ISO/IEC JTC1/SC2/WG2 N3667
L2/09-252
2009-09-14

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to encode the Pahawh Hmong script in the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Author: **Michael Everson**
Status: **Liaison Contribution**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2009-09-14**

1.0 Introduction. Pahawh Hmong is a script devised for writing the Hmong language by Shong Lue Yang (*Soob Lwj Yaj* ƜƜ̓ ƮƜ ƋƜ [ʃɔ́ŋ lɨ̂ jâ]). Shong Lue Yang was a charismatic figure among the Hmong in Laos, and was considered by many to be a kind of messiah. It is said that in 1959 the writing system was revealed to him by two supernatural messengers who appeared to him over a period of months. A full account of this is given in Smalley, Vang, and Yang 1990. Devised in Laos, Pahawh Hmong was taken to northern Thailand refugee camps, and then moved with waves of immigrants to Minnesota and California in the United States, and to Australia. The writing system itself had four Stages of development. In this document, the Romanized Popular Alphabet orthography (widely used by the Hmong in North America) is given alongside example text in Pahawh Hmong. Two features of the RPA are of note. Double vowels *ee* and *oo* indicate [ɛŋ] and [oŋ] respectively; final letters indicate tones thus:

RPA			
-b	⌈	*v́*	high-level
-m	L	*ṿ*	low-glottalized
-d	⊬	*v̬*	low-rising
-j	⊧	*v̂*	high-falling
-v	⊬	*v̌*	mid-rising,
-Ø	⊦	*v*	mid-level
-s	⊢	*v̀*	low-level
-g	⊧	*v̤*	falling-breathy

1.1 The Source Version, Pahawh Pa (*Phajhauj Paj* ƋK̇ ƝƜ ƋƜ̇ [pʰâ hâu pâ]), is not in current use. While containing the seeds of the system, in its structure and glyphs it is very different from the later Stage Versions, and was never used as a practical system for writing Hmong. It is considered a separate but related script, and is not supported by this encoding.

1.2 The Second Stage Reduced Version, Pahawh Njia Dua O (*Phajhauj Ntsiab Duas Ob* ƋK̇ ƝƜ JK ƜƜ̇ ƝȦ [pʰâ hâu ndʒía dùa ʔɔ́]), is in current use. It was taught by Shong Lue Yang in 1965-04, and is supported by the Australian Hmong Language Institute and by Hmong Script Software's ƮV ƜƜ̇ *Cwjmem* [ʈ͡ʂɨ̂ mẹ] font; fonts are also available from the Hmong Language Institute in Minnesota. The Hmong user community in Australia uses the Second Stage Reduced Version.

1.3 The Third Stage Reduced Version, Pahawh Njia Dua Pe (*Phajhauj Ntsiab Duas Peb* ƋK̇ ƝƜ JK ƜƜ̇ ƜƜ̇ [pʰâ hâu ndʒía dùa pé]), is in current use. It rationalizes some features of the Second Stage Reduced Version, and was introduced by Shong Lue Yang in 1970-08. Some members of the Hmong user

community in Minnesota use the Third Stage Reduced Version. A Third Stage font is available from Hmongwriting.com.

1.4 The Final Version, Pahawh Tsa (*Phajhauj Txha* 𖬖𖬲𖬮𖬵 𖬄𖬲𖬡 [p^{h}â hâu ts^{h}a]), is not in regular use. It is a radical simplification of the Third Stage Reduced Version introduced in January 1971 by Shong Lue Yang about a month before his assassination. Smalley *et al*. 1990 state that it is not in use as a practical system, though some people who know it use it as a kind of shorthand (and called it "shorthand" in English). The encoding proposed here can represent text written in all three of these Revisions.

The fact that Stage Two and Stage Three orthographies are both used makes character naming and placement of characters in the code table slightly problematic. In the Third Stage Reduced Version, base characters without diacritics end in *-b* or *-v* tones; these are represented by a more complex alternation of tones (*-b*, *-v*, -Ø, *-g*, *-m*) in the Second Stage Reduced Version; The easier Third Stage Reduced Version names have been used here—this does not imply a preference for either Stage, as UCS names are arbitrary. The code charts here follow the Second Stage Reduced Version ordering because we have access to a complete dictionary which follows that order.

2.0 Processing. Pahawh Hmong syllables are separated by spaces in text, and may contain one to four characters: base, base with diacritic, base + base, base with diacritic + base, base + base with diacritic, and base with diacritic + base with diacritic. Structurally, Pahawh Hmong is unique among the world's writing systems in that the vowel rime of a syllable (its vowel with or without tone diacritic) is written before the consonant onset of the syllable (its consonant with or without consonant-identifier diacritic). In the Figures 1 and 2, the structure of the words "Pahawh Hmong" (*Phajhauj Hmoob* [p^{h}â hâu ^{h}mɔ́ŋ]) is analyzed, given in Second and Third Stage Reduced Version (Final Version is identical to Third Stage Reduced Version in this example).

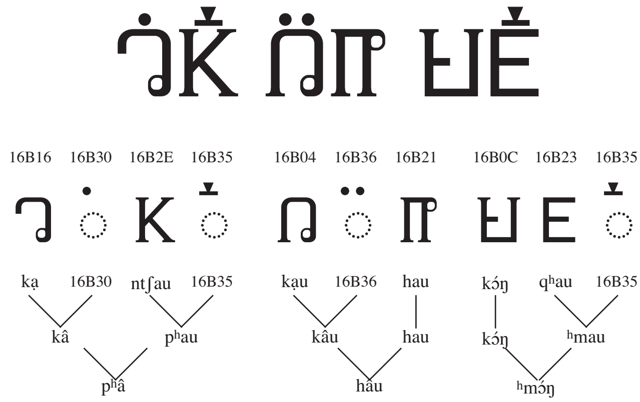

Figure 1. Second Stage Reduced Version

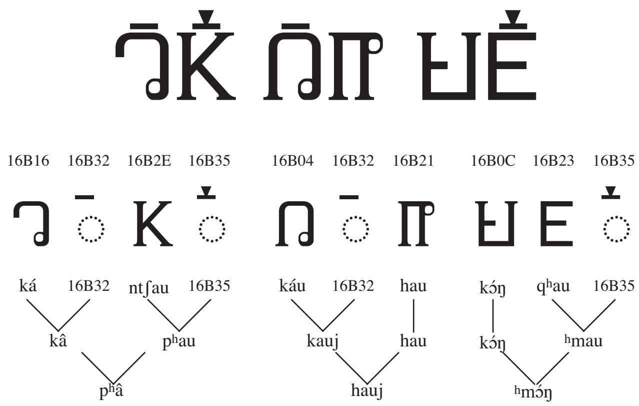

Figure 2. Third Stage Reduced Version and Final Version

2.1 Vowel rimes. Characters from 16B00–16B1B are vowel codas. Adding diacritics to these alters the tone. 16B1A–16B1B are long vowels. In Shong Lue Yang's system, Hmong Daw dialect syllables KIAB 𖬔, KIAV 𖬕, KAB 𖬖, and KAV 𖬗 are used for Hmong Leng dialect *kav*, *kav*, *kaab* or *kaav* respectively. A revision of the script by Jay Kue of Hmong Script Software includes special characters for *kaab* 𖬚 and *kaav* 𖬛 (in Second Stage Reduced Version *kaam* and *kaav*). These are atomic characters with no decomposition. In the first place, decomposition would break the one-to-four character convention for representing Hmong syllables. In the second, the addition of a (non-productive) character I would be problematic as 16B4A PAHAWH HMONG NUMBER TENS looks just like it.

2.2 Consonant onsets. Characters from 16B1C–16B2F are consonant heads. Adding diacritics to these changes the base consonant to a different, usually unrelated, consonant. Use of diacritics to affect various changes is unsystematic for the consonants. For the vowels, Stage Two Reduced Version, Stage Three Reduced Version, and Final Stage Pahawh Hmong offer an increasing rationalization of relationships, which in Final Stage Pahawh Hmong is quite systematic. The differences are orthographic, however, and do not affect the encoding. As stated above, the Stage Three Reduced Version was chosen as the basis for the character names in the encoding because it is more systematic than the Stage Two Reduced Version, and because the Final Stage is a subset of the Stage Three Reduced Version.

2.3 Combining diacritics are found at 16B30–16B36 and function in the usual way. Note that 16B34 and 16B35 could be composed (16B32 + 16B30 and 16B32 + 16B31 respectively). This encoding is not recommended (because decomposition would break the one-to-four character convention for representing Hmong syllables) but a canonical decomposition is given in the character properties. See Figure 3 for discussion of grounds for encoding these as script-specific characters.

2.4 Encoding order. Visual-order encoding should be preferred for Pahawh Hmong because it will make implementation less expensive and it is what users expect. The logical "reversal" of coda and head from the pronounced syllable does not affect the sorting algorithm, which follows visual order as well. Inputting and display are also done according to visual order. Unlike Devanagari, where a few vowel signs appear before the base consonant but should be represented phonetically in the backing store, *all* Pahawh Hmong syllables are uniformly represented as V^tC even though the pronunciation is CV^t. All current implementations employ this method of encoding.

3.0 Non-alphabetic characters are used in Pahawh Hmong.

3.1.0 Punctuation marks are found at 16B37–16B3C. Additional punctuation marks like ? ( ) . , ; : < > – — are used in Pahawh Hmong and have been unified with existing UCS characters.

3.1.1 Exclamation mark 16B38 𖬸 was invented by Pa Kao Her (*Paj Kaub Hawj* ℸՈ̊ Ո П̄Г [pâ káu hâw]) in 1985; Smalley and the Naadaa font retain a special glyph for this but the Cwjmem font either does not include it or prefers the generic exclamation mark. Shong Lue Yang also used "!".

3.1.2 Intonation mark 16B39 𖬹 indicates the sung or chanted nature of the text. It was also used by some Second Stage Reduced Version users to mark the -*d* tone.

3.1.3 Reduplication mark 16B3A 𖬺 indicates reduplication of the syllable preceding: ŰK̇𖬺 = ŰK̇ ŰK̇. *tsuag tsuag* [tʃṳa tʃṳa] 'hurry hurry'.

3.1.4 Ampersand 16B3B 𖬻 is derived from the ampersand and was also invented by Pa Kao Her. Smalley's font and the Naadaa font have a special glyph for this but in the Cwjmem font it faces the same direction as the generic ampersand.

3.1.5 Percent sign 16B3C 𖬼 is the percent sign. Smalley and the Naadaa font retain a special glyph for this but the Cwjmem font appears to modify the regular percent sign by having dots instead of rings.

3.2 Digits and numbers. 16B40–16B49 are the decimal digits 0–9. A nondecimal numeric system also exists, which makes use of 16B4A–16B50. It is not in current use. One complication is that some users employ 16B4A PAHAWH HMONG NUMBER TENS as a *zero*.

