

Title: Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2011-04-27

1 Introduction

This is a proposal to encode the Pau Cin Hau Alphabet in the Universal Character Set. The information presented here is based upon and supersedes that given in the following documents:

- N3781 L2/10-080 “Preliminary Proposal to Encode the Pau Cin Hau Script in ISO/IEC 10646”
- N3784R L2/10-092R “Defining Properties for Tone Marks of the Pau Cin Hau Script”
- N3865R L2/10-073R1 “Allocating the Pau Cin Hau Scripts in the Unicode Roadmap”
- N3960 L2/10-437 “Preliminary Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646”

Several changes have been made since N3960, which include: revision of character names; clarification of tone mark values; addition of collation and other implementation details; and the inclusion of new specimens.

There is a logographic script related to the Pau Cin Hau Alphabet. More information about it is available in “Introducing the Logographic Script of Pau Cin Hau” (N3961 L2/10-438).

2 Background

The Pau Cin Hau Alphabet is a liturgical script of the Laipian religious tradition, which emerged in the Chin Hills region of present-day Myanmar at the turn of the 20th century. The script is named after Pau Cin Hau (1859–1948), a Tedim Chin, who founded the Laipian tradition and developed the script in order to convey his teachings. In an account given by J. J. Bennison in the 1931 *Census of India* report for Burma, Pau Cin Hau stated that the characters of his script were revealed to him in a dream in 1902.¹

The script was designed to represent Tedim (ISO-639: ctd), a language of the northern branch of the Kuki-Chin group of the Tibeto-Burman family, which is spoken in Chin State, Myanmar. It is known traditionally as *pau cin hau lai* “script of Pau Cin Hau” and *tual lai* “local script”. The Tedim word *lai* means “writing” and *laipian* is the ‘script-based religion’; Pau Cin Hau himself is referred to as *laipianpa* “creator of laipian”.² The name of the script is also romanized as ‘Pau Chin Hau’ and is spelled in one source as ‘Bow-chinhow’.³ Tedim is the modern name for the language previously known as ‘Kamhow’ and ‘Sokte’.

There are two distinct writing systems associated with Pau Cin Hau and the Laipian tradition. One is a logographic script and the other is the alphabetic system discussed here. Both are attested in written and printed sources. The original script devised by Pau Cin Hau in 1902 was reformed at least twice. The logographic script represents the first reform. It consists of 1,050 characters, a number that is based upon a

¹ Bennison 1933: 217.

² Pau: 11.

³ American Bible Society 1938: 82.

traditional count of the characters used in a Laipian recitation text (see Figure 10). The final reform took place in 1931, when the logographic script was simplified.⁴ The result was the 57-character alphabet, proposed here for encoding. Apart from the derivation of the alphabet from the logographic script, these two scripts do not have any genetic relationship with any other writing system. Some characters of the alphabet resemble those found in Burmese and Latin, which may be explained by the influence of major scripts used in Chin State, as well as the presence of Western missionaries.

Pau Cin Hau’s alphabet was used for writing and publishing Laipian and Christian literature. Pau Cin Hau developed a script primer that was published in 1932 in Rangoon by Dorothy Sherrat and the Baptist Foreign Bible Society (BFBS) under the title *Kamhow-Sokte Spelling Book* (see figures 2 and 3). In 1931, the BFBS printed the “Sermon on the Mount” from the Christian text of St. Matthew in Tedim (‘Kamhow’) using the alphabet (see figure 6). These documents are hand-written. It is reported that books of Laipian ritual songs were printed in the alphabet,⁵ however, there is no documented evidence for this claim. The digitized font used in this proposal is the first typeface developed for the Pau Cin Hau Alphabet.

The use of the Pau Cin Hau Alphabet diminished with the decline of the Laipian tradition and the rise of Christian missionary activity. Indeed, very little information is available about the script. For instance, the record for BFBS’s “Sermon on the Mount” in the catalogue of Cambridge University Library offers a note that text might be in “Pollard syllabic script”. Baptist missionaries introduced Latin-based scripts for Chin languages and in many cases developed the first written forms for these languages. A Latin orthography for Tedim was introduced by Joseph H. Cope, an American Baptist missionary. It remains the regular script for Tedim. Although practice of Laipian and usage of its liturgical scripts have declined, both traditions still survive to an extent. It is reported that both the logographic and alphabetic systems are still in limited use by the Laipian community.⁶ Information on the size of the user community was unavailable to Bennison in 1931;⁷ the same is true at present. The script also enjoys some scholarly attention, as is evidenced by articles published on several websites by members of Chin-speaking communities.

3 Script Details

3.1 Script Name

The name of the script is ‘Pau Cin Hau Alphabet.’

3.2 Character Repertoire

The Pau Cin Hau Alphabet has 57 characters: 21 consonant letters, 7 vowel letters, 9 final-consonant letters, and 20 tone marks. A code chart and names list is provided in Figure 1. Some of these letters and tone marks represent sounds that are not attested in Tedim, but which do exist in other Chin languages. This suggests that the alphabet may have been created as universal script for writing Chin languages.

3.3 Character Names

Names for letters of the Pau Cin Hau Alphabet were first transliterated into Latin by Dorothy Sherrat in 1932 (Figure 2). It is likely that these names are the bases for those used by J. J. Bennison (Figure 8). These names are idiosyncratic and do not correspond to actual phonetic values. In this system, initial consonant letters

⁴ Pau: 10.

⁵ Banks 1967: 46.

⁶ Button 2009: 23 fn. 20.

⁷ Bennison 1933: 217.

are written with an ‘a’ in the Indic manner. This suggests an inherent-vowel feature that does not exist for these letters. The Latin letters in the names for aspirated consonant letters are transposed, eg. ‘*hpa*’ instead of ‘*pha*’. The names for final consonants are also misleading. Codas in Tedim can be only voiceless, liquid, or nasal consonants. Therefore, names such as *ab*, *ag*, *ad* contradict the basic phonological properties of the letters. Moreover, the use of Latin ‘a’ in the names of final consonants is not reflective of any actual value; it is likely a prosthetic element to assist in the pronunciation of these consonants in isolation.

A new naming system has been adopted for the names of proposed characters. Names for letters adhere to an alphabetic convention and are based upon phonetic values. Shown below are historical names for the letters found in Sherrat and Bennison, and the new names assigned to the letters:

| | Old | New | | Old | New | | Old | New |
|---|---------|-----|---|-------|-----|---|-----|----------|
| ᄁ | pa | P | ᄂ | tga | C | ᄃ | ua | UA |
| ᄄ | ka | K | ᄅ | ta | T | ᄆ | ia | IA |
| ᄇ | la | L | ᄈ | hta | TH | ᄉ | ab | FINAL P |
| ᄊ | ma | M | ᄋ | na | N | ᄌ | ag | FINAL K |
| ᄍ | da | D | ᄎ | hpa | PH | ᄏ | ad | FINAL T |
| ᄐ | ya | Y | ᄑ | ra | R | ᄒ | am | FINAL M |
| ᄓ | va, ba | V | ᄔ | fa | F | ᄕ | an | FINAL N |
| ᄖ | nga | NG | ᄗ | cha | CH | ᄘ | al | FINAL L |
| ᄙ | ha | H | ᄚ | á, à | A | ᄛ | au | FINAL W |
| ᄜ | ga | G | ᄝ | é, ài | E | ᄞ | ang | FINAL NG |
| ᄟ | hka, xa | X | ᄠ | i | I | ᄡ | ai | FINAL Y |
| ᄢ | hsa | S | ᄣ | aw | O | | | |
| ᄤ | ba | B | ᄥ | u | U | | | |

3.4 Encoding Order

The encoding order for the script is based upon the traditional arrangement given in alphabet charts.

3.5 Character Semantics

The semantic information for the letters and tone marks are based upon the work of Christopher Button (2009), as well as his analysis of the script and its relationship to Tedim phonology. The phonetic values of letters and tone marks are drawn from Button (2009). Information on the syllabic structure of Tedim is based upon Thang (2001).

4 Writing System

4.1 Structure

The Pau Cin Hau Alphabet is written from left to right. Consonants, vowels, and tone marks are written linearly as independent characters. The script was designed for Tedim, whose syllable canon may be described as (C₁)V₁(V₂)(C₂)T.

