Towards a History of Mon-Khmer: Proto-Semai Vowels

Gérard DIFFLOTH*

The presence of fifteen Mongolic or Tungusic languages in the center of Kyushu would be a major element in appreciating the history of the Far-East. No such languages have been found, but the Malay peninsula presents a case of this sort: there are fifteen Mon-Khmer languages spoken today in a region which is a part of the Austronesian (Malayo-Polynesian) language and culture area. The historical implications of this fact are still a complete mystery.

The structure and relationships of some of these Aslian languages have become a little clearer in recent years (Benjamin 1973a, b; Diffloth 1968, 1973b, 1975) but we still know very little about the history of the societies where they are used. There are several reasons for this: the usual methods of document-based history cannot apply to these non-literate cultures, archeologists have not been able to assign their finds to the ancestors of present-day populations, and anthropologists tend to be historically shy, perhaps for lack of a sound method.

There is, however, a historical approach for which we have an endless supply of documentation: the linguistic comparative method. Admittedly, it has its limitations: for instance, it usually does not provide us with an absolute but only a relative chronology; moreover it is most rigourous in areas like phonology which have but little intellectual content, and becomes more hazardous in the areas of syntax and semantics which the historian of ideas is more interested in.

And yet, it also has clear advantages: language change being mostly unconscious, linguistic history is not the victim of systematic distortions and propaganda as document-based history often is. Over archeology, historical linguistics has the great advantage of penetrating the more mental areas of culture and not simply the remaining material traces of human thought.

It is in this context that the history of Semai vowels which I propose here should be viewed: certainly not as an end in itself but as a beginning; after all, being able to find the exact position of someone's tongue several centuries ago is a kind of small feat, but hardly a necessity. However, historical phonology is irreplaceable because it gives us a solid chronological frame upon which we can build the history of morphology, of syntax and, last but not least, of semantics, i.e. etymology in the wider sense. Historical phonology is to the etymologist what dating is to the document-historian. The

---

* Department of Linguistics, University of Chicago

present paper can serve as a backbone for Aslian etymology, one step towards a history of Mon-Khmer civilization.

* * *

The Semai language belongs, historically, to the Senoic branch of Aslian (Benjamin 1973a, Diffloth 1975). Senoic is one of the three branches of Mon-Khmer languages spoken in Malaysia today; and Mon-Khmer is one of the two divisions of the Austro-asiatic stock, the other being Munda. In terms of number of speakers, Semai, spoken by over 16,000 people, is the most important of the Aslian languages. Some Semais live in settled villages in the lowlands of Perak, but the majority are hill dwellers in the highlands of Perak and Pahang, practicing slash and burn agriculture; they also spend much time hunting, fishing and gathering wild plants, fruits and tubers. There are even some quasi-nomadic individuals and families. The linguistic situation is even more diversified; practically every river and stream has its own dialect, depending of course on what we mean by "dialect". Semais from different settlements may have difficulties in communicating, and the difference between the most distant dialects amounts to a language barrier. No dialect of Semai has ever been described linguistically except in my own work.

Over a period of nine years[1] I have discussed with Semai speakers from 117 settlements, scattered in the entire Semai speaking area; some very briefly, others for many months, depending on the circumstances. No two villages are identical. Some only differ in the choice of certain words, while others differ in more systematic ways, for example in their phonological systems. It is these systematic differences which are historically significant because they disrupt communication far more than mere vocabulary differences. If we consider as one "dialect" all the speech forms among which no systematic differences exist (or have been found so far), we can regroup the 117 settlements into 42 dialect areas, quite a large number considering the total population; and this is not a complete coverage of Semai society. I have mapped and compared these 42 dialects and gained a picture of the linguistic system, called Proto-Semai (PS), from which they all historically descend. This, and a comparison of Semai with its relatives: Temiar, Lanoh and Semnam, and then with Jah-Hut and other Aslian languages, have produced a much sharper picture of Proto-Semai than I had in 1968.

Most of the sound changes which have taken place between Proto-Semai and the present systems have affected the Major Vowels.[2] I will now enumerate the Proto-

---

2) In Mon-Khmer linguistic terminology a Major Vowel is the vowel found in the last syllable of a word.

Semai Major Vowels, then give the correspondences to modern dialects and finally show the probable paths of development from the Proto-Semai system to those of fifteen modern Semai dialects[3] chosen for their diversity or their relevance.

The Correspondences

I) Long Vowels

Semai, like most other Mon-Khmer languages, makes an essential distinction between long and short vowels. To the outsider's ear, the long vowels do not seem especially drawn out as they are, for example in Tamil. Phonetically speaking, it is the Semai short vowels which are extra-short, while the "long" vowels are neutral, not elongated. A precise phonetic notation should mark short vowels with a ˘ diacritic, but for ease of writing, I prefer to note "long" vowels with a double graphic sign: /aa/ for [a], and the extra-short vowels with a single sign: /a/ for [ă]. There may also be some systemic justification for noting "long" vowels as consisting of two segments, but this claim will not be defended here.

Semai words containing a long Major Vowel are more than twice as numerous as words with short Major Vowels; this is the result of two factors:

1) the inventory of long vowels is nearly twice as large as that of short vowels,

2) for every long/short pair of vowels there are more words having the long than the short vowel. The wealth and variety of the examples given here will reflect this imbalance.

Proto-Semai Long Vowels

	Front	Central	Back	
			Unrounded	Rounded
Diphthong:	*ii		*ɯɯ	*uu
*iə	*ee		*ɤɤ	*oo
	*ɛɛ	(*ʌʌ)		*ɔɔ
		*aa		

Of these, the most stable, historically, are: *ii, *ɛɛ, *aa, *ɔɔ, *uu. Those five form

---

3) Semai is spoken in an area overlapping the states of Perak and Pahang, in a triangle between the towns of Ipoh (Perak), Selim River (Perak) and Kuala Lipis (Pahang).

(NE) = North Eastern Semai is located on the middle course of the Telom River (Pahang), between the confluents of the Telom with Sungai Bertam (teew bltɔp) and Sungai Sua (teew swaar).

(CW₁) = Central-West Semai₁ is in the lower part of the Sungai Batang Padang (təəw gɔɔl) basin, for example at kg. Batu Tiga (batuuʔ tigɛɛh), Sungai Landor (təəw lanɔɔr) and others (Perak).

(C₁) = Central Semai₁ is at the village of toʔ mandiiʔ in the Bidor valley (Perak). The geography of Semai dialects deserves a separate publication. The capital letters used here correspond to those in Diffloth (1968) in the following way: W = PER; NW and CW = KAM; N = BOH; E and SE = LIP they are identical to those in Diffloth (1976) and (in press) except for N = CNE, S = SW, C₁ = CSW, CW = CNW, CW₁ = SNW.

the "outer frame" of the vowel system, which remained less affected by change than the other, "inner" vowels. Even then, only Proto-Semai *ii and *uu have been transmitted to the present day without any change in any dialect.

1) *ii and *uu: the old bastion

a) *ii

Proto-Semai *ii corresponds to /ii/ in all dialects, in all positions:

(1) PS *balii^{g}ŋ   "sky"＝(N): /baliik/, (NE): /balii^{k}ŋ/.

(2) PS *tawii^{g}ŋ   "spider"＝(W): /tawiik/, (NE):/tawii^{k}ŋ/.

(3) PS *knriit   "forearm"＝(W, NW, N):/kŋriit/, (NE, E, C):/knriit/.

(4) PS *hii^{d}n   "to grow taller"＝(NW, N, W, C):/hiit/, (E, NE):/hii^{t}n/.

(5) PS *ciip   "to walk, to go"＝(all): /ciip/.

(6) PS *bhii^{b}m   "blood"＝(W, NW, N, CW, C, S): /bhiip/, (NE, E): /bhii^{p}m/, (SE): /bhii^{b}m/.

(7) PS *liiw   "to last long"＝(W, NW, S, N): /liiw/.

(8) PS *gnsiir   "husband"＝(NE, S): /gnsiir/, (NW):/gŋsiir/, (CW, W):/ŋŋsiir/, (C): /ŋnsiir/.

(9) PS *cgiil   "uncomfortable"＝(W, NW, N, C): /cgiil/.

(10) PS *jiis   "daytime"＝(W, NW, S): /jiis/, (NW): also/jiih/.

(11) PS *gii?   "to cut a tree"＝(W, C, S, NW): /gii?/.

No case of Proto-Semai *ii has been found before the final palatals: *-c, *-$^{(J)}$ɲ, *-y; this is one of the many dissimilatory patterns found in Semai; more surprising is the absence of words ending in *-iik, although words in *-ii$^{(g)}$ŋ are common.

Regarding the consonants preceding *-ii-, only *y- seems to be excluded: there are no words with a *yii initial or medial sequence, even though the sequence *yi is found, as we shall se below. The large quantity of data I collected (around 25,000 vocabulary entries) and the obvious phonetic relevance of these patterns suggest that the gaps are not accidental but systematic.

b) *uu

Proto-Semai *uu also corresponds to /uu/ in all dialects and all positions:

(12) PS *dkuuk   "helmeted hornbill, Rhinoplax vigil"＝(W, NW, C, SE): /dkuuk/.

(13) PS *kbuu^{g}ŋ   "a short snake sp., python curtus"＝(NW, N): /kbuuk/, (E): /kbuu^{k}ŋ/, (SE): /kbuu^{g}ŋ/.

(14) PS *huuc   "to climb"＝(W, NW, C, N, NE, E, SE): /huuc/.

(15) PS *bkuu^{J}ɲ   "to get up suddenly"＝(C): /bkuuc/, (NE): /kuu^{c}ɲ/, (SE): /bkuu^{J}ɲ/.

(16) PS *bahuut   "voracious"＝(NW, CW, C, N, NE, E, SE): /bahuut/.

(17) PS *luuᵈn   "to have a bump (on body)"=(NW, C): /luut/, (SE): /luuᵈn/.

(18) PS *tuup   "to spill (liquid)"=(W, NW, CW, C, NE): /tuup/.

(19) PS *cduuᵇm   "to carry (embraced)"=(W, CW, N, C):/cduup/, (E): /cduuᵖm/.

(20) PS *kuur   "to beat bark (into bark-cloth)"=(N, C):/kuur/.

(21) PS *guul   "mortar"=(W, N, NW, CW, C): /guul/.

(22) PS *kuus   "Malayan porcupine, hystrix brachyura"=(W, NW, CW, NE): /kuus/.

(23) PS *kuuy   "head"=(all dials.): /kuuy/.

(24) PS *luuh   "to raid"=(C): /luuh/.

(25) PS *cruuʔ   "sleeping mat"=(W, NW, N, C, S)"/cruuʔ/.

No cases of *-uuw have been found for Proto-Semai. But this sound sequence is not impossible: some dialects have it as a result of certain sound changes[4]

Just as in the case of *-yii- and *-yi-, there are no Proto-Semai words containing the sequence *wuu even though PS *-wu- does exist.

2) *aa: the base

Proto-Semai *aa is almost as stable as *ii and *uu: it remains /aa/ in all dialects, except in (NW) where it becomes a back /ɑɑ/. This phonetic adjustment, by itself, does not affect the system of NW vowels, but it is part of a larger chain reaction to be described below. The *aa vowel was already stable before Proto-Semai times; in fact, it goes back to Proto-Mon-Khmer *aa without any modification.

(26) PS *lntaak   "tongue"=(W, N, NE, E, SE, C, SC, S): /lntaak/, (NW, CW): /lntɑɑk/.

(27) PS *jʔaaᵏŋ   "bone"=(W, N, C, SC, S): /jʔaak/, (NW, CW): /jʔɑɑk/, (NE, E): /jʔaaᵏŋ/, (SE): /jʔaaᵍŋ/.

(28) PS *laac   "bald"=(W, N, C, E): /laac/, (NW): /lɑɑc/.

(29) PS *taaʲɲ   "to plait"=(W, N, C, SC, S):/taac/, (E):/taaᶜɲ/, (NW): /tɑɑc/.

(30) PS *cttaat   "to be stuck, glued[5]"=(W, C, E): /cttaat/, [cəttát], (N): /cʔtaat/ [cəʔtát], (NW): /cʔtɑɑt/ [cəʔtɑ́t].