3.3 Grammatical classifier. 16B51 𖭑 represents the syllable *lub* [lú], the most common grammatical classifier in the Hmong language. Smalley *et al*. 1990 give the example 𖭑 *lub npe* [lú mbe] ‘a name’. Shong Lue Yang created a sign for this because of the high frequency of the word in the language, and considering the similarity of the two characters used to write it it seem that in devising the character Shong Lue Yang was being very practical indeed.

3.4 Logographs. 16B52–16B56 are logographs naming periods of time: *xyoo* ‘year’ 𖭒, *hli* ‘month’ 𖭓, *zwj thaj* ‘date’ 𖭔, *hnub* ‘day’ 𖭕, *ntuj* ‘season’ 𖭖 respectively.

3.5 Arithmetic operators. 16B57–16B5A are arithmetic operators. Smalley *et al*. 1990 give them, but they are not found in the fonts available from the Australian and Cwjmem communities.

3.6 Logographs for clan names. 16B60–16B71 are logographs for clan names. 16B60–16B6D were devised by Shong Lue Yang, and 16B6E–16B71 were added by Chia Koua Vang (*Txiaj Kuam Vaj* [tsîa kụa vâ]).

> According to Hmong custom, men and women from the same clan cannot marry each other, and are restricted in their behavior in each other’s presence. They are perceived to be like brothers and sisters so far as the appropriateness of sexual contact is concerned, with considerably more restrictions than exist in a sibling relationship in the West. For example, men and women of the same clan should not throw the ball to each other at the Hmong New Year, a custom potentially leading to courtship; neither should they spend time alone together.…
>
> Shong Lue Yang designed the clan logographs to be sewn into garments or worn as badges, or posted on desks or doors to identify a person’s clan. This would enable people to behave appropriately. Such identification was needed in the resettlement camps in Laos to which many Hmong people had fled for protection from the communists. In those surroundings they did not know all of their neighbors, much less other people they met.
>
> It is also sometimes hard to identify a person’s clan even if you have heard the person’s name. Order of given name and clan name is not fixed. Somebody called *Vaj Yaj* ‘Vang Yang’ might belong either to the *Vang* clan or the *Yang* clan, depending on which order is being used. Under conditions where strangers are regularly encountered, it is awkward to have to ask constantly what the other person’s clan is.… (Smalley *et al*. 1990:83–84)

These characters are not in current use, but are encoded for historical reasons.

4.0 Ordering. The ordering given in Lee Nao Long et al 2001, which uses the Second Stage Reduced Version orthography, follows the relative order of the tones, namely *-b* < *-m* < *-d* < *-j* < *-v* < -Ø < *-s* < *-g* (*v́* < *v̱* < *v̤* < *v̂* < *v̌* < *v* < *v̀* < *ṿ*). All stages use this tone-based ordering—where they differ is in which *characters* they use to represent the tones. This causes difficulties, in particular for a generic ordering based on the Second Stage Reduced Version.

In the presentation below, base characters are black, letters with CIM TUB are (using Web-named colours) dark slate blue, letters with CIM SO are dark goldenrod, letters with CIM KES are dark orange, letters with CIM KHAV are dark green, letters with CIM SUAM are crimson, letters with CIM HOM are dark magenta, and letters with CIM TAUM are dark cyan.

That is, while the Second Stage Reduced Version orthography begins:

∇̈ *ké* < ∇ *kę* < ∇̇ *ke̬* < ∇̇ *kê* < ∇̄ *kě* < ᒣ *ke* < ᒣ̇ *kè* < ᒣ̄ *ke̤*

The Third Stage Reduced Version orthography begins:

∇ *ké* < ∇̇ *kę* < ∇̇ *ke̬* < ∇̄ *kê* < ᒣ *kě* < ᒣ̇ *ke* < ᒣ̄ *kè* < ᒣ̈ *ke̤*

The Final Version orthography begins:

∇ *ké* < ∇̇ *kę* < ∇̇ *ke̬* < ∇̄ *kê* < ∇̂ *kě* < ∇̍ *ke* < ∇̌ *kè* < ∇̈ *ke̤*

For the purposes of a default ordering that easily supports both Third Stage Reduced Version and Final Version orthographies, each of the consonants and each of the vowels can be given a primary weight. This applies diacritical marks used only in the Final Version orthography to vowel rimes which do not appear in that Version, but it does form a complete specification. Since the consonant onsets are uniform in all stages, the list below gives only those forms which occur.

4.1 Generic ordering accommodating Third Stage Reduced Version and Final Version orthographies:

Vowel rimes:

∇ < ∇̇ < ∇̇ < ∇̄ < ∇̂ < ∇̍ < ∇̌ < ∇̈ < ᒣ < ᒣ̇ < ᒣ̇ < ᒣ̄ < ᒣ̂ < ᒣ̍ < ᒣ̌ < ᒣ̈ <
Ʌ < Ʌ̇ < Ʌ̇ < Ʌ̄ < Ʌ̂ < Ʌ̍ < Ʌ̌ < Ʌ̈ < ꓘ < ꓘ̇ < ꓘ̇ < ꓘ̄ < ꓘ̂ < ꓘ̍ < ꓘ̌ < ꓘ̈ <
ᑎ < ᑎ̇ < ᑎ̇ < ᑎ̄ < ᑎ̂ < ᑎ̍ < ᑎ̌ < ᑎ̈ < Ø < Ø̇ < Ø̇ < Ø̄ < Ø̂ < Ø̍ < Ø̌ < Ø̈ <
Ⴎ < Ⴎ̇ < Ⴎ̇ < Ⴎ̄ < Ⴎ̂ < Ⴎ̍ < Ⴎ̌ < Ⴎ̈ < Ո < Ո̇ < Ո̇ < Ո̄ < Ո̂ < Ո̍ < Ո̌ < Ո̈ <
Ʊ < Ʊ̇ < Ʊ̇ < Ʊ̄ < Ʊ̂ < Ʊ̍ < Ʊ̌ < Ʊ̈ < Ħ < Ħ̇ < Ħ̇ < Ħ̄ < Ħ̂ < Ħ̍ < Ħ̌ < Ħ̈ <
H < Ḣ < Ḣ < H̄ < Ĥ < H̍ < Ȟ < Ḧ < Ꮑ < Ꮑ̇ < Ꮑ̇ < Ꮑ̄ < Ꮑ̂ < Ꮑ̍ < Ꮑ̌ < Ꮑ̈ <
Ƕ < Ƕ̇ < Ƕ̇ < Ƕ̄ < Ƕ̂ < Ƕ̍ < Ƕ̌ < Ƕ̈ < ω < ω̇ < ω̇ < ω̄ < ω̂ < ω̍ < ω̌ < ω̈ <
Ͳ < Ͳ̇ < Ͳ̇ < Ͳ̄ < Ͳ̂ < Ͳ̍ < Ͳ̌ < Ͳ̈ < ɷ < ɷ̇ < ɷ̇ < ɷ̄ < ɷ̂ < ɷ̍ < ɷ̌ < ɷ̈ <
ΠU < ΠU̇ < ΠU̇ < ΠŪ < ΠÛ < ΠU̍ < ΠǓ < ΠÜ < Ш < Ш̇ < Ш̇ < Ш̄ < Ш̂ < Ш̍ < Ш̌ < Ш̈ <
Ո < Ո̇ < Ո̇ < Ո̄ < Ո̂ < Ո̍ < Ո̌ < Ո̈ < ᒍI < ᒍ̇I < ᒍ̇I < ᒍ̄I < ᒍ̂I < ᒍ̍I < ᒍ̌I < ᒍ̈I <
JJ < J̇J < J̇J < J̄J < ĴJ < J̍J < ǰJ < J̈J < ʌ < ʌ̇ < ʌ̇ < ʌ̄ < ʌ̂ < ʌ̍ < ʌ̌ < ʌ̈ <
Ꞅ < Ꞅ̇ < Ꞅ̇ < Ꞅ̄ < Ꞅ̂ < Ꞅ̍ < Ꞅ̌ < Ꞅ̈ < Ƒ < Ƒ̇ < Ƒ̇ < Ƒ̄ < Ƒ̂ < Ƒ̍ < Ƒ̌ < Ƒ̈ <
Ƭ < Ƭ̇ < Ƭ̇ < Ƭ̄ < Ƭ̂ < Ƭ̍ < Ƭ̌ < Ƭ̈ < Л < Л̇ < Л̇ < Л̄ < Л̂ < Л̍ < Л̌ < Л̈ <
ΙꞄ < ΙꞄ̇ < ΙꞄ̇ < ΙꞄ̄ < ΙꞄ̂ < ΙꞄ̍ < ΙꞄ̌ < ΙꞄ̈ < ΙƑ < ΙƑ̇ < ΙƑ̇ < ΙƑ̄ < ΙƑ̂ < ΙƑ̍ < ΙƑ̌ < ΙƑ̈ .

Consonant onsets:

Ͼ < Ͼ̇ < Ͼ̍ < ∩ < ∩̇ < ∩̍ < A < Ȧ < A̍ < V < V̇ < V̍ < ƖՈ < ƖՈ̇ < ƖՈ̍ <
ΠΓ < Π̇Γ < Π̍Γ < ШΓ < Ш̇Γ < Ш̍Γ < E < Ė < E̍ < Ƀ < Ƀ̇ < Ƀ̍ < ΓΓ < Γ̇Γ < Γ̍Γ <
Ʊ < Ʊ̇ < Ʊ̍ < ƱF < Ʊ̇F < Ʊ̍F < Ч < Ч̇ < Ч̍ < Ӈ < Ӈ̇ < Ӈ̍ < ƎH < Ǝ̇H < Ǝ̍H <
Ⱥ < Ⱥ̇ < Ⱥ̍ < ᗰ < ᗰ̇ < ᗰ̍ < M < Ṁ < M̍ < K < K̇ < K̍ < Ꮐ < Ꮐ̇ < Ꮐ̍ < Ꮐ̎ .

Notice that Ꮐ̎ *dlhau* is also found.

This regular ordering accounts for all base letters and diacritic combinations—even those which are never used (unused combinations are underscored). Although CIM SO, CIM KES, CIM KHAV, CIM SUAM, and CIM TAUM are never used with consonant onsets (which is why they are not shown in here), this scheme could just as easily accommodate the sequences Ͼ < Ͼ̇ < Ͼ̇ < Ͼ̄ < Ͼ̂ < Ͼ̍ < Ͼ̌ < Ͼ̈, etc.

4.2 Generic ordering accommodating Second Stage Reduced Version orthography. Since all combinations have to be accommodated in the ordering, the previous table has been taken as a base, and only those characters that needed to be moved have been.