4.2 Vowels

The nucleus ($V_1(V_2)$) is represented using the following vowels and diphthongs:

| | | | | | | | | |
|---|---|-----|---|----|------|---|----|------|
| Ǻ | A | /æ/ | Ǻ | O | /ɔ/ | Ǻ | IA | /ia/ |
| Ǻ | E | /ɛ/ | Ǻ | U | /ʊ/ | | | |
| Ǻ | I | /i/ | Ǻ | UA | /ʊa/ | | | |

4.3 Consonants

The onset (C_1) is represented by one of the following consonants.

| | | | | | | | | |
|---|---|----------|---|----|------|---|----|-------|
| Ǻ | P | /p/ | Ǻ | NG | /ŋ/ | L | T | /t/ |
| Ǻ | K | /k/ | Ǻ | H | /h/ | Ǻ | TH | /tʰ/ |
| C | L | /l/ | Ǻ | G | /g/ | Ǻ | N | /n/ |
| Ǻ | M | /m/ | Ǻ | X | /x/ | Z | PH | /pʰ/ |
| B | D | /d/ | e | S | /s/ | Ǻ | R | /r/ |
| U | Y | /j/, /z/ | Ǻ | B | /b/ | h | F | /f/ |
| Ǻ | V | /v/, /w/ | Ǻ | C | /tʃ/ | Ǻ | CH | /tʃʰ/ |

The letters R, F, CH represent phonemes not found in Tedim.

4.4 Final Consonants

The coda (C_2) is represented using one of the following consonants.

| | | | | | | | | |
|---|---------|-----|---|---------|-----|---|----------|-----|
| Ǻ | FINAL P | /p/ | V | FINAL M | /m/ | P | FINAL W | /w/ |
| T | FINAL K | /k/ | Ǻ | FINAL N | /n/ | Ǻ | FINAL NG | /ŋ/ |
| Ǻ | FINAL T | /t/ | Ǻ | FINAL L | /l/ | O | FINAL Y | /j/ |

4.5 ‘Tone’ Marks

The tone (T) is represented using one of the 20 ‘tone’ marks. These marks are used for indicating vowel length, tone, and glottal stop and punctuation. Of these, 15 represent tones and 5 represent glottal stop.

| | | | |
|---|-------------------------|---|------------------------------|
| Ǻ | RISING TONE LONG | · | MID-LEVEL TONE SHORT |
| Ǻ | RISING TONE SHORT | : | GLOTTAL STOP OTHER |
| Ǻ | GLOTTAL VARIANT | ↗ | MID-LEVEL TONE LONG FINAL |
| Ǻ | RISING TONE LONG FINAL | ↘ | MID-LEVEL TONE SHORT FINAL |
| Ǻ | RISING TONE SHORT FINAL | ↘ | LOW-FALLING TONE LONG |
| Ǻ | GLOTTAL VARIANT FINAL | ↘ | LOW-FALLING TONE SHORT |
| Ǻ | SANDHI TONE LONG | Ǻ | GLOTTAL STOP |
| Ǻ | SANDHI TONE SHORT | ↘ | LOW-FALLING TONE LONG FINAL |
| Ǻ | SANDHI TONE LONG FINAL | ↘ | LOW-FALLING TONE SHORT FINAL |
| Ǻ | SANDHI TONE SHORT FINAL | Ǻ | GLOTTAL STOP FINAL |

The traditional arrangement of tone marks is:

| | BASIC | | | SENTENCE FINAL | | |
|--------------------|-------|-------|---------|----------------|-------|---------|
| | long | short | glottal | long | short | glottal |
| rising tone (II) | Ꞥ | Ꞥ̣ | Ꞥ̣: | ꞤꞤ | ꞤꞤ̣ | ꞤꞤ̣: |
| sandhi tone | Ꞥ̣ | Ꞥ̣̣ | | ꞤꞤ̣ | ꞤꞤ̣̣ | |
| mid-level tone (I) | | · | : | ꞤꞤ̣ | ꞤꞤ̣̣ | |
| low-falling (III) | Ꞥ̣̣ | Ꞥ̣̣̣ | Ꞥ̣̣̣: | ꞤꞤ̣̣ | ꞤꞤ̣̣̣ | ꞤꞤ̣̣̣: |

True tone marks are shown in black, while glottal-stop marks are in gray. True tone marks have two features: tone contour (basic or sandhi) and vowel length. The ‘sentence-final’ marks are contextual variants of the corresponding ‘basic’ marks that are used at the end of sentences or as punctuation.

The remaining 5 marks are used for writing codas ending in a glottal stop. Although orthographically paired with tone marks, they have no actual correlation with tones. As shown below, there is one basic glottal stop in Tedim, one variant form, and another that is used for representing other languages:

| | BASIC | SENTENCE FINAL |
|-----------|---------|----------------|
| | glottal | glottal |
| variant | Ꞥ̣̣̣ | ꞤꞤ̣̣̣ |
| non-Tedim | : | |
| regular | Ꞥ̣̣̣̣ | ꞤꞤ̣̣̣̣ |

The structure of ‘sentence-final’ marks shows influence from Indic scripts, particularly Burmese, in which basic punctuation marks are doubled to indicate major sections, eg. Ꞥ U+104A MYANMAR SIGN LITTLE SECTION is doubled to produce ꞤꞤ U+104B MYANMAR SIGN SECTION.

The repertoire of tone and glottal-stop marks indicates that the alphabet was intended for writing other northern Chin languages besides Tedim. However, as the contours of tone categories varies across the Chin languages, the semantic value of a tone mark is dependent upon language.

4.6 Digits

Script-specific digits are unattested. Latin digits are used.

4.7 Punctuation

Word boundaries are indicated using spaces. End of sentence is marked with final forms of tone marks. Latin punctuation is used. In some cases, sentence-final tone marks may be followed by a Latin period, etc. Other script-specific punctuation is unattested.

4.8 Collation

Letters are sorted as primary, tone marks as secondary. The sort order for letters is:

𐑀 A < 𐑁 E < 𐑂 I < 𐑃 O < 𐑄 U < 𐑅 UA < 𐑆 IA < 𐑇 P < 𐑈 K < 𐑉 L < 𐑊 M <
 𐑋 D < 𐑌 Y < 𐑍 V < 𐑎 NG < 𐑏 H < 𐑐 G < 𐑑 X < 𐑒 S < 𐑓 B < 𐑔 C < 𐑕 T <
 𐑖 TH < 𐑗 N < 𐑘 PH < 𐑙 R < 𐑚 F < 𐑛 CH < 𐑜 FINAL P < 𐑝 FINAL K <
 𐑞 FINAL T < 𐑟 FINAL M < 𐑠 FINAL N < 𐑡 FINAL L < 𐑢 FINAL W < 𐑣 FINAL NG <
 𐑤 FINAL Y

The collation order for tone marks differs from the encoding order. It is based upon the order of tone contours for Tedim.

· MID-LEVEL TONE SHORT < 𐑦 MID-LEVEL TONE SHORT FINAL < 𐑧 MID-LEVEL TONE LONG
 FINAL < 𐑨 RISING TONE SHORT < 𐑩 RISING TONE SHORT FINAL < 𐑪 RISING TONE LONG <
 𐑫 RISING TONE LONG FINAL < 𐑬 LOW-FALLING TONE SHORT < 𐑭 LOW-FALLING TONE SHORT
 FINAL < 𐑮 LOW-FALLING TONE LONG < 𐑯 LOW-FALLING TONE LONG FINAL < 𐑰 SANDHI
 TONE SHORT < 𐑱 SANDHI TONE SHORT FINAL < 𐑲 SANDHI TONE LONG < 𐑳 SANDHI TONE
 LONG FINAL < 𐑴 GLOTTAL STOP < 𐑵 GLOTTAL STOP FINAL < 𐑶 GLOTTAL VARIANT < 𐑷
 GLOTTAL VARIANT FINAL < : GLOTTAL STOP OTHER

4.9 Line Breaking

Line breaking should occur at spaces. Words are not broken at end-of-line and no hyphen is used or attested. No breaking may occur between a tone mark and the character that precedes it. All tone marks have the linebreaking property ‘AL’.