(31) PS *daaᵈn   "to braid (hair)"=(W, N): /daat/, (E): /daaˈn/, (SE): /daaᵈn/, (NW): /dɑɑt/.

(32) PS *paap   "to warm (self) near fire"=(W, N, E, C): /paap/, (NE): /brpaap/, (S): /ppaap/[pəpáp], (NW): /brpɑɑp/.

(33) PS *jaaᵇm   "to weep"=(W, N, C,): /jaap/, (NW): /jɑɑp/, (E):/jaaᵖm/,

---

4) /tiɲuuw/ "to see from far" (one dialect in (N) area) a borrowing from Malay: tinjau.

5) This is a fossilized verb form containing a -t- infix which is a copy of the final consonant -t. See Diffloth (1972). Many other such forms will be found in the examples given here.

(SE):/jnjaaᵇm/, (S):/jnjaap/.

(34)　PS　*bnsaaw　"wife's elder brother"=(C,　NE):/mnsaaw/,　(N):/mŋsaaw/, (NW):/bŋsaaw/.

(35)　PS　*jʔaar　"thunderstorm"=(C, N, SE):/jʔaar/, (NW):/jʔaar/.

(36)　PS　*glpaal　"shoulder"=(W, N, C, SC, S, E):/glpaal/, (NW):/glpaal/.

(37)　PS　*ʔaas　"shy, reluctant"=(W, N, C):/ʔaas/, (NW):/ʔaas/.

(38)　PS　*subaay　"cooked greens"=(W, N):/subaay/, (NW):/subaay/.

(39)　PS　*maah　"what?"=(W, N, NE, C):/maah/, (NE):/maah/.

(40)　PS　*kaaʔ　"fish"=(W, N, NE, C):/kaaʔ/, (NW):/kaaʔ/, (E):/kaʔ/.

There seem to be no restrictions on the occurence of Proto-Semai *aa with either preceding consonants or finals.

3)　The crowded back: *ɯɯ, *ɤɤ, *oo, *ɔɔ

Five back vowels (including *ɯɯ) are difficult to keep apart: there will be pressure on some of them to merge or leave the area. Both of these solutions were adopted in different Semai dialects: in the North-East the proto-vowel *oo moved towards the center, while in all other dialects *ɤɤ and *ɯɯ merged. In both cases, it is the mid-high vowels, *oo and *ɤɤ which proved to be the weak point in the system.

a)　*oo

Proto Semai *oo has become a centralised diphthong in the NE: before final *-ʔ and *-h it retained part of its former back-rounded quality and gave -ɐu̯ʔ and -ɐu̯h sequences. (see exx. 54, 55); before other finals the rounding was lost, producing a -ɐɯ̯- diphthong. The NE area is not uniform, even though it is small in acreage and population: in some settlements final -h after the -ɐu̯- diphthong became rounded, giving rise to a bilabial voiceless fricative [-ɸ] (see ex. 54). In other places, -ɐu̯ɸ went to -ɐɯ̯ɸ, but stopped short of giving [ɸ] phonemic status. These changes can be very simply described if we allow phonetic features to overlap segment boundaries: the feature of lip-rounding has simply moved forward, leaving the initial part of the vowel unrounded;

	+R		
Proto-Semai:	o	o	h

		+R	
early NE Semai:	ə	o̞	ɸ
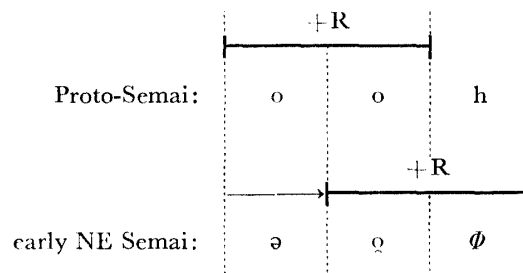
Nevertheless, the sound ɸ is almost unknown in South East Asian languages, especial-

ly as a word final consonant. Besides certain NE Semai settlements, it is also found in Lanoh and Bateg. Since Lanoh, a Senoic language, contains numerous North Aslian ("Semang", "Negrito") elements, and since the NE dialect of Semai is the closest, geographically, to the Bateg (North Aslian) area, we may assume that this unusual -∅ is of North Aslian origin, by diffusion or substratum.

There are other settlements in the NE Semai area (indicated here as: NE$_2$) where the reflexes of Proto-Semai *oo went one step further towards the center and became /əə/, merging with the reflex of Proto-Semai *ɯɯ.

In all other Semai dialects, besides (NE), the proto-vowel *oo is normally preserved. The southern dialect (S) has /oo/ except before final Alveolars (*-t, *-ᵈn, -*r, -*l, -*s) and Palatals (*-c, *-ʲɲ, -*y) where we find /waa/; see exx. (46, 50, 51, 53). Some settlements in the NC area (designated: NC$_1$) have merged *oo with *ɔɔ in all positions; this is part of the chain reaction already mentioned, and explained below.

(41) PS *prook   "rodent (general term)"=(W, NW, N, C, E, SE): /prook/, (NE):/prəɯk/, (NE$_2$):/prəək/.

(42) PS *ploo^{k}ŋ   "atap-roof"=(NW, N, C, S): /plook/, (E): /ploo^{k}ŋ/, (NE): /pləɯ^{k}ŋ/.

(43) PS *kooc   "to dig (the ground with hand)"=(W, NW, N, C, E): /kooc/, (CW$_1$):/kɔɔc/, (S): /kwaac/.

(44) PS *lckooc   "a wild rambutam sp., Nephelium glabrum"=(W, C, E): /lckooc/, (NW): /liʔkooc/, (NE$_2$): /lckəəc/.

(45) PS *sɲroo^{ʲ}ɲ   "a tree w. small green berries, a river name"=(NW): /siŋrooc/ ~/sŋrooc/, (C): /snrooc/, (NW$_1$): /sɲroo^{ʲ}ɲ/.

(46) PS *soot   "(river) to dry up"=(W, NW, N, C, E, SE): /soot/, (S): /swaat/.

(47) PS *cloo^{d}n   "the back (of the body)"=(W, NW, N, C): /cloot/, (E): /cloo^{t}n/, (NE): /cləɯ^{t}n/, (NE$_2$):/cləət/.

(48) PS *roop   "to accompany, companion"=(N, C, E, SE): /roop/, (NE): /rəɯp/.

(49) PS *coo^{b}m   "to dig (earth), to bury"=(W, NW, N, C, S): /coop/, (E): /coo^{p}m/, (NE$_2$): /cəəp/.

(50) PS *krdoor   "woman"=(W, NW, N, C, E, SE): /krdoor/, (S): /kdwaar/, (NE$_2$): /krdəər/.

(51) PS *hool   "come out, go out"=(W, NW, N, C): /hool/, (CW$_1$): /hɔɔl/, (S): /hwaal/, (NE$_2$): /həəl/.

(52) PS *doos   "to step carefully on sth."=(W, NW, N, C, E): /doos/, (CW$_1$): /dɔɔs/, (S): /dwaas/, (NE$_2$): /dəəs/.

(53) PS *bkooy   "an edible yam, Dioscorea alata"=(W, NW): /bkooy/, (S): /bkwaay/.

(54) PS *bkooh   "shiny greyish-white"=(W, NW, N, C, S):/bkooh/, (NE): /bkəuɸ/.

(55) PS *cooʔ   "dog"=(W, NW, N, S):/cooʔ/, (C, E):/coʔ/, (NE): /cəuʔ/.

Proto-Semai *oo does not occur before *-w, just as in the case of *uu, and is other wise unrestricted by finals. Again like *uu, *oo does not seem to occur after *w-; only one example has been found:

(56) PS (?) *wcwooc   "avoidance name of the pangolin, Manis javanica"=(C): /wcwooc/, [from a verb *wooc "to coil"=(C):/wooc/ which may represent a dialect borrowing and not a direct inheritance from Proto-Semai]

b) *ɤɤ

Proto-Semai *ɤɤ has merged with the reflexes of *ɯɯ in all Semai dialects except (NE). When it merges, the exact phonetic value of the resulting vowel depends on the dialect; in most places it has been centralized to [ə]; in the NW it became [ɯ], and in C: [ʉ]; only some settlements in the N area retain the original [ɤ] value, but none of these contrast with the reflexes of *ɯɯ. In the (NE) dialect, *ɤɤ turns up as /oo/. In all likelyhood, *ɤɤ has simply been rounded to /oo/ in (NE) after the older *oo had left the place vacant by becoming [əu], [əɯ] and [ə].

(57) PS *jɤɤk   "to move house"=(W, N, C):/jʉʉk/, (NW):/jɯɯk/, (E): /jəək/, (NE): /jook/.

(58) PS *pnlɤɤᵏŋ   "egg"=(W, N): /pŋlʉʉk/, (NW): /pŋlɯɯk/, (C): /pnlʉʉk/, (S): /pŋləək/, (E): /pnləəᵏŋ/, (SE): /pnləəᵏŋ/, (NE): /pnlooᵏŋ/, (NE₂): /pnlook/.

(59) PS *bɤɤc   "to lick"=(W, C): /bʉʉc/, (NW):/bɯɯc/, (E): /bəəc/, (NE): /booc/.

(60) PS *kɤɤʲɲ   "elder brother of parent, husband of elder sister of parent" =(W, N, C): /kʉʉc/, (S): /kəəc/, (E): /kəəᶜɲ/, (SE): /kəəʲɲ/, (NE₂): /kooc/.

(61) PS *bɤɤt   "to refuse (to do), to do reluctantly"=(C): /bʉʉt/, (NW): /bɯɯt/, (SE): /bəət/, (NE): /boot/.

(62) PS *bɤɤᵈn   "medicinal leaves (used ceremonially, general term)"=(W, N): /bʉʉt/, (NE): /booˡn/, (NE₂): /boot/.

(63) PS *lɤɤp   "very large round rocks (usually in river bed)"=(E): /ləəp/, (NW): /lɯɯp/, (NE): /loop/, (N) :/lʉʉp/.

(64) PS *gɤɤm   "to winnow vertically"=(W, N, C): /gʉʉp/, (NW): /gɯɯp/, (E): /gəəᵖm/, (SE): /gəəᵇm/, (NE): /gooᵖm/, (NE₂): /goop/.

(65) PS *jʊʊr   "(object) to fall"=(W, C): /juur/, (NW): /juuur/, (E): /jəər/, (NE): /joor/.

(66) PS *bʊʊl   "intoxicated"=(W, N, C): /buul/, (NW): /buuul/, (NE): /bool/.

(67) PS *pʊʊs   "to sweep"=(W, N, C): /puus/, (NW): /puuus/, (S, E, SE): /pəəs/, (NE): /poos/.

(68) PS *kʊʊy   "to roast (in embers) sth. wrapped in leaves"=(W, C): /kuuy/, (NW): /kuuuy/, (CW): /kəəy/, (NE): /kooy/.

(69) PS *btʊʊh   "to go down river by raft"=(N): /btuuh/, (NE): /btooh/.

(70) PS *gtʊʊʔ   "skin"=(W, N, C): /gtuuʔ/, (NW): /gtuuuʔ/, (S): /gtəəʔ/, (E): /gtəʔ/, (NE): /gtooʔ/.

The shortening of the vowel in (70), already noticed in (40) and (55) in the E dialect and sometimes others will be accounted for in the section on short vowels.

All finals are found after *ʊʊ-, except *-w; all initials can occur before *-ʊʊ, including *w-:

(71) PS *wʊʊy   "mythical humans, with blade-sharp elbows"=(C): /wuuy/, (NW): /wuuuy/, (NE): /wooy/.

As we noticed before, *uu and *oo cannot occur after *w-, therefore, *ʊʊ does not have, in Proto-Semai, the distribution of a rounded vowel.

c) *ɯɯ

Proto-Semai *ɯɯ appears as /ɯɯ/ in the (NW) dialect but has been centralized to /uu/ in (C) and (N), and further opened to /əə/ in most other dialects. As already mentioned, it has merged with the reflexes of *ʊʊ in all dialects except (NE). In the (NE) dialect, it is usually centralized: /uu/; for some speakers e.g. (NE₂) it has merged with the reflex of *oo into a central /əə/.