Vowel rimes:

𖬀𖬲 < 𖬀 < 𖬀𖬱 < 𖬀𖬰 < 𖬀𖬳 < 𖬀𖬴 < 𖬀𖬵 < 𖬀𖬶 < 𖬁 < 𖬁𖬰 < 𖬁𖬱 < 𖬁𖬳 < 𖬁𖬴 < 𖬁𖬵 < 𖬁𖬶 < 𖬁𖬲 <
𖬂𖬳 < 𖬂 < 𖬂𖬱 < 𖬂𖬰 < 𖬂𖬲 < 𖬂𖬴 < 𖬂𖬵 < 𖬂𖬶 < 𖬃 < 𖬃𖬰 < 𖬃𖬱 < 𖬃𖬳 < 𖬃𖬴 < 𖬃𖬵 < 𖬃𖬶 < 𖬃𖬲 <
𖬄𖬰 < 𖬄 < 𖬄𖬱 < 𖬄𖬲 < 𖬄𖬳 < 𖬄𖬴 < 𖬄𖬵 < 𖬄𖬶 < 𖬅 < 𖬅𖬰 < 𖬅𖬱 < 𖬅𖬳 < 𖬅𖬴 < 𖬅𖬵 < 𖬅𖬶 < 𖬅𖬲 <
𖬆𖬰 < 𖬆 < 𖬆𖬱 < 𖬆𖬲 < 𖬆𖬳 < 𖬆𖬴 < 𖬆𖬵 < 𖬆𖬶 < 𖬇 < 𖬇𖬰 < 𖬇𖬱 < 𖬇𖬳 < 𖬇𖬴 < 𖬇𖬵 < 𖬇𖬶 < 𖬇𖬲 <
𖬈𖬰 < 𖬈 < 𖬈𖬱 < 𖬈𖬳 < 𖬈𖬴 < 𖬈𖬵 < 𖬈𖬶 < 𖬈𖬲 < 𖬉 < 𖬉𖬰 < 𖬉𖬱 < 𖬉𖬳 < 𖬉𖬴 < 𖬉𖬵 < 𖬉𖬶 < 𖬉𖬲 <
𖬊𖬰 < 𖬊 < 𖬊𖬱 < 𖬊𖬲 < 𖬊𖬳 < 𖬊𖬴 < 𖬊𖬵 < 𖬊𖬶 < 𖬋 < 𖬋𖬰 < 𖬋𖬱 < 𖬋𖬳 < 𖬋𖬴 < 𖬋𖬵 < 𖬋𖬶 < 𖬋𖬲 <
𖬌 < 𖬌𖬰 < 𖬌𖬱 < 𖬌𖬳 < 𖬌𖬴 < 𖬌𖬵 < 𖬌𖬶 < 𖬌𖬲 < 𖬍 < 𖬍𖬰 < 𖬍𖬱 < 𖬍𖬳 < 𖬍𖬴 < 𖬍𖬵 < 𖬍𖬶 < 𖬍𖬲 <
𖬎 < 𖬎𖬰 < 𖬎𖬱 < 𖬎𖬳 < 𖬎𖬲 < 𖬎𖬴 < 𖬎𖬵 < 𖬎𖬶 < 𖬏 < 𖬏𖬰 < 𖬏𖬱 < 𖬏𖬳 < 𖬏𖬴 < 𖬏𖬵 < 𖬏𖬶 < 𖬏𖬲 <
𖬐𖬲 < 𖬐 < 𖬐𖬱 < 𖬐𖬰 < 𖬐𖬳 < 𖬐𖬴 < 𖬐𖬵 < 𖬐𖬶 < 𖬑 < 𖬑𖬰 < 𖬑𖬱 < 𖬑𖬳 < 𖬑𖬴 < 𖬑𖬵 < 𖬑𖬶 < 𖬑𖬲 <
𖬒𖬰 < 𖬒 < 𖬒𖬱 < 𖬒𖬳 < 𖬒𖬲 < 𖬒𖬴 < 𖬒𖬵 < 𖬒𖬶 < 𖬓 < 𖬓𖬰 < 𖬓𖬱 < 𖬓𖬳 < 𖬓𖬴 < 𖬓𖬵 < 𖬓𖬶 < 𖬓𖬲 <
𖬔 < 𖬔𖬰 < 𖬔𖬱 < 𖬔𖬲 < 𖬔𖬳 < 𖬔𖬴 < 𖬔𖬵 < 𖬔𖬶 < 𖬕 < 𖬕𖬰 < 𖬕𖬱 < 𖬕𖬳 < 𖬕𖬴 < 𖬕𖬵 < 𖬕𖬶 < 𖬕𖬲 <
𖬖𖬳 < 𖬖 < 𖬖𖬱 < 𖬖𖬰 < 𖬖𖬴 < 𖬖𖬵 < 𖬖𖬶 < 𖬖𖬲 < 𖬗 < 𖬗𖬰 < 𖬗𖬱 < 𖬗𖬳 < 𖬗𖬴 < 𖬗𖬵 < 𖬗𖬶 < 𖬗𖬲 <
𖬘𖬰 < 𖬘 < 𖬘𖬱 < 𖬘𖬳 < 𖬘𖬴 < 𖬘𖬵 < 𖬘𖬶 < 𖬘𖬲 < 𖬙 < 𖬙𖬰 < 𖬙𖬱 < 𖬙𖬳 < 𖬙𖬴 < 𖬙𖬵 < 𖬙𖬶 < 𖬙𖬲 <
𖬚𖬳 < 𖬚 < 𖬚𖬱 < 𖬚𖬰 < 𖬚𖬴 < 𖬚𖬵 < 𖬚𖬶 < 𖬚𖬲 < 𖬛 < 𖬛𖬰 < 𖬛𖬱 < 𖬛𖬳 < 𖬛𖬴 < 𖬛𖬵 < 𖬛𖬶 < 𖬛𖬲 .

Consonant onsets:

𖬜 < 𖬜𖬰 < 𖬜𖬶 < 𖬝 < 𖬝𖬰 < 𖬝𖬶 < 𖬞 < 𖬞𖬰 < 𖬞𖬶 < 𖬟 < 𖬟𖬰 < 𖬟𖬶 < 𖬠 < 𖬠𖬰 < 𖬠𖬶 <
𖬡 < 𖬡𖬰 < 𖬡𖬶 < 𖬢 < 𖬢𖬰 < 𖬢𖬶 < 𖬣 < 𖬣𖬰 < 𖬣𖬶 < 𖬤 < 𖬤𖬰 < 𖬤𖬶 < 𖬥 < 𖬥𖬰 < 𖬥𖬶 <
𖬦 < 𖬦𖬰 < 𖬦𖬶 < 𖬧 < 𖬧𖬰 < 𖬧𖬶 < 𖬨 < 𖬨𖬰 < 𖬨𖬶 < 𖬩 < 𖬩𖬰 < 𖬩𖬶 < 𖬪 < 𖬪𖬰 < 𖬪𖬶 <
𖬫 < 𖬫𖬰 < 𖬫𖬶 < 𖬬 < 𖬬𖬰 < 𖬬𖬶 < 𖬭 < 𖬭𖬰 < 𖬭𖬶 < 𖬮 < 𖬮𖬰 < 𖬮𖬶 < 𖬯 < 𖬯𖬰 < 𖬯𖬵 < 𖬯𖬶 .

Notice that 𖬯𖬵 *dlhau* is also found.

It should also be noted that there are two schemes for the ordering of the vowel rimes in the Second Stage Reduced Version, and no standard order for the consonant onsets. Taking only the base characters, we find:

4.3.1 Vowel rime order in Smalley 1990, www.pahauhhmong.org/yublas.html, Lee Nao Long 2001, Chia Koua Vang 2002

𖬀 < 𖬁 < 𖬂 < 𖬃 < 𖬄 < 𖬅 < 𖬆 < 𖬇 < 𖬈 < 𖬉 < 𖬊 < 𖬋 < 𖬌 < 𖬍 <
𖬎 < 𖬏 < 𖬐 < 𖬑 < 𖬒 < 𖬓 < 𖬔 < 𖬕 < 𖬖 < 𖬗 < 𖬘 < 𖬙 < 𖬚 < 𖬛 .

4.3.2 Vowel rime order in www.cwjmemhmong.info/vowels.html

𖬀 < 𖬁 < 𖬐 < 𖬑 < 𖬔 < 𖬕 < 𖬂 < 𖬃 < 𖬆 < 𖬇 < 𖬒 < 𖬓 < 𖬎 < 𖬏 <
𖬊 < 𖬋 < 𖬄 < 𖬅 < 𖬈 < 𖬉 < 𖬘 < 𖬙 < 𖬖 < 𖬗 < 𖬌 < 𖬍 < 𖬚 < 𖬛 .

4.4.1 Consonant onset order in Lee Nao Long *et al*. 2001

𖬜 𖬝 𖬞 𖬟 𖬠 𖬡 𖬢 𖬣 𖬤 𖬥 𖬦 𖬧 𖬨 𖬩 𖬪 𖬫 𖬬 𖬭 𖬮 𖬯

4.4.2 Consonant onset order in Smalley *et al*. 1990

𖬜 𖬝 𖬞 𖬟 𖬠 𖬮 𖬨 𖬤 𖬩 𖬬 𖬭 𖬪 𖬦 𖬧 𖬣 𖬡 𖬫 𖬥 𖬢 𖬯

4.4.3 Consonant onset order in www.pahauhhmong.org/yublas.html

𖬜 𖬝 𖬡 𖬞 𖬟 𖬠 𖬩 𖬦 𖬢 𖬨 𖬪 𖬤 𖬧 𖬣 𖬬 𖬫 𖬭 𖬮 𖬥 𖬯

4.4.4 Consonant onset order in www.cwjmemhmong.info/vowels.html

𖬜 𖬝 𖬠 𖬩 𖬮 𖬤 𖬣 𖬧 𖬞 𖬬 𖬦 𖬢 𖬪 𖬫 𖬨 𖬭 𖬡 𖬟 𖬥 𖬯

4.4.5 Consonant onset order in Chia Koua Vang 2002

𖬥 𖬮 𖬢 𖬡 𖬠 𖬟 𖬜 𖬬 𖬪 𖬞 𖬩 𖬨 𖬝 𖬦 𖬤 𖬭 𖬫 𖬣 𖬧 𖬯

In the code charts, the order of characters follows 5.3.1 and 5.4.1 above.

5.0 Character names. The chief problem in encoding Pahawh Hmong involves what to name the vowel rimes, because the values given to the base letters in the Second Stage Reduced Version and the Third Stage Reduced Version are not compatible. The table to the right here shows the problem: the expected order is the order of the tones (left to right then top to bottom), regardless of the shape of the glyphs. The black glyphs in the table here (without diacritics) should be the source names for the characters.

In the code chart below, the names used are Third Stage Reduced Version names based on the regular paradigm.