4.10 Character Properties

11840;PAU CIN HAU LETTER P;Lo;0;L;;;;N;;;;;
 11841;PAU CIN HAU LETTER K;Lo;0;L;;;;N;;;;;
 11842;PAU CIN HAU LETTER L;Lo;0;L;;;;N;;;;;
 11843;PAU CIN HAU LETTER M;Lo;0;L;;;;N;;;;;
 11844;PAU CIN HAU LETTER D;Lo;0;L;;;;N;;;;;
 11845;PAU CIN HAU LETTER Y;Lo;0;L;;;;N;;;;;
 11846;PAU CIN HAU LETTER V;Lo;0;L;;;;N;;;;;
 11847;PAU CIN HAU LETTER NG;Lo;0;L;;;;N;;;;;
 11848;PAU CIN HAU LETTER H;Lo;0;L;;;;N;;;;;
 11849;PAU CIN HAU LETTER G;Lo;0;L;;;;N;;;;;
 1184A;PAU CIN HAU LETTER X;Lo;0;L;;;;N;;;;;
 1184B;PAU CIN HAU LETTER S;Lo;0;L;;;;N;;;;;
 1184C;PAU CIN HAU LETTER B;Lo;0;L;;;;N;;;;;
 1184D;PAU CIN HAU LETTER C;Lo;0;L;;;;N;;;;;
 1184E;PAU CIN HAU LETTER T;Lo;0;L;;;;N;;;;;
 1184F;PAU CIN HAU LETTER TH;Lo;0;L;;;;N;;;;;
 11850;PAU CIN HAU LETTER N;Lo;0;L;;;;N;;;;;
 11851;PAU CIN HAU LETTER PH;Lo;0;L;;;;N;;;;;
 11852;PAU CIN HAU LETTER R;Lo;0;L;;;;N;;;;;
 11853;PAU CIN HAU LETTER F;Lo;0;L;;;;N;;;;;
 11854;PAU CIN HAU LETTER CH;Lo;0;L;;;;N;;;;;
 11855;PAU CIN HAU LETTER A;Lo;0;L;;;;N;;;;;
 11856;PAU CIN HAU LETTER E;Lo;0;L;;;;N;;;;;
 11857;PAU CIN HAU LETTER I;Lo;0;L;;;;N;;;;;
 11858;PAU CIN HAU LETTER O;Lo;0;L;;;;N;;;;;

```

11859;PAU CIN HAU LETTER U;Lo;0;L;;;;N;;;;;
1185A;PAU CIN HAU LETTER UA;Lo;0;L;;;;N;;;;;
1185B;PAU CIN HAU LETTER IA;Lo;0;L;;;;N;;;;;
1185C;PAU CIN HAU LETTER FINAL P;Lo;0;L;;;;N;;;;;
1185D;PAU CIN HAU LETTER FINAL K;Lo;0;L;;;;N;;;;;
1185E;PAU CIN HAU LETTER FINAL T;Lo;0;L;;;;N;;;;;
1185F;PAU CIN HAU LETTER FINAL M;Lo;0;L;;;;N;;;;;
11860;PAU CIN HAU LETTER FINAL N;Lo;0;L;;;;N;;;;;
11861;PAU CIN HAU LETTER FINAL L;Lo;0;L;;;;N;;;;;
11862;PAU CIN HAU LETTER FINAL W;Lo;0;L;;;;N;;;;;
11863;PAU CIN HAU LETTER FINAL NG;Lo;0;L;;;;N;;;;;
11864;PAU CIN HAU LETTER FINAL Y;Lo;0;L;;;;N;;;;;
11865;PAU CIN HAU RISING TONE LONG;Lo;0;L;;;;N;;;;;
11866;PAU CIN HAU RISING TONE SHORT;Lo;0;L;;;;N;;;;;
11867;PAU CIN HAU GLOTTAL VARIANT;Lo;0;L;;;;N;;;;;
11868;PAU CIN HAU RISING TONE LONG FINAL;Lo;0;L;;;;N;;;;;
11869;PAU CIN HAU RISING TONE SHORT FINAL;Lo;0;L;;;;N;;;;;
1186A;PAU CIN HAU GLOTTAL VARIANT FINAL;Lo;0;L;;;;N;;;;;
1186B;PAU CIN HAU SANDHI TONE LONG;Lo;0;L;;;;N;;;;;
1186C;PAU CIN HAU SANDHI TONE SHORT;Lo;0;L;;;;N;;;;;
1186D;PAU CIN HAU SANDHI TONE LONG FINAL;Lo;0;L;;;;N;;;;;
1186E;PAU CIN HAU SANDHI TONE SHORT FINAL;Lo;0;L;;;;N;;;;;
1186F;PAU CIN HAU MID-LEVEL SHORT;Lo;0;L;;;;N;;;;;
11870;PAU CIN HAU GLOTTAL STOP OTHER;Lo;0;L;;;;N;;;;;
11871;PAU CIN HAU MID-LEVEL TONE LONG;Lo;0;L;;;;N;;;;;
11872;PAU CIN HAU MID-LEVEL TONE SHORT;Lo;0;L;;;;N;;;;;
11873;PAU CIN HAU LOW-FALLING TONE LONG;Lo;0;L;;;;N;;;;;
11874;PAU CIN HAU LOW-FALLING TONE SHORT;Lo;0;L;;;;N;;;;;
11875;PAU CIN HAU GLOTTAL STOP;Lo;0;L;;;;N;;;;;
11876;PAU CIN HAU LOW-FALLING TONE LONG FINAL;Lo;0;L;;;;N;;;;;
11877;PAU CIN HAU LOW-FALLING TONE SHORT FINAL;Lo;0;L;;;;N;;;;;
11878;PAU CIN HAU GLOTTAL STOP FINAL;Lo;0;L;;;;N;;;;;

```

4.11 ‘Confusables’

| | |
|------------------------------------|------------------------------------|
| PAU CIN HAU LETTER K | ; 1005 MYANMAR LETTER CA |
| PAU CIN HAU LETTER L | ; 0043 LATIN CAPITAL LETTER C |
| PAU CIN HAU LETTER L | ; 1004 MYANMAR LETTER NGA |
| PAU CIN HAU LETTER M | ; 1013 MYANMAR LETTER DHA |
| PAU CIN HAU LETTER D | ; 0042 LATIN CAPITAL LETTER B |
| PAU CIN HAU LETTER Y | ; 0055 LATIN CAPITAL LETTER U |
| PAU CIN HAU LETTER H | ; 1001 MYANMAR LETTER KHA |
| PAU CIN HAU LETTER S | ; 1054 MYANMAR LETTER VOCALIC L |
| PAU CIN HAU LETTER B | ; 0190 LATIN LETTER OPEN E |
| PAU CIN HAU LETTER T | ; 004C LATIN CAPITAL LETTER L |
| PAU CIN HAU LETTER TH | ; 1027 MYANMAR LETTER E |
| PAU CIN HAU LETTER PH | ; 005A LATIN CAPITAL LETTER Z |
| PAU CIN HAU LETTER R | ; 0294 LATIN LETTER GLOTTAL STOP |
| PAU CIN HAU LETTER F | ; 0068 LATIN SMALL LETTER H |
| PAU CIN HAU LETTER O | ; 0041 LATIN CAPITAL LETTER A |
| PAU CIN HAU LETTER FINAL P | ; 1016 MYANMAR LETTER PHA |
| PAU CIN HAU LETTER FINAL K | ; 0054 LATIN CAPITAL LETTER T |
| PAU CIN HAU LETTER FINAL M | ; 0056 LATIN CAPITAL LETTER V |
| PAU CIN HAU LETTER FINAL Y | ; 004F LATIN CAPITAL LETTER O |
| PAU CIN HAU LETTER FINAL Y | ; 101D MYANMAR LETTER WA |
| PAU CIN HAU RISING TONE LONG | ; 104A MYANMAR SIGN LITTLE SECTION |
| PAU CIN HAU RISING TONE LONG FINAL | ; 104B MYANMAR SIGN SECTION |
| PAU CIN HAU MID-LEVEL TONE SHORT | ; 00B7 MIDDLE DOT |
| PAU CIN HAU GLOTTAL STOP OTHER | ; 003A COLON |

