ISO/IEC JTC1/SC2/WG2 N3616R
L2/09-145R
2009-05-04

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: **2009-05-04**

1.0 Introduction. Pahawh Hmong is a script devised for writing the Hmong language by Shong Lue Yang (*Soob Lwj Yaj* [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 campks, 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: *-b* ꜒, *-m* ꜖, *-d* ꜖꜔, *-j* ꜒꜔, *-v* ꜔꜒, *-Ø* ꜔, *-s* ꜕, and *-g* ꜒꜕ indicate the tones respectively. Another way of describing the tones is *v́* high-level, *ṿ* low-glottalized, *v̬* low-rising, *v̂* high-falling, *v̌* mid-rising, *v* mid-level, *v̀* low-level, and *v̤* falling-breathy.

1.1 The Source Version, Pahawh Pa (*Phajhauj Paj* [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* [pʰâ hâu ntʃía 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 *Cwjmem* [cɨ̂ mḛ] font; both communities have web sites today with downloadable fonts (see References below). 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* [pʰâ hâu ntʃía 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. The Hmong user community in Minnesota uses the Third Stage Reduced Version.

1.4 The Final Version, Pahawh Tsa (*Phajhauj Txha* [pʰâ hâu tsʰa]), 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.

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* [pʰâ hâu ʰmɔ́ŋ]) is analyzed, given in Second and Third Stage Reduced Version (Final Version is identical to Third Stage Reduced Version in this example).

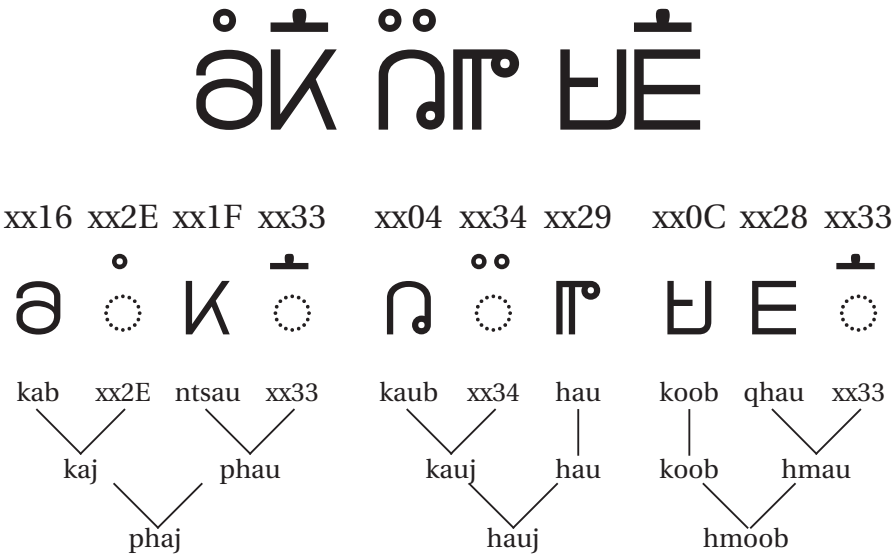


Figure 1. Second Stage Reduced Version

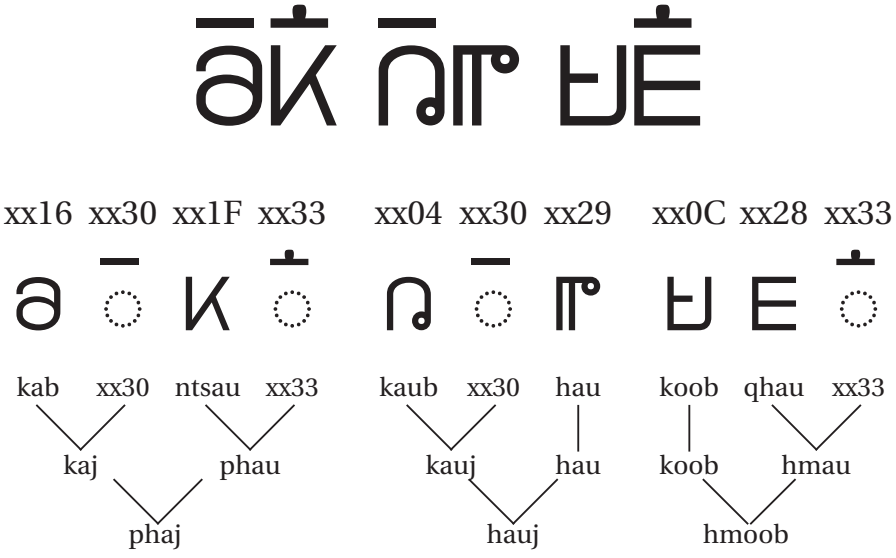


Figure 2. Third Stage Reduced Version and Final Version

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* I𖬖 and *kaav* I𖬗 (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.

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.

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

2.5.1 Exclamation mark 16B38 𖬸 was invented by Pa Kao Her (*Paj Kaub Hawj* 𖬖𖬰𖬝 𖬄 𖬎𖬰𖬧 [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 "!".

2.5.2 Exclamation 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.

2.5.3 Reduplication mark 16B3A 𖬺 indicates reduplication of the syllable preceding: 𖬌𖬲𖬤𖬺 = 𖬌𖬲𖬤 𖬌𖬲𖬤.

2.5.4 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.

2.5.5 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.

2.6 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*.

2.7 Grammatical classifier. 16B51 represents the syllable *lub* [lú], the most common grammatical classifier in the Hmong language. Smalley *et al*. 1990 give the example *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 charaters used to write it it seem that in devising the character Shong Lue Yang was being very practical indeed.