(72) PS *rawɯɯc   "carelessly"=(NE, C, N): /rawuuc/.

(73) PS *blɯɯʲɲ   "(flame) to flare up"=(NE): /bluuᶜɲ/, (E): /bləəᶜɲ/, (W): /bluuc/.

(74) PS *sɯɯt   "wax"=(NW): /suuut/, (NE₂, E): /səət/.

(75) PS *lɯɯᵈn   "to stare"=(NW): /luuut/, (N, C): /luut/, (NE₂, SC): /ləət/, (E): /ləəᵗn/.

(76) PS *cɯɯp   "to spread (intr.)"=(NW): /cuuup/, (W, C): /cuup/, (NE₂, CW, E): /cəəp/.

(77) PS *pɯɯᵇm   "to grope"=(NE):/puuᵖm/, (NE₂, CW): /pəəp/, (NW): /puuup/, (N, W): /puup/, (E) /pəəᵖm/.

(78)　PS *kʔɯɯr　"tart, sour, too salty"＝(NE, N, C, S): /kʔʉur/, (NW): /kʔɯɯr/, (E, NE₂): /kʔəər/.

(79)　PS *prgɯɯs　"a tree with edible fruit, Artocarpus sp."＝(W): /prgʉʉs/, (NE₂): /prgəəs/.

(80)　PS *bybɯɯy　"to talk while dreaming"＝(C): /bibʉʉy/, (NE₂): /bibəəy/.

(81)　PS *bsɯɯh　"to follow, pursue"＝(NW): /bsɯɯh/, (N, C): /bsʉʉh/, (NE): /bsʉʉh~bsʉh/, (S): /bsəh/, (NE₂): /bsəəh/.

(82)　PS *bhɯɯʔ　"soiled, already used"＝(NE, N, C): /bhʉʉʔ/, (NE₂): /bhəəʔ/.

The numerous gaps in distribution before final consonants may be due to limitations in my data: *ɯɯ is less frequent than *ɤɤ and the other long proto-vowels. But some of these gaps are probably structural: for example, no convincing example of *ɯɯ is found before *-w; the only candidates belong to the grammatical class of Expressives.

Ex: (83)　(?) *plɯɯw　"(Expressive) loud noise due to hitting"＝(NE₂): /pləəw/, (W, C): /pluuw/.

but the NE dialect also has

(84)　(NE): /ploow/, "(Expressive) loud noise, e.g. when hitting bamboo"

In this case, it is impossible to decide, at present, if the protoform was *plɯɯw or *plɤɤw, or if both forms existed in Proto-Semai: this sort of problem is normal with Expressives because they are lexically non-discrete and weakly productive (Diffloth, 1973a). Nevertheless these examples show that the sequences /-ɯɯw/, /-əəw/, /-oow/ are not impossible in Semai, but perhaps only found in the Expressive phonological system which is more free in combinations than the ordinary, prosaic system.

Here also, the presence of *-wɯɯ- should be noted:

(85)　*wɯɯᵈn　"to crawl"＝(NW): /wɯɯt/, (W, C): /wʉʉt/, (E): /wəəᵗn/, (NE): /wʉʉᵗn/, (NE₂): /wəət/.

The distributions of *ɯɯ, *ɤɤ, *uu and *oo are shown below:

		found in the ordinary system	found in the Expressive system
after *w-	*uu	0	0
	*oo	0	0
	*ɯɯ	+	+
	*ɤɤ	+	+
before -*w	*uu	0	0
	*oo	0	0
	*ɯɯ	0	+
	*ɤɤ	0	+

The table shows that the two proto-vowels *uɯ and *oo form a distributional class, and so do *ɤɤ and *ɯɯ. All four proto-vowels also form together a larger distributional class opposed to other vowels such as *ɔɔ. These various classes are made natural (i. e. phonetic) by our reconstruction of *ɤɤ as a back high-mid unrounded vowel. In this way, the statements of distribution with *w- and -*w can easily be made in terms of phonetic features. Thus, the reconstruction of the phonetic qualities of *ɤɤ, which was arrived at on other grounds, is nicely confirmed.

d) *ɔɔ

This proto-vowel has remained essentially unchanged since PS times in every dialect. The only modification was a slight phonetic raising in the NW and CW dialects due to the backing of *aa to /ɑɑ/. This had no effect on the system except in (CW₁) where the reflex of *oo was lowered, causing a merger. See exx. (43) (51) (52).

(86) PS *sɔɔk "hair"=(W, NW, CW, N, NE, E, SE, S): /sɔɔk/.

(87) PS *sbɔɔᵏŋ "to suck up"=(N, C): /sbɔɔk/, (E, NE): /sbɔɔᵏŋ/, (SE): /sbɔɔᵏŋ/.

(88) PS *gɔɔc "to cut (meat) back and forth"=(C, N, E): /gɔɔc/.

(89) PS *tɲrɔɔʲɲ "to fish"=(W, NW): /tiɲrɔɔc/, (E): /tɲrɔɔᶜɲ/, (N): /tɲrɔɔc/.

(90) PS *kɔɔt "to take"=(W, NW, N, NE, C, E): /kɔɔt/.

(91) PS *lɔɔᵈn "(object) to emerge"=(W, NW, NE₂): /lɔɔt/, (NE): /lɔɔᵗn/, (SE): /lɔɔᵈn/.

(92) PS *kɔɔp "to lie (face down)"=(W, NW, N): /kɔɔp/, (C, E, SE): /kpkɔɔp/.

(93) PS *krɔɔᵇm "under"=(W, NW, CW, NE₂): /krɔɔp/, (E): /krɔɔᵖm/.

(94) PS *jwjɔɔw "a bird sp. (? chestnut capped babbling thrush, Garrulax mitratus major?)"=(N, SE): /jujɔɔw/.

(95) PS *ʔɔɔr "to order, to ask"=(W, N, NW, C): /ʔɔɔr/.

(96) PS *pʔɔɔl "to cook in bamboo"=(NW, NE): /pʔɔɔl/.

(97) PS *ʔɔɔs "fire"=(W, NW, CW, N, NE, E, C, S): /ʔɔɔs/.

(98) PS *tawɔɔh "white-handed gibbon, Hylobates lar"=(SC, C, NE): /tawɔɔh/, (S): /tawɔh/.

(98') PS *sɔɔʔ "rotten"=(W, NW, CW, N, C): /sɔɔʔ/, (S): /sɔʔ/.

There seem to be no restrictions on the immediate environment of *ɔɔ; even *-wɔɔ- sequences are quite common, cf. ex. (98) and:

(99) PS *swɔɔr "bamboo sp. used for blowpipe tubes, Bambusa wrayi"=(N, C): /swɔɔr/.

The sequence -*ɔɔw is also found although it tends to be unstable, and irregularly merge with -*ɔw, as -*ɛɛy does with -*ɛy.

4) The open front: *iə, *ee, *ɛɛ

The front vowel system of Proto-Semai is less crowded than the back, and a four-degree distinction in vowel height (ii, ee, ɛɛ, aa) appears to be under no pressure to collapse or spill over; the cause of trouble however, was an outsider, the diphthong *iə.

Proto-Semai is the outcome of earlier systems such as Proto-Semai-Temiar and Proto-Senoic which apparently had two and three diphthongs respectively. Since Proto-Senoic times, there has been a steady drift to eliminate diphthongs. Proto-Semai only had one left, *iə, and later Semai dialects have all continued this trend[6] by turning *iə into a monophthong. This has created a new long vowel which had to be integrated into the neat four-degree system, causing several mergers and shifts in the front and central areas.

But let us examine first the most stable element, *ɛɛ:

a) *ɛɛ

Proto-Semai *ɛɛ has been preserved in all dialects. Only the (S) dialect has split it into two reflexes: /yaa/ before final Alveolars and Labials (*-t, *-⁽ᵈ⁾n, *-r, -*l, *-s, *-p, *-⁽ᵇ⁾m, *-w), and /ɛɛ/ elsewhere. This is similar to the split of *oo into /waa/ and /oo/ in this dialect but the environment is partly different. It gives the (S) dialect an appearance of striking difference from every other Semai dialect, but the historian can see beyond appearances.

(102)	PS *byɛɛk	"white"=(W, NW, C, SE): /byɛɛk/.
(103)	PS *hɛɛᵍŋ	"to fly"=(W, NW, CW, N, C, NE₁, NE₂, S): /hɛɛk/, (E, NE): /hɛɛᵏŋ/.
(104)	PS *lɛɛc	"to be acquainted with"=(NW, N, NE, C): /lɛɛc/.
(104′)	PS *bɛɛt	"to sleep"=(NW, C, C₁, NE, NE₁, E, SE): /bɛɛt/, (S): /byaat/, (W): /btbɛɛt/, (N): /bʔbɛɛt/.
(105)	PS *cɛɛᵈn	"cooked"=(W, NW, N, C): /cɛɛt/, (S): /cyaat/,
	*br(-n-)cɛɛᵈn	"to cook"=(NE₂): /brcɛɛt/, (E): /brncɛɛᵗn/, (SE): /brncɛɛᵈn/.
(105′)	PS *kʔɛɛp	"red centipede, Scolopendra sp."=(W, NW, CW, N, NE, NE₁, C, E): /kʔɛɛp/, (S): /kʔyaap/.
(106)	PS *cɛɛᵇm	"bird"=(W, NW, N, NE₂, C, SC): /cɛɛp/, (E): /cɛɛᵖm/, (S): /cyaap/.

---

6) One of the reasons for this steady drift is in the morphology; Senoic infixes are sensitive to the number of initial consonants in a root; opening diphthongs like iə, or uə create confusion because their first element can be interpreted in two ways: as a consonant y-, w-, or as part of the vowel; and the infix will be completely different depending on the interpretation. The only diphthongs found in Semai today are either closing diphthongs, which do not create any confusion, or opening diphthongs which have appeared in (S) where the morphology has been very much reduced.

(107) PS *lɛɛw   "to manipulate" = (CW, C, SE): /lɛɛw/.

(108) PS *chɛɛr   "bright red" = (W, NW, N, NE, C): /chɛɛr/.

(109) PS *sɛɛl   "shy, ashamed" = (N): /sɛɛl/, (NE): /slsɛɛl/, (S): /slsyaal/.

(110) PS *kɛɛs   "nuts, acorns" = (W, NW, C, C_1, NE): /kɛɛs/.

(111) PS(?) *tlɛɛy   "banana" = (NW): /tlɛɛy/, (N, C, W, S): /tley/.

(112) PS *galɛɛh   "manioc, Manihot utilissima[7]" = (N, NW, SE): /galɛɛh/.

(113) PS *plɛɛʔ   "fruit" = (W, NW, N, C, NE, S): /plɛɛʔ/, (E): /plɛʔ/.

There are apparently no restrictions on *ɛɛ with initials and finals, except perhaps with final palatals: before -*y, the vowel is usually short and raised to /e/, but not in all dialects; the same applies to *ɔɔw. Before -*c, examples are very hard to find, but do exist; the absence of a convincing example before -*ʲɲ is probably due to this rarity and not to a systematic gap.

b)   *iə

Proto-Semai *iə has a complicated history. Some speakers in the (NE) area — designated here as (NE_1) — have a front low-mid rounded vowel /œœ/ as a reflex, which contrasts with every other vowel. This sound is unique in Aslian languages and unusual, to my knowledge, for South-East Asia. This (NE_1) is the only dialect, with (C_1) cf. below, to have preserved a distinct unit for *iə; in the rest of (NE), *iə has merged with *ɛɛ to /ɛɛ/. In addition, *iə went to [ɛi] before -*ʔ in the whole (NE) area, just as *oo went to [əu] before -*ʔ. In the other Semai dialects, *iə can be reconstructed by its unique pattern of merging with other vowels: in (N, E, C), *iə has merged with *ii to /ii/, in others (W, NW, C_1, CW, SC), *iə has become /ee/, either causing a merger with the reflex of *ee as in (W, SC) or merely filling a position left empty by a previous movement of *ee towards the center as in (NW, CW), or towards a mid /EE/ as in (C_1). In the (S) dialect, *iə has split into two reflexes: /yaa/ before Alveolars (-*t, -*ᵈn, -*r, -*l, -*s) and Labials (-*p, -*ᵇm), and /ee/ elsewhere, merging there with the reflex of *ee as in (SC, CW, NW, W). Finally, in (SE), *iə has also merged with /ee/ but only after palatals initials (*c-, *j-, *s-); after the other initials it has merged to /ii/, as in the neighbouring (E) dialect.