5 References

- American Bible Society. 1938. *The Book of a Thousand Tongues*. Edited by Eric M. North. New York and London: Harper & Brothers.
- British and Foreign Bible Society. 1931. ၵၢၼ် ၵၢၼ် ၵၢၼ် ၵၢၼ် ၵၢၼ် [“The Sermon on the Mount” (Kamhow-Sokte)]. Rangoon.
- Banks, E. Pendleton. 1967. “Pau Cin Hau: A Case of Religious Innovation among the Northern Chin.” In *American Historical Anthropology: Essays in Honor of Leslie Spier*. London and Amsterdam: Southern Illinois University Press.
- Bennison, J. J. 1933. “The Pau Cin Hau Movement in the Chin Hills.” *Census of India, 1931*, vol. XI, Burma, pt. 1, ch. XI. pp. 217–218.
- Button, Christopher T. J. 2009. “A Reconstruction of Proto Northern Chin in Old Burmese and Old Chinese Perspective”. Ph.D. dissertation. School of Oriental and African Studies, University of London. <http://victoria.linguistlist.org/~lapolla/downloads/Button-Proto-Northern-Chin.pdf>
- Cambridge University Library. “Bible. N.T. Matthew. Chin, Tiddim. Selections. 1931.” <http://ul-newton.lib.cam.ac.uk/vwebv/holdingsInfo?searchId=3581&recCount=25&recPointer=4&bibId=4376880>
- Go, Khup Za. 2008. *Zo Chronicles: A Documentary Study of History and Culture of the Kuki-Chin-Lushai Tribe*. New Delhi: Mittal Publications.
- Pandey, Anshuman. 2010a. “Preliminary Proposal to Encode the Pau Cin Hau Script in ISO/IEC 10646” (N3781 L2/10-080), February 28, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3781.pdf>
- . 2010b. “Defining Properties for Tone Marks of the Pau Cin Hau Script” (N3784R L2/10-092R), May 22, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3784.pdf>
- . 2010c. “Allocating the Pau Cin Hau Scripts in the Unicode Roadmap” (N3865R L2/10-073R1), July 26, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3865.pdf>
- . 2010d. “Preliminary Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646” (N3960 L2/10-437), October 27, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3960.pdf>
- . 2010e. “Introducing the Logographic Script of Pau Cin Hau” (N3961 L2/10-438), October 27, 2010. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3961.pdf>
- Pau Chin Hau and Tham Chin Kham. 1932. ငၢၤဝါ ငၢၤၵါ ငၢၤ [Spelling Book in Kamhow-Sokte]. Toungoo: P. R. G. Press.
- Pau, Pum Khan. “Religious conversions reconsidered: Christian missions and indigenous movement among the Zo people in colonial Chin Hills.” <http://www.scribd.com/doc/25568052/Religious-Conversion-Reconsidered>
- “Pu Pau Cin Hau Lai.” Zomi Online Library, Item #200, <http://zomilibrary.com/main/items/show/200>
- Thang, Khoi Lam. 2001. “A Phonological Reconstruction of Proto-Chin”. M.A. thesis. Payap University, Chiang Mai, Thailand. http://ic.payap.ac.th/graduate/linguistics/theses/Khoi_Lam_Thang_Thesis.pdf

6 Acknowledgments

I am indebted to Christopher Button for his willingness to share his knowledge of the northern Chin language family. My understanding of the Pau Cin Hau tone marks owes a great deal to his observations and analysis of the correspondences between the tone marks and tone system of Tedim. He patiently answered my questions and overlooked my ignorance of Chin linguistics. I also thank him for providing me with materials written in the Pau Cin Hau logographic script. The project to encode the Pau Cin Hau Alphabet in the UCS would not be possible without his generous assistance.

I would also like to thank Cambridge University Libraries for providing me with copies of the “Spelling Book in Kamhow-Sokte” by Pau Cin Hau and Tham Cin Kham, as well as the Tedim “The Sermon on the Mount” published by the British and Foreign Bible Society.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

| | 1184 | 1185 | 1186 | 1187 |
|---|------------|------------|------------|------------|
| 0 | 𐌰 11840 | 𐌱 11850 | 𐌲 11860 | 𐌳 11870 |
| 1 | 𐌴 11841 | 𐌵 11851 | 𐌶 11861 | 𐌷 11871 |
| 2 | 𐌸 11842 | 𐌹 11852 | 𐌺 11862 | 𐌻 11872 |
| 3 | 𐌼 11843 | 𐌽 11853 | 𐌾 11863 | 𐌿 11873 |
| 4 | 𐍀 11844 | 𐍁 11854 | 𐍂 11864 | 𐍃 11874 |
| 5 | 𐍄 11845 | 𐍅 11855 | 𐍆 11865 | 𐍇 11875 |
| 6 | 𐍈 11846 | 𐍉 11856 | 𐍊 11866 | 𐍋 11876 |
| 7 | 𐍌 11847 | 𐍍 11857 | 𐍎 11867 | 𐍏 11877 |
| 8 | 𐍐 11848 | 𐍑 11858 | 𐍒 11868 | 𐍓 11878 |
| 9 | 𐍔 11849 | 𐍕 11859 | 𐍖 11869 | |
| A | 𐍗 1184A | 𐍘 1185A | 𐍙 1186A | |
| B | 𐍚 1184B | 𐍛 1185B | 𐍜 1186B | |
| C | 𐍟 1184C | 𐍠 1185C | 𐍡 1186C | |
| D | 𐍤 1184D | 𐍥 1185D | 𐍦 1186D | |
| E | 𐍧 1184E | 𐍨 1185E | 𐍩 1186E | |
| F | 𐍮 1184F | 𐍯 1185F | 𐍰 1186F | |

Consonants

| | | |
|-------|---|-----------------------|
| 11840 | 𐌰 | PAU CIN HAU LETTER P |
| 11841 | 𐌱 | PAU CIN HAU LETTER K |
| 11842 | 𐌲 | PAU CIN HAU LETTER L |
| 11843 | 𐌳 | PAU CIN HAU LETTER M |
| 11844 | 𐌴 | PAU CIN HAU LETTER D |
| 11845 | 𐌵 | PAU CIN HAU LETTER Y |
| 11846 | 𐌶 | PAU CIN HAU LETTER V |
| 11847 | 𐌷 | PAU CIN HAU LETTER NG |
| 11848 | 𐌸 | PAU CIN HAU LETTER H |
| 11849 | 𐌹 | PAU CIN HAU LETTER G |
| 1184A | 𐌺 | PAU CIN HAU LETTER X |
| 1184B | 𐌻 | PAU CIN HAU LETTER S |
| 1184C | 𐌼 | PAU CIN HAU LETTER B |
| 1184D | 𐌽 | PAU CIN HAU LETTER C |
| 1184E | 𐌾 | PAU CIN HAU LETTER T |
| 1184F | 𐌿 | PAU CIN HAU LETTER TH |
| 11850 | 𐍀 | PAU CIN HAU LETTER N |
| 11851 | 𐍁 | PAU CIN HAU LETTER PH |
| 11852 | 𐍂 | PAU CIN HAU LETTER R |
| 11853 | 𐍃 | PAU CIN HAU LETTER F |
| 11854 | 𐍄 | PAU CIN HAU LETTER CH |

Vowels

| | | |
|-------|---|-----------------------|
| 11855 | 𐍅 | PAU CIN HAU LETTER A |
| 11856 | 𐍆 | PAU CIN HAU LETTER E |
| 11857 | 𐍇 | PAU CIN HAU LETTER I |
| 11858 | 𐍈 | PAU CIN HAU LETTER O |
| 11859 | 𐍉 | PAU CIN HAU LETTER U |
| 1185A | 𐍊 | PAU CIN HAU LETTER UA |
| 1185B | 𐍋 | PAU CIN HAU LETTER IA |

Final consonants

| | | |
|-------|---|-----------------------------|
| 1185C | 𐍌 | PAU CIN HAU LETTER FINAL P |
| 1185D | 𐍍 | PAU CIN HAU LETTER FINAL K |
| 1185E | 𐍎 | PAU CIN HAU LETTER FINAL Y |
| 1185F | 𐍏 | PAU CIN HAU LETTER FINAL M |
| 11860 | 𐍐 | PAU CIN HAU LETTER FINAL N |
| 11861 | 𐍑 | PAU CIN HAU LETTER FINAL L |
| 11862 | 𐍒 | PAU CIN HAU LETTER FINAL W |
| 11863 | 𐍓 | PAU CIN HAU LETTER FINAL NG |
| 11864 | 𐍔 | PAU CIN HAU LETTER FINAL Y |

Tone marks

| | | |
|-------|---|------------------------------------------|
| 11865 | 𐍕 | PAU CIN HAU RISING TONE LONG |
| 11866 | 𐍖 | PAU CIN HAU RISING TONE SHORT |
| 11867 | 𐍗 | PAU CIN HAU GLOTTAL VARIANT |
| 11868 | 𐍘 | PAU CIN HAU RISING TONE LONG FINAL |
| 11869 | 𐍙 | PAU CIN HAU RISING TONE SHORT FINAL |
| 1186A | 𐍚 | PAU CIN HAU GLOTTAL VARIANT FINAL |
| 1186B | 𐍛 | PAU CIN HAU SANDHI TONE LONG |
| 1186C | 𐍜 | PAU CIN HAU SANDHI TONE SHORT |
| 1186D | 𐍝 | PAU CIN HAU SANDHI TONE LONG FINAL |
| 1186E | 𐍞 | PAU CIN HAU SANDHI TONE SHORT FINAL |
| 1186F | 𐍟 | PAU CIN HAU MID-LEVEL TONE SHORT |
| 11870 | 𐍠 | PAU CIN HAU GLOTTAL STOP OTHER |
| 11871 | 𐍡 | PAU CIN HAU MID-LEVEL TONE LONG |
| 11872 | 𐍢 | PAU CIN HAU MID-LEVEL TONE SHORT |
| 11873 | 𐍣 | PAU CIN HAU LOW-FALLING TONE LONG |
| 11874 | 𐍤 | PAU CIN HAU LOW-FALLING TONE SHORT |
| 11875 | 𐍥 | PAU CIN HAU GLOTTAL STOP |
| 11876 | 𐍦 | PAU CIN HAU LOW-FALLING TONE LONG FINAL |
| 11877 | 𐍧 | PAU CIN HAU LOW-FALLING TONE SHORT FINAL |
| 11878 | 𐍨 | PAU CIN HAU GLOTTAL STOP FINAL |

Figure 1: Proposed code chart and names list for the Pau Cin Hau Alphabet.

