

Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony

Author(s): James A. Matisoff

Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society (1990), pp. 543-559

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Bulging Monosyllables: Areal Tendencies in Southeast Asian Diachrony

James A. Matisoff
University of California, Berkeley

1. The complex monosyllable and intersegmental slopover.¹

Most language families of East and SE Asia are "monosyllabic" (in the sense that their morpheme stocks, not necessarily their words) consist overwhelmingly of monosyllables.² As I observed long ago, "There is something about the tightly structured nature of the syllable in monosyllabic languages which favors the shift in contrastive function from one phonological feature of the syllable to another."³ There is indeed much to be shifted around in these languages. We are not talking about such puny little (C)V(C) syllables as in Indonesian *rambut* 'hair' or Japanese *kokoro* 'heart',⁴ but rather such robust entities as Vietnamese *nguyên* [ɲwɤn] 'spring; source', Hmong *npluag* 'chaff' [the -g is a tonemark indicating a falling pitch with breathy voice], Thai *khriang* 'gear; apparatus', or Written Tibetan *bsnyigs* 'sediment'.

Sometimes there is a morphological basis for a syllable's phonological complexity, as in the Aslian branch of Mon-Khmer (Malaya), where various derivational processes involve the infixation of a copy ("incopyfixation") of the syllable-final consonant, leading sometimes to spectacular consonant clusters:⁵

(1) with incopyfix of final alone (roots with initial clusters)

Ci Cii V Cf ----> Ci Cf Cii V Cf
Cheq Wong (N. Asl.) *hwæc* 'whistle' / *hcwæc* 'whistling'

(2) with root-external infix plus incopyfix

(a) simple initial

Ci V Cf ----> Ci Cf N V Cf
Semai (C. Asl.) *te:w* 'river' (mass) / *twne:w* [tuni:w] 'river' (count)

(b) cluster initial

Ci Cii V Cf ----> Ci N Cf Cii V Cf
Semai *sla:y* 'swidden' (mass) / *snyla:y* [snila:y] 'id.' (count)

(3) with root-external prefix plus incopyfix

Ci V Cf ----> N Cf Ci V Cf
Semelai (S.As.) *kap* 'bite' / *npkap* 'biting'

(4) with reduplication of the initial and incopyfixation of the final

Ci V Cf ----> Ci Cf Ci V Cf

Bateg (N. Asl.) kwɔ 'grate' / kwɔkwɔ 'is grating'

Jah Hut (C. Asl.) ɔaʔ 'eat' / ɔʔɔaʔ 'is eating'

Semai laal 'stick out tongue' / lllaal [lɛlla:l] 'id.' (prog.)

Temiar (C. Asl.) hɔɔh 'follow' / hhhɔɔh [hɛhhɔɔh] 'is following'

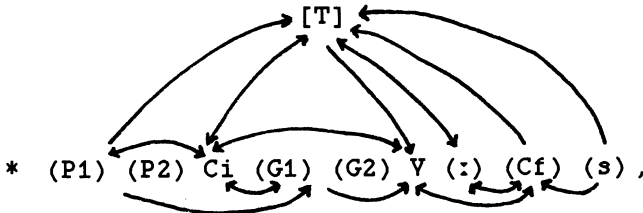
Note that these languages do not shrink from applying this process even when the root-initial and -final consonants are identical (as in the last two examples), even though this leads to four occurrences of the same consonant in the syllable!

Tibeto-Burman has never had a system of infixes⁶, yet the Proto-Tibeto-Burman [PTB] syllable canon was certainly complex enough:

[T]

* (P1) (P2) Ci (G1) (G2) V (:) (Cf) (s) ,

where P = prefix, Ci = initial consonant, G = glide, V = vowel, : = vowel length, Cf = final consonant, s = suffixal -s, and T = tone (which I consider to have been phonetically present but phonologically redundant at the proto-stage).⁷ Both synchronic and diachronic evidence shows that virtually all parts of such syllables are capable of influencing every other. We may crudely indicate this by inserting ligatures in the above formula:



A few examples of some of these types of interrelationship:⁸

(a) Cf > V

<i>PLB</i>		<i>Lahu</i>		<i>PLB</i>		<i>Lahu</i>
*-am	→	-o		*-ap	→	-oʔ
*-an	→	-e		*-at	→	-eʔ
*-aŋ	→	-ɔ		*-ak	→	-aʔ

e.g. fathom	*lam ¹	lò	rub	*sap	šò?
slave	*gywan ¹	cè	leech	*k-r-wat	vè?
you	*naŋ ¹	nò	rat	*k-r-wak	fâ?

Particularly interesting is the "circular" development shown by the PTB rhyme⁹ *-ik in the history of Burmese, where the feature of palatality has shifted back and forth from the vowel to the final consonant to the vowel:

PTB *-ik → W[ritten] B[urmese] -ac → Mod. Bs. I?,
e.g. PTB *tsik 'joint' (cf. Written Tibetan [WT] tshigs)
→ WB chac, Mod. Bs. hsi?

