Pacific Linguistics 501.
- Sidwell, Paul. 2005. *The Katuic Languages: classification, reconstruction and comparative lexicon*. Munich, Lincom Europa.
- Sidwell, Paul. 2006. Preface to Harry L. Shorto, *A Mon-Khmer Comparative Dictionary*. Canberra, Pacific Linguistics.

- Siripen Ungsitipoonporn, *A phonological comparison between Khlongphlu Chong and Wangkraphrae Chaong*. MA thesis, Mahidol University, 2001.
- Skeat, Walter W. and Charles O. Blagden. 1906. *Pagan Races of the Malay Peninsula*. 2 Vols. London: Macmillan.
- Smith, Kenneth, D. 1967. A Phonological Reconstruction of Proto Central North Bahnaric. *University of North Dakota Summer Institute of Linguistics Workpapers*.
- Smith, Kenneth, D. 1972. *A phonological reconstruction of Proto-North-Bahnaric*. Dallas, Language Data Series, Summer Institute of Linguistics.
- Smith, Kenneth, D. 1981. A Lexico-statistical study of 45 Mon-Khmer languages. In Andre Gonzalez and David Thomas (eds), *Linguistics Across Continents*. Manila SIL, pp 180-205.
- Sokolovskaia, N.K. 1978. Materialy k sravnitel'no-ëtimologicheskomu slovar'u v'etmyongskix jazikov. *Issledovanija po fonologii i grammatike vostochnyx jazikov*, Moskva, Nauka, pp126-80.
- Stevens, Edward. 1896. *A vocabulary, English-Peguan, to which are added a few pages of geographical names*. Rangoon, American Baptist Mission Press.
- Suwilai Premsirat. 2002. *Thesaurus of Khmu Dialects in Southeast Asia*. Mon-Khmer Studies-Mahidol University Special Publications No.1. Institute of Language and Culture for Rural Development, Mahidol University, Salaya.
- Svantesson, Jan-Olof. 1988. U. *Linguistics of the Tibeto-Burman Area*, 11.1:64-133.
- Theraphan L-Thongkum. 2001. *Languages of the Tribes in Xekong Province Southern Laos*. The Thailand Reseach Fund.
- Thomas, David & Robert Headley. 1970. More on Mon-Khmer subgroupings. *Lingua* 25:398-418.
- Thomas, David, & Marilyn Smith. 1967. Proto-Jeh-Halang. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 20: 157–175.
- Thomas, David. 1964. A survey of Austro-asiatic and Mon-Khmer comparative studies. *Mon-Khmer Studies* 1:49-163.
- Thomas, David. 1966. Mon-Khmer subgroupings in Vietnam. In Norman Herbert Zide, ed. *Studies in Comparative Austroasiatic Linguistics*. Indo-Iranian Monographs, V. The Hague, pp 194-202.
- Thomas, Dorothy M. 1967. *A Phonological Reconstruction of Proto-East-Katuic*. MA thesis, University of North Dakota.
- Thompson, Laurence C. 1976. Proto-Viet-Muong Phonology. In Jenner et al. (eds.) *Austroasiatic Studies*, Volume 2. Honolulu, University of Hawaii Press. pp 1113-1204.
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: two thousand years of language contact and change*. Oceanic Linguistics special Publications No. 28. Honolulu, University of Hawaii Press.
- Unchalee Singoi. 1988. *A Comparative Study of Pray and Mal Phonology*. MA thesis, Mahidol University.
- Watson, Richard, Sandra K. Watson and Cubuat. 1979. *Pacoh Dictionary: Pacoh-Vietnamese-English*. Trilingual Language Lessons, No.25, part 1. Manila, Summer Institute of Linguistics.
- Whitehead, G. 1925. *A Dictionary of the Car Nicobarese Language*. Rangoon.