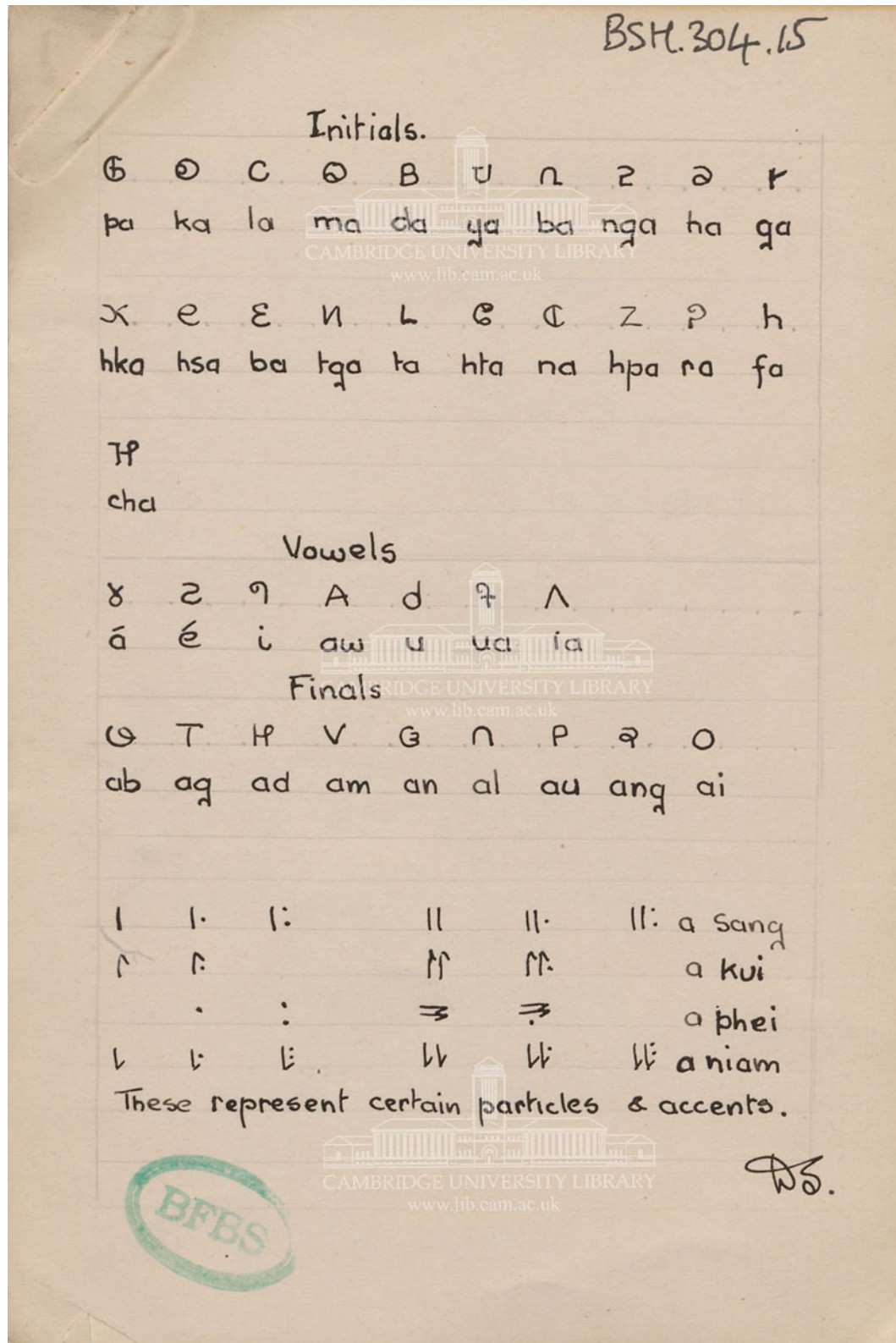


Figure 2: A plate showing the letters of the alphabet with Latin names given by Dorothy Sherrat (from Pau Cin Hau 1932). Tedim names for the tone mark groups are also given.

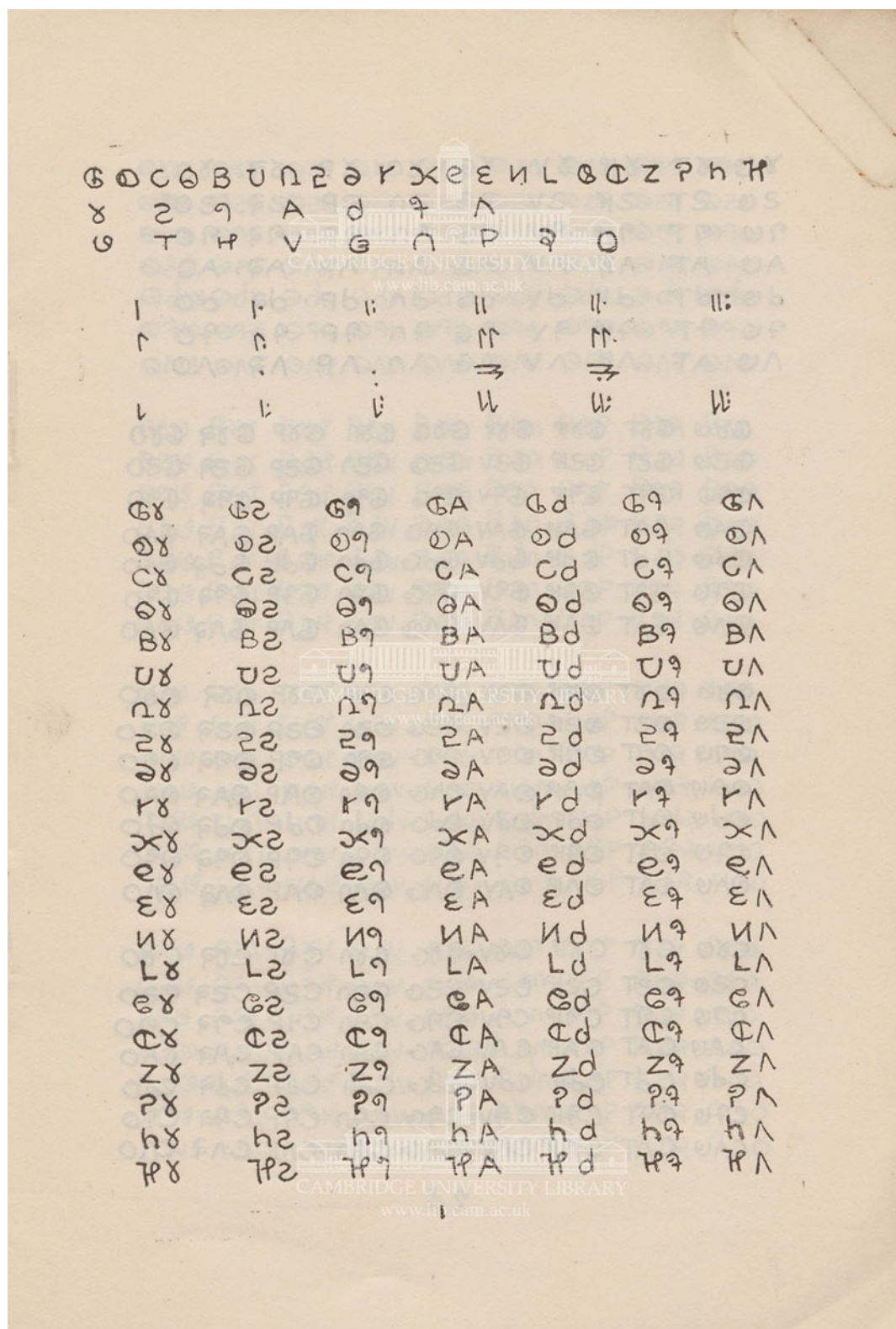


Figure 3: A page from the 'Kamhow-Sokte' primer of Pau Cin Hau (1932: 1). Shown are the letters and tone marks of the alphabet in traditional order. Also shown are consonant-vowel combinations.