Second Stage Reduced Version vowel rimes:

kéŋ	*kẹŋ*	*kẹŋ*	*kêŋ*	*kěŋ*	*keŋ*	*kèŋ*	*kẹŋ*
kí	*kị*	*kị*	*kî*	*kǐ*	*ki*	*kì*	*kị*
káu	*kạu*	*kạu*	*kâu*	*kău*	*kau*	*kàu*	*kạu*
kú	*kụ*	*kụ*	*kû*	*kŭ*	*ku*	*kù*	*kụ*
ké	*kẹ*	*kẹ*	*kê*	*kě*	*ke*	*kè*	*kẹ*
kái	*kại*	*kại*	*kâi*	*kăi*	*kai*	*kài*	*kại*
kóŋ	*kọŋ*	*kọŋ*	*kôŋ*	*kǒŋ*	*koŋ*	*kòŋ*	*kọŋ*
káw	*kạw*	*kạw*	*kâw*	*kăw*	*kaw*	*kàw*	*kạw*
kúa	*kụa*	*kụa*	*kûa*	*kŭa*	*kua*	*kùa*	*kụa*
kó	*kọ*	*kọ*	*kô*	*kǒ*	*ko*	*kò*	*kọ*
kía	*kịa*	*kịa*	*kîa*	*kĭa*	*kia*	*kìa*	*kịa*
ká	*kạ*	*kạ*	*kâ*	*kă*	*ka*	*kà*	*kạ*
kẃ	*kẉ*	*kẉ*	*kŵ*	*kw̌*	*kw*	*kẁ*	*kẉ*
káa	*kạa*	*kạa*	*kâa*	*kăa*	*kaa*	*kàa*	*kạa*

Third Stage Reduced Version vowel rimes:

kéŋ	*kẹŋ*	*kẹŋ*	*kêŋ*	*kěŋ*	*keŋ*	*kèŋ*	*kẹŋ*
kí	*kị*	*kị*	*kî*	*kǐ*	*ki*	*kì*	*ki*
káu	*kạu*	*kạu*	*kâu*	*kău*	*kau*	*kàu*	*kau*
kú	*kụ*	*kụ*	*kû*	*kŭ*	*ku*	*kù*	*ku*
ké	*kẹ*	*kẹ*	*kê*	*kě*	*ke*	*kè*	*ke*
kái	*kại*	*kại*	*kâi*	*kăi*	*kai*	*kài*	*kai*
kóŋ	*kọŋ*	*kọŋ*	*kôŋ*	*kǒŋ*	*koŋ*	*kòŋ*	*koŋ*
káw	*kạw*	*kạw*	*kâw*	*kăw*	*kaw*	*kàw*	*kaw*
kúa	*kụa*	*kụa*	*kûa*	*kŭa*	*kua*	*kùa*	*kua*
kó	*kọ*	*kọ*	*kô*	*kǒ*	*ko*	*kò*	*ko*
kía	*kịa*	*kịa*	*kîa*	*kĭa*	*kia*	*kìa*	*kia*
ká	*kạ*	*kạ*	*kâ*	*kă*	*ka*	*kà*	*ka*
kẃ	*kẉ*	*kẉ*	*kŵ*	*kw̌*	*kw*	*kẁ*	*kw*
káa	*kạa*	*kạa*	*kâa*	*kăa*	*kaa*	*kàa*	*kaa*

Final Version vowel rimes:

kéŋ	*kẹŋ*	*kẹŋ*	*kêŋ*	*kěŋ*	*keŋ*	*kèŋ*	*kẹŋ*
kí	*kị*	*kị*	*kî*	*kǐ*	*ki*	*kì*	*ki*
káu	*kạu*	*kạu*	*kâu*	*kău*	*kau*	*kàu*	*kau*
kú	*kụ*	*kụ*	*kû*	*kŭ*	*ku*	*kù*	*ku*
ké	*kẹ*	*kẹ*	*kê*	*kě*	*ke*	*kè*	*ke*
kái	*kại*	*kại*	*kâi*	*kăi*	*kai*	*kài*	*kai*
kóŋ	*kọŋ*	*kọŋ*	*kôŋ*	*kǒŋ*	*koŋ*	*kòŋ*	*koŋ*
káw	*kạw*	*kạw*	*kâw*	*kăw*	*kaw*	*kàw*	*kaw*
kúa	*kụa*	*kụa*	*kûa*	*kŭa*	*kua*	*kùa*	*kua*
kó	*kọ*	*kọ*	*kô*	*kǒ*	*ko*	*kò*	*ko*
kía	*kịa*	*kịa*	*kîa*	*kĭa*	*kia*	*kìa*	*kia*
ká	*kạ*	*kạ*	*kâ*	*kă*	*ka*	*kà*	*ka*
kẃ	*kẉ*	*kẉ*	*kŵ*	*kw̌*	*kw*	*kẁ*	*kw*
káa	*kạa*	*kạa*	*kâa*	*kăa*	*kaa*	*kàa*	*kaa*

5.1 Resolving the Vowel Rime Names

For the Second Stage Reduced Version and Third State Reduced Version vowel rimes, whose names would be most accepted and used by the communities, the vowel rime names derived from the tables shown above would be:

Glyph	*Second*	*Third*
𖬀	keem	keeb
𖬁	kee	keev
𖬂	kim	kib
𖬃	ki	kiv
𖬄	kaum	kaub
𖬅	kau	kauv
𖬆	kum	kub
𖬇	ke	kuv
𖬈	kem	keb
𖬉	*kev*	*kev*
𖬊	kaim	kaib
𖬋	kai	kaiv
𖬌	*koob*	*koob*
𖬍	*koov*	*koov*
𖬎	*kawb*	*kawb*
𖬏	kaw	kawv
𖬐	kuam	kuab
𖬑	kua	kuav
𖬒	kom	kob
𖬓	kog	kov
𖬔	*kiab*	*kiab*
𖬕	kia	kiav
𖬖	kam	kab
𖬗	*kav*	*kav*
𖬘	kwm	kwb
𖬙	*kwv*	*kwv*
𖬚	kaam	kaab
𖬛	*kaav*	*kaav*

Where these vowel rime names are identical, they are simply used as the name for the corresponding character in the code chart. Where they are not identical (identical ones are *italicized* above), a choice has to be made for the encoded character name, and the proposed choice is to use the Third Stage Reduced Version names in those cases, for consistency. In all cases, where the Second Stage and Third Stage names differ, the Second Stage name is added to the code chart as an alias, so that users of either system can easily find names appropriate to their usage.