The probable path of development from Proto-Semai *iə to its reflexes in each present dialect will be shown at the end of this paper, after we examine Nasalized vowels which are crucial for this proto-vowel.

---

7) This word could be a borrowing from Malay (cf. Mal. sĕgala-gala "everywhere", tiada bĕrgala "uncontrollable"). Many words spelled with a final -a in Malay are borrowed into Semai with an -ɛɛh ending: (CW)/batuuʔ tigɛɛh/ "a village name", (Mal. batu tiga), (N): /sumbɛɛh/ "a plant used for its red seeds, Bixa orellana" (Mal. kĕsumba, Burkill (1966) p. 333). This is apparently due to a feature of the Malay dialect from which Semais borrowed these words. But in the case of /galɛɛh/, the wide dialectal distribution of the word in Semai would argue against it being a borrowing.

(114) PS *rntiək "a plant used in making traps, Donax sp."=(C, SE): /rntiik/, (NE): /rntɛɛk/.

(115) PS *lhiaᵍŋ "saliva"=(W, NW, C₁, S): /lheek/, (N, C): /lhiik/, (E): /lhiiᵏŋ/, (SE): /lhiiᵍŋ/, (NE₂): /lheek/.

(116) PS *siəc "to steal"=(W, NW, SC, SE): /seec/, (N, E, C): /siic/, (S): /seek/, (NE): /sɛɛc/, (NE₁): /sœœc/.

(117) PS *giəc "dental decay"=(W, NW): /geec/, (N, C): /giic/.

(118) PS *riət "to wring (e. g. a cloth)"=(W, NW, SC): /reet/, (C, E, N): /riit/, (S): /ryaat/, (NE): /rɛɛt/.

(119) PS *tadiəᵈn "a small bird, grey wagtail?"=(W, NW): /tadeet/, (N, C): /tadiit/, (E): /tadiiᵗn/, (SE): /tadiiᵈn/, (S): /tadyaat/, (NE): /tadɛɛᵗn/, (NE₂): /tadɛɛt/.

(120) PS *liəp "to plait palm leaves"=(W, NW): /leep/, (C, E, SE): /liip/, (S): /lyaap/.

(121) PS *m(a)riəᵇm "how many?"=(W, NW, CW, C₁, SC): /mreep/, (C, N): /mriip/, (E): /mariiᵖm/, (SE): /mariiᵇm/, (S): /mayaap/, (NE): /marɛɛᵖm/, (NE₁): /marœœep/, (NE₂): /marɛɛp/.

(122) PS *ɲmpiər "winnowing sieve"=(W, NW): /ɲmpeer/, (SC): /ɲmpeer/, (S): /ɲmpyaar/, (C, E): /ɲmpiir/.

(123) PS *diəl "spoor"=(W, NW): /deel/, (C, E): /diil/, (NE): /dɛɛl/, (S): /dyaal/.

(124) PS *jiəs "a fruit tree[8], Artocarpus sp."=(W, NW, SE): /jees/, (E): /jiis/, (S): /jyaas/, (NE): /jɛɛs/.

(125) PS *tiə? "earth"=(W, NW, CW, SC, S): /tee?/, (N): /tii?/, (C, E): /ti?/, (NE): /tɛi̯?/.

Proto-Semai *iə is not found before -*h and -*w. There are no good examples of *iəʲɲ, but this may be a limitation of our data since *iəc is found. The presence of *iək and *iəc shows that the diphthong *iə has the distribution of an open vowel like *ɛɛ, and not that of higher vowels like *ii and *ee which do not occur before -*k and -*c.

---

8) Several speakers of Semai and Temiar have told me that this tree was called: jias in Malay. However, it is not found in the existing Malay dictionaries, which, at any rate, do not represent Malay regional dialects consistently. If the form: jias is found in some Malay dialect, it confirms our PS reconstruction *jiəs, whatever the direction of borrowing might be. There are other examples of this sort: Malays call Sungai Chĕndĕriang a river which the Semais living in that area (CW) call: /təəw cŋreek/; this name would be reconstructed as PS *cnriaᵍŋ, a phonologically well formed Semai word. The odd polysyllabic Chĕndĕriang of Malay is almost certainly a rendering of this Proto Semai form; the -dĕ- part represents an epenthetic -d- which is inserted in -nr- clusters in several conservative Semai dialects; phonetically, PS *cnriaᵍŋ was probably pronounced [cənᵈriaᵍŋ] and borrowed almost intact by Malay speakers.

The only restriction on consonants preceding *iə is the absence of *yiə-; in this respect, *iə behaves like a high front *ii or *ee.

The reconstruction of *iə as a diphthong is consistent with its ambivalent patterning: the final part [ə] behaves like an open vowel with final consonants and the initial part [i] behaves like a high vowel with initial consonants.

It should also be noted that although *ya is found in absolute initial position:

(126) PS *yah "to be day-time"=(NW, N, C, E, NE): /yah/.

*ya- cannot occur after an initial consonant. And conversely, even though the sequence *ʔiə does occur after an initial consonant:

(127) *rʔiəs "root[9])"=(W, NW, SC): /rʔees/, (N, E, SE): /rʔiis/, (S): /rʔyaas/, (NE): /rʔɛɛs/.

*ʔiə cannot occur in absolute initial position. This near complementary distribution of PS *ya and PS *iə suggests that PS *iə derives from an earlier (Proto-Senoic?) *yV̄ sequence. See also Diffloth (in press).

c) *ee

Proto-Semai *ee has been centralized to /əə/ in the (CW) dialect and further went to /ɯɯ/ in the (NW). Otherwise it has remained unchanged as /ee/.[10)] One dialect in the (C) area — designated (C_1) — has a slightly more open reflex: /EE/ which contrasts with both /ɛɛ/ from *ɛɛ and /ee/ from *iə. This (C_1) dialect has five degrees of vowel height: /ii/, /ee/, /EE/, /ɛɛ/, /aa/.

(128) PS *kdeek   "a trap for squirrels"=(W, C, N, NE, SC, S, SE): /kdeek/, (NW): /kdɯɯk/.

(129) PS *tbeeᵍŋ   "full"=(W, N, NE₂, C, S): /tbeek/, (NW): /tbɯɯk/, (C_1): /tbEEk/, (CW): /tbəək/, (E): /tbeeᵏŋ/, (SE): /tbeeᵍŋ/.

(130) PS *kaceet   "unskilled"=(N, C): /kaceet/, (NW): /kacɯɯt/.

---

9) Mon-Khmer cognates to this word cf. e. g. Diffloth (1976) exx. (54, 72, 98) do not have a -ʔ- even though the languages in question are not known to loose medial -*ʔ-. That sound is found in the whole Senoic branch: Lanoh: /rʔɛɛs/, Jah Hut: /rʔyɛs/ and also in the whole of North Aslian: Kinta' Bong: /ʔiʔyes/, Jahai: /jʔes/, Che' Wong: /jʔɛs/.

I doubt that it represents an old Mon-Khmer cluster, but rather a very early metathesis from a *ʔryɛs form. There are other examples of metathesis due to *yɛ and *wə in early times. The absence of /ʔ/ in the South Aslian forms (ibid. ex. 54), although ʔr- clusters do exist in South Aslian languages (ibid. ex. 53), suggests that in *ʔryɛs the initial *ʔ- was not original but introduced by a syllabification rule *[rəyɛs]→*[ʔəryɛs].

10) The solution, which I proposed in 1968, to reconstruct this *ee as əə and *iə as ee, was correct only for a stage more recent than Proto-Semai. In 1968, I was not aware of the (NE) dialect and had not enough evidence on nasalised vowels in (S) and (SC).

(131) PS *tee^{d}n   "(birds) gather (to feed)" = (N, C): /teet/.

(132) PS *br(m)cee^{b}m   "to hide oneself" = (W, N, C, NE$_2$): /brceep/, (NW): /brcuɯɯp/, (CW): /brcəəp/, (E): /brcee^{p}m/, (S): /brnceep/.

(133) PS *teew   "water, river" = (W, N, NE, C, SC, E, SE): /teew/, (NW): /tuɯɯw/, (CW): /təəw/, (C$_1$): /tEEw/.

(134) PS *(ta-)heer   "to drag" = (W, N, C, E): /heer/, (W) also: /taheer/, (NW): /tahuɯɯr/.

(135) PS *bnteel   "space in front of the house" = (N): /mnteel/, (NW): /bntuɯɯl/.

(136) PS *btees   "mushroom" = (W, N, NE, NE$_1$, C, SC, S, E, SE): /btees/, (C$_1$): /btEEs/, (NW): /btuɯɯs/, (CW): /btəəs/.

(137) PS *b(-a?-)hee?   "satiated" = (W, NE, C): /bhee?/, (N): /ba?hee?/, (NW): /ba?huɯɯ?/.

The absence of *-eec and *-ee$^{(J)}$ɲ is conspicuous, at least in the ordinary system; in the larger phonological system of Expressives, such finals are found:

(138) PS *rbeec   "(Expressive) unclean eyes" = (C): /rc-rbeec/.

The apparent absence of *eeh is due to a shortening process to be shown in the section on short vowels.

5) The weak center: *ʌʌ

We have already seen how the central part of the vowel system was used throughout the history of Semai to relieve pressure from the crowded front and back areas and to merge all new comers into one vowel: /əə/, or /ʉʉ/, or /uɯɯ/, depending on the dialect. But the central area was probably not empty in Proto-Semai times: it seems that there already was a Proto-Semai vowel *ʌʌ. But this area has seen so much movement that *ʌʌ has been absorbed by merging with some other proto-vowel in most dialects. The only dialects where *ʌʌ had a chance to survive were those in which *uɯɯ had retained its high quality and had attracted *ʊʊ to this high position, that is: the W, NE, C and especially the NW dialect. Unfortunately, the number of words where an *ʌʌ can be reconstructed is extremely small and this raises some doubt as to the authenticity of *ʌʌ. It would be tempting to explain these words as borrowings, for example from one of the dialects where *uɯɯ went to əə at an early period. But a comparison with Lanoh and Temiar shows that there was a well represented Proto-Senoic vowel *əə, distinct from Proto-Senoic *uɯɯ, and from which the few cases of Proto-Semai *ʌʌ all descend. The most likely conclusion is that Proto-Senoic *uɯɯ and *əə merged to *uɯɯ some time before the appearance of Proto-Semai, and that Proto-Semai itself later

borrowed the words with (PS) *ʌʌ from some other dialect which had not undergone that merger. In any event, PS *ʌʌ is too rare and irregular (cf. the (C) dialect) for me to give a full set of examples with all finals or to describe its distribution.

(139) *ha-hʌʌc "to guffaw"=(N, C, E, NE): /hahəəc/.

(140) *ktkʌʌt "(to go) backwards"=(NW): /kʔkəət/.

(141) *grʔʌʌp "to burp"=(W, NW, CW, E, NE): /grʔəəp/, (C): /grʔuup/, (S): /grʔeep/[11].

(142) *tadʌʌr "(rooster) to sing"=(W, NW, N, NE): /tadəər/, (C): /taduur/.

(143) *ʔʌʌy "address term to wife or friends"=(W, C, NE): /ʔəəy/.

(144) *bʌʌy "address term to young person"=(W, NW, N): /bəəy/.

(145) *ʔa-bʌʌh "reference term to father"=(N, NE): /ʔabəəh/.

The presence of three kinship terms (143, 144, 145) and three words with onomato-poetic connotations (but not Expressives) (139, 141, 142) in this list is probably not accidental as such words tend to be irregular. This further weakens the case for an authentic PS *ʌʌ vowel.

II) Short Vowels

The inventory and the history of Proto-Semai short vowels is more simple:

Proto-Semai Short Vowels[12]

*i	(*ʉ)	*u
*ɛ	*ə	*ɔ
	*a	

Of these, only *ʉ is not clearly established; the contrast /ə/-/ʉ/ is found only in the (NE) dialect, and not always consistently, and the number of words with *ʉ is small. In all other dialects, it has merged with *ə.