(b) G > Ci

*k → Lh. qh	PLB *ka ²	→ Lh. qhá	'bitter'
*kr → Lh. kh	PLB *krew ¹	→ Lh. kho	'horn'
*ky → Lh. ch	PLB *kyow ¹	→ Lh. cho	'sweet'
*kw → Lh. ph	PLB *kwey ²	→ Lh. phé	'dog'

(c) G > V

PLB	Lahu	Example
*-a	-a	*n-da ¹ 'fern' → Lh. dà
*-ya	-e	*bya ² 'bee' → Lh. pê
*-wa	-u	*twa ¹ 'handspan' → Lh. thu

(d) Ci > G

The initial consonant may provoke an assimilatory change in a following glide, as in the following example of "rhinokinesis":

PTB *s-myak 'eye'	→ Gyarong	temák		
[+nas]	[-nas]	>	[+nas]	[+nas]
	[+pal]			[+pal]

(e) Ci <> V

The interinfluence of initial consonant and vowel is well illustrated, e.g. on the allophonic level in Lahu. Least remarkable are unidirectional cases like Lh. /n/ → [ñ] /---i, e.g. /nī/ [ñī] 'penis', where the high front vowel palatalizes the preceding consonant -- similar rules might be expected in any language family. More striking are *bidirectional* instances of mutual conditioning of the initial and the vowel, such as obtains with four labials before /u/ and five palatals before /ɛ/:

/p ph b m/	→ [pf pfh bv mv] / -----u
/c ch j š ʃ y/	→ [ts tsh dz s z] / -----ɛ

These vowels themselves have special realizations here, with /u/ unrounding to [u] and /ɨ/ being raised to the "superhigh apical buzz" [ɿ]. Thus, /pû/ 'insect' [pʰɿ], /yɨʔ/ 'sleep' [zɿʔ]. The marriage between initial and vowel is carried to extremes in the syllable /mu/, where the vowel, after affricating the initial, is usually swallowed up entirely, yielding a syllabic nasalized labiodental spirant: /mù/ 'mushroom' [m̥ʷ].

(f) Ci ... Cf > V

Sometimes both the pre- and post-vocalic consonants contribute phonetic material to the intervening vowel, as in syllables showing "rhinoglottal transfer" in the S. Loloish language Mpi. There are no fewer than nine Mpi syllables which derive from PLB etyma with a nasal root-initial and the rhyme *-ak.¹⁰ In all these cases, the vowel has acquired a complex quality containing both a nasal and a laryngeal component, donated by the pre- and post-vocalic consonants respectively.

PLB *s-nak 'dream' > Mpi **mãʔ** (written **maŋʔ**)
 PLB *s-nak 'black' > Mpi **nãʔ**
 PLB *ʔ-ŋak 'open' > Mpi **ŋãʔ**

(g) : > Cf

Vowel length sometimes exerts a crucial influence on a post-vocalic consonant. Proto-Aslian nasals develop into Sabum homorganic stops after short vowels, but into a characteristic type of "decomposed nasals" after long vowels:

PAslian *kan → Sabum **kap**
 *kaan → **kaabm** (with "nasal decomposition")

Here the timing of the velic opening is slowed down by the extra mora of the vowel, so that a stop articulation is audible before the final nasal segment.

Similar to this is a phenomenon in Maru (Burmish branch of Lolo-Burmese), where the two open PTB rhymes with long high vowels have developed secondary final stops:¹¹

PTB *-iɿ or *-eɿ > Maru **-it**
 PTB *-uɿ or *-əɿ > Maru **-uk,**

e.g. 'four' PTB *b-liɿ > Maru **bit**, 'steal' PTB *r-kuɿ > Maru **khúk**

(h) P > T ; C_i > T ; C_r > T

Perhaps the most remarkable kind of intersegmental feature shuffling is involved in the development of phonemic tone, or *tonogenesis*.¹² In a word, contrastive tone arises as a compensatory mechanism for mergers or losses in the systems of pre- and/or post-vocalic consonants of syllables, especially the loss of a prefix or the neutralization of a voicing contrast in pre-vocalic position, or the loss of a laryngeal segment (-h or -ʔ) after the vowel. The more thoroughly monosyllabic a language family is, the more "tone-prone" it seems to be.

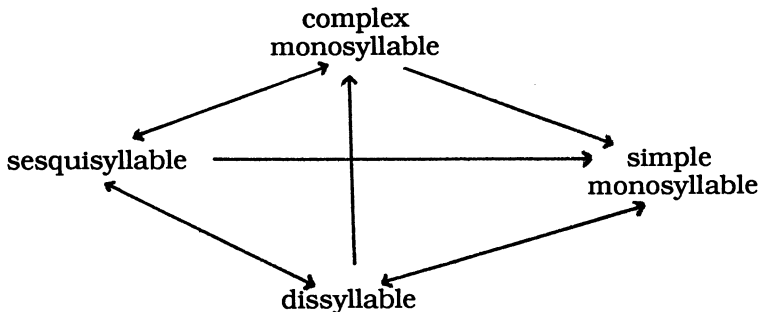
As one example, the high-rising tone /' / of Lahu can be shown to have arisen through a dissimilatory upheaval in syllables that both began and ended with a "glottal incident", e.g.

