

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

Doc Type: Working Group Document

Title: Preliminary proposal to encode the Pahawh Hmong script in the UCS

Source: Michael Everson

Status: Individual Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2009-04-16

1.0 Introduction. Pahawh Hmong is a script devised for writing the Hmong language by Shong Lue Yang (*Soob Lwj Yaj* ສອບ ລວງ ຢາຈ [ʃóŋ lɰ̃ 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. The writing system itself had four Stages of development. In this document, the Romanized Popular Alphabet orthography is given alongside example text in Pahawh Hmong. Two features of the RPA are of note. Double vowels *ee* and *oo* indicate [eŋ] and [oŋ] respectively; final letters indicate tones thus: *-b* ɿ, *-m* ɿ, *-d* ɿ, *-j* ɿ, *-v* ɿ, *-Ø* ɿ, *-s* ɿ, and *-g* ɿ indicate the tones respectively. Another way of describing the tones is *ú* high-level, *ɿ* low-glottalized, *ɿ* low-rising, *ɿ* high-falling, *ɿ* mid-rising, *ɿ* mid-level, *ɿ* low-level, and *ɿ* falling-breathy.

1.1. The Source Version, Pahawh Pa (*Phajhauj Paj* ຫັ ຫັ ຫັ [p^hâ 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* ຫັ ຫັ ຫັ ຫັ ຫັ [p^hâ hâu ntʃíá dùa ʔó]), 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 ຫັ ຫັ ຫັ *Cwjmem* [cê mɛ] font; both communities have web sites today with downloadable fonts (see References below).

1.3. The Third Stage Reduced Version, Pahawh Njia Dua Pe (*Phajhauj Ntsiab Duas Peb* ຫັ ຫັ ຫັ ຫັ ຫັ ຫັ [p^hâ hâu ntʃíá 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.

1.4. The Final Version, Pahawh Tsa (*Phajhauj Txha* ຫັ ຫັ ຫັ ຫັ ຫັ [p^hâ hâu ts^ha]), is not in regular use. It is a radical simplification of the Third Stage Reduced Version introduced in 1971-01 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.

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 coda of a syllable (its vowel with or without tone diacritic) precedes the head of the syllable (its consonant with or without consonant-identifier diacritic).

In the Figure 1, the structure of the words “Pahawh Hmong” (*Phajhauj Hmoob* [p^hâ hâu ^hmóŋ]) is analyzed, given in Second and Third Stage Reduced Version (Final Version is identical to Third Stage Reduced Version in this example).