1) PS *ʉ (?)

(146) PS *taluˤŋ "a black bee sp."=(NE): /taluᵏŋ/, (W, N, NW): /talək/.

(147) PS *ckʉt "rough to touch"=(NE): /ckʉt/, (NW, E, SE): /ckət/, (N): /cket/.

---

11) The data on the (S) dialect is not sufficient at the moment to decide if this /ee/ is the regular reflex of PS *ʌʌ. If it is confirmed, this would make the existence of a separate Proto-Semai vowel *ʌʌ almost certain: PS *ɯɯ and *ɤɤ have merged to a not very high /ʉʉ/ in the (S) dialect.

12) In 1968, I thought the short-vowel system also contained /e/ and /o/; it later turned out that these two vowels are due to Malay borrowings or to a shortening of /ee/ and /oo/ before -h and -ʔ. The example /klater/ given in 1968 is apparently from Malay kuatir "to worry" (ultimately from Arabic) and means "to insist" rather than "to quarrel".

(148) PS *sruᵈn   "to press and push (in delivery)"＝(NE): /sruᵗn/, (NW, C): /srət/.

(149) PS *yнp   "to be twilight"＝(W, C): /yəp/, (NE): /yнp/, (NE) also: /yip/.

(150) PS *cнl   "to tatoo"＝(NE): /cнl/, (C, E): /cəl/.

2)   *ə

After their merger, *н and *ə share a common history. The only later event is a shift to /e/ before Alveolar finals (-*t, -*ᵈn, -*r, -*l, -*s) in the (N) dialect. This is a mere phonetic shift since Proto-Semai does not have a vowel *e; but it eventually produced a merger with *i in case the vowel was nasalized, and some speakers do have a contrastive /e/ due to Malay borrowings.

(151) PS *tək   "to uproot, to pull out"＝(W, NW, N, NE, C, E, SE): /tək/.

(152) PS *ləᵉŋ   "Malayan flying fox, Pteropus vampyrus"＝(W, NW, N, C, S): /lək/, (E, NE): /ləᵏŋ/.

(153) PS *kət   "egg, belly"＝(W, NW, NE, E, SE): /kət/, (N): /ket/.

(154) PS *klbəᵈn   "still unripe"＝(C, S): /klbət/, (NE): /klbəᵗn/.

(155) PS *cəp   "to plant (a stick in the ground)"＝(W, NW, N, C, E): /cəp/.

(156) PS *pləᵇm   "land leech"＝(W, NW, N, C): /pləp/, (NE, E): /pləᵖm/, (SE): /pləᵇm/.

(157) PS *dər   "(fire) to flare up"＝(W, NW, CW, N, NE, C, E): /dər/, (N): /der/.

(158) PS *jəl   "to bark"＝(NW, NC, C, NE): /jəl/, (N): /jel/.

(159) PS *crəs   "rib"＝(W, NW, C, NE, SC, S): /crəs/, (N): /cres/.

(160) PS *bəh   "to make"＝(W, NW, CW, C, N, NE): /bəh/.

There are no examples of PS *ə before Palatals (-*c, -*ʲɲ, -*y); the few possible exceptions are found in the grammatical class of Expressives, for which a separate phonology and a separate history will have to be written. However, such finals are found in Temiar ordinary phonology, so this remains a problem for Proto-Senoic.

Cases of possible -*əw finals are also extremely rare and problematic.

3)   *a

Proto-Semai *a merged with *ɔ to /ɔ/ when preceded by a Labial (*p-, *b-, *m-, *w-) and followed by a Velar (*-k, -*ᵉŋ). This merger occurred in all dialects, except, once again, in the (NE) dialect.

(161) PS *maŋ   "there is"＝(W, NW, N, E, C): /mɔŋ/, (NE): /maŋ/.

(162) PS *baᵉŋ   "younger sibling (other than one's own)"＝(NW, N, C): /bɔk/,

(E): /bɔᵏŋ/, (NE): /baᵏŋ/.

(163) PS *(s)lbak "a tree (leaves used ceremonially), Ouratea crocea" (cf. Burkill (1966) p. 1642) = (NE): /lbak/, (NW, N, E, SE): /slbɔk/.

(164) PS *wak "a depression (e. g. in the ground)" = (NE): /wak/, (NW, N, E, SE): /wɔk/.

The merger has created a morphophonemic alternation whenever a Proto-Semai infix separated the *a from the initial Labial; compare (162) to the following:

(165) PS *b-n-aŋ "younger sibling (one's own)" = (W, NW, N, NE, C, S): /mnaŋ/.

The examples above can be contrasted with the following, where an original PS *ɔ preceded by a Labial and followed by a Velar is found in the (NE) and other dialects as well.

(166) PS *mɔk "a low palm, ?Curculigo latifolia?" = (NW, NE, C, SE): /mɔk/.

(167) PS *pɔk "an edible ground mushroom sp." = (NW, N, NE, C, E, SE, S): /pɔk/.

(168) PS *tapɔᵍŋ "a small pouch (e. g. for tobacco)" = (NW, N, NE₂): /tapɔk/, (E): /tapɔᵏŋ/.

(169) PS *wɔᵍŋ "stomach, hips" = (NW, N, C, SC, S): /wɔk/, (E, NE): /wɔᵏŋ/, (SE): /wɔᵍŋ/.

The merger of *a to /ɔ/ has also affected a number of Malay borrowings:

(170) (Mal. kumbang) "coconut beetle" = (S): /kumɔk/.
(171) (Mal. subang) "ear-studd" = (N): /subɔk/, (E): /subɔᵏŋ/.

Proto-Semai *a has merged with *ɛ to /ɛ/ after Palatals in the (S) dialect, otherwise, *a has remained a separate entity in all dialects, only shifting to a long front [æ:] in the (NW) and (CW) (a phonetic adjustment which is not noted here).

The examples of PS *a given below are useful to establish the Mon-Khmer affiliation of Semai (and other Senoic languages) because PS *a reflects Proto-Mon-Khmer *a and can easily be contrasted with PS *aa which reflects Proto-Mon-Khmer *aa.

(172) PS *dak "a spear-trap" = (W, NW, NE, C, SE): /dak/.
(173) PS *pnraᵍŋ "sunshine" = (NW, N): /pŋrak/, (C): /pnrak/, (E): /pnraᵏŋ/.
(174) PS *dkat "cold, feverish" = (NW, C, SC, E, NE): /dkat/.
(175) PS *daᵈn "to die" = (W, NW, N, C, NE₂): /dat/, (NE, E): /daᵗn/.
(176) PS *kap "to hold in the teeth" = (W, NW, N, E, C, S): /kap/.

(177) PS *lsaᵇm "rainy season"=(W, N): /lsap/, (E, NE): /lsaᵖm/.

(178) PS *jar "to run"=(W, NW, N, NE, C, E): /jar/, (S): /jɛr/.

(179) PS *r(a)ŋkal "ladder"=(W, NW): /rŋkal/, (N, C, SE, S): /raŋkal/.

(180) PS *cas "smoke"=(W, NW, N, C): /cas/, (S): /cɛs/.

(181) PS *bah "parent's younger brother (address)"=(W, NW, N, C, NE): /bah/.

(182) PS *daʔ "to lie down"=(W, NW, CW, N, C): /daʔ/.

Before final Palatals (-*c, -*ʲɲ) and semi-vowels (-*w, -*y) Proto-Semai *a is found only in Expressives, not in the regular phonology.

4) *ɔ

This proto-vowel has changed very little since Proto-Semai times: before -*y it has been centralized to /əy/ in the (W, NW, CW) dialects, and before -*w it has been opened to /aw/ in the (NE) and raised to /ow/ in the (C). These changes are phonetic shifts which did not create mergers in the ordinary vocabulary inherited from Proto-Semai. But the endings /əy/, /aw/, /ow/ may exist and cause secondary mergers with Expressives and borrowings.

(183) PS *lɔk "to put on, to wear"=(W, NW, N, C, E): /lɔk/.

(184) PS *dɔᵍŋ "pig-tailed macaque, Macaca nemestrina"=(W, NW, N, C, S): /dɔk/, (E, NE): /dɔᵏŋ/, (SE): /dɔᵍŋ/.

(185) PS *rɔc "to uproot"=(W, NW, N, C, E, SE): /rɔc/.

(186) PS *tɔʲɲ "to pluck"=(W, NW, N, C, NE₂): /tɔc/, (E): /tɔᶜɲ/.

(187) PS *tbɔt "to mount, to ride"=(W, NW, CW, E, NE, SE): /tbɔt/.

(188) PS *jarɔᵈn "to tiptoe"=(NW, C): /jarɔt/, (NE): /jarɔᵗn/.

(189) PS *gɔp "Malay, foreigner"=(W, NW, N, C, E): /gɔp/.

(190) PS *(kn-)tɔᵇm "right side, (kn-="from")"=(W, N, NW): /kntɔp/, (C, S): /tɔp/, (E): /tɔᵖm/.

(191) PS *litɔw "bachelor"=(W, NW, N): /litɔw/, (NE): /litaw/, (C): /litow/.

(192) PS *bagɔr "small felines"=(W, C, N, NE): /bagɔr/.

(193) PS *brɔl "blowpipe dart butt"=(W, NW, N, C, E, S): /brɔl/.

(194) PS *lgɔs "a rambutan, Nephelium lappaceum" (Burkill, (1969) p. 1571) =(W, NW, N, NE, C, E, S): /lgɔs/.

(195) PS *lɔy "to reach"=(W, NW, CW): /ləy/, (N, C, NE): /lɔy/.

(196) PS *(g)tɔh "to spit"=(W, NW): /gtɔh/, (N, E, SE, C, S): /tɔh/.

(197) PS *bɔʔ "to carry (child) on back"=(NW, N, C, S): /bɔʔ/.

Proto-Semai *ɔ has a full distribution with finals. With initials, the only restric-

tion seems to be on *wɔ sequences which can only occur in word initial position, cf. ex. (169). A similar restriction on *ya sequences has been mentioned earlier.

5) *u

This proto-vowel is stable in all dialects, like its long counterpart *uu.

(198)	PS *luk	"to laugh" = (W, NW, CW, N, NE, E, C): /luk/.
(199)	PS *juᵉŋ	"foot" = (W, NW, CW, N, C, S): /juk/, (E, NE): /juᵏŋ/.
(200)	PS *huc	"to sip" = (N): /hũc/, (W, NW, C, E, SE): /huc/.
(201)	PS *pɲrũɲ	"a rat sp. ?Rattus rajah (Thomas)" = (C, E): /pɲrũɲ/.
(202)	PS *bgut	"to snap (intr.)" = (N, NW, C, E): /bgut/.
(203)	PS *pruᵈn	"to burn (leftovers, rubbish)" = (NW, C): /prut/, (NE): /pruᵗn/.
(204)	PS *sagup	"cloud, mist" = (W, NW, CW, N, C, E): /sagup/.
(205)	PS *blhuᵇm	"a small tree with little round grey fruits" = (NE): /blhuᵖm/.
(206)	PS *dur	"to become soft, jelly-like (after a blow)" = (W, NW, C): /dur/.
(207)	PS *bul	"to wrap sth. (for protection)" = (NW, N, C): /bul/.
(208)	PS *kus	"to get up" = (W, NW, NC, N, NE, C): /kus/, (S): /b-kus/.
(209)	PS *truy	"to disturb (e. g. bees) in order to disperse" = (C): /truy/.
(210)	PS *tʔuh	"to blow the blowpipe" = (SC): /tʔuh/.
(211)	PS *nruʔ	"a grub" = (NW, N): /ŋruʔ/, (C, NE): /nruʔ/.

This proto-vowel being somewhat rare, it is not easy to find good examples for all finals: ex. (201) properly belongs to the Nasalized vowel category; but the examples chosen, even when scanty, have cognates outside Semai which support their reconstruction. The only final sequence which seems excluded is -*uw. As noted above, the sequence -*wu- does occur:

(212) PS *cawul   "to be paralyzed" = (C): /cawul/.

but -w- only serves to separate a Minor Vowel from the Major Vowel /u/. No words with the structure *CwuC have been found.