2.8 Logographs. 16B52–16B56 are logographs naming periods of time: year, month, date, day, season respectively.

2.9 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.

2.11 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.

3 Unidentified marks. 16B5B–16B5F are found in the Naadaa font and keyboard and have not yet been identified.

4 References

Cwjmem font. http://www.geocities.com/SiliconValley/Pines/5884

Naadaa font. http://www.linguistics.unimelb.edu.au/research/hmong

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

5. 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.

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

2009-05-04

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.

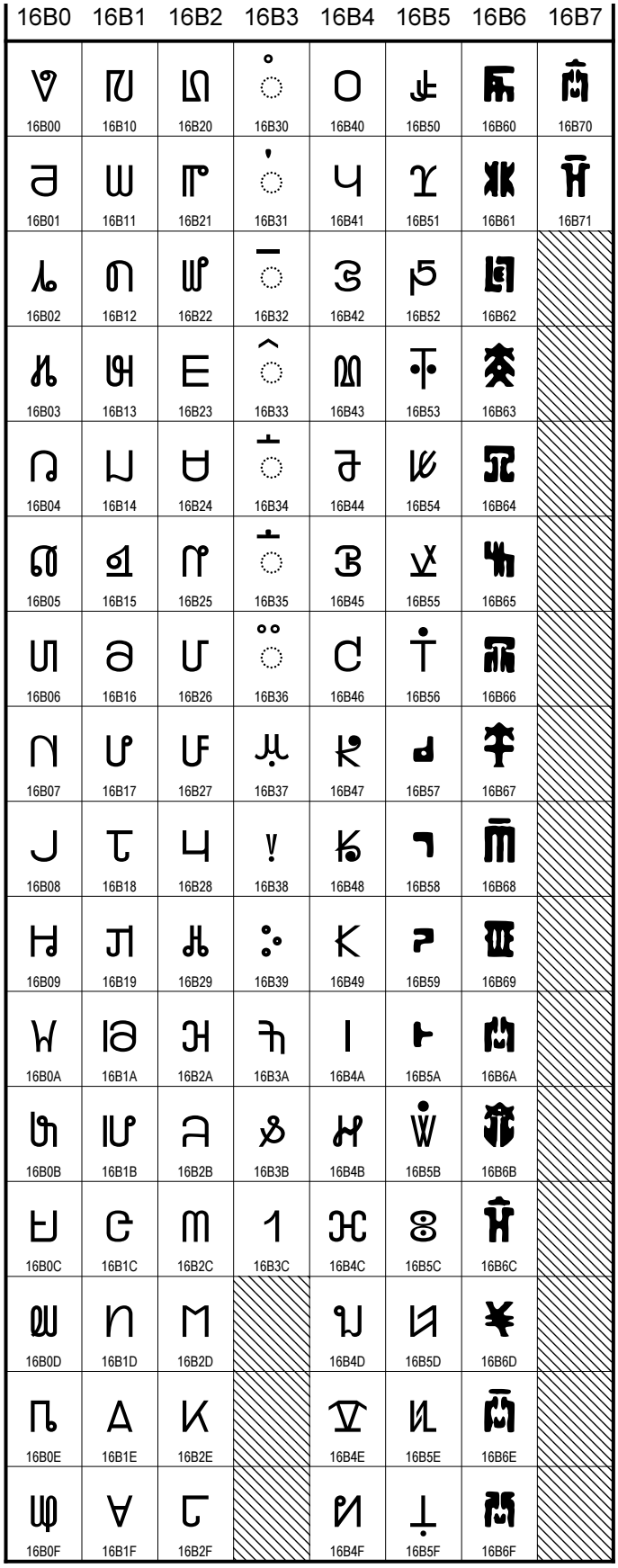
2b. If YES, with whom?

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	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	16B5B	16B6B	
C	16B0C	16B1C	16B2C	16B3C	16B4C	16B5C	16B6C	
D	16B0D	16B1D	16B2D		16B4D	16B5D	16B6D	
E	16B0E	16B1E	16B2E		16B4E	16B5E	16B6E	
F	16B0F	16B1F	16B2F		16B4F	16B5F	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 KUAV
16B12 PAHAWH HMONG VOWEL KOB
16B13 PAHAWH HMONG VOWEL KOV
16B14 PAHAWH HMONG VOWEL KIAB
16B15 PAHAWH HMONG VOWEL KIAV
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 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 vowel and consonant modifiers

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 PAHAWH HMONG CLAN SIGN YEEG
16B61 PAHAWH HMONG CLAN SIGN LIS
16B62 PAHAWH HMONG CLAN SIGN LAUJ
16B63 PAHAWH HMONG CLAN SIGN XYOOJ
16B64 PAHAWH HMONG CLAN SIGN HAWJ
16B65 PAHAWH HMONG CLAN SIGN MUAS
16B66 PAHAWH HMONG CLAN SIGN THOJ
16B67 PAHAWH HMONG CLAN SIGN TSAB
16B68 PAHAWH HMONG CLAN SIGN KHAB
16B69 PAHAWH HMONG CLAN SIGN HAM
16B6A PAHAWH HMONG CLAN SIGN VAJ
16B6B PAHAWH HMONG CLAN SIGN YAJ
16B6C PAHAWH HMONG CLAN SIGN KWM
16B6D PAHAWH HMONG CLAN SIGN VWJ
16B6E PAHAWH HMONG CLAN SIGN TSHEEJ
16B6F PAHAWH HMONG CLAN SIGN KOO
16B70 PAHAWH HMONG CLAN SIGN FAJ
16B71 PAHAWH HMONG CLAN SIGN TSWB