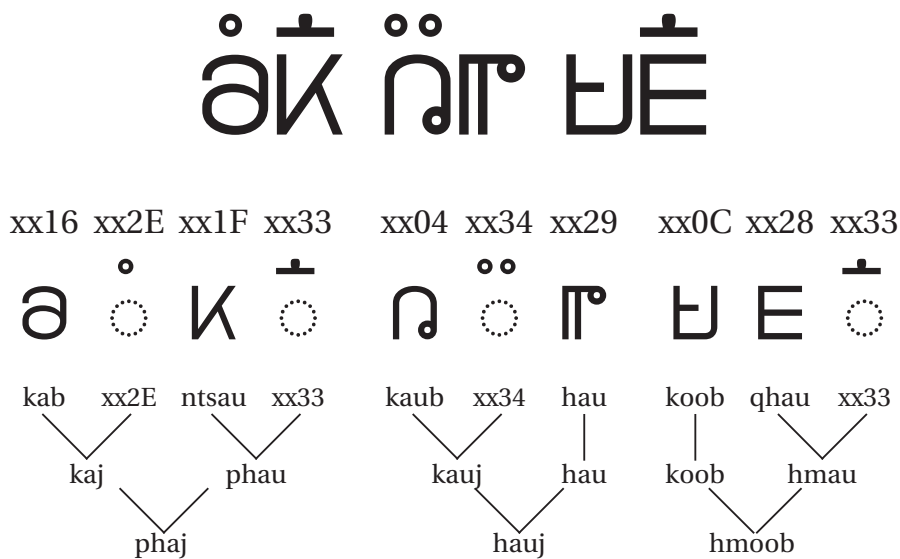


Figure 1. Second Stage Reduced Version

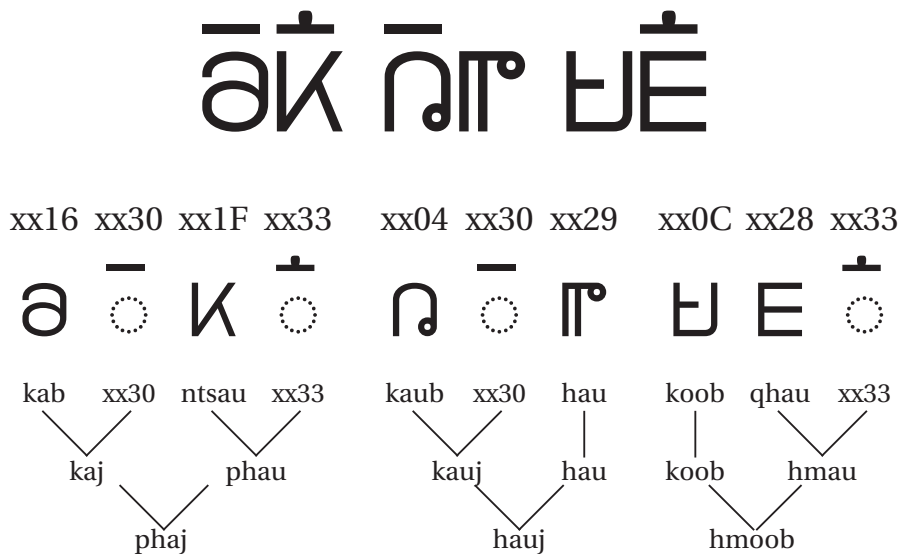


Figure 2. Third Stage Reduced Version and Final Version

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 ɭ, KIAV ɬ, KAB ǎ, and KAV ɰ 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* ɭ̄ and *kaav* ɰ̄ (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 𐌆 would be problematic as 16B4A PAHAWH HMONG NUMBER TENS looks just like it.

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.

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.

Combining diacritics are found at 16B30–16B36 and function in the usual way.

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

16B38 𐌆 was invented by Pa Kao Her (*Paj Kaub Hawj* 𐌆𐌇 𐌈 𐌉𐌊 [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 “!”.

16B39 𐌇 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.

16B3A 𐌈 indicates reduplication of the syllable preceding: 𐌈𐌉𐌊 = 𐌉𐌊 𐌉𐌊.

16B3B 𐌉 is derived from the ampersand. This character was based on the & 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.

16B3C 𐌊 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.

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*.

16B51 𐌋 represents the syllable *lub* 𐌋𐌌 [lú], the most common grammatical classifier in the Hmong language. Smalley *et al.* 1990 give the example 𐌋𐌌𐌍 *lub npe* [lú mpe] ‘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 seems that in devising the character Shong Lue Yang was being very practical indeed.

16B52–16B56 are logographs naming periods of time: year 𐌌, month 𐌍, date 𐌎, day 𐌏, season 𐌐 respectively.

A. Administrative

1. Title

Preliminary proposal for encoding the Pahawh Hmong script in the UCS

2. Requester's name

Michael Everson.

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

Individual contribution.

4. Submission date

2009-04-16

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

53.

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.

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?

	16B0	16B1	16B2	16B3	16B4	16B5	16B6	16B7
0	𐄀 16B00	𐄁 16B10	𐄂 16B20	𐄃 16B30	𐄄 16B40	𐄅 16B50	𐄆 16B60	𐄇 16B70
1	𐄈 16B01	𐄉 16B11	𐄊 16B21	𐄋 16B31	𐄌 16B41	𐄍 16B51	𐄎 16B61	𐄏 16B71
2	𐄐 16B02	𐄑 16B12	𐄒 16B22	𐄓 16B32	𐄔 16B42	𐄕 16B52	𐄖 16B62	
3	𐄗 16B03	𐄘 16B13	𐄙 16B23	𐄚 16B33	𐄛 16B43	𐄜 16B53	𐄝 16B63	
4	𐄞 16B04	𐄟 16B14	𐄠 16B24	𐄡 16B34	𐄢 16B44	𐄣 16B54	𐄤 16B64	
5	𐄥 16B05	𐄦 16B15	𐄧 16B25	𐄨 16B35	𐄩 16B45	𐄪 16B55	𐄫 16B65	
6	𐄬 16B06	𐄭 16B16	𐄮 16B26	𐄯 16B36	𐄰 16B46	𐄱 16B56	𐄲 16B66	
7	𐄳 16B07	𐄴 16B17	𐄵 16B27	𐄶 16B37	𐄷 16B47	𐄸 16B57	𐄹 16B67	
8	𐄺 16B08	𐄻 16B18	𐄼 16B28	𐄽 16B38	𐄾 16B48	𐄿 16B58	𐅀 16B68	
9	𐅁 16B09	𐅂 16B19	𐅃 16B29	𐅄 16B39	𐅅 16B49	𐅆 16B59	𐅇 16B69	
A	𐅈 16B0A	𐅉 16B1A	𐅊 16B2A	𐅋 16B3A	𐅌 16B4A	𐅍 16B5A	𐅎 16B6A	
B	𐅏 16B0B	𐅐 16B1B	𐅑 16B2B	𐅒 16B3B	𐅓 16B4B	𐅔 16B5B	𐅕 16B6B	
C	𐅖 16B0C	𐅗 16B1C	𐅘 16B2C	𐅙 16B3C	𐅚 16B4C	𐅛 16B5C	𐅜 16B6C	
D	𐅝 16B0D	𐅞 16B1D	𐅟 16B2D		𐅠 16B4D	𐅡 16B5D	𐅢 16B6D	
E	𐅣 16B0E	𐅤 16B1E	𐅥 16B2E		𐅦 16B4E	𐅧 16B5E	𐅨 16B6E	
F	𐅩 16B0F	𐅪 16B1F	𐅫 16B2F		𐅬 16B4F	𐅭 16B5F	𐅮 16B6F	