6) *i

This proto-vowel is somewhat rare in the Non-Expressive phonology of Proto-Semai, but it must be reconstructed and it does not seem to have changed in the modern dialects:

(213)	PS *bsiᵉŋ	"Dusky leaf monkey, Presbytis obscura" = (W, NW, N, C, S): /bsik/, (E, NE): /bsiᵏŋ/.
(214)	PS *dic	"all, terminated" = (W, NW, CW, N, C, SC, E, SE): /dic/,

(S): /dik/.

(215) PS *lɲsiʲɲ "gums"＝(W): /lŋsic/, (N): /liŋsic/, (C): /linsic/, (E): /lɲsiʲɲ/.

(216) PS *slit "(trace, track) to become unclear"＝(C, SE): /slit/.

(217) PS *tip "(hole) to be covered up"＝(C, NE): /tip/.

(218) PS *hniᵇm "solid, stable"＝(W, NW, N): /hnip/, (E): /hniᴾm/.

(219) PS *rasiw "to relase (e. g. a trap tenser)"＝(C): /rasiw/.

(220) PS *kayir "to shake cloth"＝(C): /kayir/.

(221) PS *twil "to bend by force"＝(C): /twil/.

(222) PS *krjis "to shake sth. off"＝(NW): /krjis/.

(223) PS *plʔis~plʔīs "fresh fish smell"＝(N): /plʔis/~/plʔīs/, (S): /plʔis/, (NW): /plʔih/~/plʔīh/, (W, CW, C): /plʔīh/.

(224) PS *krdiʔ "to remain silent, quiet"＝(NW, N, W, NE): /krdiʔ/, (W) also: /krdiiʔ/.

The sequence -*ik is excluded, as was -*iik, but -*iᵉŋ is established, and contrasts with -*iʲɲ and -*iiᵉŋ. Also excluded is the sequence -*iy, just as in the case of -*uw. As for -*is, it becomes -ih irregularly in the different dialects, but without merger since -*ih is not found.

We find *yi- in the same conditions as *wu- was found: after a Minor Vowel but not immediately after a consonant.

7) *ε

The only change which affected this Proto-Vowel was a phonetic raising to /e/ before palatal finals (-*c, -*ʲɲ, -*y) in (C, E, SE, SC). As in the case of *ə before Alveolars, this phonetic adjustment did not cause any merger, except after Nasals where it was raised further to /i/.

Proto-Semai *ε is as frequent as *ɔ, *a and *ə, and not rare like *u and especially *i. The frequency ratios are roughly: 100 each for *ε, *a, *ɔ and *ə, 50 for *u, 25 for *i.

(225) PS *lεk "to tickle (trans.)"＝(NW, CW, N, C, E, SE): /lεk/.

(226) PS *krdεᵉŋ "to warm up (self, food) near fire"＝(W, NW, C): /krdεk/, (E): /krdεᵏŋ/.

(227) PS *sεc "meat"＝(W, NW, N, NE): /sεc/, (C, E, SE): /sec/, (S): /sεk/~/sek/.

(228) PS *sbεʲɲ "mosquito"＝(W, NW, N, NE₂): /sbεc/, (NE): /sbεᶜɲ/, (C, SC): /sbec/, (E): /sbeᶜɲ/, (SE): /sbeʲɲ/, (S): /sbεk/~/sbek/.

(229) PS *gltεt "to flick (finger)"＝(NW, NE, C, E, SE): /gltεt/.

(230) PS *tεᵈn "to chop"＝(NW, C): /tεt/ (the evidence for -*ᵈn is morpho-

phonemic: (C): /tnnɛn/ "chopping)".

(231) PS *lɛp   "to finger (e. g. music instrument)"=(W, N, C): /lɛp/.

(232) PS *rnsᵉᵇm   "tip of leaves"=(SE): /rnsᵉᵇm/.

(233) PS *kbɛr   "to wink"=(W, NW, CW, N, C): /kbɛr/.

(234) PS *gɛl   "to cut firewood"=(NW, N, C, SC): /gɛl/.

(235) PS *lɛs   "to scold"=(W, NW): /lɛs/, (N): /les/.

(236) PS *bybɛy   "independently"=(NW): /bibɛy/, (C): /bibey/.

(237) PS *-rɛh   "down"=(W, NW, N, S, SC, SE): /rɛh/.

(238) PS *gɛʔ   "a little"=(W, NW, N, C): /gɛʔ/.

The contrast between -*ɛɛy, cf. ex. (111) and -*ɛy is not firmly established and may be of secondary origin; many Malay words spelled with a final -ai are borrowed into Semai with a final -ɛy.

The distinction between long and short vowels is, however, well established in spite of the imbalance in favor of long vowels already noted. The only interchange between long and short vowels occurs before final -*ʔ and -*h. In the case of -*ʔ, long proto-vowels merge with short ones into short vowels regularly in (E) and (SE), often in (S) and (SC), and occasionally in (C). As for -*h, the situation is even more irregular, long vowels being always rare in that position; however the correspondence patterns of many words with -Vh endings are those of long vowels:

(239) PS *-te(e)h   "there"=(W, N, NE, C, E, SE): /-teh/, (NW, CW): /-təh/.

(240) PS *pɤ(ɤ)h   "to blossom"=(W, NW, N): /pəh/, (C): /papəh/, (NE): /poh/, /papoh/.

(241) PS *ko(o)h   "to cut off"=(W, NW, N, E, C, S): /koh/, (NE): /kəyɸ/.

Long vowels before -h have apparently been shortened independently in several dialects, this being a very natural sound change, with much diffusion and borrowing, leaving only a few remaining cases of -VVh.

III) Nasalized Vowels

Semai nasalized vowels are of two kinds; either automatically conditioned by a preceding nasal consonant, or unpredictable by environment. Both kinds have essentially the same history, which is quite different from that of oral vowels and gives us a clue to the history of the whole vowel system.

Most the of examples which follow are of the first kind (vowel preceded by N). The second kind is less common and the nasality disappears in some dialects without apparent regularity. This is a problem which remains to be presented in detail and will

not be treated here. I will only point out that these unpredictable nasal vowels are often (but by no means always) preceded by h- or ʔ- without any other nasal element in the same word. Irregular nasalization in such words may be due to an old nasal prefix which existed in Proto-Semai and could have nasalized the vowel through the h- or ʔ- before disappearing. But let us turn to the more regular cases: vowels preceded by N.

With the exception of a few dialects, there are no mid-high nasalized vowels in Semai today. This is a rule of Semai phonology which has applied several times independently, but it is not a very old phenomenon: Proto-Semai did have mid-high vowels preceded by N, and new ones appeared in the course of time. There are two ways of eliminating such vowels: raising them to high or lowering them to mid-low. Different Semai dialects have used one or the other method, or both, at different times.

1) /Noo/

For /oo/ both changes are found: to /uu/ and to /ɔɔ/. The northern half of Semai territory (W, NW, N, NE) raises /Noo/ to /Nuu/, causing a merger with the reflex of *Nuu which is /Nuu/ in all dialects; this happened regardless of the historical origin of this /oo/: from PS *oo in (W, NW, N) or from PS *ʊʊ in (NE). The reflexes of the original *Noo were not affected by this change in the (NE) dialects: they are the regular /Nəɯ/ or /Nəʉ/ or /Nəə/. This shows that the change Noo→Nuu is recent in that area, and later then the *ʊʊ→oo shift.

In the southern half of Semai territory (CW, C, E, SE, SC, S), /Noo/ was lowered to merge with the reflex of PS *Nɔɔ which is /Nɔɔ/ in all dialects. Evidently, in the (S) dialect this merger is more recent than the diphthongization of *oo to /waa/. cf. ex. (249, 250, 252, 253, 254).

A few examples of PS *Nuu and *Nɔɔ are given first for the sake of contrast with *Noo and *Nʊʊ.

a) PS *Nuu

(242) PS *lmuuɲ "tooth" = (W, NW, CW, C, NE, E, SE, SC, S): /lmuuɲ/.
(243) PS *trnuuʔ "to give in excess" = (W, NW, CW, C, SC): /trnuuʔ/.
(244) PS *muus "to break ground to extract sth." = (W, NW, CW, C, SC): /muus/.

b) PS *Nɔɔ

(245) PS *nɔɔm "urine" = (W, NW, CW, C, N, NE, S): /nɔɔm/.
(246) PS *k(c)mɔɔc "ghost" = (W): /kmɔɔc/, (NW, N): /kiʔmɔɔc/, (C, SC, E): /kcmɔɔc/, (S): /krmɔɔc/.
(247) PS *ŋɔɔy "smell (V, N)" = (W, NW, CW, N, C, E, SE): /ŋɔɔy/.
(248) PS *crmɔɔr "legend" = (W, NW, CW, N, C, NE): /crmɔɔr/.

c) **PS *Noo**

(249) PS *noos   "heart"=(W, NW, N): /nuus/, (CW, C, SC, E, SE): /nɔɔs/,
                        (S): /nwaas/, (NE₂): /nəəs/, (NE): /nəɯs/.

(250) PS *kmoor   "caterpillar"=(W, NW, N): /kmuur/, (CW, C, E, SE): /kmɔɔr/,
                        (S): /kmwaar/, (NE₂): /kməər/, (NE): /kməɯr/.

(251) PS *kmooŋ   "calf of leg"=(W, NW, N): /kmuuŋ/, (CW, C, E, SE, SC, S):
                        /kmɔɔŋ/, (NE₂): /kməəŋ/.

(252) PS *ŋoot   "to drink"=(W, NW, N): /ŋuut/, (CW, C, SC, E, SE): /ŋɔɔt/,
                        (S): /ʔwã̈ãt/, (NW₂): /ŋəət/.

(253) PS *ɲuɲooy   "lips"=(W, NW, N): /ɲiɲuuy/, (C, SC, E, SE): /ɲiɲɔɔy/,
                        (S): /yiywã̈ãy/, (NE): /ɲiɲəɯy/.

(254) PS *k-n-oon   "child"=(W, NW, N): /knuun/, (CW, C, SC, E, SE): /knɔɔn/,
                        (S): /knwaan/, (NE₂): /knəən/.

This last example can be contrasted with the original root without nasal infix, a
rare case of Semai morphophonemics.

(255) PS *koodn   "nephew (address)"=(NW, N, C): /koot/, (E): /kooⁱn/.

d) **PS *Nɤɤ**

(256) PS *nɤɤbm   "ripe"=(W, CW, N, C): /nɯɯp/, (NW): /nɯɯp/, (E): /nəəᵖm/,
                        (NE₂): /nuup/, (NE): /nuuᵖm/.

(257) PS *krɲɤɤt   "to go backwards"=(W, E, C): /krɲɯɯt/, (NW): /krɲɯɯt/,
                        (NE): /krɲuut/.

(258) PS *lrŋɤɤr   "a small black bee"=(NW): /lrŋɯɯr/, (C): /lrŋɯɯr/, (NE):
                        /lrŋuur/.

(259) PS *smɤɤc   "stinging insect"=(W, C, E, SE, S): /smɯɯc/, (N): /snɯɯc/,
                        (NW): /smɯɯc/, (NE): /smuuc/.

This is again an occasion for morphophonemics:

(260) PS *sɤɤc   "to sting"=(W, N, C): /sɯɯc/, (NW): /sɯɯc/, (E, SE, S):
                        /səəc/, (NE): /sooc/.

In the case of *Noo there are several dialects, especially in the southern half of the
(N) area, (N₁), which have preserved a distinct /Noo/, but they are not cited here.

2) **/Nee/**

In the case of /Nee/ sequences, we also find that both treatments are possible:

raising to /Nii/ in the northern half of Semai territory and lowering to /Nɛɛ/ in the rest, but the dividing line is not the same, and the treatment of /Nee/ from PS *Nee is very different from that of /Nee/ from PS *Niǝ.

In every dialect where they occurred, both raising to /Nii/ and lowering to /Nɛɛ/ caused mergers with the reflexes of PS *Nii and PS *Nɛɛ, which are /Nii/ and /Nɛɛ/ respectively everywhere. But the patterns of merging are different and have to be stated separately for each dialect.