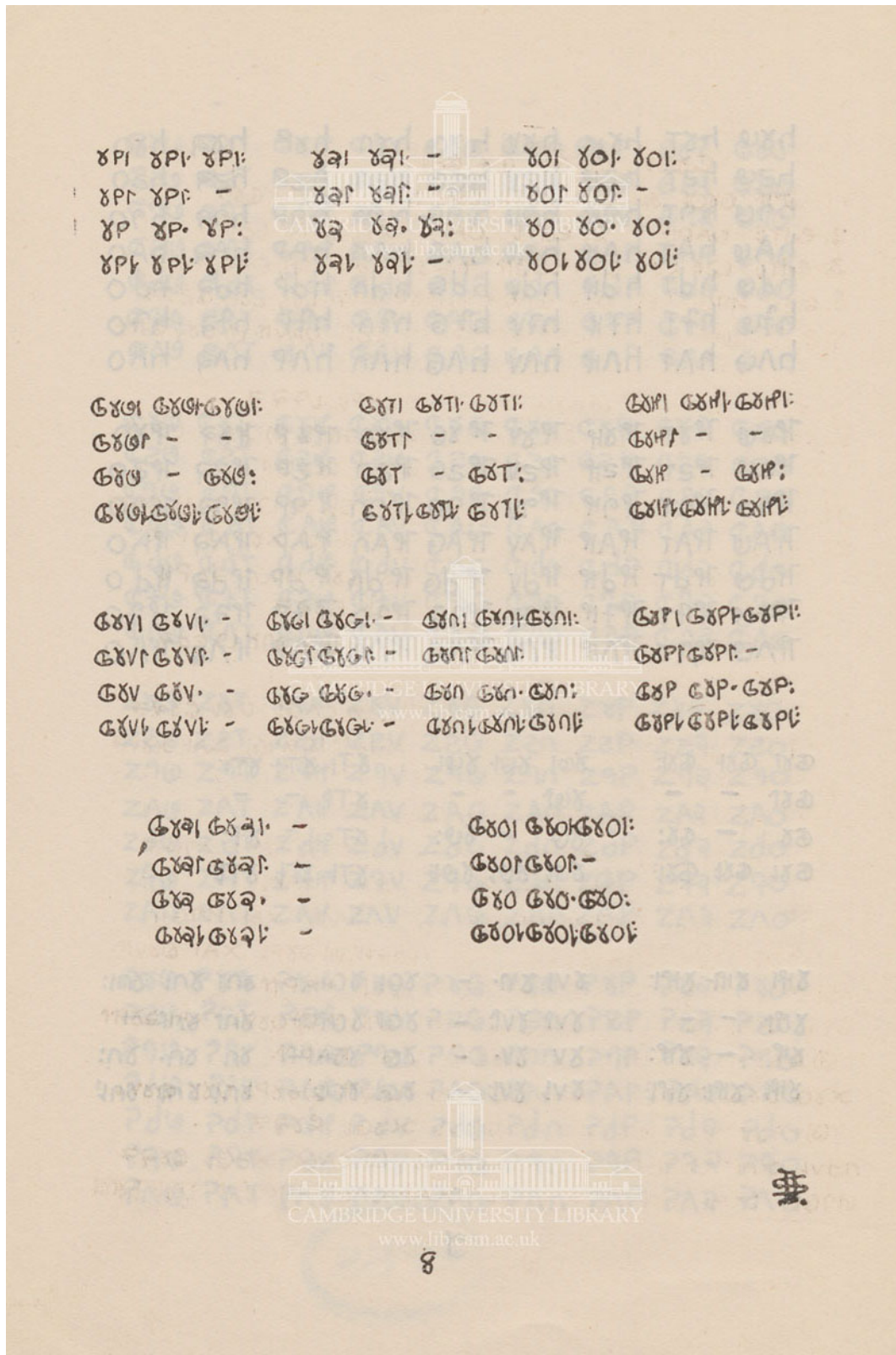


Figure 4: A page from Pau Cin Hau's primer showing tone patterns (1932: 8).

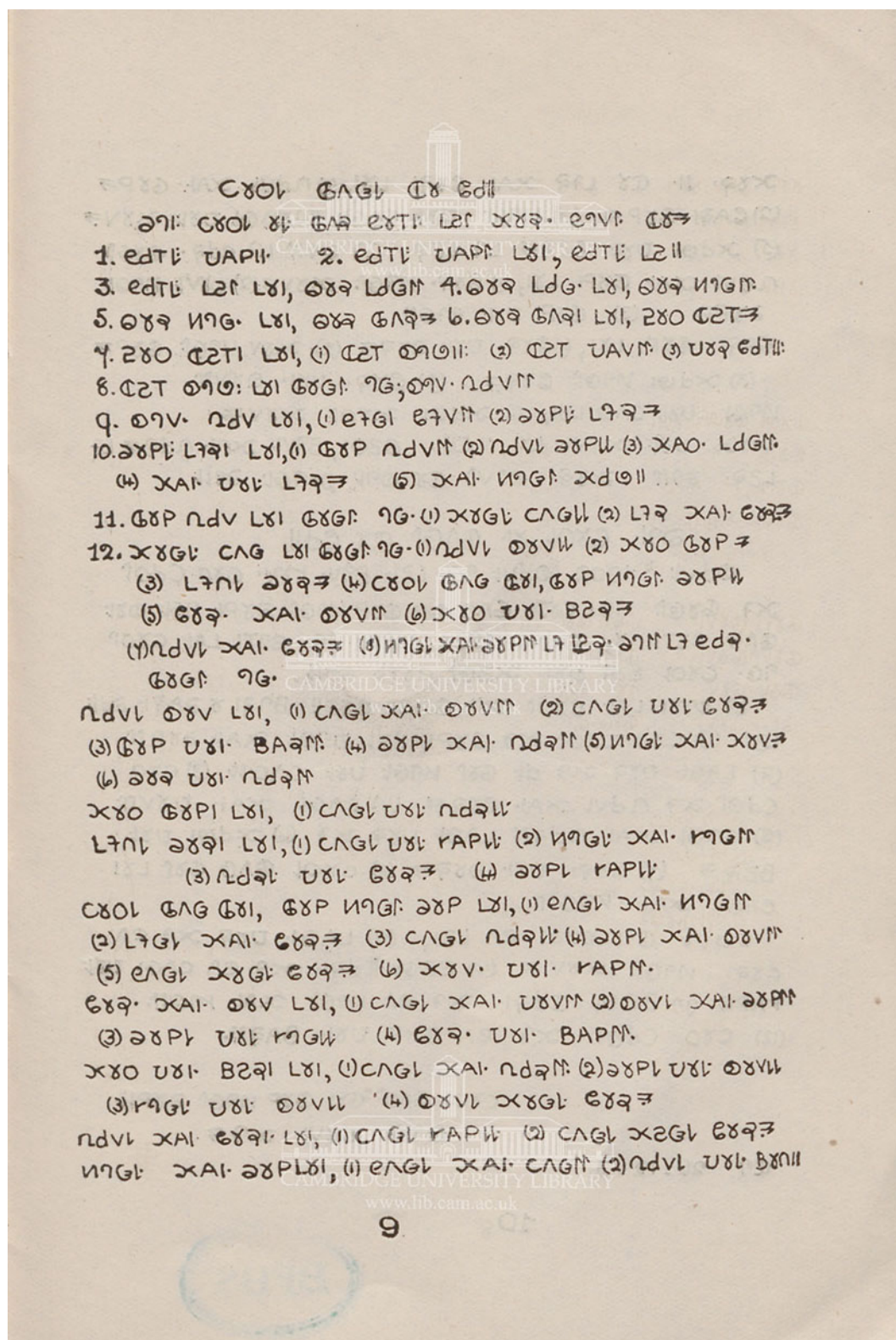


Figure 5: A page from Pau Cin Hau’s primer showing the alphabet used in running text (1932: 9).