PTB *s-kok 'outer covering' > PLB *ʔguk > pre-Lahu *ʔguʔ > Lh. qú
 PTB *s-tsik 'joint' > PLB *ʔdzik > pre-Lahu *ʔdziʔ > Lh. cǎ

Several branches of Mon-Khmer have developed a similar suprasegmental compensation known as *register* for the loss of a consonantal voicing contrast: a multiplication of the "phonation types" or kinds of voicing with which the vowel may be pronounced. The simplest register systems are two-way, e.g. contrasts between "plain" and "breathy" voice, or between "plain" and "creaky", or "breathy" vs. "creaky". Other languages have 3- or even 4-way systems which rival true tone systems in complexity.¹³

2 The expansion/contraction cycle: monosyllables, dissyllables, and sesquisyllables.

It is my central thesis that the lexemes of Southeast Asian languages undergo a millennial dance from one type of syllable-structure to another, oscillating cyclically among (1) (consonantly) *complex monosyllables* (2)(consonantly) *simple monosyllables* (3) *dissyllabic compounds* or tight collocations and (4) *sesquisyllables*.¹⁴ These developments may be diagrammed crudely as follows:



It would be too much to expect that we could trace a given morpheme through this whole cycle -- the period of oscillation is, after all, rather slow. Yet in favorable circumstances the same etymon may indeed be attested at several different stages of the cycle in a given language family. So many solid examples exist of passage from one "contiguous" stage to another, that it seems logical to believe in the reality of the entire cycle itself.

Most of these interrelationships are bidirectional, as indicated by the arrows in the diagram. We will briefly take them up one by one.

2.1 **Contractive: from complex-consonant to simple-consonant monosyllables.**

The documented history of many languages (including Chinese, Tibetan, and Burmese) shows sometimes radical simplification of an earlier consonantism: e.g. Old Chinese **tsywet** 'emerge' > Mandarin **chū**; Written Tibetan **brgyad** 'eight' > Mod. Lhasa **ce**; Written Burmese **krwat** 'leech' > Mod. Rangoon **cwa?**. This has been carried to an extreme in branches of the family like Loloish, where the rich syllable canon set up for PTB [see above] has been eroded, e.g. in Lahu to

T
(C) V ,

as in 'snake' WB **mrwe**, Lh. **vɿ**; 'rat' WB **krwak**, Lh. **fâ?**; 'eight' WB **hrac**, Lh. **hí**. Note, however, that compensatory mechanisms -- proliferation of vowels and/or tones [see above 1(h)] -- usually operate to counteract the consonantal impoverishment.¹⁵

2.2 *Expansive: from complex monosyllables to sesquisyllables.*

This development has nothing characteristically SE Asian about it, since it is basically the low-level phonetic phenomenon of *epenthesis* that is at work -- the breaking up of hard-to-pronounce consonant clusters by shwa-insertion. It seems reasonable to assume that many of the orthographic clusters of a language like Written Tibetan were broken up in actual speech by such epenthetic vowels, so that e.g. **brgyad** 'eight' must have been pronounced something like [brəgyat].¹⁶

2.3 *Expansive: from monosyllables to dissyllables*

Radical consonantal simplification can pose a serious problem for monosyllabic languages in terms of *pernicious homophony*. Keeping potentially homophonous morphemes distinct requires adding some "phonological bulk" or redundancy. One way to achieve this is to increase the number of syllables per word. Occasionally this is done by meaningless "mechanical" extrusions of the monosyllable's phonetic substance (2.31); usually, however, the strategy is to create dissyllabic compounds or collocations where each of the constituent syllables is fully meaningful (2.32).

2.31 *Dissyllables without morpheme boundary.*

Some TB languages have developed "echo vowels" which serve to resyllabify an original monosyllable into a 2-syllable sequence. The PTB root ***p-wak** 'pig' has become the dissyllabic form **wo-ko** in Lotha Naga. Similarly, PTB ***lam** 'road' has developed into **la-na** in Kokborok (Bodo-Garo group) and to **le-nu** in Khoa (Abor-Miri-Dafla group). In Garo, a medial vowel is sometimes split into two morae separated by glottal stop: PTB ***g-sat** 'kill' > Garo **soʔot**.

2.32 *Dissyllables with morpheme boundary: compounding and affixation.*

Much more important as a strategy for providing compensatory phonological bulk is the process of *compounding*. Many dialects of Southern American English have merged the rhymes /-in/ and /-en/, rendering pairs like *pin* / *pen*

homophonous. To disambiguate such cases, these dialects have introduced compounds like *stick-pin* / *ink-pin*. East and Southeast Asian languages have proceeded in exactly analogous fashion.