Note that while the spellings of these vowel rimes in Latin letters is rather different, the differences are in the final letters, which transcribe the tones for the syllables. So the actual difference in the syllables used to represent the names is just in the tones used for them.

```
16B00;PAHAWH HMONG VOWEL KEEB;Lo;0;L;;;;;N;;;;;
16B01;PAHAWH HMONG VOWEL KEEV;Lo;0;L;;;;;N;;;;;
16B02;PAHAWH HMONG VOWEL KIB;Lo;0;L;;;;;N;;;;;
16B03;PAHAWH HMONG VOWEL KIV;Lo;0;L;;;;;N;;;;;
16B04;PAHAWH HMONG VOWEL KAUB;Lo;0;L;;;;;N;;;;;
16B05;PAHAWH HMONG VOWEL KAUV;Lo;0;L;;;;;N;;;;;
16B06;PAHAWH HMONG VOWEL KUB;Lo;0;L;;;;;N;;;;;
16B07;PAHAWH HMONG VOWEL KUV;Lo;0;L;;;;;N;;;;;
16B08;PAHAWH HMONG VOWEL KEB;Lo;0;L;;;;;N;;;;;
16B09;PAHAWH HMONG VOWEL KEV;Lo;0;L;;;;;N;;;;;
16B0A;PAHAWH HMONG VOWEL KAIB;Lo;0;L;;;;;N;;;;;
16B0B;PAHAWH HMONG VOWEL KAIV;Lo;0;L;;;;;N;;;;;
16B0C;PAHAWH HMONG VOWEL KOOB;Lo;0;L;;;;;N;;;;;
16B0D;PAHAWH HMONG VOWEL KOOV;Lo;0;L;;;;;N;;;;;
16B0E;PAHAWH HMONG VOWEL KAWB;Lo;0;L;;;;;N;;;;;
16B0F;PAHAWH HMONG VOWEL KAWV;Lo;0;L;;;;;N;;;;;
16B10;PAHAWH HMONG VOWEL KUAB;Lo;0;L;;;;;N;;;;;
16B11;PAHAWH HMONG VOWEL KUAV;Lo;0;L;;;;;N;;;;;
16B12;PAHAWH HMONG VOWEL KOB;Lo;0;L;;;;;N;;;;;
16B13;PAHAWH HMONG VOWEL KOV;Lo;0;L;;;;;N;;;;;
16B14;PAHAWH HMONG VOWEL KIAB;Lo;0;L;;;;;N;;;;;
16B15;PAHAWH HMONG VOWEL KIAV;Lo;0;L;;;;;N;;;;;
16B16;PAHAWH HMONG VOWEL KAB;Lo;0;L;;;;;N;;;;;
16B17;PAHAWH HMONG VOWEL KAV;Lo;0;L;;;;;N;;;;;
16B18;PAHAWH HMONG VOWEL KWB;Lo;0;L;;;;;N;;;;;
16B19;PAHAWH HMONG VOWEL KWV;Lo;0;L;;;;;N;;;;;
16B1A;PAHAWH HMONG VOWEL KAAB;Lo;0;L;;;;;N;;;;;
16B1B;PAHAWH HMONG VOWEL KAAV;Lo;0;L;;;;;N;;;;;
16B1C;PAHAWH HMONG CONSONANT VAU;Lo;0;L;;;;;N;;;;;
16B1D;PAHAWH HMONG CONSONANT NKAU;Lo;0;L;;;;;N;;;;;
16B1E;PAHAWH HMONG CONSONANT XAU;Lo;0;L;;;;;N;;;;;
16B1F;PAHAWH HMONG CONSONANT CAU;Lo;0;L;;;;;N;;;;;
16B20;PAHAWH HMONG CONSONANT LAU;Lo;0;L;;;;;N;;;;;
16B21;PAHAWH HMONG CONSONANT HAU;Lo;0;L;;;;;N;;;;;
16B22;PAHAWH HMONG CONSONANT YAU;Lo;0;L;;;;;N;;;;;
16B23;PAHAWH HMONG CONSONANT QHAU;Lo;0;L;;;;;N;;;;;
16B24;PAHAWH HMONG CONSONANT RAU;Lo;0;L;;;;;N;;;;;
16B25;PAHAWH HMONG CONSONANT MAU;Lo;0;L;;;;;N;;;;;
16B26;PAHAWH HMONG CONSONANT NAU;Lo;0;L;;;;;N;;;;;
16B27;PAHAWH HMONG CONSONANT NLAU;Lo;0;L;;;;;N;;;;;
16B28;PAHAWH HMONG CONSONANT HLAU;Lo;0;L;;;;;N;;;;;
16B29;PAHAWH HMONG CONSONANT HNAU;Lo;0;L;;;;;N;;;;;
16B2A;PAHAWH HMONG CONSONANT CHAU;Lo;0;L;;;;;N;;;;;
16B2B;PAHAWH HMONG CONSONANT NCHAU;Lo;0;L;;;;;N;;;;;
16B2C;PAHAWH HMONG CONSONANT PLHAU;Lo;0;L;;;;;N;;;;;
16B2D;PAHAWH HMONG CONSONANT NTHAU;Lo;0;L;;;;;N;;;;;
16B2E;PAHAWH HMONG CONSONANT NTSAU;Lo;0;L;;;;;N;;;;;
16B2F;PAHAWH HMONG CONSONANT AU;Lo;0;L;;;;;N;;;;;
16B30;PAHAWH HMONG MARK CIM TUB;Mn;230;NSM;;;;;N;;;;;
16B31;PAHAWH HMONG MARK CIM SO;Mn;230;NSM;;;;;N;;;;;
16B32;PAHAWH HMONG MARK CIM KES;Mn;230;NSM;;;;;N;;;;;
16B33;PAHAWH HMONG MARK CIM KHAV;Mn;230;NSM;;;;;N;;;;;
16B34;PAHAWH HMONG MARK CIM SUAM;Mn;230;NSM;16B32 16B30;;;;N;;;;;
16B35;PAHAWH HMONG MARK CIM HOM;Mn;230;NSM;16B32 16B31;;;;N;;;;;
16B36;PAHAWH HMONG MARK CIM TAUM;Mn;230;NSM;;;;;N;;;;;
16B37;PAHAWH HMONG SIGN VOS THOM;Po;0;L;;;;;N;;;;;
16B38;PAHAWH HMONG SIGN VOS TSHAB CEEB;Po;0;L;;;;;N;;;;;
16B39;PAHAWH HMONG SIGN VOS SEEV;Lm;0;L;;;;;N;;;;;
16B3A;PAHAWH HMONG SIGN VOS NRUA;Lm;0;L;;;;;N;;;;;
16B3B;PAHAWH HMONG SIGN VOS THIAB;Po;0;L;;;;;N;;;;;
16B3C;PAHAWH HMONG SIGN VOS FEEM;Po;0;L;;;;;N;;;;;
16B40;PAHAWH HMONG DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
16B41;PAHAWH HMONG DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
16B42;PAHAWH HMONG DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
16B43;PAHAWH HMONG DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
16B44;PAHAWH HMONG DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
16B45;PAHAWH HMONG DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
16B46;PAHAWH HMONG DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
16B47;PAHAWH HMONG DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
16B48;PAHAWH HMONG DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
16B49;PAHAWH HMONG DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
16B4A;PAHAWH HMONG NUMBER TENS;No;0;L;;;;10;N;;;;;
16B4B;PAHAWH HMONG NUMBER HUNDREDS;No;0;L;;;;100;N;;;;;
16B4C;PAHAWH HMONG NUMBER TEN THOUSANDS;No;0;L;;;;10000;N;;;;;
16B4D;PAHAWH HMONG NUMBER MILLIONS;No;0;L;;;;1000000;N;;;;;
16B4E;PAHAWH HMONG NUMBER HUNDRED MILLIONS;No;0;L;;;;100000000;N;;;;;
16B4F;PAHAWH HMONG NUMBER TEN THOUSAND MILLIONS;No;0;L;;;;10000000000;N;;;;;
16B50;PAHAWH HMONG NUMBER BILLIONS;No;0;L;;;;1000000000000;N;;;;;
16B51;PAHAWH HMONG SIGN VOS LUB;Lm;0;L;;;;;N;;;;;
16B52;PAHAWH HMONG SIGN XYOO;So;0;L;;;;;N;;;;;
16B53;PAHAWH HMONG SIGN HLI;So;0;L;;;;;N;;;;;
16B54;PAHAWH HMONG SIGN ZWJ THAJ;So;0;L;;;;;N;;;;;
16B55;PAHAWH HMONG SIGN HNUB;So;0;L;;;;;N;;;;;
16B56;PAHAWH HMONG SIGN NTUJ;So;0;L;;;;;N;;;;;
16B57;PAHAWH HMONG SIGN XYEEM NTXIV;Sm;0;ES;;;;;N;;;;;
16B58;PAHAWH HMONG SIGN XYEEM RHO;Sm;0;ES;;;;;N;;;;;
```

```
16B59;PAHAWH HMONG SIGN XYEEM TOV;Sm;0;ES;;;;;N;;;;;
16B5A;PAHAWH HMONG SIGN XYEEM FAIB;Sm;0;ES;;;;;N;;;;;
16B60;PAHAWH HMONG CLAN SIGN YEEG;So;0;L;;;;;N;;;;;
16B61;PAHAWH HMONG CLAN SIGN LIS;So;0;L;;;;;N;;;;;
16B62;PAHAWH HMONG CLAN SIGN LAUJ;So;0;L;;;;;N;;;;;
16B63;PAHAWH HMONG CLAN SIGN XYOOJ;So;0;L;;;;;N;;;;;
16B64;PAHAWH HMONG CLAN SIGN HAWJ;So;0;L;;;;;N;;;;;
16B65;PAHAWH HMONG CLAN SIGN MUAS;So;0;L;;;;;N;;;;;
16B66;PAHAWH HMONG CLAN SIGN THOJ;So;0;L;;;;;N;;;;;
16B67;PAHAWH HMONG CLAN SIGN TSAB;So;0;L;;;;;N;;;;;
16B68;PAHAWH HMONG CLAN SIGN KHAB;So;0;L;;;;;N;;;;;
16B69;PAHAWH HMONG CLAN SIGN HAM;So;0;L;;;;;N;;;;;
16B6A;PAHAWH HMONG CLAN SIGN VAJ;So;0;L;;;;;N;;;;;
16B6B;PAHAWH HMONG CLAN SIGN YAJ;So;0;L;;;;;N;;;;;
16B6C;PAHAWH HMONG CLAN SIGN KWM;So;0;L;;;;;N;;;;;
16B6D;PAHAWH HMONG CLAN SIGN VWJ;So;0;L;;;;;N;;;;;
16B6E;PAHAWH HMONG CLAN SIGN TSHEEJ;So;0;L;;;;;N;;;;;
16B6F;PAHAWH HMONG CLAN SIGN KOO;So;0;L;;;;;N;;;;;
16B70;PAHAWH HMONG CLAN SIGN FAJ;So;0;L;;;;;N;;;;;
16B71;PAHAWH HMONG CLAN SIGN TSWB;So;0;L;;;;;N;;;;;
```

7. Unicode Linebreaking Properties. The vowel rimes and consonant onsets 16B00..16B2F behave like letters. The marks 16B30..16B36 behave like combining diacritics. The punctuation marks 16B37..16B38 behave like ? and ! respectively. The characters 16B39..16B3A behave like letters. The punctuation marks 16B3B..16B3C behave like & and % respectively. The digits and numbers 16B40..16B50 behave like numerals. The logographs 16B51..16B56 behave like letters. The arithmetical symbols 16B57..16B5A behave like +, –, ×, and ÷ respectively. The clan signs 16B60..16B71 behave like letters.

8. References

Cwjmem font by JYK of Hmong Script Software. http://www.cwjmemhmong.info/

Naadaa font by Vang Peng Yang. http://www.linguistics.unimelb.edu.au/research/hmong (*not working*)

Pahauh font. http://www.pahauhhmong.org/

Pahawh font. http://www.hmongwriting.com/

Chia Koua Vang. 2002. ՈՈՐ ՐՍ ԱԷ: ՌԼՈ З (*Moj Kuab Hmoob: Lawj 2*) = *Hmong Language: Second Grade*. St. Paul: Hmong National Organization.

Lee Nao Long, Yang Lee Hue, Yang Nao Shoua, Christina Eira. 2001. Ṿ ԱԷ Ḱ ՐՍ ṾՐ ɅԱ (*Keeb Hmoob Phau Kuab Meej Sim*) = *The Hmong bilingual dictionary*. [s.l., Australia]: Yang Shong Lue's Millennium Academy for the Hmong Language and Writing System.

Ratliff, Martha. 1996. "The Pahawh Hmong script", in Peter T. Daniels and William Bright, eds. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Smalley, William A., Chia Koua Vang, & Gnia Yee Yang. 1990. *Mother of writing: the origin and development of a Hmong messianic script*. Chicago & London: University of Chicago Press. ISBN 0-226-76287-4

9. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Pahawh Hmong. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.


16B0 16B1 16B2 16B3 16B4 16B5 16B6 16B7
0 16B00 16B10 16B20 16B30 16B40 16B50 16B60 16B70
1 16B01 16B11 16B21 16B31 16B41 16B51 16B61 16B71
2 16B02 16B12 16B22 16B32 16B42 16B52 16B62
3 16B03 16B13 16B23 16B33 16B43 16B53 16B63
4 16B04 16B14 16B24 16B34 16B44 16B54 16B64
5 16B05 16B15 16B25 16B35 16B45 16B55 16B65
6 16B06 16B16 16B26 16B36 16B46 16B56 16B66
7 16B07 16B17 16B27 16B37 16B47 16B57 16B67
8 16B08 16B18 16B28 16B38 16B48 16B58 16B68
9 16B09 16B19 16B29 16B39 16B49 16B59 16B69
A 16B0A 16B1A 16B2A 16B3A 16B4A 16B5A 16B6A
B 16B0B 16B1B 16B2B 16B3B 16B4B 16B6B
C 16B0C 16B1C 16B2C 16B3C 16B4C 16B6C
D 16B0D 16B1D 16B2D 16B4D 16B6D
E 16B0E 16B1E 16B2E 16B4E 16B6E
F 16B0F 16B1F 16B2F 16B4F 16B6F

The character names used for Pahawh Hmong follow the Third Stage Reduced Version orthography. Annotations give the character names in the Second Stage Reduced Version orthography.