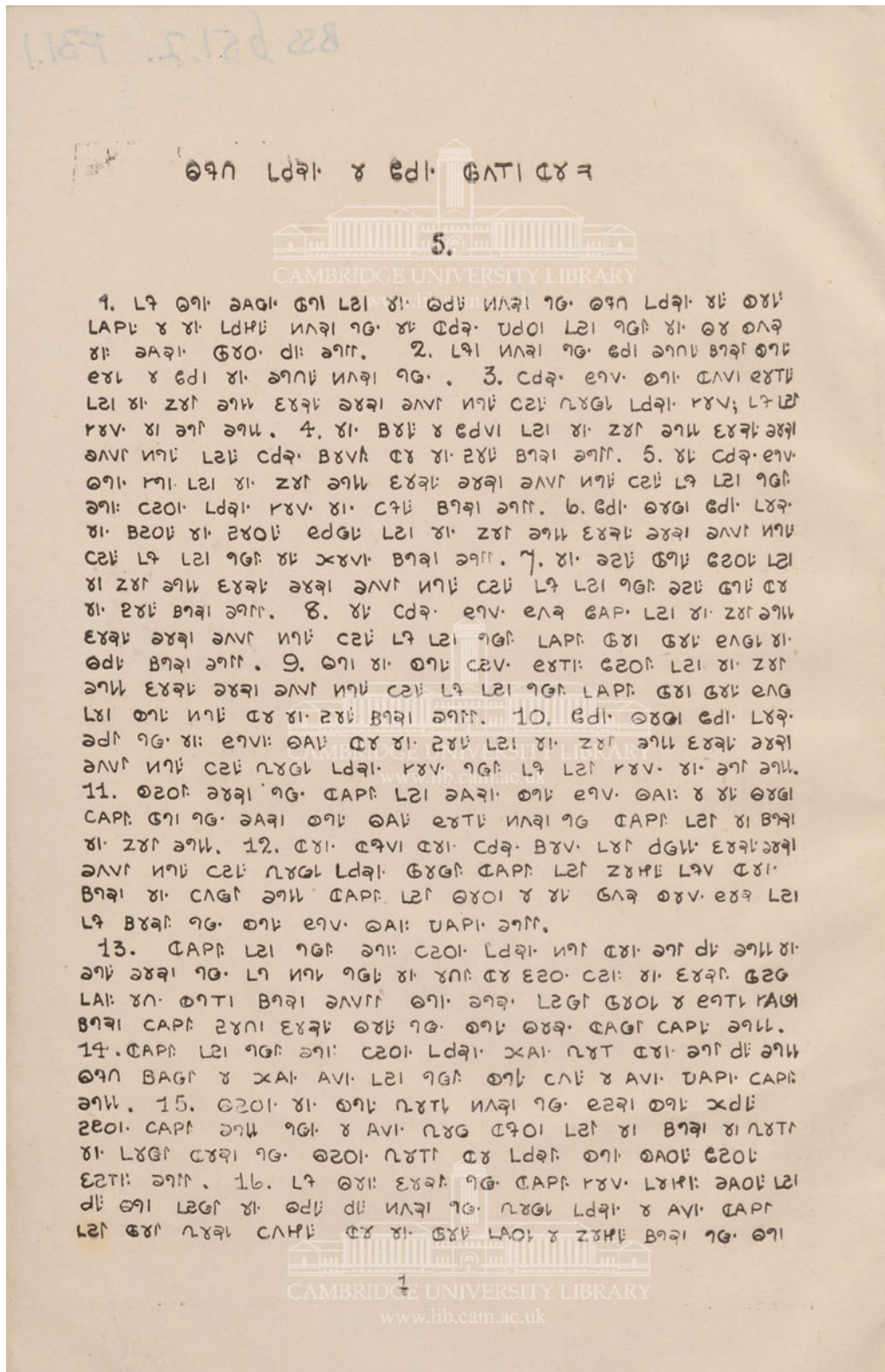


Figure 6: Page of the “Sermon on the Mount” in the Pau Cin Hau alphabet (from British and Foreign Bible Society 1931: 1).

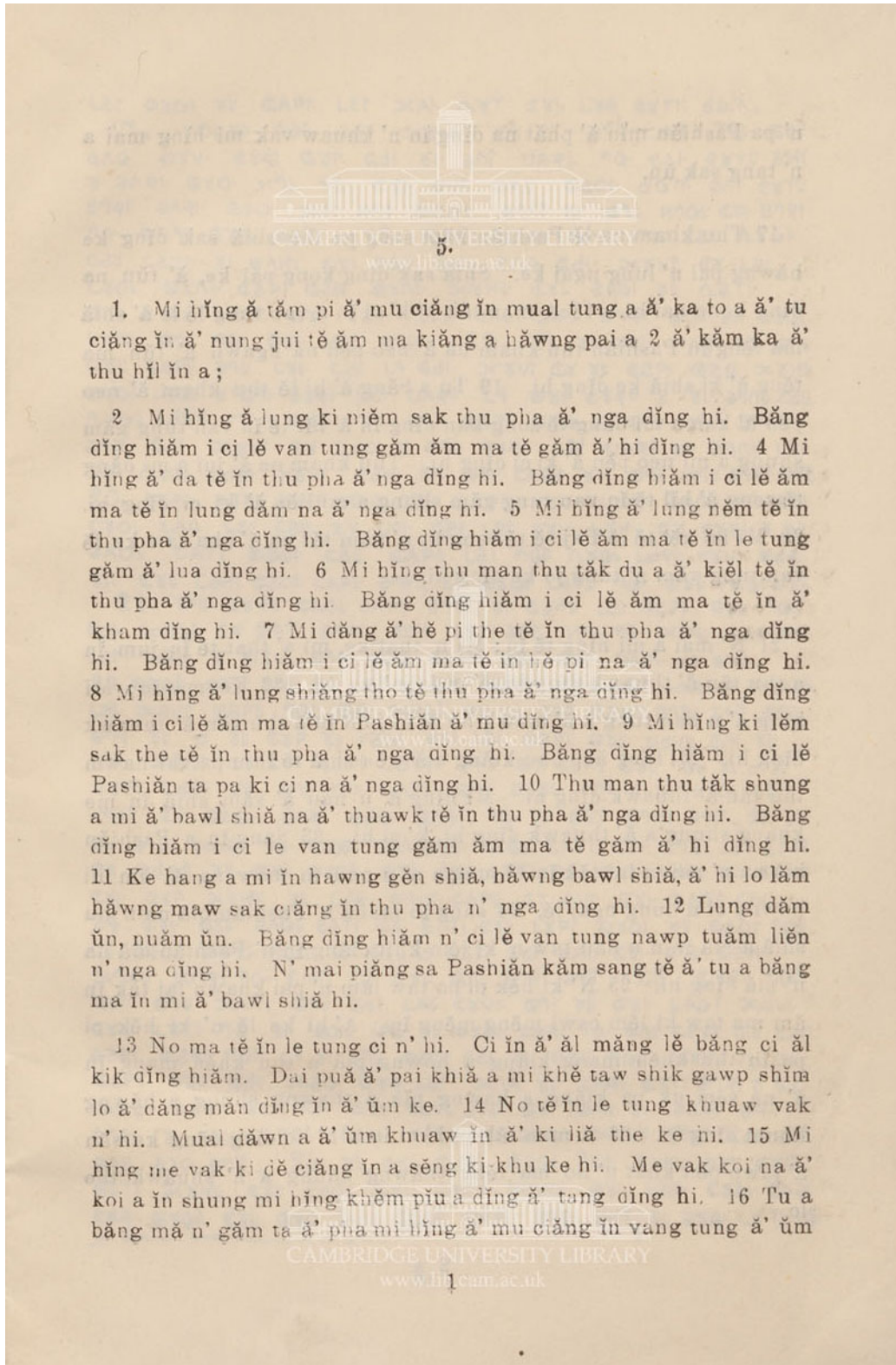


Figure 7: Page of the “Sermon on the Mount” showing Latin transliteration of the Tedim text (from British and Foreign Bible Society 1931: 2).

| THE PAU CHIN HAU SCRIPT. | | | | | | | | | | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|--|--|
| (See paragraph 120). | | | | | | | | | | | |
| G O C O B U N Z a r x e e n l o c z p h k y z 7 A d 4 A u t h v o n p q o | | | | | | | | | | | |
| I F H H H H r n . n n n f e v u w w | | | | | | | | | | | |
| GY GZ G7 GA Gd G9 GA QY QZ Q7 QA Qd Q9 QA CY CZ C7 CA Cd C9 CA OY OZ O7 OA Od O9 OA BY BZ B7 BA Bd B9 BA UY UZ U7 UA Ud U9 UA NY NZ N7 NA Nd N9 NA ZY ZZ Z7 ZA Zd Z9 ZA AY AZ A7 AA Ad A9 AA XY XZ X7 XA Xd X9 XA EY EZ E7 EA Ed E9 EA HY HZ H7 HA Hd H9 HA LY LZ L7 LA Ld L9 LA EY EZ E7 EA Ed E9 EA QY QZ Q7 QA Qd Q9 QA ZY ZZ Z7 ZA Zd Z9 ZA PY PZ P7 PA Pd P9 PA HY HZ H7 HA Hd H9 HA WY WZ W7 WA Wd W9 WA | | | | | | | | | | | |
| pa ka la ma da ya va nga ha ga xa hsa bo tga ta hta na hpa ra la cha à ài i aw u ua ia ab ag ad am an ai au ang a | | | | | | | | | | | |
| } These are tonal signs | | | | | | | | | | | |
| pá yai pi paw pu pua pia ká kai ki kaw ku kua kia lá lai li law lu lua lia má mai mi maw mu mua mia dá dai di daw du dua dia yá ya yi yaw yu yua yia vá vai vi vaw vu vua via ngá ngai ngi ngaw ngu ngua ngia há hai hi haw hu hua hia gá gai gi gaw gu gua gia xá xai xi xaw xu xua xia hsá hsai hsi hsw hsu hsua hsia bá bai bi baw bu bua bia tga tgu tgi tgaw tgu tgu tgia tá tai ti taw tu tua tia hta hta hti htaw htu htua htia ná nai ni naw nu nua nia hpá hpai hpi hpaw hpu hpua hpia rá rai ri raw ru rua ria fà fai fi faw fu fua fia chá chai chi chaw chu chua chia | | | | | | | | | | | |
| 1 | | | | | | | | | | | |

Figure 8: The Pau Cin Hau Alphabet (from Bennison (1931), reprinted in Go (2008: 94)).