Chinese is an excellent case in point. Classical Chinese, with its relatively rich consonantism, was strictly monosyllabic, with the syntactic word and the phonological syllable virtually coextensive. In phonologically eroded modern dialects like Mandarin, however, the vast majority of *words* are now dissyllabic, though almost all of them are still analyzable into monosyllabic constituent morphemes.¹⁷

In Loloish also, the consonantal simplification of monosyllables has led to homophony on a grand scale. There are e.g. at least five Lahu morphemes pronounced **ha**¹⁸ which descend from once consonantly distinct PTB etyma:

	PTB	PLB	WB	Lahu monosyllable	Lahu dissyllabic collocations
'hundred'	*b-r-gya	*ʔra ¹	ra	ha	tê ha
'moon'	*s-gla	*s-la ³	la'	ha	ha-pa
'tongue'	*s-lyā	*s-l(y)a ¹	hlyā	ha	ha-tē
'spirit, likeness, shadow'	*s-hla	*sla ³	hla' 'beautiful'	ha	ò-ha
'winnow'	*g-ya(:p)	*ʔ-ya ¹	Akha zá	ha	ha ve

[**ha** 'hundred' is not usable by itself, but functions as a 'round-number classifier' which must always be preceded by a numeral (e.g. **tê ha** 'one hundred'); the **-pa** in 'moon' is a meaningless noun suffix, ubiquitous in TB (cf. WT **ɣla-ba** 'moon')¹⁹; the **-tē** in 'tongue' looks like it once had an independent meaning, but recurs nowhere else in the language; the **ò-** prefix in 'spirit' (< PTB *ʔaŋ-) occurs as a bulk-provider before hundreds of Lahu roots; the particle **ve** in **ha ve** 'to winnow' is a nominalizer that occurs in the citation form of verbs (much like Eng. *to*), serving to distinguish verbs from any homophonous nouns.]²⁰

2.4 Contractive: from dissyllables to monosyllables.

Once we have a dissyllabic structure in close phonological juncture, the stage is set for the reverse swing of the cyclical pendulum: the fusion of the two syllables into one. Several subtypes of "trans-syllabic absorption" may be distinguished:

Sometimes the two syllables have been so thoroughly fused that they can only be teased apart by comparative evidence. I have explained the peculiar Angami Naga form *pfhə* 'bitter', where the labiodental initial stands out sharply against the simple velar attested everywhere else in TB (< **ka*), in terms of the fusion of an original dissyllable **ka-ba* (where *-ba* was a nominalizing citation particle):

**ka-ba* → **ka-wa* → **kwa* → *pfhə*.²¹

The Garo word for 'I; me' is *aŋ*, whereas most other TB languages have forms which directly reflect **ŋa*. Metathesis will not do as an explanation! It seems more likely that the well-attested TB pronominal prefix **a-* was combined with this morpheme, after which the final (originally root-)vowel was dropped:

**ŋa* → **a-ŋa* → *aŋ*

Speaking of pronominal fusions, the Mandarin pluralizing suffix *-men*, used only with pronouns and a few nouns referring to human beings, has lost its rhyme and fused its initial with the vowel of the preceding pronoun in rapid colloquial: *wō-men* 'we' > *wōm*, *nǐ-men* 'you pl.' > *nǐm*, *tā-men* 'they' > *tām*. This has incidentally had the effect of marginally reintroducing final *-m* into the Mandarin syllable canon, from which it had disappeared centuries ago by merging with final *-n* (cf. Cantonese *saam*, Mand. *sān* 'three').

New data on the Qiangic languages (TB of Sichuan) shows that the Northern dialects have a strong tendency to apocopate the vowel of the 2nd elements of compounds, leading to secondary monosyllables with final consonants (e.g. voiced spirants) that are sometimes highly untypical of TB:

	S.Qiang (Taoping)	N.Qiang (Mawo)
'seed'	<i>zue-pə</i>	<i>zəp</i>
'day after tomorrow'	<i>zue-za</i>	<i>tʃhaz</i>
'stove'	<i>sɣ-dɣ</i>	<i>syt</i>
'fifteen'	<i>xa-ŋa</i>	<i>haŋ</i>