Vowel rimes

16B00 PAHAWH HMONG VOWEL KEEB
= keem
16B01 PAHAWH HMONG VOWEL KEEV
= kee
16B02 PAHAWH HMONG VOWEL KIB
= kim
16B03 PAHAWH HMONG VOWEL KIV
= ki
16B04 PAHAWH HMONG VOWEL KAUB
= kaum
16B05 PAHAWH HMONG VOWEL KAUV
= kau
16B06 PAHAWH HMONG VOWEL KUB
= kum
16B07 PAHAWH HMONG VOWEL KUV
= ke
16B08 PAHAWH HMONG VOWEL KEB
= kem
16B09 PAHAWH HMONG VOWEL KEV
16B0A PAHAWH HMONG VOWEL KAIB
= kaim
16B0B PAHAWH HMONG VOWEL KAIV
= kai
16B0C PAHAWH HMONG VOWEL KOOB
16B0D PAHAWH HMONG VOWEL KOOV
16B0E PAHAWH HMONG VOWEL KAWB
16B0F PAHAWH HMONG VOWEL KAWV
= kaw
16B10 PAHAWH HMONG VOWEL KUAB
= kuam
16B11 PAHAWH HMONG VOWEL KUAV
= kua
16B12 PAHAWH HMONG VOWEL KOB
= kom
16B13 PAHAWH HMONG VOWEL KOV
= kog
16B14 PAHAWH HMONG VOWEL KIAB
16B15 PAHAWH HMONG VOWEL KIAV
= kia
16B16 PAHAWH HMONG VOWEL KAB
= kam
16B17 PAHAWH HMONG VOWEL KAV
16B18 PAHAWH HMONG VOWEL KWB
= kwm
16B19 PAHAWH HMONG VOWEL KWV
16B1A PAHAWH HMONG VOWEL KAAB
= kaam
16B1B PAHAWH HMONG VOWEL KAAV

Consonant onsets

16B1C PAHAWH HMONG CONSONANT VAU
16B1D PAHAWH HMONG CONSONANT NKAU
16B1E PAHAWH HMONG CONSONANT XAU
16B1F PAHAWH HMONG CONSONANT CAU
16B20 PAHAWH HMONG CONSONANT LAU
16B21 PAHAWH HMONG CONSONANT HAU
16B22 PAHAWH HMONG CONSONANT YAU
16B23 PAHAWH HMONG CONSONANT QHAU
16B24 PAHAWH HMONG CONSONANT RAU
16B25 PAHAWH HMONG CONSONANT MAU
16B26 PAHAWH HMONG CONSONANT NAU
16B27 PAHAWH HMONG CONSONANT NLAU
16B28 PAHAWH HMONG CONSONANT HLAU
16B29 PAHAWH HMONG CONSONANT HNAU
16B2A PAHAWH HMONG CONSONANT CHAU
16B2B PAHAWH HMONG CONSONANT NCHAU
16B2C PAHAWH HMONG CONSONANT PLHAU
16B2D PAHAWH HMONG CONSONANT NTHAU
16B2E PAHAWH HMONG CONSONANT NTSAU
16B2F PAHAWH HMONG CONSONANT AU

Combining diacritical marks

16B30 ◌ PAHAWH HMONG MARK CIM TUB
16B31 ◌ PAHAWH HMONG MARK CIM SO
16B32 ◌ PAHAWH HMONG MARK CIM KES
16B33 ◌ PAHAWH HMONG MARK CIM KHAV
16B34 ◌ PAHAWH HMONG MARK CIM SUAM
16B35 ◌ PAHAWH HMONG MARK CIM HOM
16B36 ◌ PAHAWH HMONG MARK CIM TAUM

Punctuation

16B37 PAHAWH HMONG SIGN VOS THOM
= question mark
16B38 PAHAWH HMONG SIGN VOS TSHAB CEEB
= exclamation mark
16B39 PAHAWH HMONG SIGN VOS SEEV
= chanting intonation
16B3A PAHAWH HMONG SIGN VOS NRUA
= reduplication
16B3B PAHAWH HMONG SIGN VOS THIAB
= ampersand
16B3C PAHAWH HMONG SIGN VOS FEEM
= percent sign

Digits

16B40 PAHAWH HMONG DIGIT ZERO
16B41 PAHAWH HMONG DIGIT ONE
16B42 PAHAWH HMONG DIGIT TWO
16B43 PAHAWH HMONG DIGIT THREE
16B44 PAHAWH HMONG DIGIT FOUR
16B45 PAHAWH HMONG DIGIT FIVE
16B46 PAHAWH HMONG DIGIT SIX
16B47 PAHAWH HMONG DIGIT SEVEN
16B48 PAHAWH HMONG DIGIT EIGHT
16B49 PAHAWH HMONG DIGIT NINE

Numbers

16B4A PAHAWH HMONG NUMBER TENS
16B4B PAHAWH HMONG NUMBER HUNDREDS
16B4C PAHAWH HMONG NUMBER TEN THOUSANDS
16B4D PAHAWH HMONG NUMBER MILLIONS
16B4E PAHAWH HMONG NUMBER HUNDRED MILLIONS
= billions
16B4F PAHAWH HMONG NUMBER TEN THOUSAND MILLIONS
= ten billions

16B50 𖭐 PAHAWH HMONG NUMBER BILLIONS
= trillions

Logographs

16B51 𖭑 PAHAWH HMONG SIGN VOS LUB
= classifier

16B52 𖭒 PAHAWH HMONG SIGN XYOO
= year

16B53 𖭓 PAHAWH HMONG SIGN HLI
= month

16B54 𖭔 PAHAWH HMONG SIGN ZWJ THAJ
= date

16B55 𖭕 PAHAWH HMONG SIGN HNUB
= day

16B56 𖭖 PAHAWH HMONG SIGN NTUJ
= season

Arithmetical symbols

16B57 𖭗 PAHAWH HMONG SIGN XYEEM NTXIV
= plus sign

16B58 𖭘 PAHAWH HMONG SIGN XYEEM RHO
= minus sign

16B59 𖭙 PAHAWH HMONG SIGN XYEEM TOV
= multiplication sign

16B5A 𖭚 PAHAWH HMONG SIGN XYEEM FAIB
= division sign

Logographs for clan names

16B60 𖭛 PAHAWH HMONG CLAN SIGN YEEG
= Yeng

16B61 𖭜 PAHAWH HMONG CLAN SIGN LIS
= Lee

16B62 𖭝 PAHAWH HMONG CLAN SIGN LAUJ
= Lor

16B63 𖭞 PAHAWH HMONG CLAN SIGN XYOOJ
= Xiong

16B64 𖭟 PAHAWH HMONG CLAN SIGN HAWJ
= Her

16B65 𖭠 PAHAWH HMONG CLAN SIGN MUAS
= Moua

16B66 𖭡 PAHAWH HMONG CLAN SIGN THOJ
= Thao

16B67 𖭢 PAHAWH HMONG CLAN SIGN TSAB
= Chang

16B68 𖭣 PAHAWH HMONG CLAN SIGN KHAB
= Khang

16B69 𖭤 PAHAWH HMONG CLAN SIGN HAM
= Hang

16B6A 𖭥 PAHAWH HMONG CLAN SIGN VAJ
= Vang

16B6B 𖭦 PAHAWH HMONG CLAN SIGN YAJ
= Yang

16B6C 𖭧 PAHAWH HMONG CLAN SIGN KWM
= Kw

16B6D 𖭨 PAHAWH HMONG CLAN SIGN VWJ
= Vue

16B6E 𖭩 PAHAWH HMONG CLAN SIGN TSHEEJ
= Cheng

16B6F 𖭪 PAHAWH HMONG CLAN SIGN KOO
= Kong

16B70 𖭫 PAHAWH HMONG CLAN SIGN FAJ
= Fang

16B71 𖭬 PAHAWH HMONG CLAN SIGN TSWB
= Chue

Figures.

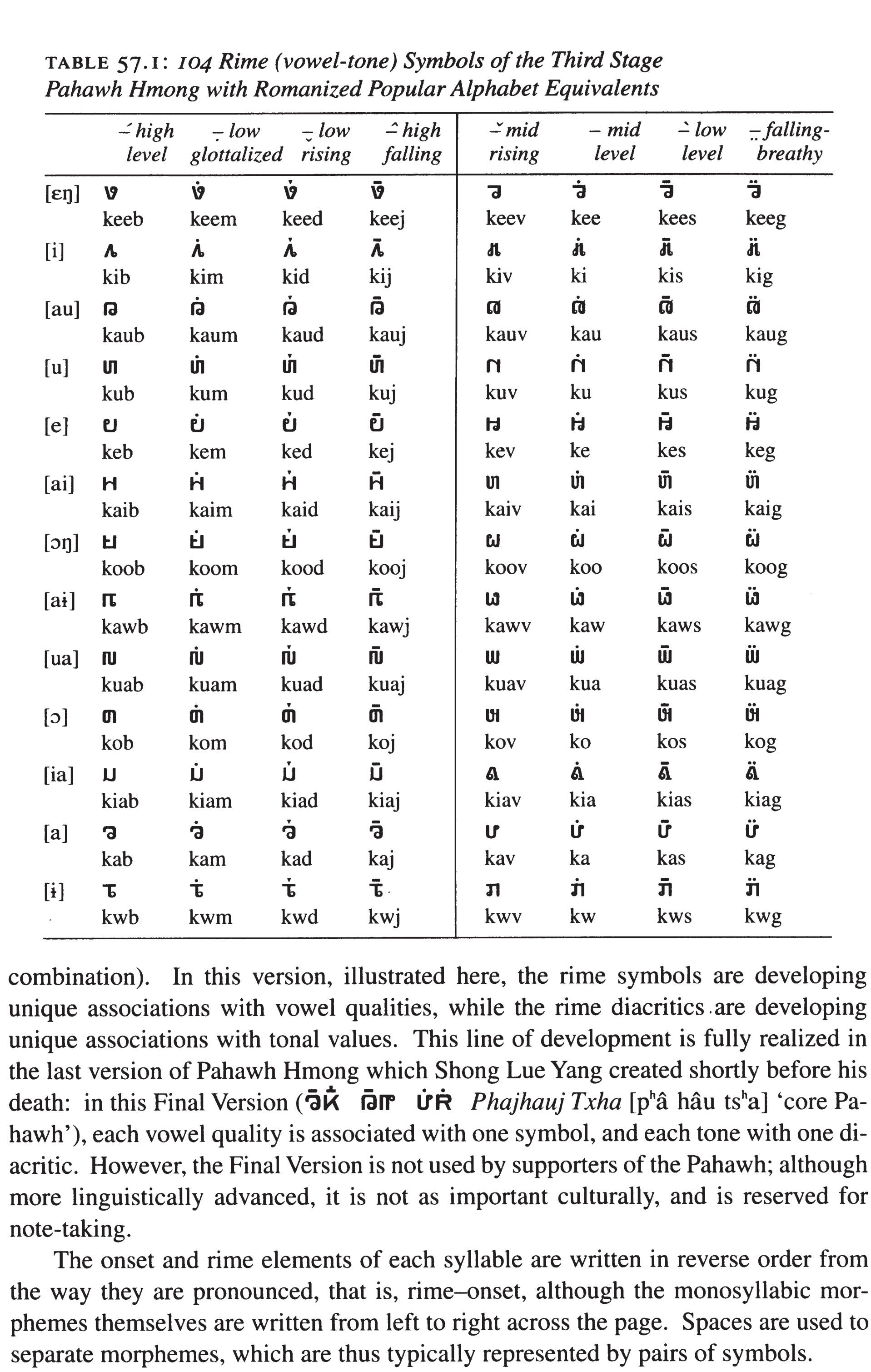

TABLE 57.1: *104 Rime (vowel-tone) Symbols of the Third Stage Pahawh Hmong with Romanized Popular Alphabet Equivalents*