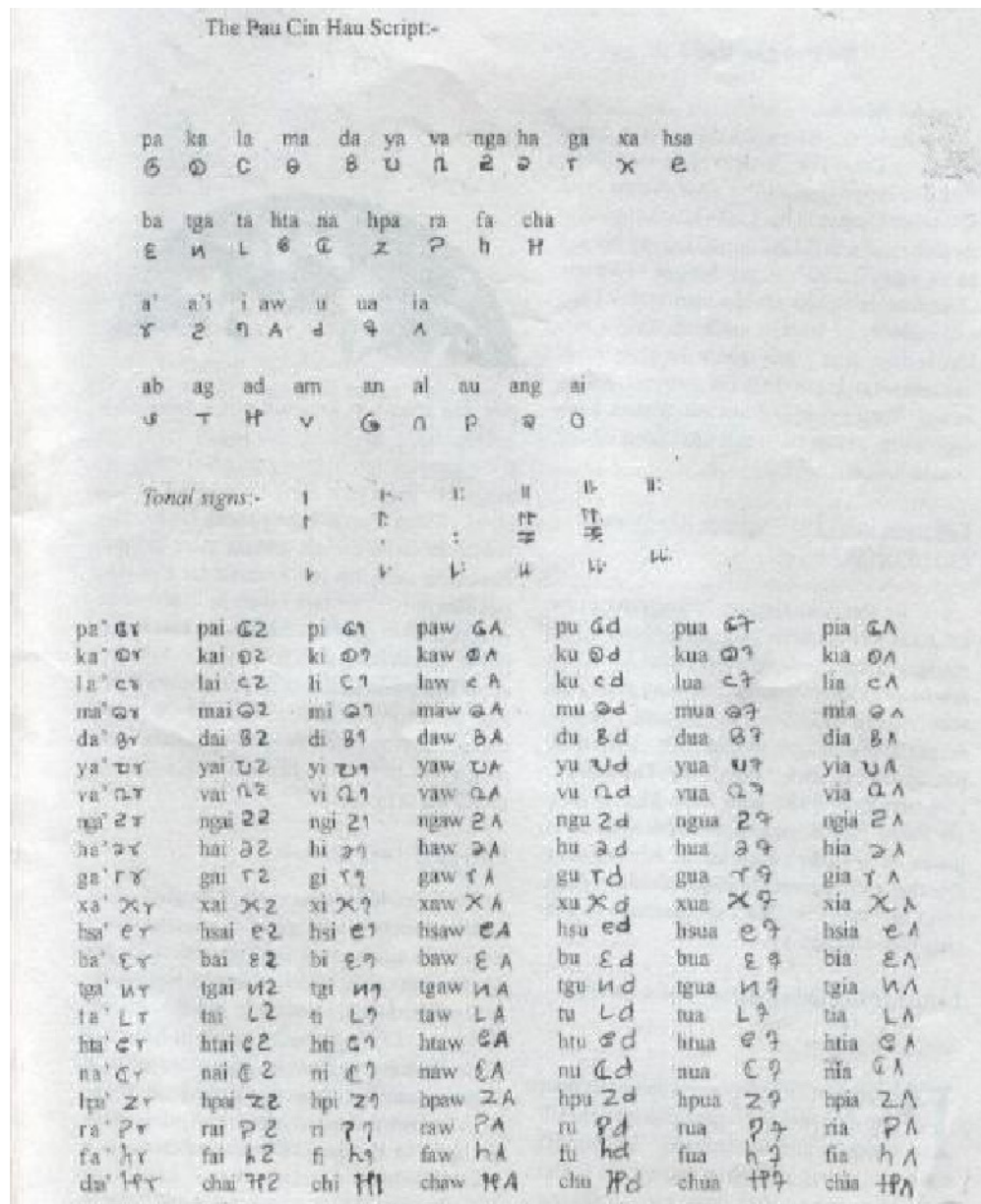


Figure 9: Characters of the Pau Cin Hau Alphabet (from “Pu Pau Cin Hau Lai”: 4).

- 1 -

1. p p^e
p ɿ ʔ^e ɿ ɾ^e ɿ ʔ^e ɿ
 1. p 4 3 e ɿ ʔ^e ɿ ɾ^e ɿ
 ʔ^e ɿ p ɿ ɾ^e ɿ ɿ ɿ
 2. 2 ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 3ɿ ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 3. ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 4. ɿ ʔ^e ɿ ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 5. ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e
 ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e ʔ^e

Figure 10: A page from a Laipian recitation text written in the logographic script of Pau Cin Hau. Courtesy of Christopher Button.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

| | |
|------------------------------------------------------------------|------------------------------------------------------------------------------------|
| 1. Title: | Proposal to Encode the Pau Cin Hau Alphabet in ISO/IEC 10646 |
| 2. Requester's name: | <i>Script Encoding Initiative (SEI) / Anshuman Pandey <pandey@umich.edu></i> |
| 3. Requester type (Member body/Liaison/Individual contribution): | <i>Liaison contribution</i> |
| 4. Submission date: | <i>2011-04-27</i> |
| 5. Requester's reference (if applicable): | |
| 6. Choose one of the following: | |
| This is a complete proposal: | <i>Yes</i> |
| (or) More information will be provided later: | |

B. Technical – General

| | | | | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|-----------------------------------------|-------------------------------------|------------------------------------|--|
| 1. Choose one of the following: | | | | | |
| a. This proposal is for a new script (set of characters): | <i>Yes</i> | | | | |
| Proposed name of script: | <i>Pau Cin Hau Alphabet</i> | | | | |
| b. The proposal is for addition of character(s) to an existing block: | | | | | |
| Name of the existing block: | | | | | |
| 2. Number of characters in proposal: | <i>57</i> | | | | |
| 3. Proposed category (select one from below - see section 2.2 of P&P document): | | | | | |
| A-Contemporary | <input checked="" type="checkbox"/> | B.1-Specialized (small collection) | <input checked="" type="checkbox"/> | 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 | | | |
| 4. Is a repertoire including character names provided? | <i>Yes</i> | | | | |
| a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? | <i>Yes</i> | | | | |
| b. Are the character shapes attached in a legible form suitable for review? | <i>Yes</i> | | | | |
| 5. Fonts related: | | | | | |
| a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? | <i>Anshuman Pandey, author of the proposal</i> | | | | |
| b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): | <i>Anshuman Pandey <pandey@umich.edu></i> | | | | |
| 6. References: | | | | | |
| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | <i>Yes</i> | | | | |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | <i>Yes</i> | | | | |
| 7. Special encoding issues: | | | | | |
| 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)? | <i>Yes, see proposal for additional details</i> | | | | |

8. Additional Information:

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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3702-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11)

C. Technical - Justification

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| 1. Has this proposal for addition of character(s) been submitted before? If YES explain | No |
| 2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents: | Yes <i>Dr. Christopher Button <chris.button@hotmail.com></i> <i>See citations in text of proposal</i> |
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: | Yes <i>See text of proposal</i> |
| 4. The context of use for the proposed characters (type of use; common or rare) Reference: | Rare <i>See text of proposal</i> |
| 5. Are the proposed characters in current use by the user community? If YES, where? Reference: | Yes <i>Limited use by Laipian community in Chin State, Myanmar</i> |
| 6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: | No |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? | Yes |
| 8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: | No |
| 12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) | No |
| 13. Does the proposal contain any Ideographic compatibility character(s)? If YES, is the equivalent corresponding unified ideographic character(s) identified? If YES, reference: | No |