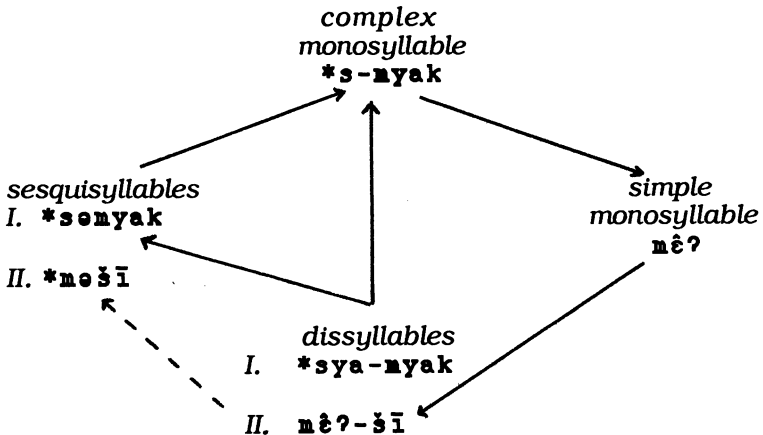
When the 2nd syllable in a collocation begins with a vowel, and when its meaning is grammaticalized or abstract (which facilitates destressing), it is susceptible of being absorbed into the vocalic nucleus of a preceding open syllable. This often leads to a "bulging" of the vowel until it is diphthongal, or at least "sesquimoral" -- a mora and a half long.²²

2.5 Contractive: from dissyllables to sesquisyllables.

Perhaps the most interesting stage in the cycle is the one where we can catch a dissyllabic compound in the process of being reduced to a sesquisyllable, but before the semantic content has quite disappeared from the unstressed minor syllable -- or at any rate while it may still be deduced from comparative data.²³ Much as the heart of a chivalrous knight would be quickened by hapless damsels in days of yore, so might we well be moved by the plight of these poor "syllables in distress". Let us quickly chase a few of them around the cycle:

2.51 EYE

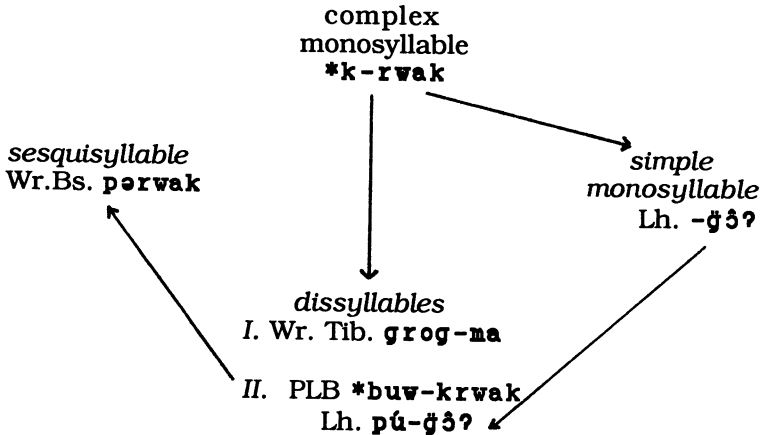
Pre-PTB ***sya-myak** > PTB/PLB ***səmyak** or ***s-myak** > Lahu **mê?** > Lahu **mê?-šī** > hypothetical future ***məšī**



Here the only directly attested forms are the Lahu ones, yet there is good evidence for all the other stages. The high

stopped tone of **mɛʔ** indicates a PLB ***s-** prefix before the nasal [Matisoff 1972, pp. 24, 58-61]. This prefix with bodypart terms is generally explained as deriving from the full morpheme ***sya** 'flesh; animal' [Benedict 1972, p. 106]. It may be supposed that the dissyllable ***sya-myak** passed through a sesquisyllabic stage ***səmyak** during the "prefixization" process. Meanwhile, the monosyllabic form **mɛʔ** is replaced in most contexts by the dissyllabic compound **mɛʔ-šī** in modern Lahu, where the second element means 'round object'. We may confidently expect that in several hundred years this compound will itself degrade into a new sesquisyllable, perhaps ***məšī**, as the carousel goes around once again.

2.52 ANT



The first syllable of the Wr.Tib. form directly reflects the PTB prototype with prefixal velar (one of the generally accepted "animal prefixes" of TB). The high-stopped tone of the Lahu syllable **-ḡḡʔ** also reflects the ***k-** prefix [Matisoff 1972, pp. 68-70], although this syllable does not occur independently in modern Lahu. Instead we find the compound **pú-ḡḡʔ**, where the first element is definitely a sandhi variant of the morpheme **pú** 'bug' (< PTB ***bəw**), which occurs in a large number of insect names. This compound formation, innovative with respect to TB as a whole, must be referred back to the PLB stage, since a sesquisyllable with this same element is directly attested in Burmese, both in the orthography and in the modern pronunciation /pəyweʔ/.

2.53 SON-IN-LAW

Wr.Bs. **semak** / Wr.Tib. **mag-pa**, Dhimal hma-wa

sesquisyllable

Wr.Bs. **semak**

*simple
monosyllable*

***ma:k**

dissyllables

Wr.Tib. **mag-pa**

PLB ***za-nak(-pa)**

PLB ***ʔaŋ-nak(-pa)**

Lahu **ò-má(-pā)**

Here the proto-monosyllable ***ma:k** was quite simple to begin with, so that at an early stage it was compounded with various morphemes in the daughter languages, especially with the suffix **-pa** (which after nouns referring to humans functioned sometimes as a masculine morpheme contrasting with **-na** 'feminine'). The minor syllable of the directly attested sesquisyllabic form in Burmese seems obviously to be a reduction of the full morpheme ***za** 'son; child'. This time the Lahu forms do not reflect the same formation as in Burmese, but only the meaningless prefix ***ʔaŋ-**, with the masculine suffix **-pā** sometimes thrown in for additional redundancy.