	high level	*low glottalized*	*low rising*	*high falling*	*mid rising*	*mid level*	*low level*	*falling-breathy*
[ɛŋ]	keeb	keem	keed	keej	keev	kee	kees	keeg
[i]	kib	kim	kid	kij	kiv	ki	kis	kig
[au]	kaub	kaum	kaud	kauj	kauv	kau	kaus	kaug
[u]	kub	kum	kud	kuj	kuv	ku	kus	kug
[e]	keb	kem	ked	kej	kev	ke	kes	keg
[ai]	kaib	kaim	kaid	kaij	kaiv	kai	kais	kaig
[ɔŋ]	koob	koom	kood	kooj	koov	koo	koos	koog
[aɨ]	kawb	kawm	kawd	kawj	kawv	kaw	kaws	kawg
[ua]	kuab	kuam	kuad	kuaj	kuav	kua	kuas	kuag
[ɔ]	kob	kom	kod	koj	kov	ko	kos	kog
[ia]	kiab	kiam	kiad	kiaj	kiav	kia	kias	kiag
[a]	kab	kam	kad	kaj	kav	ka	kas	kag
[ɨ]	kwb	kwm	kwd	kwj	kwv	kw	kws	kwg

combination). In this version, illustrated here, the rime symbols are developing unique associations with vowel qualities, while the rime diacritics are developing unique associations with tonal values. This line of development is fully realized in the last version of Pahawh Hmong which Shong Lue Yang created shortly before his death: in this Final Version (*Phajhauj Txha* [p^{h}â hâu ts^{h}a] 'core Pahawh'), each vowel quality is associated with one symbol, and each tone with one diacritic. However, the Final Version is not used by supporters of the Pahawh; although more linguistically advanced, it is not as important culturally, and is reserved for note-taking.

The onset and rime elements of each syllable are written in reverse order from the way they are pronounced, that is, rime–onset, although the monosyllabic morphemes themselves are written from left to right across the page. Spaces are used to separate morphemes, which are thus typically represented by pairs of symbols.

Figure 1a. Chart of Third Stage Revised Version vowel rimes from Ratliff 1996.

TABLE 57.2: *Sixty Onset (consonant) Symbols of the Third Stage Pahawh Hmong with Romanized Popular Alphabet Equivalents*

vau	[v-]	nrau	[ɳɻ-]	fau	[f-]
nkau	[ŋk-]	ntxau	[nts-]	rhau	[ɻʰ-]
xau	[s-]	au	[ʔ-]	nyau	[ɲ-]
cau	[c-]	ntshau	[ntʃʰ-]	txau	[ts-]
lau	[l-]	dau	[ʔd-]	dhau	[ʔdʰ-]
ntsau	[ntʃ-]	tsau	[tʃ-]	phau	[pʰ-]
hlau	[ʰl-]	zau	[ʒ-]	ntxhau	[ntsʰ-]
rau	[ɻ-]	nphau	[mpʰ-]	nphlau	[mpʰl-]
hnau	[ʰn-]	khau	[kʰ-]	ntau	[nt-]
plhau	[pʰl-]	tshau	[tʃʰ-]	pau	[p-]
nthau	[ntʰ-]	nplau	[mpl-]	nkhau	[ŋkʰ-]
chau	[cʰ-]	xyau	[ç-]	tau	[t-]
nau	[n-]	nqau	[ɴq-]	nqhau	[ɴqʰ-]
nlau	[ml-]	hnlau	[ʰml-]	gau	[ŋ-]
qhau	[qʰ-]	nyhau	[ʰɲ-]	hmau	[ʰm-]
hau	[h-]	thau	[tʰ-]	plau	[pl-]
nchau	[ɲcʰ-]	nrhau	[ɳɻʰ-]	npau	[mp-]
mau	[m-]	txhau	[tsʰ-]	qau	[q-]
yau	[j-]	ncau	[ɲc-]	sau	[ʃ-]
'au	Ø	ndlau	[ndl-]	ndlhau	[ndʰl-]

Figure 1b. Chart of consonant onsets from Ratliff 1996.

Figure 2. Chart of Second Stage Revised Version vowel rimes and consonant onsets from Lee Nao Long *et al*. 2001.

Figure 3. Chart of Second Stage Revised Version vowel rimes and consonant onsets from http://www.pahauhhmong.org/. Notice the glyph shapes of the diacritics in both the monoline font at the top and the more decorative font below. Compare these with the glyphs in the code chart. The shape of U+16B30 PAHAWH HMONG MARK CIM TUB (literally 'son mark') and U+16B36 PAHAWH HMONG MARK CIM TAUM ('beans mark') have hollow forms here, but are filled in other font styles. Similarly, the horizontal bar in U+16B32 PAHAWH HMONG MARK CIM KES ('line mark') and U+16B35 PAHAWH HMONG MARK CIM HOM ('kind mark') has a distinctive shape in the decorative font.

The Pahawh Hmong diacritics were devised by Shong Lue Yang in isolation, and have no genetic relation to similar-looking punctuation in the European tradition (DOT ABOVE, DIAERESIS, MACRON). Since it can also typically take shapes which are different from the typical shapes that European punctuation has, it would be inappropriate to attempt to unify Pahawh Hmong diacritics with characters in the General Punctuation mark. (Similar arguments were given for N'Ko diacritics.)

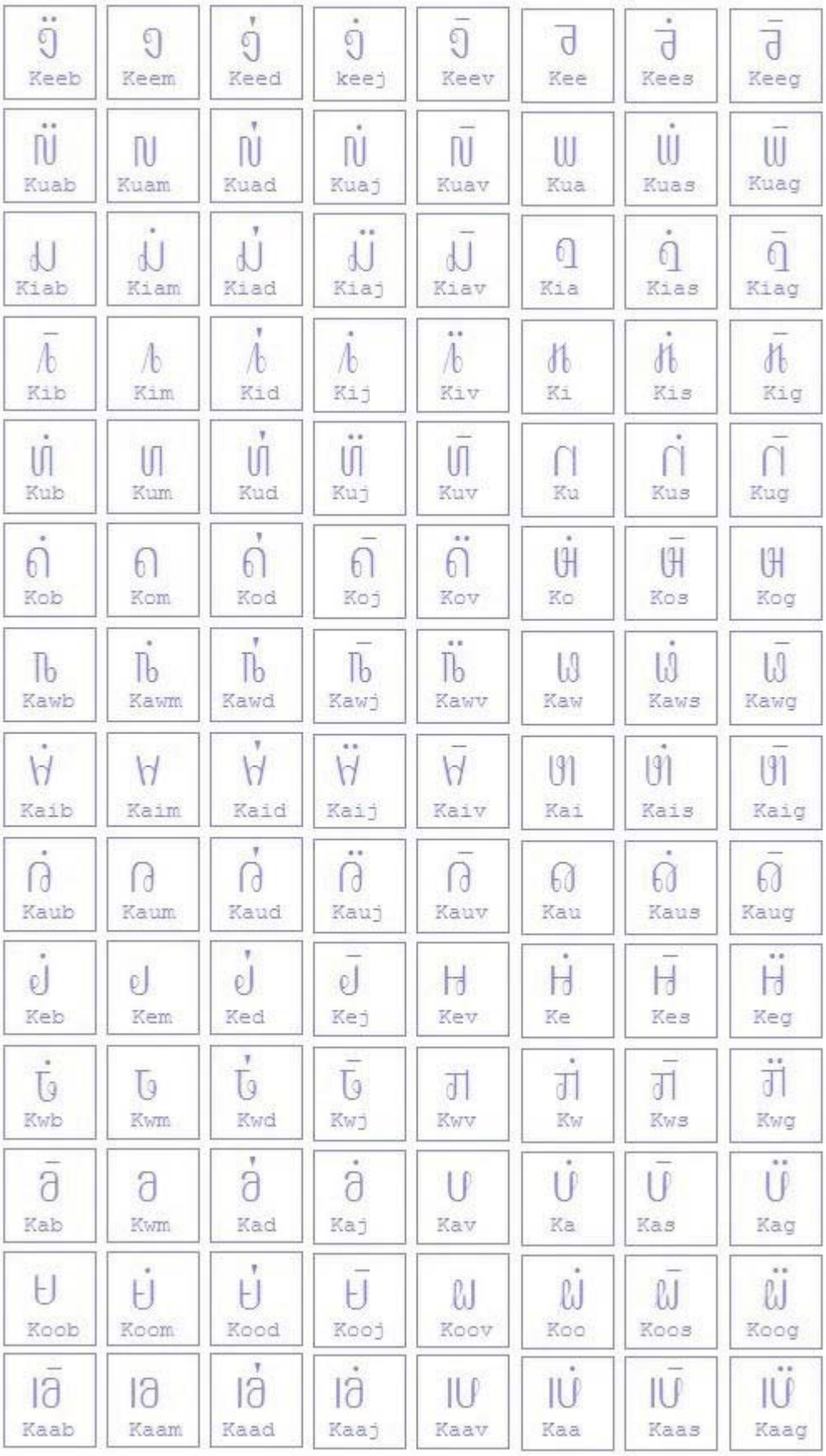

Figure 4a. Chart of Second Stage Revised Version vowel rimes from http://www.cwjmemhmong.info/.

Figure 4b. Chart of Second Stage Revised Version consonant onsets from http://www.cwjmemhmong.info/.

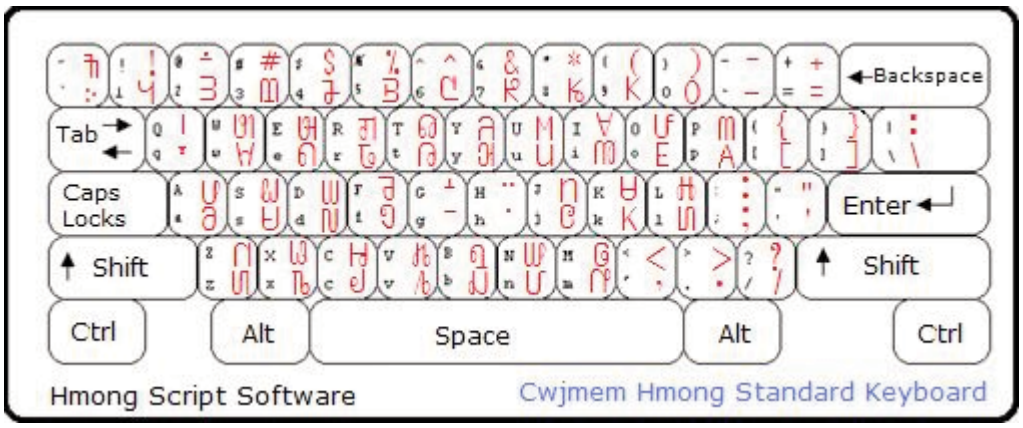

Figure 5a. Keyboard layout from **hmongscript.cwjmemhmong.info**.