In the (W) dialect, where *ee and *iǝ merged to /ee/, PS *Nee and PS *Niǝ also merged together and were then raised to /Nii/, thus merging with the reflex of PS *Nii. The result is a triple merger: *Nii, *Nee, *Niǝ→/Nii/.

In the (NW) and (CW), PS *ee was centralized to /ɯɯ/ and /ǝǝ/ respectively, merging with both PS *ɯɯ and PS *ɤɤ. The reflex of PS *Nee followed the same route; another triple merger: *Nee, *Nɯɯ, *Nɤɤ→/Nǝǝ/ in (CW),→/Nɯɯ/ in (NW). In addition, as these two dialects have PS *iǝ→/ee/, we find another merger: PS *Niǝ became /Nii/, merging with the reflex of PS *Nii.

In the (N) dialect, PS *iǝ and *ii merged to /ii/ while PS *ee remained distinct; accordingly, PS *Niǝ and PS *Nii merged to /Nii/. In the southern part (N₁) of the (N) area, PS *Nee was retained as /Nee/, just as PS *Noo was kept as /Noo/, but in the northern part, PS *Nee also became /Nii/ causing the same triple merger as in (W).

In the (NE), PS *iǝ had an independent course, going to /œœ/ or merging with PS *ɛɛ to /ɛɛ/. In this area, PS *Niǝ followed the same path, giving /Nœœ/ or merging with PS *Nɛɛ to /Nɛɛ/. In addition, PS *ee had been kept as /ee/, but PS *Nee was raised to /Nii/, merging with the reflex of PS *Nii: two separate mergers: *Niǝ, *Nɛɛ →/Nɛɛ/, and *Nee, *Nii→/Nii/.

In the (SE), PS *Niǝ, *Nee and *Nii all merged to /Nii/ as in the (N) dialect, even though *iǝ and *ii have not merged in (SE) as they did in (N).

In the (C) and (E) dialects, PS *iǝ and PS *ii merge to /ii/ as in the (N) dialect, and so do PS *Niǝ and *Nii; however PS *Nee was not raised as in (N), but lowered instead to /Nɛɛ/, merging there with the reflex of PS *Nɛɛ: two separate mergers: *Niǝ, *Nii→/Nii/, and *Nee, *Nɛɛ→/Nɛɛ/.

In the (SC) dialect, *Nee was also lowered and merged with PS *Nɛɛ into /Nɛɛ/. But, while PS *iǝ and *ee merged to /ee/, PS *Niǝ was not raised but also lowered to /Nɛɛ/, merging with PS *Nɛɛ: a different triple merger: *Niǝ, *Nee, *Nɛɛ→/Nɛɛ/.

Finally, in the (S) dialect, PS *Nee was not lowered but raised, to merge with PS *Nii to /Nii/. In addition, this dialect has the unique development of PS *iǝ and *ɛɛ to /yaa/ before Alveolars and Labials; the same thing happened to PS *Niǝ and *Nɛɛ giving /Nyaa/ sequences in these positions; elsewhere PS *Niǝ was raised to /Nɛɛ/, merging with the reflex of PS *Nɛɛ as in the neighbouring (SC) dialect. This peculiar development of PS *Niǝ and its exact parallel to that of PS *Nɛɛ suggest that

PS *iə and *ɛɛ have a partly common history in this dialect, and were phonetically similar at an early date, closer to each other than to PS *ee. We shall return to this point.

I will first give a few examples of PS *Nii and *Nɛɛ for comparison with PS *Nee and *Niə.

a) __PS *Nii__

(261) PS *kmiil "on top of" = (W, NW, N, NE, SC): /kmiil/.

(262) PS *sŋii? "awareness" = (W, NW, N, NE, C, S): /sŋii?/, (S): also: /sŋi?/.

(263) PS *manii? "rain" = (W, NE, CW, N, C, SE, SC, S): /manii?/, (SE, S) also: /mani?/, (E): /mani?/.

b) __PS *Nɛɛ__

(264) PS *nɛɛŋ "to see" = (W, NW, CW, N, NE, C, S): /nɛɛŋ/.

(265) PS *r(t)mɛɛt "yellow, turmeric" = (W, NW, C, SE): /rmɛɛt/, (N): /r?mɛɛt/, (E): /rtmɛɛt/, (S): /rmyaat/.

(266) PS *mɛɛm "breast" = (W, NW, N, C, E, SE): /mɛɛm/, (S): /myaam/.

c) __PS *Niə__

(267) PS *(ka)miəŋ "cheek" = (W, CW, N, C, C₁, E, SE): /miiŋ/, (NW): /kamiiŋ/, (NE, S): /mɛɛŋ/.

(268) PS *krmiər "a large millipede" = (N, C): /krmiir/, (SC): /krmɛɛr/.

(269) PS *ɲiəŋ "to eat uncooked rice-grains" = (C): /ɲiiŋ/, (SE): /yiiŋ/, (SC): /ɲɛɛŋ/.

(270) PS *kniə? "rat" = (W, NW, CW, N, C): /knii?/, (E, SE): /kni?/, (NE): /knɛi?/, (SC): /knɛɛ?/ also /knɛ?/, (S): /knɛ?/.

(271) PS *k-t-n-iət "in the back, after, the last" = (NW, N): /k?niit/, (W, C, E): /ktniit/, (SC): /ktnɛɛt/, (S): /knyaat/.

Here again, the nasal infix gives rise to a morphophonemic alternation when we compare (271) to its root (272):

(272) PS *kiət "bottom, buttocks" = (W, NW, SC): /keet/, (C, E): /kiit/, (NE): /kɛɛt/.

d) __PS *Nee__

(273) PS *kmeet "a small mosquito sp." = (W, N): /kmiit/, (NW): /kmuɯt/, (CW): /kməət/, (E, SC): /kmɛɛt/.

(274) PS *r-k-n-eek "orphan" = (N, SE): /rkniik/, (NW): /r?nuɯk/, (C):

/rknɛɛk/, (S): /rniik/.

(275)　PS *k-n-eeŋ　　　"tree" = (W, N, NE): /kniiŋ/, (NW): /knɯɯŋ/, (C, C₁, E, SC): /knɛɛŋ/.

The (S) dialect only has the root form for this word, which apparently no other dialect has kept:

(276)　PS *keeᵍŋ　　　"tree" = (S): /keek/.

(277)　PS *r-n-ces　　　"the face" = (W, N): /rniis/, (NW): /rnɯɯs/, (C, C₁): /rnɛɛs/.

(278)　PS *b-n-eeʔ　　　"father (one's own)" = (W, N, NE): /mniiʔ/, (NW): /mnɯɯʔ/, (CW): /mnəəʔ/, (SE): /mniʔ/, (C, SC): /mnɛɛʔ/, (C₁): /mnɛʔ/.

(279)　PS *d-aʔ-n-eeʔ　　　"the middle" = (N): /daʔniiʔ/, (S): /tahniiʔ/.

(280)　PS *k(-raʔ-)meeʔ　　　"beautiful, happy face" = (W, NE): /kraʔmiiʔ/, (NW): /kraʔmɯɯʔ/, (E): /kraʔmɛʔ/, (C): /kmɛɛʔ/.

Here also, the existence of the root without the nasal infix reveals the history of these vowels:

(281)　PS *reek　　　"to be orphan" (cf. ex. 274) = (N, C, E): /reek/.

(282)　PS *rees　　　"to be alive, erect" (cf. ex. 277) = (W, C, S): /rees/, (NW): /rɯɯs/.

(283)　PS *(ʔa-)beeʔ　　　"father" (cf. ex. 278) = (W, N): /beeʔ/, (C, N): /ʔabeeʔ/, (NW): /bɯɯʔ/.

(284)　PS *-deeʔ　　　"half" (cf. ex. 279) = (W, C): /deeʔ/, (SC): /smaʔdeeʔ/, (N): /smadeeʔ/, (NW): /dɯɯʔ/, (CW): /dəəʔ/.

In such cases, it is possible to see that certain sound changes happened before (or after) certain others, by comparing oral and nasalized vowels of the same origin.

* * *

Chronology

By extending the domain of such comparisons, to include, for instance, the history of consonants, it is possible to reconstruct in great detail the most likely course of events from the Proto-Semai system to each living dialect.　I will not retrace here every hypothesis and every piece of evidence.　These include phonetic, grammatical, and lexical arguments, as well as sociological and geographic considerations, all taken together.

In this paper, I will simply give the results which I have reached so far.

1) the (NE) dialects

The earliest change which happened in the ancestor of the present (NE) systems was the diphthongization of PS *oo. This left an empty space which was filled by PS *ɤɤ becoming /oo/. None of these changes occurred in any other ancestor of present Semai dialects, and both had to occur before the raising of nasalized vowels.

Another early change was the opening and monophthongization of PS *iə, which eventually produced /œœ/ and /εε/. This must also have happened before the raising of nasalized vowels and took place in no other proto-dialect. The most conservative dialects of the (NE) area, such as (NE$_1$), did not evolve any further, but others turned /əɯ/ to a monophthong /əə/ and merged it with the reflex of *ʌʌ, and others such as (NE$_2$) merged PS *ɯɯ as well into the all-purpose /əə/.
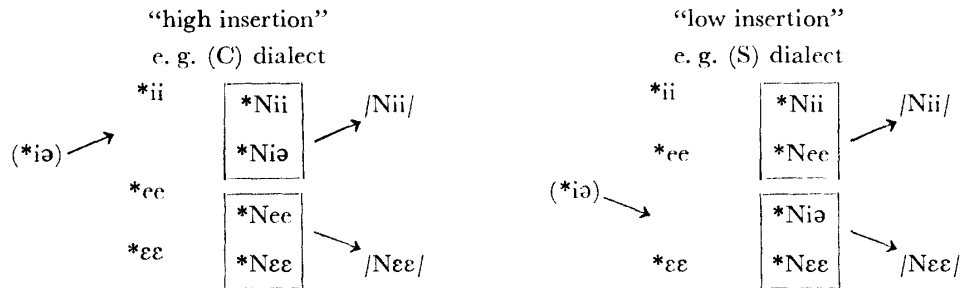
2) Low insertion vs. high insertion dialects

All the other dialects of Semai are distinguished from the (NE) by the following innovations which probably occurred only once in the proto-system from which they all descend: *ɤɤ merges with *ɯɯ to /ɯɯ/, *ʉ merges with *ə to /ə/ and *a merges with *ɔ in the environment B — K. After this common stage there was divergence: oral and nasalized *iə and *ee merged with each other and with *ii and *εε in a variety of patterns. At the time, the front vowel system (discounting short vowels) probably was:

$$\begin{array}{cc} & \text{*ii} \\ \text{*iə} & \text{*ee} \\ & \text{*εε} \end{array}$$

In at least one case, (C$_1$), we know that *iə was "inserted" as a long vowel between *ii and *ee, forcing a phonetic adjustment of *ee to a lower /EE/. This lowering explains why this *ee would have merged with *εε to /εε/ after Nasals: there is a general tendency in all languages to have fewer nasal vowels than oral ones; nasalized /EE/ simply continued its lowering movement till it merged with /εε/, reducing the number of contrasts among nasalized vowels. This merger of *Nee and *Nεε is not restricted to (C$_1$): it is common to all the (C) and (E) dialects. In the proto-(C)-(E) dialect *ee and *εε must have been near each other and could not have been separated by another vowel like the reflex of *iə: the "high-insertion" of *iə between *ii and *ee probably occurred at that time, or earlier. In fact, there is no argument against placing it at the proto-stage of the (N), (CW), (NW) and (W) dialects as well. Only the two southern-most dialects: (SC) and (S) cannot be accomodated in this manner: in both, *Niə merges with the lower *Nεε, and in (S) *Nee merges with *Nii: a "low insertion" of *iə between *ee and *εε in these dialects seems necessary to avoid a collision of the na-
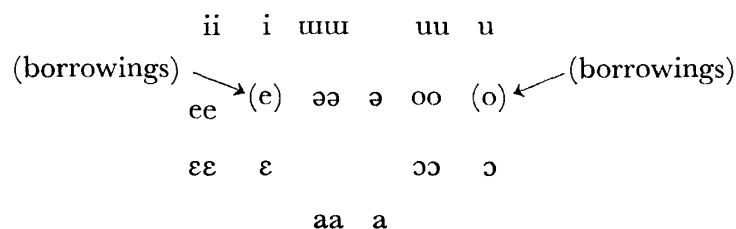
salized vowels.