2.54 LUNGS

*complex
monosyllable*
Lushai **tśuap**

sesquisyllable

***tśəwap**

*simple
monosyllables*

***tsi** 'lung'

***-wap** 'spongy'

dissyllables

pre-Lushai ***tsi-wap**

Jingpho **sin-wóp**

pre-Lahu ***ʔaŋ-tsi(-pwap)**

Lahu **ò-chî(-phô?)**

I have demonstrated at length that the complex monosyllabic form for 'lungs' in Lushai (Kuki-Chin branch of TB), **tśuap**, actually derives from a dissyllabic prototype, ***tsi-wap**, where the 2nd element is an adjectival root meaning 'spongy; porous'.²⁴ This is made clear by the Jingpho compound **sin-wóp** 'lungs' ("spongy liver"), where the first element derives from PTB ***n-sin** 'liver', and which is paradigmatically opposed to the compound **sin-jàʔ** 'liver' (lit. "solid liver"). The 2nd syllable of the Lahu form reflects ***tsi**, while the 3rd syllable might derive from a prefixed variant of the morpheme for 'spongy', ***p-wap**.²⁵

2.6 *Contraction: from sesquisyllables to monosyllables*

Finally, we may close the circle by noting that there are multiple paths by which a sesquisyllable may be reduced to a monosyllable.

Most commonly perhaps, the minor syllable is simply lost by procope. This is what happened regularly in the history of Vietnamese, once a sesquisyllabic Mon-Khmer language that became monosyllabic and highly tonal under massive and prolonged Chinese influence.

Alternatively, the shwa vocalism of the pre-syllable may disappear by syncope, creating a complex monosyllable, which is probably the scenario leading to the Lushai form for 'lungs' just discussed.

Still another possibility is a development I have called *prefix preemption*,²⁶ which occurs especially when the initial of the major syllable is a "weak" consonant (a liquid, semivowel, or nasal). In these cases the initial of the prefixal or "minor" syllable may drive it out entirely. Many examples may be given from TB (cf. Maru **bit** 'four' < PTB ***b-loy**, cited in I(g), above), and the same phenomenon may be noted in colloquial Siamese:

məlèt 'seed' (formal) > **mèt** (colloq.)
mələcɛŋ 'bug' (formal) > **mɛcɛŋ** (colloq.)

3. L'économie de la syllabe: decay and rebirth.

In conclusion, let me repeat a passage with which I ended another recent discussion of this general topic:

"The monosyllabic languages of East and Southeast Asia show an uncanny homeostatic ability to regulate themselves in cyclic swings of expansion and contraction. What is absorbed and incorporated *here* will be diffused or extruded *there*.

"The accretional or augmentative tendencies do not of course stand in a simple one-to-one replacement relationship versus the tendencies toward reduction and attrition. Things are more indirect and slow-moving than that. Nonetheless, it is hard not to believe in some kind of overarching regulatory principle which eventually ensures that things will not go too far in any one direction. There is no harm in referring to this by some functional label like 'the economy of the syllable'.

"In a more cosmic vein, these phenomena furnish one more bit of reassuring evidence that the forces of creativity have nothing to fear from the forces of destruction."²⁷

NOTES

¹A version of this paper was presented at the 15th annual meeting of the Berkeley Linguistics Society in February 1989. My thanks to John B. Lowe for help in formatting the present version. Some of this material is based upon work done at the Sino-Tibetan Etymological Dictionary and Thesaurus (STEDT) Project, supported by the National Science Foundation under Grant No. BNS-867726 and by the Division of Research Programs of the National Endowment for the Humanities, an independent federal agency, under Grant No. RT-20789-87.

²These monosyllabic families include Sino-Tibetan/Tibeto-Burman, Tai-Kadai, Hmong-Mien (Miao-Yao), and the Viet-Muong branch of Austroasiatic (Mon-Khmer). The rest of Austroasiatic is largely "sesquisyllabic" (comprising morphemes "one syllable and a half" in length)/[see below]. Only Austronesian and Japanese contain large numbers of truly dissyllabic morphemes.

³Matisoff 1973, p. 78. Henderson (1985) later dubbed this phenomenon "feature shuffling."

⁴There seems to be a rough inverse correlation between the typical number of syllables in a language's morphemes and the phonetic complexity of the individual syllables.

⁵These are elegantly discussed in Diffloth 1973, though I alone am responsible for coining the ugly term "incopyfix."