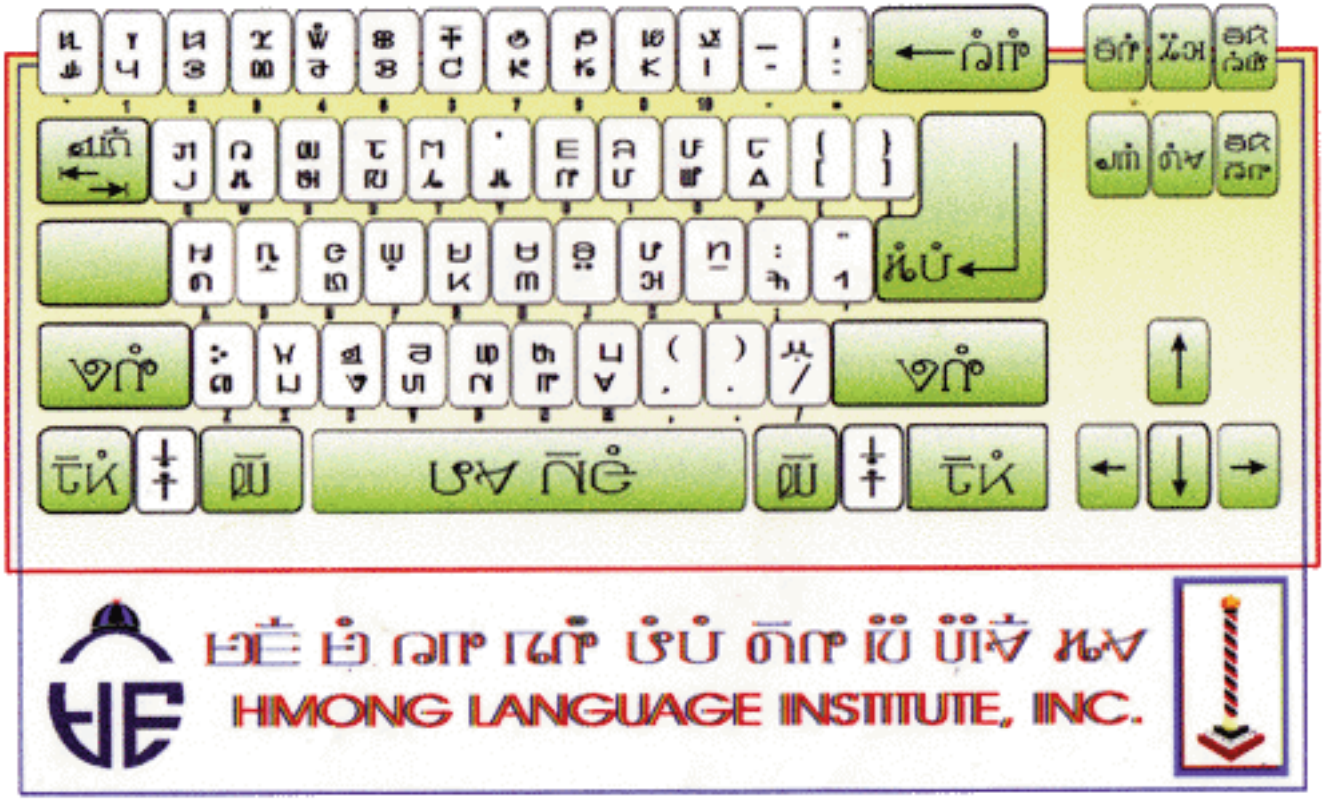

Figure 5b. Keyboard layout from **www.pahauhhmong.org**.

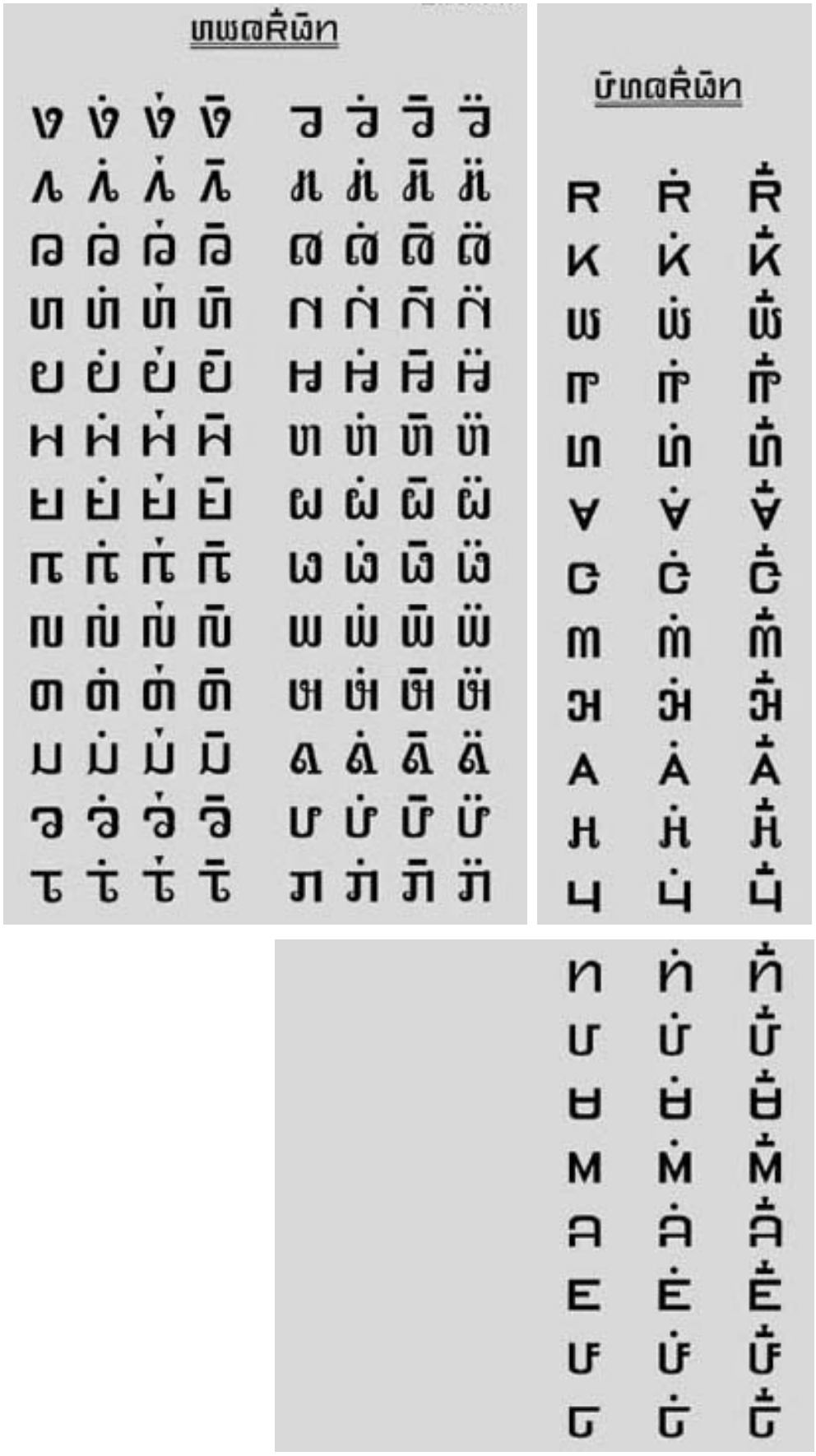

Figure 6. On the left, vowel rimes, on the right, consonant onsets, in the Third Stage Reduced Version from Chia Koua Vang's 2002 *Hmong Language: Second Grade*.

Figure 7. A page of text in the Third Stage Reduced Version from Chia Koua Vang's 2002 *Hmong Language: Second Grade*.

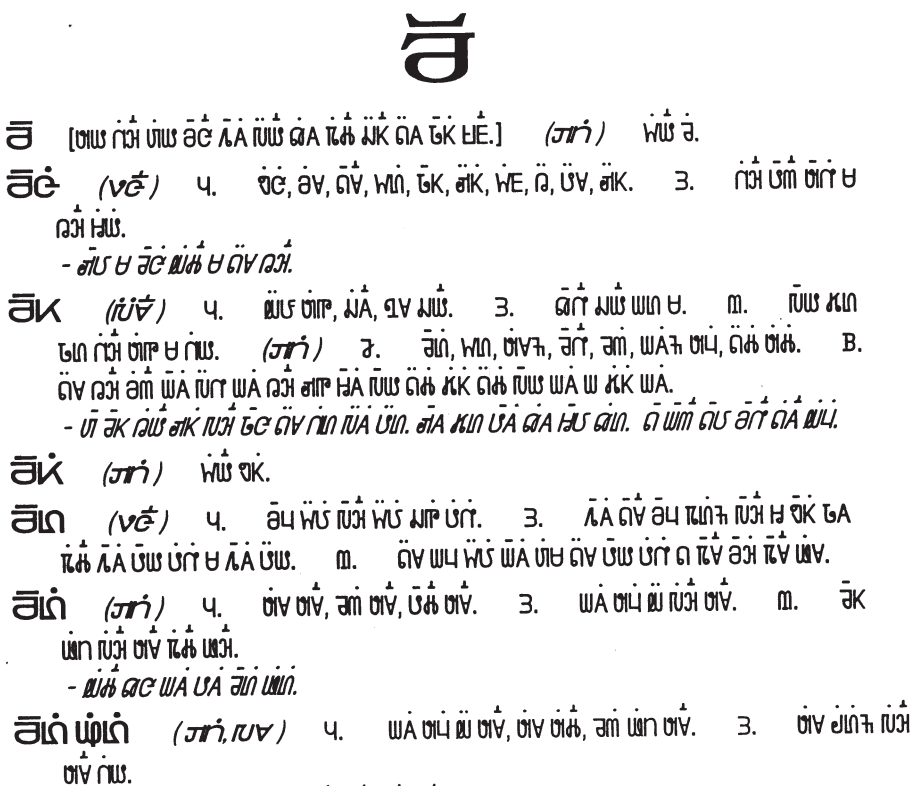

Figure 8. A page of text from a Second Stage Reduced Version dictionary (no bibliographic details).

A. Administrative

1. Title

Preliminary proposal for encoding the Pahawh Hmong script in the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project)

3. Requester type (Member body/Liaison/Individual contribution)

Liaison contribution.

4. Submission date

2009-09-14

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Pahawh Hmong.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

106.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson and Jason Glavy

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

TBD

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Historical and contemporary cultural use by Hmongs and historians of Hmong culture.
4a. The context of use for the proposed characters (type of use; common or rare)
Common.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
In Australia and in the US.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
No.
6b. If YES, is a rationale provided?
6c. If YES, reference
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
Yes.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?