"high insertion"
e. g. (C) dialect

"low insertion"
e. g. (S) dialect

The difference between "high insertion" and "low insertion" appears to be ancient, separating the two southernmost dialects from the middle dialects, i.e. all except (NE).
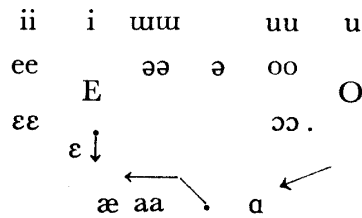
3) Chain reaction in the (NW)

A number of phonetic shifts, without mergers, in the (NW), seem to be interrelated. They are apparently due to the influence of recent Malay borrowings: in several areas, especially (W), (NW), (W) and parts of (N), Malay is spoken with some fluency by many Semais.　In those dialects, the Malay phonemes /i/ and /u/ are pronounced with the phonetic value they have in Malay, usually [e] and [o], and not merged to the existing Semai /i/ and /u/ as happens in other dialects, e.g. (C).　These new sounds, for which there already exist long Semai counterparts /ee/ and /oo/, now contrast with Semai /i/ and /u/ and have enriched the short vowel system.　Following this, in most of the (NW) and part of (CW), /ɔ/ has been pushed down by the new /o/ to a more open, and centralized [ă].　This in turn pushed /a/ forward to [æ].　However, the same pressure was working on *ɛ due to the new /e/.　There was a danger of merging *ɛ: [ɛ̆] with *a: [æ̆].　The solution has been to lengthen [æ], creating a new long vowel /æææ/.　A merger of this new long vowel with the existing /aa/ was avoided by moving the latter to a back position: /ɑɑ/, for which the short counterpart already existed.　Finally, this back /ɑɑ/ exerted pressure on /ɔɔ/ which moved up slightly.　The chain of events stopped here in most of these dialects, except in (CW₁) where this higher /ɔɔ/ merged with the reflexes of *oo.　This final merger, mentioned earlier, could only be explained by considering the total vowel system, the pressures exerted on it at certain points, and the tendency to match short and long vowels.
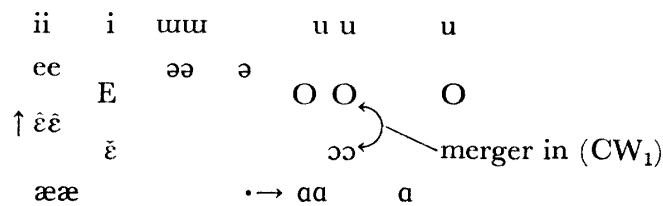
1) Proto NW-CW (oral) vowels:

2) movement of a to [æ̃]:

```
ii    i    ɯɯ        uu    u

ee          əə    ə    oo
       E                      O
εε                      ɔɔ .
      ε ↓
          ←——＼      ↙——
        æ aa    ＼  ɑ
```

3) change of æ to æ̃æ and movement of aa and εε:

```
ii    i    ɯɯ        u u     u

ee         əə    ə
       E              O O       O
↑ ε̂ε̂
       ε̃              ɔɔ   ＼merger in (CW₁)
æ̃æ              ·→ ɑɑ    ɑ
```

4) Overall chronology

I will now propose a general chronology of vowel innovations in the history of fifteen Semai dialects since Proto-Semai. The diagram given below is slightly different in nature from the loosely defined "genetic trees" often used in historical linguistics.
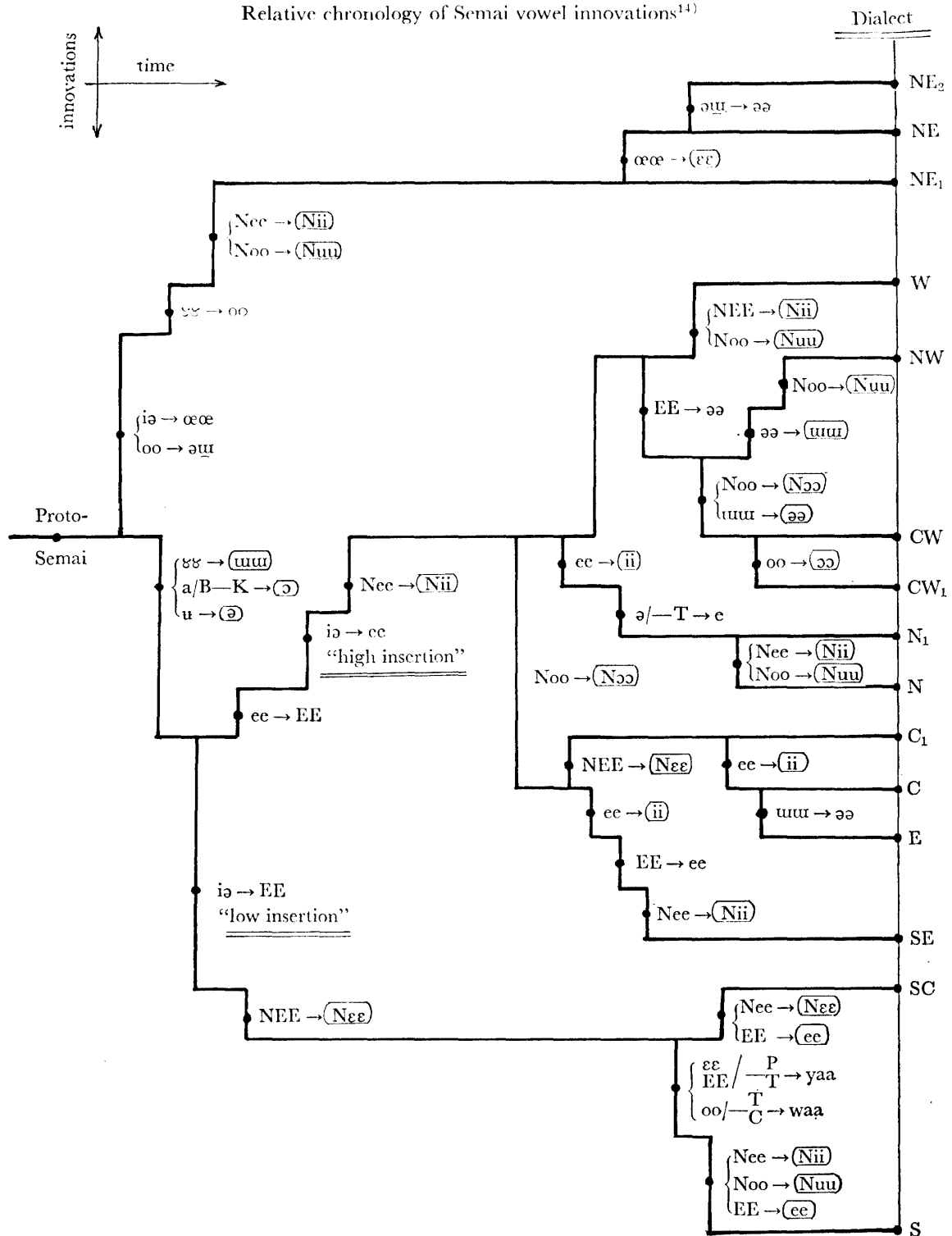
The horizontal dimension represents time; it is oriented (from left to right), and scalar (the length of the line is roughly proportional to the duration of time). The vertical dimension represents systematic innovations; it is neither oriented nor scalar; i.e. there is no significance attached to an upwards vs. a downwards direction of the line, and the length of any vertical line is irrelevant; both are determined by graphic convenience.

When several innovations are known, but their relative date is not, they are enclosed together within curly brackets on the same vertical line, pending further clarification. The points of which the lines are made represent language systems, i.e. dialects.

Every innovation creates a dialect which is different from the non-innovating one; if the latter survives, a branching appears in the diagram. The binary logic of innovation vs. lack of innovation automatically results in binary branching.

Thus, the diagram, by itself, makes no reference to the social groups or individuals using the various dialects: it only represents linguistic systems. The loosely defined "genetic trees" have often, and justly, been criticized because in every population there are usually several dialects in use, and also multidialectal individuals. This is true at present in all sections of Semai society, was probably true at the time when Proto-Semai was spoken, and presumably also at every other time in history. But the diagram which follows is not couched in these terms.[13] For example, the fact that Proto-Semai is repre-

---

13) In order to take into account bi- (or multi-) dialectal individuals, a true historical socio-linguistic diagram would have to mention not only the number, but also the identity of individual speakers.

Relative chronology of Semai vowel innovations[14]

innovations

time

Dialect

ɔɯ → əə → NE₂

NE

œœ → (εε) → NE

NE₁

{Nee → (Nii)
Noo → (Nuu)}

εε → ɔɔ

W

{NEE → (Nii)
Noo → (Nuu)}

Noo → (Nuu)

EE → əə

NW

əə → (ɯɯ)

{Noo → (Nɔɔ)
ɯɯ → (əə)}

Proto-
Semai

{iə → œœ
oo → əɯ}

ee → (ii)

oo → (ɔɔ)

CW

{εε → (ɯɯ)
a/B—K → (ɔ)
u → (ə)}

Nee → (Nii)

CW₁

ə/—T → e

N₁

iə → ee
"high insertion"

{Nee → (Nii)
Noo → (Nuu)}

Noo → (Nɔɔ)

N

ee → EE

C₁

NEE → (Nεε)

ee → (ii)

C

ee → (ii)

ɯɯ → əə

E

EE → ee

iə → EE
"low insertion"

Nee → (Nii)

SE

SC

NEE → (Nεε)

{Nee → (Nεε)
EE → (ee)}

{εε/—P/T → yaa
oo/—T/C → waa}

{Nee → (Nii)
Noo → (Nuu)
EE → (ee)}

S

14) Phonetic symbols written inside a circle indicate a merger; thus, ee→(ii) means: /ee/ changes into /ii/ and merges with the already existing /ii/. The diagram given here does not indicate all phonetic adjustments, for example those which occurred in (NW) and (W), but only those which resulted in mergers, either immediately or at some later time. Mergers are the most useful tool for chronology because they are irreversible.

sented by a single point does not constitute a claim about Semai prehistoric socio-linguistics; it simply follows from the defintion of the term "Proto-Semai"=the linguistic system (dialect) from which a certain number of living (and extinct) dialects descended through the historical process of systematic transmission and innovation.

Now, we can study Semai history and climb deeper into the past of Mon-Khmer civilization.

References

Benjamin, G. 1973a. "Austroasiatic subgroupings and prehistory in the Malay peninsula," *Austroasiatic Studies,* Oceanic Linguistics, Special publication No. 13, Univ. of Hawaii Press, Hawaii.

—————. 1973b. "An outline of Temiar grammar," *ibid.*

Burkill, I. H. 1966. *A dictionary of the economic products of the Malay Peninsula.* 2 vols. Ministry of Agriculture and Cooperatives. Kuala Lumpur.

Diffloth, G. 1968. "Proto-Semai phonology," *Federation Museums Journal,* Vol. XII. Kuala Lumpur.

—————. 1972. "Ambigüité morphologique en Semai," *Langues et Techniques, Nature et Société,* J. M. C. Thomas and L. Bernot eds. Editions Klincksieck. Paris.

—————. 1973a. "Expressives in Semai," *Austroasiatic Studies,* Oceanic Linguistics, Special publication No. 13. University of Hawaii Press, Hawaii.

—————. 1973b. "Minor syllable vocalism in Senoic languages," *ibid.*

—————. 1975. "Les langues Mon-Khmer de Malaisie, Classification historique et innovations," *Asie du Sud Est et Monde Insulindien.* VI, 4. Paris.

—————. 1976. "Proto-Mon-Khmer final spirants," *Genetic Relationship, Diffusion and typological similarities of East and Southeast Asian Languages,* Japan Society for the Promotion of Science, Tokyo.

—————. (in press) "Jah-Hut, an Austroasiatic language of Malaysia," *South East Asian Linguistic Studies,* Pacific Linguistics, Australian National University, Canberra.