⁶Pace Miller 1958.

⁷When prefixes were present we may suppose that they were often vocalized by a following shwa, so that such syllables were phonetically "sesquisyllabic" [below].

⁸We are using the symbol ">" to indicate the direction of influence. "PLB" is Proto-Lolo-Burmese, one of the major subgroups of PTB. Lahu is a member of the Central Loloish branch of Lolo-Burmese.

⁹The traditional term "rhyme" refers to the entire syllable except for the initial consonant. The "glides" /w y r l/ are inherently ambiguous with respect to this concept, sometimes functioning as if they were part of the initial but sometimes as if part of the vocalic nucleus. Thus,

PLB *wak 'pig' > Lh. vǎʔ, but PLB *twak 'emerge' > Lh. tǔʔ

PLB *s-ya¹ 'winnow' > Lh. ha, but PLB *hya¹ 'swidden' > Lh. he.

¹⁰Matisoff 1978b, pp. 22-4. All examples found so far reconstruct with either *s- or *ʔ-, two prefixes which are especially similar in their diachronic behavior in Lolo-Burmese.

¹¹These rhymes were reconstructed as *-iy and *-uw in the original version of Benedict 1972, but were later reinterpreted as *-əy and *-əw. Under either analysis they are functionally equivalent to bimoraic vowels.

An analogous process is posited for the development of the Old Chinese final semivowels *-w and *-y into the Middle Chinese final voiced stops *-g and *-d. See Benedict 1948.

¹²The term *tonogenesis* was first used in Matisoff 1970, 1973, though the phenomenon itself had been explained long before by Haudricourt (1954).

¹³Diffloth [p.c.] reports that the Chong language (of the Pearic branch of Mon-Khmer, spoken in Cambodia) has a 4-way contrast among clear / breathy / creaky-breathy / and "forced" (i.e. tight and faucalized) phonation types.

As an adjunct or alternative to "registrogenesis" many Mon-Khmer languages have developed highly complex vocalic systems to compensate for consonantal mergers. Khmer itself has 31 vocalic nuclei (including some triphthongs), while Bru (Katuic group, Vietnam) has no fewer than 41.

¹⁴The term *sesquisyllable* was introduced in Matisoff 1973 to refer to words that are "a syllable and a half" in length. Sesquisyllables consist of a fully stressed "major syllable" preceded by an unstressed "minor syllable" that usually has shwa-vocalism (e.g. *phənom*, *robiap*, *məlet*). This sort of syllable structure is especially characteristic of the Mon-Khmer family (except for Vietnamese, which has become strictly monosyllabic under Chinese influence), though it is also widespread in Tibeto-Burman and occasionally encountered in Tai-Kadai.

¹⁵In addition to these phonemic compensations (transphonologizations) resulting from regular sound change, many Lahu syllables are more complex phonetically than this bare-bones canon would suggest. Loanwords, affective vocabulary, and morphophonemic processes like vocalic fusion combine to reintroduce such features as secondary nasalization, glottalization, labial glides, long vowels, and diphthongs into the language. Any of these historically secondary features might acquire considerable importance in the future. "Once some feature is present phonetically in a SE Asian monosyllable, no matter how

redundant or trivial it may appear, it is available for future exploitation and transphonologization" [see Matisoff 1989a].

¹⁶These syllables are crucially different from those much more interesting sesquisyllables to be considered below (2.5), where the minor syllable derives from a recognizable separate morpheme.

¹⁷In a discussion of homophony in an encyclopedia article on Chinese, Y.R. Chao once concocted three little stories in Classical Chinese style consisting entirely of monosyllabic words pronounced homophonously in Mandarin (**shí**, **ji**, and **yi**, respectively, under various tones). For the Old Chinese listener, these stories would have been understandable even orally, since most of the syllables were still pronounced differently. Thus the three words of the title of the story *Ten Stone Lions* (Mand. **Shí Shí Shí**) were pronounced something like ***Dyep Dyak Syer** in OC. For the stories to be auditorily understandable to a modern Mandarin speaker, they would have to be recast using dissyllabic compounds or collocations to differentiate the now individually homophonous syllables (e.g. **shí-ge** 'ten', **shí-tou** 'stone', **shí-zi** 'lion').

¹⁸Besides sharing the same initial and vowel, these syllables are also *tonally* homophonous, all being under the mid-tone (unmarked in the transcription). Many other **ha**-morphemes occur under the other 6 tones!

¹⁹A homophonous suffix **-pa/-ba** occurred as a nominalizer after verbal roots. See the discussion of the Angami form for 'bitter', below 2.4.

²⁰Cf. also **ha-ke** 'small winnowing tray', **ha-ma(-qō)** 'large winnowing sieve'. It should be noted that many TB compounds are 3, 4, or even more syllables long, as in the "long form" of the word for 'large winnowing sieve', **ha-ma-qō**, where the 2nd syllable is a recurrent but meaningless suffix and the 3rd syllable means 'concave object'.