Vowel rimes

16B00	◌	PAHAWH HMONG VOWEL KEEB
16B01	◌	PAHAWH HMONG VOWEL KEEV
16B02	◌	PAHAWH HMONG VOWEL KIB
16B03	◌	PAHAWH HMONG VOWEL KIV
16B04	◌	PAHAWH HMONG VOWEL KAUB
16B05	◌	PAHAWH HMONG VOWEL KAUV
16B06	◌	PAHAWH HMONG VOWEL KUB
16B07	◌	PAHAWH HMONG VOWEL KUV
16B08	◌	PAHAWH HMONG VOWEL KEB
16B09	◌	PAHAWH HMONG VOWEL KEV
16B0A	◌	PAHAWH HMONG VOWEL KAIB
16B0B	◌	PAHAWH HMONG VOWEL KAIV
16B0C	◌	PAHAWH HMONG VOWEL KOOB
16B0D	◌	PAHAWH HMONG VOWEL KOOV
16B0E	◌	PAHAWH HMONG VOWEL KAWB
16B0F	◌	PAHAWH HMONG VOWEL KAWV
16B10	◌	PAHAWH HMONG VOWEL KUAB
16B11	◌	PAHAWH HMONG VOWEL KUAU
16B12	◌	PAHAWH HMONG VOWEL KOB
16B13	◌	PAHAWH HMONG VOWEL KOV
16B14	◌	PAHAWH HMONG VOWEL KIAB
16B15	◌	PAHAWH HMONG VOWEL KIAU
16B16	◌	PAHAWH HMONG VOWEL KAB
16B17	◌	PAHAWH HMONG VOWEL KAV
16B18	◌	PAHAWH HMONG VOWEL KWB
16B19	◌	PAHAWH HMONG VOWEL KWV
16B1A	◌	PAHAWH HMONG VOWEL KAAB
16B1B	◌	PAHAWH HMONG VOWEL KAAU

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 NTHAU
16B2F	◌	PAHAWH HMONG CONSONANT AU

Tone marks

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

16B36	◌	PAHAWH HMONG TONE MARK CIM TAUM
-------	---	---------------------------------

Punctuation

16B37	◌	PAHAWH HMONG SIGN 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	◌	PAHAWH HMONG NUMBER BILLIONS = trillions

Logographs

16B51	◌	PAHAWH HMONG SIGN VOS LUB = classifier
16B52	◌	PAHAWH HMONG SIGN XYOO = year
16B53	◌	PAHAWH HMONG SIGN HLI = month
16B54	◌	PAHAWH HMONG SIGN ZWJ THAJ = date
16B55	◌	PAHAWH HMONG SIGN HNUB = day
16B56	◌	PAHAWH HMONG SIGN NTUJ = season

Arithmetical symbols

16B57	◌	PAHAWH HMONG SIGN XYEEM NTXIV = plus sign
16B58	◌	PAHAWH HMONG SIGN XYEEM RHO = minus sign

- 16B59 𐄀 PAHAWH HMONG SIGN XYEEM TOV
= multiplication sign
- 16B5A 𐄁 PAHAWH HMONG SIGN XYEEM FAIB
= division sign

Uncertain signs

- 16B5B 𐄂 PAHAWH HMONG SIGN-1
- 16B5C 𐄃 PAHAWH HMONG SIGN-2
- 16B5D 𐄄 PAHAWH HMONG SIGN-3
- 16B5E 𐄅 PAHAWH HMONG SIGN-4
- 16B5F 𐄆 PAHAWH HMONG SIGN-5

Logographs for clan names

- 16B60 𐄇 PAHAWH HMONG CLAN SIGN YEEG
- 16B61 𐄈 PAHAWH HMONG CLAN SIGN LIS
- 16B62 𐄉 PAHAWH HMONG CLAN SIGN LAUJ
- 16B63 𐄊 PAHAWH HMONG CLAN SIGN XYOOJ
- 16B64 𐄋 PAHAWH HMONG CLAN SIGN HAWJ
- 16B65 𐄌 PAHAWH HMONG CLAN SIGN MUAS
- 16B66 𐄍 PAHAWH HMONG CLAN SIGN THOJ
- 16B67 𐄎 PAHAWH HMONG CLAN SIGN TSAB
- 16B68 𐄏 PAHAWH HMONG CLAN SIGN KHAB
- 16B69 𐄐 PAHAWH HMONG CLAN SIGN HAM
- 16B6A 𐄑 PAHAWH HMONG CLAN SIGN VAJ
- 16B6B 𐄒 PAHAWH HMONG CLAN SIGN YAJ
- 16B6C 𐄓 PAHAWH HMONG CLAN SIGN KWM
- 16B6D 𐄔 PAHAWH HMONG CLAN SIGN VWJ
- 16B6E 𐄕 PAHAWH HMONG CLAN SIGN TSHEEJ
- 16B6F 𐄖 PAHAWH HMONG CLAN SIGN KOO
- 16B70 𐄗 PAHAWH HMONG CLAN SIGN FAJ
- 16B71 𐄘 PAHAWH HMONG CLAN SIGN TSWB