²¹See Matisoff 1982, p. 24. The development of labiodentals from labiovelars is well attested elsewhere in TB, e.g. PTB ***kwəy** 'dog' > Lh. **phǎ** [see above 1(b)].

²²I have recently discussed this topic at length elsewhere (Matisoff 1989a, 1989b). Falling into this category are the innumerable syllables of Pekinese Mandarin with retroflex vowels that derive from a diminutive suffix that is itself an unstressed variant of the full morpheme **ēr** 'child', e.g. **niǎor** 'birdie', **xiǎo-háir** 'child', **guò-tiēr** 'potsticker dumpling'.

²³The loss of morphemic identity from one or both of the constituents of an original compound is a danger which every language faces, with many familiar examples citable from the history of English: *hussy* (< HOUSE + WIFE), *bonfire* (< BONE + FIRE), *nostril* (< NOSE + THYREL [obs.] 'hole'), *window* (< WIND + EYE), *daisy* (< DAY's + EYE), etc.

²⁴See Matisoff 1978a, pp. 113-23.

²⁵The examples given so far are all from TB, though similar cases of sesquisyllabization of compounds can easily be found in e.g. Tai-Kadai. Thus the minor syllable of Siamese **sədəa** 'navel' is derivable from **səaj** 'line; band' via the umbilical cord, while the unstressed **mə-** in many names of fruits and vegetables (e.g. **məntəaj** 'mango', **məphráaw** 'coconut', **məkhyə** 'eggplant') is a reduction of PTai ***hmaek** 'fruit'. Cf. Li 1977, pp. 92, 75.

²⁶See Matisoff 1979, p. 24.

²⁷Matisoff 1989a, p. 185.

REFERENCES

- Benedict, Paul K. 1948. "Archaic Chinese *g and *d." *Harvard: Journal of Asiatic Studies* 6, 313-37.
- _____. 1972. *Sino-Tibetan: a Conspectus*. Contributing editor, James A. Matisoff. Cambridge: Cambridge University Press.
- Diffloth, Gérard. 1973. "Minor syllable vocalism in Sinoic languages." In Philip N. Jenner, et al., eds. *Austroasiatic Studies, Part I*. (Pacific Linguistics, Spec. Pub. #13.). Honolulu: University of Hawaii Press, 229-47.
- Haudricourt, André-Georges. 1954. "De l'origine des tons en vietnamien." *Journal Asiatique* 242, 68-82.
- Henderson, Eugénie J.A. 1985. "Feature shuffling in Southeast Asian languages, or How big is a segment?" In S. Ratanakul, D. Thomas, and S. Premisrat, eds., *Southeast Asian Linguistic Studies Presented to André G. Haudricourt*, pp. 1-22. Bangkok, Mahidol University.
- Li Fang-kuei. 1977. *A Handbook of Comparative Tai*. Honolulu: University of Hawaii Press.
- Matisoff, James A. 1970. "Glottal dissimilation and the Lahu high-rising tone: a tonogenetic case-study." *Journal of the American Oriental Society* 90.1, 13-44.
- _____. 1972. *The Lolotsh Tonal Split Revisited*. Research Monograph #7, Center for South and Southeast Asia Studies, University of California, Berkeley.
- _____. 1973. "Tonogenesis in Southeast Asia." In L.M. Hyman, ed., *Consonant Types and Tone*. (Southern California. Occasional Papers in Linguistics #1.) Berkeley and Los Angeles, University of California Press, 71-96.
- _____. 1978a. *Variational Semantics in Tibeto-Burman*. Philadelphia: ISHI Publications.
- _____. 1978b. "Mpi and Lolo-Burmese microlinguistics." *Monumenta Serindica* (Tokyo), #4. 36 pp.
- _____. 1979. "Problems and progress in Lolo-Burmese: Quo Vadimus?" *Linguistics of the Tibeto-Burman Area* 4.2, 11-43.
- _____. 1982. "Proto-languages and proto-Sprach-gefühl." *Linguistics of the Tibeto-Burman Area* 6.2, 1-64.
- _____. 1989a. "The bulging monosyllable, or *The mora. the merrier*: echo-vowel adverbialization in Lahu." In J.H.C.S. Davidson, ed., *South-East Asian Linguistics: Essays in honour of Eugénie J.A. Henderson*. London: School of Oriental and African Studies, 163-97.
- _____. 1989b. "The three palatal suffixes of Tibeto-Burman." Talk presented at 22nd International Conference on Sino-Tibetan Languages and Linguistics, University of Hawaii, Honolulu (October 7).
- _____. 1989c. "Tone, intonation, and sound symbolism in Lahu: loading the syllable canon." *Linguistics of the Tibeto-Burman Area* 12.2, 147-63.
- Miller, Roy Andrew. 1958. "The Tibeto-Burman infix system." *Journal of the American Oriental Society* 78.3.